

Werkprocesbeschrijving

**Werken met technische onderzoeksdata
door vakgroep Waterbouw CiTG TU Delft**

Inhoud

1. INLEIDING.....	1
1.1. Vraagstelling	1
1.2. Leeswijzer	1
2. WERKBESCHRIJVING.....	2
2.1. Onderzoeksvoorstel schrijven	2
2.2. Uitvoeren onderzoek	2
2.3. Rapportage.....	4
2.4. Archivering.....	4
3. AANBEVELINGEN.....	6

1. Inleiding

Het onderzoek Waardevolle Data & Diensten richt zich op het inzichtelijk maken van de functionele eisen en wensen voor de centrale (duurzame) opslag van technische onderzoeksdata. Dit onderzoek wordt uitgevoerd door de TU Delft Library van de Technische Universiteit. Het onderzoek vormt een voorbereiding op de ontwikkeling van een centraal datacentrum voor de opslag van technische onderzoeksdata onder de noemer 3TU.Datacentrum, een samenwerkingsproject tussen de drie Nederlandse technische universiteiten – Technische Universiteit Delft, Technische Universiteit Eindhoven en de Universiteit Twente. De kennis en ervaring die met het onderzoek Waardevolle Data & Diensten worden opgedaan zullen de universiteiten toepassen in dit project. Het is echter tevens van toepassing voor de ontwikkeling van andere (wellicht decentrale) datacentra en wordt daarom ondersteund door SURFnet.

Om de functionele eisen en wensen in kaart te brengen voert de TU Delft Library naast desk research een casestudie uit die zich richt op de dataverzameling van onderzoeksdata, gerelateerd aan het specialisme Waterbouw. Hiertoe interviewde de TU Delft Library ervaringsdeskundigen over het gebruik van centrale data archivering. De resultaten zijn gepubliceerd in het tussenrapport “Rapportage Voorbereidend onderzoek & Good practice case”.

Het voorliggende rapport vormt een beschrijving van de werkprocessen van technische onderzoekers die geen gebruik maken van centrale digitale opslag. De leden van de projectgroep interviewden de onderzoekers van de vakgroep Waterbouwkunde van de faculteit Civiele Techniek en Geowetenschappen van de TU Delft. Gesproken werd met de universitair hoofddocent, een voormalig promovendus en een afstudeerder. Met hen werd het proces besproken en werd onderzocht welke ondersteuning een centrale digitale opslag van onderzoeksgegevens hen kan bieden.

1.1. Vraagstelling

De rapportage vormt een beschrijving van het werkproces van de onderzoekers. De centrale vraag die in dit deel van het onderzoek speelt is:

Hoe ziet uw onderzoeksproces eruit en op welke punten kan centrale opslag van digitale onderzoeksdata u ondersteunen?

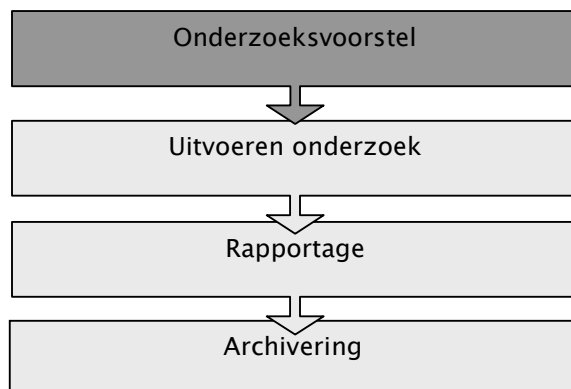
Vanuit deze onderzoeksvraag identificeren wij de processen. Door het uitvoeren van deze analyse identificeren wij mogelijke lacunes waar een centraal datacentrum een rol in kan spelen.

1.2. Leeswijzer

De rapportage beschrijft in hoofdstuk 2 de werkprocessen van de onderzoekers. De analyse van deze werkprocessen vormt tevens onderdeel van dit hoofdstuk. In hoofdstuk 3 wordt aangegeven welke toegevoegde waarde een centraal datacentrum kan bieden aan de onderzoekers.

2. Werkbeschrijving

In dit hoofdstuk gaan we dieper in op de werkprocessen van de sectie Waterbouwkunde van de faculteit Civiele Techniek en Geowetenschappen van de TU Delft. Het onderzoeksproces is te verdelen in vier onderdelen waarin een datacentrum een rol kan spelen. Figuur 1 geeft een overzicht van de verschillende onderdelen. Per onderdeel worden de activiteiten van de onderzoeker beschreven en worden aanbevelingen gedaan ten aanzien van de toegevoegde waarde van een datacentrum.



Figuur 1: Overzicht van de verschillende onderdelen van het onderzoek

2.1. Onderzoeksvoorstel schrijven

Vóór de start van het onderzoek schrijft de onderzoeker een onderzoeksvoorstel. Dit voorstel geeft een beschrijving van de wijze waarop het onderzoek uitgevoerd wordt. In het onderzoeksvoorstel zou de onderzoeker al budget dienen te reserveren voor de opslag van zijn meetdata. In veel gevallen wordt dit niet meegenomen.



