

Eindrapportage pilot – DCC voor Praktijkgericht onderzoek

Pilotgegevens	
Titel van pilot	SMART Sensordata Infrastructuur
Hogeschool	HAS green academy
Contactpersoon	Michiel Jellema
Emailadres	jem@has.nl
Datum	5-12-2022

1 – Beschrijving van pilot

- *Geef een korte beschrijving van de pilot*
- *Wat was het doel van de pilot?*
- *Welk vraagstuk of probleem heb je aangepakt en hoe (doorlopen stappen/proces)?*

Minimaal 3 lectoraten op HAS green academy maken gebruik van sensordata. Zij lopen in hun onderzoek regelmatig tegen het probleem aan dat de data-infrastructuur onvoldoende mogelijkheden biedt om de real-time sensordata uit de verschillende fieldlabs efficiënt en effectief op te slaan en in onderlinge samenhang breed toepasbaar te krijgen voor het lectoraatsonderzoek. Hierdoor wordt sensordata niet optimaal ingezet en gaat veel tijd verloren met het uitrollen en operationaliseren van sensoren. Deze tijd kan niet besteed worden aan data analyse en visualisatie. Omdat het aantal fieldlabs met sensoren voor nieuwe onderzoeksprojecten in de toekomst verder gaat toenemen is het noodzakelijk het geschetste probleem nu op te gaan lossen.

Het ultieme doel is het opstellen van een sensordata infrastructuur om de via fieldlabs verkregen ruwe sensor-data gestandaardiseerd op basis van een eenduidig metadata-model en volledig automatisch via een datadriven-workflow, die alle fasen uit het datascience-lifecycle omvat, direct toepasbaar te krijgen voor de verschillende onderzoeksvragen.

Om dit doel te bereiken zijn de volgende activiteiten uitgevoerd:

- Interviews met experts van verschillende hogescholen en SURF.
- Ontwerpen van sensordata infrastructuur en metadatamodel op basis van opgehaalde kennis en ervaringen uit interviews
- Valideren van opstelde sensordata infrastructuur en metadatamodel met experts van SURF
- Proof-of-Concept in de vorm van een fieldlab om ontworpen sensordata infrastructuur in de praktijk te valideren

2 – Resultaat & impact

- *Wat zijn de resultaten? Welk resultaat heb je bereikt? Wat is er klaar?*
- *Wat heb je geleerd?*
- *Hoe ga je hiermee verder?*
- *Wat is de impact op (het faciliteren) van praktijkgericht onderzoek?*

In het kader van het project zijn de volgende resultaten opgeleverd:

- Ontwerp en praktische beschrijving van algemeen toepasbare datadriven-workflow voor sensordata (*Ontwerp en praktische beschrijving van sensordata infrastructuur en metadata model_v1_151222.pdf*)
- Ontwerp en praktische beschrijving van metadata-model van sensor-data, gericht op datadefinitie en datakwaliteit (*Ontwerp en praktische beschrijving van sensordata infrastructuur en metadata model_v1_151222.pdf*)
- Handleiding & FAQ voor ontwerpen en implementatie datascience-workflow en metadata-model in verschillende SURF-diensten (*Handleiding sensordata_v3_151222.pdf*)
- Proof-of-concept van voorgestelde sensordata infrastructuur om de werking in de praktijk te toetsen

Op basis van het uitgevoerde onderzoek en de proof-of-concept, is het volgende geleerd en kunnen de volgende conclusies getrokken worden:

- Het is mogelijk om met open source producten en open standaarden een sensordata infrastructuur op te zetten waarmee het gehele proces van data inwinning met sensoren tot permanente database opslag en gebruik in visualisaties en analyses geïmplementeerd is
- Het netwerk van The Things Network (TTN) wordt veel gebruikt voor het opzetten van sensordata infrastructures in een HBO/academische setting
- The Things Network biedt verschillende open standaarden om de sensordata met ETL-tools op te halen
- Voor permanente opslag kan gekozen worden voor een relationele SQL-database maar er zijn ook andere NoSQL alternatieven speciaal ontwikkeld voor de opslag van timeseries (bijv. InfluxDB).
- Metadata is essentieel voor delen en verder gebruik van de ingewonnen sensordata. Bestaande metadatamodelen zijn vereenvoudigd zodat alleen de essentiële metadata opgeslagen wordt. Het metadatamodel is generiek opgezet zodat nieuwe sensoren toegevoegd kunnen worden zonder dat het model aangepast moet worden en het eenduidig toepasbaar is.
- Nieuwe sensoren kunnen eenvoudig opgenomen worden in de sensordata infrastructuur door middel van registratie in het netwerk en het toevoegen aan een gestandaardiseerde ETL-procedure.

