Digitally Transforming Scottish Environmental Protection

Company: Scottish Environment Protection Agency (SEPA)

Headquarters: Stirling, Scotland

Industry: Government

Founded: 1996

Facilities: 22, present across all of Scotland

The Scottish Environment Protection Agency (SEPA) is tasked with protecting and improving the Scottish environment through regulation, engagement, and its new regulatory strategy, One Planet Prosperity. SEPA is also responsible for flooding protection and warnings, providing a key service to the people of Scotland. To ensure SEPA meets these responsibilities and aims, it is constantly gathering data from many varied sources to help drive evidence based business and environmental decisions. A single analytics platform that can harness the value of this data and empower staff in their daily work is crucial to success. To this end, the agency went on a data science journey using TIBCO Spotfire and TIBCO Data Science software, which is now fully embedded in all aspects of its work. SEPA staff can now provide the Scottish industry and general public with powerful data-driven, web-based analytics and dashboards so they can perform their own environmental analyses and answer questions.

Historical and Current Environmental Data Gathering and Analysis

SEPA has a vast array of past and present environmental data. This includes hydrological data for rivers and rainfall dating back to the 1950s gathered at hourly or 15-minute intervals, ecological and chemistry samples dating back to the 1960s, as well as various data sources relating to how it regulates. Other examples include detailed air quality data, inspection, and compliance information of sites such as farms, sewage treatment works, landfills, and all other areas of Scottish industry. To hold this vast amount of data, SEPA uses databases such as Oracle and geographical information systems (GIS).

With such a long-term digital store of environmental data, a real challenge was how to bring these disparate data sources together so their true value and potential could be realized. SEPA searched for a single platform that staff could use at any time to help drive and make evidence-based decisions. With such a rich source of evidence, it also presented great opportunities for modeling the environment to help shape and futureproof the country's environmental needs.

The agency began a two-year trial of TIBCO Spotfire analytics. One of the key aspects of the trial was the ability to use the power of the statistical language R, as well as rapidly deploy web-based analytics to SEPA staff. Another key component was the ability to embed visuals and analytics in other web-based products. Spotfire software was able to provide all of this. It integrates with R by way of TIBCO's enterprise-grade R engine, TIBCO Enterprise Runtime for R (TERR), and includes a powerful JavaScript API allowing Spotfire software to be embedded in many areas of the organization.

With all of these benefits, including bringing all of the disparate sources of data together, the trial convinced SEPA to purchase a full Spotfire licence. Today, nine years later, over 140 Spotfire driven tools have been published to internal SEPA staff, and also to the public through three websites. Over 80% of SEPA staff use Spotfire analytics for their daily jobs.

Agile Informatics Development

With the Spotfire ability to instantly deploy any analytics or data science tool to the web, SEPA rapidly produced and gave staff access to new analytical and statistical capabilities. Because Spotfire software allows for rapid data analysis and dashboard production, SEPA implemented an Agile informatics development process to help support and manage the ever-increasing number of Spotfire tools and areas they would be used in. SEPA took the Agile ethos of collaboration and progress through small, well-defined tasks, and applied it to the informatics team for their data analysis and modeling work. This agile process enabled data science project team members to focus on key business needs and collaboration between the Spotfire analyst and business leads.

The process involved an initial discussion to set out the key questions or business needs for the analytics development. It then moved to a series of short sprints to develop the tool or dashboard in a stepwise manner, followed by testing and release. This collaborative process encouraged interaction with the business leads as early as possible during all steps, releasing versions or demonstrations to users early in the process so their feedback could be gathered and fed back into the process from the outset. Below is a diagram of the process:

