



The  
University  
Of  
Sheffield.



HR Wallingford  
*Working with water*

## EnviLOD Work Package 2—User Engagement and Case Studies

Johanna Kieniewicz (The British Library; johanna.kieniewicz@bl.uk)

Michael Wallis (HR Wallingford; M.Wallis@hrwallingford.com)

### Introduction

Environmental Science is a broad interdisciplinary subject area that spans biology, chemistry, earth sciences, physics and engineering. Because of the breadth of the subject scope, information discovery and sharing in environmental science is often a challenge, as it lacks a standardised resource description framework, such as the MeSH headings in biomedical sciences. Linked Open Data and vocabularies offer an opportunity to improve the process of information discovery and sharing through unique, machine readable interlinked open vocabularies, ultimately connecting users more efficiently to useful and relevant sources. The overall aim of the EnviLOD project is to demonstrate the value of using Linked Open Data vocabularies to the field of environmental science. In order for this value to be effectively demonstrated, it is necessary to understand not just what the vocabularies are capable of doing for information discovery in environmental science, but understanding whether those capabilities match the needs of users of environmental information. To that end, Work Package 2 represents an effort to develop detailed requirements from a user perspective.

In this context, a ‘user’ is defined as a user of flooding information, whether a researcher or practitioner. The flooding community stretches across a wide variety of sectors, including academia, consultants, internal drainage boards, government agencies, local government, and charities. Naturally, the roles of individuals in these sectors vary substantially, and as such their information needs are different, as may be the ways in which they discover information. In understanding the applicability of Linked Open Data vocabularies to the discovery of flooding information, it is necessary to understand the vocabularies that the different parts of the flooding community use when searching for information. The aim was also to understand what commonalities exist in the needs of information users across the flooding community in order to determine where support might be most effectively targeted.

Thus it was established that from a technical perspective, it is important to understand what users search for, how they phrase their queries, and their expectations on the content retrieved in order to demonstrate the utility of vocabularies to the community.

## **Methods**

Although it was initially proposed that face-to-face interviews would be used for capturing user requirements, the decision was made to use a survey, as this could allow us to gather responses from a wider range of individuals. A survey was sent out to 20 contacts within the flooding community and a link was also placed on the FlowNet Knowledge Hub website for practitioners in local authorities. Respondents were asked:

- To identify their sector
- Their role within their organisation
- An example of a flooding related topic that they had recently searched for
- The precise terminology they used in their query
- What they would expect in terms of results
- Ways in which they search for information (ex- Keyword searching, Boolean etc).

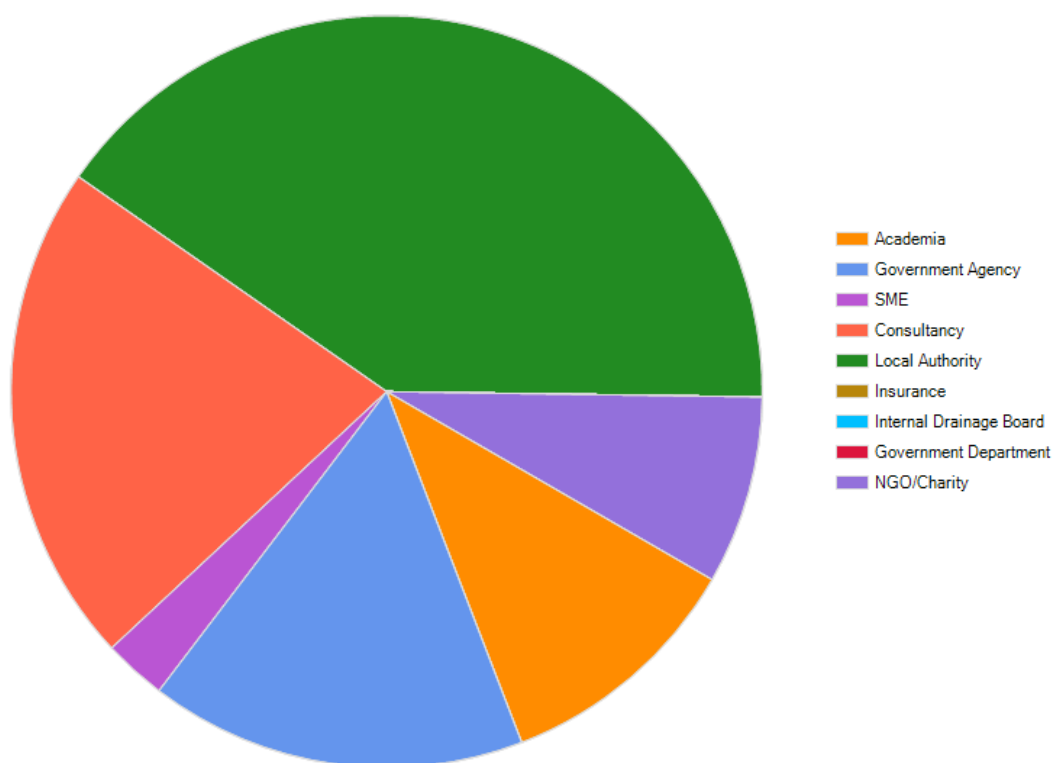
Responses were broken down by sector, and commonalities and differences in responses were identified.

