Observing Water-Related Events for Evidence-Based Decision-Making

Joao Pita Costa * *****, M. Besher Massri *, Inna Novalija *, Ignacio Casals del Busto **, Iulian Mocanu ***, Maurizio Rossi ****, Jan Šturm *, Eva Erzin *, Alenka Guček *, Matej Posinković *, Marko Grobelnik * *****

* Institute Jozef Stefan, Slovenia, ** Aguas del Alicante, Spain, *** Apa Braila, Romania, **** Ville de Carouge, Switzerland, ***** Quintelligence, Slovenia

ABSTRACT

With the awareness of a changing climate impacting our sustainability, and in line with the European Green Deal initiative or the Sustainable Development Goal 6 addressing water, the industry, society and local governments are requiring reliable and comprehensive technology that can provide them an overview to water events to anticipate problems and the tools to analyse best practices appropriate to solve them. This paper presents the NAIADES Water Observatory (NOW), a digital solution offering a series of analysis and visualisations of water-related topics, helping users to extract important insights in relation to the water sector. Taking advantage of heterogeneous data sources, from the media and social media landscape, to published research and global/local indicators. Through collaboration with local water resource management institutions, the NWO was configured to local priorities and ingests local datasets to better fit the needs of decision-makers.

CCS CONCEPTS

• Real-time systems • Data management systems • Life and medical science

KEYWORDS

Water Resource Management, Smart Water, Observatory, Water Digital Twin, Elasticsearch, Streamstory

1 Introduction

The water sector is facing rapid development towards the smart digitalisation of resources, much motivated and supported by the UN's global initiative for the Sustainable Development Goal 6. In that context, the efforts to address the specific challenges related to water management data and priorities multiply globally. There are several "digital twin" systems dedicated to water, each of which focuses on the different aspects of the digitalisation of signals to support water management companies, as well as water "observatories". These are usually meant as Geographical Information Systems that showcase the different aspects of water resources through time.

Within the scope of the European Commission-funded project NAIADES [1] focusing on the automation of the water resource management and environmental monitoring, we propose a slightly different approach that integrates heterogeneous data sources to try and solve common research questions, as well as to support water management companies in their current problems. This solution is named NAIADES Water Observatory (NWO), available at naiades.ijs.si, putting together: (i) real-time information from multilingual world news on water topics; (ii) data visualisation of water-related indicators through time, sourced from the datasets associated with the Sustainable Development Goal 6 (water) and other UN data (see Figure 1); and (iii) scientific knowledge from published biomedical research on water-related topics (e.g., water contamination).

Due to the rapidly growing awareness of the sustainability challenges that we are facing in Europe and worldwide in the context of water resource management, there has been much work done to develop systems that are able to collect information about the available water and even simulate and forecast that in the near future. But these are usually geolocation-based systems ingesting water-related data to enable real-time monitoring of resources and usage [x] [y] [z], and thus much different than the water observatory that we are proposing in this paper. The typical example is GoAigua system [4], a digital twin technology allowing, e.g., the city of Valencia to optimize its water management at the network level, improving efficiency in daily operations, plan real-time scenarios, and make some prediction on its future behaviour [5].

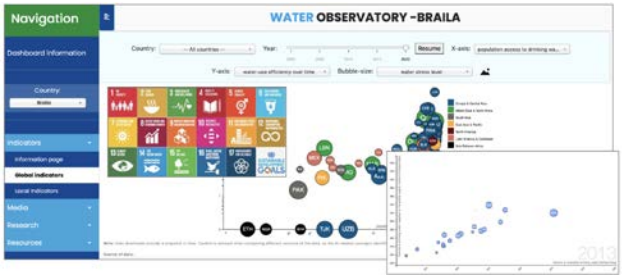

Figure 1: **Visualisation of water-related indicators within Spain to complement the global indicators view ingesting data from, e.g., U.N. and the World Bank.**

2 A data-driven solution for water events

The proposed Water Observatory enables extraction of insightful water-related information, configured to use case priorities and needs from the data integration of

heterogeneous sources. This includes information from social media when the weather is favourable for floods and the historical information from news and published research on these weather-related events and how to make better decisions to solve them.

