Stream.® ODI

Improving our data flows: an open data strategy for the water industry

The companies that have currently signed up to this strategy are:

Contents

Background

There is a growing movement towards openness in the UK water industry and progress is already being made both within individual companies and at a sector level. This comes proactively from many of the water companies themselves, including those who are members of the Stream initiative and from Ofwat, who are actively encouraging the sector to accelerate towards openness, and whose recommendations this strategy responds to.

The water sector is facing unprecedented challenges and we are committed to building trust in response to demands for greater transparency from the public and civic society. Companies are also facing increasingly complex and novel issues, many of which are driven by volatile and uncertain factors such as climate change. Coupled with increasing demand for higher standards of service at the same time as lower bills, the need for increased and faster levels of innovation and adaptability have never been higher.

It is also becoming increasingly clear that water companies cannot solve this challenge alone. The approaches and interventions that led to improved performance since privatisation are no longer sufficient on their own to address the new era challenges facing the sector. The sector needs to increase transparency and collaboration, in a systems-focused way, leveraging broader pools of innovative talent and data to help accelerate the speed at which the sector can adapt and keep pace with the challenges ahead. Opening sector data is a key enabler of this.

To go far, fast and efficiently as a sector, we recognise that we need to go together.

Our initial steps together were laying out our plans in the very first Water Sector Open Data Roadmap. The roadmap outlines the activities and outputs that we believe need collective action and adoption by the water sector to maximise the value generated, while at the same time minimising any risks and harms from open data publication.

One of our promises was to create and publish an overarching open data strategy for the water sector to articulate the collective vision and mission for the publication and use of open data, and to help align the open data strategies of each water company.

This is that strategy. It is our vision for the future of open data in the water sector, the priorities we intend to focus on first, and the commitments we make as an industry to be as responsible stewards of data as we are stewards of national infrastructure and the natural environment.

"This open data strategy for the water sector is an industry-first, and sets out an important path for water companies to take a more open approach to the valuable data that they collect and hold. We support these steps to drive innovation and transparency through publishing open data, and believe that done properly, this will help build trust in the sector, and deliver value to customers, communities and the environment."

Louise Burke, CEO, Open Data Institute

Vision

Our vision is simple:

We aspire to build an open and trustworthy data ecosystem that supports achieving our strategic outcomes for our customers, communities, industry and the environment.

We see the following themes of activity being crucial to doing so, and delivering the most value from open data.

We seek to publish with purpose.

For open data publishing to truly drive value and benefits to our customers, it is best when it is focussed on what is needed by the communities we serve, and by wider regional, sectoral and national priorities, such as helping with the cost of living and protecting the environment. We don't see open data as a compliance exercise or a nice-to-have, but a way to improve engagement, innovation and trust.

We seek to build trust and demonstrate trustworthiness.

At the core of our vision is leveraging the opportunity for open data to improve transparency in the industry. This includes our performance as a business and crucially our performance as stewards of the environment. Without better transparency, we can never hope to rebuild trust in the industry, and demonstrate the ways we are improving as individual organisations and as a sector. But we must also go beyond showing and telling, and equally have a commitment to listening.

We seek to be highly engaged in stakeholder communication and collaboration.

The voice of the customer is critical to our work, and we can only ensure that we are delivering benefits through robust engagement with the rest of society - our customers chief among those stakeholder groups. To this end we will work with communities, local authorities, and civil society to support the social needs and ambitions of the areas where we operate. We will also work with businesses, innovators and our supply chain to be part of the commercial backbone of our regional and national economies.

We seek to build an open culture alongside our data culture.

In recent years we have invested more and more in a variety of different data skills that allow us to collect, analyse and use data to improve our service delivery. We now want to take that further and improve our data literacy around the value of sharing data and publishing data openly, from daily data practitioners, up to our senior executive teams.