## **Results**

A total of 36 respondents returned the survey.

The respondents of the survey were well distributed across the flooding sector, with a slight emphasis on local authorities (likely due to the survey being posted on the FlowNet Knowledge Hub website) and no returns were received from the insurance industry (Figure 1).

What is your sector? Please tick those that apply to you.



**Figure 1:** Distribution of respondents by sector.

Search queries from survey respondents were broken down by sector, and then categorised by query type. Seven broad categories of query emerged: respondents were interested in spending (consultants, government agency), regulations (cross-sector interest), general information/data (cross-sector interest), flood risk (local authorities), flood defences (cross-sectoral interest), and specific flooding events (consultants and local government). Across all of these categories, location emerged as a universal theme (please see Appendix).

In searching for information, 92% of respondents reported using keywords; 75% reported using complete phrases (although only 25% reported entering questions into search engines). Use of Boolean operators, such as AND/OR/NOT was low (25%), as was the use of faceted navigation (8%).

### Interpretation of Survey Results and Development of Use Cases

The survey returned results that encompass a representative sample of roles, responsibilities, areas of expertise and topics related to flood risk management. The query topics can be classified in various ways but a useful typology was derived which includes:

- Flood Risk Management regulation & policy,
- Flood Management Research,
- Flood Defence (management and practice),
- Finance & spending

- Flood Risk (general -including specific flooding events)
- Other general information/data

Some caution should be placed on the interpretation of the distribution of respondents by sector. This distribution is a function of the surveyed population rather than actual sector size – although it does reflect the experience of the authors somewhat. The lack of responses from the insurance sector and the Internal Drainage Boards reflects the difficulty encountered in finding potential respondents to the survey rather than the lack of any interest, and these should not be construed as ‘no-use’ sectors.

Broadly the queries came from two main types of user:

- ‘Higher ‘ level users – consultant (researcher)/academia/government agency - seeking information on funding, policy, budgets, guidance, etc.
- ‘Local’ level users - local authority/government agency/Internal Drainage Board/consultant (practitioner) looking for information more related to local issues and practical implementation – assessment, areas at risk, protection measures, historic information on past events, imagery, monitoring technology, etc.

#### **Implications for vocabularies:**

The types of queries returned (as would be expected) imply different user perspectives and requirements. The survey responses show a wide range of query types from the specific / detailed to the more generalist 'interest' type query. These can come from any 'user type' although the specific/detailed type queries are more likely (but not exclusively) to come from FRM specialists/practitioners, and the generalist query is more likely to arise (although again not necessarily exclusively) from the lay-person/general public user. This reinforces the fact that the linked ontology's will be key to linking such queries with relevant data/information sources.

The ‘local level’ information can be typically related closely to a location and one might expect useful information pertaining to the enquiry to be returned from a ‘location’ oriented search by place, town, region, locality etc. Flood events obviously occur in specific areas, places and times and therefore would fit this type of query well if a search schema included ‘location’ as one of the tags.

Using a ‘location’ oriented search for information pertaining to the former group however may be less effective in returning precise results or in locating documents or information that may be relevant to the query. Using a ‘location’ oriented search function is also problematic if the user does not specify their location of interest in the search query.

A technical query – for example –‘engineering aspects of earth embankments’ (i.e. guidance) would potentially not have a location specific ‘key’ as such for a search operation to find – it is a general topic of no particular geographical ‘abode’ - hence a different type of vocabulary search construct (or a modification) would possibly be required for a search on such information. However such searches for technical information are rarely made ‘cold’ – i.e. typical (i.e. technical in this case) users often know that the information exists and often where it originates from (publisher/country,

etc) – so a location tag could still be sourced from such information. It would however possibly occur lower down in any hierarchical search schema for such information.

The queries returned also highlight the variety of data sources that are inferred. Some are technical questions - the answers to which would come from journal papers/conference proceedings or specialist science or engineering text books for example, others one might find amongst local authority or national government reports or amongst public information on their websites. Some queries, although related to the Flood Risk Management arena, are not exclusive to it - for example there is a query about the resolution of satellite imagery – one can imagine why the question is being asked in this context, but the answer could easily come from a source which has nothing at all to do with flooding or flood risk management.