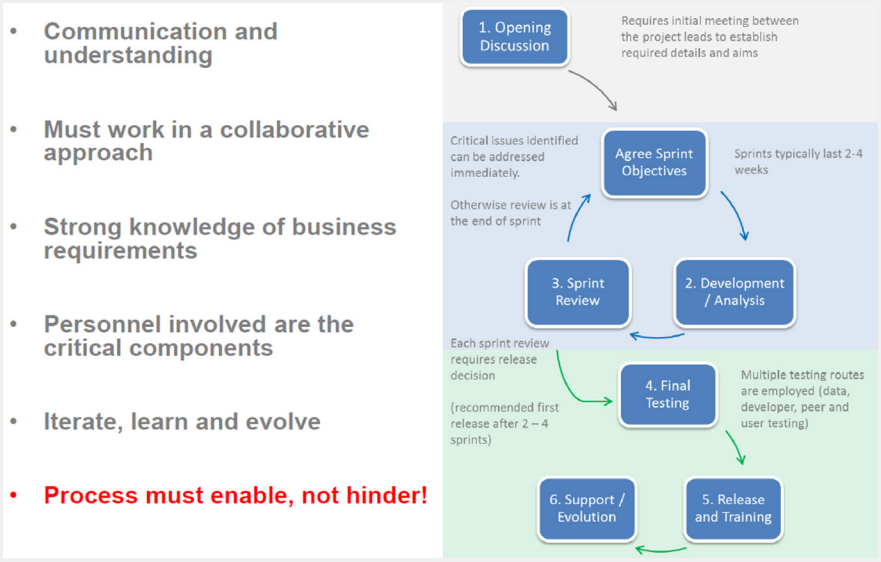

Figure 1: Summary of SEPA's agile process used for each data science project.

The Spotfire ability to cover the full data science development lifecycle: data retrieval, preparation, transformation, analysis, and final interactive dashboard or presentation meant it could easily meet the aspirations of the proposed agile process.

SEPA's Data Science Journey

Initially, the focus for analytics was confined to SEPA's use of Spotfire analytics to review and map water quality's short- and long-term trends and changes in Scotland. The analytics focus was then expanded to determine capacity of the water environment and likelihood of change using increasingly powerful statistical methods to analyze changes and incidents in water quality.

Due to the ability to rapidly produce analyses and publish them on SEPA's intranet using Spotfire capabilities, new use cases soon appeared. For example, a tool to enable staff to verify and cleanse data while sharing comments with each other and a tool to produce key performance indicators for the business and assess new applications for industry's use of the environment were created. SEPA categorized three Spotfire use cases:

1. Data verification
2. Analysis and decision-making
3. Reporting and dashboards

As the number of topics and tools covering these use cases increased, it became apparent that not only were these analytical tools empowering staff to make decisions, they were also saving crucial time. Staff were no longer battling to find data, performing analysis using traditional spreadsheet desktop tools, or implementing their own methods and statistics. Now all the data was prepared and available instantly via a browser. Staff had the power to produce standardized and agreed upon statistics, as well as delve into and analyze their data interactively as they required. SEPA soon pushed Spotfire driven analytics beyond the science department into regulation, flooding, and business areas such as estate strategy, organizational and business planning, and even equalities and staff wellbeing. One such example is a dashboard allowing managers to view crucial information and alerts for their team. The following chart shows the year-on-year increase in Spotfire usage by SEPA staff (~1,200 staff), with almost 40,000 uses of Spotfire tools in 2019.

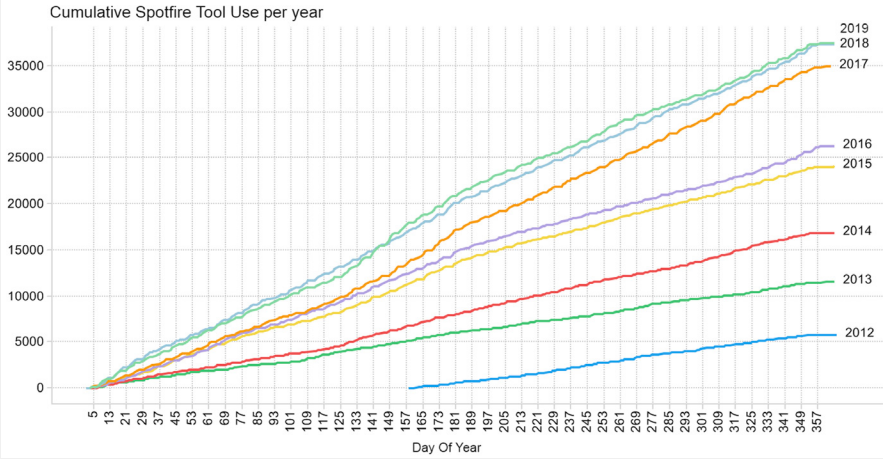

Figure 2: Cumulative internal uses of Spotfire tools each year.