Met de resultaten van dit onderzoek wordt het in praktijkgericht onderzoek eenvoudiger en eenduidiger om een sensordata infrastructuur op te zetten en real-time sensordata uit meerdere fieldlabs gecombineerd en in samenhang te gebruiken. De betrokken onderzoekers zullen in de toekomst minder tijd en aandacht nodig hebben om een goede data-infrastructuur rond sensoren te ontwikkelen.

De toegevoegde waarde van het metadatamodel is dat onderzoekers direct inzicht hebben in de sensor metadata. Goede metadata garandeert een bredere inzetbaarheid en efficiënter en effectiever gebruik van de sensordata. Zonder metadata is data validatie en het inschatten van data kwaliteit niet mogelijk. Ook is moeilijk na te gaan of sensordata bruikbaar is voor het beoogde doel.

Met de opgedane kennis zal de infrastructuur die in het kader van de proof-of-concept opgezet is, verder uitgebouwd worden. Er zullen meerdere sensoren aangekoppeld worden en er zal een dashboard gebouwd worden om real-time inzicht te krijgen in de ingewonnen data en geavanceerde data analyses en visualisaties te kunnen uitvoeren.

Ook zal vervolgonderzoek gedaan worden naar de mogelijkheid om een NoSQL timeseries database te gaan gebruiken voor permanente opslag van de ingewonnen sensordata. Het voordeel van het werken met een NoSQL timeseries database is dat deze meer flexibiliteit biedt met betrekking tot de metadata die opgeslagen kan worden en dat deze een betere integratie biedt met tools voor timeseries analyse.

3 – Diensten

- *Welke SURF-diensten waren onderdeel van de pilot?*
- *Hoe zijn deze diensten ingezet of gebruikt binnen de pilot?*

In de loop van het project werd duidelijk dat de diensten die voor de implementatie van de voorgestelde sensordata infrastructuur nodig zijn, nog niet door SURF geleverd worden. SURF Research Drive en SURF Research Cloud bieden bijvoorbeeld geen relationele database. Daarom is in overleg met de experts van SURF besloten gebruik te maken van een bestaand open netwerk (van The Things Network) en bestaande clouddiensten (van The Things Network en Azure). Om de kosten zo laag mogelijk te houden en vendor-lockin te voorkomen is er waar mogelijk gekozen voor licentievrije, open source producten en open standaarden. Alleen aan de Azure diensten zijn kosten verbonden voor data processing (voor de ETL-tooling) en data opslag (voor de PostgreSQL database).

4 – Betrokkenen en inzet

- *Wie waren betrokken bij de pilot (rollen/functionies)?*
- *Welk lectoraat(en) of kenniscentra waren betrokken bij de pilot?*
- *Welke inzet en expertise is vanuit SURF geleverd?*
- *Urenverantwoording graag aanleveren in de verstrekte Excelfile*

Het project is uitgevoerd door de volgende mensen van de HAS green academy:

- Michiel Jellema (contractmanager)
- Mark Terlien (projectleider)
- Marien de Bakker (projectmedewerker)
- Max de Visser (projectmedewerker)
- Jesper Giesbers (projectmedewerker)
- Koen Verschuren (technisch expert sensordata infrastructuur)
- 4^e jaars studenten opleiding Applied Geo-Information Science

De volgende lectoraten van de HAS green academy zijn betrokken bij de pilot:

- Lectoraat Impactvolle ruimtelijke inzichten i.o.
- Lectoraat Klimaatrobuuste landschappen
- Lectoraat Innovatieve biomonitoring
- Lectoraat Precision Livestock Farming

De volgende mensen van verschillende hogescholen en expertise centra zijn geïnterviewd:

