Harmonisation of distributed geographic datasets

A model driven approach for geotechnical & footprint data

Master of Science Thesis

A.H.H.J. van den Berg

Professor: prof.dr.ir. P.J.M. van Oosterom
Supervisors: drs. T.P.M. Tijssen
ir. B.J Hulshof

Assen, December 2004

Section GIS-technology
Geodetic Engineering
Faculty of Civil Engineering & Geo-Sciences
Delft University of Technology

Geo-Information Management
Europe Data Acquisition & Management
Shell Exploration & Production Europe

Preface

This thesis is the result of the research project I conducted at Shell EP Europe – NAM. This research is conducted and reported in order to obtain the title of the Master of Science in Geodetic Engineering at the Delft University of Technology. The aim of this research is to harmonise the different datasets in use by Shell on different locations, in such a way as to realise eventually a so-called geographic information infrastructure in which all geographic data can be shared seamlessly. Within Shell EP Europe the geotechnical and footprint datasets were taken as the case for this research.

This eight-month research couldn't be done solely on my own. Many people were directly or indirectly involved with this project, and helped, supported or distracted me when needed. Therefore I'd like to thank them all for their help and I'd like to mention some of the especially.

At Delft University of Technology I thank my professor Peter van Oosterom, for the discussions we had and advise he gave, my supervisor Theo Tijssen, and the section GISt.

At Shell EP Europe I'd like to thank my supervisor Bart Hulshof, for the guidance, confidence and steering; the whole EDAM-department: GIM, Offshore and Onshore Surveys, for their support and good-fellowship; and David van Dalen, from Fugro-Inpark, for the discussions and interest.

Besides the scientific and professional interest I'd like to thank all who supported me in my living divided between Groningen and Delft: the NAM Student Society, for all the distractions: the lunches, the cups of tea and the beers; my student house in Groningen: IDBVP, for their amicability and taking me in as one of them; and of course my friends in Delft, for still being there.

Special thanks I'd like to give to my parents and my brother, for helping me out with limitless support and crucial advice and for just always being there when needed.

Assen, December 28th, 2004

Bart van den Berg

Summary

Introduction

In this so-called information age, having control over and access to data and information is regarded vital for organisations. Within Shell Exploration & Production (Shell EP) it has turned out that 80 percent of all information used has a geographic context, holding a reference to a position on the globe. This information, better known as geographic information, is usually available as datasets stored in databases and accessible via Geographic Information Systems (GIS). However, these GISs and databases are often isolated and distributed, each of them created for one specific purpose.

In recent years the focus has started to shift from managing several stand-alone geo-information datasets to harmonised management of interoperable GISs. Through interoperability and standardisation these harmonised networks improve the availability, accessibility and usability of geographic data. Such a facility is the Geographic Information Infrastructure (GII).

Also Shell Exploration & Production in Europe (Shell EP Europe), the operating company of Shell for European countries, has made a start to harmonise the diverse and distributed offshore point datasets. This is composed in the project "aligning offshore infrastructure data-models". Commonly at the departments involved this project is referred to by the name of its data-model: Fixed And Mobile Entities version 2 (FAME-2).

Within the FAME-2 project many datasets are involved, varying in usage, disjunction, level of digitisation, number of stakeholders involved, etc. To reduce complexity the two most distinct and complex datasets are separated to form a different, but parallel project. These datasets are the geotechnical sites and the footprints:

- Geotechnical data, in regard to Shell EP Europe's offshore activities, are seabed soil data stating the composition of the seabed. They are required for the foundation of installations.
- Footprint data are derived from the imprints left by a Jack-up rig on the seabed. These data are required for they cause unevenness of the seabed, which should be avoided in the positioning of installations.

Shell EP Europe wants to have these datasets analysed and harmonised to the requirements of the corporate GII, and to have them re-integrated into the FAME-2 data-model. To come to a harmonised solution for these geotechnical datasets and footprint datasets is the main theme of this research project. Therefore the main research question is:

> *How can geotechnical & footprint data be harmonised to a data-model that fits into the corporate Geo-Information Infrastructure of Shell EP Europe?*

This research concerns the development of a harmonised data-model for geotechnical data and footprint data to implement into the corporate GII. A harmonised data-model implies that there should be concordance between the different formats of the datasets involved to allow a consistent infrastructure, and a merge from the datasets held at the different locations and departments involved, to come to a common data- and workflow.

This research has followed the phases of a development process: definition, information, modelling, implementation and maintenance phase. According to the phase, several research methods have been used: desk research, literature study, work visits, interviews and questionnaire, modelling process and usability testing.

Definition phase

In this research I have used the next definition of a GII: a facility to improves availability, accessibility and usability of geographic data for all users within all levels of an organisation. A GII enables the interoperability and harmony between geographic datasets, being composed of four main components: Datasets, Technology, Standards and Organisation.

For a corporate GII, Shell EP has set up a guiding strategy for standardised geo-information handling (Shell Geo-Information Strategy), which provides recommendations for the operating companies about the system architecture and standards to use. In short, the Geo-Information Strategy aims to realise standardised handling of geographic data, by storage in Oracle databases, by access via ArcGIS and ArcIMS, and by links to various stored documents and other non-geographic corporate datasets. Within the system architecture both internal Shell standards and external (ISO, OGC, EPSG) standards are practised.

Information phase

An inventory of the current datasets available at the different locations is set up by means of interviews, a questionnaire, analysis of work- and dataflows concerning the geotechnical data and footprint data. It gives an insight in the data-attributes, data-models and formats regarding these datasets.

Shell EP Europe used two footprint datasets, one at NAM and one at Shell Expro. There were four geotechnical datasets in use: one at NAM and three at Expro. There is also a European webpage being a metadata portal for seabed soil data.

For the geotechnical data, there was no data-model in Shell EP Europe. For the footprint data only one data-model was available, which was part of the first edition of FAME. This model was not in use, and is not suitable as Shell EP Europe-wide data-model. The absence of one single data-model for the combination of geotechnical & footprint data sets the need for a new data-model.

Modelling phase

The reason for developing data-model at all is based on the concept of the Model Driven Architecture (MDA), set up by the Object management Group (OMG). In this MDA a data-model is used as representation of the real world in the computer. Main idea of the MDA is that a data-model is the centre of an information system, and from this data-model all other functions are derived. It will result in one data-model in which all datasets are harmoniously proportioned.

An information system contains the following components: Technology, Users, Data and Data-model. The conditions for the Data-model can be derived from the analysis of the other three components. The conditions involved are: the recommended architecture of the Shell geo-information strategy; the data-attributes, dataflow and relations to the other datasets; the users' conditions as described in the workflow.

From these conditions, the dataset inventory and the definitions a data-model is created that harmonises the geotechnical data and footprint data and fits them harmoniously in the corporate GII.

Research of the possibilities to fit the data-model in the corporate GII, led to the conclusion that the geotechnical & footprint data-model could be integrated in the FAME-2 data-model. In line with the ideas of a corporate GII and harmonisation, the geotechnical & footprint data-model is therefore implemented in the FAME-2 data-model. In this model they are related to the other datasets.

Implementation phase

According to the concept of MDA – having the data-model as central axis of an information system – the developed data-model is used as starting point in the implementation process. From the data-model the implementations towards the other three components of the information system are managed. These three implementations are:

- Configuration of the Technology; the soft- and hardware to support the data-model.
- Interface for the Users; created and tested for the new system.
- Migration of the Data; from the old systems to the new system.

The technology is configured by the recommended architecture of the corporate GII-strategy. For this configuration automated tools are available to implement the data-model. One of these tools is converting an UML-structured data-model through MS Visio and ArcCatalog to an ESRI geodatabase. In this research the possibilities for using this tool for the developed data-model are investigated. It proved to be functioning. In this specific case this tool is not recommended to apply, since the use of the specific type of ESRI geodatabase is unsought for strategic reasons.

The user interface is custom developed for the data custodians. With an interface-prototype a usability test is carried out. It resulted in recommendations for this data management interface and improvements in the usability, enhancing of the corporate GII.

The migration of the data is worked out in a data-migration listing. This listing states the data-attributes of the original datasets, their related counterparts and the transformation required to come from the old datasets to the new dataset.

Maintenance phase

Based on my research done, several conditions and recommendations are given for well functioning of the system. These are split up according to the components of a GII:

- Organisation: Set up a user's guide to have a safety net for the current users and a guideline for new users. This guide should consist of at least a data dictionary and an interface walkthrough. Have a high level of communication, since there are many departments and locations involved. Set out the responsibility for both the system and the datasets clearly. There have to be one or two responsible persons for both the data and the system; more is not recommendable since there is only one system in which one set of interdependent data is stored.

- Standardisation: Standardise the workflows and data specifications; the new standardised list can be distributed to the contractors, who can deliver new data directly in the right format. Keep going with the standards set in the geo-information strategy and try to elaborate them and implement them more widely.
- Technology: Set out the responsibility for the system clearly. Use views in overcoming the gap between the data the users wish to see and the data stored in the database. Other applications can be linked to these views as well, without needing to de-normalise the tables themselves. Look at the possibilities for a totally web-based interface, for both the data-requests, the data upkeep and the Hazard Notification, which is easier to maintain and manage.
- Datasets: Clearly set out the responsibility for the data management. Fill in the data gaps as soon as possible, and keep the data as accurate as possible. Consider the use of 3rd party as available through the EU-Seased website, either by taking access to this website, or by copying relevant data to the Shell Environment.

Conclusion

In this research a model has been developed to harmonise two different kinds of distributed geographic datasets within a corporate GII. This is based on the MDA-concept, under the conditions set by the information system's components of Data, Users and Technology. The implementation of this model has led to the conclusion that this model is not only developed on theoretical and methodological concepts, but it is applicable in a real working situation. Due to the results of my research it has been introduced and implemented within Shell EP Europe.

Although it is beyond the limitations of this research project, it can be assumed that this model is also applicable in other organisations. This case, researched in Shell EP Europe, proved to be best practice. The harmonisation method used can also be applied to other distributed geographic datasets. Probably the model will need only minor adjustments to make it applicable in other organisations. More research on this can prove this.

Samenvatting

Inleiding

In dit huidige informatietijdperk is het van vitaal belang voor organisaties om toegang en controle te hebben over data en informatie. Bij Shell Exploratie & Productie (Shell EP) is gebleken, dat 80 procent van alle informatie een geografische context bevat, wat inhoudt, dat het een referentie heeft ten opzichte van de aardbol. Deze informatie, beter bekend als geografische informatie, is gewoonlijk beschikbaar in de vorm van datasets die zijn opgeslagen in databanken en toegankelijk zijn via Geografische Informatie Systemen (GIS). Echter, deze GIS'en en databanken zijn veelal geïsoleerd en verdeeld, ontworpen voor één specifiek doel.

Recentelijk is de focus begonnen te verschuiven van het beheren van verscheidene, op zichzelf staande geo-informatie datasets naar geharmoniseerd beheer van interoperabele GIS'en. Zulke geharmoniseerde netwerken bevorderen de beschikbaarheid, toegankelijkheid en gebruiksvriendelijkheid van geografische data door middel van interoperabiliteit en standaardisatie. Zo'n bevorderende faciliteit wordt aangeduid als Geografische Informatie Infrastructuur (GII).

Ook Shell Exploratie & Productie in Europa (Shell EP Europe), de werkmaatschappij van Shell voor de Europese landen, is begonnen met het harmoniseren van de diverse offshore puntdatasets. Dit is ondergebracht in het project "aligning offshore infrastructure datamodels". Gewoonlijk wordt dit project binnen de betrokken afdelingen aangeduid met de naam van zijn datamodel: Fixed And Mobile Entities versie 2 (FAME-2).

Binnen het FAME-2 project zijn verscheidene datasets betrokken, variërend in gebruik, verdeeldheid, digitaliseringgraad, hoeveelheid betrokkenen, etc. Om de complexiteit te verminderen zijn de twee meest complexe en verschillende datasets afgescheiden om een apart, doch parallel project te vormen. Deze datasets betreffen de geotechnische testlocaties en de Jack-up rig voetafdrukken:

- Geotechnische data, in relatie tot Shell EP Europe's offshore activiteiten, zijn bodemdata die de compositie van het zeebed weergeven. Ze zijn benodigd voor het bepalen van de funderingen van installaties.
- Voetafdrukdata zijn afkomstig van de afdrukken die een Jack-up rig op het zeebed achterlaat. Dit veroorzaakt gaten in het zeebed. Deze oneffenheden van het zeebed moeten worden vermeden bij het plaatsen van installaties.

Shell EP Europe wil dat deze datasets worden geanalyseerd en geharmoniseerd ten opzichte van de bedrijfs-GII, en zou ze graag gereïntegreerd hebben in het FAME-2 datamodel. Om te komen tot een geharmoniseerde oplossing voor deze geotechnische & voetafdruk datasets is dit onderzoeksproject opgezet. Dit herleidde ik tot de volgende onderzoeksvraag:

> *Hoe kunnen geotechnische & voetafdruk data worden geharmoniseerd tot een datamodel dat past in de bedrijfs-Geo-Informatie Infrastructuur van Shell EP Europe?*

Dit onderzoek betreft het ontwikkelen van een geharmoniseerd datamodel voor geotechnische data en voetafdrukdata voor implementatie in de bedrijfs-GII. Een geharmoniseerd datamodel houdt in, dat er overeenstemming moet zijn tussen de verschillende formaten van de betrokken datasets om een consistente infrastructuur toe te staan, en een samenvoeging van de datasets die

op de verschillende betrokken locaties en afdelingen worden beheerd, om uiteindelijk te komen tot een gemeenschappelijke data- en workflow.

Dit onderzoek heeft de fasen van een ontwikkelproces gevolgd: definitiefase, informatiefase, modelleerfase, implementatiefase en onderhoudsfase. In overeenstemming met de fase, zijn er verscheidene onderzoeksmethoden gebruikt: bureauonderzoek, literatuuronderzoek, werkbezoeken, interviews en enquête, modelleerproces en gebruiksvriendelijkheidtest.

Definitiefase

In dit onderzoek is de volgende definitie van een GII gebruikt: een faciliteit die de beschikbaarheid, toegankelijkheid en gebruiksvriendelijkheid van geografische data bevordert voor alle gebruikers op elk niveau van een organisatie. Een GII maakt interoperabiliteit en harmonie tussen geografische datasets mogelijk en bestaat uit vier componenten: Datasets, Technologie, Standaarden en Organisatie.

Voor de bedrijfs-GII heeft Shell EP een richtlijn (Shell Geo-Informatie Strategie) uitgezet voor gestandaardiseerde verwerking van geo-informatie. Hierin worden aanbevelingen gedaan voor de werkmaatschappijen over welke systeemarchitectuur en welke standaarden te gebruiken. In het kort is het doel van de Geo-Informatie Strategie om een gestandaardiseerde verwerking van geografische data te realiseren, door opslag in Oracle databanken, door toegang via ArcGIS en ArcIMS, en door links naar de verscheidene opgeslagen documenten en andere niet-geografische bedrijfs-datasets. Binnen de systeemarchitectuur wordt het gebruik van zowel interne Shell standaarden als externe (ISO, OGC, EPSG) standaarden aangeraden.

Informatiefase

Van de huidige datasets die beschikbaar zijn op de verschillende locaties is een inventarisatie gemaakt door middel van interviews, een enquête, analyse van work- en dataflows van de geotechnische data en voetafdrukdata. Het geeft een overzicht van de data-attributen, datamodellen en formaten van deze datasets.

Shell EP Europe gebruikt twee voetafdruk datasets, één bij NAM en één bij Shell Expro. Er worden vier geotechnische datasets gebruikt: één bij NAM en drie bij Expro. Ook is er een Europese webpagina met metadata portaal voor bodemgesteldheiddata van het zeebed.

Voor de geotechnische data was er geen datamodel in Shell EP Europe aanwezig. Voor de voetafdruk data was er slechts één datamodel beschikbaar, welke deel is van de eerste versie van FAME. Dit model werd niet gebruikt en is niet geschikt om als algemeen Shell EP Europe datamodel te dienen. Het ontbreken van één specifiek datamodel voor de combinatie van geotechnische & voetafdrukdata geeft de noodzaak voor een nieuw datamodel.

Modelleerfase

De reden om überhaupt een nieuw datamodel te ontwikkelen is gebaseerd op het concept van de Model Driven Architecture (MDA), ontworpen door de Object management Group (OMG). In de MDA wordt een datamodel gebruikt als representatie van de werkelijkheid in de computer. Het basisidee van de MDA is dat een datamodel het centrum is van een informatiesysteem, en vanuit dit datamodel worden alle andere functies afgeleid. Het resulteert in één datamodel in welke alle datasets in harmonie met elkaar zijn.

Een informatiesysteem bestaat uit de volgende componenten: Technologie, Gebruikers, Data en Datamodel. De condities voor het Datamodel kunnen worden afgeleid uit de analyse van de andere drie componenten. In deze casus worden de condities voor het datamodel bepaald door: de aanbevolen systeemarchitectuur van de Shell geo-informatie strategie; de data-attributen, de dataflow en de relaties tot de andere datasets; de condities gesteld door de gebruikers zoals beschreven in de workflow.

Van deze condities, de inventarisatie van de datasets en de definities is een datamodel gecreëerd dat de geotechnische data en voetafdruk data onderling harmoniseert en in harmonie brengt met de componenten van de bedrijfs-GII.

Onderzoek naar de mogelijkheden om het datamodel in de bedrijfs-GII op te nemen leidde tot de conclusie dat het geotechnische & voetafdruk datamodel kon worden geïntegreerd in het FAME-2 datamodel. In overeenstemming met de ideeën van een bedrijfs-GII en harmonisatie, is het geotechnische & voetafdruk datamodel dan ook geïmplementeerd in het FAME-2 datamodel. In dit model zijn deze datasets gerelateerd aan de andere datasets.

Implementatie fase

In overeenstemming met het MDA-concept – een datamodel als centrale as van een informatiesysteem – is het ontwikkelde datamodel gebruikt als uitgangspunt voor het implementatieproces. Vanuit het datamodel zijn de implementaties naar de andere drie componenten van het informatiesysteem geleid. Deze drie implementaties zijn:

- Configuratie van de Technologie; de soft- en hardware om het datamodel te ondersteunen.
- Interface voor de Gebruikers; ontworpen en getest voor het nieuwe systeem.
- Migratie van de Data; van de oude systemen naar het nieuwe systeem.

De technologie wordt geconfigureerd door middel van de aanbevolen architectuur van de bedrijfs-GII-strategie. Voor deze configuratie zijn er automatische tools beschikbaar om het datamodel te implementeren. Één van deze tools is de conversiemogelijkheid om van een in UML beschreven datamodel met behulp van MS Visio en ArcCatalog naar een ESRI geodatabase te komen. In dit onderzoek zijn de mogelijkheden voor het gebruik van deze tool voor het ontwikkelde datamodel onderzocht en is het toepasbaar gebleken. Voor deze specifieke casus is echter aangeraden deze tool niet te gebruiken, daar het gebruik van het specifieke type van ESRI geodatabase om strategische redenen niet gewenst is.

De gebruikersinterface is speciaal ontwikkeld voor de databeheerders. Op een interfaceprototype is een gebruiksvriendelijkheidtest uitgevoerd. Het resulteerde in aanbevelingen voor deze datamanagement-interface en vooruitgang van de gebruiksvriendelijkheid, wat leidt tot verbeteringen van de bedrijfs-GII.

De migratie van de data is uiteengezet in een datamigratie -overzicht. Dit overzicht geeft de data-attributen van de originele datasets weer, de gerelateerde tegenhangers en de benodigde transformatie om van de oude datasets naar de nieuw dataset te komen.

Onderhoudsfase

Gebaseerd op dit onderzoek zijn er verscheidene voorwaarden en aanbevelingen te geven voor het goed functioneren van het systeem. Deze kunnen worden uiteengezet volgens de componenten van een GII:

- Organisatie: Zorg voor een handleiding als ondersteuning voor de huidige gebruikers en als leidraad voor nieuwe gebruikers. Deze handleiding moet minstens een bestek met uitleg en een beschrijving van de interface bevatten. Zorg voor een hoog mate van communicatie, daar er veel afdelingen en locaties bij betrokken zijn. Zorg voor een duidelijk bepaalde verantwoordelijkheid voor zowel het systeem als de datasets. Er moeten één of twee verantwoordelijken zijn; meer is niet wenselijk omdat er slechts één systeem is waarin één set van onderling afhankelijke data is opgeslagen.
- Standaardisatie: Standaardiseer de workflows en dataspecificaties; een nieuwe gestandaardiseerde lijst van data-attributen kan worden doorgegeven aan de aannemers, die nieuwe data dan direct in het juiste formaat kunnen aanleveren. Ga door met het gebruik van standaarden, zoals aanbevolen in de geo-informatie strategie, en probeer dit uit te breiden en verder te implementeren.
- Technologie: Zorg voor een duidelijk bepaalde verantwoordelijkheid voor het systeem. Gebruik views om de kloof te overbruggen tussen de data-attributen die de gebruikers wensen te zien en de data-attributen zoals die worden opgeslagen in de databank. Tevens kunnen andere applicaties worden gelinkt naar deze views, zonder dat de tabellen zelf ge-de-normaliseerd hoeven worden. Kijk naar de mogelijkheden voor een interface die helemaal beschikbaar is via het web, voor zowel de bevraging van de data, de bijhouding van de data, als de ‘Hazard Notification’.
- Datasets: Zorg voor duidelijk bepaalde verantwoordelijkheid voor de data. Vul de gaten in de datasets zo spoedig mogelijk en houd de data zo accuraat mogelijk. Creëer een dataset van de 3rd party data, omdat overleg met de EU-Seased webpagina geen handeling is in de workflow, of zorg dat dit wel een standaard handeling wordt in de workflow.

Conclusie

In dit onderzoek is een datamodel ontwikkeld om twee verschillende soorten van verdeelde geografische datasets binnen een bedrijfs-GII te harmoniseren. Dit is gebaseerd op het MDA-concept, doch binnen de condities die gesteld zijn door de overige componenten van het informatiesysteem: Data, Gebruikers en Technologie. De implementatie van dit model heeft geleid tot de conclusie dat dit model niet alleen ontwikkeld is op theoretische en methodologische concepten, maar ook bruikbaar is in een realistische, werkende situatie. Door de resultaten van mijn onderzoek is dit inmiddels geïntroduceerd en geïmplementeerd bij Shell EP Europe.

Hoewel het buiten dit onderzoeksproject valt, kan worden aangenomen, dat dit model ook toepasbaar is in andere organisaties. Deze casus, onderzocht bij Shell EP Europe, bewees best practice te zijn. De gebruikte harmonisatiemethode kan ook worden toegepast op andere verdeelde geografische datasets. Waarschijnlijk heeft het model slechts kleine aanpassingen nodig om toepasbaar te zijn voor andere organisaties en datasets. Meer onderzoek kan dit aantonen.

Index

1 Introduction

NAM ET IPSA SCIENTIA POTESTAS EST

Francis Bacon, Meditationes Sacrae 11, 1597

"Knowledge itself is power" is what Francis Bacon already said at the end of the 16th century. And especially nowadays, in this so-called information age, this is regarded a basic truth. To gain knowledge one should interpret an appropriate collection of useful information. Information in its turn is derived from combining related data [33]. This is the so-called information hierarchy, which can be regarded the basis for information sciences and knowledge management [2]. Having access to and control over the data and information is what is regarded to be the backbone of many organisations [56].

Of all information used, stored and circulated, it is said that approximately 80 percent has a geographic context [54]. This means, that this information has a reference to a position on the globe, either via coordinates, address data or administrative areas. Information used and stored by geographic context is usually referred to as geographic information [18]. Geographic information is usually available as datasets stored in databases and accessible via Geographic Information Systems (GIS).

However, these GISs and databases are often isolated and distributed, each of them created for one specific purpose. From the information hierarchy comes forth that information is combined and interpreted data. Thus if these isolated and distributed geographic datasets could be combined, they could offer new possibilities for information gathering and from that new knowledge can be gained.

In recent years the focus has started to shift from managing several stand-alone geo-information datasets in stand alone or isolated network environments to harmonised management of interoperable GISs. This interoperability and harmonisation enables the sharing and accessibility of geographic data throughout networked systems. Such a facility that enables the interoperability and harmony between geographic datasets is the Geographic Information Infrastructure (GII). It improves availability, accessibility and usability of geographic data "for all users within all levels of an organisation" [38].

In this field of interoperability and standardisation several players are active. The European Petroleum Survey Group (EPSG) has standardised all coordinate reference systems and transformations that are used around Europe. The OpenGeospatial Consortium (OGC) endeavours to make spatial data available through any kind of (networked) system by specifying interoperability standards. The International Organisation for standardisation (ISO) has a special work group (ISO-TC 211) for developing standards in the field of Geographic Information and Geomatics. [48]

Also for the information industry in general there are standardisation concepts and specifications. The Object Management Group (OMG) produces concepts and specifications for interoperability. They have set up the Unified Modeling Language UML, which allows the structuring data-models using a schematic visualisation [42]. Recently they have started to develop the concept of the Model Driven Architecture (MDA). This implies that there is a data-model at the centre of an information system of which all other functions are derived, ideally resulting in just one data-model in which all datasets are harmoniously proportioned. [41]

Also within Shell Exploration & Production (Shell EP) there is the need for harmonised data management. Shell EP has set up a strategy for standardised geo-information handling [54 & 55]. It provides recommendations for the operating companies about the system architecture and standards to use. In this way it gives guidance to come to a local and regional GII at first and later to a worldwide corporate GII.

