Open data assessment report

Table of contents

	Slide
Executive Summary	4
Framework Design	18
Assessment Design	27
Industry Learnings	33
Company Assessments	51
Appendix 1 - Framework development and feedback	71
Appendix 2 - Assessment Questions	77
Appendix 3 - Progress Criteria	86
Appendix 4 - APR and EDM data requirements	95

Introduction

Terms of reference
PricewaterhouseCoopers LLP (PwC) was commissioned by Ofwat to develop and deliver an assessment of water industry's progress on Open Data in response to H2Open.

Process of work
To assess progress, we developed an assessment framework in collaboration with Ofwat, water companies and water and other industry stakeholders. Our approach to developing the framework involved the following steps:

- We conducted a desktop research of Open Data maturity models, data best practice and progress frameworks
- We organised workshops with relevant stakeholders such as government departments, regulators, companies, and data users to refine the framework and assessment criteria
- We developed a finalised assessment criteria to measure company progress on Open Data
- We created a comprehensive questionnaire for collecting information from water companies to provide evidence of their work in enabling Open Data

Following company responses, we conducted an evaluation of the submissions against the assessment criteria and prepared one-page summaries (hereinafter "one-pagers") for each company highlighting use cases of Open Data and best practices of each company.

Scope of work and limits
The scope of the work was to:

- Develop a robust framework for assessing open data maturity in the sector, which can be used for repeated assessment in future
- Engage with water industry and other stakeholders in development of the framework
- Produce a summary report of findings on progress and examples of good practices, and next steps/recommendations

The scope of our work excluded the following:

- Review or account for any progress made by companies on Open Data after end of submission period
- Review whether companies have met their legal or regulatory obligations associated with release of data
- Provide recommendations on the regulatory response for Ofwat based on the findings of the review

1. Executive Summary

We developed an open data framework and assessment approach to meet the assessment objectives set out by Ofwat

Assessment Objectives

- **Showcase the open data progress companies have made** since the H2Open paper was released in October 2021
- Enable progress across the water sector through **highlighting good practice** from both within the water sector and beyond
- Support Ofwat's next steps in encouraging further progress on open data

Guiding principles

- Draw on **existing good practice** from other sectors
- Be capable of being used as a **self-assessment tool**
- **Repeatable** in future to track progress for individual companies and across the sector
- Easily **interpretable** by users, evaluators and the public

We collaborated with a range of experts whilst developing the assessment framework

The framework was originally developed by drawing on published best practice literature on open data and the knowledge of PwC and Ofwat. The framework was then refined based on feedback collected in two workshops and a series of 1-1 sessions with wider stakeholders, water companies and data users.

Ongoing collaboration between PwC,Ofwat and Open Data experts to create a repeatable self-assessment framework

1. Initial Research

The initial framework structure identifying beneficiaries, outcomes and enablers was developed based on;

- ODI's Open Data Maturity Model's themes
- Ofgem's Data Best Practice Guidance,
- Catapult Network's Energy Data Taskforce Report
- UKRN's Infrastructure Data Sharing
- Ofwat's H2Open's Enablers

2. Wider Stakeholder Workshop

We consulted with a range of stakeholders including government departments, independent advisory bodies and regulators to refine the framework structure.

We also identified relevant activities to measure progress against enablers.

3. Water Companies Workshop

We engaged with water companies to discuss the types of questions that will be asked to assess progress against the developed open data framework.

The questions were adjusted in line with feedback and a list of potential supporting evidence to be supplied by companies was created to verify company responses.

4. 1-1 Session Input

1-1 sessions were utilised to gain deeper feedback from regulators and data users to ensure the framework represents the needs of various stakeholders.

5. Ofwat's Sign-off and Clarification session

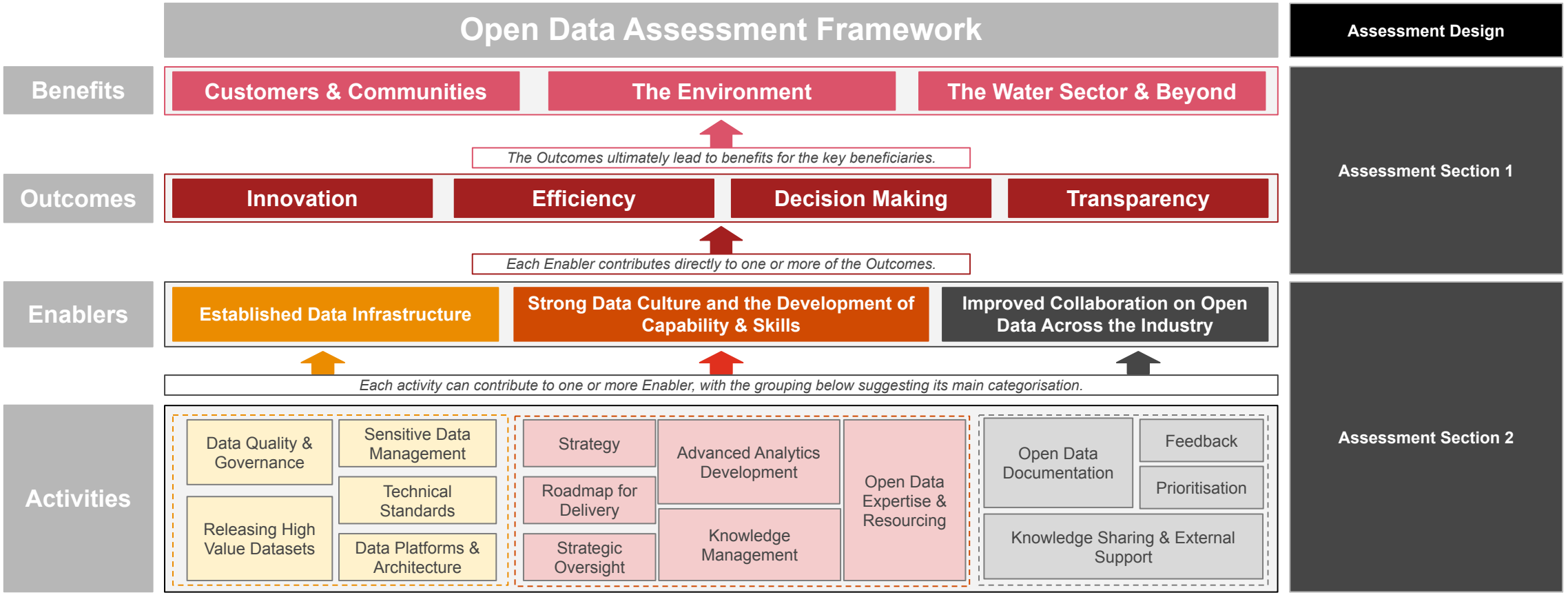
The final Ofwat approved version of the framework was sent to water companies for review and to raise any queries in a clarification session.

Companies were provided a 3 week period to respond to the framework questionnaire.

Additional details can be found on the following slides

A framework consisting of Activities, Enablers, Outcomes and Beneficiaries was devised, leading to a two-part assessment

The first section of the assessment asked open ended questions aligned to each of the Outcomes, drawing out case studies of good practice. The second section focused on the Enablers/ Activities, looking to understand the progress that has been made in facilitating Outcome achievement.

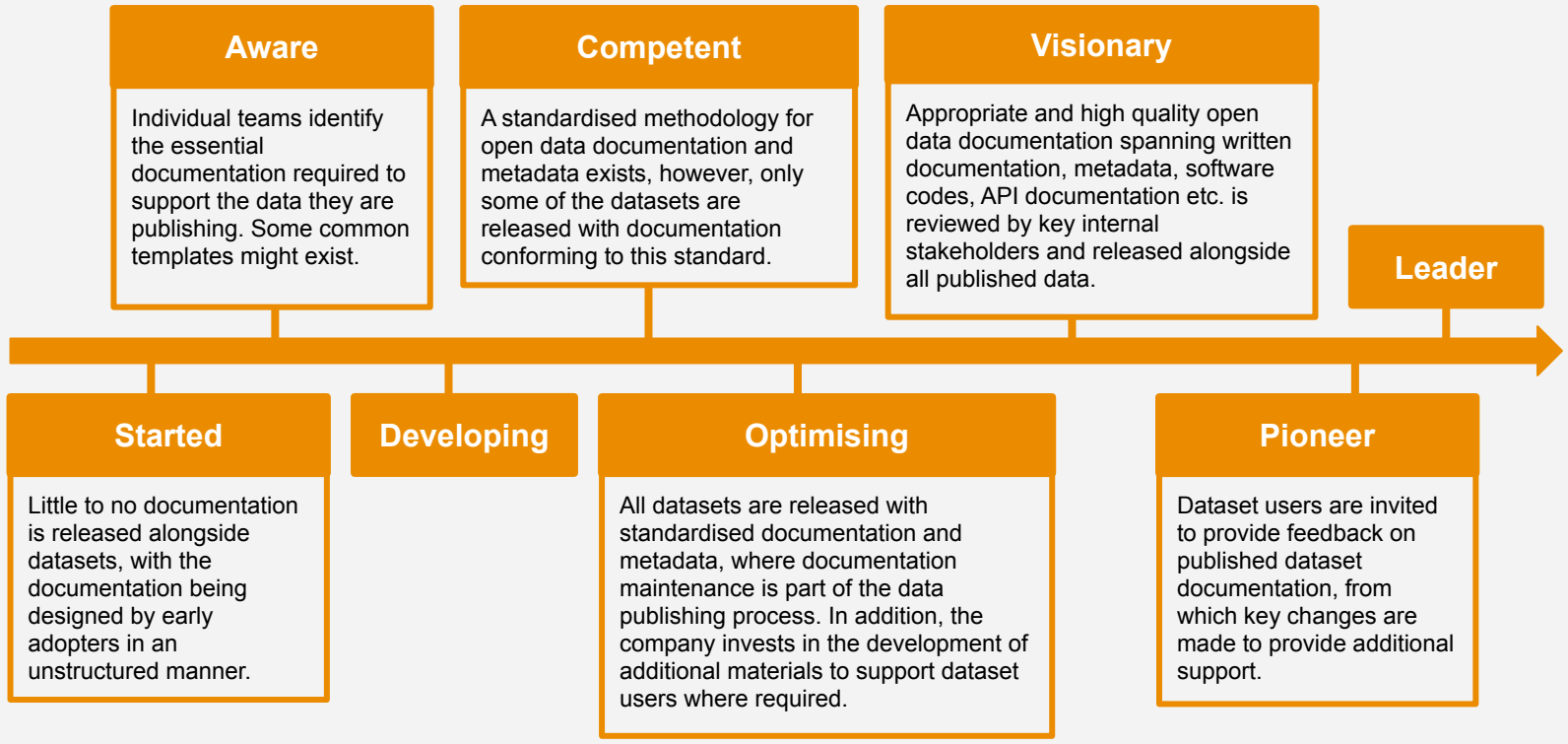


Please click here for full assessment criteria

Progress statements were used to map what good looks like for each Activity

To navigate the shared problem of conducting a progress assessment with no baseline to work from, we devised eight progress statements for each activity to define the journey from just starting to becoming a cross industry leader.

The following is an example showing the progress statements for Open Data Documentation:



Key Design Considerations

- The progress statements were **designed with cross-industry best practice in mind**. Hence, the assessment is a measure against cross-industry best practice rather than the water sector against itself
- The progress statements **provide a benchmark, enabling a repeatable assessment** to measure future progress

Note: Not all activities have a progress statement aligned with all 8 points on the scale, however, if a company response indicates it lies between two non-adjacent progress statements, then the appropriate statement from the full 8 point scale will be used to summarise their performance.

The evidence attached to assessment responses was evaluated and datasets were categorised as either shared or open (1 of 2)

PwC evaluated evidence to make sure it met the required standard, and datasets provided were categorised to be either open or shared based on the Open Data Institute (ODI) definition for open data.

Assessment Approach & Evidence Evaluation

Companies provided written response and attached evidence for each assessment question → **PwC evaluated the assessment response and evidence** → **One-pager progress summaries created based on company response and evidence**

- Potential sources of evidence were provided for each assessment question to guide companies on appropriate evidence expected to support their response.
- If evidence against a particular response was deemed insufficient, the case study or the reported dataset has not been mentioned in this report. For instance,
  - where companies have failed to provide any supporting evidence or details about the process or outcome from a given activity in their response.
  - where companies have provided broken links to datasets/supporting evidence published online.
- No additional material outside of the assessment response or supporting evidence was used in the evaluation of companies' progress on open data.
- Industry learnings were devised through aggregating the one-pager progress summaries to understand where the industry as a whole has progressed faster / where it needs to focus its efforts.

Assessment of Progress since H2Open: Companies were informed through the industry workshops and upfront in the framework questionnaire, that the assessment aims to understand the progress companies have made since H2Open. However, the individual questions did not ask about the timeframe for each Open Data related activity/case study that company submits as evidence. As a result, the responses submitted by companies includes applications of Open/Shared data from years preceding the publication of H2Open.

This assessment therefore provides an evaluation of companies progress to date, and establishes a baseline to track progress against in the future. Additionally, where possible, we have highlighted case studies that have either started or made further progress over the past year.

The evidence attached to assessment responses was evaluated and datasets were categorised as either shared or open (2 of 2)

PwC evaluated evidence to make sure it met the required standard, and datasets provided were categorised to be either open or shared based on the Open Data Institute (ODI) definition for open data.

What is open data?

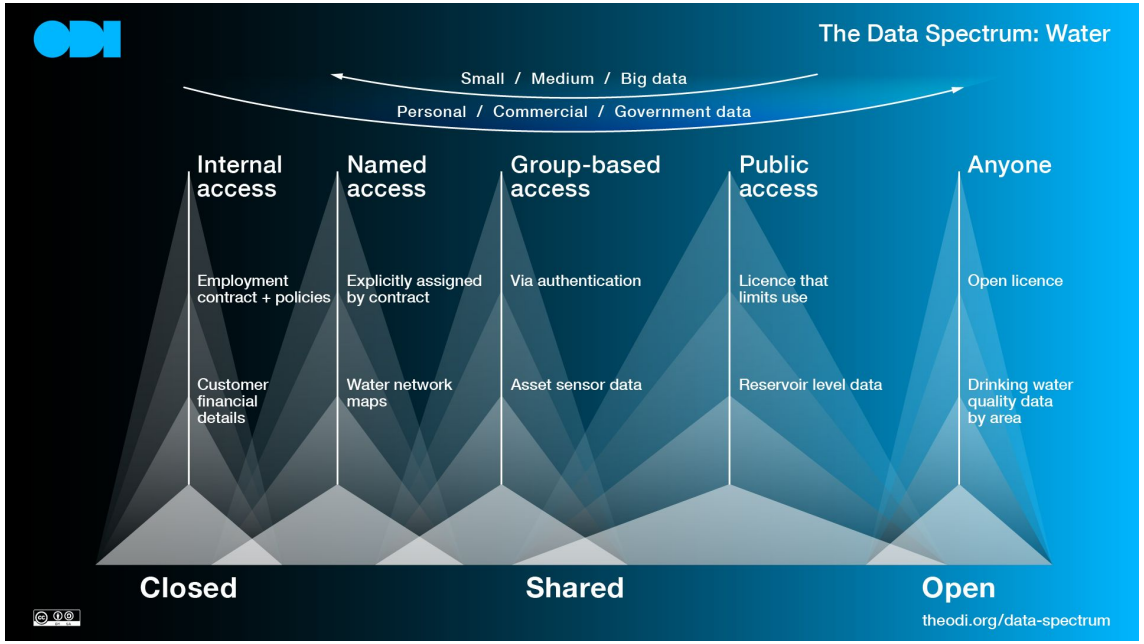
- We have adopted the Open Data Institute's definition for open data, built on the principles developed by the Open Knowledge Foundation
- The ODI defines open data as **data that anyone can access, use and share**
- The Open Knowledge Foundation, whose principles inform the ODI definition, states that open data must be made **available in common, machine - readable formats** i.e. provided in a form readily processable by a computer and where the individual elements of the work can be easily accessed and modified by users
- Open **data must be licensed**. Its licence must permit people to use the data in any way they want, including transforming, combining, and sharing it with others, even commercially

What is shared data?

- Shared data is data made available to specific groups of users
- Access is restricted to specific users via authentication

The Data Spectrum

- The diagram on the right shows the data spectrum developed by the Open Data Institute
- The spectrum is a useful tool in understanding the key differences between open and shared data



Assessment of Water Company Datasets

- The datasets provided by the water companies on the assessment were categorised into two groups - open or shared.
- Where data was categorised as shared, it was typically due to having restricted access (i.e. not public) or being published in a format that is non-machine readable e.g. pdfs. Our dataset review process first investigated whether the dataset submitted by companies were publicly accessible or not. If a dataset was publicly accessible, a further check was conducted to review if the dataset was machine readable.
- Barring Thames Water's API portal, which provided a data license, an open data license was not discoverable for any of the datasets provided by the water companies as part their response to this framework.
- Therefore, by the definition of ODI, only Thames Water's EDM data published on their API portal will be considered as 'Open'. However, to highlight the number of datasets published by water companies that are **openly accessible** and **published in a common, machine-readable format**, **we categorise any such dataset as open despite a lack of a supporting open data 'license'**.

Our assessment shows that where benefits are being delivered to key beneficiaries, this is typically only through shared data

Water companies have shared data with organisations both within the water sector and beyond to create solutions that provide benefits to the key beneficiaries identified in the assessment framework - Customers & Communities, the Environment, and the Water Sector & Beyond. However, there has been limited evidence of open data releases and hence few benefits achieved from it.

Customers & Communities

Customers & Communities, benefit in a number of ways; including:

- **Identification of financially vulnerable customers** and providing targeted services/tariffs **through shared data agreements** between water companies, Department for Work & Pension (DWP) and other utilities companies
- Applications focusing on **sharing maintenance, leakage and water quality information**, e.g. South East Water's In Your Area (IYA) service, are providing increased transparency for customers
- **Near real-time alerts on storm release activity** for example, Southern Water's Beachbouy App*, where alarms and sensors trigger alerts to the public if there is a storm release event

The Environment

In response to the current issues facing the water sector, the environment has been a key focus area.

- All wastewater companies have **released Event Duration Monitoring** (EDM) data (please see slide 96 for details). Two companies have taken the next step of making their Event Duration Monitoring (EDM) data more user-friendly. Anglian Water and Thames Water have **released the locations of their storm overflows** alongside their EDM data on **interactive maps** on their websites.
- **Wessex Water's Intelligent Sewers Competition***, led to the development of a predictive analytics tool in collaboration with StormHarvester that can identify potential sewer blockages earlier. The solution identified >80 blockages over the past year.

The Water Sector & Beyond

Cross-sectoral collaboration with the water sector on data-sharing initiatives is providing benefits to the whole the economy

- **Streetworks Monitoring and Evaluation tool** developed by Affinity Water in collaboration with Greater London Authority (GLA) has helped in reducing multiple road closures/diversion through sharing of data on planned works programmes with other utilities.
- Initiatives like the **Sewer CCTV AI project***, with multiple companies contributing CCTV footage, help to establish a sector wide repository of CCTV labelled images that companies can use to train their own AI models. These models **replace the manual process of coding defects with an automated solution thereby providing up to 20% of time savings for survey and coding operators**.

Note: Case study examples highlighted with an asterisk (*) were also mentioned in the H2Open publication (link)

There have only been a small number of open datasets released

Companies have achieved benefits, through varying outcomes, primarily through sharing data with key partners or by merging external third party open datasets with their internal data. A common understanding of the value of open data now exists across the industry but few open datasets have been released.

Regular data publications have led to trialling of methods to release data: Companies have used regular data publications to trial methods of releasing data (e.g. Annual Performance Report (APR) and EDM data, see slide 96 for further details), but not all companies have fully met the requirements for Open Data; for instance, by not releasing supporting metadata alongside APR data releases. However, these publications have prompted some companies to trial methods of releasing data that have the potential to facilitate additional data releases in the future. (See evidence on slide 36)

Companies are open data users more than open data providers: Companies are commonly creating benefits for the key beneficiaries by merging data from outside their company with their own data, i.e. they are keen users of open data but have made limited progress in opening of their own datasets. (See evidence on slide 36)

Data is shared rather than making it publicly accessible: When releasing data, companies are generally sharing data with key partners over releasing open datasets with unrestricted public access. Of the datasets referenced by the companies on their assessment responses, 39% (81 of 210) datasets were deemed to be shared due to being restricted to group-based access as per the ODIs open data definition (slide 10).

There is a lack of machine-readable open data: Whilst there are some data releases beyond APR and EDM data, these are commonly not done in a machine readable format. For example, through interactive maps with no way of downloading the data are commonplace. Of the datasets referenced by the companies on their assessment responses, 26% (54 of 210) datasets were deemed to be shared due to not being in machine readable format.

Progress in developing open data Enablers has been limited across the water sector, with larger companies further along their open data journey relative to smaller water only companies (1 of 2)

The water sector is at the beginning of its open data journey, with the majority of companies being categorised as Started, Aware or Developing for most Activities.

Enabler Level Learnings

Overall:

- Assessment of the open data Enablers show that the **water sector is still early in its open data development**, with the majority of companies achieving between 'Started' and 'Developing' for most Activities.
- Companies achieving higher progress statements are typically larger and are further along on their digitalisation journeys. Hence, general **Enabler progress is likely driven by broader digitalisation efforts and not solely by open data initiatives.**

Established Data Infrastructure

- **Investment into Data Platforms & Architecture is the most common theme** across the assessment, with many companies currently working towards transitioning towards a cloud-based architecture. This investment is often part of a wider digital transformation with only a few companies making explicit consideration towards how open data may be facilitated on the platform in the future.
- **Releasing High Value Datasets is an area of relative weakness**, where companies have rarely made data open and often fail to provide sufficient metadata or documentation to support releases.
- Some **companies reference that they are engaging with wider collaborative groups** to define/ adopt supporting frameworks and standards. For example, common Platforms, Data Governance Frameworks and common Technical Standards. However, limited evidence was provided by companies to show progress in these areas.

