HYDROGRAPHISCHE NACHRICHTEN

Journal of Applied Hydrography 10/2023 HN 126

BIM und digitale Zwillinge

DEUTSCHE HYDROGRAPHISCHE GESELLSCHAFT e.V.
DHyG

DOI: 10.23784/HN126-05

Managing a growing ocean of data

An article by MEIKE KLISCHIES, ADRIAN NEUMANN, FRITHJOF HENNEMANN and PHILIPPE TAKE

Ocean data become increasingly important and data volumes are continuously expanding. This raises the demand for smart data management applications that provide scalability and flexibility to adjust to the needs of the project, stakeholders, users and data. This article presents the TrueOcean marine data platform, developed and operated by North.io GmbH. The platform emerged from research collaborations targeting unexplored ordnances (UXO) in the coastal waters and is now available for commercial application. TrueOcean combines cloud services, web technology, big data processing and advanced data and user management to support and optimise maritime projects. Its cloud-based infrastructure allows to easily scale projects as they proceed, data volumes increase and data types change. Its modular nature streamlines workflows and provides the necessary flexibility to adjust to the dynamic maritime project environment. On-going research collaborations introduce new cutting-edge technologies into the platform and will expand its tool palette, including artificial intelligence driven processing and analysis modules.

marine data platform | cloud technology | smart data management | big data | AI-driven analysis
Meeresdatenplattform | Cloud-Technologie | smartes Datenmanagement | Big Data | KI-gesteuerte Analyse

Meeresdaten gewinnen immer mehr an Bedeutung und die Datenvolumina vergrößern sich ständig. Daran gekoppelt ist ein wachsender Bedarf an intelligentem Datenmanagement, das die nötige Skalierbarkeit und Flexibilität liefert, um den Ansprüchen von Projektzielen und -partnern, von Nutzern und den Daten selbst gerecht zu werden. Dieser Artikel stellt TrueOcean vor, eine marine Datenplattform, die von der North.io GmbH entwickelt und betrieben wird. Hervorgegangen aus Forschungsprojekten zum Thema Munitionsaltlasten in Küstengewässern, ist die Plattform mittlerweile in kommerzieller Nutzung. Dabei kombiniert TrueOcean Cloud-Dienste, Web-Technologien und Big-Data-Analysen mit modernem Anwendender- und Datenmanagement, um maritime Projekte effizienter zu gestalten. Die zugrunde liegende Cloud-Infrastruktur ermöglicht eine einfache Skalierung von Projekten mit wachsenden Datenvolumina und sich verändernden Datentypen. Der modulare Aufbau der Plattform bietet die nötige Flexibilität für maritime Projekte und ihr dynamisches Umfeld. Die Palette an verfügbaren Modulen wird durch Forschungskooperationen stetig erweitert und neueste Spitzentechnologien wie das automatisierte Bearbeiten und Analysieren von Daten mittels künstlicher Intelligenz können verfügbar gemacht werden.

Authors
Meike Klischies, Adrian Neumann, Frithjof Hennemann and Philippe Take work at North.io GmbH in Kiel.

ptake@north.io

Introduction

The UN Ocean Decade (oceandecade.org), initiatives such as the Seabed 2030 project (Mayer et al. 2018) and the societal transition towards renewable energies with a growing number of offshore wind farms and underwater infrastructure (Bugnot et al. 2021) have been moving ocean data into the spotlight (e.g., Aschenbrenner and Winkler 2019). Ocean data has become increasingly available and data volumes are continuously expanding (Fig. 1; Felden et al. 2023; Carbotte et al. 2022; Tanhua et al. 2019; Wölfl et al. 2019). This raises the demand for the application of new technologies, including cloud computing, big data analytics and artificial intelligence (e.g., Mellone et al. 2022; Schwartz-Belkin and Portman 2023; Ziolkowski et al. 2022).

Cloud technology distributes computing and storage, which makes projects scalable and allows to adjust the data management to the needs of a project. At the same time, cloud solutions provide the accessibility and transparency needed in growing projects with an increasing number of users and stakeholders. Nevertheless, it requires a deep understanding of geoinformatics, algorithms, security and scalability to harvest the cloud potentials.

To ensure unlimited scalability of geospatial and maritime data projects, TrueOcean combines cloud technology with web-based software services. The cloud services provide the infrastructure for secure, transparent and scalable data and user management. TrueOcean's software services tailor geodata management to the individual demands of a project and can be adjusted to dynamic for growing project environments.