Aanbeveling

- Het datacentrum moet als opslaglocatie aangeduid worden bij de start van het onderzoek
- Een deel van het onderzoeksbudget moet gereserveerd worden voor opslag van data in het datacentrum

2.2. Uitvoeren onderzoek

Het uitvoeren van onderzoek is te verdelen in twee hoofdactiviteiten.

1. De onderzoeker genereert allereerst een dataverzameling
2. De onderzoeker start met de analyse en bewerking van de verzamelde gegevens.

2.2.1. Dataverzameling

Voor het vakgebied van de sectie Waterbouw kan het doen van onderzoek zeer kostbaar zijn. Dataverzameling vindt (in de meeste gevallen) plaats op basis van het monitoren van de waterstroom door een waterbak. De onderzoeker richt de waterbak in op basis van zijn onderzoekswensen. Het gaat hierbij voornamelijk om laboratoriummetingen wat altijd kortdurende experimenten zijn. De sectie doet geen langdurige metingen van een proces. Het betreft zowel meetdata (signalen in volt) als beelddata (signalen in afbeeldingen; visueel verschil

tussen twee momenten in tijd). Met het verzamelen van de beelddata worden circa 100 afbeeldingen per seconde opgeslagen. Sinds kort zijn ook veldgegevens als moment/event opnames door het nemen van foto's of door gebruik te maken van elektrische pulsen onderdeel van de dataverzameling.

Wanneer de onderzoeker de dataverzameling aanlegt door het maken van foto's ontstaat een grote hoeveelheid data. De waterstroom wordt in dit geval vele malen per seconde gefotografeerd, waardoor de stroom zichtbaar is. Bij een dataverzameling met elektrische pulsen is de omvang van de verzamelde hoeveelheid data veel minder, maar desondanks aanzienlijk.

Tijdens het onderzoek slaat de onderzoeker de gegevens op een speciale server op. De data verplaatst hij vervolgens naar zijn persoonlijke computer voor het uitvoeren van de analyse. Het behoud van de gegevens wordt in dit geval via een lokale back-up gewaarborgd. De onderzoekers maken in veel gevallen tijdens het onderzoek back-ups op mobiele datadragers als tapes, cd's of dvd's. Wanneer het een dataverzameling door middel van foto's betreft, is een back-up op een externe harde schijf noodzakelijk, gezien de grote hoeveelheid meetdata.

Deze manier van data verzamelen en back-up zorgt ervoor dat data moeilijk toegankelijk is en niet beschikbaar voor vakcollega's. Oude data worden regelmatig overschreven. Hierdoor vindt hergebruik van data zelden tot nooit plaats.

Geproduceerde data heeft voor de vakgroep Waterbouw op internationaal niveau toegevoegde waarde. Het International Association of Hydraulic Engineering and Research (IAHR) heeft getracht in deze behoefte te voorzien met het inrichten van een centrale database voor de opslag van onderzoeksdata. Deze database wordt echter nauwelijks gebruikt door de beperkte vindbaarheid en werking.



Aanbeveling

- Het datacentrum moet voldoende capaciteit bieden voor zowel meetdata als fotodata
- Het datacentrum biedt een beveiligde omgeving waar onderzoeksdata opgeslagen staat
- Het datacentrum maakt hergebruik van onderzoeksdata mogelijk
- Het datacentrum moet gebruiksvriendelijk worden ingericht
- Het datacentrum moet over een goede vindbaarheid van gegevens beschikken
- Het datacentrum dient (op lange termijn) een internationale scope te hebben

2.2.2. Uitvoeren analyses

Na het verzamelen van de data voert de onderzoeker een analyse uit. Onderdeel van de analyse is het doen van berekeningen met de data. De ruwe data worden hierbij omgezet naar bewerkte data. Bewerkte data vragen een stuk minder opslagcapaciteit. Deze data worden gebruikt in scripties en publicaties. Andere wetenschappers kunnen geïnteresseerd zijn in de bewerkte data.

Het is mogelijk dat zich fouten voordoen in de bewerkte data (circa 1 % van de gevallen). In dat het geval moet de ruwe meetdata opnieuw geanalyseerd te worden. Dit gebeurt echter zelden. Fouten dienen handmatig te worden opgespoord door de onderzoeker. Geautomatiseerde controle is in de meeste gevallen niet mogelijk omdat het om een interpretatie van de onderzoeksdata gaat.

Naast de ruwe data slaat de onderzoeker voornamelijk de bewerkte data op. De onderzoekers hebben geen richtlijn en er zijn geen afspraken om te regelen hoe zij deze data op moeten slaan. Behalve een meetrapport – waarin de meetopstelling en wijze van verkrijgen van de onderzoeksdata staat beschreven – is geen metadata beschikbaar die direct aan de onderzoeksdata is gekoppeld. Ook de bewaartermijn van onderzoeksdata staat niet vast.