Heatmap showing the Companies' Progress Statements by Activity

		Started	Aware	Developing	Competent	Optimising	Visionary	Pioneer	Leader
Established Data Infrastructure	Data Governance	5	1	3	5	2	0	0	0
	Data Platforms & Architecture	2	0	3	6	4	1	0	0
	Data Quality	1	5	5	2	2	1	0	0
	Releasing High Value Datasets (APR Data)	3	5	5	3	0	0	0	0
	Releasing High Value Datasets (Immediate Challenges. E.g. EDM)	9	1	4	1	1	0	0	0
	Sensitive Data Management	3	10	1	2	0	0	0	0
	Technical Standards	5	2	3	3	3	0	0	0
Improved Collaboration on Open Data Across the Industry	Knowledge Sharing & External Support	2	3	2	8	1	0	0	0
	Open Data Documentation; Feedback	4	8	3	0	0	1	0	0
	Prioratisation	5	7	4	0	0	0	0	0
Strong Data Culture and the Development of Capabilities & Skills	Developing Advanced Analytics	5	5	2	2	2	0	0	0
	Knowledge Management	4	4	5	3	0	0	0	0
	Open Data Expertise & Resourcing	5	4	6	1	0	0	0	0
	Roadmap for Delivery	5	7	4	0	0	0	0	0
	Strategy; Strategic Oversight	3	7	5	1	0	0	0	0

Progress in developing open data Enablers has been limited across the water sector, with larger companies further along their open data journey relative to smaller water only companies (2 of 2)

The water sector is at the beginning of its open data journey, with the majority of companies being categorised as Started, Aware or Developing for most Activities.

Enabler Level Learnings

Improved Collaboration on Open Data Across the Industry

- Assessment responses indicated that most companies do not have **feedback loops between companies and data users and is an area of weakness** for water companies to improve on.
- **Knowledge Sharing & External Support is an area of relative progress**, with 8 companies scoring competent. This is driven by involvement with collaborative initiatives, however, outputs from these groups were difficult to verify.
- Evidence provided suggests that there are **no collaborative initiatives bringing all 16* water companies together** in relation to open data.

Data Culture and the Development of Capabilities & Skills

- **Strategy, Strategic Oversight & Roadmap for delivery is an area of weakness**, with only 7 companies reporting to have a general data strategy and only 5 of these reporting to having elements specific to open data contained within it.
- **Progress in Developing Advanced Analytics was more common in larger companies** - albeit it focused on the use of other's open data over efforts to open their own data - whereas smaller companies typically opt to leverage external capabilities to support them.

Heatmap showing the Companies' Progress Statements by Activity

		Started	Aware	Developing	Competent	Optimising	Visionary	Pioneer	Leader
Established Data Infrastructure	Data Governance	5	1	3	5	2	0	0	0
	Data Platforms & Architecture	2	0	3	6	4	1	0	0
	Data Quality	1	5	5	2	2	1	0	0
	Releasing High Value Datasets (APR Data)	3	5	5	3	0	0	0	0
	Releasing High Value Datasets (Immediate Challenges. E.g. EDM)	9	1	4	1	1	0	0	0
	Sensitive Data Management	3	10	1	2	0	0	0	0
	Technical Standards	5	2	3	3	3	0	0	0
Improved Collaboration on Open Data Across the Industry	Knowledge Sharing & External Support	2	3	2	8	1	0	0	0
	Open Data Documentation; Feedback	4	8	3	0	0	1	0	0
	Prioratisation	5	7	4	0	0	0	0	0
Strong Data Culture and the Development of Capabilities & Skills	Developing Advanced Analytics	5	5	2	2	2	0	0	0
	Knowledge Management	4	4	5	3	0	0	0	0
	Open Data Expertise & Resourcing	5	4	6	1	0	0	0	0
	Roadmap for Delivery	5	7	4	0	0	0	0	0
	Strategy; Strategic Oversight	3	7	5	1	0	0	0	0

*Note: Bristol Water and South West Water provided a joint submission in response to the framework questionnaire and have been considered as one for the purpose of this report.

There are lessons to be learned from sectors that are further along on their open data journeys

The transport and energy sectors have made significant progress on their open data journeys and can be looked towards for examples of a greater open data maturity. The development of cross-industry frameworks, centralised open data portals and increased collaboration with data users could help the water sector realise greater benefits.

Examples of good practice

Development of cross-industry frameworks and data sharing templates

- The **Rail Technical Strategy (RTS)** highlighted the transport sectors ambitions for improving access and sharing of data across the railway industry. An essential component was the **Information Management Framework (IMF)**, which included the development of a data standards route map, data-sharing templates and a framework for identifying 'high-value' datasets.
- **Developing cross-industry data sharing templates and prioritisation frameworks** has enabled the railway sector to move quickly to make stronger use of open data, and has supported the creation of the **Rail Data Marketplace** (RDM); which will combine fragmented sources of railway data to form one digital service.

Utilising data portals and APIs to release data

- Data portals in the energy sector not only provide a centralised source for data users to find multiple datasets, but also act as a feedback mechanism to request additional dataset. For example: the **National Grid's Connected Data Portal** or the **Open Energy Data Portal.**
- The rail sector is releasing data collected from various monitoring kits through an **Application Programming Interface (API)**. API development has promoted the standardisation and led to the development of third party products and services; for example, travel planner apps.

Collaborating with data users and utilising the feedback loop

- The **Open Data Communities** programme, run by the Open Data Institute, supports collaboration between data users and local authorities to identify and publish open data that is important to local communities. The initiative has been successful in providing useful data to local residents, businesses and organisations, leading to increased transparency and engagement.
- **OpenActive**, a community-led initiative which aims to help people get active using open data, has built an open data ecosystem of over 100 organisations across the sport and physical activity sector, including almost 70 data publishers. It has achieved this through active engagement with individuals working within the sport and physical activity sector.

Learnings for the water sector

Adoption of cross-industry frameworks in the water sector

- The water sector should look to **develop/ adopt similar cross-industry templates and frameworks** to accelerate their path towards a more open data enabled business model.
- The companies involved in Stream and CastCo are working towards defining some of these frameworks. Further development of frameworks across the water sector with sufficient external collaboration would help the water sector progress further.

Developing data portals and APIs to release data

- The development of **open data portals** for the water sector could accelerate innovation by centralising and signposting the data currently available to potential data users. A data portal could also support the development of an established feedback loop.
- Thames Water have developed the first **generic open data API** seen in the water sector. Greater activity in this area, and **sharing of learning**, will see the water sector make considerable progress in open data and support the creation of apps and services in the future.

Engage with data users to identify priority datasets and to incorporate feedback

- Water companies should look to engage with communities and data users to help identify priority data sets. Companies can benefit from having a defined and repeatable approach in place to gather feedback from their data releases.
- Wessex Water's Marketplace platform provides a good example of how water companies can both share data and solve common challenges in collaboration with data users.

Next Steps / Recommendations

There are a number of next steps that water companies can take to increase the level of progress on open data across the water sector.

Recommendations to accelerate release of open datasets

As progress in releasing of datasets for public access has been limited across the water sector, companies should look to -

1. **Identify priority datasets and set stretching targets to release them such that they accelerate the development of companies' general open data capabilities to support further data releases in the future. Pilot releases of datasets should be conducted and a 'learn by doing' approach should be adopted to iteratively improve on each subsequent data release.**
2. **Review datasets that have been shared with only specific user groups or organizations and where appropriate assess the steps required to make these openly available.**
3. **Identify datasets that have been released for public but not in machine readable format and convert these to appropriate formats which can be readily processed by a computer and individual elements can be accessed and modified by users e.g. releasing the data in .csv format.**
4. **Release datasets with an open license to encourage use and re-use of the information freely and flexibly e.g. License to share and adapt the data provided by Thames Water on its API Portal.**

Next Steps / Recommendations

There are a number of next steps that water companies can take to increase the level of progress on open data across the water sector.

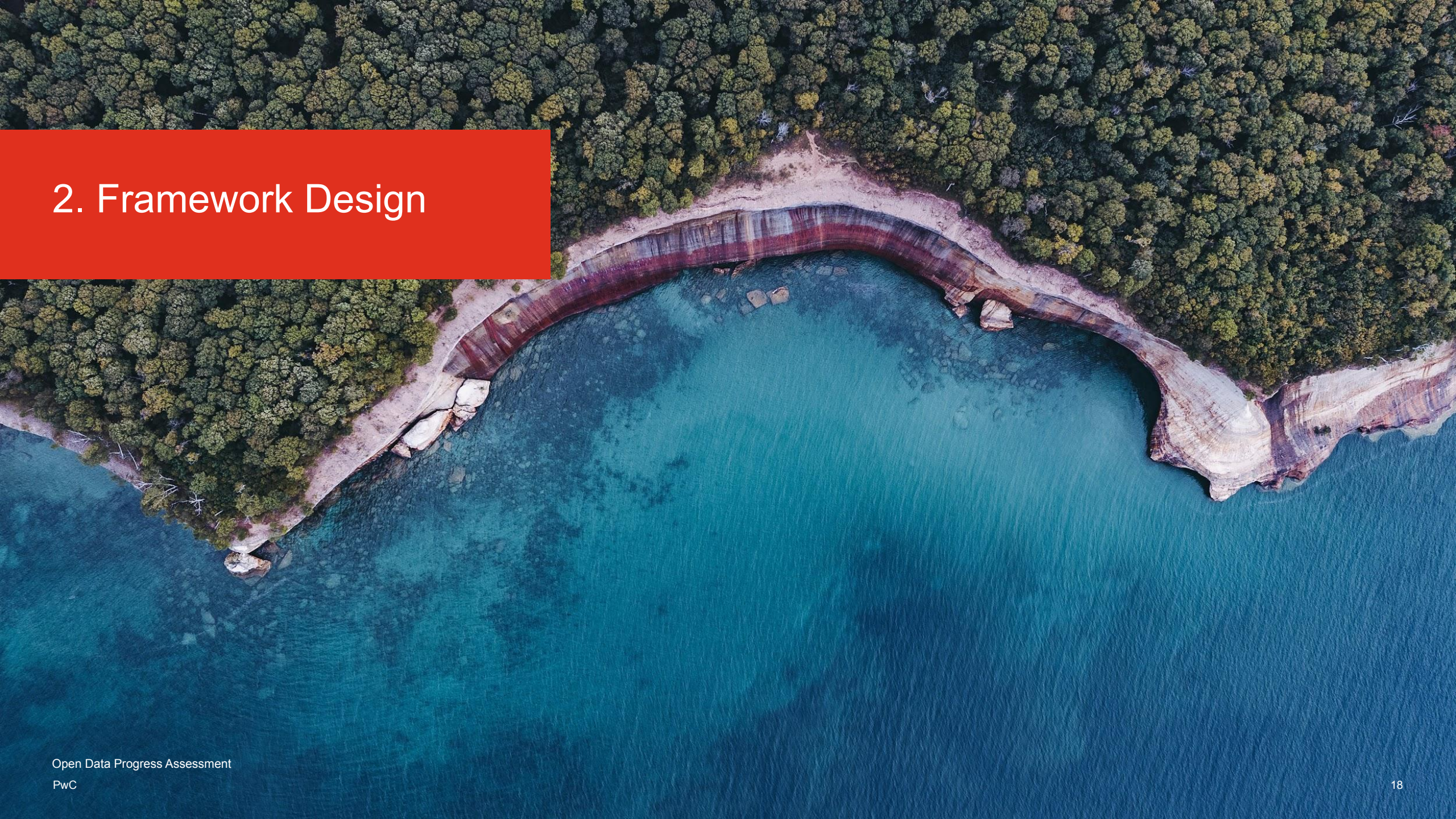
Recommendations on development of Enablers

Water companies have started to lay the foundations for open data by investing in data infrastructure, but further progress should focus on development of enablers in a balanced and integrated way -

5	**Companies should develop comprehensive open data strategies, clear roadmaps for delivery with associated delivery KPIs, and set up sufficient accountability and strategic oversight to ensure company board and C-suite buy-in e.g. by tying delivery of strategy to executive performance.**
6	**Companies should develop a culture of open data in their organisation e.g. by developing and delivering open data literacy training to create greater understanding of open data across all levels of their organisation.**
7	**Industry-wide collaboration is helpful for development of data portals, facilitation of knowledge sharing and development of common standard that enhance interoperability of data. However, such initiatives should be supported with adequate funding, resources and active participation from all companies in order to be effective. Further, priority-based, measurable and time-specific targets should be set to ensure maximum progress is achieved through these collaborative efforts. Such collaboration should also involve engagement with independent experts to draw on learnings from other sectors. Importantly, companies should not rely solely on collaborative groups for ensuring progress.**
8	**Companies should actively engage with data users and have a well-defined feedback loop, especially to identify priority datasets and improving quality of data release and associated documentation.**

For recommended next steps on each activity in our framework, please see the detailed enabler learnings section.

2. Framework Design



There are three key open data beneficiaries

The open data framework aims to assess the progress water companies have made in delivering value to customers & communities, the environment, to the water sector and beyond.

Customers & Communities:

Customers & communities benefit from open data through the development of a wider range of products and services: for instance, through provision of additional support to specific communities (e.g. vulnerable customers). Open data can also provide benefits to wider interest groups: for example, academics, researchers and NGOs). Additionally, open data can help build trust amongst customers by holding industry accountable for their performance (e.g. on leakage, water quality etc.).

The Environment:

The water industry holds important and relevant data that can help inform effective climate policy and enhance coordination, enable innovation and support investment decisions. Opening of water industry data (for example: leakage, EDM and other environmental data) can aid in the development of long-term projects and solutions that protect the environment: for instance, collaborating and building of cross sectoral data sets can help to deliver predictive forecasting, better understanding of the state of the environment and helping adapt to climate change.

Water Companies & Other Industry Players:

Open data can require an initial capital and resource investment but provides opportunities to find substantial efficiency savings for the companies and the sector as a whole: for example, through increased information and insights surrounding performance of assets. In addition, open data can be used to collaborate more effectively with innovators, government, and research institutions to share knowledge and resources in order to develop better solutions.

Four key Outcomes support the delivery of benefits

The framework supports the assessment of companies against four key Outcomes that were outlined in the H2Open paper. These outcomes support the transformation of the companies' day-to-day services for customers & communities and the environment. In addition, they enhance the accountability of the sector to the regulators, government and the public, and support companies in meeting the long-term challenges present in the water sector through increased collaboration and partnerships.

Innovation

- Stimulate innovation opportunities by drawing in and leveraging new skills and technologies.
- Encourage innovation by third party companies, academics, researchers and NGOs, through the merging and linking of datasets to form larger set of data capable of deeper analysis and insights.

Efficiency

- Provide increased efficiency through fresh thinking and the use of new insights, such as performance of assets allowing for predictive maintenance.
- Provide opportunities to increase efficiency through streamlining processes.

Decision Making

- Enable improved decision making at the business level, through informing current operations decisions and strategy.
- Enable improved decision making at the sector and policy level.
- Improve sector-wide decision making: for example, through benchmarking different manufacturer's pump reliability.

Transparency

- Support building trust in the water sector through the increased transparency provided by the inherent openness of open data.
- Deliver increased accountability of the water sector to its customers & communities, the government and the regulators.

Outcome realisation is driven through three key Enablers

The framework adopts the three key enablers identified by the H2Open paper as elements that support progress towards open data.

Established Data Infrastructure

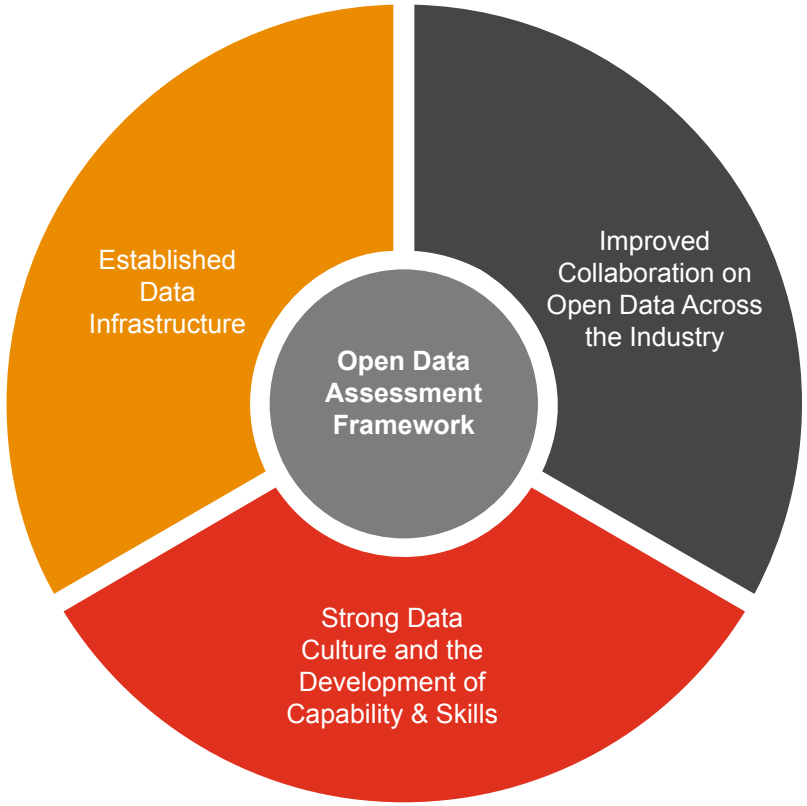
The level of data quality, completeness and common standards maturity, enabled by people, processes and platforms set up in a logical manner to achieve open data objectives and the companies' wider data strategy.

Strong Data Culture and the Development of Capability & Skills

The level of open data capabilities & skills in an organisation, built through staff training and targeted resourcing of open data capabilities. In addition, the culture built to support open data within the company, sector and beyond, including the existence of an open data strategy linked to organisational strategy with C-suite and board-level support to encourage further development.

Improved Collaboration on Open Data Across the Industry

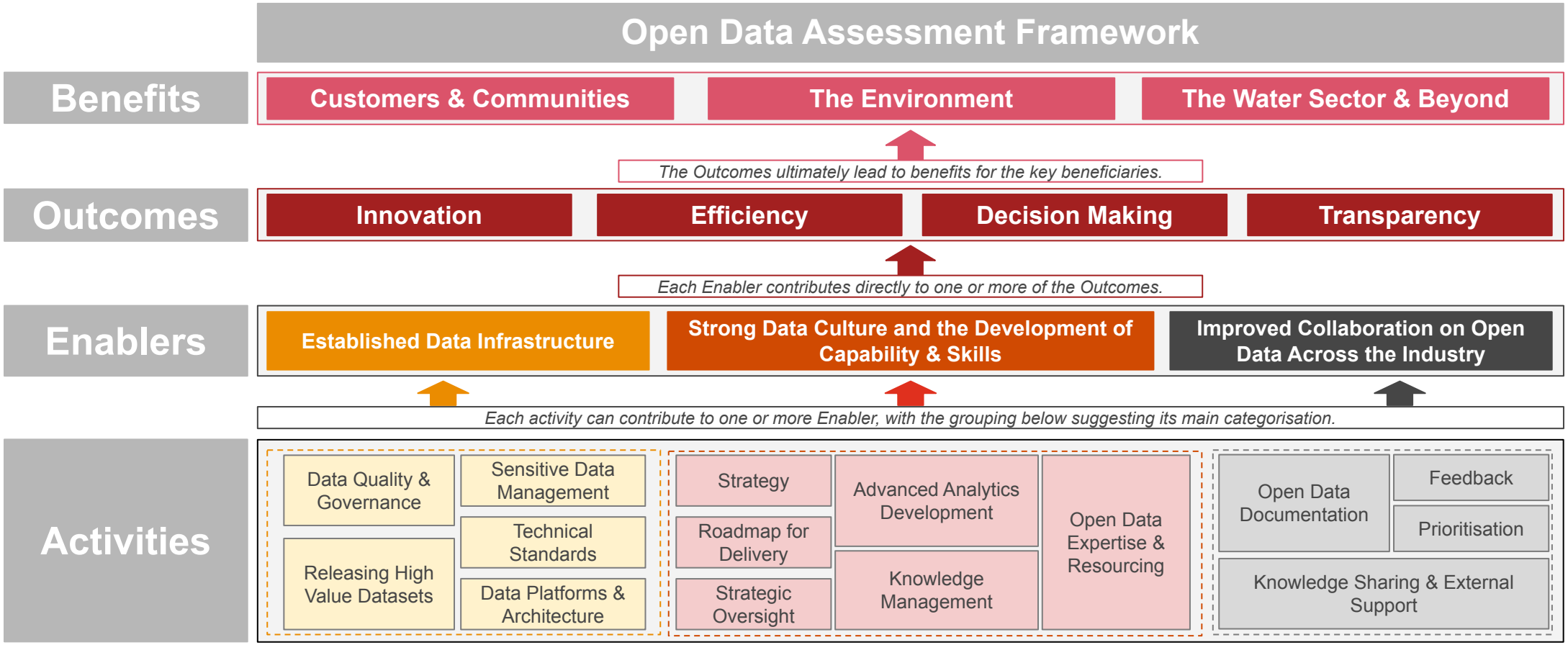
The level of collaboration across the companies and their communities (for example: local government, NGOs, academics, researchers, smaller industry players etc.), encompassing a culture of feedback and knowledge sharing, and prioritisation of high-value use cases of open data.



Note: **Water companies can realise benefits** for customers & communities, the environment and the water sector **before all elements of the Enablers are in place**

The open data progress framework

Across the three Enablers, we have identified a number of key Activities that support releasing high-quality datasets that lead to benefits for customers and communities, the environment, the water sector and beyond.



Enablers: Established data infrastructure

The level of data quality, completeness and common standards maturity, enabled by platforms set up in a logical manner to achieve open data objectives.

Outcomes

Efficiency

Data Governance processes and appropriate platforms are in place to capture and store data efficiently.

Transparency

Data is Open by Default with suitable processes in place to identify and anonymise personal data, protect commercially sensitive data and ensure security is not compromised.

Enabler

Activities

Data Infrastructure

Data Quality & Governance:

The organisation has a standard open data governance process in place and works to continually improve the process over time in line with technological developments. In addition, the organisation is working in tandem with dataset users to continuously improve data quality and maintenance.

Technical Standards

The organisation has a defined strategic approach for managing its technical standards, updating these standards in line with technological trends to ensure common standards are followed and preferentially adopting applicable open standards over closed or bespoke alternatives.

Data Platforms & Architecture:

The organisation has considered open data in the development of its data architecture / business intelligence (BI) operating model to best support the use of open data. In addition, relevant platforms are in place to optimise the use of open data: for example, by using platforms that share / host data, perform ETL (Extract, Load, Transform) / integration, allow data visualisation, enable advanced analytics

Releasing High-Value Datasets:

The organisation identifies and publishes high value datasets in a usable format, adopting an ‘open by default’ culture: thus enabling the development of innovative new business models, products and services that bring value to the industry and its customers.

Sensitive Data Management:

The organisation has clear processes in place to identify and anonymise personal data, and ensure that risks pertaining to commercially sensitive data are mitigated. In addition, sufficient triaging of existing internal and external datasets is performed to mitigate risks associated with linking and merging open data datasets: for example, to mitigate security risks.