- Leo Remeijn (Hogeschool Rotterdam)
- Ton Ammerlaan (Hogeschool Arnhem-Nijmegen)
- Roderick Peters (Hogeschool Arnhem-Nijmegen, Lentse Luchten)
- Andries van Dongen (Avans Hogeschool)
- Huibert-Jan Lekkerkerk (NHL Stenden)
- Jeroen van der Pluijm (Hanze Hogeschool Groningen)
- Michel Grothe (Geonovum)

De volgende experts van SURF zijn bij het project betrokken:

- David Salek (sensor expert)
- Robert Griffioen (sensor expert)
- Sophie de Koning (relatiemanager)

5 - Lessons learned

- *Zijn er lessons learned?*
- *Zijn er aandachtspunten of verbeterpunten?*

De lessons learned zijn beschreven in hoofdstuk 2 (Resultaat & impact).

Aandachtspunten voor sensorprojecten in het algemeen zijn:

- Zorg dat er voldoende budget is voor de aanschaf van sensoren om een fieldlab in te richten. Een eenvoudige LoRaWAN sensor om temperatuur te meten inclusief batterij en antenne kost al snel 50 euro. Prijzen voor gateways beginnen bij 250 euro.
- Let op dat je sensoren met de juiste frequentie aanschaft (868MHz voor Europa)
- Houd er rekening mee dat The Things Network niet overal in Nederland goede dekking heeft. Als er geen dekking is, zal er naast een sensor ook een gateway aangeschaft moeten worden.
- Naast TTN zijn er alternatieve gesloten commerciële netwerken die wel 100% dekking claimen in Nederland (bijv. het KPN LoRa netwerk).
- Maak zo veel mogelijk gebruik van open standaarden om vendor lockin te voorkomen en zorg dat je eigenaar van je data blijft.
- Maak zo veel mogelijk gebruik van bestaande software en clouddiensten. De kosten hiervoor wegen in het algemeen niet op tegen de kosten voor het zelf ontwikkelen van software of het zelf beheren van servers
- Besteed voldoende aandacht aan permanente opslag van ingewonnen data, analyse en visualisatie zodat gebruik van data in de toekomst ook geborgd is.

6 - Kennisdeling en disseminatie

- *Hoe zijn de resultaten van het pilot beschikbaar gesteld binnen de eigen hogeschool?*
- *In welke vorm is de opgedane kennis beschikbaar gesteld aan andere hogescholen?*

De proof-of-concept is opgezet door 4^e-jaars studenten van de opleiding Applied Geo-information Science van de HAS green academy. De proof-of-concept zal verder uitgewerkt worden in het kader van het project Smart Climate Adaptive Environment van de HAS green academy. Dit project heeft tot doel het klimaatrendement van groene daken te bepalen. Hiervoor worden een aantal groene daken verspreid over Nederland van sensoren voorzien. Op basis van de ingewonnen data zal het klimaatrendement van de groene daken ten opzichte van traditionele naburige zwarte daken gekwantificeerd worden. De geplande activiteiten (installeren sensoren, ontwikkelen dashboards en visualisaties, kwantificeren klimaatimpact) zullen uitgevoerd worden door studenten van

verschillende opleidingen in het kader van hun specialisatie of afstuderen. Hierbij wordt de kennis gebruikt die in deze pilot verkregen is. Het Digital Innovation Lab van de HAS green academy is verantwoordelijk voor het beheer van de sensordata infrastructuur.

De infrastructuur is generiek opgezet zodat deze ook door andere lectoraten van de HAS gebruikt kan worden, bijvoorbeeld voor het monitoren van hydro-fysische parameters in stroomgebieden binnen het lectoraat Klimaatrobuuste landschappen.

Om de opgedane kennis te delen en beschikbaar te stellen aan andere hogescholen zodat deze ook in andere praktijkgerichte onderzoeken gebruikt kan worden, wordt er op 26 januari 2023 een afsluitende workshop gegeven waarbij o.a. de deelnemers van het DCC-PO en andere hogescholen en de lectoren en onderzoekers van de HAS uitgenodigd zijn.

De documenten met de beschrijving van de sensordata infrastructuur en het metadatamodel en de handleiding met de FAQ voor het inrichten van de beschreven sensordata infrastructuur zullen via DCC-PO van SURF beschikbaar gesteld worden aan andere hogescholen en geïnteresseerden.