Shell EP Europe, the operating company of Shell for European countries, has made a start with this and is actively trying to bring the various GISs and geographic databases together. One example of this is the harmonisation of the offshore point dataset. This harmonisation implies both concordance between the different formats of the datasets involved to allow a consistent infrastructure, and a merge from the datasets held at the different locations and departments involved, to come to a common data- and workflow. The harmonisation of the Shell EP Europe's offshore point datasets is composed to form the project "aligning offshore infrastructure data-models". Commonly at the departments involved this project is referred to by the name of its data-model: Fixed And Mobile Entities version 2 (FAME-2).

Within the FAME-2 project many datasets are involved, varying in usage, disjunction, level of digitisation, number of stakeholders involved, etc. To reduce complexity the two most distinct and complex datasets are separated to form a different, but parallel project. These datasets are the geotechnical sites & the footprints.

Geotechnical data, in regard to Shell EP Europe's offshore activities, are seabed soil data stating the composition of the seabed. They are required for the foundation of installations.

Footprint data are derived from the imprints left by a Jack-up rig on the seabed. These data are required for they cause unevenness of the seabed, which should be avoided in the positioning of installations.

Shell EP Europe wants to have these datasets analysed and harmonised to the requirements of the corporate GII, and – if possible – to have them integrated into the FAME-2 data-model. In order to come to a harmonised solution for these geotechnical & footprint datasets, this research project was set up. In alignment with the Model Driven Architecture it aims to develop and to describe the harmonisation process for the involved datasets and data management. This combination leads to the following research question:

How can geotechnical & footprint data be harmonised to a data-model that fits into the corporate Geo-Information Infrastructure of Shell EP Europe?

The aim of this research is to develop a harmonised data-model for geotechnical & footprint data that can be implemented into the corporate GII. The aim for a harmonised data-model implies that there should be concordance between the different formats of the datasets involved to allow a consistent infrastructure, and a merge from the datasets held at the different locations and departments involved, to come to a common data- and workflow.

To research this problem, the harmonisation process is aligned to the development process. For each of the (development) process phase a research question can be posed, supporting the research of the harmonisation process of the geotechnical & footprint data. These questions and related process phases are:

1 Definition phase
What does the corporate GII for Shell EP Europe look like?

2 Information phase
Which geotechnical & footprint data and related data-models are used within Shell EP Europe?

3 Modelling phase
What should a new data-model look like?

4 Implementation phase
How can the new data-model be implemented into the corporate GII of Shell EP Europe?

5 Building phase & maintenance phase
What recommendations can be done to manage geotechnical & footprint data in the corporate GII of Shell EP Europe, using the new data-model?

As can be derived from the questions, the focus of this project is on the first phases of the development process. For the latter phases recommendations will be done, this is because these latter phases will be executed by Shell EP Europe itself. These first phases are also of importance for alignment with the Model Driven Approach, since in the MDA the data-model is regarded the axis of the information system. To come to that, these phases and research questions are handled by different research methods: desk research, literature study, work visits, interviews and questionnaire, modelling process and prototype evaluation.

The research questions are answered in the following chapters:

At first chapter 2 deals with Shell EP Europe and the corporate GII. It informs about the different views on GIIs in general, about the Shell Company, the departments involved and the way geo-information is managed within Shell EP and Shell EP Europe.

In chapter 3 the current geotechnical & footprint data management is explained. This is done by means of an inventory of the current datasets available at the different locations, by an analysis of the interviews and questionnaire concerning the users' wishes and by an analysis of the work- and dataflow of the geotechnical & footprint data.

Chapter 4 then describes the newly designed data-model for the geotechnical data and the footprint data.

In chapter 5 the implementation possibilities for the data-model are stated. It holds a special focus on the possibilities of using MS Visio and ArcCatalog to convert UML-models to actual database tables.

Chapter 6 describes the recommendations for the data management interface coming amongst others from the user testing of a prototype version.

Chapter 7 deals with the data migration plan and the recommendations for the system management.

Chapter 8 finalises by stating the conclusions and recommendations.

2 Geo-information management within Shell EP in Europe

It is often said [48] and in some cases even estimated [54], that of all information used, stored and circulated, approximately 80 percent has a geographic context. All this information that has a reference to a position on the globe, either via coordinates, address data or administrative areas, is called geographic (geo-) information [18]. This geo-information is mostly maintained in isolated data sets and accessible through disjoint Geo-Information Systems (GISs), built for a specific purpose.

In recent years the focus has shifted from managing several single geo-information datasets to harmonised management of interoperable GISs. This harmonisation endorses the sharing and accessibility of data throughout network systems, thus facilitating and supporting so-called geo-information infrastructures. Such a geo-information infrastructure (GII) provides a basic function for geographic data "discovery, evolution and application for users and providers within all levels" of the organisation [38].

This chapter is divided into three major parts. In section 2.1, geo-information infrastructures in general are dealt with. Information about Shell and NAM is provided in section 2.2, and in section 2.3 both elements are joined to describe the geo-information management in Shell EP Europe and thus answer the first research question: What does the corporate GII for Shell EP Europe look like.

2.1 Geo-information infrastructures

According to Hopkinson [22], Spatial Data Infrastructure (SDI) has been the 'buzz' word of the year 2003 and she states that it is "not an issue that will stop buzzing in near future". However already since 1994, the OpenGeospatial Consortium (OGC) has been developing and promoting interoperability standards to make spatial data available through any kind of (networked) system [48]. Though, what then is exactly meant with an SDI? That is what Dale [10] wonders in the same edition as Hopkinson: "It is all things to all people". He therefore compares the SDI concept to the fairytale's emperor's new cloths.

These examples make clear that the concept of a SDI, or GII as it is referred to as well, is still a relatively fresh concept of which not yet a single definition exists. Various descriptions, definitions and component settings are going around:

- "GII is a setup for the efficient distribution, integration, and exploitation of geographic information to enhance its availability, accessibility and use. It can be described as a set of institutional, technical, and economical arrangements to support the availability of relevant, up-to-date, and integrated geo-information" [5].
- A GII concerns the technology, policies, standards and human resources, necessary to acquire, process, store, distribute and improve the use of geo-information" [29].
- "Coordinated actions of nations and organisations to promote awareness and implementation of complimentary policies, common standards, and effective mechanisms

for the development of interoperable digital geographic data and technologies to support decision making at all scales for multiple purposes” [37].

- “The technology, policies, standards and human resources necessary to acquire, process, store, distribute and improve utilization of geospatial data” [15 in 37].
- “The relevant base collection of technologies, policies and institutional arrangements that facilitate the availability of and access to spatial data. The SDI provides a basis for spatial data discovery, evolution, and application for users and providers within all levels of government, the commercial sector, the non-profit sector, academia and by citizens in general” [38].
- “Geospatial Data Infrastructure encompasses the networked geospatial databases and data handling facilities, the complex of institutional, organizational, technological, human en economic resources which interact with one another and underpin the design, implementation, and maintenance of mechanisms facilitating the sharing, access to, and responsible use of geospatial data at an affordable cost for a specific application domain or enterprise” [18].

What all these definitions have in common is that it is all about facilitating the improvement of availability, accessibility and usability of geographic, spatial or geospatial data. Which in general terms are the same, and is data about the “geographic location and characteristics of natural or constructed features and boundaries on the earth” [15]. A working definition can then be that a SDI or

GII is a facility improvement for enhancing the availability, accessibility and usability of geo-information.

In this research project, the use of the term Geo-Information is preferred above the term Spatial Data. This is based on the infrastructure dealing with so-called ‘geo-referenced’ information, which is information related to a location in reference to the earth. The second reason is since the data are not the reason for creating an infrastructure; the reason is to gain information from the interoperable systems, which is more than simply data. Therefore the term GII has the preference above the term SDI.

Now that it is clear what a GII is intended to be, it is necessary to have the components set out clearly. Here as well, there are several compositions described:

- Organizational dimension; information dimension; technology dimension [37].
- Technology; Policy & Institutional Arrangements and Strategies; People [33].
- Data; Standards; Policies; Technologies; Procedures [43].
- Authentic geo-data sets; Geo-data processing services; Interoperability standards [44]; sometimes elaborated with (wireless) networks [45].
- Users of geo-information; Geo-information producers; Geo-data products; Marketplace for geo-information; Price settings; Political commitment; Interoperability framework; Policy; Forums [29].
- Geo-data; Technology; Standards; Policy and organisation [18 in 6].

Coming from the working definition as well, it is clear that a GII is about geo-information. And since information is derived from data [33] (see chapter 1 as well), the first component should be the (geo-) datasets, as described by Groot & McLaughlin [18 in 6]. These datasets should preferably be authentic datasets, since such authenticity safeguards the unambiguity of the data in respect to their content, data-model, quality and so on [46].

From the above listed components and from the fact that most data nowadays are in digital formats, the component technology [18 in 6] should be taken into account as well. This technology-components consists of the soft- and hardware, (wireless) networks and geo-data processing services, as described by Van Oosterom [46].

Furthermore the working definition is about the availability, accessibility and usability. Especially the latter is done for the users, who can be part of any organisation. Derived from the above components, the organisation will then be regarded the third component.

To improve availability and mostly accessibility a harmonisation of or interoperability between formats is required. This calls for a standardisation policy, of both the datasets, the technology to use and the organisational work- and dataflows is required. Therefore the concept of standards can be seen as the fourth component of a GII [46 & 18 in 6].

Compared to the listings of GII components above, the here mentioned components are comparable to the components as named by RAVI [18 in 6], with the regard that the term policy is dropped due to its ambiguity.

Although the definition and components of the GII are now cleared out, a few things need to be mentioned for the understanding of the concept of a GII. Ultimately the ideal for the concept of GII is to have all of the world's geo-information available and accessible from any system in the world. The main thought behind the GII-concept is that is should encompass more than a single dataset or database. [38] It therefore embraces the idea of sharing information through interoperable systems, to reduce duplication and increase efficiency [5, 37, 43 & 44]. Such interoperability is expressed as the "exchange of data in an open format" [22].

2.2 Shell & NAM; the business, the operating unit and the department

This section give a brief insight in the company, the business and the departments and units involved in this research. Unless stated otherwise, the information in this section is based on the various Shell websites [57] concerning these topics.

2.2.1 Royal Dutch / Shell Group

The Royal Dutch / Shell Group of Companies is an alliance between the two parent companies N.V. Koninklijke Nederlandse Petroleum Maatschappij and the 'Shell' Transport and Trading Company plc. This alliance started in 1907 when both companies decided to join. The two parent companies do not engage in operational activities, but hold the shares in the Royal Dutch / Shell Group (figure 2.1).

Until recently the parent companies were two separate companies with both an interest in the group. This double ownership caused unclear responsibilities resulting in a proposal for reorganisation of the group's structure in fall of 2004. In this the N.V. Koninklijke Nederlandse

Petroleum Maatschappij and the 'Shell' Transport & Travelling Company plc will unify under a one new parent company called 'Royal Dutch Shell plc' (figure 2.2). In June 2005, the Annual General Meeting will decide upon the envisaged change in structure.

The new company will be incorporated in the UK and tax resident in the Netherlands, as well as that the 60:40-balance of the interests in the group will be reflected by a conversion of the shares, where the total amount of N.V. Koninklijke Nederlandse Petroleum Maatschappij's shares will be converted to 60% of the new company's shares and the total amount of the 'Shell' Transport & Trading Company plc.'s shares will be converted to 40% of the new company's shares.

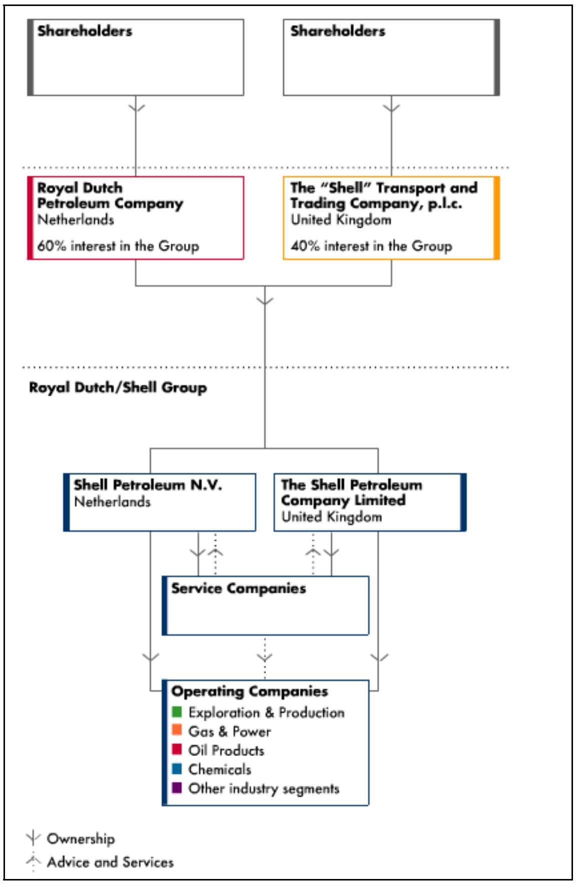


Figure 2.1: Structure of the Royal Dutch / Shell Group of Companies [52]

Within the Royal Dutch / Shell Group there are a numerous group of companies. These are divided into service companies, operating companies and the two holding companies: Shell Petroleum N.V. and Shell Petroleum Company Limited. These holding companies hold the interests in the service and operating companies. The service companies provide services and advise to the companies within the group. The operating companies execute the actual energy related operations. The operating companies are dived into five businesses:

- Exploration and Production
- Gas and Power
- Oil Products
- Chemicals
- Other Industry segments, a.o. Renewables

Each branch of the operating companies has its own responsibilities as is it a stand-alone enterprise, though able to use the experiences of the other companies within the Royal Dutch / Shell Group.

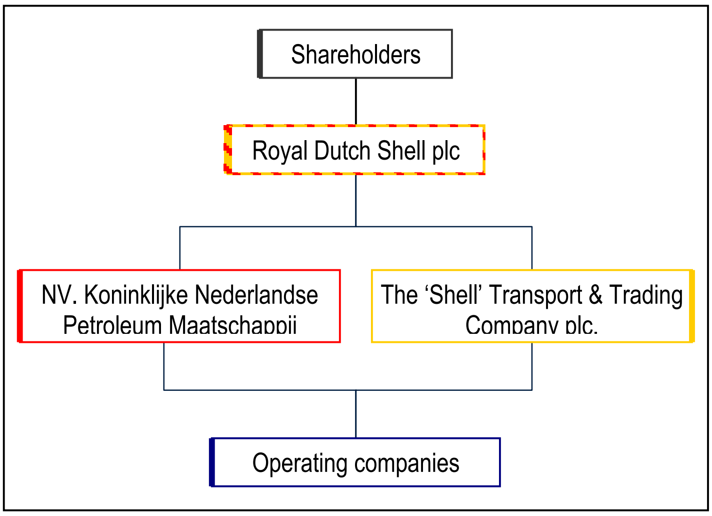


Figure 2.2: Reorganisation of the Royal Dutch / Shell Group of Companies [57]

2.2.2 NAM as part of Shell Exploration & Production

Shell Exploration and Production (EP) has activities in over 40 countries all over the world. It is responsible for a core activity of the Royal Dutch / Shell Group of Companies. Within the oil and gas industry this is referred to as upstream business and contains amongst others the exploration, production and transport of hydrocarbons (oil and gas). Downstream business, on the other hand, is the refining and sale of oil and gas products.

At the start of 2004, the internal structure of Shell EP was reorganised. From having national based operating units, the organisational structure changed into a global business divided in five regions (figure 2.3), all operating along global processes. The intended result will be that people will "work much more effectively together across the world by standardising processes, sharing learning, eliminating duplication, focussing recourses and speeding decisions" [51 in 4].

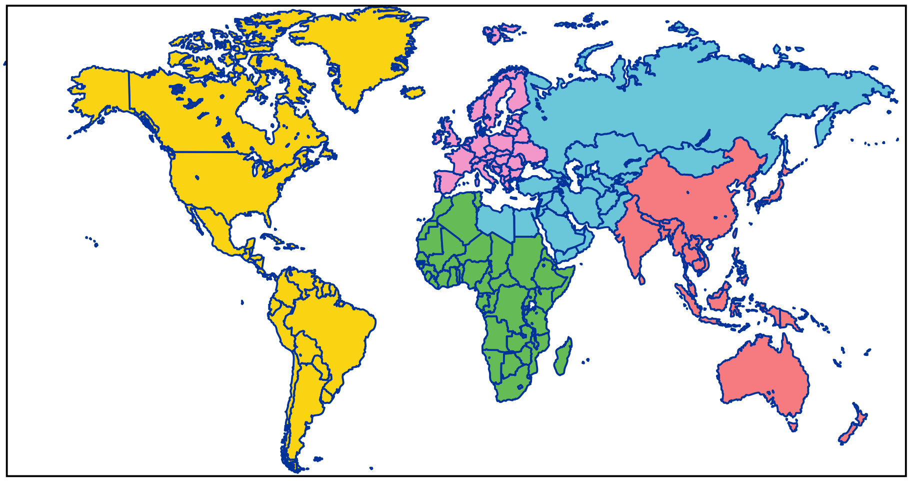


Figure 2.3: Regional division of Shell Exploration & Production [57]

One of the national operating units is NAM, Nederlandse Aardolie Maatschappij B.V. It was founded in 1947 as a joint venture of Shell and Esso (nowadays Exxon Mobile) to explore and produce gas and oil, firstly onshore in the Netherlands and later offshore as well. Due to the reorganisation in Shell EP, NAM is now part of Shell EP's European organisation (Shell EP Europe). Within EP Europe the main players are:

- NAM in Assen, managing gas operations, operations and operations in the southern North Sea.
- Shell Expro in Aberdeen, managing production and technical operations in the central North Sea and the Atlantic Margin.
- Norske Shell in Risavika near Stavanger, managing exploration and operations in the northern North Sea area.

Shell EP Europe holds also interests in 5 other countries: These countries involved are: Austria, Denmark, Germany, Ireland, and Italy.

2.2.3 Geo-Information Management within Data Acquisition & Management

Within the European departments of Shell EP, the department focussing on geographic-related information is EPE Data Acquisition and Management (E-DAM). Within the department there are two expertise domains: Sub-surface data management and Geomatics. The latter exists of the teams: Onshore Surveys, Offshore Surveys and Geo-Information Management.

The Geo-Information Management team is responsible for the geographic component of the Shell EP business information. It is both geographic data provider and consultant: providing for instance data access, cartographic services and geodetic advice concerning coordinate reference systems. In the workflow of the users of geo-information, they either use data directly from the corporate databases, or copied into personal or project databases. Therefore it is in the team's responsibility to watch over the quality and integrity of the corporate databases and thus safeguard the quality and integrity of the data.

2.3 A GII for Shell EP Europe

In the past years the focus of geographic information management changed from managing several local databases to combining these databases into a harmonised information collection. In governmental and public administration work field this is poured into so-called national and regional geo-information infrastructures. In the field of enterprises is spoken of corporate geo-information infrastructures [18]. These geo-information infrastructures (GIIs) include all policy, organisational, technical, legal and financial arrangement needed to support a standardised access to the geographic data [20]. Benefits of having a GII are amongst others the interoperability, avoidance of duplicate data and prevention of data inconsistency.

According to the definition of a GII, it has effects on the facilities already available. This section describes the available facilities within Shell EP Europe that can assemble a corporate GII. This is described by the four components of a GII: Datasets, Technology, Organisation and Standards [18 in 6].

Datasets

Within Shell EP Europe the department GIM is responsible for the geographic components of all types of data, as is explained in section 2.2.3. These data are various. They differ from data concerning the location of a single point object to the boundaries of three-dimensional basins. Currently these data are stored in corporate databases, personal and project databases. The data stored in the corporate databases are regarded as accurate, reliable and extra checked, comparable to the authentic datasets as described by Van Oosterom [46]. From these data map layers are constructed in ArcGIS or in ArcIMS. In the personal or project databases these map layers are mainly stored as geodatabases, used for specific purposes. Some of the data that are stored in either the corporate or the project databases have links referring to data stored in other systems, such as document-management systems (Figure 2.7).

Technology

Since October 2004 Shell has a global contract with ESRI for the use of ESRI GIS-products within the Shell Group. This implies a way forward for the use of GIS, both on web and on desktop within the company. This is supported by the global network consisting of interconnected local servers and local networks that are used throughout Shell. Off course this is a restricted network surrounding.

From the data management side, the data are stored in databases, mainly Oracle, both spatial and relational, however other database programmes, such as MS Access, are used as well. These databases are linked to each other, directly or by copy-management procedure, and to other corporate systems. An example of how geographic platform data are linked to other platform data can be seen in figure 2.4. Some of non-geographic corporate databases worth mentioning are: SAP, for the asset, financial and project data and LiveLink, a document management system for storage and retrieval of documents. An overview of the recommended architecture concerning the geo-information management is given in figure 2.8.

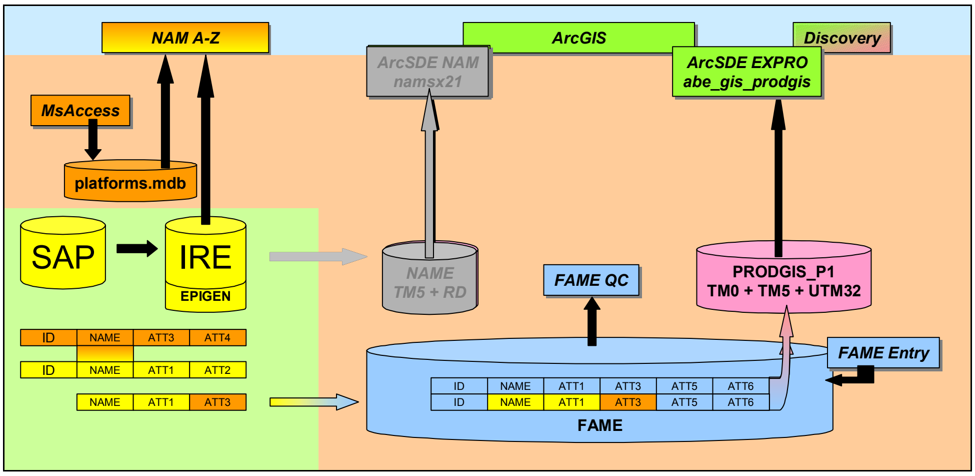


Figure 2.4: Platform data, interrelated between FAME and other systems [11]

Organisation

The departments dealing with the geo-information within Shell are mostly mentioned in section 2.2.3. However, since the access to and availability of GIS on desktop and especially on web are open for all, there are many more users of geo-information spread through all of the company. In a survey done it came out that there are more than 1000 GIS users within Shell worldwide. Keeping such widespread GIS access in mind, it might be worthwhile to compare the corporate GII efforts of such a company as Shell not to a simple corporate GII, but rather to a national or regional GII. See figure 2.5 for a differentiation between the levels of a GII.

In this case of the geotechnical & footprint data, there are less users and locations involved. Though the ones involved are still widespread, both by geographic location, by department and intended use. An overview of the departments and locations involved is to be found in figure 2.6.

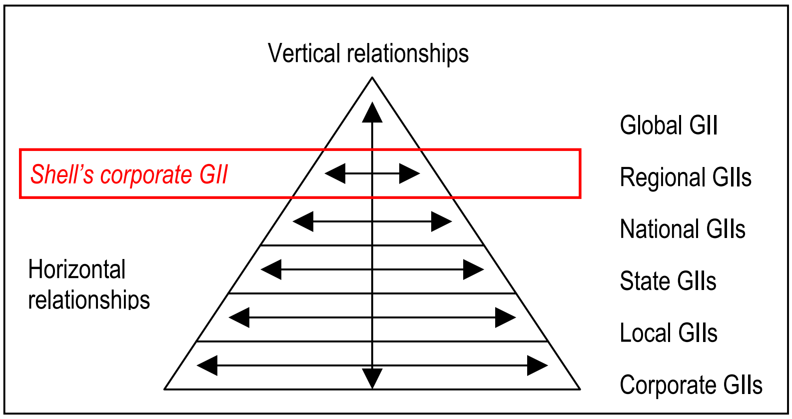


Figure 2.5: Pyramid of differentiation levels of GIIs [50]

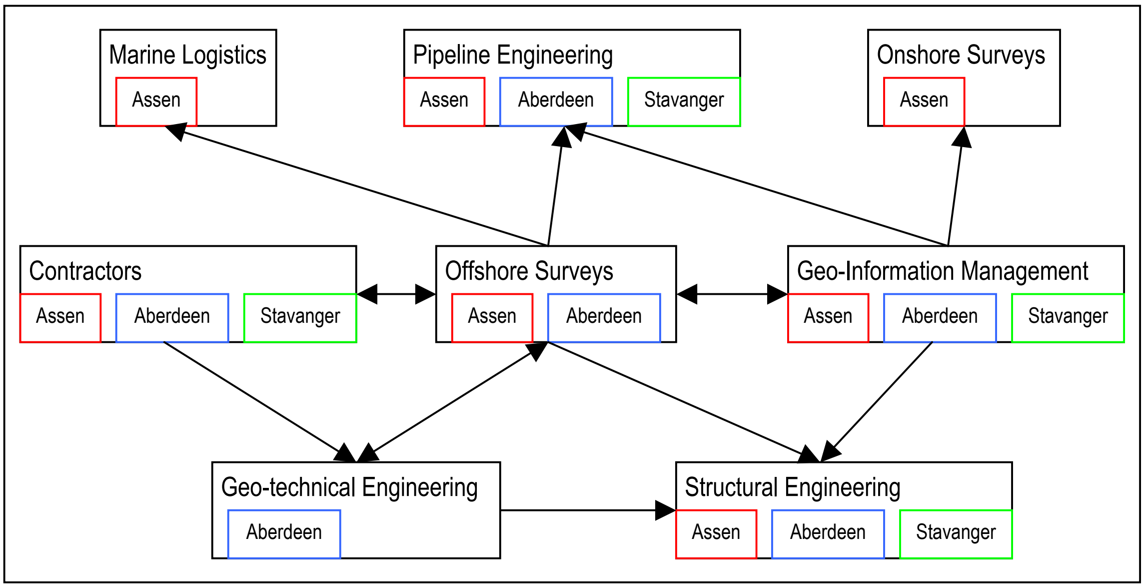


Figure 2.6: Departments and locations involved in geotechnical & footprint dataflow

Standards

Standards usually deal with how to handle the data, how to set up accessibility and how to organise work procedures. Major players in standardisation in the geo-information work field are the International Standards Organisation (ISO) and the OpenGeospatial Consortium (OGC). Also the European Petroleum Survey Group (EPSG) brings forth standards for the usage of coordinate reference systems in the oil industry.