Aanbeveling

- Het datacentrum biedt een opslagmogelijkheid voor ruwe data tijdens het onderzoek
- Het datacentrum biedt een opslagmogelijkheid voor bewerkte data tijdens het onderzoek
- Het datacentrum maakt versiebeheer van de analyses mogelijk
- Het datacentrum moet een uniforme en zo uitgebreid mogelijke rapportage bevatten van de metadata

2.3. Rapportage

De rapportage van het onderzoek kan in meerdere vormen gedaan worden. In veel gevallen maakt het onderzoek deel uit van een promotie of een ander wetenschappelijk onderzoek. Tussentijds worden onderdelen van het onderzoek gepubliceerd in de vorm van artikelen voor wetenschappelijke tijdschriften. In deze publicaties vindt in toenemende mate een verwijzing plaats naar de meetdata van het onderzoek. Andere onderzoekers kunnen hierdoor refereren aan de meetdata. Dit levert de betreffende onderzoeker een citatie op.



Aanbeveling

- Het datacentrum biedt een ontsluitingsmogelijkheid voor data, waarmee verrijkte publicaties mogelijk zijn
- Het datacentrum moet een locatie zijn waaraan gerefereerd kan worden, waardoor wetenschappers citaties kunnen halen

2.4. Archivering

Na afloop van het onderzoek slaat de onderzoeker de verzamelde data vaak op een mobiele datadrager op. Deze dragers (floppy, tape, cd, dvd) zijn snel verouderd en de opslag is verouderd. Hierdoor raakt veel onderzoeksdata verloren. Wanneer naar aanleiding van een publicatie een controle op de analyse uitgevoerd moet worden, dient het gehele onderzoek herhaald te worden. Er bestaat geen standaardtermijn voor de opslag van de onderzoeksdata. Een historische analyse van de data is hierdoor niet mogelijk.

Voor de analyse maakt de onderzoeker gebruik van speciale analysesoftware. Indien een onderzoeker een nieuwe analyse wil uitvoeren op een dataset, dient de meting herhaalbaar te zijn. Dit is vaak niet mogelijk doordat de gebruikte software verouderd – en hierdoor niet meer beschikbaar – is. De gesproken onderzoekers zouden graag de bijbehorende software tot hun beschikking hebben.

Het eventueel vrijgeven van de gearchiveerde data gebeurt na publicatie van het onderzoek. Deze werkwijze hanteren de onderzoekers om te voorkomen dat derden de onderzoeksdata eerder publiceren, waardoor het eigen onderzoek aan waarde inboet. Het format dat de onderzoeker hanteert in de opslag van de data ligt niet vast. Hierdoor is data niet altijd

gemakkelijk te interpreteren indien de data hergebruikt of getoetst wordt. Ook voor het beschrijven van de metadata bestaan momenteel geen standaarden.



Aanbeveling

- Het datacentrum moet duurzame opslag van data garanderen
- Het datacentrum kan mogelijk de analysesoftware aanbieden bij de opslag van data
- Het datacentrum moet een structuur bieden voor de opslag van data
- Het datacentrum voorziet in een template als standaard voor de opslag van metadata

3. Aanbevelingen

In de werkprocesbeschrijving hebben we laten zien waar een datacentrum toegevoegde waarde kan hebben voor het uitvoeren van (wetenschappelijk) onderzoek en de archivering van de onderzoeksdata. Onderstaande tabel geeft een overzicht van deze aanbevelingen. De aanbevelingen vormen de input voor een centrale discussie met onderzoekers van de drie technische universiteiten, als onderdeel van het uitvoeren van het project 'Waardevolle Data & Diensten'.



Aanbevelingen

- Het datacentrum moet als opslaglocatie aangeduid worden bij de start van het onderzoek
- Een deel van het onderzoeksbudget moet gereserveerd worden voor opslag van data in het datacentrum
- Het datacentrum moet voldoende capaciteit bieden voor zowel meetdata als fotodata
- Het datacentrum biedt een beveiligde omgeving waar onderzoeksdata opgeslagen staan
- Het datacentrum maakt hergebruik van onderzoeksdata mogelijk
- Het datacentrum moet gebruiksvriendelijk worden ingericht
- Het datacentrum moet over een goede vindbaarheid van gegevens beschikken
- Het datacentrum dient (op lange termijn) een internationale scope te hebben
- Het datacentrum biedt een opslagmogelijkheid voor ruwe data tijdens het onderzoek
- Het datacentrum biedt een opslagmogelijkheid voor bewerkte data tijdens het onderzoek
- Het datacentrum maakt versiebeheer van de analyses mogelijk
- Het datacentrum moet uniforme en zo uitgebreid mogelijke metadata bevatten
- Het datacentrum biedt een ontsluitingsmogelijkheid voor data, waarmee verrijkte publicaties mogelijk zijn
- Het datacentrum biedt een locatie waaraan gerefereerd kan worden, waardoor wetenschappers citaties kunnen halen
- Het datacentrum kan duurzame opslag van data garanderen
- Het datacentrum kan mogelijk de analysesoftware aanbieden bij de opslag van data
- Het datacentrum biedt een structuur voor de opslag van data
- Het datacentrum voorziet in een template als standaard voor de opslag van metadata