Enablers: Strong data culture and the development of capability & skills

The level of open data capabilities & skills present in an organisation and the culture built to support these.

Outcomes

Innovation

Innovation is driven through careful knowledge management, expertise development and sharing of high-value use cases of open data.

Decision Making

Open data is at the forefront of the organisation's decision making, with open data capabilities informing current operational decisions and strategy, prioritising the development of open data capabilities to maximise its future use.

Enabler

Strong Data Culture and the Development of Capability & Skills

Activities

Strategy

The organisation is considering open data as a key element of its overall organisational strategy, both incorporating open data into its strategic vision and by leveraging their current open data capabilities to inform operational decisions.

Strategic Oversight:

The organisation has clear lines of ownership of the open data strategy, with board-level endorsement, setting the stage for a high-level of support for open data and to drive its continued development.

Roadmap for Delivery:

The organisation has a clear roadmap for developing its open data capabilities, with targets set and reviewed / adjusted over time to continue to stretch the organisation in line with its strategy.

Enablers: Strong data culture and the development of capability & skills

The level of open data capabilities & skills present in an organisation and the the culture built to support these.

Outcomes

Innovation

Innovation is driven through careful knowledge management, expertise development and sharing of high-value use cases of open data.

Decision Making

Open data is at the forefront of the organisation's decision making, with open data capabilities informing current operational decisions and strategy, prioritising the development of open data capabilities to maximise its future use.

Enabler

Strong Data Culture and the Development of Capability & Skills

Activities

Advanced Analytics Development:
The organisation has enabled the development of innovative assets through fostering skills in advanced analytics: for example, machine learning, and adjusting their Business Intelligence (BI) model accordingly.

Open Data Resourcing & Expertise:
The organisation actively works to enhance its staff's open data knowledge, capabilities and skills.

Knowledge Management:
The organisation actively encourages the sharing of open data knowledge and experiences internally: for example, through internal group presentations and knowledge transfer policy to prevent knowledge loss during operational changes.

Enablers: Improved collaboration on open data across the industry

The level of cross-company and cross-sector collaboration, encompassing a culture of feedback and knowledge sharing will drive forward open data.

Outcomes

Innovation

Knowledge sharing is leveraged to highlight the value derived from open data and inspire new and innovative use cases.

Transparency

Sharing of information and insights across the industry can help increase public awareness and increase accountability of the water sector to its customers & communities, the government and the regulator.

Enabler

Improved Collaboration on Open Data Across the Industry

Activities

Open Data Documentation:

The organisation releases high-quality metadata / documentation with its published resources. Sufficient governance is in place alongside the development of this documentation to ensure that it is in line with the organisation's common standards.

Feedback:

A well established feedback loop is in place, allowing the organisation to collect feedback on its published resources and make appropriate changes to meet user needs. In addition, the organisation provides feedback to external dataset providers to help shape the external open data resources into meeting their user needs.

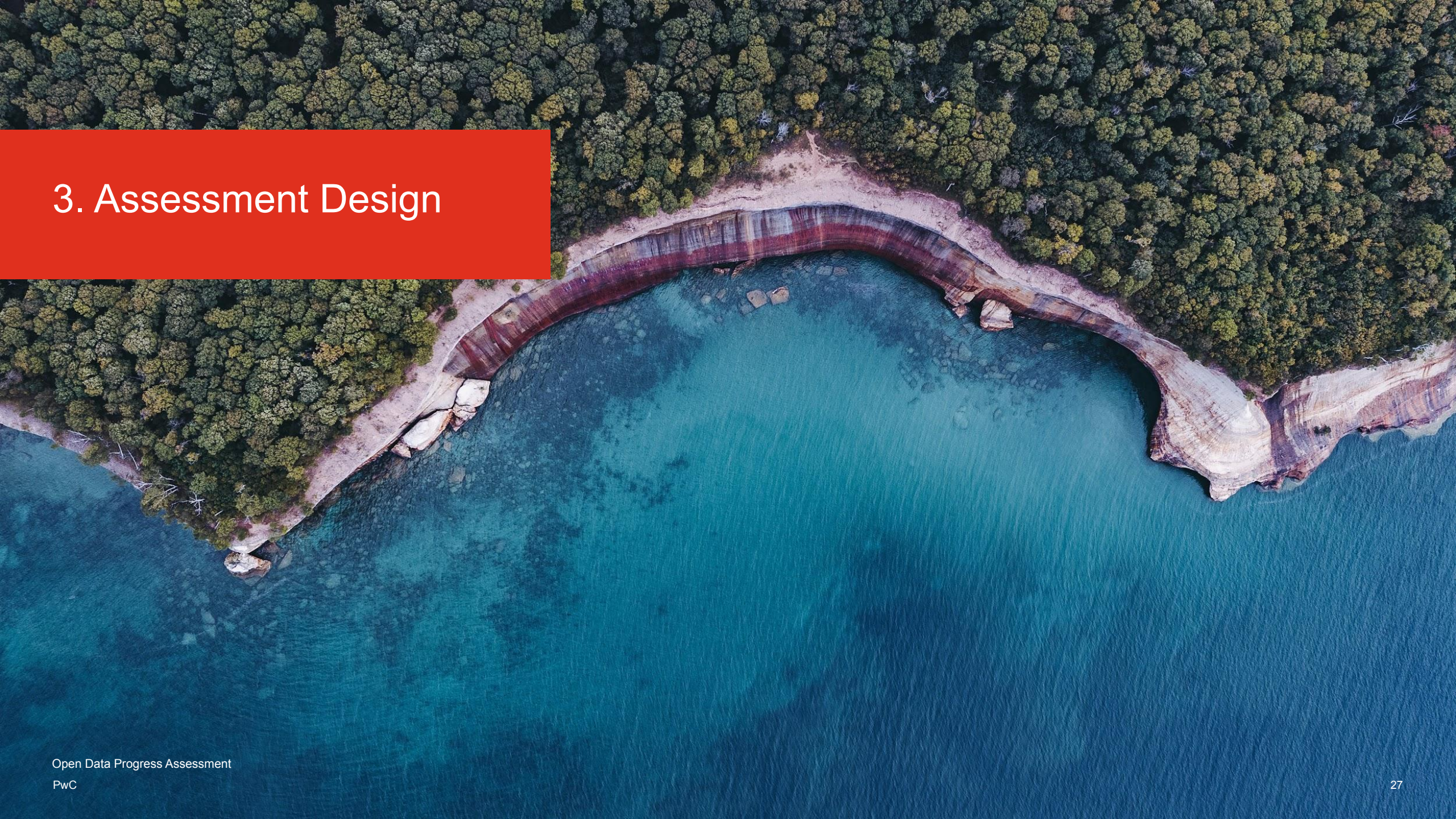
Prioritisation:

The organisation works to increase the engagement users have with its published resources by identifying and prioritising high-value datasets. Prioritisation can be established routinely tracking metrics related to their open data, collaboration with the user community and highlighting the value the organisations published resources can provide.

Knowledge Sharing & External Support

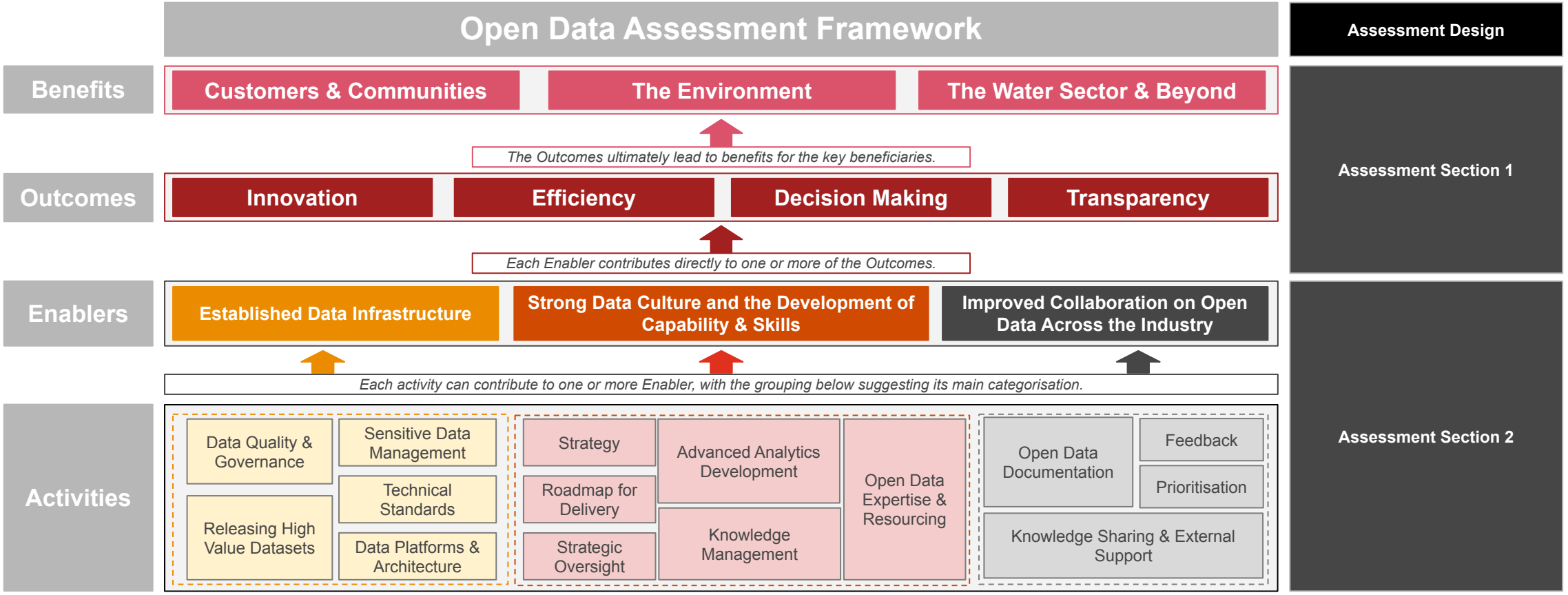
The organisation shares its open data knowledge externally, within the water sector and beyond, to highlight the value derived from open data and inspire new and innovative use cases. In addition, the organisation provides information and training sessions with potential dataset users to ensure they are aware of the open data that exists and are supported in its use.

3. Assessment Design



Open data progress framework: assessment design

The assessment has two sections. The first section asks open ended questions aligned to each of the Outcomes, drawing out examples of good practice. The second section is focused on the Enablers/ Activities, looking to understand the progress that has been made in facilitating Outcome achievement.



Open Data Progress Assessment
PwC

Note, Section 3 in the assessment questionnaire providing space to provide additional information. Questions asked can be found here.

Open data progress framework: assessment design

The assessment looks to understand the progress made in open data, as well as draw out examples of good practice and areas for improvement. Section 1 draws out key case studies demonstrating value to the beneficiaries. Section 2 uses the progress criteria to understand the progress that has been made and set a baseline for future assessments.

	Assessment Method	Question Design	Assessment Output
Assessment Section 1 *(Outcome Aligned Questions)*	• Assess each response on the quality of evidence provided	• 1 long format question per Outcome. *(200-300 Words)* • Supporting evidence with commentary describing relevance	• Examples of best practices to be shared with industry
Assessment Section 2 *(Activity / Enabler Aligned Questions)*	• Assess each response against its associated progress statements • Please see Progress Criteria section for more details	• 1 shorter format question per Activity. *(100-200 words)* • Supporting evidence with commentary describing relevance	• Progress statement by activity • Examples of good practice • Areas for improvement

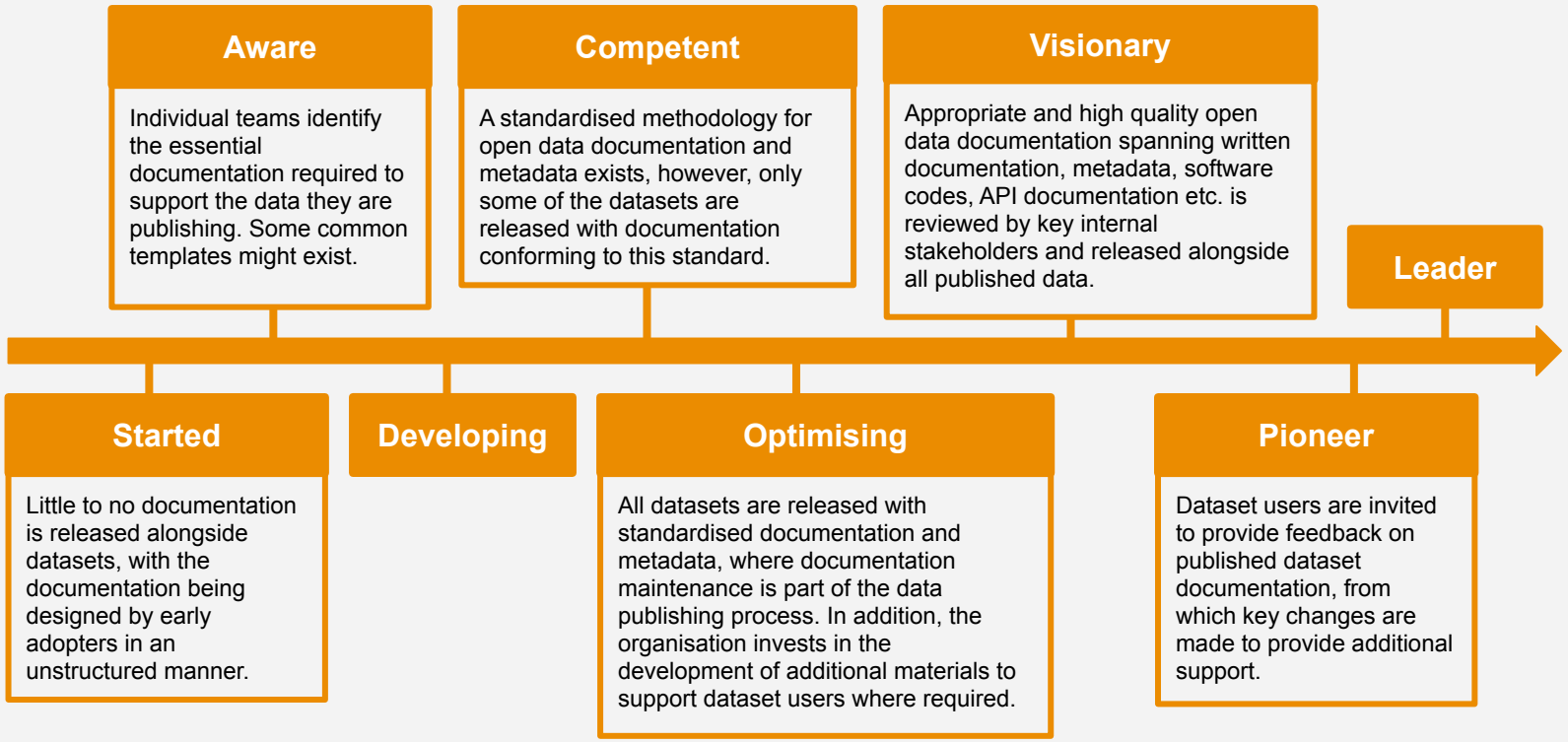
Note, a Section 3 can be seen in the assessment questionnaire providing space to provide additional information. Questions asked here can be found here.

Please click here for full assessment criteria

Progress statements helped map out what good looks like

To navigate the shared problem of conducting a progress assessment with no baseline to work from, we devised eight progress statements for each activity to define the journey from just starting to becoming a cross industry leader.

The following is an example showing the progress statements for Open Data Documentation:



Key Design Considerations

- The progress statements were designed with cross-industry best practice in mind. Hence, they provide a method of assessing the water sector against examples of best practice across all industries.
- The progress statements provide a benchmark, enabling a repeatable assessment to measure future progress.

Note: Not all activities have a progress statement aligned with all 8 points on the scale, however, if a company response indicates it lies between two non-adjacent progress statements, then the appropriate statement from the full 8 point scale will be used to summarise their performance.

The evidence attached to assessment responses was evaluated and datasets were categorised as either shared or open (1 of 2)

PwC evaluated evidence to make sure it met the required standard, and datasets provided were categorised to be either open or shared based on the Open Data Institute (ODI) definition for open data.

Assessment Approach & Evidence Evaluation

Companies provided written response and attached evidence for each assessment question → **PwC evaluated the assessment response and evidence** → **One-pager progress summaries created based on company response and evidence**

- Potential sources of evidence were provided for each assessment question to guide companies on appropriate evidence expected to support their response.
- If evidence against a particular response was deemed insufficient, the case study or the reported dataset has not been mentioned in this report. For instance,
  - where companies have failed to provide any supporting evidence or details about the process or outcome from a given activity in their response.
  - where companies have provided broken links to datasets/supporting evidence published online.
- No additional material outside of the assessment response or supporting evidence was used in the evaluation of companies' progress on open data.
- Industry learnings were devised through aggregating the one-pager progress summaries to understand where the industry as a whole has progressed faster / where it needs to focus its efforts.

Assessment of Progress since H2Open: Companies were informed through the industry workshops and upfront in the framework questionnaire, that the assessment aims to understand the progress companies have made since H2Open. However, the individual questions did not ask about the timeframe for each Open Data related activity/case study that company submits as evidence. As a result, the responses submitted by companies includes applications of Open/Shared data from years preceding the publication of H2Open.

This assessment therefore provides an evaluation of companies progress to date, and establishes a baseline to track progress against in the future. Additionally, where possible, we have highlighted case studies that have either started or made further progress over the past year.

The evidence attached to assessment responses was evaluated and datasets were categorised as either shared or open (2 of 2)

PwC evaluated evidence to make sure it met the required standard, and datasets provided were categorised to be either open or shared based on the Open Data Institute (ODI) definition for open data.

What is open data?

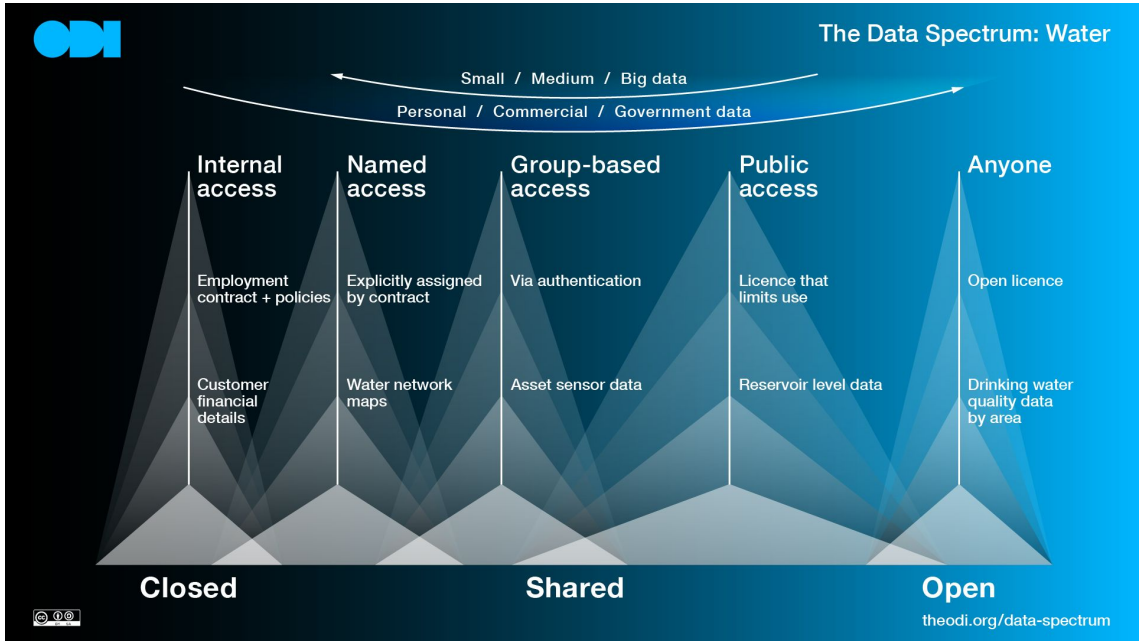
- We have adopted the Open Data Institute's definition for open data, built on the principles developed by the Open Knowledge Foundation
- The ODI defines open data as **data that anyone can access, use and share**
- The Open Knowledge Foundation, whose principles inform the ODI definition, states that open data must be made **available in common, machine - readable formats** i.e. provided in a form readily processable by a computer and where the individual elements of the work can be easily accessed and modified by users
- Open **data must be licensed**. Its licence must permit people to use the data in any way they want, including transforming, combining, and sharing it with others, even commercially

What is shared data?

- Shared data is data made available to specific groups of users
- Access is restricted to specific users via authentication

The Data Spectrum

- The diagram on the right shows the data spectrum developed by the Open Data Institute
- The spectrum is a useful tool in understanding the key differences between open and shared data



Assessment of Water Company Datasets

- The datasets provided by the water companies on the assessment were categorised into two groups - open or shared.
- Where data was categorised as shared, it was typically due to having restricted access (i.e. not public) or being published in a format that is non-machine readable e.g. pdfs. Our dataset review process first investigated whether the dataset submitted by companies were publicly accessible or not. If a dataset was publicly accessible, a further check was conducted to review if the dataset was machine readable.
- Barring Thames Water's API portal, which provided a data license, an open data license was not discoverable for any of the datasets provided by the water companies as part their response to this framework.
- Therefore, by the definition of ODI, only Thames Water's EDM data published on their API portal will be considered as 'Open'. However, to highlight the number of datasets published by water companies that are **openly accessible** and **published in a common, machine-readable format**, **we categorise any such dataset as open despite a lack of a supporting open data 'license'**.

4. Industry Learnings

Our assessment shows that where benefits are being delivered to key beneficiaries, this is typically only through shared data

Water companies have shared data with organisations both within the water sector and beyond to create solutions that provide benefits to the key beneficiaries identified in the assessment framework - Customers & Communities, the Environment, and the Water Sector & Beyond. However, there has been limited evidence of open data releases and hence few benefits achieved from it.

Customers & Communities

Customers & Communities, benefit in a number of ways; including:

- **Identification of financially vulnerable customers** and providing targeted services/tariffs **through shared data agreements** between water companies, Department for Work & Pension (DWP) and other utilities companies
- Applications focusing on **sharing maintenance, leakage and water quality information**, e.g. South East Water's In Your Area (IYA) service, are providing increased transparency for customers
- **Near real-time alerts on storm release activity** for example, Southern Water's Beachbouy App*, where alarms and sensors trigger alerts to the public if there is a storm release event

The Environment

In response to the current issues facing the water sector, the environment has been a key focus area.

