OPEN UP

Open data opportunities for water

A project from

THE WATER REPORT *POLICY REGULATION COMPETITION*

YorkshireWater

"Dear New Yorkers, It's this simple: data makes our city run better. It optimises the bus route you take in the morning; finds the perfect spot for a new park; places fire hydrants in the most useful location on a city street; and drives so many of the decisions, big and small, we make every day." So says Bill de Blasio, mayor of New York in a report published at the turn of the year charting the city's "groundbreaking" results since launching its 'Open Data for All' programme in 2015. The report also sets out the city's ambition, over the course of this decade, to evolve open data "from publishing to problem solving" including by empowering citizens to co-create.

Then there's Transport for London (TfL), which has for over a decade now released timetables, service status and disruption information in an open format for anyone to use, free of charge. A 2017 report by Deloitte found over 600 apps used by 42% of Londoners had been created, providing better travel experiences; there were annual economic gains of up to to £130m; social benefits including healthier lives from more cycling and walking; and environmental benefits from reduced emissions and better air quality.

With such positive results from open data there for the taking, surely it's worth water companies, with their manifold challenges as the new decade dawns, considering what democratising data could do for them?

WHAT'S TO GAIN?

Transparency and trust

A virtually guaranteed result would be a boost to legitimacy from increased transparency. The desire to empower customers and communities to hold it to account, and therefore to build trust, is a key reason Yorkshire Water is rapidly moving to an 'open by default' position, working in partnership with the Open Data Institute (ODI) Leeds and the Data Mill North publishing platform (see box overleaf).

Richard Emmott, director of corporate affairs, has a vision that being an "open book" could create a cohort of what he terms "citizen regulators" scrutinising and challenging performance and perhaps enabling formal regulation to take a step back. He admits it's "no more than an idea" at present, but has a vision that as Yorkshire Water releases all its Performance Commitment data over the coming AMP, customers will at some point gain the ability to create their own personalised service dashboards, tracking the company's performance in areas which matter to them the most.

Transparency is absolutely the reason for

Southern Water's imminent new move in the open data space. In March, it will launch a new section of its website which will act as a one-stop-shop for information on its environmental performance. There will be raw datasets supplemented by what head of business assurance Barry Cleasby calls an "interactive, easy access front end", on metrics like pollution incidents, bathing water compliance and wastewater effluent, as well as carbon emissions and river levels.

Cleasby says making this data available to all was something the company volunteered for as a direct result of the need to rebuild trust in the wake of last year's £126m penalty from Ofwat for serious sewage treatment failures and deliberate misreporting. "If we're not transparent, we will struggle to build trust over time," he comments. "We are acutely aware of the movement to become more transparent...we see it as an opportunity, the right way to go."

In fact, he says, the company is now very much looking forward rather than back. "Rather than being about data analytics, it's morphed to become something more fundamental about how we engage with our customer base...it's driving our overall comms strategy and we want to build a more evidence-based discussion...It's not about any single product, it's about how you live your corporate life."

The March launch is phase one. The phase two ambition is both to grow the amount and type of data published and to use it to drive interaction with customers. Prominent will be supporting customers to reduce consumption to an average 100 litres a day, under Southern's Target 100 programme.

Engagement and co-creation

So it is quite easy to see how greater transparency through open data has the potential to stimulate and alter engagement with service users. At its most successful, this engagement would become co-creation – hence NYC's focus for the 2020s on shifting 'from publishing to problem solving'. But even baby steps can make a difference.

Electricity distribution network operator Northern Powergrid partnered with the ODI Leeds in 2019 and the first project was to explore how to make the Distribution Future Energy Scenarios (DFES) the company had to produce – an outline of potential future events – more engaging and interactive. The usual route was to produce PDF reports, "primarily using internal or closed sources of data, created in isolation from other organisations that face the same challenges," ODI Leeds reports. "However, these methods prevent meaningful collaboration or feedback."

