

Glaciar: Software de Visualización de Datos de Recursos Hídricos y Atmosféricos

Pablo Inchausti, Pablo Romanos, Mariana Barturen, y Bibiana Rossi

Universidad Argentina de la Empresa,
Facultad de Ingeniería y Ciencias Exactas, Buenos Aires, Argentina
{pinchausti, promanos, mbarturen, birossi}@uade.edu.ar

Resumen. Se desarrolló un software de visualización para facilitar el análisis de la calidad de los recursos hídricos y atmosféricos a partir de datos públicos del medio ambiente. Se describen los sistemas de calidad del aire y del agua que el software permite integrar en las visualizaciones. Se explica el procedimiento de extracción de los datos desde cinco plataformas de datos abiertos. El software se distribuye en formato open source para facilitar su extensión, por ejemplo, en la visualización de datos generados por soluciones de IoT o recolectados mediante el uso de drones equipados con sensores.

Palabras claves: open data, visualización, recursos hídricos, recursos atmosféricos.

1 Introducción

La gestión de los recursos naturales es un ámbito específico donde las tecnologías de la información contribuyen a generar información de valor para la toma de decisiones. Gracias a la evolución tecnológica, los datos son cada vez más abundantes en volumen, en variedad y se generan a mayor velocidad. En medio de un entorno complejo para obtener información de valor, las herramientas de visualización aportan claridad, siempre que reflejen con veracidad los datos que están representando [1].

La visualización de los datos de los recursos hídricos y atmosféricos presenta desafíos que es necesario abordar con el asesoramiento de especialistas en medio ambiente y conociendo la bibliografía especializada. Para realizar visualizaciones adecuadas se debe entender, por ejemplo, como se determina la calidad para el uso sustentable de los recursos naturales, como se seleccionan las variables de un sistema de calidad o como varían los valores guía según el uso definido como apto para un recurso natural.

Respecto a los niveles de calidad, se deben considerar estándares internacionales, como los definidos por la Agencia de Protección Ambiental de los Estados Unidos (US EPA) [2] o las guías de calidad de la Organización Mundial de la Salud (OMS) [3]. En relación al lugar donde fueron tomados los datos, se debe tener en cuenta que dice la normativa nacional, provincial o municipal respecto a la calidad de los recursos naturales en su ámbito de competencia.

Y como parte del análisis, después de representar los datos de las variables hídricas y atmosféricas como una serie temporal en un gráfico, resulta importante poder identificar sus atributos, es decir, la tendencia, la estacionalidad, los ciclos, el "ruido", la presencia de valores atípicos (los "outliers"), las discontinuaciones repentinas, además de la necesidad de poder ajustar adecuadamente la escala de los datos [4].

1.1 Contribución y Aplicaciones

La apertura de los datos públicos genera oportunidades de innovación, pero distintos obstáculos por causas tecnológicas, políticas o legales repercuten en una mala calidad de los datos públicos que dificulta el proceso de generar valor. [5]

El objetivo del presente trabajo es ofrecer una herramienta de "visualización de datos de recursos hídricos y atmosféricos" para facilitar el proceso de análisis de datos públicos del medio ambiente, utilizando variables y sistemas de calidad predefinidos.

La herramienta obtiene los datos a través de procesos automatizados de extracción, transformación y carga (ETL) y dispone de distintos sistemas de calidad con valores guía que el usuario puede integrar en las visualizaciones. La herramienta también permite explorar las series temporales con funcionalidades como la selección de rangos, el manejo de escalas y la definición de valores límites para identificar o excluir los "outliers".

El diferencial de la herramienta es su especialización en la visualización de datos públicos del medio ambiente, con funcionalidades de valor añadido para el análisis de este tipo de datos.

Entre sus aplicaciones, la herramienta permite incrementar la transparencia de los datos públicos al implementar facilidades de análisis y visualización que no son ofrecidas en los portales que distribuyen los datos. La herramienta se puede utilizar en soluciones de IoT (Internet de las Cosas) para visualizar, por ejemplo, datos de monitoreo o recolectados mediante el uso de drones equipados con sensores.

1.2 Variables Definidas por Alcance

Las variables definidas para determinar los niveles de calidad se detallan en la Tabla 1:

Tabla 1. Variables de calidad definidas en el alcance

Variable	Abreviatura	Unidad de medida	Símbolo	Calidad
Temperatura	Temp.	Grados celsius	°C	Agua
Grado de acidez	pH	Escala de pH. Rango 0 a 14	pH	Agua
Oxígeno disuelto	OD	Miligramos por litro	mg/L	Agua
Potencial de reducción	ORP o Redox	Milivoltios	µV	Agua
Conductividad	Cond.	Microsiemens por cm.	µS/cm	Agua
Monóxido de Carbono	CO	Partes por millón	ppm	Aire
Dióxido de Carbono	CO ₂	Partes por millón	ppm	Aire
Monóxido de Nitrógeno	NO	Partes por billón	ppb	Aire
Dióxido de Nitrógeno	NO ₂	Partes por billón	ppb	Aire

La selección de variables hídricas se basó en el antecedente de trabajos similares. La empresa Libelium, fabricante de dispositivos de IoT, utilizó estas cinco variables en el monitoreo de la calidad del agua del Acuario Fluvial de Zaragoza, el más grande de su tipo en Europa [6]. Asimismo, las variables temperatura, pH, ORP y conductividad formaron parte de una propuesta de IoT para un "Sistema inteligente de monitoreo de la calidad del agua" [7].