- All wastewater companies have **released Event Duration Monitoring** (EDM) data (please see slide 96 for details). Two companies have taken the next step of making their Event Duration Monitoring (EDM) data more user-friendly. Anglian Water and Thames Water have **released the locations of their storm overflows** alongside their EDM data on **interactive maps** on their websites.
- **Wessex Water's Intelligent Sewers Competition***, led to the development of a predictive analytics tool in collaboration with StormHarvester that can identify potential sewer blockages earlier. The solution identified >80 blockages over the past year.

The Water Sector & Beyond

Cross-sectoral collaboration with the water sector on data-sharing initiatives is providing benefits to the whole economy

- **Streetworks Monitoring and Evaluation tool** developed by Affinity Water in collaboration with Greater London Authority (GLA) has helped in reducing multiple road closures/diversion through sharing of data on planned works programmes with other utilities.
- Initiatives like the **Sewer CCTV AI project***, with multiple companies contributing CCTV footage, help to establish a sector wide repository of CCTV labelled images that companies can use to train their own AI models. These models **replace the manual process of coding defects with an automated solution thereby providing up to 20% of time savings for survey and coding operators**.

Note: Case study examples highlighted with an asterisk (*) were also mentioned in the H2Open publication (link)

There have only been a small number of open datasets released

Companies have achieved benefits, through varying outcomes, primarily through sharing data with key partners or by merging external third party open datasets with their internal data. A common understanding of the value of open data now exists across the industry but few open datasets have been released.

Regular data publications have led to trialling of methods to release data: Companies have used regular data publications to trial methods of releasing data (e.g. Annual Performance Report (APR) and EDM data, see slide 96 for further details), but not all companies have fully met the requirements for Open Data; for instance, by not releasing supporting metadata alongside APR data releases. However, these publications have prompted some companies to trial methods of releasing data that have the potential to facilitate additional data releases in the future. (See evidence on slide 36)

Companies are open data users more than open data providers: Companies are commonly creating benefits for the key beneficiaries by merging data from outside their company with their own data, i.e. they are keen users of open data but have made limited progress in opening of their own datasets. (See evidence on slide 36)

Data is shared rather than making it publicly accessible: When releasing data, companies are generally sharing data with key partners over releasing open datasets with unrestricted public access. Of the datasets referenced by the companies on their assessment responses, 39% (81 of 210) datasets were deemed to be shared due to being restricted to group-based access as per the ODIs open data definition (slide 10).

There is a lack of machine-readable open data: Whilst there are some data releases beyond APR and EDM data, these are commonly not done in a machine readable format. For example, through interactive maps with no way of downloading the data are commonplace. Of the datasets referenced by the companies on their assessment responses, 26% (54 of 210) datasets were deemed to be shared due to not being in machine readable format.

Despite not always fully meeting data release requirements, the progress made by companies should help with future releases

In terms of releasing open datasets, companies have generally focused on meeting the APR & EDM data publication requirements. Trying to meet these requirements has prompted companies to trial and develop capabilities and solutions that will be useful for additional data releases in the future.

Outcome Level Learnings

Regular data publications have led to trialling of methods to release data

Companies have used regular data publications to trial methods of releasing data (e.g. Annual Performance Report (APR) and EDM data, see slide 96 for further details), but not all companies have fully met the requirements for Open Data; for instance, by not releasing supporting metadata alongside APR data releases. However, these publications have prompted some companies to trial methods of releasing data that have the potential to facilitate additional data releases in the future.

Companies are open data users, not providers

Companies are commonly creating benefits for the key beneficiaries by merging data from outside their company with their own data. I.e. they are keen users of open data whilst being limited in what they make open themselves. These activities are seen more with larger companies, typically the Water and Sewage Companies (WaSCs), who have more developed advanced analytics capabilities.

Supporting Evidence

- Thames Water's generic Open Data API, whilst still in its infancy and only currently being used for EDM data, shows a **method of releasing additional datasets in a highly usable and standardised manner.**
- Wessex Water has invested in developing a sewer monitoring alarms systems ('Intelligent Sewers') that uses **near-real time feed of EDM date** to provide **predictions and alerts on blockage formations.**

- Please see company one-pagers for multiple examples of how external datasets (e.g. Met Office data, Open Banking data and Environment Agency data) are being used to deliver benefits for the key beneficiaries.

Companies are prioritising sharing data over publishing open data

Of the 210 datasets reported by the companies, 65% were categorised as shared data due to having either restricted access or not being in a machine readable format. The majority of the remaining 35% of datasets (the open datasets*) cover EDM data and/or APR tables.

	Group-based shared access over full public access	Lack of machine-readable open data
Outcome Level Learnings	When releasing data, companies are generally sharing data with key partners over releasing open datasets with unrestricted public access. Companies tend to see this as an important step on the journey towards open data, however, outside of a few explicit examples (e.g. United Utilities releasing their Peat Depth as open data after sharing it with Natural England), it is unclear where shared data has led to the release of open data.	Whilst there are some data releases beyond APR and EDM data, these are commonly not done in a machine readable format. For example, through interactive maps with no way of downloading the data are commonplace. Data users, particularly those looking to create products and solutions using the data, would benefit from moving towards more machine readable content in the future.
Supporting Evidence	• Of the datasets referenced by the companies on their assessment responses, **39% (81 of 210**) datasets were deemed to be shared** due to being restricted to group-based access as per the ODIs open data definition. • There are multiple examples of benefits derived from shared data in the response highlights by company section of this report. For example, Affinity Water's work with the Greater London Authority, Cadent Gas and Openreach, in which they shared data on planned works.	• Of the datasets referenced by the companies on their assessment responses, **26% (54 of 210) datasets were deemed to be shared** despite being publicly accessible due to not being in machine readable format. • There are multiple examples of datasets that have been released in a non machine readable format across the assessment. For instance, sewer boundary maps often released by Water and Sewage Companies (WaSCs).

*The datasets categorised as 'open' here includes those may not have a supporting open data license published alongside the data, as explained in slide 10

**This report only reviews whether a published dataset is available in a machine-readable format and does not investigate whether datasets that are restricted to group-based access are machine-readable. Therefore, it's possible that of the 83 datasets that were restricted to group-based access, some may or may not be shared in a machine-readable format.

Progress in developing open data Enablers has been limited across the water sector, with larger companies further along their open data journey relative to smaller water only companies (1 of 2)

The water sector is at the beginning of its open data journey, with the majority of companies being categorised as Started, Aware or Developing for most Activities.

Enabler Level Learnings

Overall:

- Assessment of the open data Enablers show that the **water sector is still early in its open data development**, with the majority of companies achieving between 'Started' and 'Developing' for most Activities.
- Companies achieving higher progress statements are typically larger and are further along on their digitalisation journeys. Hence, general **Enabler progress is likely driven by broader digitalisation efforts and not solely by open data initiatives.**

Established Data Infrastructure

- **Investment into Data Platforms & Architecture is the most common theme** across the assessment, with many companies currently working towards transitioning towards a cloud-based architecture. This investment is often part of a wider digital transformation with only a few companies making explicit consideration towards how open data may be facilitated on the platform in the future.
- **Releasing High Value Datasets is an area of relative weakness**, where companies have rarely made data open beyond publishing of APR and EDM data, and often fail to provide sufficient metadata or documentation to support releases.
- Some **companies reference that they are engaging with wider collaborative groups** to define/ adopt supporting frameworks and standards. For example, common Platforms, Data Governance Frameworks and common Technical Standards. However, limited evidence was provided by companies to show progress in these areas.

Heatmap showing the Companies' Progress Statements by Activity

		Started	Aware	Developing	Competent	Optimising	Visionary	Pioneer	Leader
Established Data Infrastructure	Data Governance	5	1	3	5	2	0	0	0
	Data Platforms & Architecture	2	0	3	6	4	1	0	0
	Data Quality	1	5	5	2	2	1	0	0
	Releasing High Value Datasets (APR Data)	3	5	5	3	0	0	0	0
	Releasing High Value Datasets (Immediate Challenges. E.g. EDM)	9	1	4	1	1	0	0	0
	Sensitive Data Management	3	10	1	2	0	0	0	0
	Technical Standards	5	2	3	3	3	0	0	0
Improved Collaboration on Open Data Across the Industry	Knowledge Sharing & External Support	2	3	2	8	1	0	0	0
	Open Data Documentation; Feedback	4	8	3	0	0	1	0	0
	Prioratisation	5	7	4	0	0	0	0	0
Strong Data Culture and the Development of Capabilities & Skills	Developing Advanced Analytics	5	5	2	2	2	0	0	0
	Knowledge Management	4	4	5	3	0	0	0	0
	Open Data Expertise & Resourcing	5	4	6	1	0	0	0	0
	Roadmap for Delivery	5	7	4	0	0	0	0	0
	Strategy; Strategic Oversight	3	7	5	1	0	0	0	0

Progress in developing open data Enablers has been limited across the water sector, with larger companies further along their open data journey relative to smaller water only companies (2 of 2)

The water sector is at the beginning of its open data journey, with the majority of companies being categorised as Started, Aware or Developing for most Activities.

Enabler Level Learnings

Improved Collaboration on Open Data Across the Industry

- Assessment responses indicated that most companies do not have **feedback loops between companies and data users and is an area of weakness** for water companies to improve on.
- **Knowledge Sharing & External Support is an area of relative progress**, with 8 companies scoring competent. This is driven by involvement with collaborative initiatives, however, outputs from these groups were difficult to verify.
- Evidence provided suggests that there are **no collaborative initiatives bringing all 16* water companies together** in relation to open data.

Data Culture and the Development of Capabilities & Skills

- **Strategy, Strategic Oversight & Roadmap for delivery is an area of weakness**, with only 7 companies reporting to have a general data strategy and only 5 of these reporting to having elements specific to open data contained within it.
- **Progress in Developing Advanced Analytics was more common in larger companies** - albeit it focused on the use of other's open data over efforts to open their own data - whereas smaller companies typically opt to leverage external capabilities to support them.

Heatmap showing the Companies' Progress Statements by Activity

		Started	Aware	Developing	Competent	Optimising	Visionary	Pioneer	Leader
Established Data Infrastructure	Data Governance	5	1	3	5	2	0	0	0
	Data Platforms & Architecture	2	0	3	6	4	1	0	0
	Data Quality	1	5	5	2	2	1	0	0
	Releasing High Value Datasets (APR Data)	3	5	5	3	0	0	0	0
	Releasing High Value Datasets (Immediate Challenges. E.g. EDM)	9	1	4	1	1	0	0	0
	Sensitive Data Management	3	10	1	2	0	0	0	0
	Technical Standards	5	2	3	3	3	0	0	0
Improved Collaboration on Open Data Across the Industry	Knowledge Sharing & External Support	2	3	2	8	1	0	0	0
	Open Data Documentation; Feedback	4	8	3	0	0	1	0	0
	Prioratisation	5	7	4	0	0	0	0	0
Strong Data Culture and the Development of Capabilities & Skills	Developing Advanced Analytics	5	5	2	2	2	0	0	0
	Knowledge Management	4	4	5	3	0	0	0	0
	Open Data Expertise & Resourcing	5	4	6	1	0	0	0	0
	Roadmap for Delivery	5	7	4	0	0	0	0	0
	Strategy; Strategic Oversight	3	7	5	1	0	0	0	0

*Note: Bristol Water and South West Water provided a joint submission in response to the framework questionnaire and have been considered as one for the purpose of this report.

Before looking at the detailed enabler findings, it is important to recognise there are several collaborative initiatives ongoing in the water sector

Companies made reference to collaborative projects they engage with as part of their assessment responses, the most common examples given are:

Project Name	Stream	Catchment Systems Thinking Cooperative (CaSTCo)
Project Ambitions	Stream originally set out to design and deliver the 'network of data pipes' needed to share useful industry datasets in a secure, standardised and easy to access way. More recently, however, they have re-prioritised to focus on developing user personas, designing technical architecture and agreeing common data formats. As of January 2023, Stream were in a 20-week blueprint phase involving 11 of the 16* water companies in England and Wales. See full list of participants below.	The Catchment Systems Thinking Cooperative (CaSTCo) aims aims to ensure there is a robust evidence base for tackling environmental challenges with direct support for local evidence gathering and community engagement in eight demonstration catchments, and a national framework of standardised tools and training. CaSTCo involves 12 of the 16 water companies in England and Wales. See full list of participants below.
Participants	**Led By:** Northumbrian Water **Participants:** Anglian Water Group, Thames Water, Severn Trent Water, Scottish Water, SES Water, United Utilities, South West Water, Dwr Cymru Welsh Water, Southern Water, Yorkshire Water, Sia Partners, Aiimi, Open Data Institute, Costain.	**Led By:** United Utilities **Participants:** The Rivers Trust, Thames Water Utilities Ltd, South West Water Ltd, Southern Water Services Ltd, Dŵr Cymru, Anglian Water Services Ltd, Severn Trent Water Ltd, Wessex Water, Yorkshire Water Services Ltd, Northumbrian Water Ltd, Affinity Water, South East Water, Earthwatch Institute, The Freshwater Biological Association, The Zoological Society of London, UK Centre for Ecology and Hydrology, Cardiff University, University of Exeter, Natural Course EU Life Integrated Project, Hummingbird Technologies

*Note: Bristol Water and South West Water are considered as one company in this report following the acquisition of Bristol Water by South West Water.

Enabler aligned learnings: established data infrastructure

Data Platforms and Architecture

Started	Aware	Developing	Competent	Optimizing	Visionary	Pioneer	Leader
2	1	3	6	3	1	0	0

Industry Assessment

- Investment into platforms and architecture is the most common theme across the assessment, with many companies currently working towards transitioning towards a cloud-based architecture. This investment is often part of a wider digital transformation and has a knock on effect to the level of progress seen elsewhere on the assessment.
- Water only companies are typically at an early stage in their digital transformation journey and as a result are closer to the beginning of their journey in this area.
- Where already in operation, full utilisation of cloud-based platforms and their features is typically still in its infancy. I.e. the data sharing or advanced analytics capabilities are not being used to their full potential.
- A common approach has been developed between Anglian, Northumbrian, Thames, Southern and Wessex Water to move towards using Azure Databricks to drive standardisation, cataloguing and provision of data for private, shared and open datasets.
- Several companies mention waiting for Stream to provide clarification on a common set of platforms to use when approaching opening their data.

Potential Next Steps

- Companies currently transitioning towards cloud-based architecture should complete this transformation whilst continuing to consider developing their open data capabilities. I.e. ensuring open data capabilities are a key requirement throughout systems design.
- Companies already using cloud-based architecture should look to explore the associated features most relevant to open data. E.g. data sharing, ETL, integrations and advanced analytics.
- Companies should use their existing systems to deliver open data capabilities.

Enabler aligned learnings: established data infrastructure

Data Governance

Started	Aware	Developing	Competent	Optimizing	Visionary	Pioneer	Leader
5	1	3	5	2	0	0	0

Industry Assessment

- Companies with limited progress in this area typically see data governance as a security-focused activity only. Typically, they are at the stage of having employed a Data Protection Officer (DPO), who is in turn in the process of/ has set up data governance policies and processes focused on security only.
- Companies that have a higher-level of strategic oversight also tend to have a more established data governance structure. For example, through the existence of a data governance steering committee, and certain roles and responsibilities such as data stewards and data owners.
- Companies that are part of stream reference that they are engaging with Stream to define a data governance framework.

Potential Next Steps

- A reasonable next step for those early on in their journeys as well as those that have progressed further should be to look to incorporate c-suite / board-level support for data governance, with sufficient KPIs in place for effective monitoring and tracking.
- Companies with limited governance currently in place should look to define roles and responsibilities associated with data governance. For example data stewards, data owners and to introduce data governance steering committees. These groups can then look to continually improve data governance practices over time.

Data Quality

Started	Aware	Developing	Competent	Optimizing	Visionary	Pioneer	Leader
1	5	5	2	2	1	0	0

Industry Assessment

- The primary focus across the industry has been on addressing data quality issues present in master data. Several companies are looking to invest in data quality management tools such as Azure Data Factory or Google's Dataplex to support these initiatives.
- Those that have made an average amount of progress on their data quality journeys are typically addressing data quality issues at the data sourcing stage.
- Companies showing the most progress have defined rules for data quality profiling at later stages of the data lifecycle. For example, Northumbrian Water look to address data quality lapses at the Extract Load Transform (ETL) stage.
- Companies that are part of stream reference that they are engaging with Stream to define a set of data quality standards.

Potential Next Steps

- Companies at the beginning of their data quality journey should look to develop processes to address data quality issues at the data gathering stage. With a reasonable next step looking towards incorporating sufficient data quality gates and opportunities to address lapses throughout the data life cycle.
- Companies should look to define and publish their own data quality standards, this can be used to determine whether they are meeting their own standards or not.
- Companies should not let data quality issues stop them from making data releases, and instead release data with supporting documentation explaining quality issues. Data user feedback loops can also be leveraged to improve data quality over time.

Enabler aligned learnings: established data infrastructure

Sensitive Data Management

Started	Aware	Developing	Competent	Optimizing	Visionary	Pioneer	Leader
3	10	1	2	0	0	0	0

Industry Assessment

- As required by the General Data Protection Regulation (GDPR), companies have specific roles and processes in place (i.e. data protection officer and data protection policies) to support with sensitive data management. Whilst these roles have a wider remit than just open data, they are seen as the logical owners of sensitive data management.
- At present, companies are generally focused on sensitive data management as being primarily focused on anonymisation, with less thought taken towards triaging internal/ external datasets to evaluate risks of sensitive data creation through merging of multiple data sources.
- Companies that are part of stream reference that they are engaging with Stream to define sensitive data management best practice.

Potential Next Steps

- If they are not already in place, companies should look towards developing key sensitive data management processes and policies and ensure they are visible and communicated to all employees.
- Companies should ensure that they are considering all forms of sensitive data when developing mitigation plans. I.e. not just personal data but also data that might have commercial, IP and national security implications.
- Companies should strive to have clear triage processes in place to ensure that data does not become sensitive through linking and merging, both considering internal and existing open datasets.

Technical Standards

Started	Aware	Developing	Competent	Optimizing	Visionary	Pioneer	Leader
5	2	3	4	2	0	0	0

Industry Assessment

- Companies earlier on in their journey are generally not considering technical standards at this stage, and are instead focusing of system design and implementation in line with their digitalisation journeys.
- Companies that have progressed further in this area are following internal technical standards supported by reasonable internal justification. For example, United Utilities follow ISO 19115 and 19157 standards for geographical data metadata and data quality standards.
- There is currently little agreement across the industry on specific common standards for open data, and companies commonly note that they are waiting for collaborative groups, such as Stream, to inform the industry on a set of common technical standards to adopt going forwards.

Potential Next Steps

- Companies should look to collaborate with other water companies and beyond the water sector to define or agree on a set of common standards that are suitable for all.
- Whilst waiting for agreement on common standards, companies should not be deterred from open data releases, but rather release data and collect feedback from data users to learn through iterations of data releases to determine which standards are most appropriate.

Enabler aligned learnings: established data infrastructure

Releasing High Value Datasets

Started	Aware	Developing	Competent	Optimizing	Visionary	Pioneer	Leader
12*	6*	9*	4*	1*	0	0	0

Industry Assessment

- Companies have published APR data on their websites in excel format, but most companies have not fully met the requirements; for instance, by not releasing supporting metadata alongside APR data releases.
- In line with new section 141C Water Industry Act 1991 companies have published annualised EDM data annually (from 2021) and are working towards making this data available in near real-time. Progress in this area will likely accelerate the development of the required infrastructure to support further open data releases in the future.e.g. Thames Water have developed a generic open data Application Programming Interface (API) which is currently used to publish EDM data, but also forms the basis for future data releases in a standardised manner, and provides data users with the ability for easy data refreshes; a significant step forward from hosting excel tables on their website and allows for more powerful products and solutions to be created by third party users.
- Certain companies are going further to release high value data to meet the current challenges faced by the water sector. For example, Anglian Water and Thames Water have both published an interactive map of sewer overflows on their website. Wessex Water created a Marketplace platform which is the central hub that hosts and provides open access to different datasets including APR and EDM data.
- Data publications are most commonly curated data as part of a product or solution rather than raw data for re-use. Some exceptions to that would include Thames Water's open data API and Wessex Water's Marketplace platform which enable re-use of data published on their platform.
- Those achieving the highest progress statements in this area are looking for ways to standardise their data releases. For example, Northumbrian Water are adopting the ODI Certification Scheme to ensure appropriate metadata and support is provided to data users.

Potential Next Steps

- Companies should look to develop a standardised organisation-wide process for the release of published resources, building in key checks to ensure that their quality meets data user needs.
- In addition to curated datasets incorporated in a product or solution, companies should strive to release raw datasets for use by 3rd parties.
- Companies should look towards adopting common practices such as applying the FAIR principle to data releases. I.e. it being Findable, Accessible, Interoperable and Reusable (FAIR).
- Companies making progress in this area should look to share their learnings with the industry and beyond to support the development of open data industry wide.

** The assessment contained two questions on releasing high-value datasets, hence the total scores add up to 32 for the 16 water companies.*

Enabler aligned learnings: improved collaboration on Open Data

Prioritisation

Started	Aware	Developing	Competent	Optimizing	Visionary	Pioneer	Leader
5	7	4	0	0	0	0	0

Industry Assessment

- Companies are typically making prioritisation decisions in-house and without external influence, commonly determined by their assessment of perceived potential use cases, and perceived-value vs. effort.
- Companies performing particularly well in this area engage with communities and data users to help inform prioritisation decisions. For instance, Severn Trent Group's work on the River Pledges, where they interviewed tens of stakeholders, categorised them into four groups and developed a set of requirements for each.

Potential Next Steps

- Companies should make sure that there is a defined and repeatable approach in place with regards to engaging with the dataset user community both before and after the release of data.
- Companies should consider tracking metrics related to the effectiveness of their data release, identifying ways in which they can measure the impact of their open datasets. This findings can then inform future prioritisation decisions.

Open Data Documentation

Started	Aware	Developing	Competent	Optimizing	Visionary	Pioneer	Leader
4	8	3	0	0	1	0	0

Industry Assessment

- In general, open data documentation is in its infancy across the industry, commonly being thought of as one of the final stages in completing a data publication. In these cases it is unlikely that a standardised methodology towards documentation being followed.
- Companies often refer to waiting on development of their data strategy to help define their approach to documentation.
- The more advanced companies have a defined approach towards releasing basic documentation. E.g. Thames Water have a dedicated documentation portal in which they provide field descriptions and types, as well as provide a communication channel through asking data users to register.

