Data-Driven Decision Support for Flood Risk Management

16th April 2018
Flood Sprint
Industry Perspective

• How do we systematically integrate a wide range of data sets and models?

• How do we deal with the different approaches to understanding and communicating risk?

• How do we promote discussions around uncertainty?

• Current software development approaches are rigid
Flood Sprint
Research Questions

• How to bring together heterogeneous data sets

• Tracking data provenance

• Multi-perspective data queries

• Make uncertainty explicit and visible

• Demonstrate agile ways of working
Story-Led Demonstrator
(Used to illustrate and evaluate key research ideas)

Technologies:
- Jupyter
- Shiny App
- Web Site / Web App
- SPARQL (++)
- Linked Data
- RDF
- RDFS / OWL
- R / Python / Fortran
- Operating Context API
- Filesystem storage
- Cloud service
  - AWS S3
  - Azure Storage
- Data microservice

Conceptual Stack:
- User Interface
- Query
- Semantic Enrichment
- Model Service
- Data Storage

Key Benefits to Illustrate:
- Collaborative Managed Sharing
- Same system, single query
  - Multi-Perspective
  - Check-Box Approach
- Flood Risk from Multi-Perspectives
  - Provenance
- Models-on-demand
  - Reproducibility / Reuse
  - Adaptability / Model Selection
- Work across heterogeneous datasets
  - Generate missing data 'on-demand'
  - Describe data requirements using 'complete' ontology
  - Reflective / Intelligent System
Demonstration Use Cases

Exploratory Analysis
- What data do I have/not have?

Multi-Perspectives
- How does national and local knowledge compare?

Risk Analysis
- What is the risk of flooding to a spatial area?
Multiple Perspectives – looking at the data challenges

The 4 V’s of data

Variety
What is an ontology?

An ontology is a formal, explicit specification of a shared conceptualization.

Formally describe a ‘domain’ [5]

“an explicit specification of a conceptualization” [4]

Extensible
Scalable
Flexible

Capture
Store
Reuse

Knowledge

Designing an ontology

Building a Schema on top of the heterogeneous data sets

Perform richer querying across the data sets

Using Ontologies

A river

has

Dyke

subject

predicate

object

Triple
Proposed Framework: Integration of Data Using a Linked Data Model
Ontology Design: **Environmental Impact Assessment (EIA) ontology**
Ontology Design: Environmental Impact Assessment (EIA) ontology

Axiom definitions for the concept “River_Flood”
Data Transformation Process for Structured Data:

(Converting Excel/CSV Files into RDF form)
Data Transformation Process for Unstructured Data:

(Converting PDF Files (Section 19 Report) into RDF form)

- Data Slicing
- Data Conversion to Text
- Text Conversion to SPO

Using Python

Semantic Enrichment of Data

Data Conversion to RDF
- Data Loading
- Data Querying

Using Natural Language Processing
Sliced Tadcaster PDF File Converted into Text File-Object (SPO) using Natural Language Processing.
<table>
<thead>
<tr>
<th>subject</th>
<th>predicate</th>
<th>object</th>
<th>jbarcode</th>
<th>gaugelevels</th>
<th>propertypeuri</th>
<th>properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wharfe</td>
<td>has</td>
<td>GaugeName:Addingham Peak stage: 2.463 Time:07:30 Date:26 Dec Rank- InRecord:5 RecordLength:43 CurrentorPreviousHighest:2.541 Jan 1982</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wharfe</td>
<td>has</td>
<td>GaugeName:Collingham stage:5.246 Time:18:30 Date:26 Dec RankInRecord:1 RecordLength:26 Dec CurrentorPreviousHighest:4.704 Autumn 2000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wharfe</td>
<td>has</td>
<td>GaugeName:Tadcaster stage:4.51 Time:01:00 Date:27 Dec RankInRecord:1 RecordLength:27 CurrentorPrevious-Highest:3.79 Autumn 2000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WHARFE</td>
<td></td>
<td></td>
<td>4023</td>
<td>damage:Gau</td>
<td>damage:Q50g</td>
<td></td>
</tr>
<tr>
<td>WHARFE</td>
<td></td>
<td></td>
<td>4023</td>
<td>damage:Gau</td>
<td>damage:Q75g</td>
<td></td>
</tr>
<tr>
<td>WHARFE</td>
<td></td>
<td></td>
<td>4023</td>
<td>damage:Gau</td>
<td>damage:Q100g</td>
<td></td>
</tr>
<tr>
<td>WHARFE</td>
<td></td>
<td></td>
<td>4023</td>
<td>damage:Gau</td>
<td>damage:Q200g</td>
<td></td>
</tr>
<tr>
<td>WHARFE</td>
<td></td>
<td></td>
<td>4023</td>
<td>damage:Gau</td>
<td>damage:Q1000g</td>
<td></td>
</tr>
<tr>
<td>WHARFE</td>
<td></td>
<td></td>
<td>1</td>
<td>damage:Gau</td>
<td>damage:Q5g</td>
<td></td>
</tr>
<tr>
<td>WHARFE</td>
<td></td>
<td></td>
<td>1</td>
<td>damage:Gau</td>
<td>damage:Q10g</td>
<td></td>
</tr>
<tr>
<td>WHARFE</td>
<td></td>
<td></td>
<td>1</td>
<td>damage:Gau</td>
<td>damage:Q20g</td>
<td></td>
</tr>
</tbody>
</table>

Information from Section 19 Report

Information from 2 different Excel spreadsheets
Risk Analysis

Motivation

• Provide more information than just a single number

• Ability to have flexible queries

• Expose any inconsistencies and assumptions in the risk modelling

• Want to explicitly show the uncertainty around flood risk management decisions

• Show information about potential flood defence schemes
Flexible Querying of Expected Annual Damage (EAD)

- Expensive hydraulic model
- Run for a selected number of design events
- Corresponding damage calculated for these events
- Use this information to calculate expected annual damage
Flexible Querying of Expected Annual Damage Estimates

- Onset of Flooding
- Truncation Error
- New Event Information
- Uncertainty
- Cost Benefit Analysis

![Graph showing damage versus return period]

- Optimistic
- Midpoint
- Pessimistic
Onset of Flooding – Risk in Newark
Example Jupyter Output

Expected Annual Damage

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Optimistic</th>
<th>Midpoint</th>
<th>Pessimistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current Defence Scheme</td>
<td>£2,990,942.12</td>
<td>£3,312,795.07</td>
<td>£3,751,284.98</td>
</tr>
<tr>
<td>50 Year Defence Scheme</td>
<td>£1,892,017.33</td>
<td>£2,022,983.93</td>
<td>£2,203,772.42</td>
</tr>
<tr>
<td>100 Year Defence Scheme</td>
<td>£1,605,100.95</td>
<td>£1,721,811.84</td>
<td>£1,887,153.00</td>
</tr>
<tr>
<td>200 Year Defence Scheme</td>
<td>£1,332,623.57</td>
<td>£1,409,444.48</td>
<td>£1,497,270.56</td>
</tr>
<tr>
<td>500 Year Defence Scheme</td>
<td>£737,728.62</td>
<td>£762,208.60</td>
<td>£767,279.36</td>
</tr>
</tbody>
</table>
Biodiversity

Research Areas

• Relating large scale weather to biodiversity
• Software adaptability and evolution
• Data-centric view of a specific place
• Exposing the uncertainty related to measuring biodiversity