How TrueOcean came into play – the TrueOcean history

TrueOcean emerged during North.io's work in research regarding unexploded ordnance (UXO)

in Germany's coastal waters (e.g., daimonproject.com). The research projects had to combine different processing workflows and handling of big data, which revealed a lack of capabilities in terms of cloud services, big data processing and highly automated processes in the maritime world (Wehner and Frey 2022). This grounds on several factors, including the complexity of underwater data sets, the heterogeneity of sensor systems, an unconsolidated market with a large variety of small and medium enterprises, high dynamics in the development of technology and the lack of standardisation. North.io addressed all these challenges and invented a totally new way of maritime big data handling by developing and applying the TrueOcean platform technology in a modular structure (Fig. 2).

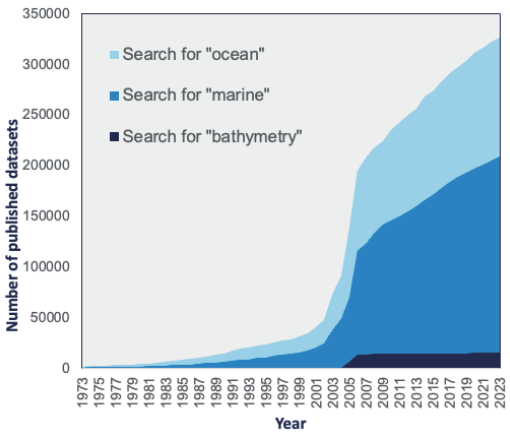

Fig. 1: Number of data sets published and accessible in PANGAEA (an open access library for archiving, publishing and distributing georeferenced data from earth and environmental sciences: Felden et al. 2023) showing the growing amount of available marine data in a single repository. Today's search for the keyword »ocean« results in 326,310 individual data sets (excluding 5,234 unpublished ones), for »marine« in 208,774 data sets (excluding 4,103 unpublished ones) and »bathymetry« in 15,189 data sets (excluding 156 unpublished ones). The figure considers only the number of data sets available, not their volumes, which may differ significantly

Strengths and bottlenecks of the TrueOcean solution

Conventional or open source geodata software offer a broad and manifold palette of tools that can easily overwhelm the user. Finding the appropriate workflow is typically a complicated task (e.g., Kastrisios et al. 2023). It becomes a true challenge, when multiple stakeholders collaborate in projects or when the project becomes long-term, data volumes increase, and new data formats, new hardware or (unwanted) software updates must be integrated.

The approach of »Software as a Service« puts the needs of the users and stakeholders into focus and tailors the software tools and resources accordingly (Fig. 2). The flexibility of this modular structure streamlines collaboration, especially in the long term and in large projects where multiple