Together, the project team visualised the DFES on maps, and created tools to allow users to interrogate and interact with the underlying data. Paul Connell, founder of ODI Leeds, speaks excitedly of "killing reports". The team also created a data standard for DFES that others could use, allowing for greater collaboration between companies like Northern Powergrid and public-facing bodies like local authorities.

Southern is not alone in facing drought and water conservation challenges, and on that subject alone most would agree that publishing data in a way that informs and engages customers and allows them to explore choices and consequences around water use would help consumption reduction efforts. More generally, the sector is not short of challenges that customer and community engagement would help with, be it to plan investment better, save money, boost performance or improve service.

Collaboration and innovation

In turn, the potential for greater stakeholder engagement and interaction edges us towards two clear needs of the water sector in 2020: collaboration and innovation. Ofwat's senior director of strategy and planning John Russell is part of the top team that has just set water companies the most stretching final determinations ever. He is enthusiastic about the role open data could play in supporting successful delivery in AMP7. He describes it as "one of the key tools in the toolbox to rise to the challenges Ofwat has set" and passionately urges the sector to share information openly – with each other, the supply chain and others – to "get their teeth into data and solve big problems".

He is explicit: "Open data is a foundation stone of collaboration and innovation" with the potential to both tackle specific challenges and, perhaps even more importantly, "nurture a more innovative mindset" which would open doors to all sorts of possibilities. Now, let's not pretend that is in any way straightforward. Yorkshire Water's Emmott points out that even when a company charges headlong towards being open by default, there are good old fashioned procedural matters that get in the way of seizing ideas from all-comers. "Ideally we'd use open data to stimulate ideas from the supply chain," he mulls, explaining though that open data doesn't conform with traditional tender approaches. "So we've got some way to go to come up with a good way of commercialising this." He suggests as an example posing a specific question and awarding a cash prize, then perhaps sharing the IP with the successful company.

Nevertheless, the doors that open data opens for innovation are crucial to the business case for it. In fact, to ODI Leeds' Connell, the transparency benefit of open data is "a hygiene issue – regulated businesses should be transparent," the potential for innovation is where the excitement is, and the potential for a "step change" in performance. No wonder, he says, Ofwat has been clear it wants the industry to be "much more ambitious".

NO MORE HOARDING PLEASE

WHERE ARE WE NOW?

So, where is the water sector on open data at the moment? You don't have to look that far to find examples of companies dipping their toes into the open data space. Scottish Water established an open data process and plan back in 2017. That same year, Southern Water collaborated with a team of nine organisations including Google on a project known as DataWell, to build something that had the ability to ingest, model, transform, blend, visualise and export data. DataWell was found to be viable and scalable and within a year the project was handed to Ofwat to take forward on the grounds it would benefit from industry-wide engagement.

Today, a number of companies have open initiatives underway. These include Wessex Water, whose new online 'Marketplace' makes specific datasets freely available and asks for third parties to pitch ideas; and Northumbrian Water, which is partnering with Ordnance Survey and others to create the country's first underground utilities map.

There is also the start of collective action. Companies have collaborated, outside of Water UK, to form a Water Data Task Force (WDTF). This involves Defra, the Environment Agency and Ofwat as well as all the companies. Taskforce chair, and Affinity Water's asset modelling, tools and data programme manager Max Gamrat, explains open data is one of three key areas of focus for the group. It is also looking at digital twinning and data sharing, and among its objectives are to stimulate conversation in the sector on the potential of better data use; to scope out possibilities; to centralise and coordinate actions – such as what datasets should be published nationally; and to engage on behalf of the industry with government and other stakeholders. Modelled on the Energy Data Taskforce, the water group is nine months old and has an 18 month roadmap.

So while water clearly lags behind the likes of the public sector which is open by mandate, it is not, in the expert view, that much of a laggard among like sectors. ODI Leeds' Connell says energy, for instance, "is not miles ahead," rather that "everyone's on a journey" and there has been a "huge acceleration" in interest across the board recently in how data is used and understood. Southern Water's Cleasby is optimistic about the potential for change: "We're all on this journey. Once one or two first movers leap, that will set the standard for the sector as a whole."