Just as importantly we need to understand the risks of doing so, and what mitigations we can take to make data as open as possible while preserving privacy, protecting sensitive data and safeguarding against the misuse of data. We will embed data ethics into our data culture to better understand the potential consequences to our customers of opening, or not opening, the data that we hold.

We seek to create a future where open data is business-as-usual.

Eventually, our vision is for open data to be just another part of daily business life, like data analytics or customer outreach. Not everyone will be an open data expert, but it should be known in relevant parts of the business, with teams engaged as needed to regularly publish new and update existing data. 'Open data' as a concept shouldn't be our focus, but the value it delivers to customers, communities, industry and the environment should be well known and documented.

Principles

In 2016, the 'FAIR Guiding Principles for scientific data management and stewardship' were published in Scientific Data. The authors intended to provide guidelines to improve the **F**indability, **A**ccessibility, **I**nteroperability, and **R**eusability of digital assets. FAIR data represents a best practice for data usability globally.

As part of the Water Sector Open Data Roadmap, the sector is keen to go beyond the FAIR Principles, to ensure that data being published is both **O**pen and **E**thical. Therefore, the FAIR Principles have been expanded upon to the FAIROE Principles which all water companies will adhere to.

The following steps are taken directly from the FAIR Principles, while 'Open' aligns with The Open Definition from the Open Knowledge Foundation and World Wide Web Consortium (W3C) standards, and 'Ethical' aligns with best practice from the Open Data Institute (ODI):

Findable

To use data it must first be found. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services.

F1. Data is assigned a globally unique and persistent identifier.

F2. Data is described with rich metadata.

F3. Metadata clearly and explicitly include the identifier of the data they describe.

F4. Data is registered or indexed in a searchable resource.

Accessible

Once the user finds the required data, they need to know how it can be accessed.

A1. Data is retrievable by their identifier using a standardised communications protocol.

A1.1 The protocol is open, free, and universally implementable.

A1.2 The protocol allows for an authentication and authorisation procedure, where necessary.

A2. Metadata is accessible, even when the data are no longer available.

Interoperable

Data is more valuable when combined with other data. In addition, the data needs to interoperate with applications or workflows for analysis, storage, and processing.

I1. Data uses a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. Data uses vocabularies that follow FAIR principles.

I3. Data includes qualified references to other data.

Reusable

The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/ or combined in different settings.

R1. Data is richly described with a plurality of accurate and relevant attributes.

R1.1. Data is released with a clear and accessible data usage licence.

R1.2. Data is associated with detailed provenance.

R1.3. Data meets domain-relevant community standards.

Open

Data will be published according to the ODI open data definition - to be accessed, used and shared by anyone, at a three-star level of linked open data:

O1. Data is published on the web under a licence conformant with the open definition.

O2. Data is published as machine-readable structured data.

O3. Data is published in a non-proprietary format.

Ethical

All data will be published in accordance with relevant laws. Beyond this, the positive and negative impact of collecting, using and sharing data, and relevant mitigations, will be assessed throughout all stages of the data lifecycle, with findings transparently and openly published.

E1. Data is published alongside a risk register that captures the risks and mitigations of publishing the dataset.

E2. Mitigations against risks are clearly implemented, and proof of this is easily accessible to data users.

Priorities

We understand that we are at an early stage in our open data journey as a sector, and there is much to do to achieve our vision. We therefore believe we need to prioritise the following areas to move towards our vision as efficiently and effectively as possible.

Publishing data to support use cases that are high value to our customers.

Through engagement with our customers and other key stakeholders we will learn what they need the most from our open data activities and prioritise publishing the data to deliver that, starting with reviewing the underlying data supporting existing public information. This approach will drive customer-oriented innovation and policymaking, help build trust in our activities, and be done in a collaborative way across the sector to benefit communities across the entire UK. We have already begun this process, publishing data about regional water quality and storm overflows, two topics important to our customers, society and the environment.

Supporting collaborative data initiatives.