All there different kinds of standards can be divided into several types. From organisational point of view a division into internal and external standards can be made. The first are the collection of standards set up by the organisation itself, as internal agreements on how to work; the latter are standards adopted from standardisation organisations such as the ones mentioned above.

Besides from organisational point of view also a division can be made into type of usage. Firstly a division can be made into low-level and high-level standards. The first are important for the technology: the interoperability of the computer systems and should provide basic infrastructure and hardware functionality. The high level standards primarily deal with the dataset, the database design, and data exchange and presentation [61 in 18]. Comparing these to the GII-components, one can conclude that there should then also be (external) standards, dealing with the organisational issues.

A second, less abstract, division divides standards into seven main categories, allowing each of them to hold more detailed specific standards. These main categories are [18]:

- Hardware and network standards
- System administration standards
- Software and application standards
- Data format standards
- Data compilation and update standards
- Product presentation standards
- System access and data/product distribution standards.

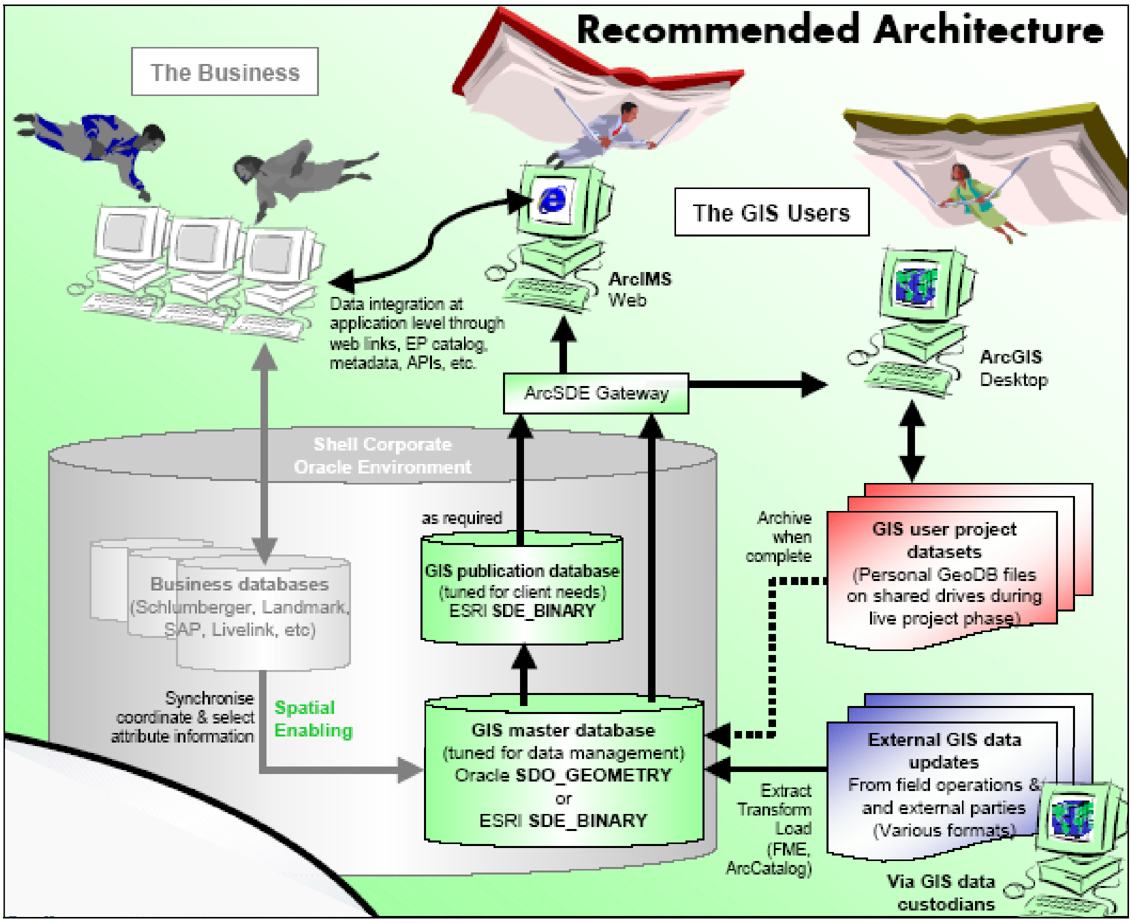


Figure 2.7: Recommended geo-information architecture by Shell Geo-Information Strategy [55]

Within Shell standards are seen as being part of the overall Geo-Information Strategy. This strategy is discussed and decided upon by the GIS Technical Advisory Panel, existing of GIS experts from all Shell offices. The highlights of the strategy are given in box 2.1, whereas figure 2.7 shows the recommended architecture. In the recommended architecture the geographic data are made available via GIS as well as the multimedia data that are interrelated with the geographic data. Worth to be mentioned are the use of Oracle as Database Management System (DBMS) and that Safe FME is recommended as the tool for extraction and transformation operations in geo-information management. ArcGIS is the GIS environment Shell has chosen as standard. Within Shell the upgrade to ArcGIS version 9.0 has just started. On the side of data visualisation has Shell created a standard legend to provide standardised symbology supporting the cartographic uniformity.

Not listed in the Geo-information strategy, but vital to the standardisation is the choice of Shell EP Europe to go to one standard coordinate reference system for the whole of the operating unit: WGS'84. The first system to be using this will be the recently developed FAME-2 system, in which offshore point data are stored and published. In the new data-model six fields describe the coordinate data:

- Two fields for the converted coordinates in WGS'84
- Two fields for the source coordinates in planar projection
- One field for the source coordinate reference system, given in EPSG-coding
- One field for the EPSG coordinate transformation code

If these mentioned standards used by Shell are then compared to the standard types listed above, the results will be the listing provided in box 2.2.

Box 2.1: Shell Geo-Information Strategy, derived from and based on [55]

Shell's Geo-Information Strategy is based on the storage and management of geo-information using industry-standard GIS, database and IT technology. Shell has strategic partnerships and global licensing arrangements with key vendors including Oracle Corporation, ESRI, and Safe Software.

- *Oracle* is Shell's corporate database standard and a world leader in database technology. It also provides geospatial data types (SDO_GEOMETRY), which are recommended as master GIS data store if circumstances require (see Recommended Architecture).
- *ESRI* supplies Shell's chosen GIS technology and a de-facto standard in the EP industry. ESRI's extensive product suite includes *ArcGIS* (desktop GIS), ArcIMS (web-enabled GIS) & ArcSDE (geodatabase and gateway).
- *Safe Software* provides *FME* (Feature Manipulation Engine), a tool for Extract-Transform-Load operations on GIS data, supporting all common GIS and spatial (database) formats.

On top of the above tools, the GIS TAP also recommends and monitors Shell-internal compliance to the following standards:

- *Shell Global Infrastructure* compliance is mandatory for GIS software deployment. The managed Windows environment (GI-D) requires all applications to be scripted for push-deployment.
- *Shell Standard Legend* (SSL) for cartographic symbology.
- *ISO* standard for GIS metadata (*ISO 19115*).
- *OpenGeospatial Consortium* (OGC – Shell is a member) standards for web and catalogue services.
- *European Petroleum Survey Group* (EPSG) standards for coordinate reference system definitions and transformations.

Conclusion

The above-explained components of a GII are related to the Shell structure of dealing with geographic information. These can be regarded as the conditions for a corporate geo-information infrastructure for Shell. In their turn these conditions set as well preconditions for the harmonisation of datasets within. Also the above-described components of a corporate GII for Shell then set both constraints and possibilities for the design and implementation of a new data-model.

To have these constraints and possibilities clarified more, especially for the geotechnical and footprint data, chapter 3 gives an overview of the datasets involved in order to have them give an insight in the conditions set by the data themselves, the data- and workflow, and the data-models currently used in Shell EP Europe.

Box 2.2: Shell EP Europe's standards divided by standard type, based on [18]

Technology (low-level) standards:

- Hardware and network standards
  - Architecture / Technology standards (see figure 2.7)
  - Shell Global Infrastructure (GI-D) (box 2.1)
- System administration standards
  - Map-layer standard templates
  - Standardised folder structure
- Software and application standards
  - Oracle (box 2.1)
  - ESRI's ArcGIS, ArcIMS, ArcSDE (box 2.1)
  - Shell Global Infrastructure (GI-D) (box 2.1)
  - Safe Software's FME (box 2.1)
  - OpenGeospatial Consortium (OGC) (box 2.1)

Dataset (high-level) standards:

- Data format standards
  - Coordinate reference system standard:
    - European Petroleum Survey Group (EPSG) (box 2.1)
    - WGS' 84 (as described above)
  - ISO 19115 (box 2.1)
  - Safe Software's FME (box 2.1)
  - OpenGeospatial Consortium (OGC) (box 2.1)
- Data compilation and update standards
  - Safe Software's FME (box 2.1)
  - OpenGeospatial Consortium (OGC) (box 2.1)
- Product presentation standards
  - Shell Standard Legend (SSL) (box 2.1)
  - Visualisation standards (standard map templates)

Organisational standards:

- System access and data/product distribution standards
  - Shell Global Infrastructure (GI-D) (box 2.1)

3 Inventory of geotechnical & footprint datasets

This chapter gives an overview and analysis of the geotechnical & footprint data and the underlying data-models that at the time of the project were in use in Shell EP Europe. Also it provides an insight in the data- and workflow concerning these datasets.

The geotechnical & footprint data are maintained in diffuse and diverse databases, spreadsheets, paper and digital reports at the various locations. They all are point data with associated attributes, documents and hyperlinks. The inventory and overview of the current situation, current data management and users' demands is obtained through the following data gathering methods:

- Interviews with the stakeholders
- Questionnaire amongst the stakeholders
- System analysis
- Data analysis
- Listing table from previous stakeholder engagement session

The outcomes of this information gathering are given in appendix C. In this chapter these outcomes are analysed to form an inventory. A distinction is made between the analysis and inventory datasets in section 3.1, the data formats and usage in section 3.2 and the desired data attributes in section 3.3 and the current data-models in section 3.4.

3.1 Available geotechnical & footprint datasets

Geotechnical data are seabed soil data stating the composition of the seabed. Geotechnical data are needed for the installation of platforms, pipelines and other installations, as well as for the positioning of jack-up rigs. They are required to determine the fundaments and possible subsidence of the structures. Examples of geotechnical data, in scanned documents, can be found in appendix B.1, figures B.1.4a-d.

Within Shell EP Europe approximately 4000 points are registered, roughly half on the UK side and half on the Dutch side. Every year approximately 30 new points are added to the existing dataset.

Figure 3.1: A Jack-up rig

Footprint data, or spudcan footprint data, are derived from the imprints left by a jack-up rig on the seabed. A jack-up rig is a floatable drilling platform, see figure 3.1. Such a rig is dragged into place and towed into its position. If on the right position it jacks itself up on its legs. These legs leave footprints on the seafloor, which can be up to 15 metres deep. (For an example see figure 3.2) The foot of such a rig leg is called a spudcan, hence also the name spudcan footprints.

If a rig would be placed too near to a set of footprints, the rig could slide into one of these holes, subside and sink.

The data collected on these footprints are besides the depth and position of the imprint, also information about the scour penetration, which is the erosion of the seabed around the rigs legs and the actions taken to prevent or stabilise it.

These jack-up rigs can only be operated in waters with a depth up to 100 metres. In deeper waters, such as the northern North Sea, semi-submersible rigs are used that are anchored to the seabed and leave no imprints.

Within Shell EP Europe roughly 1200 points are registered, of which 800 single footprint points on the UK side and 400 footprint sets on the Dutch side. Every year approximately 80 new single footprint points are added to the existing dataset. A note should be made that these single footprint points always are part of the set of footprints a jack-up rig has left behind, they thus comes in sets of three, four or six single footprints, depending on the number of legs a jack-up rig got.

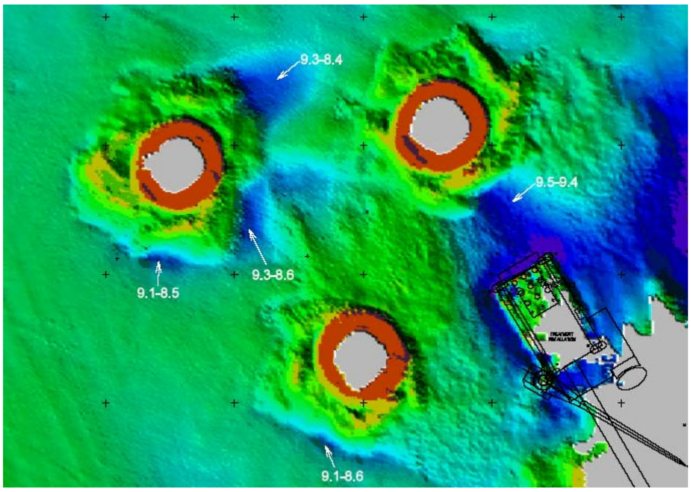


Figure 3.2: Bathymetry impression of footprints

The geotechnical & footprint data that are currently stored in Shell EP Europe are maintained in various diverse data formats at the various locations. To get an insight in what is stored at the different departments both data analysis was done and several interviews and work visits were held. The data gathering from both kinds of resources resulted in a data inventory (see tables 3.1 and 3.2), while the basic outcomes of the interviews can be found in appendix C.1.

Geotechnical data

The geotechnical data inventory (table 3.1) shows three geotechnical datasets in use in the United Kingdom, while in the Netherlands there is one dataset, and in Norway the data are stored within the greater project dataset that they are part of. One of the British datasets, the Excel-files with data from the South North Sea, has an overlap with the Dutch dataset. For all datasets it can be concluded, that some data already is in digital format, but most of the data do not have the right links to the related scanned reports, or do only have links to paper reports. And probably there is still more geotechnical information in the paper reports that reside in the archive.

Table 3.1: Shell EP Europe's geotechnical datasets

Shell Expro - United Kingdom	Dataset for UK continental shelf is maintained in: ▪ Excel-files, divided in Northern (NNS), Central (CNS) & Southern (SNS) North Sea (see figure B.1.1) ▪ ArcGIS layer based on MS Access database (figure B.1.3) ▪ OpenWorks (older dataset). Data flowchart with regard to the relationship lineage of the datasets can be seen in figure 3.5
NAM – The Netherlands	Dataset for Dutch continental shelf and UK SNS, stored in flat Oracle tables and, via copy management procedures, viewable in ArcGIS (figure B.1.2). With a hyperlink to borehole log (figures B.1.4a-d.) stored in LiveLink.
Norske Shell - Norway	There is no stand-alone management of the geotechnical data. The data are stored with the project data that they were acquired for.
EU-Seased	EU-Seased: European website for "seabed samples from the ocean basins and European continental waters […] held at European institutions", a geotechnical data metadata-portal, with links to the owners. These data will be used as 3rd party background layer [14]. (figure 3.3)

Footprint data

As the inventory in table 3.2 shows for footprints, less datasets are involved compared to geotechnical data. However, the data stored are stored as different entities in the different locations. In the United Kingdom the footprint data are stored in one dataset as single footprints. In the Netherlands there are two datasets: the reports of the rig positioning and a dataset in which the imprints of one rig on one position are stored as a set: one set per rig position. Norway does not have any footprint dataset, due to the water of the northern North Sea being too deep to use jack-up rigs. Furthermore there has been a stakeholder engagement session approximately 1.5 years ago, which resulted in a table listing possible and desired attributes for the storage of footprint data. This listing from this previous stakeholder engagement session can be regarded a first attempt for harmonisation.

Table 3.2: Shell EP Europe's footprint datasets

Shell Expro - United Kingdom	Single footprint points, stored in Oracle and available via copy management procedures as ArcGIS layer. For examples see figures B.2.3a-c.
NAM – The Netherlands	▪ NAM scour-penetration dataset, stored in MS Access database and accessible through intranet in a table-view (figure B.2.1). No geometries available. ▪ NAM assessment reports for Jack-up rigs in analogue and digital format, available in archive and LiveLink.
Norske Shell - Norway	No Norwegian dataset for footprints.
Stakeholders	Stakeholders requirements table from engagement sessions, based on stakeholders engagement sessions about 1.5 years ago. This table is not a used dataset, but at that time proposed, but not implemented. It can be found in table B.2.1 [1].

EU-Seased website

From the interviews the website of EU-Seased came up to be a possible source of third party (external) geotechnical data. The stated website acts as a metadata portal for European marine sedimentological information. It is a central server that has metadata stored about amongst others the gathering method and the data owner. In contrast to a clearinghouse, in which the client queries the metadata for availability of and access to the data [18], this website does not give digital access to the data themselves. Instead the website offers the metadata in which is stated at which organisation the data can be acquired.

This website is a merge of three meta-database-projects in the field of sedimentology, funded by the European Commission:

- Eurocore: holding metadata about seabed samples from the ocean basins, held at European institution, universities and marine stations.
- EUMarSIN: European Marine Sediment Information Network on the Internet, containing marine sediment meta-databases of the Geological Surveys of the EU-countries and Norway.
- EuroSeismic: European Marine Seismic Metadata and Information Centre. [14]

Within the EU-Seased website, a differentiation is made between seismic data on one hand (EuroSeismic) and seafloor data on the other hand (Eurocore & EUMarSIN). The latter entry, holding metadata about the "seabed samples from the ocean basins and European continental waters […] held at European institutions" [14] is of importance here. Within the metadata portal the metadata can be queried alphanumerically or by geographic location (figure 3.3).

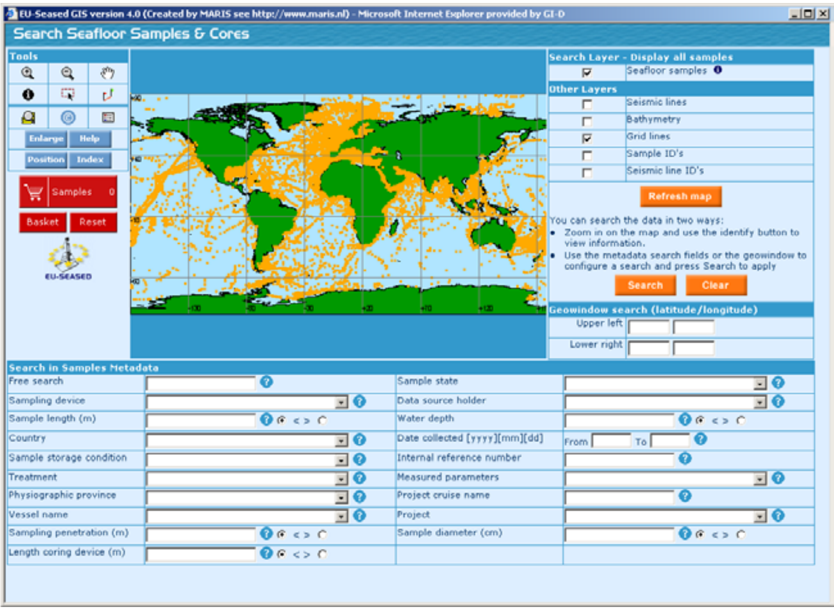


Figure 3.3: EU-Seased metadata-portal for search by geographic location [14]

Currently no use is made of this (meta-) dataset. However from the interviews it became clear that there is a desire to make use of the 3rd party (=non-Shell) geotechnical data (table C.1.2; person # 14). The users would like to have these 3rd party data stored and available for querying within the Shell-system. A possibility for implementation is to have a certain selection of the metadata that is available on the EU-Seased-website visualised as ArcGIS-layer and stored together with the Shell-owned geotechnical data.

3.2 Geotechnical & footprint data formats and usage

As already stated above, the current data are stored in a variety of different formats. To be able to form a harmony between the datasets involved and to make a basis for the later described migration plan, also an inventory is made of the different formats the datasets are currently stored and used in. Furthermore, for setting up a working data-model it is necessary to know how the data are queried, what these queries are and how users would like to have access to the data. Therefore the dataflow and workflow are described as well. The inventories stated here are the results of the analysis of the datasets and the interviews. The detailed analysis of the latter ones can be found in respectively table C.1.1 and C.2.1.

Data formats

The data formats that the datasets currently are stored in are already briefly mentioned in tables 3.1 & 3.2. In a short summary these are given as in table 3.3. All these different formats have to be harmonised and poured into the format as proposed by the Shell Geo-Information Strategy, which is explained in section 2.3.

Table 3.3: Current formats of geotechnical & footprint datasets in Shell EP Europe

Geotechnical data formats	Footprint data formats
▪ MS Excel ▪ MS Access ▪ OpenWorks ▪ Oracle flat tables ▪ Paper reports ▪ Scanned reports	▪ MS Access ▪ Oracle flat tables ▪ Paper reports ▪ Scanned reports

As in table 3.1 can be seen, there is a striking diversity of the formats of the geotechnical data in the UK. From the interviews held (table C.1.1), it became clear that the different geotechnical datasets are interrelated and descend from the same sources. This is expressed in figure 3.4: The original geotechnical data are coming from paper reports that are supplied, usually by contractors. Sometimes these reports are scanned and stored in LiveLink, the document-management system. The data from these paper reports used to be stored in Epigen, a database system. Some years ago, all data from this system was migrated to the OpenWorks database system, where the geotechnical data were stored as being a type of well data. At a certain point it was decided by geotechnical engineering that the way the data were stored in OpenWorks was not usable for their work, on which was decided to have all geotechnical data copied -once- to excel-files, which from that point onward were then directly updated from the paper reports. A bit later, there came the wish from Offshore Surveys to have the geotechnical data included in the Hazard Notification programme (see geotechnical & footprints usage). Therefore it needed to be converted to a map layer in ArcGIS, for which it was requested to have the data stored in an Oracle database. Thus the data were copied from OpenWorks to Oracle. More recently the updating of the data in the OpenWorks database has frozen and the Oracle database and GIS-layer has not been updated since.

A Consequence of this lineage of relationships is that combining these datasets will cause redundant entries. However due to the harmonisation request, a merge of these datasets is desired. How this should be dealt with is one of the data migration issues that are discussed in chapter 7.

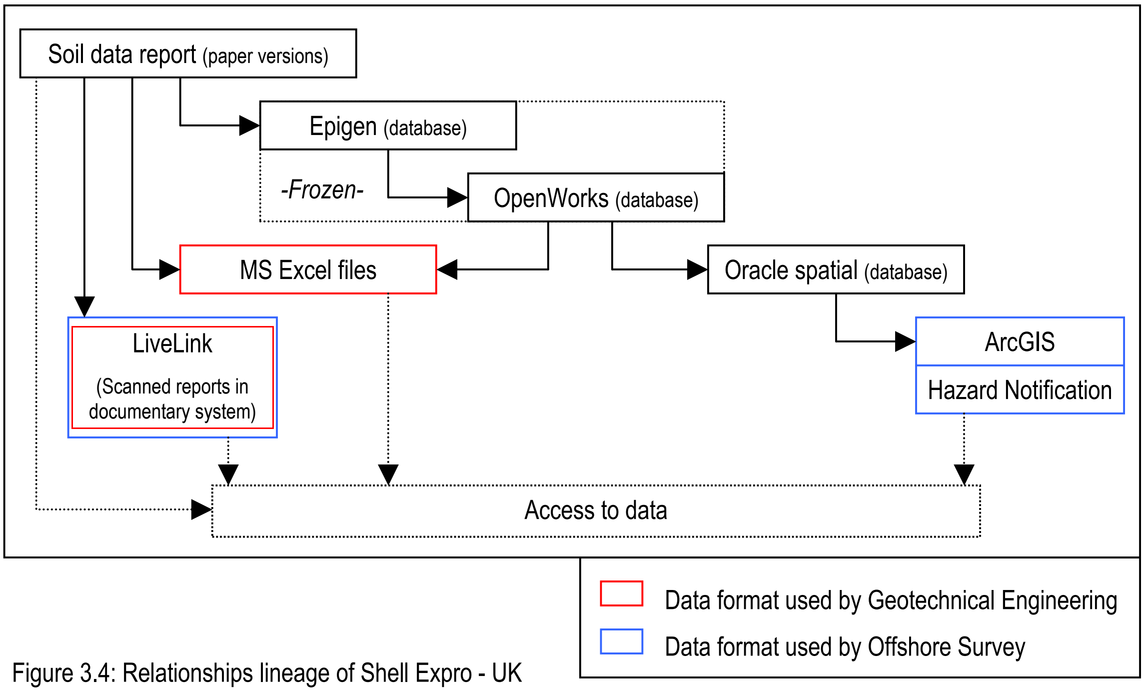


Figure 3.4: Relationships lineage of Shell Expro - UK geotechnical datasets

Geotechnical & footprint data usage

From the analysis of the interview answers regarding the workflow and dataflow of the geotechnical & footprint data, an indirect link between the geotechnical data and the footprint data appeared to exits (figure 3.5). The geotechnical data are used in the process of positioning installations, such as jack-up rigs. During the planning stage an estimate is made of the depth of the imprints on the seabed, which is called the predicted leg penetration. When the rig is positioned and installed, the legs and feet (spudcans) will leave imprints on the seabed: the footprints. The actual depth of the footprints is called the actual penetration. Both the predicted penetration and the actual penetration are additional information to the geotechnical data: When another installation is positioned in the vicinity, the geotechnical data in combination with the predicted and actual penetration is used to form an updated composition of the seabed and leads to better future predictions.

