

## D3.5 Standaardprotocol voor het behandelen van big data

*Interview met Christelle Al Haddad*

Het conceptuele kader van het i-DREAMS-platform integreert aspecten van monitoring (zoals context, bestuurder, voertuig, taakcomplexiteit en handelingsvermogen) om een Veiligheids-Tolerantie-Zone (VTZ) voor het rijden te ontwikkelen. Interventies in het voertuig en interventies na de rit helpen de VTZ te handhaven en geven feedback aan de bestuurder. Dit conceptuele kader zal worden getest in simulatorstudies en testen in het verkeer in België, Duitsland, Griekenland, Portugal en het Verenigd Koninkrijk (VK) met ongeveer 600 deelnemers die auto-, bus-, vrachtwagen- en spoorwegbestuurders vertegenwoordigen.

Tijdens de experimenten zullen grote hoeveelheden gegevens worden gegenereerd, afkomstig van de verschillende instrumenten voor gegevensverzameling en van verschillende vervoerswijzen en landen. Het doel van dit rapport is dan ook de nodige protocollen voor de behandeling van deze "Big Data" aan te reiken. Meer bepaald zal dit rapport:

- een methodologie bieden voor de behandeling van big data, op basis van leerervaringen uit eerdere studies/projecten; met name naturalistische rijstudies (NRS) in Europa
- voorzien in standaardprotocollen voor het behandelen van big data, waarbij i-DREAMS-experimenten worden geïnstrueerd over de procedures die moeten worden gevolgd om de verzamelde gegevens zo goed mogelijk te behandelen, met inachtneming van de nodige regelgeving en ethische overwegingen.

De standaardprotocollen moeten gedurende het hele project voortdurend worden gecontroleerd en in overeenstemming zijn met het [Gegevens-Beheers-Plan](#) (GBP)<sup>1</sup> en het Comité voor Gegevens- en Kennisbeheer<sup>2</sup>. De projectpartners in de verschillende landen zijn verantwoordelijk voor hun eigen gegevensverzameling en zijn verplicht de juiste standaarden te volgen, in overleg met hun nationale en lokale autoriteiten. Over het geheel genomen moet rapport 3.5 een levend document zijn, dat waar nodig in de nodige stappen wordt bijgewerkt, en dat als richtsnoer dient voor de beste aanpak van de big data die in de loop van het project worden gegenereerd.

**Christelle, u bent de auteur van dit rapport. U begon uw werk met het beoordelen van eerdere naturalistische rijstudies (NRS) in Europa? Welke projecten heeft u precies bekeken en zijn ze vergelijkbaar met wat u in i-DREAMS doet?**

CHRISTELLE AL HADDAD: *"Wij hebben 10 Europese naturalistische rijstudies<sup>3</sup> bekeken en ons gericht op de wijze waarop zij de*

<sup>1</sup> De tweede en derde updates van de GBP zijn vertrouwelijk.

<sup>2</sup> Het Comité voor Gegevens- en Kennisbeheer is een toezichthoudend orgaan dat zich in het bijzonder bezighoudt met het beheer en de bescherming van gegevensstromen.

<sup>3</sup> De volgende naturalistische rijstudies werden geëvalueerd: [AOS](#) (2007-2009), [SeMiFOT](#) (2008-2009), [euroFOT](#) (2008-2011), [TeleFOT](#) (2008-2012), [INTERACTION](#) (2008-2021), [2BeSAFE](#) (2009-2011), [PROLOGUE](#) (2009-2011), [UDRIVE](#) (2012-2016), [Track & Know](#) (2019-2022).



*gegevensverzameling, de gegevensvoorbereiding, de gegevensopslag en de juridische en ethische overwegingen hebben aangepakt. Het is belangrijk te vermelden dat de meeste van deze studies gericht waren op personenauto's en vrachtwagens en in sommige gevallen op gemotoriseerde tweewielers (motorfietsen en scooters). Hoewel de onderzochte projecten niet altijd hetzelfde toepassingsgebied hebben als i-DREAMS (auto's, vrachtwagens, bussen en spoorvoertuigen), spreekt het vanzelf dat het van cruciaal belang is rekening te houden met de lessen die uit deze projecten kunnen worden getrokken.”*

**In het rapport wijdt u een heel hoofdstuk aan methoden voor het beheer van Big Data. Over welke methoden heeft u het precies?**

