



UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Scientific Information Management in Collaborative Experimentation Environments

Kaletas, E.C.

Publication date
2004

[Link to publication](#)

Citation for published version (APA):

Kaletas, E. C. (2004). *Scientific Information Management in Collaborative Experimentation Environments*.

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Samenvatting*

Er bestaat een aantal bepalende krachten voor onderzoeksactiviteiten die ter ondersteuning dienen voor wetenschappelijke experimenten en toepassingen. De revolutionaire technologische vooruitgangen in laboratoriuminstrumenten, en technieken die enorme hoeveelheden gegevens genereren en meer ingewikkelde experimenten mogelijk maken, kunnen geschaard worden onder de 'aantrekkelijke krachten'. De technologische vooruitgangen in de ICT, daarentegen, behoren tot de 'afstotende krachten'.

De aantrekkelijke krachten bereiden de weg voor voorheen onuitvoerbare experimenten, voor nieuwe oplossingsmethoden en -technieken, en derhalve voor nieuwe resultaten. Echter, de aantrekkelijke krachten resulteren tegelijkertijd in grotendeels geautomatiseerde, doch zeer ingewikkelde experimenten, die grote hoeveelheden gegevens genereren en dientengevolge geavanceerde oplossingen vereisen voor gegevensbeheer en -analyse.

Hier staat tegenover dat, hoewel de afstotende krachten reeds een belangrijke rol spelen in het ondersteunen van wetenschappelijke experimenten en toepassingen, een gecombineerd gebruik van beschikbare technologieën wetenschappers zeker zal helpen bij het beter exploiteren van hun experimentele resultaten, alsook bij het verkrijgen van nieuwe inzichten in het probleem domein.

In deze dissertatie refereert de term *collaboratieve experimenteeromgevingen* (CEO's) naar geïntegreerde oplossingen en ondersteunende omgevingen, die alle aspecten van experimenten ondervangen, en die wetenschappers van nut zijn in elke levensfase van een experiment. Een CEO biedt derhalve de vereiste combinatie van de beschikbare (aantrekkelijke en afstotende) technologieën. In een complexe omgeving als een CEO moeten vele kwesties onder de loep genomen worden, waaronder de representatie en het beheer van experimenten, de aan deze experimenten gerelateerde gegevens, en de functionaliteit voor het beheer van experimenten en gegevens. Vooral het beheer van wetenschappelijke informatie is één van de meest belangrijke kwesties, omdat het in alle stappen van het experimenteerproces is vereist.

Het beheer van informatie in een CEO introduceert verscheidene uitdagingen, zoals het modelleren en beheren van ingewikkelde experimentele procedures, en van heterogene gegevens. In deze context is de belangrijkste doelstelling van dit proefschrift: het analyseren van de eisen voor wetenschappelijke experimenten in verschillende onderzoeksdisciplines, en het aandragen van generieke en herbruikbare oplossingen voor de aan informatiebeheer gerelateerde problemen waar wetenschappers gedurende hun experimenten mee te maken hebben.

Om deze doelstelling te bereiken is eerst een uitgebreide studie uitgevoerd naar verschillende experimenten en toepassingen in experimentele wetenschapsdomeinen. De studie omvatte life sciences experimenten in microarray en confocale microscopie, natuurkundige

*Vertaling door Frank Seinsträ

experimenten in materiaalanalyse, en verkeerssimulaties. Daarnaast zijn vele experimenten en toepassingen bestudeerd op basis van de literatuur, waaronder medische toepassingen en additionele natuurkundige toepassingen. Deze studie heeft geleid tot de identificatie van de aan informatiebeheer gerelateerde problemen waar wetenschappers mee te maken hebben tijdens hun experimenten.

Gebaseerd op de studieresultaten en de geïdentificeerde problemen is een analyse uitgevoerd om het eisenpakket, waaraan ter ondersteuning van wetenschappers tegemoet gekomen moet worden, te identificeren en classificeren. De classificatie heeft geresulteerd in twee groepen van eisen, namelijk: gebruikerseisen en standaard ICT-eisen. De eerste categorie bevat de behoeften en verwachtingen van gebruikers van het virtueel laboratorium. De tweede categorie is daarentegen meer oplossingsgeoriënteerd en bevat eisen voor de standaard ICT infrastructuur die vereist is om op een juiste wijze te kunnen voldoen aan de gebruikerseisen.