Our sector is not new to collaboration, but with the ability to properly leverage open data to help tackle society's problems, we need to build and maintain the right data infrastructure. Through Stream and the National Environment Hub we are publishing a number of datasets that support use cases to improve customer and environmental health. But collaborative infrastructure is about more than just publishing datasets, we are also adopting and setting data standards for greater interoperability and adopting policies, guidance and strategies to keep our efforts aligned.

Aligning our strategic approaches to open data and data sharing.

Each water company faces their own challenges and has their own contexts, but there are wider systemic issues and opportunities that unite us. This sector-wide open data strategy will help strengthen our individual company open data strategies and other digital strategies that we publish, by affirming our collective support and vision to build an open and trustworthy data ecosystem. By collaborating at this strategic level we will work together to tackle large social issues such as the cost of living and climate crises, and seize the opportunities of the latest developments in net zero and digital technologies, such as AI.

Embedding, defining and sharing best practices within and across industries.

Our sector is not the first to undertake open data, allowing us to learn from others, such as government and the energy sector. From them we are adopting and adapting best practices around assessing and mitigating risks, improving reusability and data ethics. By working as a sector we are sharing with each other, innovating together, and building off of and improving on existing processes such as conducting open data triage and assessing open data maturity.

Improving data literacy within our organisations.

The water sector is full of engineers, scientists and others that possess a range of data skills, using and analysing data every day. But our data literacy - thinking critically about data in different contexts and examining the positive and negative impact of data collection, use and sharing - is still developing. By working with each other and our partners, we will build our internal expertise in data literacy, not just to be better data practitioners, but to get the most out of data for our customers, society, and the environment.

Commitments

The core of our strategy is what we as the sector will commit to do for our customers, the communities we operate in, and the natural environment. Therefore, both collectively and as individual organisations, we commit to:

1. Publish appropriate data that is valuable for our customers, society and the environment. We will explore the value of open data through a social, economic and environmental lens and will aim to drive benefits to communities and businesses, but never in a way that causes harm or is otherwise illegal, unethical or irresponsible. We will do so transparently and collaboratively as part of embedding a more open culture.

2. Set out individual release schedules or roadmaps for data publishing. Different companies are different sizes, with different open data maturity levels and different levels of resources. We can't all aim to publish at the same cadence but each organisation can set itself an ambitious cadence for its own situation.

3. Embed common metrics for monitoring and evaluation. We will establish common metrics such as published and downloaded datasets, number of known reuses and qualitative user feedback, to measure ourselves by. We will use these metrics to drive continuous improvement in our activities to ensure value to customers, society and the environment.

4. Collaborate across the sector on open data priorities. We want to go far, fast and efficiently as a sector, so we need to go together. We will publish data together where it is more impactful to do so, and collaborate on open data standards, including quality standards such as accuracy and completeness.

5. Collaborate openly beyond data with information, insights and best practices. We can be open with more than our data, and share insights and feedback with each other so that improving one company helps improve us all. We also commit to publish all non-sensitive best practice guidance openly under a Creative Commons license.

6. Support the Stream initiative, either as full members, or as supportive stakeholders. Stream is a voluntary initiative that welcomes all UK water

companies. All companies should engage with Stream around open data publishing and standardisation, even if they are not full members.

7. Upskill our teams on open data and other relevant forms of data literacy. We will leverage existing resources within the sector and beyond to create organisational training plans that bring in knowledge on the value of open data, and how to approach legal and ethical considerations in the collection, use and sharing of data. Where relevant, we will leverage free online resources, expert training services and peer learning through our networks.

8. Collaborate beyond the water sector on open data to drive additional value to our customers, society and the environment. Being open with our data means working with other sectors, especially energy, and other physical infrastructure providers through initiatives like the National Digital Twin Programme and the National Underground Asset Register.

9. Be responsible open data users and innovators ourselves. We can use open data from the water sector and beyond to provide better services too. We will do so legally, ethically and responsibly, and emulate best practices, including attributing, giving constructive feedback, and helping others publish reuses and case studies.