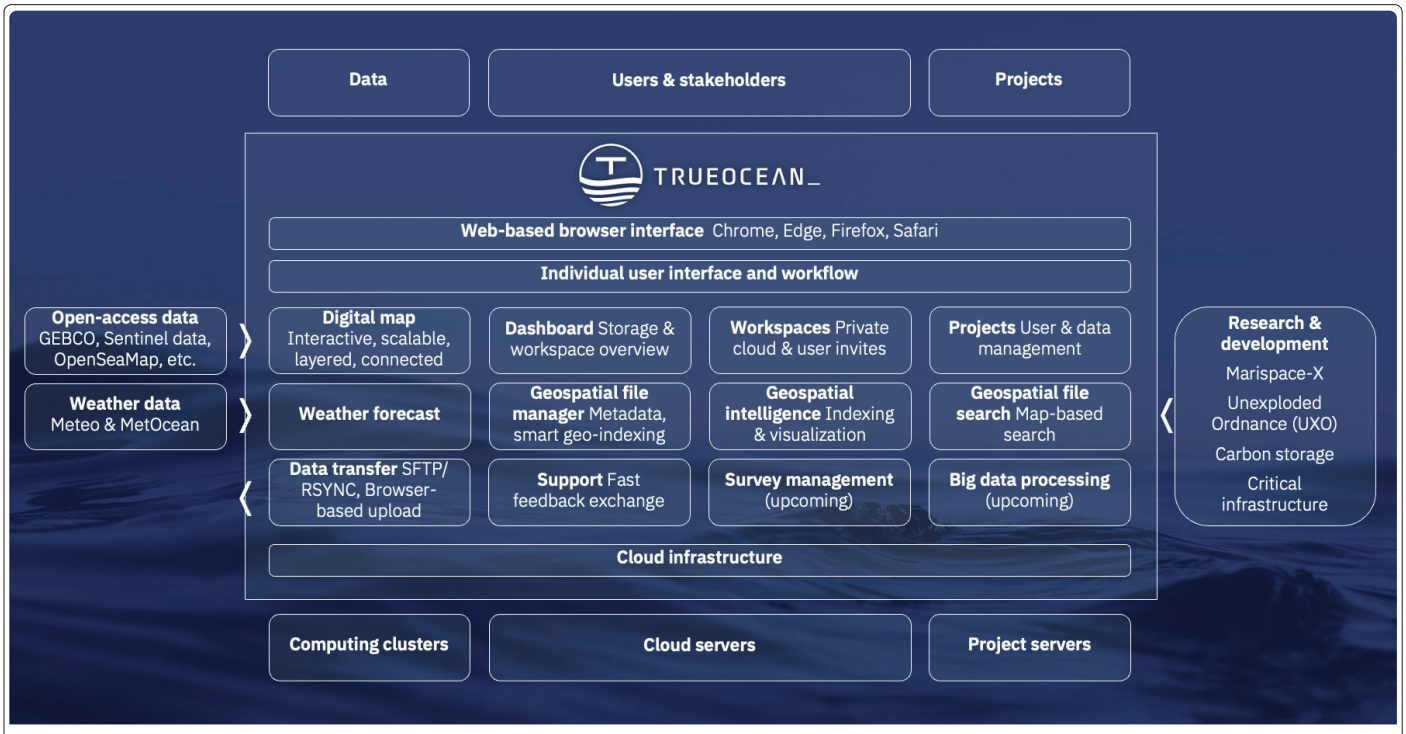

Fig. 2: Concept of the TrueOcean marine data platform: Built upon a cloud infrastructure, its modular nature allows to tailor the tools to the needs of the users, stakeholders, projects and data. In addition to commercial applications, intensive research collaborations guarantee a continuing development and implementation of new technologies into the platform

stakeholders are involved. The approach may limit the options to solve problems individually, but it encourages to document and share each issue via the »Support« module (Fig. 2). This way, the software also acts as a communication hub in collaborative projects.

An additional challenge for the scalability of projects arises from the diversity and heterogeneity of ocean data (Carbotte et al. 2022; Wehner et al. 2022). The diversity affects the data formats, but also the according metadata information (if available) and data management. It originates, amongst other factors, from the diversity of highly specialised underwater sensors and the heterogeneously applied post-processing workflows. Scalability requires extracting essential information and translating it in a standardised and fast manner. For this, TrueOcean builds upon decentralised cloud technologies, its years of experience dealing with ocean data and continuing research collaborations (Fig. 2).

A bottleneck for web-based services is the low internet bandwidth offshore. Therefore, TrueOcean supports a variety of compression options, beyond the classical ZIP-encoding. These options significantly reduce the data volume, especially for large files, but also allow them to be included into large computing environments for upscaling (e.g., Parquet-files and compressions to be built into the Apache Spark environment; »Geospatial file manager« module, Fig. 2). Due to the deep understanding of TrueOcean in maritime data management and analytics, the software can already convert a large variety of common sensor formats (e.g., GSF, XTF, ALL, KMALL, S7K, JSF, RAW, SDF, LAS/LAZ) and compress ocean data in a standardised, scalable manner. Further, the TrueOcean web-view (»Geospatial intelligence« module) enables a first inspection of the data before downloading or sharing it (via the »Data transfer« module). This way, TrueOcean fosters a highly selective data transfer that will free bandwidth from the back-and-forth traffic of quality control checks. Also, the cloud provides transparency across all collaborators involved by having a central repository with version control (»Projects« module). This also frees bandwidth, as it prevents the circulation of similar data products in parallel. The cloud service also guarantees the security of data, which becomes particularly important for data regarding critical infrastructure or UXO.

Use case: TrueOcean for offshore wind farms

The efficiency of offshore wind farms relies on the longevity of the infrastructure. To ensure such efficiency and longevity, both careful planning and constant monitoring of the infrastructure are crucial. Both produce large volumes of data that come in various formats and structures. Therefore, data management and workflows must be able to cope with the large and growing data volumes, but at the same time, must adapt to the shift in scope over time. The modular structure of TrueOcean enables to tailor the data pipelines to the needs of the project and of the collaborators. The modules connect by standardised data formats (Wehner et al. 2022) and the common cloud infrastructure. Developed tools range from classical quality control and conversion to tools for automated analyses and predictions supported by trained artificial intelligence (AI) models. AI-modules are currently in development and will include, for example, risk assessments or automated, dynamic track planning for autonomous survey campaigns that assist the decision-making process. Such AI-modules also allow for specific training to be further adjusted to the needs of a project.