Potential Next Steps

- Companies towards the earlier stages of progress in this area should consider developing a standardised methodology for open data documentation and metadata.
- Companies that have made more progress here could begin to explore expanding their documentation to incorporate not only written documentation, but also metadata, software codes, API documentation etc.
- Companies looking to maximise their impact in this area could look towards creating a feedback loop with dataset users to continually improve the quality of their documentation.

Enabler aligned learnings: improved collaboration on Open Data

Knowledge Sharing and External Support

Started	Aware	Developing	Competent	Optimizing	Visionary	Pioneer	Leader
2	3	2	8	1	0	0	0

Industry Assessment

- Where companies are making the most progress in this area, they are commonly knowledge sharing as part of larger events. For example, Northumbrian Water's 'Innovate East' inaugural innovation festival.

Potential Next Steps

- Companies should be more transparent and share knowledge as they work on open data.
- Companies should look to ensure they are transparent and actively inform potential dataset users about data they have made open or is available to be shared.
- The industry should consider creating a single forum for all companies to engage with to help support knowledge sharing and progress.
- Companies should look to actively engage with open data initiatives across the industry and beyond to promote further collaboration. For example, hackathons, CaSTCo, Stream etc.

Feedback

Industry Assessment

Note, We did not have a specific question for 'Feedback' activity in our framework questionnaire, as collecting feedback/having feedback loops is expected across a range of activities e.g. data quality and governance, sensitive data management, knowledge management etc. Therefore no progress statement has been assigned against this activity.

- Activities that would benefit from a well defined feedback loop - for example, prioritisation, data quality and documentation - have typically not been well supported by a feedback loop between the companies and the data users to date.

Potential Next Steps

- Companies should consider how they can bring key stakeholders into processes associated with certain activities to ensure that they are meeting user needs and continually adapting to changing requirements.

Enabler aligned learnings: strong data culture and the development of capability & skills

Strategy

Started	Aware	Developing	Competent	Optimizing	Visionary	Pioneer	Leader
2	9	4	1	0	0	0	0

Industry Assessment

- Data strategy was seen to be a key indicator, along with strategic oversight, of progress across the rest of the framework as a whole.
- Companies towards the beginning of the progress statements for data strategy typically do not have a defined data strategy in place. Often citing its development as the next stage on their journey and something that will be included in their 2024 price review (PR24) submissions.
- Companies with a defined data strategy have generally included limited consideration towards open data. Exceptions to this trend do exist, however, such as Northumbrian Water who have defined a set of five guiding principles for open data initiatives and its future within their company.
- Companies that are part of stream reference that they are engaging with Stream to define open data strategies.

Potential Next Steps

- Companies that have not developed an open data strategy since the H2Open recommendation should look to prioritise its development.
- Companies should avoid relying on collaborative projects to define their data strategy and instead update it in the future if required.
- Companies should consider how open data can help them in achieving their organisational goals, thus aligning their open data strategy with their wider organisational strategy.

Strategic Oversight

Started	Aware	Developing	Competent	Optimizing	Visionary	Pioneer	Leader
2	9	4	1	0	0	0	0

Industry Assessment

- Companies scoring lowest here typically do not provide details of strategic oversight of key processes. Often this is an indicator of relative immaturity across other key activities such as strategy and governance.
- Companies performing particularly well in this area show ownership of key processes and business functions, such as data governance and or data strategy delivery. Commonly there is also the recent introduction of open data specific steering groups to support continued progress.
- Companies showing the most progress in this area have ensured senior level (C-suit or board level) strategic oversight. For example, Northumbrian Water Executive Leadership Team, chaired by their CEO, provides oversight for their open data strategy.

Potential Next Steps

- Companies should ensure that there is senior leadership level oversight and support of open data, ensuring it is a priority agenda item going forwards.
- If not already present, companies should look to define clear lines of responsibility and allocate sufficient budget towards the development/ delivery of their open data strategy.
- Companies should consider how they can track the delivery of their open data strategy and where possible tie its delivery to key executives performance.

Enabler aligned learnings: strong data culture and the development of capability & skills

Roadmap for Delivery

Started	Aware	Developing	Competent	Optimizing	Visionary	Pioneer	Leader
5	7	4	0	0	0	0	0

Industry Assessment

- The quality of companies' roadmaps for delivery of their open data strategy varies quite drastically. Companies that have not set an open data strategy yet, understandably, do not have a plan for delivery and tend to refer to their plans making an appearance in PR24.
- Companies which are further along on their digital transformation journeys as a whole tend to also have considered the delivery of open data more carefully.
- Companies often look towards stream as a vehicle for progressing medium and long-term open data ambitions.

Potential Next Steps

- Companies should develop a clear and accountable roadmap for the delivery of their open data strategy. I.e. A roadmap with clear time scales, accountability, targets and KPIs to help track them etc. Companies should also consider sharing these on their website.
- Companies should consider tying the delivery of their open data strategy to key executives performance.

Knowledge Management

Started	Aware	Developing	Competent	Optimizing	Visionary	Pioneer	Leader
4	4	5	3	0	0	0	0

Industry Assessment

- Knowledge management is most commonly thought of as something to be completed at the end of a project or when their is operational change. Limited structured processes are in place to manage this transfer of knowledge but rather relies on role-level knowledge being passed between individuals.
- Companies achieving higher levels of progress here typically have higher levels of data governance. As a result, data governance artefacts, amongst other materials, are transferred as part of a structured process when required.
- Several companies suggest that they are looking to invest in the development of a data catalogue in the near future to aid with knowledge management.

Potential Next Steps

- Companies should look to understand the data that they hold so as to be able to logically address knowledge management, amongst other activities. This could be achieved by Investment in a data catalogue.
- If not already in place, companies should strive to have a organisation-wide policy/ processes in place surrounding the capturing and transferring of open data knowledge. Considerations for the entire project or dataset life cycle would be accounted for rather than restricted these policies/ processes to the end of a project or where there is operational change.

Enabler aligned learnings: strong data culture and the development of capability & skills

Open Data Resourcing and Expertise

Started	Aware	Developing	Competent	Optimizing	Visionary	Pioneer	Leader
5	3	6	1	1	0	0	0

Industry Assessment

- Companies at the 'Developing' stage of progress in this area are typically providing ad hoc data literacy training to their employees. This training currently focuses on general data literacy with no specific reference to open data.
- Smaller companies, commonly Water Only Companies, with much smaller IT teams have stated that they are not in a position to focus on developing open data expertise through targeted resourcing.
- Companies that are further along in their journey, and whom have the resources to do so, are developing centralised data teams through targeted resourcing of specific skills associated with open data. For example, Southern Water are currently building a team of data architects, data engineers, data modellers and data scientists.

Potential Next Steps

- Companies earlier on in their open data journeys should look to develop centres of excellence of open data expertise. These centres of excellence can then be used as a platform to build common knowledge around open data across all levels of the company.
- Companies should look towards incorporating open data specific skills into their resourcing decisions when hiring new staff.
- Companies should develop and deliver open data literacy training to create greater understanding of open data across all levels of their company.

Advanced Analytics

Started	Aware	Developing	Competent	Optimizing	Visionary	Pioneer	Leader
5	5	2	2	2	0	0	0

Industry Assessment

- 7 companies do not provide details of an internal advanced analytics team, indicating that they are not yet at the stage of incorporating Advanced Analytics into their organisations.
- 10 companies in total report to have internal data teams (inc. Advanced Analytics capabilities).
- The companies most advanced in this area have reasonably well established Advanced Analytics teams. E.g. Severn Trent / Hafren Dyfrdwy with 10+ data engineers and data scientists. However, these teams focus on the development of products and solutions using a combination of their own data and external open data, rather than making their data openly available for public consumption.

Potential Next Steps

- In addition to developing advanced analytics products and solutions, companies should increase their efforts to open their own datasets up for others to use.
- Companies should review their procurement processes to ensure there is a low barrier to entry for external parties looking to engage with them in open data asset creation.
- Companies with the means to do so should continue to build out their advanced analytics skills through training and targeted resourcing.

There are lessons to be learned from sectors that are further along on their open data journeys

The transport and energy sectors have made significant progress on their open data journeys and can be looked towards for examples of a greater open data maturity. The development of cross-industry frameworks and centralised open data portals could help the water sector realise greater benefits.

Examples of good practice

Development of cross-industry frameworks and data sharing templates

- The **Rail Technical Strategy (RTS)** highlighted the transport sectors ambitions for improving access and sharing of data across the railway industry. An essential component was the **Information Management Framework (IMF)**, which included the development of a data standards route map, data-sharing templates and a framework for identifying 'high-value' datasets.
- **Developing cross-industry data sharing templates and prioritisation frameworks** has enabled the railway sector to move quickly to make optimum use of open data, and has supported the creation of the **Rail Data Marketplace** (RDM); which will combine fragmented sources of railway data to form one digital service.

Utilising data portals and APIs to release data

- Data portals in the energy sector not only provide a centralised source for data users to find multiple datasets, but also act as a feedback mechanism to request additional dataset. For example: the **National Grid's Connected Data Portal** or the **Open Energy Data Portal**.
- The rail sector is releasing data collected from various monitoring kits through an **Application Programming Interface (API)**. API development has promoted the standardisation and led to the development of third party products and services; for example, travel planner apps.

Collaborating with data users and utilising the feedback loop

- The **Open Data Communities** programme, run by the Open Data Institute, supports collaboration between data users and local authorities to identify and publish open data that is important to local communities. The initiative has been successful in providing useful data to local residents, businesses and organisations, leading to increased transparency and engagement..
- **OpenActive**, a community-led initiative which aims to help people get active using open data, has built an open data ecosystem of over 100 organisations across the sport and physical activity sector, including almost 70 data publishers. It has achieved this through active engagement with individuals working within the sport and physical activity sector.

Learnings for the water sector

Adoption of cross-industry frameworks in the water sector

- The water sector should look to **develop/ adopt similar cross-industry templates and frameworks** to accelerate their path towards a more open data enabled business model.
- The companies involved in Stream and CastCo are working towards defining some of these frameworks. Further development of frameworks across the water sector with sufficient external collaboration would help the water sector progress further.

Developing data portals and APIs to release data

- The development of **open data portals** for the water sector could accelerate innovation by centralising and signposting the data currently available to potential data users. A data portal could also support the development of an established feedback loop.
- Thames Water have developed the first **generic open data API** seen in the water sector. Greater activity in this area, and **sharing of learning**, will see the water sector make considerable progress in open data and support the creation of apps and services in the future.

Engage with data users to identify priority datasets and to incorporate feedback

- Water companies should look to engage with communities and data users to help identify priority data sets. Companies can benefit from having a defined and repeatable approach in place to gather feedback from their data releases.
- Wessex Water's Marketplace platform provides a good example of how water companies can both share data and solve common challenges in collaboration with data users.

5. Company Assessments

Guide - How to interpret our One-Pager Summaries

Key findings brought out in a summary subheading.

Case Study

Example Case Study Heading

Details of case study. This case study will be an example of good practice pulled out from the assessment response.

Areas of Most Progress

Area of most progress 1
Details of how the company has made progress. This progress area will be aligned to an activity.

Area of most progress 2
Details of how the company has made progress. This progress area will be aligned to an activity.

Area of most progress 3
Details of how the company has made progress. This progress area will be aligned to an activity.

Guide on how to interpret the datasets table

- The information to update this table is obtained from company responses to Q21 (see slide 85) which asks for evidence of all open/shared datasets that have been published and datasets the companies have committed to publish.
- The evidence provided by companies in response to this question is reviewed against the Open Data definition (see slide 10) to determine whether it is considered as Open/Shared. As explained in slide 10, a dataset is **considered as open if it is publicly accessible and published in a common, machine-readable format**, even if such a dataset does not include a supporting 'open data' license.
- Note if a company has provided a single link to their company website containing various Environmental Performance, then all files listed on that website are considered as a single dataset unless the company has listed each dataset separately in their evidence submission.

Enabler	Activity	Progress Statement
Established Data Infrastructure	Data Governance	Progress Statement Achieved by Company against Each Activity
	Data Platforms & Architecture	
	Data Quality	
	Releasing High Value Datasets	
	Releasing High Value Datasets 2	
	Sensitive Data Management	
	Technical Standards	
Improved Collaboration on Open Data Across the Industry	Knowledge Sharing & External Support	
	Open Data Documentation; Feedback	
	Prioratisation	
Strong Data Culture and the Development of Capabilities & Skills	Developing Advanced Analytics	
	Knowledge Management	
	Open Data Expertise & Resourcing	
	Roadmap for Delivery	
	Strategy; Strategic Oversight	

Open Datasets	Shared Datasets	Committed Datasets
# Open	# Shared	# Committed to release

Click here to see definition for Open and Shared data

Affinity Water Assessment Response Highlights

Affinity Water are yet to make any data open beyond publishing of APR data, primarily sharing data instead of making their data open at this stage. However, they have laid important foundations for future open data initiatives by investing in key cloud-enabled platforms and setting plans to formalise their open data strategy in PR24.

Case Study

Reduced road closures, journey times and carbon impact by sharing planned works programmes

Affinity Water, along with the Greater London Authority, Cadent Gas & Openreach, won the Best Collaborative Works Award at the Street Works UK Awards 2022 for a project in which they shared data on planned works programmes. The project ultimately enabled the coordination of activities and reduced the negative impact of planned works. For example, by reducing road closures/diversions, the impact on residents, and journey times, as well as leading to a smaller carbon footprint.

Areas of Most Progress

Investing in Platforms & Architecture
Affinity Water have invested in updating their information architecture and platforms, where they are transitioning from legacy systems to a Data Lake and associated Lake House architecture. Standardised integration tooling has been selected to merge data from various systems and Qlik, the analytics platform, has been selected to support analytics development. Affinity water plan to develop consumable APIs in the future.

Creating a Data Catalogue to support Knowledge Management
Currently Affinity Water are using MS Teams and Atlassian's Confluence Wiki Pages for process and knowledge sharing, with data governance artifacts that provide ownership down to the attribute level. To further develop in this area, Affinity Water are in the process of developing a Data Catalogue platform that will provide knowledge throughout the entire lifecycle of their data.

Establishing internal data experts and advanced analytics capabilities
Affinity Water have been building a team skilled in IT transformation internally. This has led to the development of internal analytics capabilities, whereby the team has solved important problems such as using machine learning (ML) to predict Per Capita Consumption (PCC)[1] in real-time across the network.

Enabler	Activity	Progress Statement
Established Data Infrastructure	Data Governance	Started
	Data Platforms & Architecture	Competent
	Data Quality	Aware
	Releasing High Value Datasets	Started
	Releasing High Value Datasets 2	Started
	Sensitive Data Management	Aware
	Technical Standards	Started
Improved Collaboration on Open Data Across the Industry	Knowledge Sharing & External Support	Developing
	Open Data Documentation; Feedback	Aware
	Prioratisation	Developing
Strong Data Culture and the Development of Capabilities & Skills	Developing Advanced Analytics	Aware
	Knowledge Management	Started
	Open Data Expertise & Resourcing	Started
	Roadmap for Delivery	Started
	Strategy; Strategic Oversight	Started

Open Datasets	Shared Datasets	Committed Datasets
1	9	4

[1] Note: PCC forecasting case study was also mentioned by Affinity Water in the H2Open engagement with Ofwat (link)

Anglian Water Assessment Response Highlights

Anglian Water have made several datasets open, particularly those associated with their PR19 submissions and EDM data. However, they are still primarily realising benefits through data sharing agreements over open data. They have made good progress in developing their enablers, investing in data infrastructure and tools for managing data quality.

Case Study

Supporting land owner's decision making through sharing of abstraction management data

Having built weather stations across the region for use in abstraction management, Anglian Water have shared the data collected by them with Rhiza, a digital agronomy and precision farming service. Rhiza have taken the data and enriched it with localised weather data, amongst other open data sources, to ultimately create an app that supports determining when there are optimal crop spraying conditions.

In addition, the app helps increase the landowner's understanding of the impact they are having on the quality of the watercourse and how they can better manage the volume of water they need to take, thus reducing the impact downstream.

Areas of Most Progress

Developing a future focused digital ecosystem
Anglian water has invested in a logical Enterprise Data Architecture Model that defines, catalogues and models relationships across data assets. An Azure Data and Cloud platform is complimented by a number of associated platforms (Databricks, SAP BTP & Esri ArcGIS etc.), enabling the secure integration, hosting and visualisation of data, as well as to enable specific tasks such as machine learning, Enterprise Resource Planning (ERP) data-driven hosting and geospatial modelling.

Adopting external technical standards to enhance interoperability
Anglian Water follow an 'Integrate, reuse and adapt' approach when integrating external data by adopting their standards to enhance interoperability. For example, they use Unique Property Reference Number (UPRN) data, the common standard, as a reference when sharing property data with Utilities or Councils.

Leveraging Data Governance structures to support Data Quality profiling
Anglian Water has a well defined, hierarchical structure with defined roles & responsibilities for delivering data governance on priority data sets, including data owners and stewards who follow an agreed corporate quality and cleansing process where data is measured using agreed quality dimensions. In addition, Anglian Water uses the Enterprise Data Architecture Model to maintain data quality by identifying data definitions and relationships to processes and discrepancies as new data systems are built or integrated.

Enabler	Activity	Progress Statement
Established Data Infrastructure	Data Governance	Optimising
	Data Platforms & Architecture	Optimising
	Data Quality	Visionary
	Releasing High Value Datasets	Started
	Releasing High Value Datasets 2	Started
	Sensitive Data Management	Competent
	Technical Standards	Optimising
Improved Collaboration on Open Data Across the Industry	Knowledge Sharing & External Support	Competent
	Open Data Documentation; Feedback	Aware
	Prioratisation	Developing
Strong Data Culture and the Development of Capabilities & Skills	Developing Advanced Analytics	Optimising
	Knowledge Management	Competent
	Open Data Expertise & Resourcing	Competent
	Roadmap for Delivery	Aware
	Strategy; Strategic Oversight	Developing

Open Datasets	Shared Datasets	Committed Datasets
30*	15	0

*Note: 22 of 30 datasets listed as Open are associated with PR19 submission documents published in September 2018.

Open Data Assessment Framework
PwC

Dŵr Cymru Water Assessment Response Highlights

Dŵr Cymru have not made any data open beyond publishing of EDM data, instead focusing on sharing data and making use of their advanced analytics capabilities. In addition, they have developed their open data enablers, particularly Established Data Infrastructure, and are driving understanding of open data across their company.

Case Study

Sharing Sewer CCTV labelled images to train more accurate AI models[1]

Dŵr Cymru ran a project that used CCTV labelled images, captured from sewer surveys, to train an AI model capable of detecting pipework defects. The project showed that it is possible to replace manual processes with an automated solution, enabling operational colleagues to spend the time on more important tasks.

Since the project, Dŵr Cymru has been working with (and sharing data with) the Water Research Centre (WRC) who are leading a project to establish a sector wide repository of sewer CCTV labelled images that water companies can use to train their own AI models.

Areas of Most Progress

Building knowledge of open data at all levels of the company
In April 2022 Dŵr Cymru launched the Dŵr Data Academy, their internal data literacy initiative intended to support the learning and development of data skills across the entire organisation. The initiative has delivered two training modules – one of which provided opportunity for training participants to hear more about their Strategy – which includes 'Open Data' as a key objective.

Sharing knowledge to develop internal capabilities
Dŵr Cymru has held multiple knowledge share sessions with data teams from other water companies, focusing on analytical solutions and practices, and providing colleagues the platform to discuss ways of working and to share technical knowledge around certain projects and challenges. In addition, Dŵr Cymru has strong partnerships in place with several Universities based in Wales, which has enabled professional development opportunities for colleagues through tailored training and BSc/MSc degrees.

Creating formalised data governance structure to support data quality initiatives
Dŵr Cymru has formalised 10 data domains (e.g., customer, asset, finance etc.), with each domain having a nominated Data Owner, and several Data Stewards, who act as champions for the data governance initiative. Data Owners are accountable and responsible for data governance activities undertaken in their domain, and for reporting on progress to the Data Governance Group. Data Stewards typically work with data on a day-to-day basis, establish requirements, and continuously assess data quality associated with their data domain.

Enabler	Activity	
Established Data Infrastructure	Data Governance	Competent
	Data Platforms & Architecture	Optimising
	Data Quality	Competent
	Releasing High Value Datasets (APR Data)	Developing
	Releasing High Value Datasets (Immediate Challenges. E.g. EDM)	Developing
	Sensitive Data Management	Aware
	Technical Standards	Competent
Improved Collaboration on Open Data Across the Industry	Knowledge Sharing & External Support	Competent
	Open Data Documentation; Feedback	Aware
	Prioratisation	Aware
Strong Data Culture and the Development of Capabilities & Skills	Developing Advanced Analytics	Optimising
	Knowledge Management	Competent
	Open Data Expertise & Resourcing	Aware
	Roadmap for Delivery	Aware
	Strategy; Strategic Oversight	Aware

Open Datasets	Shared Datasets	Committed Datasets
1	51	5

[1] Note: Sewer CCTV case study was also mentioned by Dŵr Cymru in the H2Open engagement with Ofwat (link)

Hafren Dyfrdwy Assessment Response Highlights

HAFREN DYFRDWY severn dee

Hafren Dyfrdwy have primarily focused on shared data over open data and are active consumers of open data published by other companies. They have laid the foundations for open data by investing in the key technologies and architecture.

Case Study

Supporting vulnerable customers through establishing data sharing agreements

Hafren Dyfrdwy, as part of the Severn Trent group, have supported 6600 customers over the past 12 months by establishing data sharing arrangements with 12 external organisations; such as Equifax, Notts Energy Partnership and Severnside housing. These partnerships have allowed a dynamic view of customers' financial status and supported the early identification of vulnerable customers, providing the opportunity to step in and provide support as early as possible.

In addition, to tackle the risks associated with lead service pipe contamination, Hafren Dyfrdwy have collaborated with Sheffield University to use property data, housing records and other geographical information to predict the likelihood of properties being at risk of contamination.

Areas of Most Progress

Laying the data architecture foundations
Hafren Dyfrdwy are in the process of laying important architectural foundations to support open data by migrating from on-premise data centers to Azure hosted centralised cloud architecture; 65% of their core applications have been migrated to date.

Developing Advanced Analytics using external Open/ Shared Data
With a team of 10 internal data engineers and data scientist, Hafren Dyfrdwy are building an Advanced Analytics team that is able to develop tools that bring benefits to the key beneficiaries. For example, as can be seen by their use of 3rd party data on trade sludge volumes, traffic and weather information to enable optimisation of sludge collection routes and reduce their carbon footprint.