*CHRISTELLE AL HADDAD: “Het omvat alle methoden voor gegevensverzameling, -voorbereiding (inclusief gegevensverwerking), opslag, toegang en uitwisseling in naturalistische rijstudies. Veel van de methoden die we in het rapport beschrijven zijn een weergave van gezond verstand, maar natuurlijk zijn ze meer dan dat. Ze zijn gebaseerd op wat we hebben geleerd van eerdere projecten en de doelen die we met ons project willen bereiken. We hebben methoden ontwikkeld om de verzameling, (voor)verwerking, opslag, toegang en uitwisseling van gegevens te stroomlijnen, waarbij we uiteraard rekening hebben gehouden met de verschillende ethische en juridische overwegingen.”*

**Kunt u ons enkele voorbeelden geven van de methoden die u in i-DREAMS zult toepassen, bijvoorbeeld met betrekking tot het verzamelen van gegevens?**

*CHRISTELLE AL HADDAD: “Natuurlijk, deze methoden hebben te maken met de manier waarop wij gegevens verzamelen en welke regels wij moeten volgen. Wij zullen bijvoorbeeld in alle experimenten hetzelfde dataverzamelingssysteem gebruiken om problemen met incompatibiliteit van gegevens te beperken. Aangezien de meeste van onze gegevens, verzameld met sensoren, kwantitatief zijn en worden aangevuld met extra lagen van gegevens waaronder enquête en GPS-gegevens, wordt het een uitdaging om deze op de juiste manier te verwerken. We gebruiken specifieke technieken om persoonlijke informatie correct te beveiligen alvorens ze aan het experiment te koppelen. En natuurlijk beginnen we pas met het verzamelen van gegevens nadat we de respectieve ethische commissies en functionarissen voor gegevensbescherming in de verschillende locaties waar gegevens worden verzameld, hebben geraadpleegd en hun goedkeuring hebben gekregen. En nadat we de ondertekende toestemming van de deelnemers hebben ontvangen om aan het onderzoek deel te nemen, zodat we hun gegevens voor onderzoeksdoeleinden kunnen verwerken. Een ander belangrijk aspect dat moet worden vermeld, is de overeenkomst inzake gegevensuitwisseling die werd opgesteld voor de uitwisseling van gegevens tussen de partners.”*



### Dit roept bij mij veel nieuwe vragen op. U verwijst naar een specifiek dataverzamelingssysteem ... Welk systeem is dat?

CHRISTELLE AL HADDAD: “Voor het verzamelen van gegevens moeten wij een onderscheid maken tussen simulatorgegevens en gegevens van testen in het verkeer. Voor deze laatste gebruiken wij ons zogenaamde i-DREAMS-systeem. Dit systeem bestaat uit verschillende apparaten die door Cardiold (een van onze technische partners in i-DREAMS) op de markt worden gebracht en worden gebruikt om gegevens te verzamelen en vervolgens in real time interventies in het voertuig uit te voeren. De verschillende apparaten zijn verbonden met een gateway die informatie van andere componenten verzamelt en centraliseert en zorgt voor de gegevensverbinding en -overdracht.



Figuur 1: i-DREAMS Monitoring technologieën

Voor de interventies na de rit gebruiken we de technologie van OSeven PC (een andere technische partner in i-DREAMS). Dit is een geavanceerde, op Android en iOS gebaseerde smartphone-applicatie die ook het rijgedrag van personen controleert aan de hand van verschillende parameters. De gegevens van de Oseven PC-app worden samengevoegd met de Cardiold-technologie voor een nauwkeurigere voorspelling van het rijgedrag. Verder is er nog andere app-technologie afkomstig van UHasselt, de projectcoördinator, die de bestuurder feedback geeft over zijn persoonlijke prestaties en uitdagingen en de bestuurder motiveert via gamificatiefuncties.”

### Welke praktische lessen hebt u geleerd van de NRS-studies die u hebt geëvalueerd?

CHRISTELLE AL HADDAD: “De belangrijkste lessen met betrekking tot gegevensverzameling zijn de volgende: het is raadzaam om het aantal voertuigmodellen dat wij met ons gegevensverzamelingssysteem uitrusten tot een minimum te beperken om de complexiteit van installatie en de-installatie te verminderen. Verder hebben we geleerd hoe belangrijk het is de verantwoordelijkheden op het gebied van codering, verwerking en analyse te centraliseren om consistentie in de dataset te creëren. Met betrekking tot gegevensopslag hebben we heel wat praktische zaken geleerd, zoals het belang van voorbereiding en verrijking van de gegevens en het opslaan van gegevens in open formaten om de toegankelijkheid te vergroten. Natuurlijk is er ook het belang van een systematisch back-upsysteem en een duidelijke definitie van gegevens zodat iedereen dezelfde taal spreekt. De overdracht van gegevens van lokale naar centrale servers moet elektronisch en zonder manuele interventie gebeuren. Bestanden die wel manuele interventie vereisen (bv. papieren vragenlijsten) moeten worden opgeslagen nadat ze zijn omgezet naar de elektronische versie. En