En la selección de variables atmosféricas, se incluyeron contaminantes reconocidos por sus efectos sobre la salud por la OMS [3] e incluidos dentro del índice AQI (Air Quality Index) definido por EPA [8], y el contaminante dióxido de carbono (CO₂) reconocido por sus efectos nocivos sobre el medio ambiente al ser uno de los principales gases de efecto invernadero (GEI) [9].

1.3 Calidad del Agua

La calidad del agua puede ser determinada a través de múltiples sistemas que seleccionan diferentes variables (químicas, físicas y/o biológicas) para decidir si el agua es apta para un determinado fin, por ejemplo, el consumo humano, el uso industrial, el riego o la conservación de la vida acuática [10].

En la Tabla 2 se enumeran los sistemas de calidad del agua considerados con los valores guía de las variables.

Tabla 2. Valores guía en los sistemas de Calidad del Agua

Sistema	Variable	Min.	Prom.	Max.	Unidad	Ref.
Valores guía en bibliografía Ref.	Temp.	15,00	25,00	35,00	°C	[11]
	pH	6,50	7,00	8,00	pH	[12]
	OD	7,54	9,00	12,75	mg/L	[11]
	Redox	100,00	150,00	200,00	µV	[7]
	Cond.		100,00		uS/cm	[7]
OMS (Std. drinking water)	Redox	650,00			µV	[13]
ACUMAR Res. 46-E/2017	Uso I A	Apta protección biota y uso recreativo c/contacto directo.				[14]
		Temp.		35,00	°C	
		pH	5,00	6,00	9,00	pH
	Uso I B	OD	5,00		mg/L	
		Apta protección biota.				
		Temp.		35,00	°C	
	Uso II	pH	5,00	6,00	9,00	pH
		OD	5,00		mg/L	
		Apta para actividades recreativas con contacto directo.				
	Uso III	Temp.		35,00	°C	
		pH	5,00	6,00	9,00	pH
		OD	5,00		mg/L	
	Uso IV	Apta para actividades recreativas sin contacto directo.				
		Temp.		35,00	°C	
		pH	5,00	6,00	9,00	pH
	Uso V	OD	4,00		mg/L	
		Apta para actividades recreativas pasivas.				
		Temp.		35,00	°C	
	Uso VI	pH	6,00	9,00	pH	
		OD	2,00		mg/L	
		Apta para actividades recreativas pasivas.				

La calidad del agua también puede ser determinada a nivel general utilizando índices de calidad del agua (ICA) que reducen la información a un único valor dentro de una escala y por lo tanto resulta más sencillo de comprender y comparar por el público en general y las autoridades políticas. Uno de los índices más difundidos es el NFS-WQI que utiliza nueve variables con distinto peso para obtener su valor. [15]

Los valores guía varían según el uso definido como apto para el recurso natural. Por ejemplo, la Resolución 46-E/2017 de ACUMAR (Autoridad de Cuenca Matanza Riachuelo) distingue entre los diferentes tipos de uso a la protección de la biota o para realizar actividades recreativas con y sin contacto directo. [14]

A continuación se describen las variables del alcance con sus valores guía.

Temperatura: en aguas naturales se considera como condiciones estándares en valores de 25 °C (± 10 °C o 15 °C) para a una atmósfera de presión o con diferencias de unas pocas atmósferas [11]. La temperatura influye en gran medida en la calidad del agua, por ejemplo, afectando la cantidad de oxígeno disuelto (OD) y otros gases disueltos en agua que disminuyen con el aumento de la temperatura. El OD es necesario en cantidades suficientes para la supervivencia de peces y muchas otras formas de vida acuática [12].

Oxígeno Disuelto (OD): los valores de equilibrio se encuentran tabulados en la bibliografía y es una función que depende de la temperatura y la presión. Por ejemplo, para aguas dulces a 15 °C de temperatura y a una atmósfera de presión, el valor de equilibrio de OD es de 7,54 mg/L (miligramos por litro) [11].

Grado de acidez o pH: describe la condición ácida o básica (también llamada “alcalina”) de los líquidos y se mide en una escala de 0 a 14, siendo 7 el valor neutro. En aguas naturales el valor de pH generalmente oscila entre 6,5 y 8,5 [12].

Potencial de oxidación-reducción (Redox o ORP): es un índice de la capacidad oxidante (valores positivos) o reductora (negativos) de una solución. Para la vida acuática son adecuados valores de ORP entre 100 μ V y 200 μ V (micro voltios) [7]. Para el agua potable, la Organización Mundial de la Salud (OMS) estableció en el año 1971 un valor de ORP de 650 μ V indicando que "provocará la casi instantánea desactivación de los virus incluso en altas concentraciones" [13].

Conductividad: es la capacidad del agua para conducir una corriente eléctrica y se relaciona directamente a la salinidad, que aporta más iones disueltos en agua. En el agua potable, la conductividad está en el orden de los 100 μ S/cm (micro Siemens por centímetros) y en agua de mar, entre los 55.000-60.000 μ S/cm. Valores por fuera de esas escalas pueden indicar contaminación [7].

1.4 Calidad del Aire

Las variables atmosféricas seleccionadas en el alcance, (CO, CO₂, NO, NO₂) son consideradas en diversos sistemas para determinar los niveles de calidad en el aire.