This is complemented by data ingested from global and local indicators (i.e., datasets at regional level), showcasing the observation of water-related datasets linked to SDG 6 at global and country levels that can help us observe changes and trends. The NAIADES Water Observatory enables the user to explore the information provided by published science and the success stories that can be used in decision-making and water education at the local level (i.e., showcasing the resources and problematics of the region).

In this approach, the water data sensing is done over dynamic open data sources that serve as digital sensors (news, social media, indicators, publications, weather forecasts). This data is then integrated and visualised, each in its tab, addressing specific topics of interest. The observatory is thus composed of all that heterogeneous data coming in at different frequencies. The interactions between those data sources to solve common problems make it a Water Digital Twin. The envisioned examples include the analysis of best practices in water events in, e.g. Braila, identified in the news and explored over the published research, or the alerts triggered by weather conditions and observed over social media on a water event. The questions we are trying to solve with this innovative technology are, e.g., if we can predict water shortages in a certain region given the historical data; or if we can identify early signals of water-related problems from social media (see Figure 2).

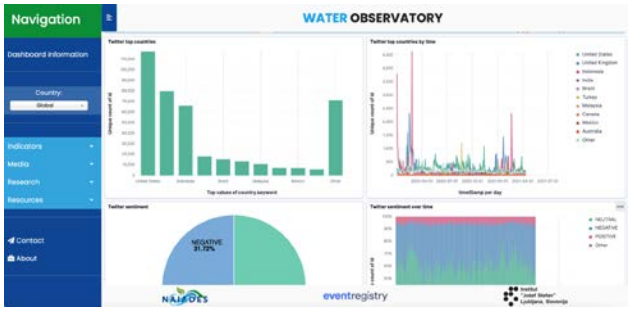

Figure 2: **Analysis of the sentiment in water-related posts in Twitter and the relation to consumer satisfaction and water-related events**

All of the views of this observatory, each of which represents digital solutions on their own, are configured to the local priorities of the NAIADES users as a Proof of Concept, showing that each can address specific conditions.

- Indicators: adding to the global UN indicators, we are ingesting curated open datasets that have regional information about water topics of interest to the stakeholder
- Media: each location has its own news and social media streams configured to priorities and aspects of the news that stakeholders define as topics of interest (e.g. floods)
- Research: similarly to the media sources, the research topics allow for some customisation to fit the needs of the local user better
- Resources: the natural resources information provided for exploration is geolocated to the regions of interest to the user of the platform

It is relatively easy to include new use cases and corresponding workspaces after the discussions on user priorities that will allow us to configure the information presented and making it meaningful.

3 Addressing the challenges of tomorrow

With the range of views provided at the observatory, the problems addressed can be of complex nature and cover a range of concerns and workflows. The different ICT capabilities available across the water sector require intuitive and meaningful technologies to ensure the usefulness of the contribution to the Community. The target users of the NWO seem to belong to three main scenarios with different workflows that can be supported by the developed technology:

1. Water resource management: using the provided information in the resolution of problems related to weather events to understand how their actions are perceived by the consumers and to explore successful scenarios in similar cases
2. Local governments: to help evidence-based decision-making using open data, better synchronise to SDG6 and other guidelines and evaluate commitments in time
3. General public: for water education with a local context, in aspects that matter to the local population, based on parts of the Water Observatory that can be open to public

The priorities in the European Union are rapidly changing towards sustainability and environmental efficiency, transversally to most domains of action. The European Commission's Green Deal [3] aiming for a climate-neutral Europe by 2050 and boosting the economy through green technology provides a new framework to understand and position water resource management in the context of the challenges of tomorrow. The NAIADES Water Observatory will not only contribute to the improvement of European sustainability in water-related matters but will also assign the local actors on the water resource management an active role in that. The NAIADES Water Observatory provides the user of the NAIADES platform, as earlier extensively discussed, with the global and local insight that can be transformed into business intelligence, and help companies to steer their strategies towards customer satisfaction. We will be

describing selected views of this observatory through the verticals (or views) News – Indicators – Biomedical, first at the level of the specific dashboards that constitute the tabs in the online instance, and then by the extended exploratory instances, including public instances and APIs, for each of the three verticals.