Data Verification Use Case: Scottish Pollutant Release Inventory (SPRI)

Each year, SEPA collects data from operators in Scotland to populate the Scottish Pollutant Release Inventory, which is reported to Europe and published on SEPA's and Scotland's Environment website.

This dataset has data from over 800 operators covering ~200 pollutants emitted to air and water. Previously this data was validated via a series of spreadsheets passed between SEPA regulatory teams and the SPRI administration team. It was a

lengthy and difficult process given the amount of data and issues of integrity always associated with traditional file-based formats that are not centralized.

In 2011, the introduction of Spotfire analytics gave SEPA the ability to replace this process with an interactive web-based suite of tools. These tools were able to highlight issues in data instantly and consistently, while making the process of reviewing and browsing the data much faster and more efficient. Now the staff had one place to view the data. SEPA added the ability to add questions and comments to enable discussion and the recording of discussions. By increasing communication, staff could concentrate their time on improving the quality of the data being reported while also gaining a better understanding of what operator and sector changes had occurred versus previous years. While the main focus of this effort was to greatly improve data quality and methods of collaborating, it was found to substantially improve efficiency. SEPA saved around 66% of time versus previous years. As a result, staff had extra time to verify the data fully, allowing for better discussions between staff and operators. In future years, time saved increased to over 80% as staff became more familiar with the new tools and processes.

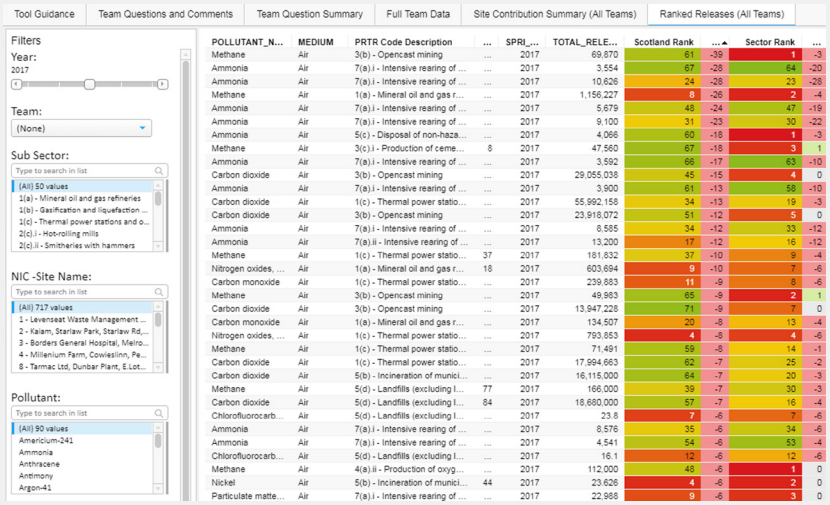

Figure 3: SPRI dashboard highlighting key changes in emissions data in Scotland.

Analysis and Decision-making Use Case: Data Analysis and Visualization of the Environment

SEPA has collected millions of chemistry and ecology samples and supporting data, dating back to the 1960s. SEPA scientists needed a way of analyzing and interrogating this sample data, which they spend considerable time and effort in collecting and storing.

Due to the fact that the existing software was unable to meet this need, SEPA developed a Spotfire analysis tool called Data Analysis and Visualization of the Environment (DAVE). This tool allows anyone to rapidly retrieve the data recorded for any site, samples from SEPA's Laboratory Information Management System (LIMS), and other relevant data sources, and combine and analyze these as time series charts. They can perform statistical analysis such as outlier checking, trend analysis, and step change analysis, as well seasonality and distribution tests. Many of these more complex statistical calculations and metrics are performed by calling out to TIBCO Data Science software to compliment the power of Spotfire analytics.