Los sistemas de calidad del aire tienen un punto de referencia en el año 1970 en los Estados Unidos, cuando se promulga la ley federal denominada “Clean Air Act” que regula sobre la contaminación del aire y se le requiere a la Environmental Protection Agency (EPA), fundada en ese mismo año, la tarea de establecer los NAAQS (National Ambient Air Quality Standards) para los contaminantes considerados perjudiciales para la salud pública y el medioambiente. La EPA define los NAAQS para seis contaminantes principales: Monóxido de Carbono (CO), Plomo (Pb), Dióxido de Nitrógeno (NO₂),

Ozono (O_3), Partículas en suspensión (PM) y Dióxido de Azufre (SO_2) y los denomina como el "criterio de contaminantes del aire". [2]

Para facilitar la comunicación con el público en general, la EPA define el índice AQI (Air Quality Index) que se calcula a partir de los cinco contaminantes del criterio y se traduce a una escala con colores verdes, amarillos y rojos que indican los efectos de la calidad del aire sobre la salud [8].

Por su parte, la Organización Mundial de la Salud (OMS) publica los valores de referencia de los contaminantes PM, O_3 , NO_2 , SO_2 y CO en las "Guías de calidad del aire de la OMS" [3] y en "Environmental Health Criteria 213: Carbon Monoxide" [16].

En la República Argentina, la "LEY 1356 de Calidad Atmosférica" define valores de referencia en el ámbito de aplicación de la Ciudad Autónoma de Buenos Aires [17] que afectan a los datos de "Buenos Aires Data". En el ámbito de aplicación de la provincia de Buenos Aires, la "Ley 5965 Protección a las fuentes de provisión y a los cursos receptores de agua y a la atmósfera" [18] afecta los datos asociados al Municipio de Bahía Blanca.

En la Tabla 3 se enumeran los sistemas considerados y los valores guía de las variables de calidad del aire seleccionadas.

Tabla 3. Valores guía en los sistemas de Calidad del Aire

Sistema	Variable	Min.	Prom.	Max.	Unidad	Ref.
GCBA Ley 1356	CO			8,00	ppm	[17]
	NO_2			200,00	ppb	
BS. AS. Ley 5965	CO			9,00	ppm	[18]
	NO_2			200,00	ppb	
EPA. AQI Index	CO	0,00	4,50	9,40	ppm	[8]
	NO_2	0,00	54,00	101,00	ppb	
OMS (EHC 213)	CO	0,00		9,00	ppm	[16]
OMS (Guía 2015)	NO_2	0,00		101,00	ppb	[3]

2 Orígenes de Datos

Los orígenes de datos secundarios se seleccionaron buscando variedad de escenarios para validar la plataforma y que sean relevantes dentro de la República Argentina (RA).

El primer origen de datos proviene de la Capital Federal de la RA y su red de cuatro estaciones de monitoreo ambiental [19]. El segundo origen es sobre la calidad del aire del Municipio de Bahía Blanca que tiene un importante Polo Petroquímico en sus proximidades [20]. El tercer origen son datos del Río Charles monitoreado por la EPA cada 15 minutos [21]. El cuarto origen son tres dataset de la Base de Datos Hidrológica Integrada (BDHI) de la RA, con datos desde el año 2011 [22]. El quinto origen es el registro de emisiones de CO_2 desde 1850, por países y por sectores, publicados en el sitio del World Resources Institute (WRI) [9].

Tabla 4. Orígenes de datos secundarios

Organización	Origen de datos	Datos	Ref.
GCABA	Calidad del Aire. Buenos Aires Data	Aire	[19]
- Estaciones:	La Boca, Centenario, Córdoba, Chacabuco		
Mun. BB	Calidad del Aire. Datos Abiertos Mun. Bahía Blanca	Aire	[20]
US EPA	Datos de calidad del agua de Charles River	Agua	[21]
MIOPyV	Base de Datos Hidrológica Integrada (BDHI)	Agua	[22]
- Río estación:	1. Paraguay - Puerto Pilcomayo, Pcia. de Formosa 2. Gualeguay - Villaguay, Pcia. de Entre Ríos. 3. Paraná de las Palmas - Zárate, Pcia. de BsAs		
WRI	Emisiones CO ₂ . Climate Data Explorer	Aire	[9]
- Países:	Argentina, Bolivia, Brasil, Chile, Alemania		
- Sectores:	Electricidad, Transporte, Construcción, Otros		

2.1 Datos Públicos Abiertos

Los orígenes de datos secundarios provienen de plataformas de datos abiertos o de portales que distribuyen sus datos siguiendo los principios de “Open Data”.

El Open Data es un movimiento que logra su consolidación en el año 2009 y entre sus principales promotores se destaca Tim Berners-Lee, el inventor de la WWW, que en una conferencia TED del año 2009 animó a su audiencia a exigirles a sus gobiernos “Raw Data Now” (“Datos Crudos Ahora”). En ese mismo año, los gobiernos de los Estados Unidos (data.gov) y el Reino Unido (data.gov.uk) lanzaron los primeros portales de datos abiertos, un ejemplo que luego es imitado por muchos otros países. [5]

En el año 2010, Tim Berners-Lee, propuso una escala de cinco estrellas, que se describe en la Tabla 5, para evaluar la calidad en Open Data, otorgando más estrellas a medida que los datos están disponibles de formas más amplia y fácil de reutilizar [23]:

Tabla 5. Esquema de 5 estrellas de Open Data. [23]

Estrellas	Descripción
☆	Datos disponibles en línea con una licencia abierta (en cualquier formato).
☆☆	Datos disponibles en un formato estructurado, ejemplo CSV, Excel.
☆☆☆	Datos disponibles en formatos no propietarios, ejemplo CSV.
☆☆☆☆	Datos con identificadores únicos de recursos (URI) que se pueden vincular.
☆☆☆☆☆	Datos vinculados con otros datos para proveer contexto.