TrueOcean's future

With its built-in ability for scalability, TrueOcean is set up for the growing interest in the ocean environment and the increase in ocean data. To foster a data-driven, informed and sustainable use of this sensitive environment, North.io collaborates with scientific partners and established tech and big-data companies, such as Intel and Nvidia. These collaborations develop new, cutting-edge modules for an informed interpretation of the increasing amount of data. Already today, classical tools are replaced by AI-driven and big-data modules that perform faster and with less resources than the classical ones. TrueOcean's newest modules can calculate quality control metrics directly on the data-rich point cloud of raw multibeam echo sounder data (Wehner and Frey 2022). Calculations directly on point clouds provide control metrics for each individual data point and avoid interpolation or smoothing effects that are commonly introduced by gridding the data into raster products. Further, the analysis of the raw point cloud allows an immediate quality check after data acquisition and skips the classical, time-consuming post-processing including manual inspection of the data. In addition, the precision of conducting quality control on point clouds is way higher since no data reduction approaches e.g., aggregation for raster representation were conducted.

The manual inspection of multibeam data is a huge limiting factor on productivity. Therefore, a planned collaboration project aims to replace the manual inspection with an AI-module that automatically »cleans« the raw multibeam point data from outliers. Such multibeam data typically provide the basis of each ocean endeavour and become increasingly available, also in open-access data compilations, such as the Global Multi-Res-

olution Topography project (gmrt.org; Ryan et al. 2009). As multibeam data are easy to acquire, raw data volumes are typically large. The AI-approach will increase the efficiency of multibeam data processing, so the focus can be moved towards interpreting multibeam data. //

References

Aschenbrenner, Marie; Gordon M. Winder (2019): Planning for a sustainable marine future? Marine spatial planning in the German exclusive economic zone of the North Sea. Applied Geography, DOI: 10.1016/j.apgeog.2019.102050

Bugnot, Ana B.; Mariana Mayer-Pinto; Laura Airoldi et al. (2021): Current and projected global extent of marine built structures. Nature Sustainability, DOI: 10.1038/s41893-020-00595-1

Carbotte, Suzanne M.; Suzanne O'Hara; Karen Stocks et al. (2022): Rolling Deck to Repository: Supporting the marine science community with data management services from academic research expeditions. Frontiers in Marine Science, DOI: 10.3389/fmars.2022.1012756

Felden, Janine; Lars Möller; Uwe Schindler et al. (2023): PANGAEA – Data Publisher for Earth & Environmental Science. Scientific Data, DOI: 10.1038/s41597-023-02269-x

Kastrisios, Christos; Noel Dyer; Tamer Nada; Stilianos Contarinis; Jose Cordero (2023): Increasing Efficiency of Nautical Chart Production and Accessibility to Marine Environment Data through an Open-Science Compilation Workflow. ISPRS International Journal of Geo-Information, DOI: 10.3390/ijgi12030116

Mayer, Larry; Martin Jakobsson; Graham Allen; Boris Dorschel et al. (2018): The Nippon Foundation – GEBCO Seabed 2030 Project: The Quest to See the World's Oceans Completely Mapped by 2030. Geosciences, DOI: 10.3390/geosciences8020063

Mellone, Gennaro; Ciro Giuseppe de Vita; Enrico Zambianchi et al. (2022): Democratizing the computational environmental marine data science: Using the High-Performance Cloud-Native Computing for inert transport and diffusion Lagrangian modelling. 2022 IEEE International Workshop on Metrology for the Sea. DOI: 10.1109/MetroSea55331.2022.9950862

Schwartz-Belkin, Inbar; Michelle E. Portman (2023): A review of geospatial technologies for improving Marine Spatial Planning: Challenges and opportunities. Ocean & Coastal Management, DOI: 10.1016/j.ocecoaman.2022.106280

Tanhua, Toste; Sylvie Pouliquen; Jessica Hausman et al. (2019): Ocean FAIR Data Services. Frontiers in Marine Science, DOI: 10.3389/fmars.2019.00440

Wehner, Daniel; Sergius Dell; Adrian J. Neumann; Jann Wendt (2022): MARISPACE-X: Digitalizing the ocean – the future of maritime data in the European federated data infrastructure GAIA-X. The International Hydrographic Review, DOI: 10.58440/ihr-28-a10

Wehner, Daniel; Torsten Frey (2022): Offshore Unexploded Ordnance Detection and Data Quality Control – A Guideline. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, DOI: 10.1109/JSTARS.2022.3200144

Wehner, Daniel; Torsten Frey; M. Kampmeier (n.d.): Researchers Involve Stakeholders to Define Unexploded Ordnance Survey Requirements

Wölfl, Anne-Cathrin; Helen Snaith; Sam Amirebrahimi et al. (2019): Seafloor Mapping – The Challenge of a Truly Global Ocean Bathymetry. Frontiers in Marine Science, DOI: 10.3389/fmars.2019.00283

Ziolkowski, Tobias; Agnes Koschmider; Peer Kröger; Colin Devey (2022): Outlier quantification for multibeam data. Informatik Spektrum, DOI: 10.1007/s00287-022-01469-w