That is something Ofwat would love to see, Russell explains. He observes "good signs of activity" in the sector but says these are "all at a relatively early stage" and in his view, water is behind the progress made in energy. "We need to come together more effectively," he says. "Our overall message is it needs to be turbo-charged now...No more hoarding please, there's a common need to get better at all these things."

BEING AN OPEN BOOK COULD CREATE A COHORT OF CITIZEN REGULATORS

YORKSHIRE WATER'S OPEN DATA JOURNEY

Yorkshire Water committed to open data in earnest in 2018 through a partnership with the Open Data Institute Leeds. Yorkshire made the choice on the back of incoming PR19 challenges. Director of corporate affairs Richard Emmott explains the company could see how to make some improvements, "but could not see how to get all the way" – so sought outside help.

Between March and May 2018, Yorkshire Water published three data sets: 12 months of leakage data, pollution incident data from the last five years, and new leakage data derived from acoustic monitoring devices inside water mains. There were events and a hackathon. Emmott says: "We deliberately didn't have a plan or a roadmap, it was a case of try it and see what happens". He remarks the company could either be termed "agile" or as an "anarchic disruptor!"

There have been some significant wins from opening up these first datasets. Emmott offers the example of an algorithm to eliminate false positives and negatives from the sound files generated by the acoustic monitoring devices, and a Shazam-type app for listening for leaks.

Since then, Yorkshire Water hasn't looked back. Datasets open freely and available to all, some on the company website, others on Leeds City Council's open data site, Data Mill North, include:
■ 75m lines of leak data in 2019.
■ All water resources data behind its Water Situation Report. This is published monthly and started in the mid 2018 drought. Emmott says: "Again it was the transparency point, as we were asking the public to save water, it made sense to show them reservoir levels, demand levels and the like."
■ Water quality sampling data, searchable by post code.
■ The last three years of Environment Performance Assessment reports, including self-reported status on each incident and categorisations of all non-compliant treatment works.
■ The last five years of all 26 performance commitments (PCs) as detailed within each Annual Performance Report (APR). To support this, by the end of February, the intention is for all 2018/19 APR data to be published. In AMP7, the company has committed to release regular supplementary information on its 43 PCs.

Emmott estimates the company is "six to nine months away" from its goal of being 'open by default' – the only exceptions to the open data policy will be personal identifiable data and information with critical infrastructure security implications. And in the new AMP, the company will turn its attention to publishing data on a more dynamic basis (towards real time rather than static "annual report" type data). "We've not got the technology to do that currently... we're scoping IT requirements."

Emmott adds it also wants to get more people engaged with the data and associated events like hackathons.

WHAT IS OPEN DATA?

According to the Open Data Institute, open data is data that's available to everyone to access, use and share. It should be easy to access. It's not the same as shared data, which is defined groups sharing information with each other. The Institute encourages both open data and open working, which means publicising projects early on and asking for input throughout. The ODI's mantra is to be "radically open".
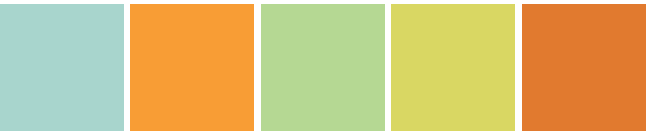
KEY QUESTIONS

Regulator's role

As ever in this sector, almost regardless of the topic, there is the question of how much should it fall on the regulator to drive change? Russell is clear that Ofwat has provided ample encouragement – among other things through the PR19 challenge, the new £200m Innovation Fund where being open by default is an application criteria, and through its new long term strategy – but that it is for the sector to lead. "It's always the hope of the regulator that the industry will do things," he remarks.

ODI Leeds' Connell thinks Ofwat has hit the right note. The regulator, he says, has been "pretty clear" on the need for innovation, greater transparency and operational performance transformation. "If you wait for the regulator to tell you what to do, it's the wrong way to do it." Yorkshire's Emmott agrees. "The challenge is set," he says. "The industry just needs to think about two things: innovation and public transparency. Publicly owned companies have to be transparent, why should we be any different?"

