

# ODANext

## Oceanografische Gegevensverzameling: Het Volgende Tijdperk

Project B2/202/P2/ODANext

### SAMENVATTING

#### Inleiding en Achtergrond

Het ODANext-project, getiteld " Oceanographic Data Acquisition: The Next Age," werd geïnitieerd om in te spelen op de veranderende behoeften in oceanografisch onderzoek met de ingebruikname van het nieuwe Onderzoeksschip (RV) Belgica. De RV Belgica vertegenwoordigt een aanzienlijke upgrade in de mariene onderzoeksfaciliteiten van België, uitgerust met geavanceerde sensoren en wetenschappelijke apparatuur, ontworpen om enorme hoeveelheden gegevens uit verschillende mariene omgevingen te verzamelen.

De verouderde systemen van zijn voorganger bleken ontoereikend om het volume en de complexiteit van de door het nieuwe schip gegenereerde gegevens te verwerken. Om de capaciteiten van het schip optimaal te benutten en om te voldoen aan Europese richtlijnen zoals INSPIRE, en het ondersteunen van open data, richtte het project zich op het herontwerpen van de systemen voor gegevensverzameling, -beheer en -verwerking.

#### Doelstellingen en Methodologie

Het primaire doel van het project was het creëren van een geautomatiseerde, efficiënte gegevensworkflow van de gegevens van de RV Belgica, waarbij werd gewaarborgd dat de data gestandaardiseerd, toegankelijk, breed verspreid en van hoge kwaliteit zijn. Dit omvatte:

1. Het opzetten van een continue gegevensstroom van sensor naar gebruiker.
2. Het verrijken van metadata waar nodig.
3. Het optimaliseren van gegevensopslag en het zorgen voor veilige back-ups.
4. Het integreren van de gegevens in relevante open science repositories.

De gehanteerde, flexibele en dynamische methodologie maakte gebruik van de MoSCoW-prioriteringstechniek en Agile-principes om in te spelen op de uitdagingen die de nieuwe technologie van het schip met zich meebracht. Het project legde de nadruk op naleving van de FAIR-principes (Findable,

Accessible, Interoperable, and Reusable) om de bruikbaarheid en duurzaamheid van de verzamelde gegevens te maximaliseren.

## **Belangrijke Ontwikkelingen en Resultaten**

### **1. Gegevensverzameling en Beheer aan Boord:**

- De acquisitiesystemen van het schip werden bijgewerkt, waarbij verouderde hardware werd vervangen door moderne alternatieven die de grote hoeveelheid door de sensoren van het schip gegenereerde gegevens aankonden. Verder werden de data-acquisitiefrequenties geoptimaliseerd om tijdsvertragingen te minimaliseren en enkel data met kennispotentieel op te slaan. We werkten ook aan de optimalisatie van de frequenties van gegevensverzameling om mogelijke vertragingen in de gegevens te verminderen en .

- Een visuele interface is aan boord geïmplementeerd met behulp van de open-source software Grafana. Het biedt een eenvoudige en betrouwbare manier om sensorgegevens in real-time te controleren, waardoor vroegtijdige detectie van sensorstoringen mogelijk is.

### **2. Gegevensoverdracht en -opslag:**

- Het project implementeerde efficiënte systemen voor gegevensoverdracht van het schip naar het vasteland, waarbij gebruik werd gemaakt van V-SAT-technologie, die betrouwbaar en kosteneffectief bleek te zijn.

- Een nieuwe database-architectuur werd ontwikkeld met behulp van PostgreSQL, geoptimaliseerd met TimescaleDB voor het beheren van tijdreeksgegevens, wat resulteerde in robuuste prestaties naarmate de database groeide. De databasestructuur volgt een internationaal erkende standaard ontwikkeld voor sensorgegevens, namelijk de Open Geospatial Consortium SensorThings-standaard.

- Gegevens van de vorige RV Belgica zijn geformatteerd en overgedragen naar de nieuwe database om naadloze toegang te garanderen tot de gegevens van beide RV Belgica-schepen.

### **3. Optimalisatie en Kwaliteitscontrole:**

- Verschillende optimalisaties werden doorgevoerd in de gehele data acquisitie- en verwerkingspipeline om ervoor te zorgen dat het systeem het toenemende gegevensvolume zonder achterstanden kon verwerken. We werkten aan de verzamelingsfrequenties om de groeisnelheid van de database te beperken, de hardware aan boord en aan wal, de databasestructuur met ad-hoc indexen voor snelle gegevensherstel, enz.

- Een geautomatiseerde kwaliteitscontroleprocedure is ontwikkeld in Python en online gepubliceerd op GitHub, waarmee gegevensanomalieën kunnen worden geïdentificeerd en gecorrigeerd. De procedure voert meerdere controles uit, waaronder drempelwaardelimiten, markering van gradiënten en pieken, en validatie van geografische locaties. Het is een eerste controle die vroegtijdige identificatie van slechte gegevenskwaliteit mogelijk maakt en moet worden aangevuld met een visuele validatie. Dit is uitdagend vanwege de hoeveelheid door het onderzoeksvaartuig gegenereerde gegevens.

#### **4. Verspreiding en Waardering:**

- Het project integreerde met succes de nieuwe data-acquisitiesystemen in bestaande platforms zoals de RV Belgica-website, GOSUD, SeaDataNet en INSPIRE, waardoor gegevens van zowel de oude als de nieuwe schepen toegankelijk waren voor onderzoekers en het publiek. Met deze prestatie demonstreerden we het belang van het online beschikbaar stellen van de gegevens via gestandaardiseerde API's (Application Programming Interface). Klanttoepassingen kunnen eenvoudig verbinding maken met het API-eindpunt en gegevens opvragen, zonder menselijke tussenkomst.

- Er zijn inspanningen geleverd om de gegevens te standaardiseren met behulp van gekoppelde gegevensconcepten en gecontroleerde woordenlijsten, waardoor de interoperabiliteit van gegevens voor machine-to-machine gegevensuitwisseling werd verbeterd en aan internationale normen werd voldaan.

#### **Impact en Toekomstig Werk**

Het ODANext-project heeft de capaciteiten voor gegevensbeheer die verband houden met de RV Belgica aanzienlijk verbeterd, waardoor het zich positioneert als een toonaangevend platform voor marien onderzoek. De systemen en workflows die tijdens het project zijn ontwikkeld, zullen een efficiëntere en bredere verspreiding van wetenschappelijke gegevens mogelijk maken, wat wereldwijde onderzoeksinspanningen en beleidsvorming zal ondersteunen. Het project benadrukte ook het belang van voortdurende optimalisatie en het potentieel voor toekomstige upgrades om de data-acquisitiesystemen van het schip verder te verbeteren.

#### **Trefwoorden**

FAIR-principes, Gegevensverzameling, Data-acquisitie, Gegevensverspreiding, Interoperabiliteit, Mariene Gegevens, API, Onderzoeksschip, Waterkwaliteit