The predominance of the use of keywords and phrases by users (as opposed to Boolean operators or faceted navigation) in search engines also infers something about the users (at least from this sample) – essentially that any search engine design should be focussed on keyword and phraseological instigation.

### **Use Cases**

The implications of the survey for use cases suggest that for the purposes of this project, testing of the search tool(s) could be usefully targeted at the main sectors including Local Authorities, Consultants and Government Agencies. These sectors contain both higher level and local level users and their queries are likely to largely reflect those from other less well represented sectors as queries from these 3 sectors featured in all of the query topic areas.

### **Summary**

The EnviLOD user survey identified the following potential use cases for vocabularies to be tested in Work Package 3.

- Returning results for geographically specific queries. Beyond keyword recognition, this use case includes proximity and recognition of geographic entities that are implied, but not stated within the query (for example in a query for flooding in SW England, identifying towns such as Exeter within that region, without it being explicitly articulated in the query).
- Answering non-open-ended queries. In this case, the user is asking for a specific piece of information, which might pertain to a budget, specific piece of legislation, flood levels in a particular locality, etc. For example: What is the annual flood defence expenditure in The Netherlands? These are questions that can be definitively answered, and to which semantic search algorithms can likely be easily trained.
- Answering open-ended queries. In this case, a user is conducting research with an aim of learning more about a particular topic. In this case, there is no definitive 'answer' to the query—the question is answered once the user has established that s/he has sufficient information on the topic. For example: What are some examples of community engagement relating to flood risk management? These questions are likely more difficult for a LOD

approach to add value—but nevertheless represent an important type of question asked by survey respondents.

In general, users were found to prefer Google-style keyword searches, or searches in which they could pose a question above other types of searches. The amount of subject specific jargon used in their queries depended on the nature of the question that was asked, as well as the job held by the individual who was asking it. As such, the LOD vocabularies used in this work need to be flexible, enabling generalist queries, while also allowing subject-specific queries, where possible.

## Appendix

Below, we present the responses to the question in the EnviLOD survey: **Please think back to a time you recently searched for information on a particular topic, or a topic on which you need to search for information. What question were you trying to answer?** The aim of this question was to get a broad sense of the types of questions members of the flooding community have in mind, when conducting a search for information. Responses to this question are broken down by sector and by topic of the query.

Blue- Local Authorities

Green- Government Agency

Red- Consultants

Black- Academic

## Spending

Flood Defence Budgets (Google)

Flood Defence Spending Policy (google)

How is flood defence spending prioritised in non UK countries?

What is the annual expenditure on flood risk management in The Netherlands?

## Regulations

How is flood management organised and regulated in Northern Ireland?

flood defence prioritisation -UK -Defra (Google)

How are flood defences assessed?

What are the Insurance industry expecting in terms of flood protection

What standards are the Water Companies supposed to be working to.

## General Information/Data

Contact details at the local authority

Information in local paper

Information in new civil engineer

Where can I obtain digital surface models of flood defences?

What digital images and LiDAR data are in the public domain covering the **area of our interest**?

How has ground penetrating radar been used for embankment surveys?

What resolution of imagery is available from satellites?

Where can I find photographs of flooding in **Boscastle**?

All the above with **country name** preceding the query term

Availability of data for undertaking flood studies in different countries (eg hydrometric data, Poland)

Data that can be used to assess the magnitude of a flood (eg peak level, peak flow and flood volume for a fluvial flood at Carlisle; peak tidal water levels for a tidal flood in Essex)

### **Flood Risk**

How many businesses are at risk of flooding in **Oxfordshire**?

How many people in the county have signed up to flooding warnings from the Environment Agency

Examples of community engagement relating to flood risk management

The top ten flood risk areas in **Oxfordshire**

### **Flood Management (Engineering perspective)**

What case studies are there where natural flood management techniques have been applied?

Property level protection measures

Engineering aspects of earth embankments

**Location** and levels of flood defences in Edinburgh

**Locations** where property level flood defences have been used during a flood.

### **Information on specific flooding events**

**Local authority area** that flooding took place



Location date and flooding

What is publically available information on SW flooding.

Where are the Main Rivers?

Where has surface water flooding taken place since 2007

Are there critical infrastructures that have previously been flooded

When was there flooding of the River Axe at Chard Junction?

Is there any history of flooding in Radlett?

Was there flooding in Warminster in 2007 and if so what was the cause?

Where is the floodplain near Aylesbury?

What is the extent of the floodplain in Radlett?

Information on historic floods at specific locations (eg Cork, Ireland).