The WDTF's Gamrat, though, thinks the regulator needs to provide support in terms of setting a framework for practical action for the whole industry. "For open data to move forward, we need a regulator-led and supported agenda on publishing data – this has been brought up numerous times by water companies involved with the WDTF. Something like a delivery mechanism, a platform and a set of standards for industry wide data publishing, which can stimulate opportunities and inform the wider public on water usage and water infrastructure, for example. Ofgem have taken this sort of approach in the gas and electricity markets."

He adds: "The innovation fund is a great opportunity but we also need, as an industry, to up our game in

terms of working together for publishing datasets and making them easily accessible, something that will need Ofwat's support – think of it as a need to modernise the sector and move into the realm of digital reporting."

What about all the obstacles?

You don't have to look far to find difficulties that might put water companies off the open data path. Data protection, security issues, competition law, competitive advantage, data ethics, data quality and the prospect of negative commentary if performance isn't up to scratch or the data is misinterpreted are all legitimate considerations.

There are also practical issues such as the absence of common data standards and finding a platform to publish on. Cleasby says the former is felt keenly by water companies, as regional, relatively isolated, businesses. By comparison, "banking has a massive imperative to have commonality, to interconnect".

Gamrat says the WDTF is working on ideas through which members will be able to share datasets (so shared, rather than open data). Connell reckons for about £20,000 each, companies could build a good platform. "Open data should not be very expensive," he says.

Emmott says Yorkshire only hit two main obstacles: rules about publishing information on critical national infrastructure and GDPR. Both were surmountable, the former with half an hour on Google and the second with "lots of time spent anonymising data". Ofwat's Russell thinks that's the right attitude. There are difficulties, "but the question is, do you lean into the problem?" In his view, the problems are "absolutely solvable".

The thorniest nuts to crack seem to be cultural. Mindset comes up time and time again, be it because companies are conservative by nature, afraid of exposing their shortcomings, or afraid of losing competitive edge. There's no easy answer to that, but Emmott points to "significant

senior level commitment" providing a "mandate for change" from the top at Yorkshire. Likewise, Cleasby says his CEO Ian McAulay's corporate transformation agenda following the wastewater penalty has provided impetus at Southern: "Because of our specific challenges, we're moving quickly in this space."

Where to start

As for where to start, Emmott has some simple advice: "Just go ahead and do some stuff." Connell agrees: "Do it, just start. Have a problem statement, publish some datasets." He says you don't need a well crafted multi-month plan, or to wait for your peers in the industry.

As for which datasets to start with, Ofwat's director of data and modelling Andy Duff says: "There's something in just getting it going, so the low hanging fruit is attractive in that sense". He offers the example of acoustic monitoring data, which is not well understood at present so there's "potential for lots of innovation". But equally, Duff says, companies should look at "powerful datasets" like flow data that go to the heart of industry problems like leakage and supply interruptions. He thinks there's "huge potential for better resilience" if such datasets were scrutinised across borders. The ODI's Connell's top pick is data to crack leakage. "It's a challenge for all, just start publishing it – why would you not?"

Format wise, there are options even if the sort of dynamic data TfL publishes are out of reach in the short term. Connell's advice is "people used to working with data probably just want raw data" but the ODI also encourages companies to provide a simple tool to aid visualisation and presentation where they can.

Emmott concludes: "I'd be delighted if other companies got out there, took a few risks, started trying things... maybe end up trying to outdo each other on their bravery." He adds, as final reassurance: "We've shown the sky doesn't collapse around your ears."

DO IT, JUST START. HAVE A PROBLEM STATEMENT, PUBLISH SOME DATASETS

HAVE A VIEW? INVITATION TO JOIN THE DEBATE

The Water Report and Yorkshire Water will be co-hosting a round table in Spring to discuss the issues raised in this paper and get sector views and experiences. To express an interest in taking part, please email karma@thewaterreport.co.uk