En el presente trabajo, los orígenes de datos secundarios utilizados verifican que sus datos están (1) disponibles en la Web bajo una licencia abierta (2) publicados en un formato estructurado y (3) utilizan un formato abierto como CSV o JSON. Es decir, un valor de tres estrellas en la escala de calidad de Open Data de Tim Berners-Lee.

3 Materiales y Métodos

3.1 Procesos ETL

La Arquitectura de Datos del presente trabajo se nutre de los procesos ETL (extracción, transformación y carga) asociados a cada origen de datos secundario. Los estados y las acciones de los procesos ETL se describen en la Fig. 1.

Estado	Acción	Fuentes de datos secundarias					Espacio de Trabajo
		BS AS Data	Bahía Blanca	Charles River	BDHN	Emisiones CO2	
0 Inicial		ETL 1	ETL 2	ETL 3	ETL 4	ETL 5	No determinado
	Limpiar: - Espacio de Trabajo	>>	>>	>>	>>	>>	
1 Limpio		<<	<<	<<	<<	<<	Vacio
	Definir URL's	1 (una) url	6 (seis) url	4 (cuatro) url	3 (tres) url	1 (una) url	
	Descargar URL's	descargar	descargar	descargar	descargar	descargar	
	Descomprimir	unzip file			unzip file	unzip file	Archivos
2 Descargado		<<	<<	<<	<<	<<	Descargados
	Normalizar: - Títulos				Títulos en: tercera fila	Títulos en: segunda fila	
	Normalizar: - punto decimal	Normalizar a: punto decimal			Normalizar a: punto decimal		
	Normalizar: - separador de campo				Normalizar a: separador coma		
	Unir archivos		misma serie: unificar 6 archivos	misma serie: unificar 4 archivos			
	Definir series	4 estaciones => 4 series			3 estaciones => 3 series	5 países => 5 series	
	Separar series	separar series			separar series	separar series	
	Columnas: - indicar AirQ	CO, NO2	CO, NO, NO2			CO2	
	Columnas: - indicar WaterQ			Temp, ph, OD	Temp, ph, OD, Redox, Cond		
	Columnas: - Extraer Variables	>>	>>	>>	>>	>>	Archivo
3 Transformado		<<	<<	<<	<<	<<	Normalizado
	- Conexión a la BD	>>	>>	>>	>>	>>	
	- Persistir Series	>>	>>	>>	>>	>>	
4 Almacenado		Nueva Base de Datos Poblada					

Fig. 1. Transiciones de estados y acciones de los cinco procesos ETL sobre sus orígenes de datos.

Los procesos ETL inician con la limpieza del área de trabajo para aislarse de ejecuciones anteriores y luego descargan los archivos desde la (o las) url del origen de datos. Si el archivo descargado está comprimido, se realiza una acción de descomprimir.

Los procesos ETL continúan con acciones de transformación para normalizar a su archivo de origen, por ejemplo, eliminando filas extra de títulos o traduciendo el formato de los números al uso de punto decimal. Puede suceder que varias series de datos estén contenidas en un mismo archivo, como en el caso de los dataset de Buenos Aires Data por utilizar cuatro estaciones de monitoreo [19]. También puede suceder que una misma serie esté segmentada en varios archivos, como en el caso del dataset de Bahía Blanca, que utiliza un archivo diferente para cada año [20]. En el primer caso, el proceso ETL1 separa las cuatro series del archivo y en el segundo caso, el proceso ETL2 reconstruye la única serie unificando los archivos.

Una vez extraídas las series de datos, cada proceso ETL realiza las transformaciones necesarias sobre su serie para converger al formato y la estructura de los datos definida para el almacenamiento en la base de datos de la plataforma "Glaciar".

En la Figura Fig. 2 se muestra el detalle del funcionamiento del proceso ETL1. La implementación permite la ejecución automatizada y repetible de los procesos ETL.

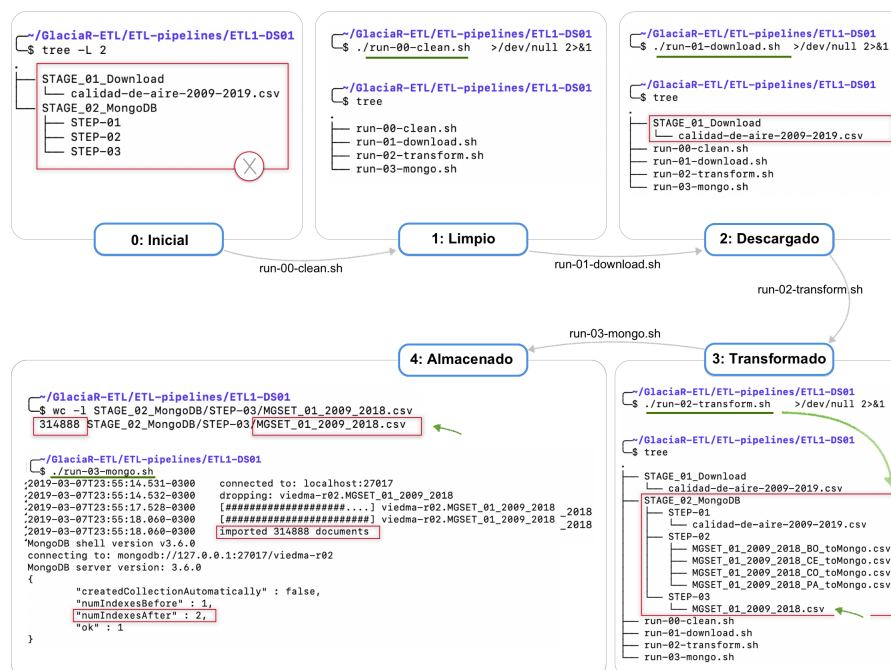


Fig. 2. Detalle de implementación del proceso ETL1 asociado al origen "Buenos Aires Data".