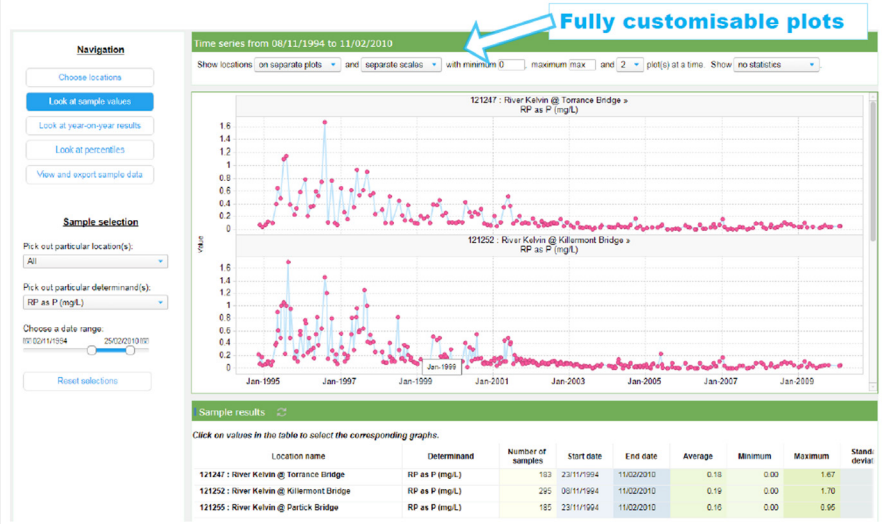

Figure 4: Live on-demand statistical analysis of river and loch water quality data.

All charts are fully customizable so users can visualize data the way they need to see it. They can easily export their analyses as graphs or data, whichever they require. To find data of interest, users are able to search for any location in Scotland by using various filters such as river catchment, team responsible for an area, etc. They can also browse a map of the SEPA sampling network and visually select the sampling locations

of interest for analysis. This means users can now analyze samples down to an entire river to understand the changing conditions and environmental pressures on that river. This tool has become the most used informatics tool in SEPA. It was part of the suite of Spotfire tools that resulted in SEPA winning the 2015 S-Labs award for use of informatics in the laboratory.

This tool puts the entirety of Scotland's sampling at any scientist's fingertips, allowing them to concentrate on research and science rather than battling with unsuitable and difficult tools previously used to retrieve, manipulate, and wrangle data into a usable form. For instance, even the simple task of extracting data for retrieval requests now takes a matter of just minutes.

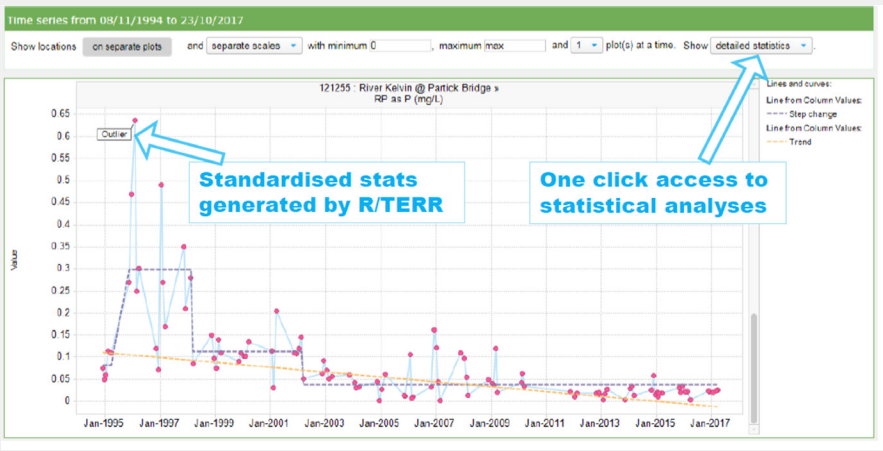

Figure 5: Spotfire software provides one-click access to standardized statistics through TIBCO Data Science software.

Analysis and Decision-making Use Case: Daily Bathing Water Quality Predictions

SEPA provides daily water quality predictions at multiple EU designated bathing waters in Scotland. Initially SEPA used a simple Excel trigger level process. However, with the advent of the new 2016 EU Bathing Waters Directive, which introduced more stringent water quality criteria, as well as multiple year assessments of bathing waters via percentiles, the need to improve and automate predictions was clear.