Sharing knowledge through key innovation project and partnerships
Hafren Dyfrdwy have been actively engaging with the water sector and beyond to better their understanding of open data and accelerate their progress. For example, they have been actively contributing to the Stream open data initiative for the past 2 years to exchange learnings and collectively investigate opportunities and issues with sharing water data. In addition, they have been engaging with other more developed sectors (UKPN), framework partners (Atkins), software vendors (Microsoft) etc. to help define their open data strategy.

Enabler	Activity	Progress Statement
Established Data Infrastructure	Data Governance	Competent
	Data Platforms & Architecture	Developing
	Data Quality	Aware
	Releasing High Value Datasets	Aware
	Releasing High Value Datasets 2	Started
	Sensitive Data Management	Aware
	Technical Standards	Competent
Improved Collaboration on Open Data Across the Industry	Knowledge Sharing & External Support	Competent
	Open Data Documentation; Feedback	Developing
	Prioratisation	Started
Strong Data Culture and the Development of Capabilities & Skills	Developing Advanced Analytics	Competent
	Knowledge Management	Developing
	Open Data Expertise & Resourcing	Developing
	Roadmap for Delivery	Aware
	Strategy; Strategic Oversight	Aware

Open Datasets	Shared Datasets	Committed Datasets
4	11	11

Open Data Assessment Framework
PwC

Northumbrian Water Assessment Response Highlights

Northumbrian Water are yet to make any datasets open beyond publishing of APR and EDM data, and have instead been primarily focused on sharing data and using other's open data. However, Northumbrian Water have set out to foster progress across the industry by enabling collaboration; for example through their Innovation Festival and their leadership of Stream.

Case Study

Supporting collaboration across the industry to solve problems

In July 2022 the Northumbrian Water Innovation Festival hosted the 'Big River' data hack. This hack, which was supported by the Stream initiative, sought to explore ways to publish storm overflow spills data in near real time whilst also looking at how to enhance that information for users by combining it with rich contextual data. This collaboration was important to overcome specific hurdles such as: 1. How to validate data quickly enough for near real-time publication? 2. How to combine data in a way that is easily accessible and understandable? 3. How can the combined data be used by Water company? and, 4. Can the combined data be used to assess the potential impact of spills. The data hack was attended by 46 participants from 24 different organisations from across public and private sectors along with two sector regulators.

Areas of Most Progress

Creating an open data strategy and fostering support from the top
Northumbrian Water have developed an open data strategy with 5 guiding principles to ensure there is continued progress on their open data journey. This data strategy, and its delivery, is overseen by an executive leadership team chaired by the CEO fostering support for open data at the highest level of the company.

Addressing Data Quality at the data discovery phase
Data quality is assessed at the data discovery phase as part of Northumbrian Water's formal analytics. Any quality lapses are communicated to the relevant data stewards and efforts are taken to iteratively improve the data collection methods. Quality lapses are subsequently addressed at the Extract, Load, Transform (ETL) stage. In addition, efforts are being taken to address master data quality issues.

Creating a platform for Knowledge Sharing
Northumbrian Water have provided a platform to bring together people and data at their Innovation Festivals. These events have provided the opportunity to further the conversation around open data and support its development across multiple industries; as can be seen by the case study outlined above.

Enabler	Activity	Progress Statement
Established Data Infrastructure	Data Governance	Competent
	Data Platforms & Architecture	Competent
	Data Quality	Optimising
	Releasing High Value Datasets	Developing
	Releasing High Value Datasets 2	Started
	Sensitive Data Management	Aware
	Technical Standards	Aware
Improved Collaboration on Open Data Across the Industry	Knowledge Sharing & External Support	Competent
	Open Data Documentation; Feedback	Aware
	Prioratisation	Aware
Strong Data Culture and the Development of Capabilities & Skills	Developing Advanced Analytics	Aware
	Knowledge Management	Aware
	Open Data Expertise & Resourcing	Developing
	Roadmap for Delivery	Aware
	Strategy; Strategic Oversight	Developing

Open Datasets	Shared Datasets	Committed Datasets
2	3	4

Portsmouth Water Assessment Response Highlights

Portsmouth Water are yet to make any data open beyond publishing of APR data, however, they have delivered value to their customers through sharing specific point-in-time datasets. In addition, they have bridged resourcing gaps, e.g. Advanced Analytics, by working with others whilst they look to develop their capabilities and infrastructure.

Case Study

Sharing data with Southern Water to support high-quality customer service

CON29DW is a report required for every residential property conveyancing transaction in England and Wales. Consisting of 28 question, it covers how a property is connected to the mains network (for waste and clean water), company owned assets within the property boundary and details of the risks and history of sewer flooding or low pressure. Southern Water (SRN), as the waste water services supplier for Portsmouth Waters' (PW) customers, need accurate and timely information from PW to ensure that they are able to complete CON29DW within the required 3-day turnaround time. Hence, enabling a good level of service to new customers. To support this, data requests are sent twice a day from SRN to PW. These requests are subsequently downloaded, the relevant data is extracted from several PW sources, and it is then returned to SRN along with the required asset/ boundary map.

Areas of Most Progress

Leveraging partnerships to deliver advanced analytics
As a smaller water-only company, Portsmouth Water do not currently have an in-house advanced analytics team. However, to support the delivery of benefits to their customers through fewer disruptions, the environment through reduced leakage and to themselves by enabling targeted maintenance, they have partnered with Inflowmatix and are in the process of developing the country's first digital twin network model.

Meeting with key stakeholder to support data prioritisation
Portsmouth Water regularly meet with a number of their key stakeholders (Havant Thicket, Portsmouth Uni, River Ems, CAB or Regulators) and use the actions from these meetings to inform which data should be shared. Hence, ensuring that the data they do provide is of the highest priority to the data-users.

Building open data requirements into investment decisions
Being early in their digital transformation journey, Portsmouth water are looking to invest in a corporate CRM system. Along with their pre-existing IT infrastructure, Portsmouth Water hope that these investments will better support the release of open data.

Enabler	Activity	Progress Statement
Established Data Infrastructure	Data Governance	Started
	Data Platforms & Architecture	Started
	Data Quality	Aware
	Releasing High Value Datasets	Aware
	Releasing High Value Datasets 2	Started
	Sensitive Data Management	Started
	Technical Standards	Started
Improved Collaboration on Open Data Across the Industry	Knowledge Sharing & External Support	Started
	Open Data Documentation; Feedback	Started
	Prioratisation	Aware
Strong Data Culture and the Development of Capabilities & Skills	Developing Advanced Analytics	Started
	Knowledge Management	Started
	Open Data Expertise & Resourcing	Started
	Roadmap for Delivery	Started
	Strategy; Strategic Oversight	Aware

Open Datasets	Shared Datasets	Committed Datasets
1	0	0

SES Water Assessment Response Highlights

SES Water are yet to make any data open beyond publishing of APR data. As a smaller company, they have been working to develop their platforms and architecture to support open data.

Case Study

Sharing incident management data with Thames Water to benefit SES' customers

As SES Water is a Water Only Company, Thames Water provide wastewater management to SES' customers. SES Water shares their Priority Services Register Customer data with Thames for Incident management. This allows Thames Water to contact SES' customers in the instance that there is an incident.

Areas of Most Progress

Laying the foundations for shared/ open data through platforms investment
SES Water have invested in a data warehouse hosted on the Azure platform. Tools such as PowerBI are used to share report versions of data.

Addressing data quality at the source
SES focus on addressing data quality at the source through data validation. To progress in this area they are looking to invest in a tool such as Aperture Data Studio to perform bulk data quality checks and perform data quality profiling.

Collaborating with others to expand their open data knowledge
SES Water state that they have not yet had much experience with opening up their data and rather look towards leveraging others experience to help support them. With this in mind, they feel their involvement with the Stream project will provide the much needed learnings to guide them on their journey.

Enabler	Activity	Progress Statement
Established Data Infrastructure	Data Governance	Started
	Data Platforms & Architecture	Competent
	Data Quality	Started
	Releasing High Value Datasets	Started
	Releasing High Value Datasets 2	Started
	Sensitive Data Management	Started
	Technical Standards	Started
Improved Collaboration on Open Data Across the Industry	Knowledge Sharing & External Support	Competent
	Open Data Documentation; Feedback	Started
	Prioratisation	Started
Strong Data Culture and the Development of Capabilities & Skills	Developing Advanced Analytics	Started
	Knowledge Management	Started
	Open Data Expertise & Resourcing	Started
	Roadmap for Delivery	Started
	Strategy; Strategic Oversight	Started

Open Datasets	Shared Datasets	Committed Datasets
1	0	0

Severn Trent Assessment Response Highlights

Severn Trent have laid the foundations for open data by investing in the key technologies and architecture to support open data's use within the company. They are active consumers of open data and whilst they have already made a number of datasets open, they indicate that they will be increasing their publishing activity in the near future.

Case Study

Leveraging open and shared datasets to improve operational efficiency and customer service

Severn Trent's.data science strategy includes using external shared data sets and combining them with internal operational data to mitigate enterprise risks, drive efficiencies and improve customer service. One of the notable solutions created was the sludge tanker optimisation model that used predicted sludge production rates with third-party data on trade sludge volumes, traffic and weather information to optimise sludge collection routes. This led to a reduction in deployed tankers, mileage travelled, and carbon footprint.

Another solution where the company shares its data with other water sector stakeholders, is their live sludge capacity tracker which is updated daily providing a 7-day forecast of treatment capacity at their key trading sites for treatment of third party waste. This has reduced number of manual enquiries and optimised trading site deliveries.

Areas of Most Progress

Laying the data architecture foundations
Severn Trent are in the process of laying important architectural foundations to support open data by migrating from on-premise data centers to Azure hosted centralised cloud architecture; 65% of their core applications have been migrated to date.

Developing Advanced Analytics using external Open/ Shared Data
With a team of 10 internal data engineers and data scientist, Severn Trent are building an Advanced Analytics team that are able to develop tools that bring benefits to the key beneficiaries. For example, as can be seen by their use of 3rd party data on trade sludge volumes, traffic and weather information to enable optimisation of sludge collection routes and reduce their carbon footprint.

Sharing knowledge through key innovation project and partnerships
Severn Trent have been actively engaging with the water sector and beyond to better their understanding of open data and accelerate their progress. For example, they have been actively contributing to the Stream open data initiative for the past 2 years to exchange learnings and collectively investigate opportunities and issues with sharing water data. In addition, they have been engaging with other more developed sectors (UKPN), framework partners (Atkins), software vendors (Microsoft) etc. to help define their open data strategy.

Enabler	Activity	Progress Statement
Established Data Infrastructure	Data Governance	Competent
	Data Platforms & Architecture	Developing
	Data Quality	Aware
	Releasing High Value Datasets	Aware
	Releasing High Value Datasets 2	Started
	Sensitive Data Management	Aware
	Technical Standards	Competent
Improved Collaboration on Open Data Across the Industry	Knowledge Sharing & External Support	Competent
	Open Data Documentation; Feedback	Developing
	Prioratisation	Started
Strong Data Culture and the Development of Capabilities & Skills	Developing Advanced Analytics	Competent
	Knowledge Management	Developing
	Open Data Expertise & Resourcing	Developing
	Roadmap for Delivery	Aware
	Strategy; Strategic Oversight	Aware

Open Datasets	Shared Datasets	Committed Datasets
9	15	9

South East Water Assessment Response Highlights

south east water

Whilst South East Water have published their water resources management plan (WRMP) and APR datasets, their primary focus has been on shared data to date. However, investments made in platforms and architecture, and their plans to develop an open data strategy may accelerate their progress.

Case Study

Building trust in their community through increased transparency

South East Water, in partnership with DigDat, have delivered an In Your Area (IYA) service to their customers and communities. By sharing information around planned works, streetworks, maintenance, reported leaks and events, all overlaid on an interactive map, customers are able to plan accordingly and ultimately reap benefits such as improved traffic management. Site visitors are also able to report leaks using the interactive map, supporting improved prioritisation of works.

In addition, visitors are able to view water quality data such as hardness levels and full water quality test results over the past 12 months, creating openness in the pursuit of greater trust.

Areas of Most Progress

Developing an open data strategy and roadmap for delivery
South East Water are approaching the development of their open data capabilities in a methodical manner. They are in the process of developing their open data strategy and are looking to engage a 3rd party for additional support. This work will ultimately inform a roadmap for delivery.

Utilising structured systems and a three step assurance process to ensure high-quality data
South East Water generate most of their shared data through structured systems, thus providing the opportunity to address data quality issues at the source when first generating/ collecting the data. Prior to any shared data releases, South East Water then follow a three step assurance process consisting of internal reviews, external audit and executive review prior to release.

Developing platforms and architecture to meet the needs of open data
South East Water are looking to move away from traditional forms of data generation (Excel, PDF etc.) and towards cloud-enabled architecture to support open data. They have developed a three step plan in conjunction with Google to deliver a cloud-enabled Proof of Value (PoV) that will focus on: 1. Extract, Transform, Load (ETL) & Data warehousing, 2. Analytics, and; 3. Advanced Analytics (machine learning & artificial intelligence).

Enabler	Activity	Progress Statement
Established Data Infrastructure	Data Governance	Developing
	Data Platforms & Architecture	Competent
	Data Quality	Aware
	Releasing High Value Datasets	Developing
	Releasing High Value Datasets 2	Developing
	Sensitive Data Management	Aware
	Technical Standards	Aware
Improved Collaboration on Open Data Across the Industry	Knowledge Sharing & External Support	Aware
	Open Data Documentation; Feedback	Aware
	Prioratisation	Started
Strong Data Culture and the Development of Capabilities & Skills	Developing Advanced Analytics	Started
	Knowledge Management	Aware
	Open Data Expertise & Resourcing	Started
	Roadmap for Delivery	Aware
	Strategy; Strategic Oversight	Aware

Open Datasets	Shared Datasets	Committed Datasets
2	12	0

South Staffs Water Assessment Response Highlights

Whilst South Staffs have published Business Plan tables, WRMP and APR data, they are primarily focusing on shared data at the moment. South Staffs are currently in the process of conducting a data strategy review which will feed into their PR24 business plans and will incorporate elements of open data.

Case Study

Benchmarking industry costs to identify efficiency gains

South Staffs Water shared their data with Water Services Association of Australia (WSAA) which enabled benchmarking of costs against industry peers at a sufficiently disaggregated and comparable level to enable the utility's management to identify potential opportunities for improvements and efficiency gains.

The output was an online Tableau report which allows users to view over 100 different cost categories with various normalisers. In addition, companies have access to an 'International Report' which compares the benchmark results for the UK participants with those of a cohort of major Australian water utilities. South Staffs is also participating in same project in 2021-2022

Areas of Most Progress

Assessing datasets' risk
South Staffs Water data undergoes annual consultations, reviews and risk assessments, internally and externally. Datasets identified as medium and high risk are then evaluated and areas requiring significant attention over the year to come are highlighted.

Developing data governance
South Staffs Water have assigned data owners who are experts in the data's area of relevance to support with governance. In addition, APR data releases are signed off by board members.

Establishing plans for the future
South Staffs Water's internal 'Data Improvement Plan' aimed at identifying issues and establishing consistent approach to data management will be implemented prior to Oct 2023. Their open data strategy focuses on improving data quality, creating new risk assessments, improving governance, data prioritisation and increasing accessibility.

Enabler	Activity	Progress Statement
Established Data Infrastructure	Data Governance	Started
	Data Platforms & Architecture	Started
	Data Quality	Developing
	Releasing High Value Datasets	Developing
	Releasing High Value Datasets 2	Started
	Sensitive Data Management	Started
	Technical Standards	Started
Improved Collaboration on Open Data Across the Industry	Knowledge Sharing & External Support	Started
	Open Data Documentation; Feedback	Started
	Prioratisation	Started
Strong Data Culture and the Development of Capabilities & Skills	Developing Advanced Analytics	Started
	Knowledge Management	Started
	Open Data Expertise & Resourcing	Started
	Roadmap for Delivery	Started
	Strategy; Strategic Oversight	Developing

Open Datasets	Shared Datasets	Committed Datasets
3	3	0

Open Data Assessment Framework
PwC

South West Water Assessment Response Highlights

Whilst South West Water are yet to publish any open datasets beyond publishing of APR and EDM data, they have made investments in their platforms and architecture, and have drafted a set of guiding principles upon which they will be able to develop their open data strategy.

Case Study

Identifying vulnerable customers through establishing data sharing agreements

Identification of vulnerable customers to be added to their Priority Services Register (PSR) has presented a key challenge for South West Water, hence limiting the support they are able to provide to those in need. South West Water have worked to establish data sharing agreements with Western Power Distribution and Scottish & Southern Electricity to enable the identification of vulnerable customers and to reduce the effort required with individual registration applications. This method of customer identification has contributed to finding 17% of their current PSR. In further support of vulnerable customers, South West Water have utilised shared data from the Department of Work & Pensions and data purchased from CACI to identify customers eligible for means tested benefit, ultimately allowing them to provide additional support.

Areas of Most Progress

Developing an open data vision to support data strategy creation
South West Water have drafted a comprehensive vision with 8 guiding principles (including elements of collaboration, governance & data quality etc.) to support the creation of their open data strategy.

Laying the foundations for open data through platforms investment
South West Water have invested in developing their digital ecosystem. They use API enabled core systems (SAP & Oracle), with Power BI for analytics and Python/ Django as an integration/ web development tool.

Working with Stream to establish technical standards
South West Water have established principles as part of their open data vision that allow them to validate technical standards. Going forwards, this will enable them to continue collaborating with Stream participants to establish common standards and increase interoperability.

Enabler	Activity	Progress Statement
Established Data Infrastructure	Data Governance	Aware
	Data Platforms & Architecture	Competent
	Data Quality	Developing
	Releasing High Value Datasets	Aware
	Releasing High Value Datasets 2	Started
	Sensitive Data Management	Aware
	Technical Standards	Developing
Improved Collaboration on Open Data Across the Industry	Knowledge Sharing & External Support	Aware
	Open Data Documentation; Feedback	Aware
	Prioratisation	Aware
Strong Data Culture and the Development of Capabilities & Skills	Developing Advanced Analytics	Started
	Knowledge Management	Developing
	Open Data Expertise & Resourcing	Aware
	Roadmap for Delivery	Aware
	Strategy; Strategic Oversight	Aware

Open Datasets	Shared Datasets	Committed Datasets
2	3	0

Southern Water Assessment Response Highlights

Southern Water have not made any datasets open beyond publishing of APR and EDM data and are currently primarily sharing data instead of making it open. However, they have made progress in developing their enablers and have laid the foundations for future open data initiatives by investing in suitable data platforms.

Case Study

Sharing data with specialist geospatial and drone operator companies to identify previously unmapped drainage

Southern Water's in-house innovation team shared data associated with known sewer and drainage assets with two companies, a geospatial company and a drone operator, to determine how their solutions could help refine, improve, and validate existing data. In the test location, the solution by the geospatial company predicted 44% more pipe assets than were currently mapped. Additionally, the drone operator explored whether manhole locations could be identified to assist in identifying new unknown networks or by validating inferred networks - they were successful in helping identify drainage that was not previously mapped.

Areas of Most Progress

Developing the capabilities for open data through targeted resourcing
To support the development of their data capabilities, and as a result their open data expertise, Southern Water have started to hire more mature data roles such as data architects, data engineers, data modellers, data governance specialists and data scientists.

Improving Knowledge Management through the development of a data catalogue
Having introduced owners for specific sharepoint data elements, Southern Water are now in the process of developing a data catalogue to support adopting common definitions, demonstrate data lineage and curate reusable data.

Ensuring data users inform the prioritisation of datasets
Southern Water identifies priority datasets for delivery through interactions with specific groups through their Stakeholder and/or Customer Engagement teams. For example, a need was identified for MPs and their constituents to be able to view where investments are being made across the constituency. This led to the development of an interactive and open map that served not only political stakeholders but also councils, national parks and county levels.

Enabler	Activity	Progress Statement
Established Data Infrastructure	Data Governance	Started
	Data Platforms & Architecture	Competent
	Data Quality	Developing
	Releasing High Value Datasets	Competent
	Releasing High Value Datasets 2	Aware
	Sensitive Data Management	Aware
	Technical Standards	Developing
Improved Collaboration on Open Data Across the Industry	Knowledge Sharing & External Support	Aware
	Open Data Documentation; Feedback	Started
	Prioratisation	Aware
Strong Data Culture and the Development of Capabilities & Skills	Developing Advanced Analytics	Aware
	Knowledge Management	Developing
	Open Data Expertise & Resourcing	Developing
	Roadmap for Delivery	Started
	Strategy; Strategic Oversight	Started

Open Datasets	Shared Datasets	Committed Datasets
2	3	0

Thames Water Assessment Response Highlights

Thames Water have laid the foundations for open data by developing an Open Data API platform with built in feedback loop and usage tracking capabilities. This has already been used for publication of EDM data and forms the basis for future data releases in a standardised manner, providing data users the ability to develop products and solutions.

Case Study

Thames Water have utilised external tide data to reduce false positive discharge alerts

Thames Water's DAM (Discharge Alert Manager) provides a near real-time capability to monitor their STW (Sewage Treatment Works) and CSO (Combined Sewer Overflows) discharge points. By merging their data with the Environment Agency Tide Gauge API, they have been able to reduce false positive discharge alerts and make their operation more efficient.

Furthermore, their Open API architecture provides the capability of near real-time and batch publications of EDM data. At the time of writing of this report, the API hosts both historical storm discharge activity and a snapshot of current storm discharge activity.

Areas of Most Progress

Developing a generic open data API to support accelerated data releases
Thames Water have extended their core systems to include a generic open data Application Programming Interface (API) to provide a scalable solution to publish open data in a highly usable format. Their solution adopts standards present in their core data systems, allows easy tracking of usage metrics, enables a feedback loop between Thames and the API's users, and provides clear documentation and data definitions.

Promoting Data Security
Thames have adopted several common and bespoke data standards. Their API uses JSON via REST API as their primary data format, which allows for easy integration with other systems. They have adopted the ISO8601 date/time standard to ensure consistency in their data and make it easy for users to interpret, and they use OSGB coordinates to accurately represent the location of their data points.

Using Data Quality Scorecards
Thames Water have established a network of data guardians who work to establish a minimum standard for legitimate use of their data. To support these efforts, Data Quality Scorecards have been created for critical data domains, including customer, property, and meter (there are currently 12 in total), and are updated on a monthly basis so ensure continual improvement.