*natuurlijk is het ook belangrijk om bij het nadenken over de gegevensarchitectuur rekening te houden met de verwerkingssnelheid en de koppelingen met databanken.*

*Wat ten slotte de juridische en ethische kwesties betreft, moet rekening worden gehouden met de AVG, naast alle nationale en lokale voorschriften, die van land tot land kunnen verschillen. Daarom wordt de deelnemers gevraagd een formulier 'geïnformeerde toestemming' te ondertekenen voordat zij aan het experiment beginnen en zorgen wij ervoor dat de nodige procedures zijn ingevoerd om de persoonsgegevens te beschermen."*

**Ik kan me voorstellen dat het bij het uitwerken van alle noodzakelijke procedures belangrijk is dat alle betrokken partners een duidelijk beeld hebben van hun eigen verantwoordelijkheden in dit grote geheel. Kunt u kort ingaan op deze verantwoordelijkheden?**

*CHRISTELLE AL HADDAD: "Het is belangrijk om een onderscheid te maken tussen technologiepartners en gegevensverzamelingspartners, maar ook tussen simulatorexperimenten en experimenten in het veld.*

*CardioID en Oseven PC zijn onze technologiepartners. DriveSimSolutions (DSS) is onze simulatorpartner. DSS integreert technologieën en levert simulatoren die gegevens loggen, die lokaal worden opgeslagen op de simulator-pc. Partners die een andere rijnsimulator gebruiken (NTUA en LOUGH) moeten een soortgelijke aanpak volgen.*

*De partners die testen in het verkeer organiseren, zijn UHasselt, TUM, NTUA, BARRA en LOUGH. Zij zijn verantwoordelijk voor de logistiek bij het opzetten en opvolgen van de experimenten op hun respectieve locaties. De gegevensverwerkers (UHasselt, LOUGH, TUM, TUD en NTUA) krijgen toegang tot de verzamelde gegevens om de hypothesen afgeleid van de onderzoeksvragen te analyseren en te testen. De UHasselt verwerkt de gegevens ook via het post-trip interventiekader door op basis van de verzamelde gegevens scores en interventies te genereren. Ik kan natuurlijk meer diepgaande informatie geven over al deze verantwoordelijkheden, maar dit geeft een beetje een idee van de grote lijnen."*

**Als al die gegevens verzameld zijn, hoe gaat u dan verder?**

*CHRISTELLE AL HADDAD: "Voor de simulatortesten worden de gegevens automatisch lokaal opgeslagen. Partners zijn vrij om hun eigen opslagsystemen (databases, bestandssystemen) te kiezen voor lokale opslag. Lokaal verwijst naar systemen die niet toegankelijk zijn via een API. De verzamelde gegevens worden gepseudonimiseerd (met een plan om de gegevens na afloop van de experimenten volledig te anonimiseren) zodat de persoonlijke informatie wordt beschermd."*



**Wat betekent API?**

CHRISTELLE AL HADDAD: *“API is het acroniem voor Application Programming Interface, een software-intermediair waarmee twee toepassingen met elkaar kunnen praten. API's zijn mechanismen waarmee twee softwarecomponenten met elkaar kunnen communiceren aan de hand van een reeks definities en protocollen. Het softwaresysteem van het weerbureau bevat bijvoorbeeld dagelijkse weergegevens. De weer-app op uw telefoon "praat" met dit systeem via API's en toont u dagelijkse weerberichten op uw telefoon. Hetzelfde principe wordt toegepast op de verzamelde i-DREAMS-gegevens. Binnen i-DREAMS wordt een backend database of een backoffice component ontwikkeld om ruwe en verwerkte gegevens van andere componenten centraal op te slaan. Partners kunnen vervolgens via een web-API gemakkelijk toegang krijgen tot gegevens uit de backoffice.”*

