How to optimally employ citizen scientists to improve water quality?

Stream. www.streamwaterdata.co.uk

Stream is a collaboration between UK water companies, supported by the Open Data Institute, industry and civil society partners with a vision to unlock the potential of water data to benefit customers, society, and the environment.

The question

Across the Open Data Institute and industry and civil society partners, there is a wealth of information about the health of our waterways, but it is distributed, fragmented and often inaccessible, constraining its full potential. Water companies, regulators and others are keen to use this data more smartly to identify optimal actions to improve water quality.

Detailed, distributed, real-time data collection is expensive and often too late. For example, following a storm, rainwater run-off from roads or urban spaces contains many contaminants that make it into our rivers and ultimately the sea. But the time course of the consequent effect of this pollution on river health is hard to measure without massive, time-consuming and expensive data collection campaigns.

Data from citizen science initiatives could complement official data collection from water companies, regulators and more. In taking this approach, there is an opportunity to create a more holistic understanding of waterways in the UK and supporting better decision making. Stream is working with charities such as Stormwater Shepherds who want to empower citizen scientists help. Currently however, there are barriers to this happening in a routine and systematic way.

- The data collected by citizen scientists is currently underutilised because it is of unknown quality and can't be easily accessed or joined up because it is not standardised and therefore is captured in a range of different formats and collection methods.
- There is no easy way for citizen scientists to identify what to sample - where, how, and when - in order to achieve maximum impact. In addition, there is no openly available way to see all potential pressures and risks displayed in an easy-to-use format to help inform citizen science planning.

The problem to be tacked by the Study Group is specifically targeted at the second of these barriers. Can we understand the spatiotemporal processes involved in the degradation of water quality following different types of contaminations? What are the different signatures of pollution we would expect to see following, for example following a wastewater pipe overflow or leak? or run-off from roads, the urban environment or

from agriculture? How long does it take for such pollution to flux through our river systems? and where does it end up?

Ultimately, we want to understand

1. Can we solve the inverse problem of a water contamination event to understand from the collection of water quality data what event happened and where, in order to suggest to water companies and regulators what remedial action should be taken?
2. Can we direct citizen scientists to the optimal places and times to collect data to have the best chance of solving these inverse problems?

Previous work on this problem

Working with the Open Data Institute, Stream have collated a wide number of open-source datasets of real-time and historic water quality information across the UK. Drawing inference from these data was a subject investigated during of Open Data Week in March 2025. As part of that week's events, in collaboration with Stream, VKEMS ran a small Virtual Study Group (VSG). There, some first steps were taken at simple mathematical models of road pollution run off and of river system analysis. The findings of that VSG will be made available to the Study Group. In the week before the Study Group, Stream will be running a hackathon and design sprint in collaboration with Northumbrian Water and other organisations, to better understand how to curate, standardise, and make water quality data accessible to all via simple APIs. These findings will also be made available.