Enabler	Activity	Progress Statement
Established Data Infrastructure	Data Governance	Optimising
	Data Platforms & Architecture	Visionary
	Data Quality	Optimising
	Releasing High Value Datasets	Competent
	Releasing High Value Datasets 2	Optimising
	Sensitive Data Management	Competent
	Technical Standards	Optimising
Improved Collaboration on Open Data Across the Industry	Knowledge Sharing & External Support	Optimising
	Open Data Documentation; Feedback	Visionary
	Prioratisation	Developing
Strong Data Culture and the Development of Capabilities & Skills	Developing Advanced Analytics	Developing
	Knowledge Management	Competent
	Open Data Expertise & Resourcing	Developing
	Roadmap for Delivery	Developing
	Strategy; Strategic Oversight	Competent

Open Datasets	Shared Datasets	Committed Datasets
2	1	3

United Utilities Assessment Response Highlights

United Utilities are primarily focusing on shared data over opening up their data, only making one dataset open beyond publishing of APR and EDM data. However, their work with the Open Data Institute (ODI) in performing a current state analysis and developing an open data strategy suggests an increased clarity on how to drive open data forwards.

Case Study

Tailoring payment plans and expediting affordability assessments for vulnerable customers[1]

United Utilities' use open banking data to verify customer's income and evidence of benefits payments to improve the accuracy of their affordability assessments. Their solution has streamlined customer eligibility for reduced-rate social tariffs, where previously customers would need to manually collate their income and expenditure information including evidence of benefit receipt in preparation for a telephone affordability assessment. An assessment can now be completed in 15 minutes with improved accuracy compared to the old process which required a week of administrative effort. In addition, it increases first time completion rate and provides customers with a decision on tariff eligibility immediately such that a sustainable payment plan can be agreed.

Areas of Most Progress

Building a dedicated data governance team & monitoring data quality
United Utilities have a dedicated team responsible for Data Governance, which reports into the Head of Data & Analytics. The team has implemented a framework, with assigned owners and stewards, which are responsible for decisions related to access, use and distribution of data. Governance groups actively monitor data quality information and work on continuous improvement of systems, processes and data.

Collaborating with the Open Data Institute (ODI) to develop an open data strategy
United Utilities have enlisted the ODI to help them develop an open data strategy that is aligned to open data best practice. To date they have produced a strategy which sets out their ambitions, aims and desired outcomes from creating, opening and sharing data and they expect their strategy document to be formally reviewed and approved by the Executive in January 2023. It will then be published on their website.

Providing public data and promoting company transparency
With help from the ODI, United Utilities completed a maturity assessment to help understand their priorities and determine their next steps. In addition, they also created an open data working group and ensured CEO & CFO support. The company is also working on development of a new public platform as part of their Better Rivers, Better North initiative to provide near real time data feeds for any overflow operations.

Enabler	Activity	
Established Data Infrastructure	Data Governance	Competent
	Data Platforms & Architecture	Optimising
	Data Quality	Competent
	Releasing High Value Datasets (APR Data)	Aware
	Releasing High Value Datasets (Immediate Challenges. E.g. EDM)	Developing
	Sensitive Data Management	Aware
	Technical Standards	Optimising
Improved Collaboration on Open Data Across the Industry	Knowledge Sharing & External Support	Developing
	Open Data Documentation; Feedback	Aware
	Prioratisation	Aware
Strong Data Culture and the Development of Capabilities & Skills	Developing Advanced Analytics	Developing
	Knowledge Management	Aware
	Open Data Expertise & Resourcing	Aware
	Roadmap for Delivery	Developing
	Strategy; Strategic Oversight	Developing

Open Datasets	Shared Datasets	Committed Datasets
3	3	0

[1] Note: Supporting of vulnerable customers by building on open banking data was also mentioned by United Utilities in the H2Open engagement with Ofwat (link)

Wessex Water Assessment Response Highlights

Wessex Water have made ten datasets open and have provided the foundations for others to innovate using their data. They have made progress in developing their enablers, particularly Established Data Infrastructure, and look to leverage their Data Access Layer (DAL) to share more data in the future.

Case Study

Developing smarter sewer monitoring alarms that detect sewer blockages earlier and prevents flooding and pollution[1]

Wessex water launched a challenge on their Marketplace website asking if sewer alarms could be made smarter. The Bath-catchment based challenge consisted of a brief of their aims and shared upto two years of data with contextual information from their sewer depth monitoring devices. The aim was to identify sewer blockages earlier to prevent flooding and pollution. The trial highlighted that the solutions could meet the brief by providing additional insights into the sewer network. A contract was ultimately, awarded to 'StormHarvester' to monitor all event duration monitored (EDM) storm overflows.

Areas of Most Progress

Creating a data-focused work culture
Wessex Water invested in an accredited training academy and currently have a level 3 Data Technician apprentice and a level 6 Data Scientist apprentice. Additionally, Wessex's water industry apprenticeships, delivered by YTL Wessex Academy, have the use of data monitoring and analysis embedded within the programme. They plan to expand their apprenticeship training with the introduction of a new 'big data' iLearn (Internal Wessex training).

Data Platforms and Architecture
Wessex Water developed the capability of a Data Access Layer (DAL) that set the foundation for easy to access and easy to share data with potential for API connection as the next step. A third-party tool, IntoZetta is also being used that enhanced governance and responsibility for data. They also created a blueprint methodology for using one of their biggest datasets of Telemetry and have a roadmap to increase the datasets held within DAL.

Knowledge Sharing and External Support
Wessex Water have been working closely with the wider market on specific datasets and challenges/hackathons to develop greater learning opportunities. A Marketplace mailing list is maintained and all opportunities are advertised to the entire list. LinkedIn and Aiimi were also used for this purpose.

Enabler	Activity	Progress Statement
Established Data Infrastructure	Data Governance	Developing
	Data Platforms & Architecture	Developing
	Data Quality	Developing
	Releasing High Value Datasets	Developing
	Releasing High Value Datasets 2	Developing
	Sensitive Data Management	Aware
	Technical Standards	Started
Improved Collaboration on Open Data Across the Industry	Knowledge Sharing & External Support	Competent
	Open Data Documentation; Feedback	Developing
	Prioratisation	Aware
Strong Data Culture and the Development of Capabilities & Skills	Developing Advanced Analytics	Aware
	Knowledge Management	Aware
	Open Data Expertise & Resourcing	Developing
	Roadmap for Delivery	Developing
	Strategy; Strategic Oversight	Developing

Open Datasets	Shared Datasets	Committed Datasets
10	0	0

[1] Note: Intelligent sewer depth monitoring case study was also mentioned by Wessex Water in the H2Open engagement with Ofwat (link)

Yorkshire Water Assessment Response Highlights

Whilst Yorkshire Water are yet to make any datasets open beyond publishing of APR and EDM data, they have made progress in developing their open data enablers and have successfully shared data with key partners, laying the foundations to release additional data.

Case Study

Supporting growth initiatives and addressing inequality by sharing data with Leeds City Council

Leeds City Council has been developing the Social Progress Index (SPI), designed to act as a catalyst and drive action by presenting reliable environmental and social outcome data across the city's wards through a user-friendly tool; hence, enabling organisations in the region to better target inequalities across the city's communities and deliver more efficient and inclusive growth initiatives. Yorkshire water have shared data with Leeds city council such that the SPI now includes meter penetration, mains repairs and water quality contacts at the LSOA level. Yorkshire Water plan to share more data in the future based on user feedback.

Areas of Most Progress

Leveraging a data reservoir to share data with partner organisation in real-time.
Whilst Yorkshire Water have a range of platforms, they leverage a single cloud-based Data Reservoir to gather, manage and present their data, ensuring that 'a single version of truth' is available. Azure is used to acquire and share data with partner organisations through batch and real-time delivery.

Using tool based common standards whilst retaining flexibility towards future common standards
Yorkshire Water are leveraging tools based common standards where appropriate, e.g. Microsoft Azure standards or APIM standards for integration, however, their method of storing data within their Data Reservoir allows the adaptation to any set of standards that may be determined in the future.

Carrying out internal & external assurance to validate data releases
Yorkshire Water's internal assurance model identifies data providers, data owners and data managers who each take responsibility for carrying out checks. Data reporting processes are documented in standardised procedure templates designed to not only record the reporting steps but to also to state reporting RACIs, business risks, and any interdependencies of reporting. These procedures are signed off annually by an internal assurance manager before external assurance takes place.

Enabler	Activity	Progress Statement
Established Data Infrastructure	Data Governance	Developing
	Data Platforms & Architecture	Optimising
	Data Quality	Developing
	Releasing High Value Datasets	Competent
	Releasing High Value Datasets 2	Competent
	Sensitive Data Management	Developing
	Technical Standards	Developing
Improved Collaboration on Open Data Across the Industry	Knowledge Sharing & External Support	Competent
	Open Data Documentation; Feedback	Aware
	Prioratisation	Developing
Strong Data Culture and the Development of Capabilities & Skills	Developing Advanced Analytics	Aware
	Knowledge Management	Developing
	Open Data Expertise & Resourcing	Aware
	Roadmap for Delivery	Developing
	Strategy; Strategic Oversight	Aware

Open Datasets	Shared Datasets	Committed Datasets
2	6	0

Thank you

This publication has been prepared for general guidance on matters of interest only, and does not constitute professional advice. You should not act upon the information contained in this publication without obtaining specific professional advice. No representation or warranty (express or implied) is given as to the accuracy or completeness of the information contained in this publication, and, to the extent permitted by law, PricewaterhouseCoopers LLP, its members, employees and agents do not accept or assume any liability, responsibility or duty of care for any consequences of you or anyone else acting, or refraining to act, in reliance on the information contained in this publication or for any decision based on it.

© 2023 PricewaterhouseCoopers LLP. All rights reserved. 'PwC' refers to the UK member firm, and may sometimes refer to the PwC network. Each member firm is a separate legal entity. Please see www.pwc.com/structure for further details.

Appendix 1
Framework Development and Feedback

We collaborated with a range of experts whilst developing the assessment framework

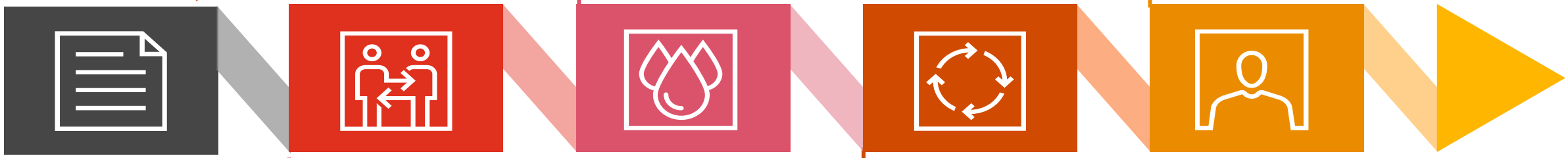
The framework was originally developed by drawing on published best practice literature on open data and the knowledge of PwC and Ofwat. The framework was then refined based on feedback collected in two workshops and a series of 1-1 sessions with wider stakeholders, water companies and data users.

Ongoing collaboration between PwC,Ofwat and Open Data experts to create a repeatable self-assessment framework

1. Initial Research

The initial framework structure identifying beneficiaries, outcomes and enablers was developed based on;

- ODI's Open Data Maturity Model's themes
- Ofgem's Data Best Practice Guidance,
- Catapult Network's Energy Data Taskforce Report
- UKRN's Infrastructure Data Sharing
- Ofwat's H2Open's Enablers



2. Wider Stakeholder Workshop

We consulted with a range of stakeholders including government departments, independent advisory bodies and regulators to refine the framework structure.

We also identified relevant activities to measure progress against enablers.

3. Water Companies Workshop

We engaged with water companies to discuss the types of questions that will be asked to assess progress against the developed open data framework.

The questions were adjusted in line with feedback and a list of potential supporting evidence to be supplied by companies was created to verify company responses.

4. 1-1 Session Input

1-1 sessions were utilised to gain deeper feedback from regulators and data users to ensure the framework represents the needs of various stakeholders.

5. Ofwat's Sign-off and Clarification session

The final Ofwat approved version of the framework was sent to water companies for review and to raise any queries in a clarification session.

Companies were provided a 3 week period to respond to the framework questionnaire.

Additional details can be found on the following slides

Workshop #1: Industry stakeholders

We held an interactive workshop with a wide group of stakeholders from within the water sector and beyond to collect feedback on the framework's structure and design. This led to several structural changes and the inception of new activities.

Objectives

1. To ensure comprehensiveness and logic of the open data progress framework - Outcomes, Enablers, and Activities
2. To gain knowledge about wider open data initiatives.
3. To understand what good looks like in the context of open data in the water sector.

Attendees: Industry stakeholders from various organisations; Open Data Institute (ODI), Ofgem, Geospatial Commision, Welsh Government, Environment Agency, CC Water, British Water, Zuhlke

Actioned Feedback

Knowledge Sharing & External Support

Activity was created to reflect the need for the industry to build awareness of their open datasets and to provide support to potential external users.

Stakeholders' quotes

"Training! Both technical and non-technical, at a variety of levels"

Releasing High Value Datasets

This activity was moved from Collaboration to Infrastructure due to its technical nature.

"Data is FAIR - Findable, Accessible, Interoperable and Reusable"

Sensitive Data Management

Sensitive data management was extracted from sitting with the Governance and Data Quality Activity to form its own activity; highlighting its importance.

"Data security, clear legal agreements (is part of governance, but crucial)"

Workshop #2: Water companies

We held an interactive workshop with 16 major water companies to collect feedback on the framework, the assessment design and the assessment questions. Key feedback surrounding the potential evidence they could supply was collected.

Objectives

1. To socialise and gather feedback on the open data framework
2. To socialise and gather feedback on the assessment criteria
3. To socialise and gather feedback on the assessment questions

Attendees: Representatives from 15* of the 16 major UK water companies, both WoCs & WaSCs.

** South Staffordshire Water plc. were unable to attend*

Actioned Feedback

Potential Sources of Evidence

The framework now suggests potential evidence which firms can supply to demonstrate their progress

Feedback

Due to the importance of the feedback loop between companies and data users, feedback was expanded into an independent Activity

Advanced Analytics Development

Creating Innovative Assets was renamed to Advanced Analytics Development to differentiate it from the Innovation Outcome.

Shared Data

Clarification is provided in a glossary alongside the assessment to show how shared data fits into the open data journey.

Stakeholders' quotes

"Project plan, strategy documents, show & tell/screen grabs "

"Collaboration with other Water Co.s"

"Examples of open data products"

"Clarity on what constitutes open data, shared data is a form of open data but might not fit the mould but could demonstrate the benefits"

1-1s: Discussions with regulators and data users on future expectations

We held 1-1 discussions on the specifics of the updated framework and to understand the expectations from water companies by data users. This helped to finalise the framework and informed its future usability.

Objectives

1. To collect feedback on the framework structure
2. To understand the expectations of regulators and data users from water companies
3. To collect feedback from wider stakeholders and ensure that the framework represents the views and needs of various stakeholders

Attendees: Representatives of Ofgem, Water UK & DEFRA

Actioned Feedback

Evidence Verification

Some questions might elicit 'PR' type responses and present evidence verification challenges. The questionnaire was updated to provide companies the option to list any external contact where relevant to substantiate claims if evidence alone is not of a high enough standard.

Strong Data Culture and the Development of Capability & Skills

The enabler now emphasises the need for the whole organisation to be involved in open data

Transparency

Scoring criteria now accounts for transparency between the water companies themselves and the water companies and customers

Stakeholders' quotes

"Some responses will be 'PR'

"The whole organisation must understand the value of open data"

"Companies must be transparent to each other and the customers"

List of stakeholders engaged in workshops and 1-1s

The tables below show a list of all stakeholders who participated in the workshops and 1-1s

Stakeholders for Workshop 1

Organisation
Open Data Institute
Geospatial Commission
Welsh Government
Environmental Agency
CC Water
British Water
MOSL

Stakeholders for Workshop 2

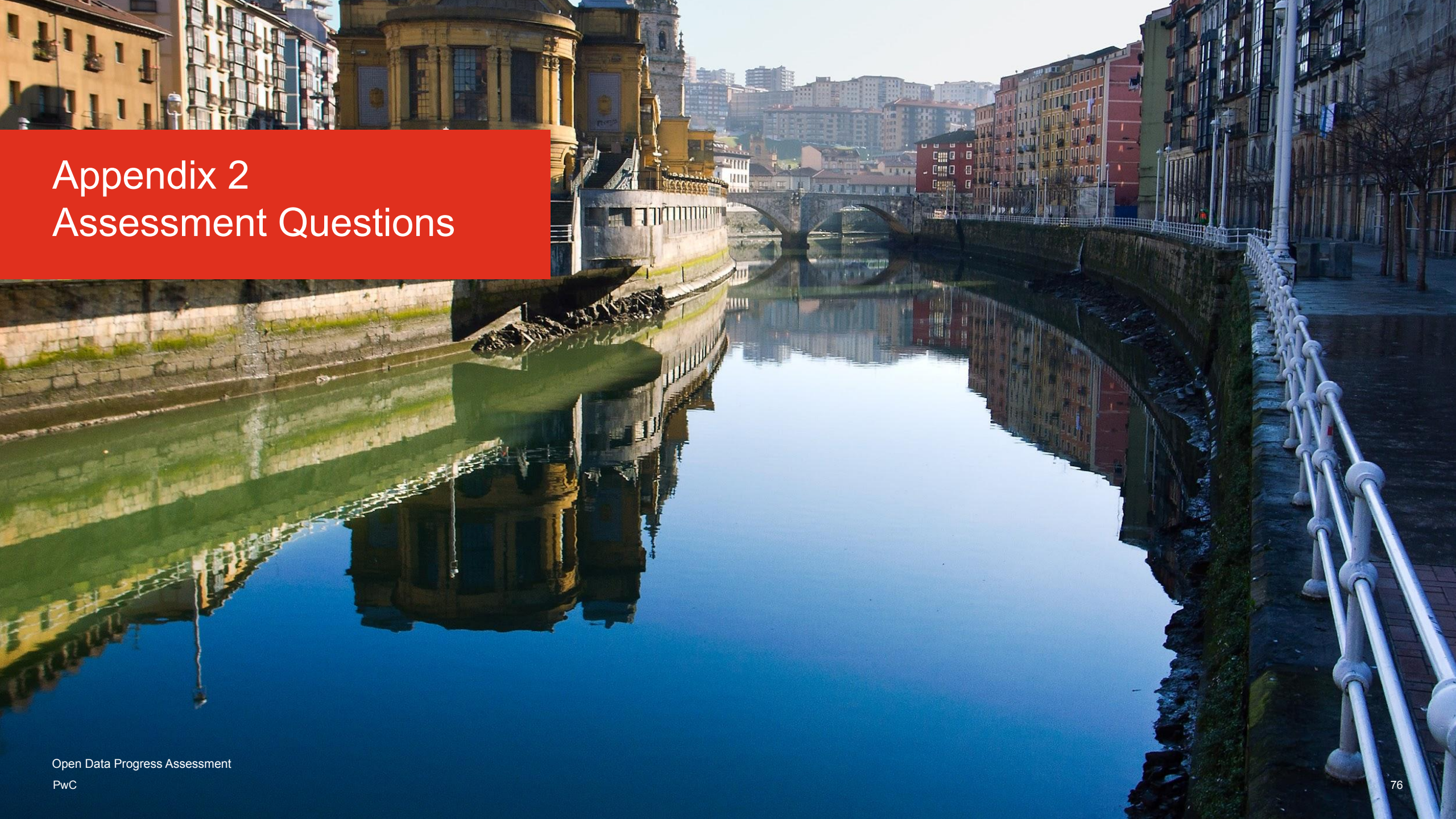
Organisation
Affinity Water Limited (AFW)
Anglian Water Services (ANH)
Bristol Water Plc (BRL) (acquired by South West Water)
Northumbrian Water Limited (NES)
Dwr Cymru Welsh Water
Pennon Water Services Limited
Portsmouth Water Limited (PRT)

Organisation
SES Water
Severn Trent Water Limited (SVE)
Wessex Water Services Limited (WSX)
South East Water Limited (SEW)
Southern Water Services Limited (SRN)
Thames Water Utilities Limited (TMS)
United Utilities Water Limited (UUW)
Yorkshire Water Services Limited (YKY)

Stakeholders for 1-1s

Organisation
Ofgem
Water UK
DEFRA

Appendix 2
Assessment Questions



Outcome aligned progress check questions.

There will be one progress check question per Outcome, designed for water companies to provide evidence of the Outcomes they have achieved and the benefits these have brought to the key beneficiary groups. A long format response of between 200-300 words will be expected, supported by relevant attachments.

Innovation	**1**	Please provide a case study with supporting evidence that describes how you have collaborated externally by providing open/ shared data to stimulate innovation or used external open/ shared data in house to develop an innovative solution. Please provide evidence of how this innovation has brought benefits for your customers & communities, the environment, and the water sector and beyond.	
Efficiency	**2**	Please describe a case study with supporting evidence that shows how open/ shared data has brought new insights and fresh thinking to your organisation that have led to efficiency gains: for example, predictive maintenance. How has this benefitted your organisation, the customers & communities, the environment and wider stakeholders?	
Decision Making	**3**	Please provide details of how the use of open/ shared data has informed a key decision within your organisation that would not have been possible without open/ shared data: for example, through informing a key operational decision or enabling the delivery of your organisational strategy.	
Transparency	**4**	Please elaborate on how you are currently sharing data or using products driven by open/ shared data to identify opportunities to build trust among your stakeholders. Please provide supporting evidence of how this has led to improved relationships with your stakeholders.	
All Outcomes	**5**	How is your organisation using open/ shared data to deal with the immediate issues that face the water sector today? For example, by using Event Duration Monitoring (EDM) or leakage data released in compliance with The Environment Act, or by opening your sewage data to improve transparency.	

Potential Sources of Evidence

Visuals** | **Feedback & Testimonials** | **Verifiable Statistics** | **Cost Saving** | **Time Savings** | **Process Mapping** | **Evidence of Published Data** | **Evidence of External Data Usage** | **Performance on ODIs** | **Evidence of Stakeholder/ Community Engagement** | **Improvement in Service Provisions** | **Artifacts / Links** | **Perceived Benefit

Enabler aligned progress check questions.

There will be 2-3 progress check question per enabler, each tied to one or more specific activities that enable outcome realisation, and ultimately benefits realisation . A shorter format response of 100-200 words is expected per question, supported by relevant attachments.