De volgende stap die genomen is alvorens te komen tot het ontwerp en de ontwikkeling van een informatiebeheerplatform voor CEO's, was een grondige studie en analyse van de state-of-the-art op dit gebied. Deze studie bestond uit twee onderdelen:

1. Onderzoek naar bestaande en opkomende technologieën, paradigma's en standaarden, die de ontwikkeling van ondersteuningsomgevingen voor complexe experimenten mogelijk maken (de zogenaamde 'enablers'). Het onderzoek omvatte informatiemodellen en -standaarden, gedistribueerde informatiebeheersystemen, resource management technologieën, en andere gerelateerde technologieën, paradigma's en hulpmiddelen, zoals workflow management en virtuele organisaties. Deze gebieden zijn gekozen op basis van het eerder geïdentificeerde pakket van eisen.
2. Een uitgebreid overzicht van verschillende ondersteuningsomgevingen voor complexe experimenten, in combinatie met een beschrijving van de wijze waarop wetenschappelijk informatiebeheer in elk hiervan is aangepakt. Het beschrijft elke ondersteuningsomgeving en geeft inzicht in de aanwezige functionaliteit gegeven de specifieke focus op informatiebeheer. Tevens zijn de beschreven ondersteuningsomgevingen geëvalueerd aan de hand van verschillende criteria, waaronder het geïdentificeerde eisenpakket, de aanwezige functionaliteit, de toegepaste 'enablers', en de gebruikte implementatietechnologieën. Het overzicht en de evaluatie geven een beeld van de state-of-the-art in wetenschappelijk informatiebeheer.

Het onderzoek naar wetenschappelijke experimenten, het overzicht van bestaande en opkomende wetenschappelijke informatiebeheeraanpakken, alsmede het aangegeven pakket van eisen, vormen gezamenlijk de basis voor de definitie van het raamwerk voor informatiebeheer in een CEO. Omdat informatiebeheer in alle fasen van wetenschappelijke experimenten vereist is, is het informatiebeheerraamwerk in een CEO direct van invloed op alle andere CEO-componenten. Voornamelijk door het modelleren van wetenschappelijke experimenten wordt de basis gelegd voor een juiste beschrijving van de overige aspecten van een CEO. Als gevolg van dit gegeven is een experimentmodel geïntroduceerd dat in staat is heterogene wetenschappelijke experimenten op *uniforme* wijze te representeren, zodanig dat het alle aspecten van deze experimenten omvat, van experimentele procedures tot aan computationele bewerkingen. Een ander belangrijk CEO-component, dat verantwoordelijk is voor het presenteren van de door de CEO aangeboden functionaliteit en de door de CEO gemanipuleerde gegevens/informatie, is de CEO-gebruikersomgeving. In een multi-disciplinair CEO is het essentieel dat de gebruikersomgeving de wetenschappers *uniforme* en *herbruikbare* interfaces biedt. Uniforme interfaces en herbruikbare componenten voor de CEO-gebruikersomgeving

worden verkregen uit de uniformiteit en uitbreidbaarheid van het gedefinieerde experiment-model. Naast de CEO-gebruikersomgeving is het beheer van informatie in een CEO eveneens op het experimentmodel gebaseerd. Diverse gegevensmodellen zijn ontwikkeld voor de representatie van verschillende soorten gegevens/informatie die door een CEO worden gegenereerd en gemanipuleerd. Naast gegevensmodellen voor informatie aangaande wetenschappelijke experimenten, bevatten de ontworpen gegevensmodellen modellen voor informatie over gebruikers, toegangsrechten, beschikbare bronnen, het delen en hergebruiken van objecten, en eigenschappen van schermweergave. Functionaliteit voor opslag en beheer van de gemodelleerde gegevens/informatie is eveneens ontworpen.

Om de ontworpen functionaliteit te ondersteunen is een architectuur voor het informatiebeheerplatform ontworpen en geïmplementeerd binnen de VLAM-G collaboratieve experimenteeroomgeving. Het ontwerp van de architectuur behandelt verscheidene uitdagingen in wetenschappelijk informatiebeheer, van het bieden van uniforme gebruikersinterfaces voor uiteenlopende functionaliteit, tot aan modulariteit, openheid, en flexibiliteit. Door modulariteit en uniforme interfaces wordt onderlinge afhankelijkheid tussen componenten voorkomen, en wordt uitbreiding van de architectuur mogelijk als dit nodig mocht zijn. De implementatie van het VIMCO-informatiebeheerplatform toont de toepasbaarheid aan van de voor de CEO voorgestelde gegevens- en functionaliteitsmodellen.

Tot slot wordt het gebruik gepresenteerd van de concepten, de modellen, en het ontwikkelde informatiebeheerplatform, in een werkelijk bestaande toepassing. De DNA microarray-applicatie toont de toepasbaarheid aan van het ontwikkelde platform in een complex experiment in het life sciences onderzoeksgebied.