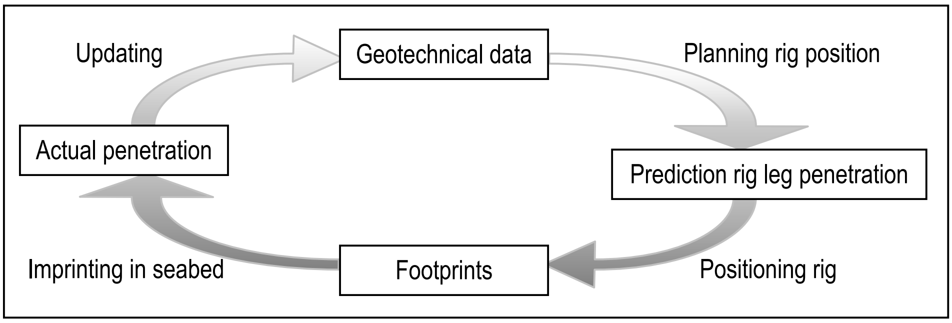


Figure 3.5: Data- and workflow relationship between geotechnical data and footprint data

The way the datasets are used in the workflow depends mainly on how the data are queried. In the workflow, the use of both geotechnical data and footprint data are quite basic and generally alike. The main usage is querying for visualisation of the information and for the location of the entities (geographic position). Primarily this comes to whether the data points are within a certain distance of an object in focus and if so, what the data-attributes are. Such queries are mainly done by means of a so-called Hazard Notification.

The Hazard Notification is a programme used to gather information for the planning of any offshore activity. The programme is custom build as querying add-on for ArcGIS. Its function is to automatically select all offshore data layers and analyse if there are any obstructions within a certain radius of the planned activity area. If so it should acquire the attribute data of these obstructions and write them to a report.

To start with, in the Hazard Notification programme an object or geographic position is to be selected. Around this position a buffer-query is created: every feature from all layers that is within the variable radius of the buffer is selected. Every selected feature set is written as an attribute table to a MS Word-document, including a map and an image-file. In this way this text document contains the attribute information of all entities within the given radius, which is an input for further planning.

When there are any geotechnical data points within the search radius, they are used as first impression information of the seabed soil. In the geotechnical data-attribute-table in the Hazard Notification document the links to the geotechnical reports can be found. With the existing geotechnical information and the information about the intended activity, acquisition of new geotechnical samples can be planned. From the existing and new geotechnical data a soil chart is created, which is used to give insight in the foundation needed for any new object.

When there are footprints within the search radius, their location is required to avoid in the positioning of a new object. Also the shape of the footprints in three dimensions is required, since footprints have an altering effect to the seabed surface. For example if a jack-up rig would be placed in or on the edge of a former set of footprints, it could subside, capsize or even sink.

Following from the described data usage, both geotechnical data and footprint data are used as indirect reference information for offshore activities. This causes certain indifference amongst the users to only use those data that are at hand and do not search for more. The users settle for the information coming from the Hazard Notification and from the paper reports that are at hand. They trust the data to be accurate, complete and up to date.

3.3 Current geotechnical & footprint data-models

As a harmonisation process differs from a development process starting from scratch, one of the main differences is the lineage and ancestry of datasets. The same can be noticed with regard to the data-models. These already existing data-models can be regarded a legacy and can be used as best practice or starting point for the new data-model.

The footprint data initially were taken into account in the first version of the FAME data-model. This first edition of FAME (FAME-1) was only created for Shell Expro, the British Shell office, and is expressed as UML diagram (appendix D.1). This FAME-1 data-model also includes that different coordinate reference systems can occur within one table. However, to visualise features via ArcGIS a uniform projection within the dataset is needed, therefore the dataset needed to be re-projected by Safe FME before publication.

Although there is an entry option for footprint data in the first version of FAME, these data were never actually stored in it. Still this FAME-1 data-model serves as a basis for the current FAME phase 2 (FAME-2) development, which is intended to store offshore point data for the whole of Shell EP Europe.

The geotechnical data have never been implemented into a separate and suited data-model, since they were not regarded of such importance. When stored in the Epigen and the OpenWorks environments the geotechnical data were treated as if they were similar to oil or gas well data and the attributes were made to fit into the attribute list of the well data. And thus no suitable previous data-model of the geotechnical data exists.

3.4 Desired geotechnical & footprint data

From the workflow and data analysis it became clear, that not all datasets and data-attributes were regarded as equally important and usable. Therefore the work-visits, interviews and questionnaire were used also to get an insight into the requirements for geotechnical & footprint data-attributes.

As seen in the previous section, the users only use the data that are at hand and do not actively seek out for more extensive or more accurate data. But although it may seem contradictory, the users certainly are interested in more extensive and more accurate data. This concludes to users desiring to have the data directly at hand, the more extensive and accurate the better, but data should not require much effort to obtain. A good solution for this is to have easily accessible harmonised datasets.

Onshore geotechnical data

The above described data analysis also led to the discovery of a set of onshore geotechnical data. From the interview the Onshore Surveys department leader, it came out that there certainly is a need for onshore geotechnical data. But there is only a need for Dutch onshore geotechnical data, since in Shell EP Europe onshore surveys are only done in the Netherlands.

At the Dutch government's organisation of TNO-NITG a very large and accurate database on onshore geotechnical data is kept. For the Onshore Surveys team it is cheaper and more convenient to acquire these data from the Dutch government when needed, instead of maintaining and managing a database themselves. This is already the case for the current workflow. Therefore this handful of Dutch onshore geotechnical data points have not need to be migrated to a new system and can be deleted.

Also in the UK datasets, there are 116 onshore geotechnical data points. The data originate from the British Geological Survey and were acquired a long time ago. The data points are quite old as well, ranging from 1897 till 1978. According to both the Onshore Survey team and Offshore Survey team there is no direct need for these data point are to have them transferred to a new system. However, since it concerns a whole, coherent and not too large dataset and it is unclear whether these data have users, it is recommendable to not discard the data and migrate them as part of the geotechnical dataset to the new environment.

Questionnaire

The actual data that the users say that is required is an almost infinite list. The most of these data are currently written in the reports, either digitally available or in paper. Having these reports digitised and correctly linked would be sufficient accessibility for most of the data. From the data that are more explicitly needed, a concept data-attribute listing is made (box 3.1). This listing has been sent to the users to give feedback upon in the form of a questionnaire.

The questionnaire asked of the questioned to state the importance of the data-attributes, to add data-attributes and to comment on their definitions. The importance of the data-attributes was ranked on a scale ranging 1 to 4, with 1 as most important ("highly important, without it dataset will be useless") and 4 as least important ("nice to have, but can do without"). This scaling is based on and analogue to the MoSCoW-hierarchy that is used in the Dynamic System Development Method (DSDM).

This MoSCoW-hierarchy is normally used as a tool in the DSDM development process, where it is used in combination with the concept of time boxing. The purpose of it is to set an action list for a given period, the time box. Within this action list a distinction is made between what is really important and need to be done and what is less important and can wait for a later iteration step or a later product version. In this MoSCoW is an acronym and stands for:

- Must have: Requirements that are fundamental to the system. Without this minimum usable subset, the system will be unworkable and useless.
- Should have: Important requirements, without which the system will still be useful and usable.
- Could have: Requirements that can more easily be left out, without harming the system.
- Want to have (but not will not have this time round): Valuable requirements that, if needed, can wait till a later development phase [58].

In this questionnaire the MoSCoW-hierarchy is used to obtain an insight in the importance of the data and to get possible obligatory fields for the data entry interface. In this the 'must haves' will then be the obligatory fields and the 'want to haves' can be left out in the design of the data-model.

The questionnaire was sent to 12 stakeholders who are direct users of the data (table C.1.1). Eighth of the stakeholders filled in the questionnaire, resulting in a 67% response and since the responses (appendix C.2) came from different departments and locations, the answers can be regarded as representative.

The outcome of the questionnaire gives an insight in what the users wishes concerning the data attributes are. Striking results are that the users care little about the attributes that are solely required for data management, e.g. UID or SSL codes. However this cannot be regarded as surprising or asking for attention. One of the results that does require attention, is the wish to explicitly see the coordinates as published attribute data, and not just as a point on a map. This has implications for the design of the user interface, where these coordinates thus should be presented

Conclusion

The current data-models for geotechnical & footprint data and the way the datasets are structured do not meet the requirements expressed by the users (section 3.3). Although they use the data available at hand, there is the desire to have improved the availability, accessibility and usability of the data. This can be derived by harmonising the datasets into a new data-model that is in line with the corporate GII. This data-model will the increase efficiency in the work- and dataflow and will reduce redundancy. How this model is created and implemented is described in the following chapters.

Box 3.1: Geotechnical & footprint data-attribute listings, based on table C.2.1

Geotechnical data-attributes:	Footprint set data-attributes:	Footprint data-attributes:
▪ UID ▪ Borehole name ▪ X coordinate ▪ Y coordinate ▪ EPSG code ▪ Boring type/ method ▪ Contractor ▪ Block code ▪ Date ▪ Depth (m) ▪ SSL ▪ Report-link ▪ Last updated ▪ Updated by ▪ Status ▪ Epigen code ▪ Water depth (m) ▪ Remarks ▪ Country name ▪ Platform name ▪ Well name	▪ Footprint Set UID ▪ Rig name ▪ Start date ▪ Block name ▪ End date ▪ Spudcan type ▪ SSL ▪ Well name ▪ Platform name ▪ Last updated ▪ Updated by ▪ Layout of spudcans ▪ Spudcan penetration survey ▪ Bathymetry data ▪ Remarks ▪ Anchor / jack-up charts ▪ Rig move reports ▪ Water depth ▪ JIM reports ▪ Rock gravel dump reports	▪ Footprint UID ▪ X coordinate ▪ Y coordinate ▪ EPSG code ▪ Spudcan UID ▪ Leg name ▪ Predicted penetration per leg (m) ▪ Penetration per leg (m) ▪ Spudcan shape (radius) ▪ Pre-loads ▪ Scour protection (tonnage) ▪ Date of scour protection ▪ Scour protection type ▪ Last updated ▪ Updated by ▪ Remarks

4 A new data-model for geotechnical & footprint data management

In this chapter the newly designed data-model for geotechnical & footprint data is described. From the analysis and inventory in chapter 3 it shows that there is a discrepancy between the dataset and dataflow on one hand and the users' and organisation's needs and wishes on the other hand. In the harmonisation process to come to a new data-model not just the data, but also the stakeholders' wishes and needs (section 3.3) and the boundary constraints set up by the Shell geo-information strategy (section 2.3) are taken into account.

During the project of harmonising the geotechnical & footprint datasets, it became clear, that the new geotechnical & footprint data-model was compatible to the also newly created FAME-2 data-model. This FAME-2 data-model takes in all other offshore point datasets (installations, features, buoys and rigs) and is based on the FAME-1 data-model. Because this compatibility, the geotechnical & footprint data-model is integrated with the FAME-2 data-model into one single data-model for all European offshore point datasets. This process and the results are described in section 4.4. Before that, section 4.1 states the needs for a new data-model as concluded from the usage described in chapter 3. Section 4.2 describes the model conditions that influence the design of a new data-model. The new data-model is structured according to the Unified Modeling Language (UML). An introduction to what UML is and how it can be used for the visualisation and structuring of data-models, is given in section 4.3. Finally section 4.5 gives the structure of the new data-model.

4.1 The need for a new data-model

In chapter 3 we noticed the discrepancy between the existing geotechnical & footprint datasets, and the current data-models not meeting the set requirements. Therefore the necessity of a new data-model for the geotechnical & footprint datasets and their data management was concluded. The various reasons for such a new model has multiple reasons, of which the main reasons are described in this section.

Firstly there is no existing data-model for the current combination of these two datasets. As is described in section 3.3, there is a current data-model for the footprint data for the British Shell office. However this data-model is not directly applicable for the counter-dataset of the Dutch footprint data. In addition to that, the geotechnical data have never had a real data-model, neither in the UK, nor in the Netherlands.

But what is then the actual added value of a data-model? According to Longley, Goodchild, Maguire & Rhind [31] a data-model is required as representation of selected aspects of the real world in the computer, since the reality is infinitely complex, while computers are finite. From that comes, that when using a data-model it is easier to retrieve the data and interpret them to extract information. By means of a data-model a list of attributes is defined for the data. This list is standardised for all data entries, which makes it easier for the user to make use of the dataset.

Having the reality and data combined in a data-model opens new possibilities in data and system management. The main concept behind a data-model centred type of data management comes from the Model Driven Architecture (MDA) that is being developed by the Object Management Group (OMG) [7]. The OMG is a non-profit, open-membership consortium that produces specifications and interoperability standards for the computer industry [42]. The most well known specifications of OMG are: CORBA and UML (see section 4.4), with now added to that is MDA. Underlying to the concept of MDA are four principles [7]:

- Well-defined models are the cornerstones to understanding complex systems
- Building a system can be organised around a set of models
- Integration of and transformation between models can be automated by tools
- Broad adaptation of models requires standards and openness

Within the MDA-concept there are different models at different levels of platform-independency. At the highest level there is the so-called Platform Independent Model (PIM), which is the "formal specification of the structure and function of a system that abstracts away technical details" [41]. This PIM is regarded the central axis of the system (figure 4.1). To have the PIM implemented it needs to be transformed to one or more Platform Specific Models (PSM). These are specified to the rules and terms of the platform's implementation technology, e.g. a database model. The last step is to transform the PSM to code that is the systems basic language [27]. In practice this means that all the databases, interfaces, copy-management procedures, and so on, can be derived from the central PIM. Both the PIM and the PSMs should be written in a well-defined language, to assure consistency and precision. Here the Unified Modeling Language (UML) comes in as a standardised solution for structuring and visualising the data-models.

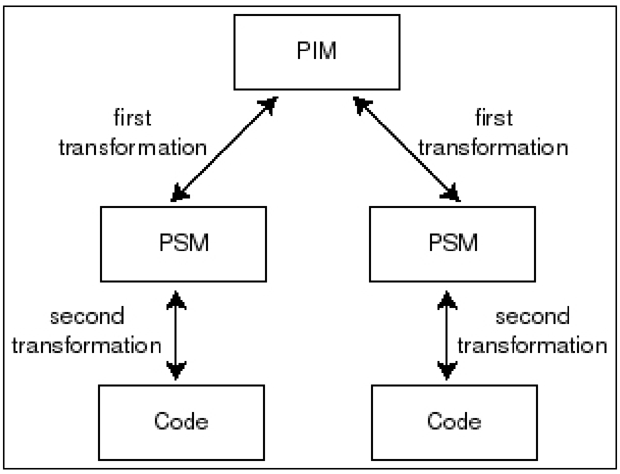


Figure 4.1: Model Driven Architecture (MDA) process steps [27]

4.2 Conditions regarding a new data-model

To be able to make a description of the new data-model, or of any of the modelling phases, it is necessary to know the conditions it has to comply to. These conditions are set by the different components of an information system. These components and their conditions are described and analysed in section 4.2.2. First in section 4.2.1 the modelling steps to derive a data-model from the real world are described. The combination of the information system's components and the modelling phases results in the description of the actual conditions in section 4.2.3.

4.2.1 Modelling steps to come from reality to a data-model

For designing a new data-model the set of raw data-attributes should be transmitted into a structured data-model. In the concept of MDA (Model Driven Architecture) the aim is to use so-called 'Object Oriented' modelling. This means that the data-model consists of objects that are representative for objects of the real world. To come to such an object oriented representation the real world is to be conceptualised into a set of objects (or entities) of interest [31] (figure 4.2). This set is what is sometimes referred to as the universe of discourse. This universe of discourse is then conceptualised into more abstract and more detailed models, resulting in a data-model that can be implemented into the system. This conceptualisation process is usually divided into several stages. Each stage results in a more detailed model.

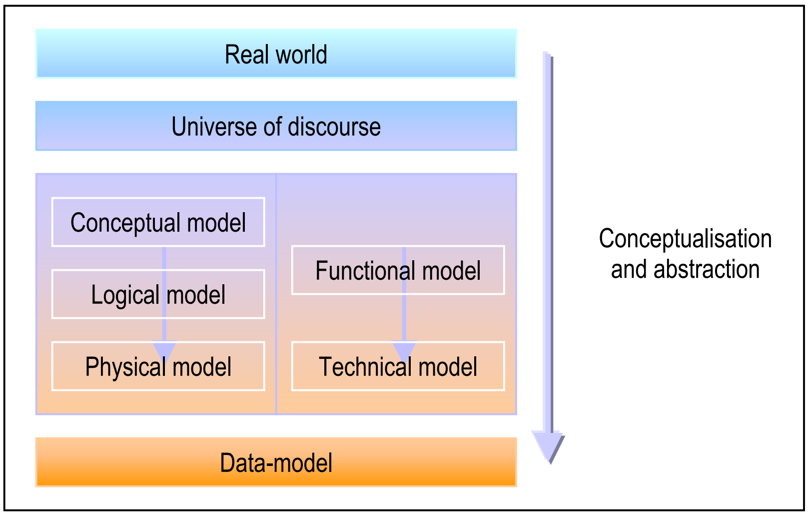


Figure 4.2: Modelling steps by level of conceptualisation and abstraction [31 & 58]

Commonly the modelling process is divided into three stages: the conceptual model, logical model and physical model [49] (figure 4.2). "The conceptual model is a human oriented, often partially structured, model of selected objects and processes that are thought relevant to a particular problem domain. The logical model is an implementation-oriented representation of reality that is often expressed in the form of diagrams and lists. The physical model portrays the actual application in a GIS, and often comprises tables stored as files or databases" [31]. Often the physical model is also referred to as technical model and usually in simple models, the modelling stages of the conceptual model and logical model are combined [49].

The other stages visualised in figure 4.2 are outcomes of the stages of the Dynamic System Development Method (DSDM). In this method usually two models are created, first a functional model as axis for further development, which is further on in the process translated to a more detailed model. The function model describes the entities themselves and their relation to each other. The more detailed model, goes into more detail and describes the attributes and the format they will be in, and is then the starting point for the implementation phase [58]. This more detailed model can thus be regarded to match the technical model, as is shown in figure 4.2

A different approach is set up by Bemelmans [3]. Instead of using the level of detail of the modelling level as guidance, he sets the stages by the aspect of the information (figure 4.3). The aspects used are pragmatics, semantics and syntax, coming from the field of semiotics, which is the study of the functions of signs and symbols [63]. Added to those the model is regarding the technical aspects to form the link to the technical system. These four stages combined form the process of semiotic modelling as proposed by Bemelmans [3].

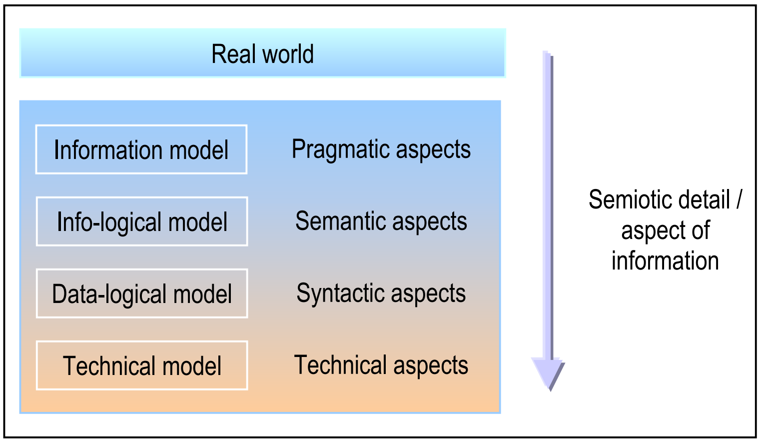


Figure 4.3: Information System modelling by semiotic aspect by Bemelmans [3]

Regarding the Model Driven Architecture and its possibilities (section 4.1), using that it results in using a modelling process based on level of abstraction. In the case of the geotechnical and footprint data-modelling process the choice has been made to follow such a line of abstract modelling, using a two stage process (conceptual model – technical model), since the complexity of the number of datasets is limited, and the MDA is thought to hold possibilities for the later development phases.

4.2.2 Components of an information system

In order to come to a balanced and structured new data-model the conditions for the new model should be clearly defined. These conditions can be described by the components of an information system that set them up.

Bemelmans [3] uses a distinction into four components of an information system (figure 4.4). When it is clear what the settings of and constraints on these components will be, the conditions for the overall information system will become clear. The components described by Bemelmans are:

- Data and knowledge base: the collection of data and knowledge
- Programme base: specific programmes and general-purpose models
- Man-machine-interaction component or dialogue component; existing of:
    - Database management system; meant for input, removal, alteration and querying of data
    - Model base management system; meant for adding, removal, alteration and use of models
    - Data generating management system: the action- en presentation language, with which user state commands and receives output.
- System interface: the programmes and hardware meant for interaction with other information systems.

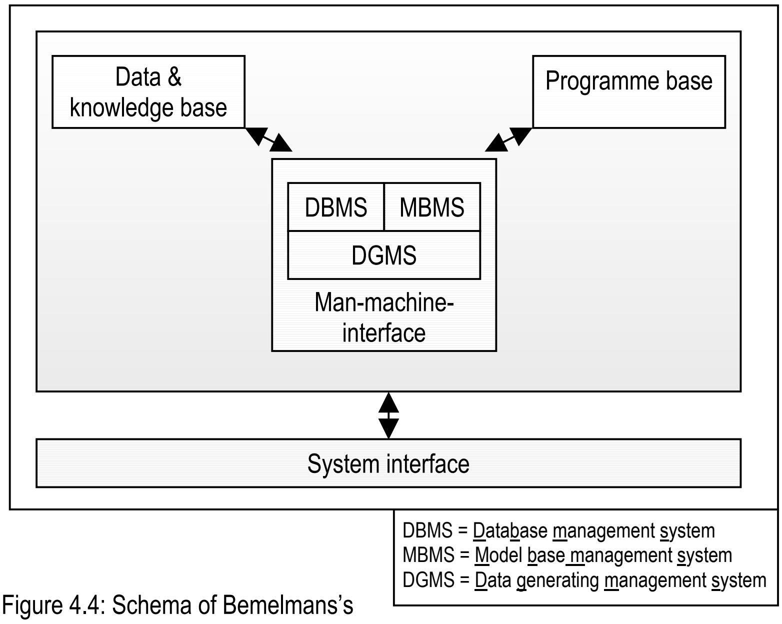


Figure 4.4: Schema of Bemelmans's model of information system components [3]

A different approach is used by Martin & Leben [35]. They have set up a so-called 'Information Systems Pyramid'. In this model the function of an information-system in an organisation is described by a pyramid (see figure 4.5). The sides of the pyramid are defined by the different components in the organisation:

- Data; that is used by the organisation in the information system.
- Activities; that the organisation carries out using the data.
- Technology; that is used in implementing information systems activities.

Whereas the horizontal layers correspond with the modelling phases described in section 4.2.1 and are defined by the level of detail and are:

- Strategy: overall strategic planning done by top management and information system executives
- Analysis: logical model about the fundamental data that are required.
- Design: detailed design of the data and all related issues.
- Construction: physical construction of databases and programmes.

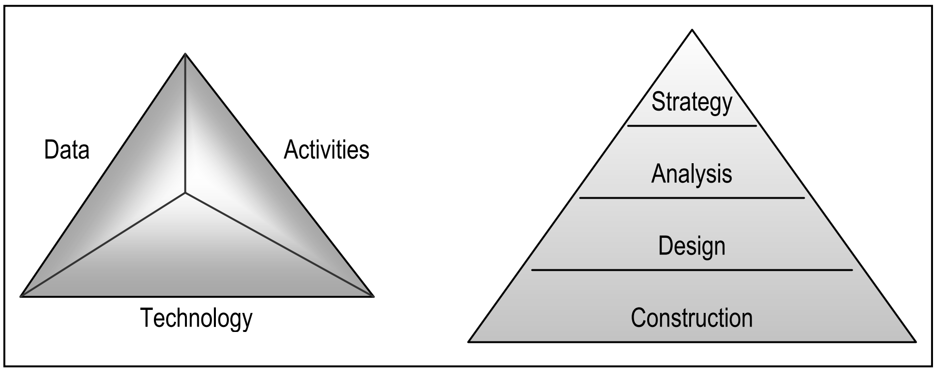


Figure 4.5: Martin & Leben's information systems pyramid, sides and layers [35]

A third way to analyse the conditions for a new data-model that the different components of an information system set up, is by looking at the components of the Geo-Information Infrastructure (GII). As also described in section 2.1, the components of a GII are, according to Groot & McLaughlin [18 in 6]:

- Geo-data
- Technology for storage, dissemination and use
- Standards for description, exchange and integration
- Policy and organisation.