Established Data Infrastructure

6	Please describe what you have done to adapt your organisation's core data platforms and architecture to support the use of open/ shared data? Why did you decide to do it in this manner and how has it supported external stakeholders' use of your published resources?	***Evidence of data model standardisation | Evidence of data platforms adaptation: hosting, sharing, ETL, visualisation, cloud, advanced analytics (for example: machine learning) | Technological investments | Investment decision collaboration to reduce bifurcation***	Data Platforms & Architecture
7	Please describe how you have adopted common or bespoke data standards (for example, standardised format) to increase interoperability of your open/ shared data and to ensure your published resources meet user needs. What are the key standards that you have adopted and why?	***Technical standards documentation | Data strategy reference to technical standards approach | Evidence of collaboration to define common standards | Evidence of continual technical standards reviews | Evidence of sharing best practice more widely***	Technical Standards
8	Please describe your organisation's data governance structure; specifically its forums, roles and responsibilities. If applicable, what have adaptations you made to support open data/ shared datasets?	***Evidence of data owners & stewards | Evidence of governance structure | Evidence of board level support | Evidence of KPI tracking & monitoring | Governance process documentation | Evidence of feedback from data users leading to changes in processes | Development of thought leadership sharing best practice***	Data Governance

Enabler aligned progress check questions.

There will be 2-3 progress check question per enabler, each tied to one or more specific activities that enable outcome realisation, and ultimately benefits realisation . A shorter format response of 100-200 words is expected per question, supported by relevant attachments.

Established Data Infrastructure

9	Please describe your organisation's approach to ensuring data quality is robust enough to meet user needs. If applicable, what adaptations have you needed to implement to support open data?	Potential Sources of Evidence	*Data quality prioritisation schedule/ rules | Process documentation | Data quality KPI tracking and monitoring | Evidence of sharing data quality best practice externally*	Data Quality
10	Please describe your organisation's policy and any relevant processes in place to ensure data is secure. Please make reference to how you are progressing towards an 'Open by Default' by i) identifying and anonymising sensitive personal data, ii) obfuscating commercially sensitive data and, iii) triaging internally and externally generated open/ shared datasets to mitigate security risks associated with merging the data.	Potential Sources of Evidence	*Data sharing policy | GDPR policy | Cyber security policy | Data protection policy | Data control & governance documentation | Evidence of security steering groups | Evidence of rule for triaging data for aggregation threats*	Sensitive Data Management
11	Please describe the actions your organisation has taken to ensure the Annual Performance Report (APR) data meets the best practice guidelines (i.e. publishing data in excel, publishing metadata and methodology changes) and can be reused and accessed by all? What were the key challenges you faced in meeting open data requirements?	Potential Sources of Evidence	*Link to published APR data*	Releasing High Value Datasets

Enabler aligned progress check questions.

There will be 2-3 progress check question per enabler, each tied to one or more specific activities that enable outcome realisation, and ultimately benefits realisation . A shorter format response of 100-200 words is expected per question, supported by relevant attachments.

Established Data Infrastructure	**12**	Please describe the actions your organisation has taken for release of high value datasets such as leakage, event duration monitoring (EDM) and other environmental data. Please provide details of any supporting narrative that is provided alongside the release of the datasets. How frequently are these datasets updated? If no such dataset has been released, please explain challenges faced in release of such datasets and how these are planned to be addressed for a future release.	Releasing High Value Datasets
		Potential Sources of Evidence	***Links to published environmental datasets | Evidence of supporting narrative***

Enabler aligned progress check questions.

There will be 2-3 progress check question per enabler, each tied to one or more specific activities that enable outcome realisation, and ultimately benefits realisation . A shorter format response of 100-200 words is expected per question, supported by relevant attachments.

Strong Data Culture and the Development of Capability & Skills

13	Please outline the main principles of your organisation's strategic vision for open/ shared data. Please describe the strategic oversight in place to ensure the successful implementation of your open data strategy.	***Reference to open data's role in company's organisational strategy | Reference to open data in wider data strategy | Evidence of ownership of key processes | Evidence of key executives' performance being tied to delivery***	Strategy ——— Strategic Oversight
14	Considering your organisation's open data strategy, what steps have you taken to communicate and implement this strategy across all levels of your organisation to create a culture of open data? Please outline the key milestones you have achieved since H2Open's publication?	***Evidence of KPI tracking | Evidence of sufficient budget allocation | Evidence of key executives' performance being tied to delivery | Evidence of open data steering group***	Roadmap for Delivery
15	How do you ensure that institutional knowledge around open/ shared data is preserved during operational changes? For example, by ensuring it is preserved through changes in dataset ownership or staff turnover.	***Knowledge transfer policy | Investment in knowledge transfer capabilities | Internal presentations | Knowledge transfer training | Reference to knowledge transfer in data strategy | Evidence of external collaboration / sharing best practice***	Knowledge Management

Potential Sources of Evidence

Enabler aligned progress check questions.

There will be 2-3 progress check question per enabler, each tied to one or more specific activities that enable outcome realisation, and ultimately benefits realisation . A shorter format response of 100-200 words is expected per question, supported by relevant attachments.

Strong Data Culture and the Development of Capability & Skills

16

Please describe your organisation's data capabilities/ company-wide data literacy in relation to supporting the Open Data agenda. How are you developing these capabilities / making provisions to resource these skills? How have your efforts progressed to date with respect to your capability development journey?

Potential Sources of Evidence

Evidence of data literacy training: for example: licensing, technology, common knowledge, operational requirements, strategic objectives etc.| Evidence of centers of excellence | Evidence of resourcing and budget allocation for developing open data capabilities | Evidence of engagement with external Open Data networks.

Open Data Expertise & Resourcing

17

Please describe how you are facilitating external collaboration with advanced analytics experts / developing internal advanced analytics capabilities: for example, machine learning, in your organisation.

Potential Sources of Evidence

Evidence of training | Evidence of engagement with external experts | Thought leadership | Internal & external presentations| Evidence of roles with elements dedicated to open data advanced analytics

Developing Advanced Analytics

Enabler aligned progress check questions.

There will be 2-3 progress check question per enabler, each tied to one or more specific activities that enable outcome realisation, and ultimately benefits realisation . A shorter format response of 100-200 words is expected per question, supported by relevant attachments.

Improved Collaboration on Open Data Across the Industry

18

Please describe the internal processes and engagement with external stakeholders your organisation has in place to identify priority data sets that lead to the greatest benefits for customers & communities, and with wider stakeholders?

Potential Sources of Evidence

Evidence of processes | Evidence of engagement with data user community | Evidence of data usage metrics tracking | Evidence of collaborating with others to determine priority data sets

Prioritisation

19

Please describe the how your organisation ensures that the documentation that supports your open/ shared datasets meets user needs. What standards do you adhere to? And what proportion of your shared / open data complies with these standards?

Potential Sources of Evidence

Documentation methodology | Standardised documentation template | Evidence of a feedback loop showing how data users inform changes in documentation

Open Data Documentation

Feedback

20

Describe how your organisation is increasing engagement with the wider community on open data, both within the industry and beyond, to i) learn from others, ii) to share your own open data experiences, and iii) provide additional support to external data users? What benefits have this community-enabled approach brought to your organisation?

Potential Sources of Evidence

Evidence of hosting/ attending events | Evidence of engaging with other open data initiatives | Evidence of making external parties aware of what data your company is publishing | Evidence of additional support provided to data users | Evidence of sharing high-value use cases of open data your organisation has identified

Knowledge Sharing & External Support

Additional Questions

Three additional questions will be asked to capture supporting information that does not sit elsewhere within the progress framework. These will explore existing / future open data datasets, provide an opportunity for companies to tell about anything else that doesn't sit within the assessment questions, and to explore what the companies have learnt over the past 12 months and what they intend to do next.

21	Please: I. Provide evidence for all your existing shared / open datasets. And if applicable, how frequently are you updating these datasets? II. Share with us which datasets you have publicly committed to releasing in the future	
22	If you feel the existing questions did not give you an opportunity to showcase all of your open data progress, please share any additional information in the box below.	
23	Please tell us what you have learnt in the past 12 months and what you intend to do next regarding open data.	

Appendix 3
Progress Criteria

What good looks like: Enablers / Activities

An 8 point scale has been devised to assess the progress water companies have made against the Enablers / Activities. At the start of their open data journey, companies would be expected to sit at 'Started', graduating through the statements as progress is made. Not all activities have a progress statement aligned with all 8 points on the scale, however, if a company response indicates it lies between two non-adjacent progress statements, then the appropriate statement from the full 8 point scale will be used to summarise their performance.

Started	The organisation has made some progress with open data, largely driven in an unstructured manner by individuals within the organisation and external expert support. Open data initiatives are seen as R&D exercises with few supporting processes and little to no consideration for knowledge management and making published resources accessible to outside users.
Aware	Key processes and defined technical standards have started to emerge, with individual business units being the driving force behind progress in line with their specific objectives. There is no organisation wide open data strategy in place and as a result the key benefits of open data are not commonly understood across the organisation.
Developing	An open data strategy has been defined with set measurable targets to facilitate its delivery. Governance structures have started to emerge, knowledge management is being considered and there is an understanding of the key areas in which the organisation needs to focus its efforts to progress along their open data journey.
Competent	The organisation has progressed towards a standardised data model supporting open data, including defined processes surrounding dataset publication and managing data standards. Delivery of the open data strategy has a dedicated budget and clear lines of responsibility. Open data awareness is being developed internally and engagement with external dataset users is starting to emerge.
Optimising	Open data policy and strategy align with organisational objectives. Knowledge of open data exists across all levels of the organisation, within which staff are leveraging existing data platforms to harness open data's power. Standardised documentation, metrics tracking, feedback collection and releasing datasets according to the FAIR principle is supporting the delivery of the key outcomes / benefits.
Visionary	Open data is seen as a key element of their organisational strategy, with key executives performance tied to its delivery. End to end governance processes, including advanced sensitive data identification and anonymisation, are in place to support an 'Open by Default' approach. A proactive approach to updating technical standards, documentation, knowledge sharing and prioritisation has been adopted.
Pioneer	The organisation is a class leader in its practice of data governance, quality management and common standards; driving a cross-industry approach where possible. It invests in new technologies to better utilise open data and sees feedback from key stakeholders as essential to further development of open data.
Leader	The organisation is sharing its class leading practices with other open data players, engaging within the water sector and beyond to foster networks of expertise, promote the use of open data by highlighting its value, and teaching others to utilise their methods to derive benefits for the key beneficiaries.

What good looks like by activity

Enabler	Activity	Progress Statement	Progress
Established Data Infrastructure	Data Platforms & Architecture	The organisation holds data in a central location.	Aware
		The organisation's data is structured into a standardised data model, with consideration taken towards Open Data usage.	Competent
		The organisation is effectively using their current data platforms (hosting, sharing, integration, ETL, visualisation, cloud, advanced analytics etc.) to support the use of open data. These platforms have been organised in a logical and scalable manner using Reproducible Analytical Pipeline (RAP) best practice to best draw out insights / share data externally.	Optimising
		The organisation is collaborating with others in the open data space to ensure that their investments are in line with industry standard and not creating bifurcation.	Visionary
		The organisation is investing in new technologies to specifically help them achieve their open data goals and embed the use of open data into the day-to-day running of their organisation: for example, using cloud technologies to automate the processing and publication of open data with the highest levels of security and monitoring.	Pioneer
	Data Quality & Governance	Primary datasets have basic data governance structures in place: including, data owners and stewards, clear definitions for fields/ attributes for each dataset and data quality profiling.	Developing
		The majority of datasets have data governance structures in place, with a common suite of hierarchies for each master data set and rules built to prioritise data cleaning initiatives. There is board level support with sufficient KPIs in place for effective monitoring and tracking.	Optimising
		End-to-end data governance processes are in place across all datasets, which were created through consultation with dataset users and are consistently reviewed to address any governance or quality lapses.	Visionary
		There is a well defined feedback loop in place, allowing dataset users to highlight the challenges they are facing with data quality, enabling the organisation to continually update their data quality and governance processes to best support them.	Pioneer
		The organisation is a class leader in its practice of data governance and quality management, developing thought leadership to highlight value adding areas of open data to companies both within the water sector and beyond.	Leader
	Releasing High Value Datasets	Some datasets (shared or public) are being released and updated regularly but there are no common standards for doing so.	Started
		There is a standardised process for the release of published resources, with bespoke approaches present for some individual projects.	Developing
		All datasets are released according to a standardised organisation-wide process, and are released according to a priority based schedule. Archived datasets are also provided to allow historical analysis.	Competent

What good looks like by activity

Enabler	Activity	Progress Statement	Progress
Established Data Infrastructure	Releasing High Value Datasets (Continued)	All datasets are released according to the FAIR principal. Findable, Accessible, Interoperable and Reusable.	Optimising
		The organisation routinely collects metrics monitoring the release process, looking to optimise/ automate the processes where applicable.	Visionary
		The organisation shares its learning and methods for releasing high quality datasets within the water sector and beyond. E,g, through external presentations and/or thought leadership.	Leader
	Sensitive Data Management	Key processes are in place to identify and anonymise/ obfuscate sensitive personal or commercial data, and mitigate risks associated security related data, of published resources.	Aware
		The organisation has clear triage processes in place to ensure that data does not become sensitive through linking and merging with existing internal datasets.	Competent
		Advanced sensitive data identification, anonymisation and obfuscation is in place to ensure an 'Open by Default' approach is taken towards all datasets.	Visionary
		The organisation has clear triage processes in place to ensure that data does not become sensitive linking and merging with all existing open datasets both within the water sector and beyond.	Pioneer
		The organisation shares its approach to sensitive data practices within the industry and beyond to promote the opening of data.	Leader
	Technical Standards	Defined technical standards are applied to all datasets.	Aware
		The organisation has a defined strategic approach towards managing data standards. Common standards applicable to the water sector and other primary users are adopted over closed bespoke standards when possible: for example, API standards and Government's Digital Service (GDS).	Competent
		The organisation collaborates within the the water sector and beyond to inform their technical standards, hence preventing non-coherent standards across the industry: for example, by participating in Stream or similar initiatives.	Optimising
		The organisation continually refreshes its technical standards in line with technological advancement, current best practice and feedback collected from dataset users.	Visionary
		The organisation drives the creation/ adoption of common standards across the water industry and beyond by collaborating with other industry actors to define best practice.	Pioneer

What good looks like by activity

Enabler	Activity	Progress Statement	Progress
Strong Data Culture and the Development of Capability & Skills	Advanced Analytics Development	The organisation contracts external SMEs to use utilise advance analytics products (for example: Machine Learning) to solve specific problems.	Started
		The organisation has reviewed their procurement processes to ensure there is a low barrier to entry for external parties looking to engage with them in open data asset creation.	Aware
		Internal roles dedicated with aspects dedicated towards the use/ development of Open Data assets have started to emerge, with advanced analytics capabilities being developed in house to solve problems and bring benefits to the key beneficiaries.	Developing
		The organisation has a dedicated roles/ teams of advanced analytics SMEs tasked with solving high-priority challenges aligned to the key beneficiaries.	Optimising
		The organisation's in house advanced analytics experts actively engage with other players in the water sector and beyond to share advanced techniques / methods of analysis that provide benefits to the key beneficiaries: for example, Thought Leadership	Leader
	Knowledge Management	Individual teams capture information related to open data activities, primarily focused on knowledge transfer at the end of a project or with staff turnover rather than across the full lifecycle of the project of dataset.	Started
		There is organisation-wide policy in place surrounding the capturing and transferring of open data knowledge, with consideration for the entire project or dataset lifecycle rather than at the end.	Developing
		There is clear and well documented policy in place that is easily accessible to all staff members, that supports the usage of open data data assets and to reduce the existence of technical debt.	Competent
		Continuous investment in knowledge management is seen, with it forming a core element of their open data practice.	Optimising
		The organisation actively encourages the sharing of knowledge surrounding open data through internal presentations, continual training and by highlighting high value use cases internally.	Visionary
		The organisation engages externally within the water sector and beyond to share / learn from others' experiences of using open data and by identifying high value use cases.	Leader

What good looks like by activity

Enabler	Activity	Progress Statement	Progress
Strong Data Culture and the Development of Capability & Skills	Open Data Resourcing & Expertise	Early pockets of internal open data expertise are driven by self-taught internal early adopters, with reliance on external expertise for key skills.	Started
		There is no shared understanding of open data across the organisation but pockets of expertise are developing through training & support made available based on project need.	Aware
		The organisation is aware of where efforts need to be focused to support the development of open data expertise and is providing training around open data topics to some teams, largely focusing on operational skills: for example, licensing, technology etc. In addition, centers of excellence have started to emerge and are supporting and mentoring other staff.	Developing
		The organisation is actively building open data literacy: for example, common knowledge, operational requirements and strategic insights; at all levels of the organisation and are ensuring staff are clear on how open data impacts their personal responsibilities.	Competent
		The organisation has clear plans, with sufficient budget allocation, to support the resourcing of open data expertise.	Optimising
		Open data literacy exists across all levels of the organisation, with a shared understanding of how open data contributes towards achieving strategic objectives. Clear centers of excellence are formalised and act as a main contact point regarding open data.	Visionary
		The organisation fosters the creation of networks of open data expertise both within the organisation and with external communities in the water sector and beyond.	Leader
	Roadmap for delivery	Open data initiatives are being driven by early adopters in an unstructured and non-coherent manner.	Started
		The organisation has set measurable targets for the implementation of their data strategy in relation to open data.	Developing
		Key KPIs are tracked in relation to implementation targets for the organisation's data strategy. These are continuously reviewed and adjusted over time to stretch the organisation to achieve their open data goals and ensure they remain consistent with organisational strategy.	Competent

What good looks like by activity

Enabler	Activity	Progress Statement	Progress
Strong Data Culture and the Development of Capability & Skills	Strategic Oversight	Ownership of key processes: for example, data governance, are clear and well understood.	Developing
		There is clear responsibility and budget for the delivery of the open data strategy at a senior management level.	Competent
		The delivery of the open data strategy is frequently reviewed by C-level and board-level actors, and performance of key executives is tied to its delivery.	Visionary
	Strategy	Initial data releases are seen as R&D exercises or are driven by external influence: for example, Ofwats request for water companies to publish ARP data in excel format on their websites.	Started
		Individual business units recognise the benefits of Open data in their own activities. However, there is no coherent organisation-wide strategy in relation to open data.	Aware
		The organisation has defined an open data strategy within their wider data strategy, which includes elements addressed in other activities such as data governance and technical standards.	Developing
		The organisation has aligned their open data policy and strategy with organisational objectives.	Optimising
		The organisation sees open data as a key element to achieving their organisational strategy. This vision is supported by their data strategy with specific considerations towards open data.	Visionary

What good looks like by activity

Enabler	Activity	Progress Statement	Progress
Improved Collaboration on Open Data Across the Industry	Feedback	Judged by other activities as a higher level of progress: for example, in Ways of Working, Advanced Analytics Development, Knowledge Management, Open Data Resourcing & Expertise, Quality & Governance, Sensitive Data Management & Technical Standards	N/A
	Open Data Documentation	Little to no documentation is released alongside datasets, with the documentation being designed by enthusiastic early adopters.	Started
		Individual teams identify the essential documentation required to support the data they are publishing. Some common templates might exist.	Aware
		A standardised methodology for open data documentation and metadata exists, however, only some/ the high-value datasets are released with documentation conforming to this standard.	Competent
		All datasets are released with standardised documentation and metadata, where documentation maintenance is part of the data publishing process. In addition, the organisation invests in the development of additional materials to support dataset users where required.	Optimising
		Appropriate and high quality open data documentation spanning written documentation, metadata, software codes, API documentation etc. is reviewed by key internal stakeholders and released alongside all published data.	Visionary
		Dataset users are invited to provide feedback on published dataset documentation, from which key changes are made to provide additional support.	Pioneer
	Prioritisation	Some ad-hoc attempts are made at assessing the impact of potential dataset releases.	Aware
		There is a defined and repeatable approach for engaging with the dataset user community both before and after the release of data. The organisation begins to use demand for data from dataset users to guide release of specific datasets.	Competent
		The organisation routinely tracks the metrics related to the effectiveness of their data release, identifying standard ways of measuring the datasets impact. These learnings inform the prioritisation of future dataset releases.	Optimising

What good looks like by activity

Enabler	Activity	Progress Statement	Progress
Improved Collaboration on Open Data Across the Industry	**Knowledge Sharing & External Support**	The organisation attends knowledge sharing events hosted by others within the water sector and beyond to understand their open data learnings.	**Aware**
		Organisation is actively engaging with open data initiatives within the water sector: for example, Stream, Hackathons etc.	**Competent**
		The organisation regularly ensures external parties are aware of the data that they have made open, including specific support in accessing their opened data.	**Optimising**
		Organisation is actively engaging with open data initiatives across other sectors.	**Visionary**
		The organisation regularly host knowledge sharing events to highlight the high-value use cases of open data.	**Pioneer**

Appendix 4
APR and EDM Data Requirements

Legal requirements and regulatory expectations associated with the publication of APR and EDM data by Water Companies

Requirements on Event Duration Monitoring (EDM) data publication

Currently, under section 141D of the Water Industry Act 1991 the Environment Agency (EA) publishes an annual report on storm overflows (containing EDM data) annually by 1 April. Under section 141C of the Act, companies are also required to publish annual reports on storm overflows annually by 1 April. The Act stipulates that the reports published by companies in general must be "readily accessible to the public".

In March 2022, Ofwat asked wastewater companies how they planned to tackle storm overflow and river quality issues. As part of their response, some companies committed to release live sewage discharge notifications or near-real-time data on storm overflows by 2023 (Northumbrian, South West, Thames, and United Utilities), 2024 (Wessex), and 2025 (Dwr Cymru, Yorkshire).

Expectations on Annual Performance Reporting (APR) Data publication

In March 2022 Ofwat set out expectations (link) for companies to make 2021/22 APR data available in a way that can be reused and accessed by all, by publishing their APR tables in an Excel format on their website and publishing metadata such as publication date, version information and the file type, and supporting documentation such as explaining changes in methodology from previous years.

Thank you

This publication has been prepared for general guidance on matters of interest only, and does not constitute professional advice. You should not act upon the information contained in this publication without obtaining specific professional advice. No representation or warranty (express or implied) is given as to the accuracy or completeness of the information contained in this publication, and, to the extent permitted by law, PricewaterhouseCoopers LLP, its members, employees and agents do not accept or assume any liability, responsibility or duty of care for any consequences of you or anyone else acting, or refraining to act, in reliance on the information contained in this publication or for any decision based on it.

© 2023 PricewaterhouseCoopers LLP. All rights reserved. 'PwC' refers to the UK member firm, and may sometimes refer to the PwC network. Each member firm is a separate legal entity. Please see www.pwc.com/structure for further details.