3.2 Tecnologías Utilizadas

En el desarrollo se utilizaron productos open source sin incurrir en costos de software de terceros, haciendo posible que la plataforma "GlaciaR" también sea distribuida en formato open source para beneficio de los proyectos que la integren en su solución.

Entre los productos comerciales incluidos está la librería de gráficos "amChart" [24] que ofrece una licencia de uso gratuito sin limitaciones, aún en proyectos open source, si se mantiene visible un pequeño logo de atribución. El servicio Google Maps [25] para la georreferenciación de los datos también se utilizó dentro de los niveles de uso gratuito.

La arquitectura física de los componentes del software se organizó siguiendo el modelo propuesto por el denominado "MEAN Stack" [26] que utiliza en todas las capas tecnologías basadas en Javascript para reducir la complejidad del desarrollo, es decir:

- **Angular:** en la construcción del frontend. [27]
- **Node.js:** en la construcción de los servicios del backend. [28]
- **Express:** para implementar la API REST. [29]
- **MongoDB:** como medio de persistencia de documentos en formato JSON. [30]

En la interfaz de usuario se utilizó, además, Bootstrap [31] para el comportamiento responsivo en distintos dispositivos y la plantilla de aplicaciones ngx-admin [32].

El entorno de ejecución es provisto por plataformas que ofrecen los servicios de procesamiento y almacenamiento en la nube y fueron utilizados dentro de los niveles de uso de sus capas gratuitas. Estas plataformas son:

- **Heroku:** plataforma de ejecución de aplicaciones en la nube que provee el entorno de ejecución para el frontend en Angular y el backend en Node.js. [33]
- **MongoDB Atlas:** ofrece instancias de la base de datos Mongo como servicio. [34]
- **GitHub:** el repositorio de código fuente se utilizó integrado a la plataforma Heroku para el despliegue automático aplicando prácticas “DevOps” de CI/CD. [35]

3.3 Software Open Source

El nombre "Glaciar" que recibe la herramienta alude simbólicamente a la capacidad que tienen los glaciares de guardar en su interior valiosos registros de la calidad del aire y del agua a lo largo de los siglos.

El software se encuentra publicado en la dirección www.glaciar.org con su código fuente compartido en GitHub [36] como se observa en la Fig. 3. La distribución se realiza de forma libre y gratuita bajo Licencia Apache 2.0 que es una de las licencias permisivas más utilizadas en proyectos open source [37].

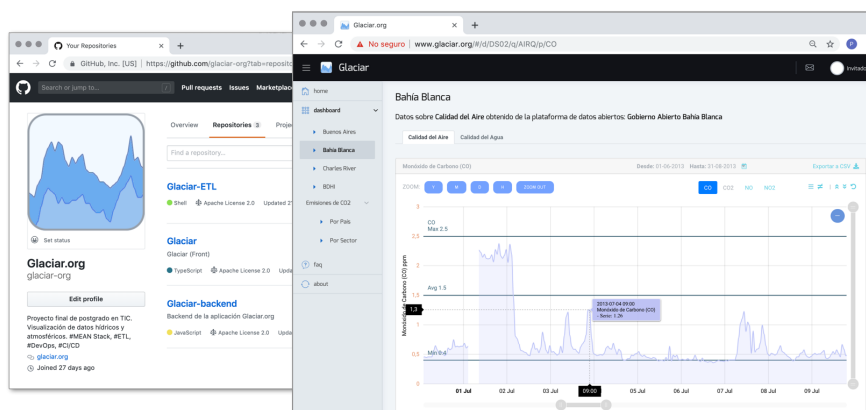


Fig. 3. Publicación: código fuente en GitHub y dirección web "www.glaciar.org"

3.4 Entrevistas con expertos

Desde el inicio del proyecto se reconoció la importancia de contar con el asesoramiento de expertos que puedan guiar sobre cómo se determina la calidad en los recursos naturales y, por ende, que funcionalidades y visualizaciones en la plataforma ofrecen mayor valor.

Para indagar sobre esta problemática, siguiendo un enfoque cualitativo y recolectando datos por medio de entrevistas, se consultaron a dos expertos en medio ambiente con grado de doctorado y un experto en análisis de datos con experiencia en el dominio. Las entrevistas realizadas fueron de tipo semiestructurada, es decir, se proponía una guía, pero existía la libertad de ampliar conceptos con preguntas adicionales. [38]

En la primera parte de las entrevistas, los expertos respondieron a preguntas genéricas sobre el concepto de calidad en el medio ambiente, los sistemas y las variables. Hubo consenso en que la calidad depende del propósito de uso, por ejemplo, en el caso del agua, recreación, consumo, acuicultura, etc. Sobre los sistemas de calidad, la primera Dra. destacó la importancia del sistema que considera el "estado trófico" de los cuerpos de agua sobre la base del fósforo y nitrógeno e hizo mención a los estándares de la EPA y la OMS para calidad del aire y del agua. La segunda Dra. comentó su experiencia determinando la calidad del agua a través de imágenes por satélite midiendo variables como la turbidez del agua, la concentración de clorofila y la producción primaria.