A new differentiation of information system's components

All three above-mentioned models can be used to analyse the conditions on the data-model. These models described by Bemelmans, Martin & Leben and Groot & McLaughlin have many similarities. However neither of them can be used to make a linkage to the Model Driven Architecture (MDA). The main condition to analyse an information system's conditions in regard to the concept of MDA is having the data-model as centred component. It therefore is required to set up a new differentiation of information system components. This set of components is based on the differentiations described above, though having a special component guaranteeing the link to the concept of MDA. In order to come to this, table 4.1 structures the components of the above-mentioned models by similarities, to form new components that can be compatible within the MDA-concept. This new set of components can then be used to analyse the conditions set for the new data-model for the geotechnical & footprint data.

Table 4.1: comparison of analysis components

Bemelmans	Martin & Leben	Groot & McLaughlin / GII	MDA compatible differentiation
'Data & knowledge base'	'Data'	'Geo-data'	Data
'Man-machine-interface' component	'Activities'	'Policy and organisation' and part of 'Standards'	Users
Part of 'system interface', 'Programme base' and 'Man-machine-interface'	'Technology'	'Technology' and part of 'Standards'	Technology
'Programme base' component	Combination of the side 'Data' and the layer 'Analysis'	Part of 'Geo-data', 'Technology' and 'Standards'	Data-model

In this new differentiation an information system is composed of four interrelated components (figure 4.6). These components are derived from the three above mentioned models combined with desk research and a requirements analysis, and can be defined as:

- Data: data-attributes, dataflow, relations between the datasets
- Users: users' requirements, workflow, user interface
- Technology: software and hardware, network functionalities, programming languages
- Data-model: the part of reality modelled in the information system, in any stage or modelling phases (as described in section 4.2.1)

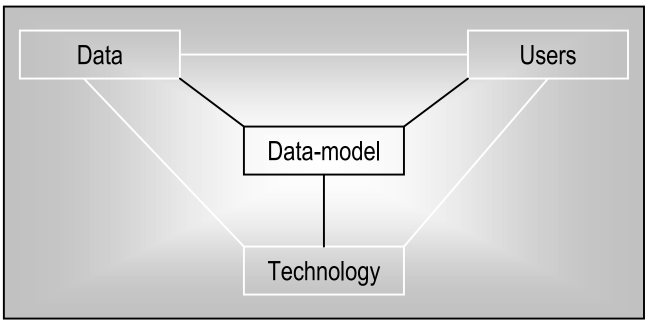


Figure 4.6: Information system's relations between data, data-model, user and technology

To come to a properly working information system, there should be harmony between the components (figure 4.6). This has the result, that if three of the four components are known, they set the conditions for the fourth component.

In the case of this research project: the users' wishes and demands are known (section 3.3), the data limitations are known (section 3.2) and the system's technical frame is known (section 2.3), and thus the conditions for the data-model are set and are described in section 4.2.3.

This harmonic proportioning of the components in the information system should then also be regarded when either of the components is altered or is implemented into the system. Thus if one of the four components needs to be implemented, the harmony with the other three should be restored. When in this project the data-model is to be implemented, a re-proportioning of the other components is needed. This is described in the next chapters:

- Implementation of 'data-model' to 'system' = configuration of the system (chapter 5)
- Implementation of 'data-model' to 'user' = user interface (chapter 6)
- Implementation of 'data-model' to 'data' = data migration (chapter 7)

4.2.3 Conditions set up by the information system's components

The actual conditions that are created by the information system's components depend on the mutual relations between the components. As said above, if the other components and the relations are known, the data-model component's conditions can be derived.

The conditions given by the data themselves are mainly the data-attributes. These are listed in box 3.1. Other conditions are coming from the relations to the other datasets, these are visualised in the conceptual model in figure 4.8. Two special issues worth mentioning concerning the footprint data are, that there is a constraint on the number of footprints belonging in one set. Since Jack-up rigs only have either three, four or six legs, there can only be three, four or six footprints left at one rig position. Secondly there is the assumption taken into the data-model, that all leg and spudcans of one rig are of the same type and diameter. This implies, that all footprints in a set are alike, except for their position.

The users' conditions regarding the data-model are expressed in the workflow description (section 3.2), the usability of the interface and the results from the questionnaire (appendix C.2). In the questionnaire-results, there is the users' wish (section 3.3) to see a certain set of data-attributes when using the interface (box 3.1). This set of data-attributes, however, differs from the data-attributes the data-model logically would have stored in the specific entity tables of the database. These latter are usually referred to as the technical specifications, which in turn relate to the abstraction level of the technical model (figure 4.2). This means that there is a gap between the data-attributes that are stored in the database tables and the data-attributes that should be displayed in the user-interface. To overcome this gap, the database tables need to be interrelated. By means these relationships, selections of different tables can be combined in order to display the data-attributes the user desires. This combination of attributes can be done on the fly or preconfigured, either in the database by e.g. views, or in the user-interface by e.g. joints, or in the publishing the data from the database to fit the interface format, e.g. by de-normalisation.

The technical conditions are set by the Shell geo-information strategy, which is described in section 2.3. They are alike the recommended architecture that is described in this geo-information strategy and consist in short of an Oracle database, ArcGIS as geographic information system, and links to other information and documentary systems, such as SAP and LiveLink.

Besides the components solely, the relations between the other three components can be treated as the conditions for the new data-model as well:

The relationship between the data-component and the users-component is mainly described in chapter 3 by means of the interviews and the questionnaire. This is expressed, amongst others, by the dataflow schemas, workflow description and query handling (section 3.2). In short this comes to that only the data are used that are at hand. The data are regarded as indirect reference information, that is supplied through the Hazard Notification programme.

The relationship between the users and the technology is partly described in the workflow description in section 3.2. Furthermore the Shell Geo-Information Strategy, which is described in section 2.3, defines this relation too: the main user interface is ArcGIS, where the data are visualised to the user. To have access to the documents, either the paper archive or the LiveLink documentary system is used.

The relation between technology and data is also greatly determined by the Shell Geo-Information Strategy. Adding to that are the data formats as described in section 3.2. The data are stored in several formats (table 3.3) but desired is to have them stored in one central Oracle database. Further the data should be published and visualised via ArcGIS and ArcIMS holding links to the appropriate documents that are stored in LiveLink.

4.3 Introduction to the Unified Modeling Language

The Unified Modeling Language (UML) is an object-oriented schematic language with standardised meanings of symbols and relations. It gives the possibility to describe a data-model in a visual and schematic way and can be interpreted by the standardised meanings of the symbols and relations. This modelling language was accepted in 1997 by the Object Management Group (OMG) to become a standard in the design modelling [61 & 62]. The concept of the UML is based on object-oriented technology, which means that it tries to identify objects of the real world and describe them as such, including their functionalities, data and attributes [36].

Within the Model Driven Architecture (MDA) UML is seen as the tool for structuring and visualising the models at the different levels. But UML is more than a visualisation tool: it is a structured language. It is regarded as standardised language for modelling both platform independent (PIM) and platform specific models (PSM). These are part of the Model Driven Architecture (MDA). The OMG even regards UML as vital for utilising the MDA-concept [41].

In line with the MDA, UML as standardised language can therefore be interpreted by various programs. On of the possibilities is to have UML-schemas translated into relational databases. One option would be the conversion of an UML-schema by a combination of Microsoft (MS) Visio and ArcCatalog from ESRI, into a relational (geo)database [13]. A brief introduction of the use of UML, based on the OGC and ISO (draft-)specifications is given in box 4.1 and figure 4.7.

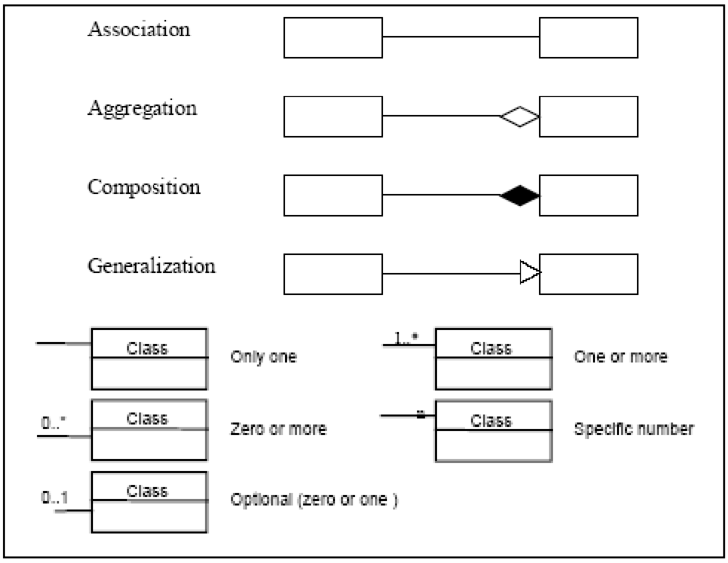


Figure 4.7: UML notations and relationships [24 & 47]

Box 4.1: UML notations & relationships [24 & 47]

The diagrams that appear are presented using the Unified Modelling Language (UML) static structure diagram. The UML notations used in are described in figure 5.1. The relationships presented are described as follows:

An association is used to describe a relationship between two or more classes. UML defines three different types of relationships, called association, aggregation and composition. The three types have different semantics. An ordinary association shall be used to represent a general relationship between two classes. The aggregation and composition associations shall be used to create part-whole relationships between two classes. The direction of an association must be specified. If the direction is not specified, it is assumed to be a two-way association. If one-way associations are intended, the direction of the association can be marked by an arrow at the end of the line.

An aggregation association is a relationship between two classes in which one of the classes plays the role of container and the other plays the role of a containee.

A composition association is a strong aggregation. In a composition association, if a container object is deleted, then all of its containee objects are deleted as well. The composition association shall be used when the objects representing the parts of a container object cannot exist without the container object.

Generalization
A generalization is a relationship between a super-class and the subclasses that may be substituted for it. The super-class is the generalized class, while the subclasses are specified classes.

4.4 Integrating the geotechnical & footprint data-model with the FAME-2 data-model

After designing the technical data-model for geotechnical & footprint data, it was compared to the recently redeveloped FAME-data-model (i.e. the FAME-2 version). It turned out, that both data-models could be integrated almost seamlessly and could make use of a common user interface. This comes forth to the concept of a corporate GII: a seamless integration - and in this specific case even a merge - of corporate datasets working harmoniously and being interoperable with other corporate information systems. Through this concept of a corporate GII and regarding the usability for Shell EP Europe, it would be very much desired and unavoidable to combine both data-models. The result is that the process of harmonising geotechnical & footprint data has become integrated and implemented within the FAME-2 harmonisation process.

This integration has also consequences for itinerary set up for this project as formulated in chapter 1. The research scope, questions and activities will be kept unchanged. However, the itinerary of the FAME-2 harmonisation process has much correspondence with the geotechnical & footprint data harmonisation itinerary. As a consequence, the research scope will, where possible, extend to the whole of the offshore point datasets into account.

This integration of the geotechnical & footprint data-model with the FAME-2 data-model should be considered a boundary constraint regarding the implementation of the data-model towards the technical components of the information system. Though on the other hand the implementation of the geotechnical & footprint data-model can be seen as a first implementation step towards the corporate GII.

Although the geotechnical & footprint data-model is taken in into the FAME-2 data-model, the harmonisation process of the first was at that point further advanced than the FAME-2 project. When the geotechnical & footprint data-model was completed, the FAME-2 model was just in modelling phase. The advantage hereby was that the geotechnical & footprint data-model could be merged without iterating the modelling phase, because certain aspects of the FAME-2 model could be adapted to overcome the small differences between the geotechnical & footprint data-model and the preliminary FAME-2 data-model. Disadvantage was that there was not yet a set FAME-2 data-model and that discussions were still going on.

One of the first actions to be taken was to integrate the geotechnical & footprint entities into the FAME-2 model at conceptual level. This is visualised in figure 4.8, presenting all the FAME-2 entities, including the geotechnical & footprint datasets. When the conceptual model of the elaborated FAME-2 data-model was created, the technical details could be entered, resulting for the geotechnical & footprint data in the technical model as visualised in figure 4.9, where the visualisation is elaborated with the entities they are directly related to.

4.5 Structure of the new data-model

With the use of the UML, the new data-model can be structured and visualised. The new conceptual model of the FAME-2 dataset is visualised in figure 4.8, which is a reduced image of figure D.2.2. This conceptual model is based on the relationship model of FAME-2 (figure D.2.1). The UML-visualisation of the conceptual model represents mainly the entities or objects representing the real world, their relations and the tables that they will be stored in.

Also the technical model can be visualised in an UML-schema (figure 4.9, reduced image of figure D.3.1). This visualisation of the technical model only presents the geotechnical data and the footprint data. These two datasets are elaborated with the datasets they are directly related to (a.o. rig tables and document tables). This technical model also serves as input for the automatic database generation with help of UML, MS Visio and ArcCatalog. The possibilities for automatic database creation from UML models are described in chapter 5.

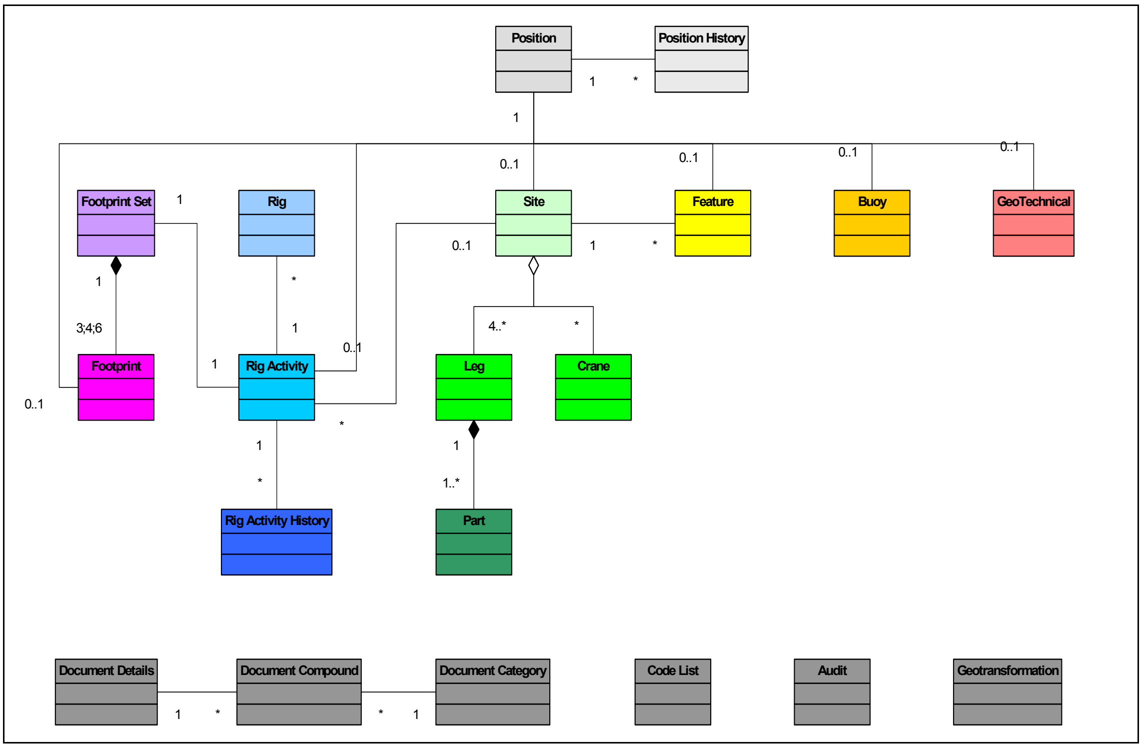


Figure 4.8: FAME-2 conceptual model (reduced image of figure D.2.2)

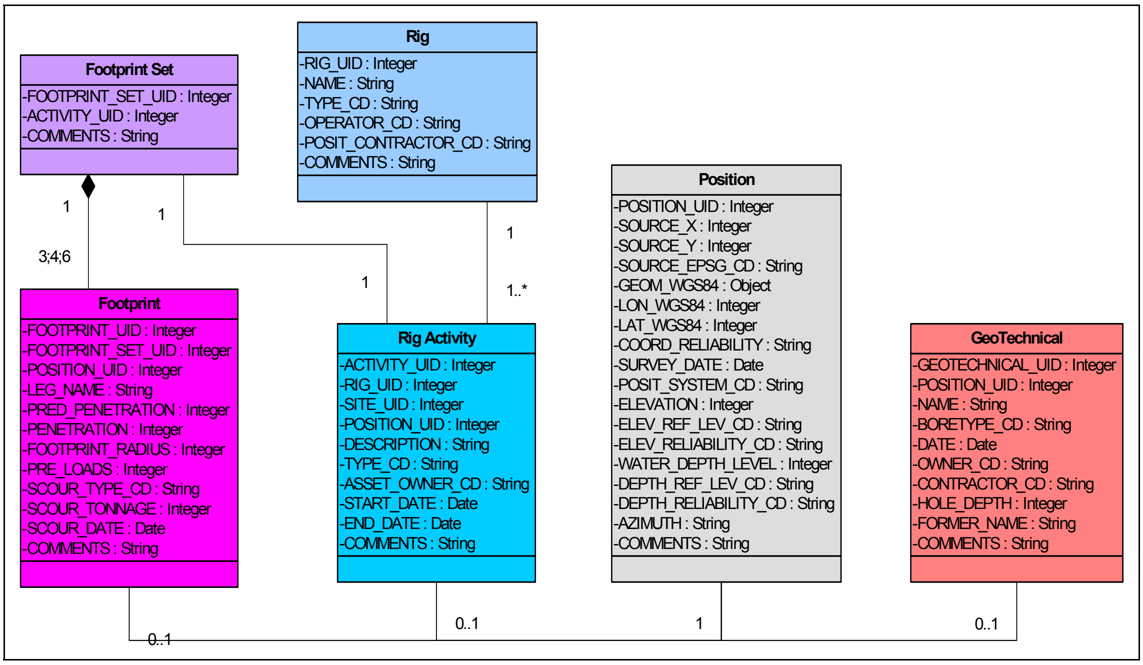


Figure 4.9: geotechnical & footprint technical data-model, elaborated with Rig and Rig Activity datasets (reduced image of figure D.3.1)

UML ambiguities

For the visualisation of the structure of conceptual or technical model in an UML-schema, the literature [34 & 36] comes up with several ways for how to use UML-tools. However in the process of visualising the structure of the extended geotechnical & footprint technical data-model, some issues came up that the modelling language was not unambiguous about. Two special indistinct situations that came up are worth mentioning.

Firstly in the situation visualised in figure 4.8, several classes have a link to 'Position'. However one position-instance can relate to only one of the classes and within that class only to one instance of that entity. In case two classes are related to 'Position' a {x-or}-constraint could be used. A constraint is described as a special type of control, limitation or restriction, sometimes Boolean, often an (in)equality relation, between two or more values or elements [36 & 63]. A {x-or}-constraint, being in UML a dotted line between two associations and the keyword '{x-or}', states that an instance of only one of the associates can be related to the subject-instance [34]. However this constraint can only be used between two class-instances. In case of figure D.2.1 and the relations towards 'position', there are six classes. If a {x-or}-constraint needs to be implemented in this case, there need to be 15 of these constraints made.

Another option to cope with this constraint is that all the classes are modelled to be subclasses of a (abstract) super-class type. For this super-class 'Position' can be chosen, from which the feature classes will then inherit its attributes, as can be seen in figure 4.10. An extended example of this can be seen in figure D.2.3.

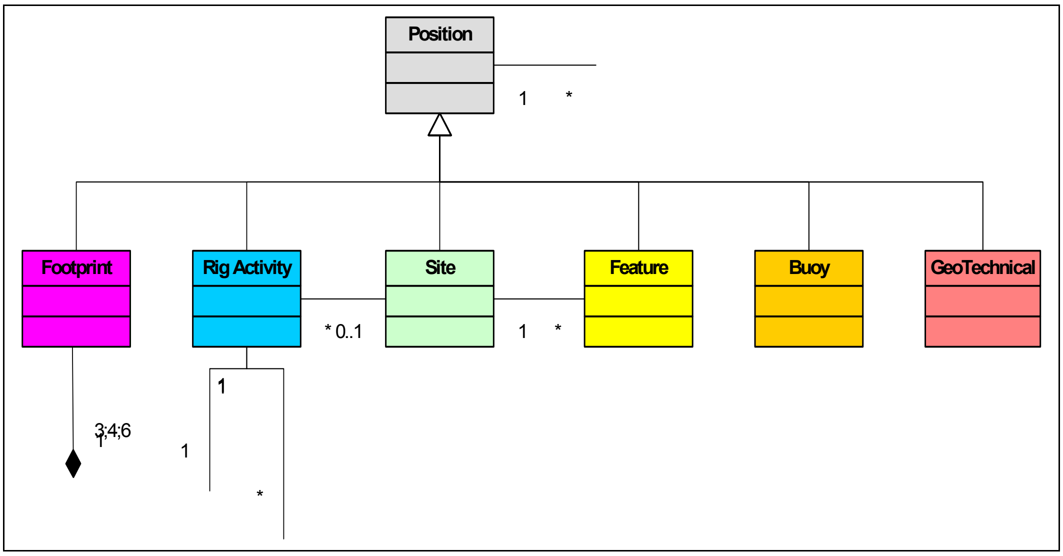


Figure 4.10: FAME-2 conceptual model with Position as super-class (reduced image of figure D.2.3)

The second ambiguity originates from the definition of the concept 'association class'. This is a class type that is identified with and linked to an association. This association class is used when the association has attributes, or has operations, or is associated itself to other classes [42]. It is, however, difficult to make a clear distinction between the situation where two classes are directly associated by means of an association class, and the situation where two classes are indirectly associated, having an attribute class as intermediate. In the case visualised in figure D.2.2, the class 'Rig Activity' can be seen as an intermediate class between several classes, most importantly being the association between 'Position' and 'Rig'. Or it can be seen as an association class between 'Position' and 'Rig', being also associated with 'Site' and 'Footprint Set'.

A possible solution for this ambiguity lies in the creation of the model: In the modelling process (as described in section 4.2.1) choices have been made about which objects / entities to take in the model. Since a rig is a mobile object it does not have a lasting position. In that case the choice could have been to model the position of the rig as an association between 'Rig' and 'Position', and thus as an association class. However a rig also has a certain assignment to perform on a certain position. Now this assignment can be modelled as an entity of its own, and thus be the attribute class 'Rig Activity'. This latter has been the case in this particular situation.

Conclusion

Following the Model Driven Architecture (MDA) (section 4.1) this data-model is the basis for further system development. Combined with the information system's components (section 4.2.2), the MDA sets the path for further (ideally automatic) implementation of the data-model into the system. The way this is done for the geotechnical & footprint data is described in the next chapters.

5 Implementation of the data-model: UML possibilities for automatic database creation

Following from section 4.2.2, a good information system needs to have a balance between its components. The consequence for the implementation phase is that when one component is designed, altered or re-engineered and implemented in the information-system, then the components need to rebalance in order to become a well-functioning system again.

The balancing of the components can be described by the relations that are laid between these components. In this case of designing a new data-model these relations can be identified as follows (see section 4.2.2 as well):

- Data-model -> Data = data migration
- Data-model -> Users = user interface
- Data-model -> Technology = configuration

This chapter deals latter case of these three relations, the technological configuration. It gives an overview of the configuration possibilities for a newly created data-model to be implemented in the technological components that are supported by the corporate GII. This is done by looking at the possibilities for automatic database creation from the data-model, described as UML-schema. An introduction to UML is given in section 4.3. Section 5.1 describes how the programmes MS Visio and ArcCatalog can be of help in this conversion process. The technical issues in using these two programmes are explained in section 5.2 and sections 5.3 give conclusions about the use of the automatic creation of a database via Microsoft (MS) Visio and ArcCatalog using UML as a modelling language.

5.1 Converting UML to database via MS Visio and ArcCatalog

To be able to convert the UML model created to a database several steps need to be done. These steps are schematised in figure 5.1. First an UML schema should be created. In the choice for programmes, there were the possibilities to use either IBM's Rational Rose or Microsoft's Visio. The choice is made to use the latter, since this was the easiest to acquire the right version of. The schema that is used in this test is presented in appendix E.1. The testing started with a very simple version of this schema and is expanded step-by-step to the schema in figure E.1.1.

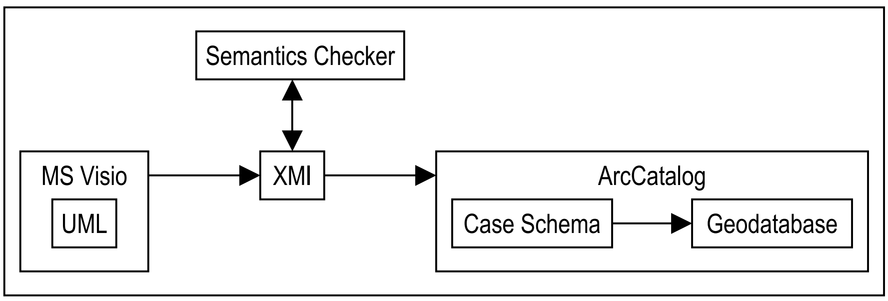


Figure 5.1: UML to Geodatabase conversion trajectory

After the UML schema is created in MS Visio according to the structure of UML and the rules set up by the ESRI-template, the schema can be exported to an XMI-file, using an add-on-functionality for MS Visio.

XMI (XML Metadata Interchange) is a metadata format that enables interchange of data and metadata between tools, middleware and repositories. It links modelling-formats such as UML with metadata formats, such as XML (eXtensible Mark-up Language) [53].

Directly following the UML to XMI conversion, it is necessary to check the XMI-file with the 'Semantics Checker'. This functionality, created by ESRI, checks the model for errors. If errors are found, a report is created, that points to the errors involved. When no errors are found, the model is – according to ESRI's programme – semantically correct and ready to be converted to a geodatabase schema. Examples of this Semantics Checker and how to use it are given in the ESRI literature [32 & 13]. A selection out of the XMI-file as in-between result of the conversion testing is to be found in appendix E.2.