To tackle this challenge the agency employed machine learning through the use of decision trees using bagging and boosting on a small test set of bathing waters. This was done using R and a library called rpart which produces decision trees.

After the work was published (*Development and use of modelling techniques for real-time bathing water quality predictions*), SEPA needed to take this prototype code and develop and deploy the models for production for the entire country.

To achieve this, it used Spotfire and TIBCO Data Science software to build a clean and simple user interface. This allowed calling the model optimizer with various parameters and settings for any bathing water required. Models could be produced without the need to edit or run code, simply by using Spotfire analytics alone. The model optimizer also produced many models as it tested various parameters, as well as multiple scoring methods to assess the success of those models. By combining Spotfire and TIBCO Data Science software, SEPA massed produced models and tested new settings and data easily and efficiently, and it can also visualize the results, making them much more usable. A key aspect of these Spotfire visualizations was that non-technical business leads could review the models proposed and assess which model met the needs of public health and Scottish bathing waters' policy.

In the 2019 bathing water season, SEPA was in its fourth year of running these models every day to predict bathing water quality. The daily predictions were performed by an Oracle system that calls TIBCO Spotfire Statistical Services software to run the TERR/R models as requested and return the prediction. This provides visitors to Scotland's bathing waters with accurate advice on bathing water quality each day. Spotfire software is then used to track performance and review models for future learning and tuning.

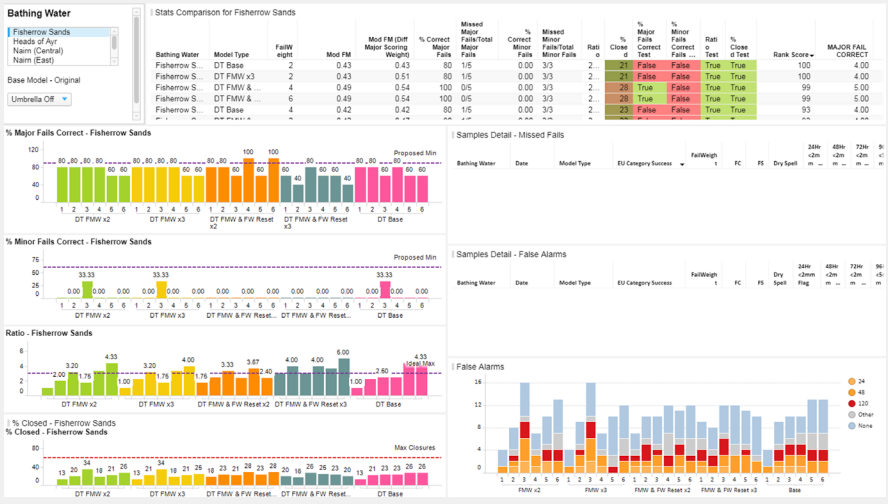

Figure 6: Dashboard for reviewing bathing water model performance used for management decisions.

Reporting and Dashboards Use Case: Managers Dashboard

As Spotfire impact and popularity increased, SEPA's data scientists were using it in new areas such as business planning, estate strategy, and finance. Spotfire use also expanded into the human resources department. Initially, it was used to view absence and work patterns to understand decreases or increases in work productivity and other interesting trends. Analyses were also used to help train and target areas for managers to ensure their teams were healthy and working in a safe and supported environment.

As the analyses grew in popularity, SEPA needed a way for managers to access this crucial staff information on demand, and for the data to be as up-to-date as possible. To achieve this, the agency developed a web service method that allowed Spotfire software to interrogate the HR system safely and securely. Now a manager could inspect all aspects of their team such as contract, employment, absence, holiday, and training information in one place. They could also compare these statistics to other teams, and to SEPA as a whole. SEPA developed a Spotfire rules engine to alert managers to pertinent tasks or issues upon viewing the dashboard. For example, a manager could quickly see if a staff's health and safety training was expiring, they had not used enough holiday allowance, or a contract was expiring.

Manager feedback on this dashboard was extremely positive with new reports being added regularly such as working patterns, information on base offices, and team grade splits. The dashboard is now one of the most used Spotfire tools in SEPA.