En la segunda parte de las entrevistas se buscó la opinión de los expertos respecto al aporte de valor de la plataforma "Glaciar". Entre los aspectos positivos, la primera Dra. destacó que compila información, que permite ver tendencias y que la visualización es dinámica. La segunda Dra. destacó que es fácil de utilizar, que las visualizaciones resultan atractivas, y que se ofrece de forma abierta y global. El analista de datos destacó que "las herramientas que se especializan en un dominio generan valor porque permiten hacer foco en el análisis en vez de la limpieza de los datos y la programación". Entre los aspectos a mejorar, se recomendó, incorporar funcionalidades de "estacionalidad" para comparar períodos de tiempo diferentes: por ejemplo, mes a mes, semana a semana, año a año o agrupar los datos por estaciones anuales, o temporadas de lluvias o de sequías.

En la tercera parte de las entrevistas, se busca relevar de los expertos que funcionalidades debería tener la plataforma para ofrecer valor agregado al análisis. La primera Dra. sugiere que el usuario disponga de sistemas de referencia como los estándares de la EPA y la OMS. La segunda Dra. indica la posibilidad de descarga de los datos, e información general de cada dataset, como su licencia de uso. El analista de datos destaca la importancia del estudio como series temporales y el análisis de los valores atípicos.

4 Resultados

Se describe a continuación las funcionalidades de la plataforma para facilitar el análisis de los datos del medio ambiente.

Manejo de las series temporales: la librería "amChart" seleccionada facilita el uso dinámico de múltiples series. De forma adicional, se implementa la selección de fechas, de variables y de distintos tipos de zoom (anual, mensual, diario o por hora). También se implementa la exportación de los datos, el escalado vertical y la opción de ver u ocultar los umbrales. En la Fig. 4 del dataset de "Buenos Aires Data" se detallan las opciones.

Sistemas de calidad y umbrales: los sistemas de calidad con los valores guía de las Tabla 2 y Tabla 3 se encuentran presentes en la plataforma. El usuario puede integrar los sistemas en las visualizaciones y calibrar los valores mínimo, promedio y máximo, como se observa en la Fig. 5.

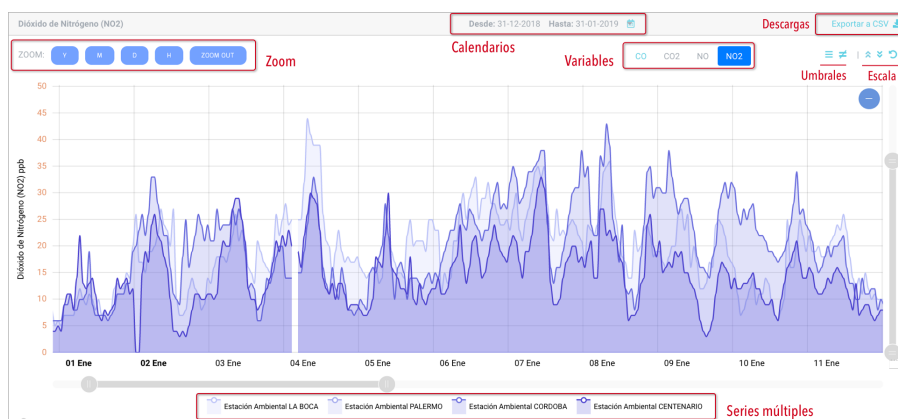


Fig. 4. Opciones sobre series: zoom, escalas, rangos, series múltiples, cambio de variables.

En cada sistema de calidad, se incluye un vínculo externo que indica el fundamento de los valores guía.

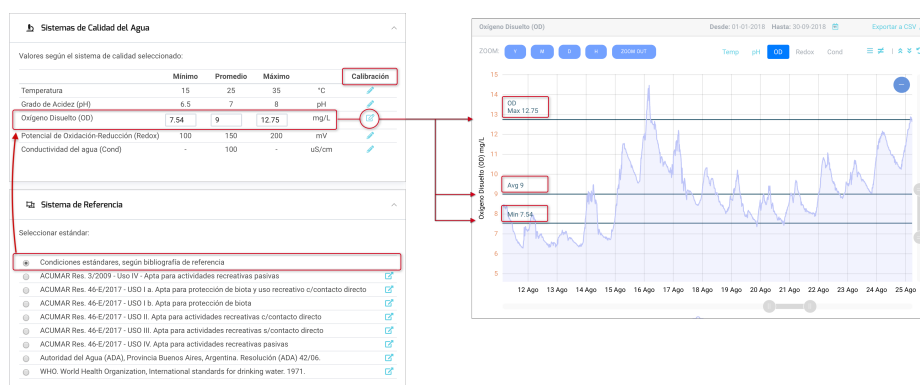


Fig. 5. Ejemplo de visualización y calibración de los umbrales en la variable "Oxígeno Disuelto".

Información sobre los dataset: se brinda información sobre el rango de fechas abarcado y la frecuencia de capturas. También se ofrece información respecto a la densidad de los datos, es decir, la cantidad de registros por año, por mes o por día y la ubicación georreferenciada de las estaciones de captura. En la Fig. 6 del dataset "Buenos Aires Data", por ejemplo, se observa que hay 2.880 registros en el mes de abril del 2014 y la ubicación georreferenciada de las cuatro estaciones de monitoreo.

Visualización y descarte de valores atípicos: el usuario dispone de la posibilidad de definir umbrales inferiores y superiores que considere como "outliers" para descartarlos o destacarlos, como se observa en el ejemplo de la Fig. 7. Una posible extensión del trabajo es incorporar algoritmos automáticos de detección de outliers en las series temporales.

