

ODANext

Oceanographic data acquisition: the next age

Project B2/202/P2/ODANext

RÉSUMÉ

Introduction et Contexte

Le projet ODANext, intitulé "Oceanographic Data Acquisition: The Next Age" (Acquisition de Données Océanographiques : La Nouvelle Ère), a été lancé pour répondre aux besoins évolutifs de la recherche océanographique avec la mise en service du nouveau navire de recherche (RV) Belgica. Le RV Belgica représente une avancée significative dans les capacités de recherche maritime de la Belgique, étant équipé de capteurs de pointe et d'une multitude d'équipements scientifiques collectant une grande quantité de données.

Cependant, les systèmes hérités utilisés par son prédécesseur étaient insuffisants pour gérer le volume et la complexité des données générées par le nouveau navire. Ainsi, le projet s'est concentré sur la refonte des systèmes d'acquisition, de gestion et de traitement des données pour s'assurer que les capacités du navire soient pleinement exploitées, soutenant ainsi la recherche de pointe et la conformité avec les directives européennes telles qu'INSPIRE et Open Data.

Objectifs et Méthodologie

L'objectif principal du projet était de créer un flux de données automatisé et efficace capable de traiter les données du RV Belgica, garantissant que les données soient standardisées, accessibles, diffusées et de haute qualité. Cela incluait :

1. Établir un flux continu de données des capteurs jusqu'aux utilisateurs.
2. Enrichir avec des métadonnées pour décrire précisément les capteurs utilisés et les données générées.
3. Optimiser le stockage des données et assurer des sauvegardes sécurisées.
4. Publier les données dans les répertoires internationaux pertinents.

La méthodologie adoptée était dynamique et flexible, utilisant la technique de priorisation MoSCoW et les principes Agile pour s'adapter aux défis posés par la technologie du nouveau navire. Le projet a mis l'accent sur le respect des principes FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable) pour maximiser l'utilité et la longévité des données collectées.

Développements Clés et Résultats

1. Acquisition et Gestion des Données à Bord

- Les systèmes d'acquisition de données du navire ont été mis à jour, remplaçant le matériel obsolète par des alternatives capables de gérer le volume important de données générées par les capteurs du navire. Nous avons également travaillé sur l'optimisation des fréquences d'acquisition des données pour ne stocker que les données potentiellement intéressantes.

- Une interface visuelle a été installée à bord en utilisant le logiciel open-source Grafana. Elle offre un moyen simple et fiable de vérifier les données des capteurs en temps réel, permettant ainsi une détection précoce des dysfonctionnements des capteurs.

2. Transfert et Stockage des données

- Le projet a mis en place un système efficace de transfert de données entre le navire et le rivage en utilisant la technologie V-SAT, fiable et peu coûteuse.

- Une nouvelle architecture de base de données a été développée en utilisant PostgreSQL, optimisée avec l'extension TimescaleDB pour gérer les données chronologiques, garantissant une performance robuste. La structure de la base de données respecte une norme reconnue internationalement pour les données des capteurs, à savoir la norme SensorThings de l'Open Geospatial Consortium.

- Les données de l'ancien RV Belgica ont été formatées et transférées dans la nouvelle base de données pour assurer un accès centralisé aux données des deux navires.

3. Optimisation et Contrôle de Qualité

- Diverses optimisations ont été effectuées tout au long de la chaîne d'acquisition et de traitement des données pour s'assurer que le système puisse gérer efficacement l'augmentation du volume de données. Nous avons travaillé sur les fréquences d'acquisition pour limiter la croissance de la base de données, le matériel informatique à bord et à terre, la structure de la base de données avec les index ad hoc pour une récupération rapide des données, etc.

- Une procédure de contrôle de qualité automatisée a été développée en Python et publiée en ligne sur GitHub, permettant l'identification des données erronées. La procédure effectue plusieurs vérifications, notamment les limites de seuil, la détection des gradients et des pics, et la validation des géolocalisations. Il s'agit d'une première vérification qui permet une identification précoce des données de mauvaise qualité et qui doit être complétée par une validation visuelle. Cette deuxième validation n'est pas appliquée systématiquement en raison de la quantité de données générées par le navire de recherche.

4. Diffusion et Valorisation

- Le projet a connecté avec succès le nouveau système de données à des plateformes existantes comme le site web du RV Belgica, GOSUD, SeaDataNet et INSPIRE, garantissant que les données des anciens et nouveaux navires soient accessibles aux chercheurs et au public. Avec cette réalisation, nous avons démontré l'importance de publier les données en ligne en utilisant des API standardisées (Interface de Programmation d'Application). Les applications clientes pouvaient facilement se connecter au point d'accès de l'API et obtenir les données dont elles ont besoin, sans nécessiter d'intervention humaine.

- Des efforts ont été faits pour standardiser les données en utilisant notamment des vocabulaires contrôlés, améliorant l'interopérabilité pour l'échange de données entre machines et la conformité aux normes internationales.

Impact et Travaux Futurs

Le projet ODANext a considérablement amélioré les capacités de gestion des données associées au RV Belgica, le positionnant comme une plateforme de premier plan pour la recherche marine. Les systèmes développés au cours du projet permettront une diffusion plus efficace et plus large des données scientifiques, soutenant les efforts de recherche et les prises de décisions politiques. Le projet a également souligné l'importance de l'optimisation continue et le potentiel de futures améliorations pour renforcer encore les systèmes de gestion des données du navire.

Mots-Clés

Principes FAIR, Acquisition de données, Diffusion des données, Interopérabilité, Données marines, API, Navire de recherche, Qualité de l'eau.