Figure 3: **The global view of the pilot 1 over usage and data sources.**

These dashboards come together to provide the user with a global perspective in real-time, where five different tiers of usability are made available (see Figure 3). The tiers allow for the extended usability of the Water Observatory, Transversally to the data sources available.

4 System description and architecture

The NWO offers user exploratory dashboards for the further investigation over news, to get deeper into the indicators ingested, and to explore the biomedical research on water contamination in detail. Moreover, each of the three dashboards have versions built to be exposed by, e.g., iframe through a publicly available channel that can be used for integration in high management KPI-monitoring dashboards. Furthermore, we also offer a part of the information in these through APIs easily integrable with our own systems.

The *Indicators* view provides the user with interactive data exploration tools that allow for the KPI-monitoring over several water-related topics that include the SDG 6, the World Bank Open Data, the UN data, etc. In this module we also ingest regional data sources that include local indicators, addressing the user's priorities. Considering their well-established data types, the data integration is possible and, whenever limitations appear due to lack or poor quality of the data, the dataset is pre-processed to allow for data completion (whenever possible), or at least the improvement of data quality.

The *Media* view provides the user with the real-time news monitoring over water-related topics (such as Water Scarcity and Water Contamination), and the analysis of water-related tweets based on data visualisation modules. Based on the news engine *Eventregistry* [7] this view provides the system with a continuous stream of news articles, sourced from RSS-enabled sites across the world. From the data management module the real-time news data is accessed by the news dashboard that can be configured by the NAIADES user to tune the topics of interest in the configuration web app. To further explore a water-related topic, the NWO provides a dashboard for the analysis of social media posts in Twitter (see Figure 2), collected in a real-time frequency, where sentiment is analysed, related concepts are extracted and it is possible to access the raw tweets or apply several filters.

Finally, the biomedical module allows for the exhaustive exploration of water contamination information from scientific research articles published worldwide and available through the MEDLINE biomedical open dataset [9] and the Microsoft Academic Graph [8]. The MEDLINE dataset is collected from the official FTP source made available by the North American National Library of Medicine (NLM) over an XML dump and uploaded to the elasticSearch data management system through a python script, the Microsoft Academic Graph dataset is collected from an Azure container with the data biweekly updated by the Microsoft Research team. The data management is based on the elasticsearch technology [2, useful for both the interactive data visualisations and the Indicators Explorer view. The latter allows the NAIADES user to explore the raw data through template visualisations, use a Lucene-based query that can leverage the loaded metadata, and easily build visualisation modules that can define a new dashboard of data visualisation modules. The dataset is then called over and HTTP API by the SearchPoint technology [6] to load the dataset and respective metadata. thus allowing for powerful Lucene-based queries and further interaction over a movable pointer. This will lead to the refinement of the search of information that can then be extended over the Biomedical Explorer, which feeds over the same dataset through Kibana, but also allows for the analysis of raw data, or the easy construction of data visualisation modules from templates, and for an interactive data visualisation dashboard. All the mentioned dashboards can be made publicly available through, e.g., iframe to be integrated in high-management KPI monitors.

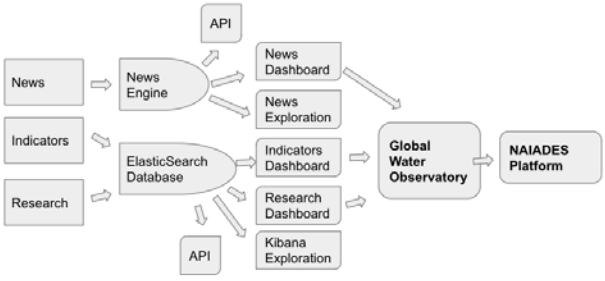

Figure 4: **System architecture of the NAIADES Water Observatory showcasing the relation between used technologies and NOW views**

5. Conclusions and further work

In this paper we discussed the technological development and research opportunities motivated by the emerging need to support decision-makers with evidence from open data that can retract best practices and answer questions from the collected data, bringing the digitalisation of the water sector to a new level.