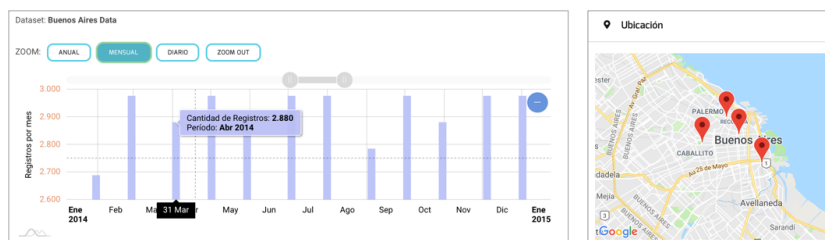


Fig. 6. Densidad de los datos y georreferenciación de las estaciones de captura.

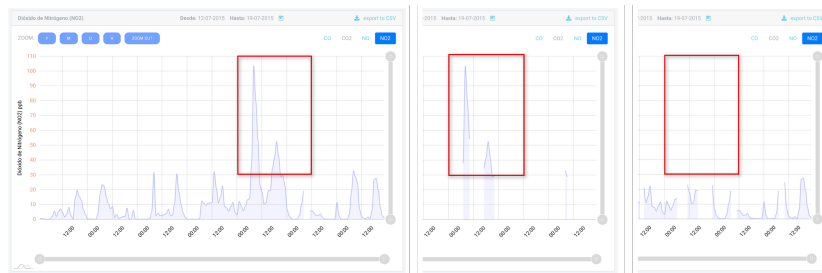


Fig. 7. Definición de outliers mayores a 30 ppb para NO_2 . Resaltado y descarte.

5 Conclusiones y Discusión

La apertura de los datos públicos representa una oportunidad para construir nuevas soluciones, pero el proceso de extracción de información de valor puede resultar complejo por el esfuerzo que demanda tanto entender los datos como procesarlos.

En el proceso del desarrollo de una herramienta para simplificar el análisis de los datos medioambientales se observa la complejidad que encierra el significado del término "calidad" cuando se refiere al uso sustentable de los recursos naturales. En la búsqueda de los sistemas de referencia para determinar la calidad, se extraen de la bibliografía especializada valores guías que, a través de la herramienta, el usuario tiene disponible.

En cuanto a los datos, el volumen no resultó ser un factor significativo, pero sí su variedad, que demandó procesos ETL dedicados a cada origen para hacer converger los datos a un formato común, en la base de datos de la plataforma.

En la construcción de las funcionalidades de visualización, para maximizar el valor, se tuvieron en cuenta las recomendaciones de los expertos que aportaron con su experiencia.

Finalmente, el código del producto "Glaciar" compartido como proyecto open source abre las puertas a la colaboración de nuevos datos, sistemas, funcionalidades y también a la adaptación de la plataforma en nuevos proyectos que decidan integrarla.

Agradecimientos. Los autores agradecen a la Universidad Argentina de la Empresa (UADE) y al Instituto de Tecnología (INTEC) por el apoyo brindado en el presente trabajo realizado como proyecto final de "Maestría en Tecnología y Comunicaciones", articulado en el Proyecto de Investigación "Visualización de Datos Hídricos y Atmosféricos" (PI18T04).

Referencias

1. Monino, J.L., Sedkaoui, S.: Big Data, Open Data and Data Development, vol. 3. John Wiley & Sons, Inc. (2016)
2. US EPA: NAAQS Table (2014). [En línea]. Disponible en: <https://www.epa.gov/criteria-air-pollutants/naaqs-table>, [Accedido: 17-07-2019].
3. OMS. Organización Mundial de la Salud: Guías de calidad del aire de la OMS relativas al material particulado, el ozono, el dióxido de nitrógeno y el dióxido de azufre. In: WHO/SDE/PHE/OEH/06.02. Ediciones de la OMS (2005), https://apps.who.int/iris/bitstream/handle/10665/69478/WHO_SDE_PHE_OEH_06.02_spa.pdf
4. Chatfield, C.: Time-Series Forecasting. Chapman and Hall/CRC (2001). <https://doi.org/10.1201/9781420036206>
5. Stagars, M.: Open Data in Southeast Asia: Towards Economic Prosperity, Government Transparency, and Citizen Participation in the ASEAN. Springer International Publishing (2016). <https://doi.org/10.1007/978-3-319-32170-7>
6. Libelium: Water Quality Monitoring in Europe's Largest Fluvial Aquarium (2015). [En línea]. Disponible en: <http://www.libelium.com/water-quality-monitoring-europe-largest-fluvial-aquarium-zaragoza>, [Accedido: 17-07-2019].
7. Prasad, A.N., Mamun, K.A., Islam, F.R., Haqva, H.: Smart water quality monitoring system. In: 2015 2nd Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE). pp. 1–6 (Dec 2015). <https://doi.org/10.1109/APWCCSE.2015.7476234>
8. US EPA: 40 CFR Appendix G to Part 58 - Uniform Air Quality Index (AQI) and Daily Reporting. [En línea]. Disponible en: <https://www.govinfo.gov/app/details/CFR-2016-title40-vol6/CFR-2016-title40-vol6-part58-appG>, [Accedido: 17-07-2019].
9. WRI. World Resources Institute: CAIT Climate Data Explorer (2017). [En línea]. Disponible en: <https://cait.wri.org>, [Accedido: 17-07-2019].
10. UNEP/WHO: Water Quality Monitoring: A practical guide to the design and implementation of freshwater quality studies and monitoring programmes. Edited by Jamie Bartram and Richard Ballance (1996)
11. Hem, J.D.: Study and interpretation of the chemical characteristics of natural water. United States Government Printing Office, Washington, 3a ed. (1985)
12. Vigil, K.M.: Clean Water: An Introduction to Water Quality and Water Pollution Control. Oregon State University Press, 2a ed. (2003)
13. WHO. World Health Organization: International standards for drinking water. H.M.S.O., London, 3a ed. (1971)
14. ACUMAR. Autoridad de Cuenca Matanza Riachuelo: Resolución 46-E/2017 (2017). [En línea]. Disponible en: <http://servicios.infoleg.gob.ar/infolegInternet/anexos/270000-274999/273042/norma.htm>, [Accedido: 17-07-2019].
15. Tyagi, S., Sharma, B., Singh, P.: Water quality assessment in terms of water quality index. American Journal of Water Resources **1**, 34–38 (01 2013). <https://doi.org/10.12691/ajwr-1-3-3>
16. WHO. World Health Organization: EHC. Environmental Health Criteria 213: Carbon Monoxide. 2a ed. (1999)
17. GCBA. Gobierno de la Ciudad de Buenos Aires: DECRETO N° 198/GCABA/06 (2006). [En línea]. Disponible en: https://www.buenosaires.gob.ar/areas/leg_tecnica/sin/normapop09.php?id=83624&r1=1, [Accedido: 17-07-2019].
18. Poder Ejecutivo de la Provincia de Buenos Aires: DECRETO 3395/96 (1996). [En línea]. Disponible en: <http://www.gob.gba.gov.ar/legislacion/legislacion/96-3395.html>, [Accedido: 17-07-2019].