**Hoe komen lokaal opgeslagen gegevens in die backoffice database terecht?**

CHRISTELLE AL HADDAD: *“Zoals gezegd worden de voertuiggegevens automatisch beschikbaar gesteld aan de partners, via de ontwikkelde gecentraliseerde backoffice. Daarvoor hoeven de partners niets handmatig te uploaden. Aanvullende informatie over de experimenten, die de partners eventueel moeten verstrekken ter ondersteuning van de analyse, moet echter handmatig in een specifiek formaat naar de backoffice server worden geüpload. Dit geldt ook voor de gegevens van de vragenlijst, maar ook voor de bestanden met simulatorgegevens.*

*Andere partners hebben toegang tot de gegevens via de ontwikkelde web-API van de server. Een gezamenlijke overeenkomst tussen de partners over "het delen van persoonsgegevens" regelt hoe persoonsgegevens tussen de consortiumpartners worden gebruikt tot het einde van het project. Bovendien is het de bedoeling om na afloop van het project een geanonimiseerd deel van de gegevens toegankelijk te maken (uiteraard met inachtneming van de AVG en andere regelgeving), zodat andere onderzoekers er hun voordeel mee kunnen doen.”*

**U zei al dat persoonsgegevens te allen tijde worden beschermd. Wat doet u precies om dit te garanderen?**

CHRISTELLE AL HADDAD: *“De gegevens worden eerst gepseudonimiseerd. Dit betekent dat aan elke deelnemer een unieke identificatiecode wordt toegekend. Het document dat de identifieer aan de deelnemer koppelt, wordt niet op een (centrale cloud) server opgeslagen, maar lokaal. Toegang tot dit document is strikt beveiligd en persoonsgegevens worden niet langer bewaard dan noodzakelijk (maximaal 5 jaar na afloop van het project). Na deze 5 jaar worden de gegevens geanonimiseerd. Dit betekent dat de eerder aangewezen unieke identificatiecode die gegevens in de databanken van de partners verbindt met de persoonsgegevens van de deelnemer, wordt vervangen door een willekeurig getal. Er is dus geen verband meer tussen de gegevens en de persoonlijke informatie. Wanneer dit is gebeurd, zijn (een deel van) de gegevens klaar om beschikbaar te worden gesteld in een open-source platform, wat een van de doelstellingen van het project is.”*



Rapport 3.5 maakt deel uit van WP3:  
*Operational design of i-DREAMS*

[Download het rapport hier](#)

**Mijn laatste vraag heeft betrekking op de toegankelijkheid van de gegevens. Welke stappen zitten vervat in uw procedure voor toegang tot gegevens?**

CHRISTELLE AL HADDAD: *“Er zijn een paar dingen waar we speciaal op letten. Eerst definiëren we verschillende gebruikerstypes, elk met verschillende toegangsrechten (bv. superadmin, admin, gebruiker...). De gepseudonimiseerde gegevens worden toegankelijk gemaakt voor de consortiumpartners volgens de bepalingen in gezamenlijke gegevensovereenkomsten. Onze onderzoeksgegevens volgen de FAIR-richtlijnen<sup>4</sup>. Geaggregeerde data van CardioID wordt beschikbaar gesteld via een web API en een geanonimiseerd deel van de data (enkele datasets) wordt aan het einde van het project beschikbaar gesteld en aangeboden aan derden (conform AVG) op een digitaal archief.”*

Wat me vooral bijblijft van ons gesprek is dat er een enorm veel details zijn waar u rekening mee moet houden om alle aspecten die met big data te maken hebben aan te pakken. Ik ben onder de indruk van hoe doordacht u bent geweest bij elke stap. Ik wens u veel succes met alle experimenten!

Edith Donders  
Discom Manager

---

<sup>4</sup> FAIR staat voor Findable (vindbaar), Accessible (toegankelijk), Interoperable (interoperabel) en Reusable (herbruikbaar)

## Onderzoeker in de kijker



**CHRISTELLE  
AL HADDAD**

Afgestudeerd als *Burgerlijk Ingenieur* aan de *Amerikaanse Universiteit van Beiroet (2015)*, en als *Transport ingenieur* aan de *Technische Universiteit van München (2018)*.

Sinds 2019 werkzaam aan de *Technische Universiteit München*,  
*leerstoel Transportation Systems Engineering*.

Gepassioneerd door *cultuur, eten, mensen, talen en muziek (zingen in het bijzonder) ...*

Taken in i-DREAMS: *Leidinggeven aan het operationele werkpakket binnen i-DREAMS (WP5), met name aan de dataverzameling in Duitsland, maar ook de coördinatie van de verschillende taken waar TUM verantwoordelijk voor is (backend, algoritme ontwikkeling, etc.).*