The potential to ingest complementary local data and configure global sources to parameters addressing local priorities provides a local dimension that is being explored close to the priorities of the NAIADES data providers within water resource management institutions. It will also be exploring the insights driven by the appropriate aspects of chosen datasets, e.g., between news data and focused interactions through Twitter for weather-related events when the weather is likely to be favourable to their cause (see Figure 5). There are many systems that can collect business intelligence data, but we believe that the "digital twin"-type of insight is in the interaction between these data streams.

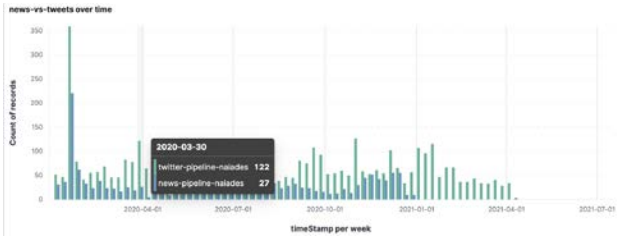

Figure 5: **Preliminary data analysis of the relation between news and tweets on water-related events and their relations with other topics (e.g., weather).**

Further development to the NAIADES Water Observatory, will be providing the users with tools to explore the impact of natural resources as, e.g., the weather, as well as predictions on the levels of the available bodies of water, based on ingested weather data from the ECMWF (on humidity, temperature and rainfall) and other open data sources. This will help the users to have some insight on the impact of the climate crisis in regions that directly relate to their water resources. We will use a sophisticated engine - Streamstory [6][10] - to explore the states of that weather-related data and short/medium term predictions on aspects of that data (see Figure 6).

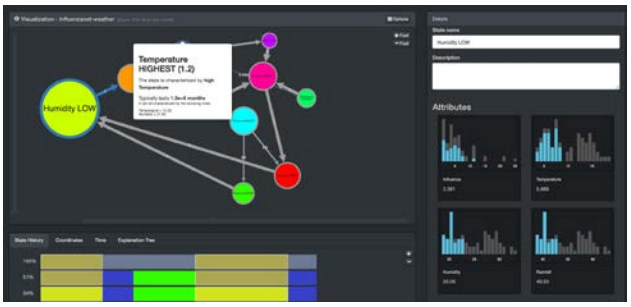

Figure 6: **The multi time-series analysis of the weather parameters, using Markov chains in complex data visualisation through the Streamstory technology** [9]**.**

ACKNOWLEDGMENTS

We thank the support of the European Commission on the H2020 NAIADES project (GA nr. 820985).

REFERENCES

[1] CORDIS, "NAIADES Project". [Online]. Available: https://cordis.europa.eu/project/id/820985 [Accessed 1 9 2020].
[2] Elasticsearch, "Elasticsearch," 2020. [Online]. Available: https://www.elastic.co/elasticsearch/. [Accessed 1 9 2020].
[3] European Commission, "European Green Deal," 2019. [Online]. Available: https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal_en. [Accessed 1 9 2020].
[4] Idrica, "GoAigua: Smart Water for a Better World," 2020. [Online]. Available: https://www.idrica.com/goaigua/. [Accessed 1 9 2020].
[5] Idrica, "Digital Twin: implementation and benefits for the water sector," 19 2 2020. [Online]. Available: https://www.idrica.com/blog/digital-twin-implementation-benefits-water-sector/. [Accessed 1 9 2020].
[6] Institute Jozef Stefan, "Streamstory". [Online]. Available: http://streamstory.ijs.si/. [Accessed 26 8 2021]
[7] G. Leban, B. Fortuna, J. Brank and M. Grobelnik, "Event registry: learning about world events from news," Proceedings of the 23rd International Conference on World Wide Web, pp. 107-110, 2014.
[8] Microsoft, "Microsoft Academic Graph". [Online]. Available: https://www.microsoft.com/en-us/research/project/microsoft-academic-graph/. [Accessed 26 8 2021]
[9] National Library of Medicine, "MEDLINE". [Online]. Available: https://www.nlm.nih.gov/medline/medline_overview.html. [Accessed 26 8 2021]
[10] L. Stopar, P. Škraba, M. Grobelnik, and D. Mladenić (2018). StreamStory: Exploring Multivariate Time Series on Multiple Scales. IEEE transactions on visualization and computer graphics 25.4: 1788-1802.