19. DGCIyGA. Dirección General de Calidad Institucional y Gobierno Abierto: (2019). [En línea]. Disponible en: <https://data.buenosaires.gob.ar>, [Accedido: 17-03-2019].
20. MBB. Municipio de Bahía Blanca: Portal de Datos Abiertos del Municipio de Bahía Blanca (2019). [En línea]. Disponible en: <http://datos.bahiablanca.gob.ar>, [Accedido: 17-03-2019].
21. US EPA: Basic Information about the Charles River Buoy (2015). [En línea]. Disponible en: <https://www.epa.gov/charlesriver/basic-information-about-charles-river-buoy>, [Accedido: 17-07-2019].
22. SIPH. Secretaria de Infraestructura y Política Hídrica: Red Hidrológica Nacional - República Argentina (2019). [En línea]. Disponible en: <http://bdhi.hidricosargentina.gov.ar>, [Accedido: 11-03-2019].
23. Berners-Lee, Tim: Linked Data - Design Issues (2006). [En línea]. Disponible en: <https://www.w3.org/DesignIssues/LinkedData.html>, [Accedido: 17-07-2019].
24. amCharts: JavaScript Charts & Maps (2019). [En línea]. Disponible en: <https://www.amcharts.com/>, [Accedido: 11-03-2019].
25. Google LLC: APIs de geolocalización | Google Maps Platform (2019). [En línea]. Disponible en: <https://cloud.google.com/maps-platform>, [Accedido: 17-07-2019].
26. Dayley, B., Dayley, B., Dayley, C.: Node.js, MongoDB and Angular Web Development: The definitive guide to using the MEAN stack to build web applications. Pearson Education (2017)
27. Google LLC: Angular (2019). [En línea]. Disponible en: <https://angular.io>, [Accedido: 17-07-2019].
28. Node.js Foundation: Node.js (2019). [En línea]. Disponible en: <https://nodejs.org>, [Accedido: 17-07-2019].
29. StrongLoop, IBM, expressjs.com contributors: Express - Node.js web application framework (2017). [En línea]. Disponible en: <https://expressjs.com>, [Accedido: 17-07-2019].
30. MongoDB, Inc.: The most popular database for modern apps (2019). [En línea]. Disponible en: <https://www.mongodb.com>, [Accedido: 17-07-2019].
31. Otto, Mark and Thornton, Jacob and Bootstrap contributors: Bootstrap (2019). [En línea]. Disponible en: <https://getbootstrap.com>, [Accedido: 17-07-2019].
32. Akveo LLC: ngx-admin: Free Open Source admin dashboard template based on Angular, Bootstrap (2018). [En línea]. Disponible en: <https://akveo.github.io/ngx-admin>, [Accedido: 08-05-2019].
33. Salesforce.com: Cloud Application Platform | Heroku (2019). [En línea]. Disponible en: <https://www.heroku.com/>, [Accedido: 08-05-2019].
34. MongoDB, Inc.: MongoDB Atlas (2019). [En línea]. Disponible en: <https://cloud.mongodb.com>, [Accedido: 08-05-2019].
35. Rossel, S.: Continuous Integration, Delivery, and Deployment: Reliable and faster software releases with automating builds, tests, and deployment. Packt Publishing (2017)
36. GitHub, Inc: Glaciar.org (2019). <https://github.com/glaciar-org>, [Accedido: 17-07-2019].
37. ASF. The Apache Software Foundation: Apache License, Version 2.0 (2004). [En línea]. Disponible en: <https://www.apache.org/licenses/LICENSE-2.0>, [Accedido: 17-07-2019].
38. Hernández Sampieri, R., Fernández Collado, C., Baptista Lucio, P.: Metodología de la investigación. McGraw-Hill Education, 6a ed. (2014)