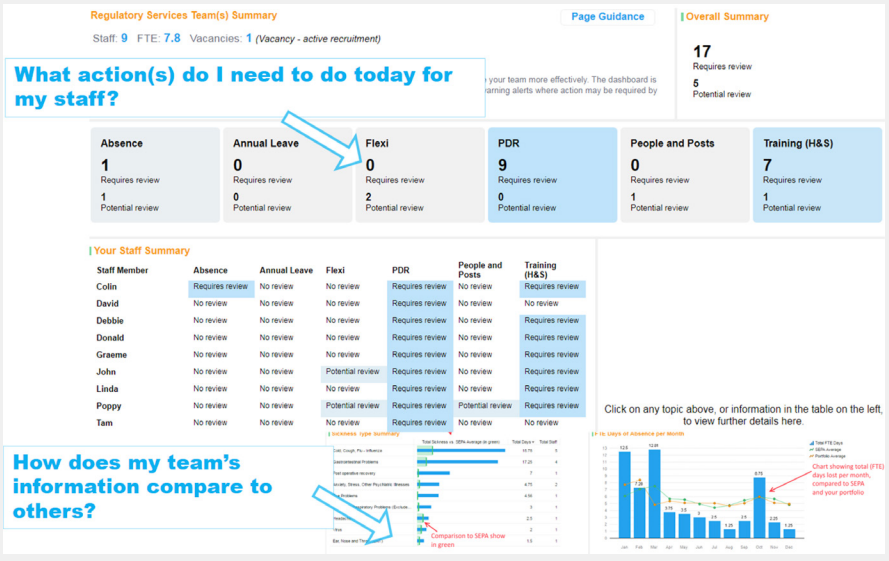

Figure 7: Manager's dashboard giving live actions required and supporting information.

Reporting and Dashboards Use Case: Water Environment Hub

SEPA is required to publish a six-year plan for river basin management planning through the EU Water Framework Directive. Historically, this plan was published as a PDF with several hundred pages of static text, maps, and tables that were difficult to read and impossible to translate for a reader's own context.

The agency wanted to modernize and provide an interactive narrative of Scotland's river basin management plan, while making it much more useful for everyday analysis and decision-making. It also needed to meet the legislative requirements of the European directive.

To achieve this, the team build a highly interactive Spotfire quick analysis tool. This provided the Scottish industry, consultants, and the general public with a question driven way to view and analyze the water environment. A user can ask "What was the condition in 2014 for water quality in my area?" or "Where are the pressures on the environment and when will they be resolved for my geographical area of interest?" The tool fit any level of technical skills or interest.

Using Spotfire scripting capabilities, such as integration with HTML and JavaScript, SEPA provided users with a modern user experience with all the visuals and information they need in a single click. The Spotfire Web Player API allows SEPA to link tools, which provides additional context for the user.

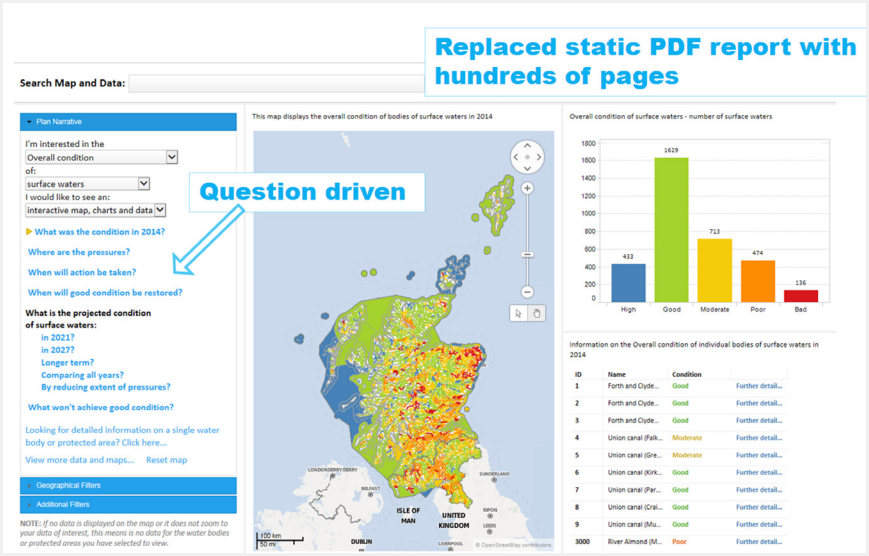

Figure 8: Scotland's Water Environment Hub covering SEPA's six-year water environment plan. https://www.sepa.org.uk/data-visualisation/water-environment-hub/