The last steps of the UML to DB conversion take place by means of ArcCatalog functionality. First a geodatabase should be selected in which the schema will be created. In most cases this geodatabase needs to be created first. This can be either a MS Access database or an Oracle database via ArcSDE; both types were used in this testing series.

This last conversion step is done by ArcCatalog's 'Schema Wizard'. In this 'Schema Wizard' the geodatabase and the XMI-file have to be specified. The latter is then converted to a geodatabase schema, as can be seen in figure 5.2.a. For each of these tables or relationships the properties can be checked and some can still be altered (figure 5.2.b).

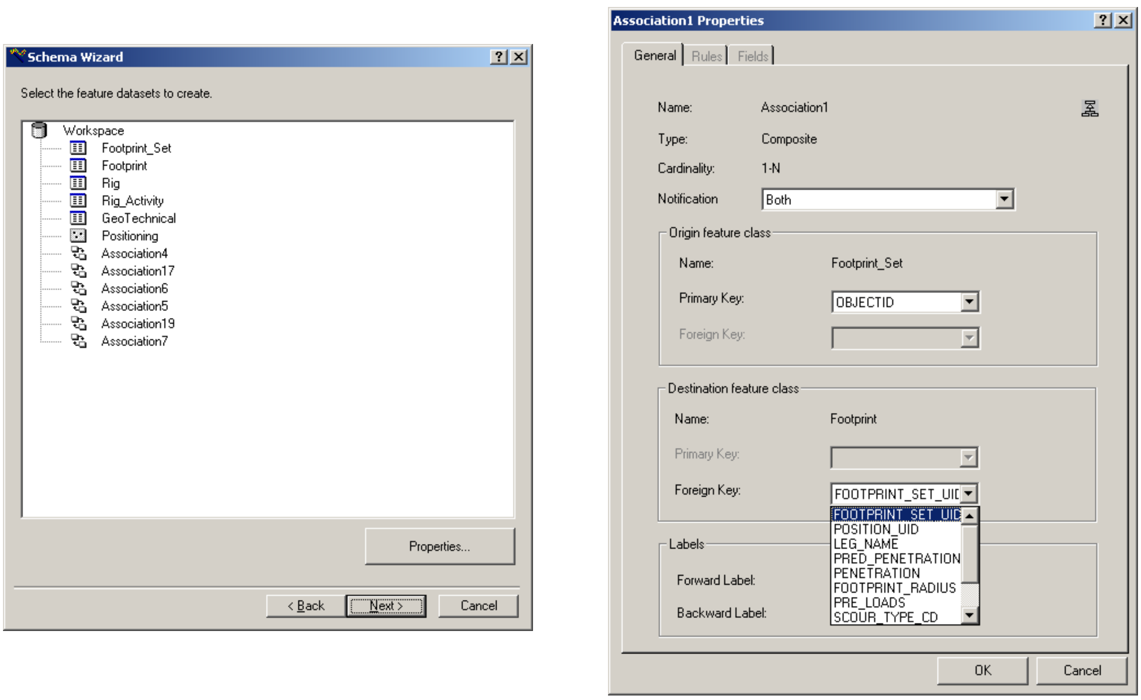


Figure 5.2: ESRI ArcCatalog's Schema Wizard: geodatabase schema and association properties

Note that in the properties for association 1, the composition relation, which states that the footprint set is made up by three, four or six footprints, is converted to a 1:N-composition. This means that a limiting relationship constraint is lost and should be fit in elsewhere, whether in the database or in the data management interface. The cause of this shortcoming lies in the conversion from XMI to the geodatabase, since this constraint is still available in the XMI-file. This can be seen in appendix E.2, where a part of the XMI-file is displayed with the 3,4,6-constraint highlighted.

Once all desired alterations are done, the Schema Wizard summarises the schema that will be implemented into the geodatabase. On approval, this is displayed as a log file (see appendix E.3) and the geodatabase in created. Now it can be populated.

5.2 Technical issues of the conversion by MS Visio and ArcCatalog

If a UML to DB conversion by means of MS Visio and ESRI ArcCatalog is intended, several technical issues have to be dealt with:

One of the first issues that came up was about the versioning of the programmes involved. When wanting to use the needed functionalities of ArcCatalog an ArcEditor licence is required. And with MS Visio as well: UML-schema creation can be done in MS Visio version 2000 and later versions, however the conversion and linkage with ArcCatalog can only be done in version 2000 'Enterprise' or higher. The version of the MS Visio programme used in this case is 2003 Professional.

In addition to the specific version of MS Visio there is also the need for add-on functionalities to be able to create correct XMI-files from UML-schemas. The UML to XMI conversion functionality is offered by Microsoft. ESRI provides the other two required add-ons: a special UML-template and the 'Semantics Checker'. The UML-template should be installed in a specific folder, stated in a short manual hidden on the ESRI support site. The same applies to the Semantics Checker, which even should be installed in the folder, where the XMI-files are stored.

The geodatabase to create the schema for needs to be available before inserting the schema by the Schema Wizard. This means, that both an empty database and the database connection need to be established on forehand, without directly creating the database schema in the creation procedure.

5.3 Conclusions regarding UML to database conversion

Although the idea of having the database automatically derived from the data-model, as is described in the MDA-concepts (section 4.1) offers many possibilities [41], however the current conversion of UML-schemas in MS Visio to ESRI geodatabases, narrows it down to a specific and non-open format. The main limitation comes from the database format that can be used as output: only an ESRI geodatabase is applicable, stored in either MS Access or Oracle.

Although there are some technical issues to overcome before the conversion toolset is working properly, after having created the UML-schema, there are just three functionality buttons to push before the database schema is applied (two in MS Visio and one in ArcCatalog). Thus once it is known what to do, the path is very easy. Although in future it might be handy to have all three functions combined behind only one button, facilitated from ArcCatalog.

But even if a geodatabase is desired, the main bottleneck will be in the time and effort needed to create an appropriate UML-schema in MS Visio that acknowledges the rules set by the ESRI-template. If a personal MS Access geodatabase is required, it often is a small project database with data derived from existing source. Such a database is then easier created in ArcMap, by adding layers to a blank workspace and saving it as personal geodatabase. If an ArcSDE geodatabase in Oracle format is desired, it often is a large departmental or even corporate database. In this case this UML to DB conversion tool can come in handy. But setting the right values for all relationships and classes to comply with the ESRI template could take as much - or even more - effort as creating the database schema through SQL.

If this UML to DB conversion toolset would be enhanced, the major improvements could be gained at easing the limitations that the ESRI template puts on the UML modelling possibilities. Or to say it differently: supporting more general UML design as input for the conversion toolset. If it were easier to design a database by UML and converting it to a (geo)database schema, it would better compete with creating a database schema by programming code, such as SQL.

6 Data management interface prototyping and testing

This chapter describes the implementation step of the ‘Data-model’ towards the ‘Users’, as coming from the model described in section 4.2. This implementation step consists mainly of the design and construction of a user interface for handling and managing the Data, Data-model and the Technology. For the newly set up overall FAME-2 data-model a distinction is made between interfaces for data consumers and data custodians. For the latter the data management interface is designed, prototyped and tested on usability. More information about this interface is given in section 6.1. The usability testing possibilities and the actual testing done are described in section 6.2. The conclusions of the testing for the interface are stated in section 6.3.

6.1 Introduction to the data management interface

Following the MDA-concept the user interface should be derived from the platform independent model (PIM). The aim is to have this done automatically. In the combination of the information system’s components (section 4.2.2) and the MDA, the interface is then the resulting visualisation of the implementation step of the ‘Data-model’-component towards the ‘Users’-component. However, currently there are no tools available to have this automatically derived, therefore it is custom designed.

This custom designed data management interface is created by a contractor from Fugro-Inpark, and is designed and created in MS Visual Studio using Visual Basic. This data management interface’s purpose is to facilitate the loading, editing and reviewing of the data (to be) stored in the database. A custom design for the data management interface is selected, since the data custodians, who are the main users of the system, are not familiar with the Oracle and SQL functionalities and because it is better usable to have a custom built, simple interface.

Concerning the user interference to the data, there are actions involved. In figure 6.1 the dataflow with respect to the users’ influence is stated. From this dataflow it becomes clear, that there are two major points of interference between the users and the data, thus the two major aspects of the ‘user interface’. Firstly when adding the data to the system dataset and secondly when retrieving the data. In the first case the users adding the data can be referred to as data custodians, in the latter case they can be thought of as data consumers. The interface for the data customers will be facilitated along the defined paths of the corporate GII strategy, by use of ArcGIS, ArcIMS or other GIS-facilitating programmes from ESRI. For the data custodians a data management interface is custom designed.

Before the data that is entered or altered in the data management interface will be written to the database, a quality check needs to be done on the data-entries. In this Quality Check (QC) a colleague of the person entering the data checks whether all entries are correct. If correct, the hold on the data is released and the data are stored in the database.

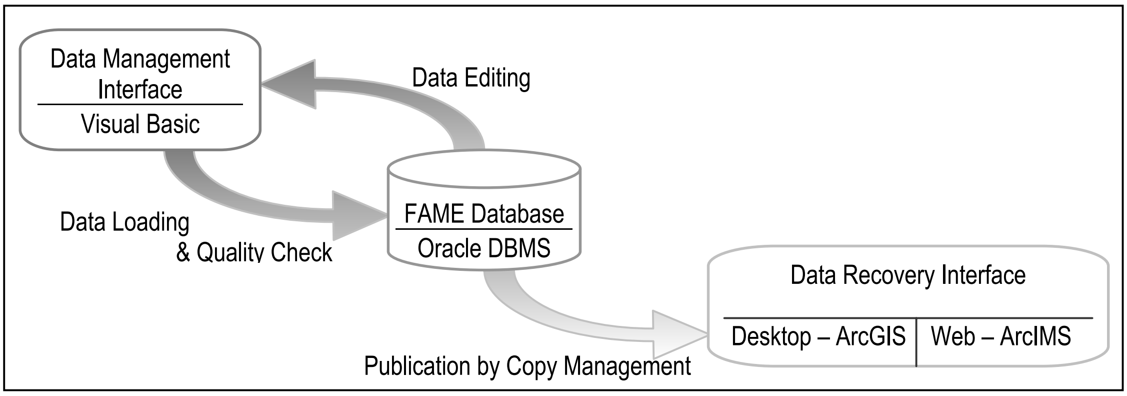


Figure 6.1: Dataflow with respect to user interference

From the database the data are copy-managed from the database-format to the required GIS-format overnight. This copy-management step is done for multiple reasons. The most important are, that the data stored in that database should be regarded as correct. Therefore it is required that direct access to that database is limited and controlled. This explains the QC-step as well. Another issue, which is already stated in section 4.2.3, is that the data stored in the database are in normalised tables, with not each of them having geometries stored. If map layers were to be created directly from these tables, a lot of data could not be linked. To require the correct information needs then linkage by 'hand' and would thus be hard to retrieve.

6.2 Interface usability testing

This section describes the usability testing done at the data management interface, which is created for the new FAME-2 system. Testing of the data management interface should be done to functionality and usability. This testing is done with a prototype of the data management interface and it is tested mainly on usability with regard to functionality and interface layout. The characteristics of usability to test the interface on are described in section 6.2.1. Section 6.2.2 gives with the possible test methods for usability testing. The characteristics and methods used in the actual test are stated in section 6.2.3.

6.2.1 Usability characteristics to test on

When an interface is to be tested on usability, it should be clear what is meant with that specific term. In literature there are several definitions of usability and usability testing, as they are described below. They differ in extent of detail, of context and in having a system or user-focus. At first sight these definitions might even seem about two different concepts, since some focus on the concept of usability itself, while others only regard it as a method of testing.

A quite sober definition of usability testing is described by Hom [21]. This system-focussed definition regards usability as “carrying out experiments to find out specific information about a design” [21].

At the other side there are also very user-focussed definitions. IRM state that usability is “the extent to which the intended user can meet his or her goals using the system being tested” [23]. A bit more detailed user-focussed definition comes from Clairmont, Dickstein & Mills: “the degree to which a user can successfully learn and use a product to achieve a goal” [9].

What comes up in these definitions is, that the user-focussed definitions are about the concept of usability and the system-focussed definitions are dealing with usability testing. It would, however, be more convenient to have the concept of usability defined in such a way, that if an interface would be tested on it, it would be tested on the same characteristics as the concept were defined by.

A less one-sided definition comes from Levi & Conrad, who say that usability testing is “the process by which the human-computer interaction characteristics of a system are measured and weaknesses are identified for correction” [30]. Doing so, they define the concept of usability to be set by human-computer interaction characteristics. However, according to Nielsen, human-computer interaction is a much more broader concept than ‘only’ usability [39].

Along the same path as Levi & Conrad, but less abstract, Gaffney defines usability testing as a “technique for ensuring that the intended users of a system can carry out the intended tasks efficiently, effectively and satisfactory” [16]. The same characteristics are also used and even extended by TNO. In their usability tests they determine usability “in terms of effectiveness, efficiency, satisfaction and learnability” [59]. These definitions bridge that gap between the system-focus and the user-focus. They define characteristics of the concept on which an interface can be tested.

A similar definition comes from Nielsen, who says that usability is “how well users can use the functionality” [39] of a system. Added to it he places the concept of usability in a broader context and characterises it by the following categories (figure 6.2):

- Learnability: how easy the system is to learn
- Efficiency: how efficient the system is to use
- Memorability: how easy the system is to remember
- Error rate: how many errors are made
- Satisfactory: how pleasing the system is to use.

This wider context is set in regard to acceptability and usefulness. Acceptability is “whether the system is good enough to satisfy all the needs and requirements of the users”. Acceptability is combined of social acceptability and practical acceptability; this latter one is made up of several categories as well. One of these categories is usefulness. Usefulness is “whether the system can be used to achieve some desired goal”. Grudin [19 in 39] divides usefulness into utility and usability, however both of them are relative to the specific user and the specific task. Where usability is as described above, utility is “whether the functionality of the system in principle can do what is needed” [39]. By this definition, utility is comparable to the characteristic effectiveness as Gaffney and TNO describe it above. However these regard it as a characteristic being part of usability, whereas Nielsen regards it complementary, but independent.

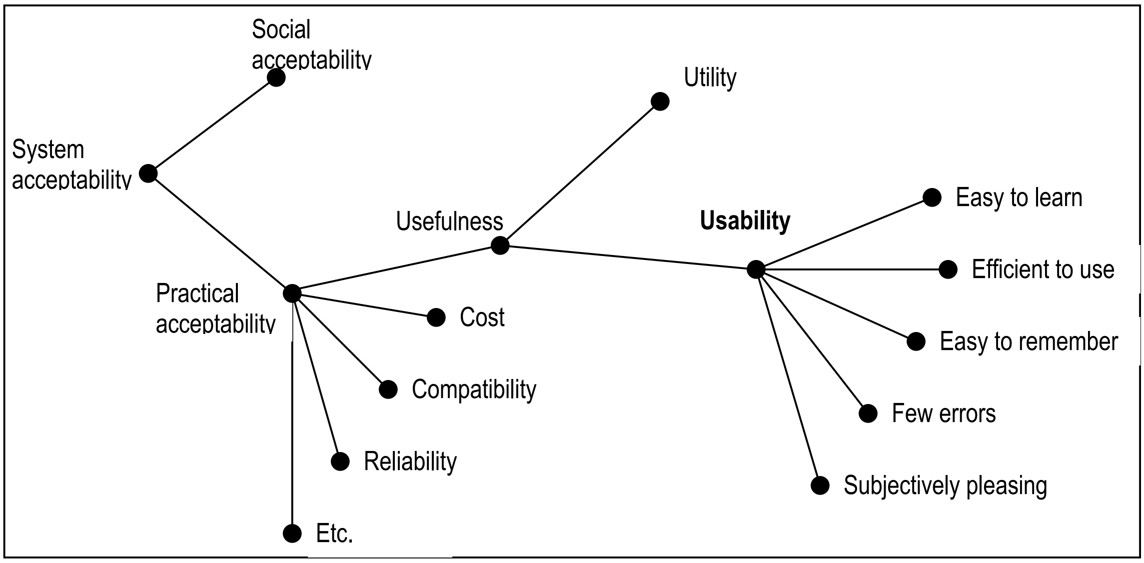


Figure 6.2: Attributes and categories of system acceptability, according to Nielsen [39]

6.2.2 Usability testing methods

Now that it is clear on which characteristics the data management interface is to be tested on, a choice for a test method and format should be made. A first differentiation is made to the different styles of testing:

- Exploratory testing; examines a system for areas of user confusion, mistake or irritation. It is performed with no particular notions about specific problems or problem areas. Their main goal is to improve overall usability.
- Threshold testing; measures whether specific system functionalities meet predetermined goals. This is a pass or fail testing by means of the performance characteristics of the system.
- Comparison testing; compares two or more designs. It is measured by usability characteristics which design suits the users' needs best [30].

In the case of the FAME-2 data management interface an exploratory test will be held (see section 6.2.3).

Besides different styles of testing, there are also several methods available for usability testing, either with or without users. Levi & Conrad [30] describe four separate methods, the first three of them are also described by Clairmont, Dickstein & Mills [9]. Nielsen [39] and Butz & Kruger [8], however recommend combinations of some of these separate methods. The intended methods are:

- Card Sorting, is used for hierarchy or structural indications, where users should pile up cards that represent concepts, functionalities or menus. These piles should be explained and named by the user [9 & 30].

- Heuristic Evaluation: a team of experts explore the system and identify usability problems and its compliance with usability principles. An example of these principles can be seen in box 6.1. A variant of this is the Usability Inspections or Cognitive Walkthroughs, were a team of designers moves through tasks like a user would, looking for problems [9 & 30].
- Scenario Based Testing: Users are given scenarios that represent the majority of the functionality and simulate real-life usage. The users behaviour while executing the scenario is observed and analysed to improve the usability of the product. According to Dumas & Redish [12 in 30], it exists of five components:
  - The goal is to improve the usability of a product
  - Testers represent real users
  - Testers do real tasks
  - User behaviour and commentary are observed and recorded
  - Data are analysed to diagnose problems and recommended corrections[9 & 30].
- Questionnaire: investigates "less quantifiable aspects of the interface design that contribute to users' subjective feelings of satisfaction or frustration" [30].

As described above, Butz & Kruger [8] and Nielsen [39] recommend combinations of methods. According to Butz & Kruger, usability testing should be using at least the methods of a User Satisfaction Questionnaire and a User Testing. This User Testing is "applied experimentation in which developers check that the system being developed is usable by the intended user populated for their tasks" [8]. One can thus say, that they recommend a usability test to have at least a scenario based test and a questionnaire.

Nielsen has a special method that is called 'discount usability engineering'. According to Nielsen it is composed of four different techniques: User and task observation; Scenarios; Simplified thinking aloud; Heuristic evaluation [39]. However, according to above methods, user and task observation, scenarios and simplified thinking aloud are all components of a Scenario Based Test, though being a bit more specialised. Thus the 'discount usability engineering method' can be regarded as composed of a Scenario Based Test and a Heuristic Evaluation.

Box 6.1: Nielsen's Heuristic Evaluation principles [40]

The ten general principles of heuristics:
- Visibility of system status
- Match between system and the real world
- User control and freedom
- Consistency and standards
- Error prevention
- Recognition rather than recall
- Flexibility and efficiency (ease) of use
- Aesthetic and minimalist design
- Help users recognise, diagnose and recover from errors
- Help and documentation

6.2.3 Usability characteristics and methods used

The characteristics of the concept of usability that are tested upon in this FAME-2 data management interface usability test, are the five characteristics described by Nielsen as given in section 6.2.1. These characteristics are chosen, since they are the most concrete and elaborated, and therefore the most workable. Adding to this is also the wider context that the concept and the characteristics are placed in.

Concerning the usability test methods, a combination of the methods of Nielsen and of Butz and Kruger is used. In the usability testing both the functionality had to be tested as well as the layout. Described by the usability characteristics used, the functionality can be expressed in terms of Efficiency and Error rate; the layout is more covered by the terms of Satisfaction as well as partly the terms of Learnability and Memorability. In the usability test run, these are split following the method of Butz and Kruger. The usability test for FAME-2 user interface consisted of:

- Scenario Based (user) Testing; being representative for the work process and having a few open questions in advance and afterwards to get an overview of users' expectations.
- User Satisfaction Questionnaire; held directly after the scenario based test, having open questions about users' expectations and satisfaction by means of Likert scales and semantic differential scales. A major part of this user satisfaction questionnaire exists of open questions about the layout of the interface.
- Heuristic Evaluation; in which the scenario testing is done by the author as well, with the focus on the heuristic principles (box 6.1).

Regarding the questions in the test and the questionnaire, they will be mainly of open questions, resulting in the test question being comparable to in-depth non-structured interviews. This type of questions results from the style of testing chosen for the FAME-2 data management interface: an exploratory test (section 6.2.2). To use and exploratory test style is chosen, due to the availability of only one prototype and only two main users: one main user is part of the Geo-information Management team and is based in Aberdeen, the other main user is an Offshore Surveyor, based in Assen. Since there is no genuine user-community, comparison testing and testing by statistical thresholds will not deliver the required information about the usability of the system. Comparison testing has no added value due to the availability of only one prototype and no alternatives.

6.3 Conclusions regarding the data management interface

From the results of the different components of the usability test (appendix F.1) several conclusions can be derived about the prototype of the data management interface. The main conclusions of this usability testing are:

- The tests were representative.
- The overall satisfaction and expectations results can be regarded neutral to positive. Keeping in mind the coming of improvements to the present prototype then the expectations will be mostly met and the satisfaction will be of good level.

- Explanation to the users about the linkage between the FAME-2 data management interface and ArcGIS is needed.
- Explanation to the users about the linkage between the FAME-2 data management interface and LiveLink is needed.
- Actions desired are listed accordingly the related list, indexed according to the MoSCoW-concept (see appendix F.2).

For efficient implementation of the changes suggested, a differentiation to importance should be made in regard to these changes. This differentiation is done by means of the MoSCoW-concept. This concept, which is also mentioned earlier in this thesis, is an acronym for Must, Should, Could and Want to have's. Where the first is an inevitable need for action, the latter one is an indication of what would be nice if there are any possibilities for it.

As general conclusion concerning the testing of the data management interface can be said, that -although the prototype was functional and was according to the functional specifications- the users' expectations were not completely met. To put it in terminology of the usability concepts: the Learnability, Memorability and the Error rate were sufficiently, however the Efficiency and Satisfaction were only meagre and could still be improved.

The main improvement to be done concerns the structure of the tabular series 'Rig' - 'Rig activity' - 'Footprint set' - 'Footprint'. When these tables will be structured differently, having as first entry option 'Rig activity', with a direct link to the sub-tables 'Footprints' and 'Rig' (and the table 'Footprint set' dropped), the efficiency and satisfaction would greatly improve, as can be concluded from the usability testing.

7 Data migration and management

This chapter describes the possibilities for the data migration and population of the databases and the possibilities for managing the system afterwards. These possibilities are described by means of a migration plan and recommendations. Section 7.1 describes this data migration plan and the recommendations for the data migration and population of the database. Section 7.2 gives possibilities and recommendations for system maintenance.

7.1 Data migration and database population

After having designed a new system, in order to make it usable data has to be added to it. As a motor needs fuel to generate movement, an information system needs data to generate information. Since there are already data available from the former datasets, it seems obvious to use these data again as a starting set. However to type in all data from the former datasets would be a tedious drudgery, therefore this should mostly be automated. In order to have this efficiently done, a data migration plan is set up, to link the sources and targets of the datasets and to guide in the transformation. Such a data migration plan exists of several elements:

- Overview of source data sets and the target database
- Relations between the source data-attributes and their target data counterparts
- Required translations and transformations
- Data selection criteria
- Method of transfer
- Risks and the quality check

An overview of the source datasets can be found in appendix B.1 & B.2. The overview of the target database is given in appendix D.3. The relations between the source data-attributes and their target data counterparts are given in appendices G.1 & G.2. Vice versa, the relation between the target data-attributes and their source data counterparts are stated in appendix G.3.

Required transformations

When it is known how the data-attributes of the former datasets relate to the new dataset, the required transformations have to be determined. The transformations and migration of the former datasets to the new FAME-2 dataset will be done using the Safe FME workbench-programme. The main translations and transformations in the geotechnical & footprint datasets concern the transformations from the UK measurement system (feet and inches) to the metric system. Given the coordinate data the transformations needed are from the various separate coordinate reference systems to the new EP Europe-wide standard of WGS'84. The reason for coming to use one standardised coordinate reference system for storage of coordinate data is explained in section 2.3. More details about the coordinate reference systems used are given in appendix A.

Data selection criteria

Data selection criteria are important in cases where it is unclear which data are relevant to take along into the new system. These criteria can also help when double data entries are to be dealt with. To distinguish between the data entries and to compare the data to find the most suitable and correct datasets can be done following different procedure steps:

- One option is to first execute the transformation and migration steps. After that the coordinates of the different datasets are compared. The double data entries can then be identified and if desired merged or deleted.
- A second option is again to go first through the coordinate transformation. But now, having all the coordinates in the same coordinate reference system, they are compared before migration. The double entries can then be selected and merged or deleted before or during the migration step.

The choice of which path to follow depends on the possibilities of the data migration programme. Also the choice for either data merge or data deletion can influence when to compare the datasets.