This tool combined more than 50 datasets, provided a full narrative of the river basin management plan, and met all the EU directive legislative requirements for data publication. It also allowed for an update to be added to the narrative where users can now ask and visualize progress.

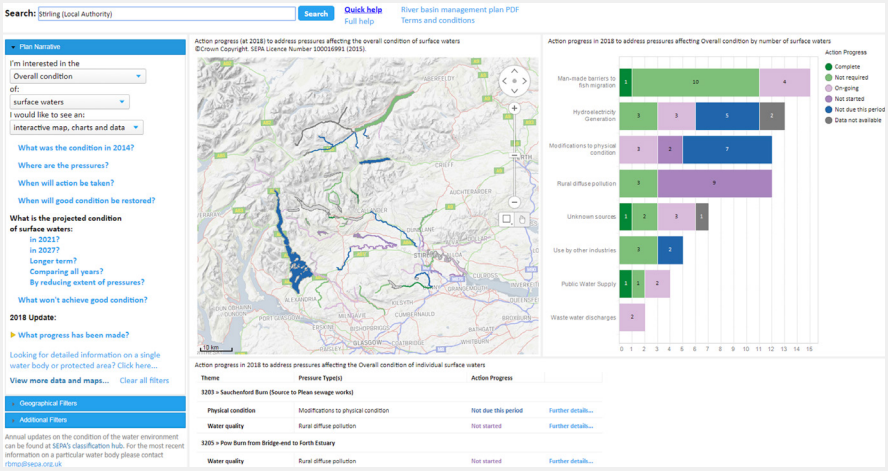

Figure 9: Updated dashboard in 2018 showing progress to date on the six-year environmental plan. https://www.sepa.org.uk/data-visualisation/water-environment-hub/?topic=Progress&display=Data

This tool is now the most used public Spotfire tool SEPA has produced and is used every day by multiple users with nearly 7,500 uses in the last six months.

TIBCO Spotfire and TIBCO Data Science Impact in SEPA

In summary, TIBCO Spotfire key components and ability to cover the whole data science lifecycle has become embedded in SEPA's everyday workings, playing a key role in decision-making, data analysis, and model production. Spotfire software has helped transform SEPA into a data-driven, evidence-based organization. With over 80% of staff now using web-based Spotfire tools through the Spotfire Web Player, the software has gained a great reputation not only for ease of use but also for high performance, rapidly produced tools. Through publishing a large suite of tools, SEPA staff and also the general public and industry, have instant access to numerous datasets from a variety of sources, with the ability to analyze the data as well as standardize its visualization and statistically review any findings. Through the Spotfire powerful analytics and visual engine, combined with the ability to call TERR/R statistical

functions and use the Spotfire Web Player JavaScript API, SEPA has been able to deploy Spotfire capabilities throughout the entire internal organization and three public websites used by external customers, stakeholders, and the general public.

With over 140 tools published and over 35,000 uses of Spotfire tools each year by SEPA staff, TIBCO Spotfire impact is widespread. It is very popular with SEPA data scientists as a desktop tool and a means to publish analysis for others. Future plans include using Spotfire Python and TIBCO Data Science capabilities to run more machine learning and artificial intelligence (AI) methods, using the potential of real-time data from sensors, earth observation and imagery data, and other Internet of Things (IoT) data. Employing these new and modern techniques through Spotfire software, while deploying to the cloud, will further increase the impact of data science and TIBCO software for SEPA.

What can TIBCO make possible for you? Talk to an expert today at tibco.com/contact-us/sales

Global Headquarters
3307 Hillview Avenue
Palo Alto, CA 94304
+1 650-846-1000 TEL
+1 800-420-8450
+1 650-846-1005 FAX
www.tibco.com

07Sep2021