As stated above as well, a choice has to be made whether to merge data entries or to delete duplicate data or to do both depending on which data-attribute is at stake. The main consideration in this paradigm is coming from the policy the organisation has regarding the deletion of old datasets. If the organisation is keen on safeguarding all old datasets and is reluctant to throw away data, it will probably choose to transform all data from all datasets. If the choice has been made to only take along the most accurate and up-to-date data a way has to be found to select the right dataset for it. To be able to select the right data to be migrated to the new system the criteria can be found in the metadata of the datasets. Especially the quality parameters can be of help here, which are [18]:

- Data lineage
- Consistency of the data
- Completeness of the data
- Semantic accuracy of the data
- Temporal accuracy of the data
- Positional accuracy
- Attribute accuracy

Another possibility to distinct between duplicate data in the new combined datasets is by adding these metadata to the data when migrating, so that users themselves can choose which data to use. However this requires from the users to have a high-level awareness of metadata and to have all the metadata available.

In the case of the geotechnical & footprint datasets these quality parameters are used to make a selection of the data sources and which data source is preferred above another. In case of the footprint data, there are only a Dutch and an UK dataset (see section 3.1). However, the completeness of the Dutch dataset is the most important issue: there are no coordinate data stored in that dataset, therefore the data would hardly be traceable and quite useless if migrated to the new system. Thus only the UK dataset will be migrated. For a short period it can be useful to keep the Dutch dataset available in its old system.

Considering the geotechnical datasets, there are one Dutch and three UK datasets. The Dutch dataset is available in GIS and can be migrated without major issues. The UK geotechnical data is available in OpenWorks and has not been updated since it was loaded in

1998. This data is currently the only geotechnical data available in GIS as part of the OpenWorks well copy-management. However, the GIS dataset is elaborated with more than 100 onshore and a few offshore data-points, but in the copy-management some of the fields are not taken along. Besides those two dataset, since 1998, geotechnical data has been maintained in a set of excel spreadsheets. All three datasets need to be migrated, however they need to be compared and before.

Method of transfer

As stated above, both the transformation and migration of the datasets will be done using the Safe FME Workbench programme. This programme also facilitates the transfer of the data from the old databases to the new FAME-2 database. Within this programme the former databases are selected in the format they exist in. These are stated as input giving all attributes, which in their turn can then be selected, either directly to move or to be transformed before moving to the database. The transformations required in the case of the geotechnical & footprint datasets are described in appendix G.

Quality check

Although the data migration process and the transformation steps should be executed without any problems, it is always wise to have an objective check whether all data of the new dataset are right. Such a check can be split into two types: a quantitative check and a qualitative check. The first checks whether the number of records (objects) afterwards are equal to the number of records of input, the latter checks whether the content of the data-attributes is correctly without error. In the methods of QC-ing, there are several kinds, each depending on the software platform involved in the migration and data publishing process:

- ArcGIS: comparison of old a new data layers by overlay or spatial queries. This is mainly useful to compare the geometric differences.
- FME: an independent check of the FME-mapping script; this checks whether the intended transformations and migration steps are correctly stated.
- SQL / Oracle: a scripting by SQL statement to compare automatically the related tables of before and afterwards.
- Analogue: To compare the separate tables and entries by hand.

Each of these options can be executed on the full datasets, on random samples, or on predetermined samples.

Regarding the geotechnical & footprint data a three-step quality-check is recommended: Firstly preceding the migration an independent check of the FME-mapping script should be done. After the migration an analogue check of predetermined samples is to be done by the user in the data management interface. The final quality-check should be after publication of the data by using an ArcGIS overlay check. The SQL check is not recommended, since the entries in the code-lists are allowed to differ from the entries of the former datasets. If then a comparison is done between the new and old datasets, these differences in code-list entries will then come up as errors as well. Although the proposed three-step quality check might seem rather thorough, it does safeguard the quality. Adding up to that is the fact that when the data are correct, they would pass the quality checks without causing much effort for action.

7.2 Data management

Giving recommendations for the management of the geotechnical & footprint data in the corporate GII can be split up according to the components of a GII [18 in 6]. They are elaborated with some special migration issues concerning the FAME-2 datasets.

Organisation

- Set up a user's guide to have a safety net for the current users and a guideline for new users. This guide should consist of at least a data dictionary and an interface walkthrough.
- Have a high level of communication, since there are many departments and locations involved, as can be seen in figure 2.6.
- Set out the responsibility for both the system and the datasets clearly. There have to be one or two responsible persons for both the data and the system; more is not recommendable since there is only one system in which one set of interdependent data is stored.

Standardisation

- Standardise the workflows and data specifications; the new standardised list can also be distributed to the contractors that deliver new data, who can deliver the new data directly in the right format.
- Keep going with the standards set in the geo-information strategy and try to elaborate them and implement them more widely.

Technology

- Set out the responsibility for the system clearly.
- Use views in overcoming the gap between the data the users wish to see and the data stored in the database. Other applications can then be linked to these views as well, without needing to de-normalise the tables themselves
- Look at the possibilities for a totally web-based interface, for both the data-requests, the data upkeep and the Hazard Notification, which is easier to maintain and manage.

Datasets

- Clearly set out the responsibility for the data management.
- Fill in the data gaps as soon as possible, and keep the data as accurate as possible.
- Consider the use of 3rd party data as available through the EU-Seased website, either by taking access to this website, or by copying relevant data to the Shell Environment.

Coordinate reference system

Implementation of the new system holds many consequences for the data management of the data types. One major consequence is the choice to use WGS'84 as new standard for the storage of all coordinate data. Although it is very recommendable to have a standard set for this, it requires a lot of communication towards the user why it has changed and more importantly *that* it is changed. The choice for WGS'84 is coming from the strategy of Shell to try to make any changes fit in a global extent, if possible. The WGS'84 coordinate reference system is best suited for this, since it is both a global reference system and as it is used in GPS-measurement, it is gaining importance.

In the use of WSG'84 in the FAME-2 data-model, there are six fields describing the coordinate data: two fields are given to the source coordinates, one field describes the source coordinate reference system, one field gives the EPSG-code concerning the transformation used in coming from the source coordinates to WGS'84 coordinates, and two fields state the coordinates in WGS'84.

Data gaps

Concerning the data gaps a choice should be made whether the new data-attributes will be only entered for new data entries or if it will also be done for the migrated data. In the latter case the best is to set up a working plan about when and how to fill these gaps. Then it should be decided upon if the data gaps for all migrated data are to be filled, or that there is a limiting factor, depending on one or more of the above mentioned migration selection criteria. In the case of the footprint data the decision has been made to only fill in the new data entries and not to fill the data gaps of the migrated data. This has to do with the temporal character of the footprint data, as these outdate relatively fast, due to scouring effects and erosion. Concerning the geotechnical data the data gaps will be filled, but only for data younger than approximately two years. This is not because of the temporal character, but because of the amount of handwork needed to fill all the data gaps.

Phasing out old datasets

The other issue is about phasing out the old datasets. The choice when to phase out the dataset depends on both the dataset itself, it's use and the data migration. Once it is clear that all data are correctly migrated and the new system is up-and-running, in principal the old dataset can be phased out. However, as stated above, in the case of the footprint data, it can be useful to keep the Dutch dataset available for a short period. This is based on the fact that the data from this dataset will not be migrated to the new system and will therefore still be required to be available as a separate dataset. When to phase out this particular dataset is to be set clearly: until it is outdated or until a set date.

8 Conclusions and recommendations

This chapter states the conclusions and recommendations that follow from this research project. In sections 8.1, the conclusions are described, mostly derived from the research questions as they were presented in chapter one. Section 8.2 gives recommendations, both in general about the research topics and in specific for the department of Geo-Information Management within Shell EP Europe.

8.1 Conclusions

The conclusions of this research project are given by means of answering the research questions stated in chapter 1. The answers to these research questions lead to the answer of the main question of this research.

What does the corporate Geo-Information Infrastructure for Shell EP Europe look like?

In short it can be stated as the storage of all sorts of geographic data in mainly Oracle databases, accessible via ArcGIS and ArcIMS, and linked to various stored documents and other non-geographic corporate datasets. The corporate GII is based on the Geo-information strategy, which recommends this architecture, including the use of both internal Shell standards and external, ISO, OGC and EPSG standards. The corporate GII is more elaborately as is described in section 2.3.

Which geotechnical & footprint data and related data-models are used within Shell EP Europe?

Shell EP Europe used two footprint datasets, one at NAM and one at Shell Expro. There were four geotechnical datasets in use: one at NAM and three at Expro. There is also a European webpage being a metadata portal for seabed soil data.

For the geotechnical data, there was no data-model in Shell EP Europe. For the footprint data only one data-model was available, which was part of the first edition of FAME. This model was not in use, and is not suitable as Shell EP Europe-wide data-model.

The geotechnical & footprint data and the related data-models are as described in chapter 3 and appendix B.

What should a new data-model look like?

The new data-model is derived from the conditions set by the information system's components: technology, users and data. Combined with the concept of MDA, such a model can be used in the implementation process, having the data-model as central axis.

The conditions to come to such a model are explained in chapter 4. In short they are the recommended architecture of the Shell geo-information strategy (section 2.3), the data-attributes (box 3.1), dataflow and the relations to the other datasets (figure 4.8) and the users' conditions as described in the workflow (section 3.2). From these conditions, the dataset inventory and the definitions a data-model is created that harmonises the geotechnical data and footprint data and fits them harmoniously in the corporate GII.

Research of the possibilities to fit the data-model in the corporate GII, led to the conclusion that the geotechnical & footprint data-model could be integrated in the FAME-2 data-model. In line with the ideas of a corporate GII and harmonisation, the geotechnical & footprint data-model is therefore implemented in the FAME-2 data-model. In this model they are related to other datasets. An impression of the FAME-2 conceptual model is to be found in figure 4.8 and a more elaborate technical model in figure 4.9.

How can the new data-model be implemented into the corporate GII of Shell EP Europe?

According to the concept of MDA – having the data-model as central axis of an information system – the developed data-model is used as starting point in the implementation process (chapter 4). The model can be implemented regarding the information system's components: Data-model, Data, Users, and Technology. From the data-model the implementations towards the other three components of the information system are managed. These three implementations are:

- Configuration of the Technology; the soft- and hardware to support the data-model.
- Interface for the Users; created and tested for the new system.
- Migration of the Data; from the old systems to the new system.

The details of how these steps can be taken can be found in respectively chapter 5, 6 & 7.

What recommendations can be done to manage geotechnical & footprint data in the corporate GII of Shell EP Europe, using the new data-model?

Recommendations for the management of the geotechnical & footprint data in the corporate GII can be split up according to the components of a GII [18 in 6]:

Organisation: Set up a user's guide to have a safety net for the current users and a guideline for new users. This guide should consist of at least a data dictionary and an interface walkthrough. Have a high level of communication, since there are many departments and locations involved, as can be seen in figure 2.6. Set out the responsibility for both the system and the datasets clearly. There have to be one or two responsible persons for both the data and the system; more is not recommendable since there is only one system in which one set of interdependent data is stored.

Standardisation: Standardise the workflows and data specifications; the new standardised list can also be distributed to the contractors that deliver new data, who can deliver new data directly in the right format. Keep going with the standards set in the geo-information strategy and try to elaborate them and implement them more widely.

Technology: Set out the responsibility for the system clearly. Use views in overcoming the gap between the data the users wish to see and the data stored in the database. Other applications can then be linked to these views as well, without needing to de-normalise the tables themselves. Look at the possibilities for a totally web-based interface, for both the data-requests, the data upkeep and the Hazard Notification, which is easier to maintain and manage.

Datasets: Clearly set out the responsibility for the data management. Fill in the data gaps as soon as possible, and keep the data as accurate as possible. Consider the use of 3rd party as available through the EU-Seased website, either by taking access to this website, or by copying relevant data to the Shell Environment.

How can geotechnical and footprint data be harmonised to a data-model that fits into the corporate Geo-Information Infrastructure of Shell EP Europe?

In this research a model has been developed to harmonise two different kinds of distributed geographic datasets within a corporate GII. This is based on the MDA-concept, under the conditions set by the information system's components of Data, Users and Technology. The implementation of this model has led to the conclusion that this model is not only developed on theoretical and methodological concepts, but it is applicable in a real working situation. Due to the results of my research it has been introduced and implemented within Shell EP Europe.

8.2 Recommendations

Besides the recommendations for the data management, as stated above, other recommendations can be given as well. These can be split into recommendations regarding Geo-Information Management within Shell EP Europe (section 8.2.1) and general, non Shell-focussed recommendations, regarding further scientific research (section 8.2.2).

8.2.1 Recommendations for Shell EP Europe

The use of UML as structured design language and tool can help in communication between parties involved in the harmonisation process. Since UML is a standardised schematic language it can be interpreted more easily than a personally created chart with an explanation report on the side. These models can then be used as framework models for dealing with the datasets.

Additional to UML is concept of the Model Driven Architecture (MDA). Having a set of platform independent models (PIMs) as core to managing the datasets can help in keeping an overview of these datasets as well as the way they are structured and related. When tools for automatic derivation and transitions from PIMs to PSMs become better and more usable, these can then be utilised in the data management.

In the design or harmonisation process, there is a need for complete, unambiguous and solid functional specifications. If these are not clear, it is not clear what possibilities and functions the new system should have. Since all deeper models rely on the functional specifications, an alteration in these would cause much work to be redone. This implies also that a complete user requirements survey is needed and implemented in the functional specifications before the design of the system is started.

In the FAME-2 project it came out that the functional specifications as listed by the users, were not as solid as would be required. It became clear that the users were not able to set such specifications themselves. Therefore it is recommended for future projects to help the users set up such functional specifications, especially when project involve users from so many different locations and departments. This can be done by an in-depth user consult, to make sure that the functional specifications do reflect all of the users' wishes.

Consider the set up of an official corporate GII. For such a corporate GII directives should be put up, enhancing the interoperability, standardisation and accessibility of geographic information. Such a corporate GII can be based on the components described in section 2.3 with now not having the architecture recommended but set as technology component of the corporate GII. Also having (corporate) datasets appointed as authentic datasets for guaranteeing the data quality could improve the reliability of the corporate GII. Within the organisation there should then be good communication and advertisement about the corporate GII, for obtaining user support. And the use of standards can be more elaborated and preferably be made compulsory.

8.2.2 Recommendations for further scientific research

Although it is beyond the limitations of this research project, it can be assumed that the developed model is also applicable in other organisations. This case, researched in Shell EP Europe, proved to be best practice. The harmonisation method used can also be applied to other distributed geographic datasets. Probably the model will need only minor adjustments to make it applicable in other organisations. More research on this can prove this.

Since more and more organisations are applying GIS and are moving from older versions to new modern GIS and database environments, the focus should move from designing completely new systems to systems based on a legacy. This is often done through combining datasets, which can be referred to as a harmonisation process.

Already in this research project a few striking differences between the development process and the harmonisation process came up: In the information phase of the harmonisation process also the current situation, systems and workflow should be considered important input. The construction phase has to be extended with a data migration process, including data population of the missing fields.

This raises research possibilities about how to deal with such legacy, what differences there are between a new development process and a harmonisation process, and how a harmonisation process could best be set out.

Due to the fact that many data are already stored in digital format, moving them to a new system requires decisions on data migration. In this respect it is also important to review the risks of such data migration and, also with the merge of datasets in mind, a need for clear criteria for data selection on which data-attributes can be tested for (relative) suitability to be migrated to the new system. As is stated in chapter 7, these can be based on or linked to the metadata of the dataset and of the data-attributes.

Possibilities for automatic database creation from UML-schemas: The MS Visio-ArcCatalog conversion, described in chapter 5, is one possibility but there may well be good alternative processes and softwares. More research into the various possibilities for the conversion of UML-schemas to database tables and there advantages for specific applications could be useful.

The differentiation of an information system into the four components Data, Users, Technology, and Data-model (chapter 4) is derived from various other models and tested for this specific case. For this case it was a useful, proper and suitable tool to analyse the effects of the various components of an information system. However it is preferred to have it applied in more and various cases to screen and perhaps enhance the model.

Within data management of larger systems, that have the data storage facilities split from the data usage programmes, there is a difference between the data-attributes stored in the database table and the data-attributes the user wants to see in the map layer. In terms of object modelling this comes to a difference between the specifications on the level of the conceptual model and those on the level of the technical model. The data are usually stored in normalised database tables, however the data needed by the user in the GIS are required in de-normalised layers.

There are several possibilities to overcome this, either in the domain of the data user or that of the data manager. Some of these possibilities are based on relationships between objects, allowing selections of different tables to be combined in order to display the data-attributes the user desires. This combination of attributes can be done on the fly or preconfigured, either in the database by e.g. views, or in the user-interface by e.g. joints, or by publishing the data from the database to fit the interface format, e.g. by de-normalisation. However it could come in handy to have sorted out what all possibilities are and which are best to bridge this gap. Further research could be done to investigate optimum choices for different situations.

Many geographic data are stored and available via GIS nowadays. But still many data are only available in reports, analoguous and scanned or digital. Often it is desired to have a link between the data available through the GIS and these reports. However the current possibilities are not optimal for the use of more than one related document which is stored seperately from the geographic data. From a users point of view it is desirable to have all related documents available via links in GIS and to have this functioning smoothly. There are still possibilities to enhance this.

Epilogue

While finishing this thesis, the results and conclusions of my research have already been implemented and applied. Two of these applications are worth mentioning, hence this epilogue.

The data-model has been integrated with the FAME-2 data-model. This integrated model, as described in this thesis, was a merge of the two former data-models. By testing the usability of this new, integrated model I concluded that a part of the entities became redundant and an adjustment was possible. In fact this adjustment was a simplification of the data-model: The entity Footprint Set should be split up and divided amongst the entities Rig, Rig Activity and Footprints (figure 9.1).

As a result of the model driven approach the adjustment of the data-model led to adjustments in the implementation. For the user interface this implied a redevelopment of the menus. This was combined with the recommendation from the usability testing to reduce the number of screens. These were reduced from 7 to 4, being at first:

Rig -> Rig Activity overview-> Rig Activity -> Footprint Set overview -> Footprint Set -> Footprint overview -> Footprint

and now changed to:

{ Rig
Rig Activity -> Footprint overview -> Footprint

The second application resulted from the recommendations of the UML-conversion testing. It led to the decision, that the department GIM will start using UML as common language to structure, visualise and communicate the different data-models. The department GIM also will keep track of the developments concerning the Model Driven Architecture and use automatic conversion from data-models for publishing data, creating interfaces or building databases, whenever possible.

These applications show, that the data-model developed in this research is applicable for Shell EP Europe and, with minor adjustments, can be applied in other organisations. The harmonisation method used can also be applied to other distributed geographic datasets.

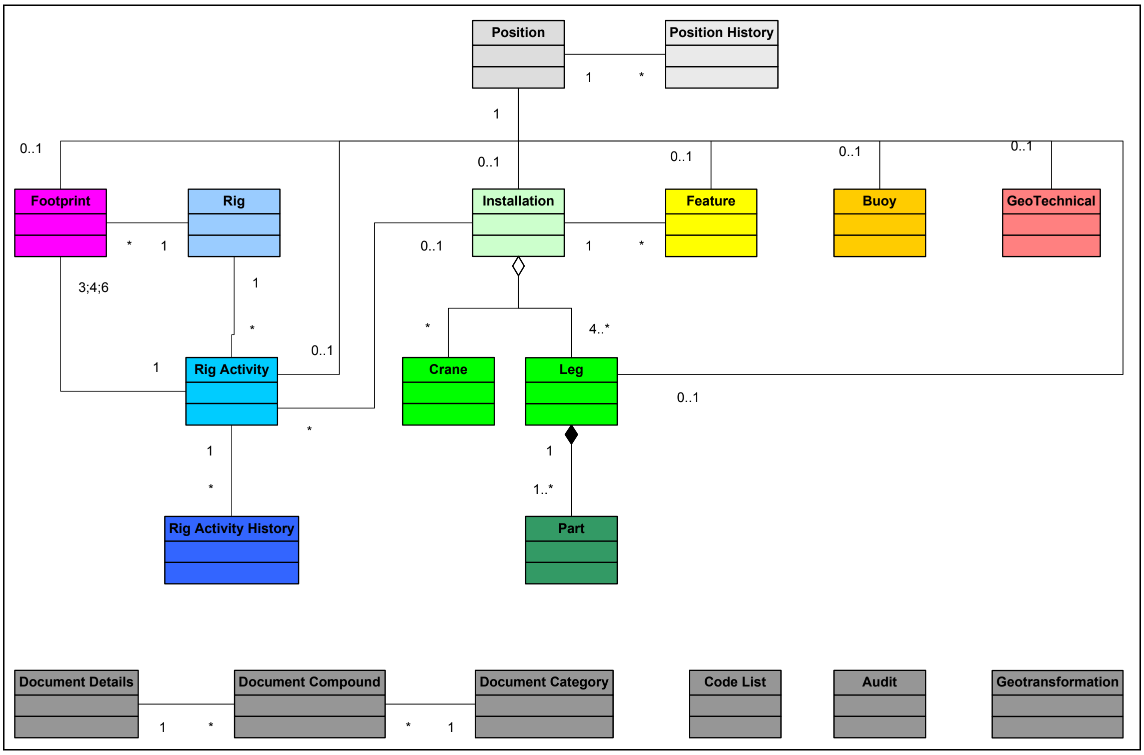


Figure 9.1: Altered FAME-2 conceptual model as basis for the FAME-2 system (21-12-2004)

Literature

-1- Al-Harty, M., P. Evarie & M. van Nes, 2004, *Aligning offshore infrastructure data models project plan*, Shell Exploration & Production – Europe - Data Acquisition and Management (Shell EP E-DAM).

-2- Bellinger, G., D. Castro & A. Mills, 2004, *Data, information, knowledge & wisdom*, in http://www.systems-thinking.org, vis. 09-2004 – 12-.2004.

-3- Bemelmans, T.M.A., 1999, *Bestuurlijke informatiesystemen en automatisering*, 7th ed., Kluwer bedrijfsinformatie.

-4- Berg, G.J.J.B. van den, 2004, *Investeren in R&D: aangenaam kennis maken?*, MSc thesis, Groningen: Rijksuniversiteit.

-5- Bishr, Y., M. Radwan & M. Molenaar, 1998, *Aspects of interoperability: a GII perspective*, Proceedings of 1st AGILE conference, Enschede: International Institute for Aerospace Survey and Earth Sciences (ITC).

-6- Bregt, A., 2002, *Space for geoinformation*, presentation on 6th GSDI conference 2002, Budapset, vis. 22-06-2004, http://www.gsdi.org/docs2002/GSDI6/PPT/bk240902.ppt.

-7- Brown, A., 2004, *An introduction to model driven architecture part I: MDA and today's systems*, IBM, the Rational Edge, vis. 12-2004, http://www-106.ibm.com/developerworks/rational/library/3100.html.

-8- Butz, A. & A. Kruger, 2003, *User interface design*, Fluidum, Munchen University, department of Computer Science, http://w5.cs.uni-sb.de/~butz/teaching/uid-ws03/.

-9- Clairmont, M., R. Dickstein & V. Mills, 1997, *Testing for usability in the design of a new information gateway*, university of Arizona library, vis. 5-11-2004, http://dizzy.library.arizona.edu/library/teams/access9798/lft2paper.htm.

-10- Dale, P., 2003, *The emperor's new clothes*, column in GeoInformatics, Vol. December 2003, pp. 33

-11- Dalen, D. van, 2004, *Aligning offshore infrastructure data models*, presentation for NAM Geo-Information Management, Fugro-Inpark

-12- Dumas, J.C., & J.C. Redish, 1993, *Practical guide to usability testing*, Norwood: Ablex publishing co. in [30].

-13- Environmental Systems Research Institute (ESRI), 2003, *Building geodatabases with CASE tools*, whitepaper, ArcGIS 9: designing geodatabases with Visio, Redlands: ESRI Support Center, vis. 09-2004 – 11-2004, http://support.esri.com, page submitted: 29-09-2003.

-14- EU-Seased.net, 2003 & 2004, http://www.eu-seased.net, vis. 05-2004 – 12-2004.

-15- Executive Order 12906, 1994, US Government, FGDC Website, in [37]

-16- Gaffney, G., 1999, *Information & design: usability techniques series*, vis. 05-11-2004, http://www.infodesign.com.au.

-17- Gregorius, T., 2000, *Coordinate systems and map projections*, presentation on Shell Geomatics awareness course, September 2000, Shell / NAM.

-18- Groot, R. & J. McLaughlin, 2000, *Geospatial data infrastructure; concepts, cases and good practice*, Oxford: Oxford University Press.

-19- Grudin, J., 1992, *Utility and usability: research issues and development contexts*, in: Interacting with Computers, vol. 4, nr. 2, pp. 209-217.

-20- Holland, P., 1999, *The strategic imperative of a global spatial data infrastructure*, Cambridge Conference for National Mapping Organisations, Cambridge, vis. 17-05-2004, http://www.gsdi.org/docs1999/stratim.html.

-21- Hom, J.T., 1996, *Usability methods toolbox*, vis. 5-11-2004, http://jthom.best.vwh.net/usability/general.htm.

-22- Hopkinson, F., 2003, *Spatial data infrastructures – Buzzing around!*, column in GeoInformatics, Vol. December 2003, pp. 9.

-23- Information Resource Management (IRM), 2002, *Usability testing*, University of Colorado, vis. 5-11-2004, http://www.cu.edu/irm/stds/usability.

-24- International Organisation for Standardisation (ISO), 2001, *Geographic information – Metadata*, Draft International Standard, ISO/TC 211 DIS 19115, OpenGeospatial Abstract Specification Topic 11: Metadata, version 5, in [48].

-25- Kadaster & Adviesdienst voor geo-informatie van Rijkswaterstaat, 2000, *Het stelsel van de Rijksdriehoekmeting en het European Terrestrial Reference System 1989; gebruik en definities*, Apeldoorn, Kadaster, vis. 04-08-2004, http://www.rdnap.nl.

-26- Kennedy, M. & S. Kopp, 2000, *Understanding map projections*, Redlands: Environmental Systems Research Institute (ESRI).

-27- Klasse Objecten, 2004, *What is the model driven architecture?*, vis. 12-2004, http://www.klasse.nl.

-28- Kraak, M.J. & F.J. Ormeling, 1996, *Cartography; visualization of spatial data*, Harlow: Longman, pp. 81-86.

-29- Lemmens, M.J.P.M., 2001, *An European perspective on geo-information infrastructure (GII) issues*, GIS Development, vis. 16-06-2004, http://www.gisdevelopment.net/policy/gii/gii0007.htm.

-30- Levi, M.D. & F.G. Conrad, 2002, *Usability testing of World Wide Web sites*, US department of Labor, Bureau of Labor Statistics Research Papers, http://www.bls.gov, vis. 5-11-2004, http://stats.bls.gov/ore/htm_papers/st960150.htm.

-31- Longley, P.A., M.F. Goodchild, D.J. Maguire & D.W. Rhind, 2001, *Geographic information systems and science*, Chichester: John Wiley & Sons, pp. 184-203.

-32- MacDonald, A., 1999, *Building a geodatabase*, Redlands: Environmental Systems Research Institute (ESRI).

-33- Magel, H., 2003, *Spatial information management for sustainable development*, Lecture at Surveyors Week 2003, Kingston, Jamaica, http://www.fig.net/news/news_2003/jamaica/jamaica_2003.htm.

-34- Marshall, C., 1999, *Enterprise modeling with UML: designing successful software through business analysis*, Reading: Addison Wesley Longman.

-35- Martin J. & J. Leben, 1989, *Strategic information planning methodologies*, Englewood Cliffs, Prentice-Hall, pp. 26-28.

-36- Microsoft Inc., 2003, Microsoft Office Visio Professional 2003 help menu and help website http://office.microsoft.com/nl-nl/assistance, vis. 09-2004 – 11-2004.

-37- Nebert, D, 2002, *SDI workshop 2002*, presentation at Lakewood SDI Workshop, http://www.fgdc.gov/clearinghouse/sdi/nebert1.ppt, vis. 09-2004 – 12-2004.

-38- Nebert, D.D., 2004, *The GSDI cookbook; developing spatial data infrastructures*, 2nd ed., Global Spatial Data Infrastructure secretariat, vis. 22-06-2004, http://www.gsdi.org.

-39- Nielsen, J., 1994, *Usability engineering*, San Francisco: Morgan Kaufmann, pp.17-37.

-40- Nielsen, J. & R.L. Mack, 1994, *Usability inspection methods*, New York: John Wiley & Sons, at http://www.useit.com/papers/heuristic/heuristic_list.html; www.useit.com, vis. 10-2004 – 12-2004.

-41- Object Management Group (OMG), 2001, *Model Driven Approach: A technical perspective*, OMG, http://www.omg.org/mda/presentations, in [42]

-42- Object Management Group (OMG), 2004, *About the Object Management Group*, OMG, vis. 19-10-2004, last update: 19-08-2004, http://www.omg.org.

-43- O'Neil, R., 2003, *Getting started building a spatial data infrastructure; the Canadian experience*, Santiago de Chile: Seminar on Spatial Data Infrastructures]

-44- Oosterom, P., W. Quak, T. Tijssen & E. Verbree, 2000, *The architecture of the geo-information infrastructure*, in: Proceedings of UDMS 2000: 22nd urban and regional data management symposium, Delft: Urban Data Management Society.

-45- Oosterom, P.J.M. van, 2001, *De geo-database als spin in het web*, Inaugurele rede, Delft: University of Technology.

-46- Oosterom, P.J.M. van, 2003, *Vision for the next decade of GIS technology; a research agenda for the TU Delft, the Netherlands*, Delft: University of Technology.

-47- OpenGeospatial Consortium (OGC), 2003, *OpenGIS abstract specification, topic 2: Spatial referencing by coordinates*, vers. 2, ed. R. Nicolai, ref nr. OGC 03-073r1, in [48]

-48- OpenGeospatial Consortium (OGC), 2004, http://www.opengeospatial.org, vis. 12-2004.

-49- Putra, S.Y., L. Wenjing & P.P. Yang, 2003, *Object-oriented GIS data modelling for urban design*, http://www.gisdevelopment.net/application/urban/overview/ma03093pf.htm, Map Asia, on http://www.gisdevelopment.net.

-50- Rajabifard, A., 2001, *SDI Hierarchy – from local to global SDI initiatives*, presentation on SDI Seminar, Tsukuba, Japan: University of Melbourne, Australia; http://www.gsi.go.jp/PCGIAP/tsukuba/seminar/presen_sdi.pdf on http://www.gsi.go.jp, Geographical Survey Institute of Japan]

-51- Royal Dutch / Shell Group press statement, 26-03-2003, *Exploration & Production and Gas & Power strategy presentation*, in [4]

-52- Royal Dutch Petroleum Company, 2004, Annual Reports and Accounts 2003 in [57].

-53- Rijkswaterstaat, Meetkundige Dienst, 2002, *Data management is metadata management: trendrapportage data management Rijkswaterstaat*, vis.04-11-2004, http://www.wadi.nl/datamanagementismetadatamgt.html.

-54- Shell GIS Technical Advisory Panel, 2003, *Spatial blueprint*, T. Gregorius.

-55- Shell GIS Technical Advisory Panel, 2004, *Geo-information strategy 2004-2008*, T. Gregorius, June 2004.

-56- Shell International Exploration & Production, 1998, *Spatial information management direction and strategy*, SIEP.98-5432.

-57- Shell International, 2004, http://www.shell.com, vis. 05-2004 – 12-2004.

-58- Stapleton, J., 2003, *DSDM: business focused development*, 2nd ed., London: Addison-Wesley.

-59- TNO Human Factors, 2002, http://www.usabilitytesting.nl, vis. 10-12-2004.

-60- Urban and regional information systems association (URISA), 1998, *Spatial information technology, standards and system integration*, ed. P. Croswell, Park Ridge: URISA.

-61- Warmer, J. & A. Kleppe, 2001, *Praktisch UML*, 2nd ed., Amsterdam: Addison-Wesley.

-62- Webster-dictionary, 2004, *Webster online dictionary*, Interapple Inc., vis. 7/2004 – 10/2004, http://www.webster-dictionary.org/definition. Sources used: Webster's 1913 Dictionary; WordNet Dictionary (WordNet 2.0, 2003); Dictionary of Computing (GNU: http://www.foldoc.org); Acronyms (GNU, 2002).

Abbreviations

DBMS	DataBase Management System
DGMS	Data Generating Management System
DSDM	Dynamic System Development Method
EDAM	Europe Data Acquisition & Management
ESRI	Environmental Systems Research Institute
EU-Seased	European Union Sea sedimentological website
FAME	Fixed And Mobile Entities
FAME-1	~ Version 1
FAME-2	~ Version 2
GII	Geographic Information Infrastructure
GIM	Geo-Information Management
GIS	Geographic Information System
HN	Hazard Notification
IBM	International Business Machines Corporation
IRM	Information Recourse Management
ISO	International Organisation for Standardisation
MBMS	Management System
MDA	Model Driven Architecture
MoSCoW	Must, Should, Could & Want to haves
MS Access	Microsoft Access
MS Excel	Microsoft Excel
MS Visio	Microsoft Visio
MS Word	Microsoft Word
NAM	Nederlandse Aardolie Maatschappij Dutch Oil Company
OGC	OpenGeospatial Consortium
OMG	Object Management Group
Oracle SDO	Oracle Spatial Data Option
PIM	Platform Independent Model
PSM	Platform Specific Model
QC	Quality Check / Quality Control
RAVI	Stichting Ravi Netwerk voor Geo-informatie 'Foundation Ravi Network for Geo-information'

SDI	Spatial Data Infrastructure
SDM	System Development Method
Shell EP Europe	Shell Exploration & Production Europe
SQL	Structures Query Language
SSL	Shell Standard Legend
TNO	Nederlandse organisatie voor Toegepast-Natuurwetenschappelijk Onderzoek ‘Dutch organisation for Applied-Scientific Research’
TNO-NITG	~ Nederlands Instituut voor Toegepaste Geowetenschappen ‘ ~ Dutch Institute for Applied Geo-sciences’
UID	Unique IDentifier
UML	Unified Modeling Language
WGS’84	World Geodetic System 1984
XMI	XML Metadata Interchange
XML	Extensible Mark-up Language

List of figures

List of tables

Tables

Boxes

Appendices

Appendix A: Coordinate reference systems and map projections

This section describes the basic principles of different map projections and coordinate reference systems and those that are used within Shell EP Europe. In section A.1 a brief introduction is given to coordinate reference systems. Section A.2 gives an overview of map projections. Section A.3 concludes with the coordinate reference systems and map projections used within Shell EP Europe.

A.1 Coordinate reference systems

Describing positions on the surface of the earth and defining their interdependence can be done by several geo-referencing methods; by name, local coordinates, order, global coordinates, topology or frames. [28]

Within geography related trades it has always been common to use (global) coordinates to describe positions on the earth. This can be done by three different methods [25 & 26]:

- Cartesian geocentric coordinates (X, Y, Z); using three perpendicular axes with the origin in the centre of the earth.
- Geographical coordinates (φ ,λ ,h); using angles and height, originating in the centre of the earth and measured from respectively the equator and the prime meridian.
- Planar (map) coordinates (x, y); using two perpendicular axes, being a planar projection of (a part of) the spheroidal earth.

The latter ones are commonly referred to as map projections, which are described in section A.2. The second ones, the geographical coordinates, are what usually is meant with the term coordinate reference systems. To be able to make use of a coordinate reference system, the origin and orientation of the systems axes' size and shape need to be defined. This set of parameters defining the coordinate reference system (mostly an ellipsoid) is called (geodetic) datum. [26]

A.2 Map projections

Since the earth resembles a spheroid, mostly geographical coordinates are used, when a three-dimensional reference system is desired. This representation of a spheroidal earth is also referred to as a globe. Though for using coordinates in a map, a planar projection is needed [25]. Such a planar projection is the representation of a spheroidal earth in two dimensions, as if a flat map sheet were wrapped around the earth's surface. Since that does not fit properly, the representation of the earth's surface on a plane will be distorted, either in shape, area, distance or direction. Trying to compensate for that different map projections will try to preserve different distortions. [26]

Map projections can be categorised by four main types, differentiating them by [26 & 28]:

- Shape of projection plane
- Projection aspect (rotation of the plane regarding the globe)
- Number of tangency lines (or point) from the plane on the globe
- Preservation of distortion.

The different shapes of the projection plane are: cones, cylinders and (regular) planes (figure A.2), giving respectively a conical, cylindrical and azimuthal projection. The latter can be distinguished into type of focus: the focal point can be either in the centre of the globe, at the opposite end of the globe or at an infinite point. This gives respectively a gnomonic, stereographic or orthographic projection (figure A.1). [26 & 28]

The position of the projection plane regarding the spheroid can differ. Such a position is called an aspect. The different projection aspects are: normal, transverse and oblique (figure A.2). The difference between the projection aspects has its consequences for position of the line(s) of tangency. [26 & 28]

The point or line of contact where the projection plane touches the spheroid is called the line (or point) of tangency. Along this line there is no distortion, and the closer to the line, the smaller the distortion will be. If the projection plane only touches the globe, there will be a line of tangency (or in case of an azimuthal projection, a point of tangency). This is called a tangent projection. If the projection plane cuts through the globe, there will be two lines where the projection plane touches the globe, and thus two lines with no distortion. Such a projection is called a secant projection. [26 & 28]

Since any projection of a globe onto a plane gives distortion, the aim of the projections is to minimise one or more of those distortions. From these projection characteristics a distinction can be made into: equal-area, equidistance, conformal, or with no distortion preservation. The equal-area projection preserves the surface-area, but distorts shape, angle and scale. The equidistance projection preserves the distances between points, but does not maintain scale. The conformal projection preserves local shape, though distorts size. [26 & 28]

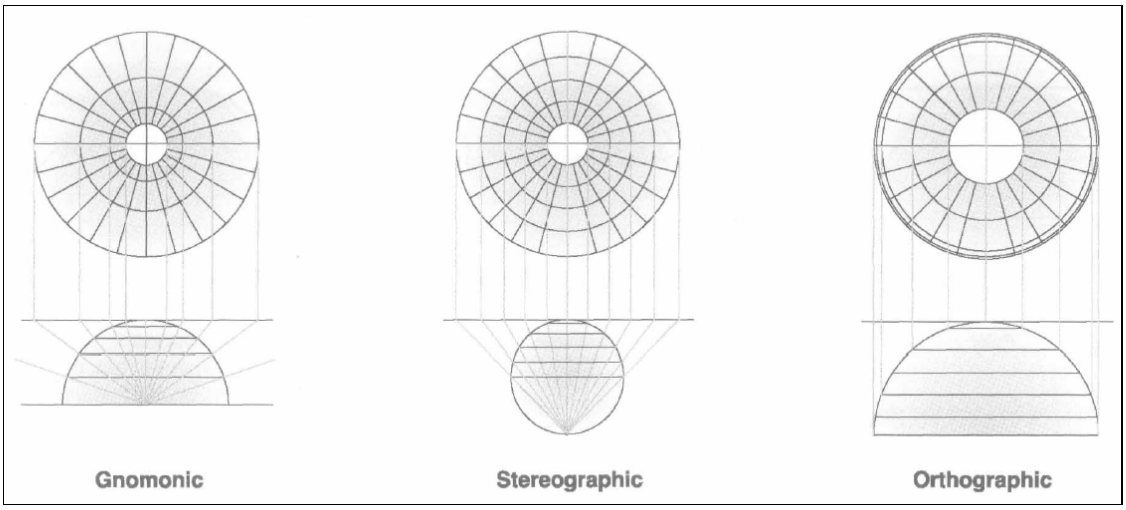


Figure A.1: Three types of azimuthal projections [26]

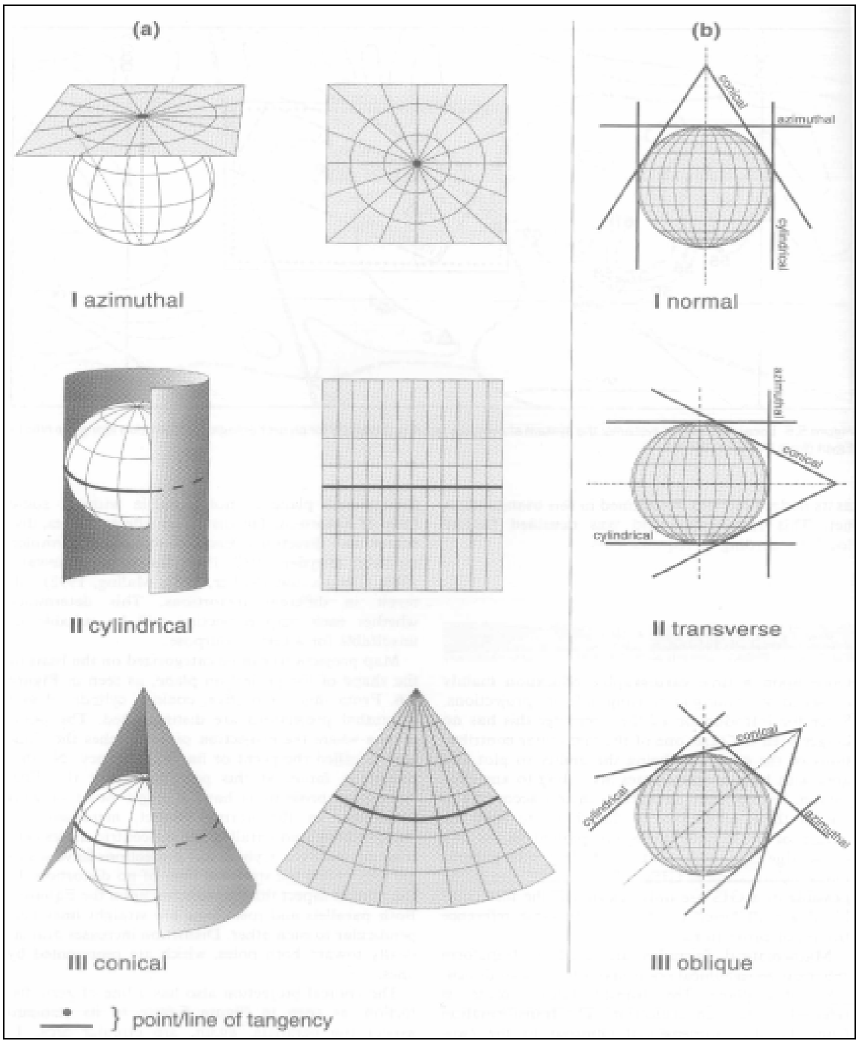


Figure A.2: Map projections: [28]
a) Projection plane: I azimuthal projection; II cylindrical projection plane; III conical projection plane.
b) Projection aspects: I normal aspect; II transverse aspect; III oblique aspect.

A.3 Coordinate reference systems and map projections in Shell EP Europe

Within Shell EP Europe different national operating units are merged. Every operating unit uses projection systems tailored to the locations or prescribed by the governments. And some have even different map projections for different parts of the country or continental shelf. The map offshore projections that are regularly used within Shell EP Europe can be seen in figure A4.

Concerning the (offshore) North Sea area, mainly three coordinate reference systems are used:
- Transverse Mercator projection at 0 degrees (TM0)
- Transverse Mercator projection at 5 degrees east (TM5)
- Universal Transverse Mercator projection at 3 degrees east (UTM31)

The first is used on the British continental shelf, the second on the Dutch continental shelf and the latter on the Norwegian part of the North Sea. All three of them have the same European Datum from 1950 (ED50) as reference datum.

The European Datum of 1950 was developed in the 1940s and is based upon the Hayford 1909 ellipsoid. It is a so-called locally best fitting ellipsoid, which means that it is the best fitting ellipsoid to approach the shape of Europe on the earth. Besides that, ED50 is the legal standard for operations in the North Sea. [17]

The Transverse Mercator projection, which is used by Shell EP Europe in the North Sea area, is a transverse cylindrical projection with the line of tangency along a meridian (also known as Gauss-Krüger projection). It is a conformal projection, thus preserving local angles. [26]

The Universal Transverse Mercator projection is a specially adapted type of the Transverse Mercator projection. Here the globe is divided into 60 north and south zones, each with a span of 6 degrees latitude and 8 degrees longitude, references by a letter along the meridians and a number along the parallels. [28] (See figure A5)

One coordinate reference system for storage in EP Europe

The new coordinate reference system that will be used for storage of point data in EP Europe is WGS'84. The 'World Geodetic System 1984' is a datum that defines a global reference frame. This frame, ellipsoid, tries to fit to the shape of the whole earth, rather than to fit to one region only. This datum is therefore used for worldwide referencing, e.g. by the Global Positioning System (GPS). Due to this global usability, and the link to GPS, it is a suitable choice for a unique coordinate reference system in EP Europe.

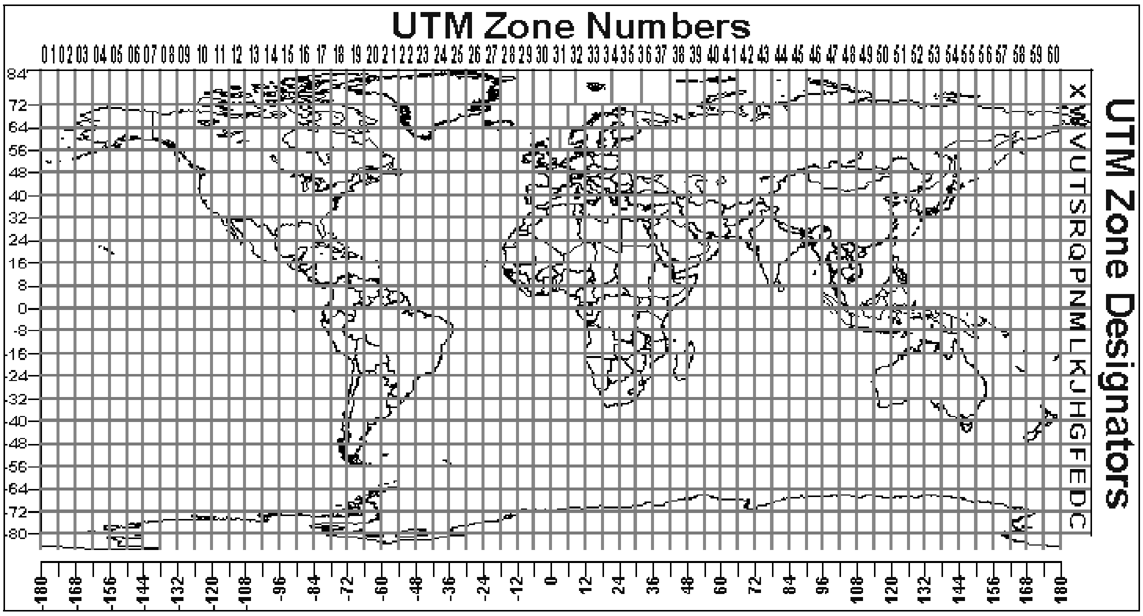


Figure A5: Universal Transverse Mercator projection [17]

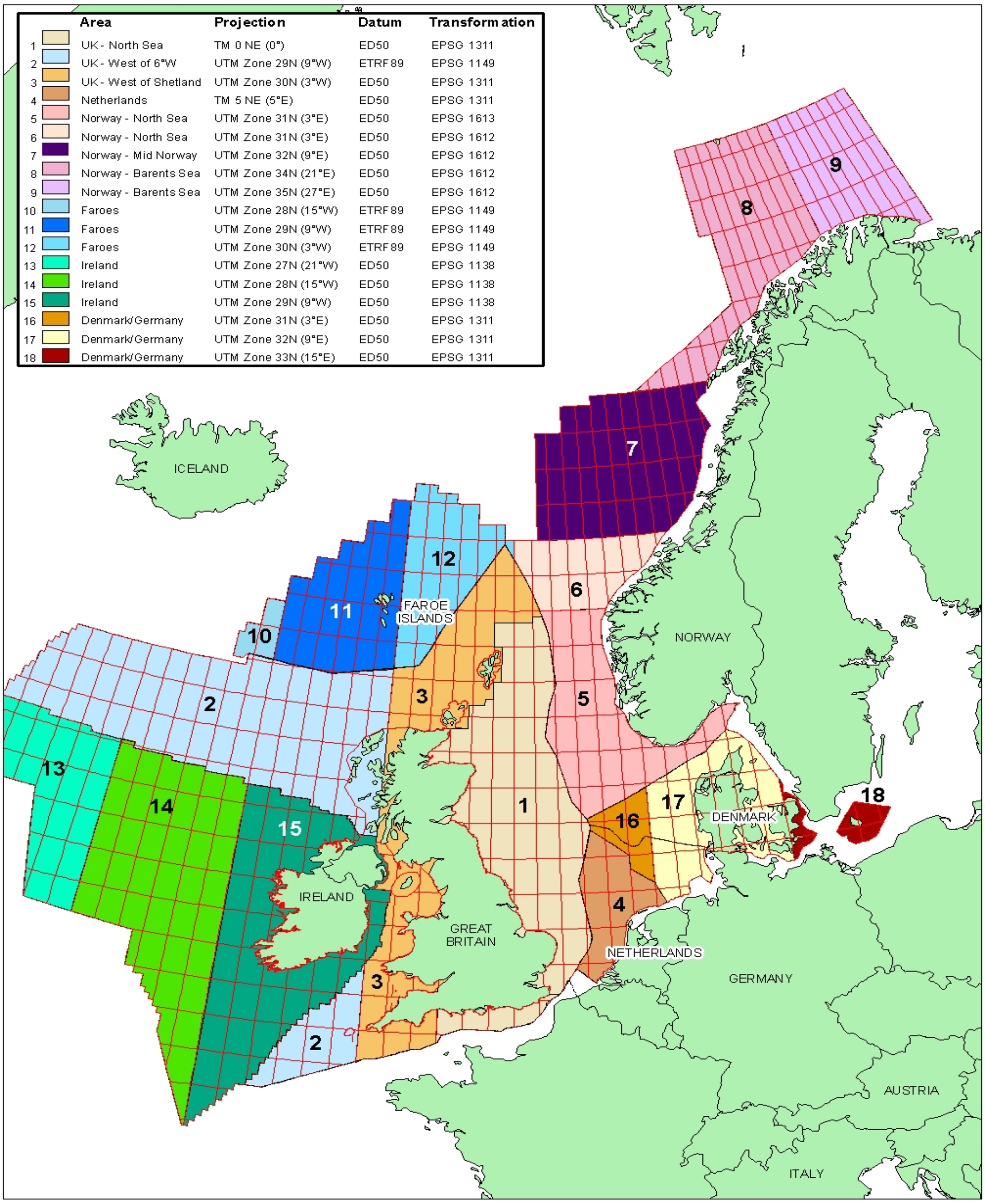


Figure A4: Shell EP Europe North-West Europe offshore coordinate reference systems [57]

Appendix B: Examples of the data inventory

This section holds examples of the datasets described mostly in the inventory of chapter 3. It is divided into firstly an overview of the geotechnical datasets in section B.1 and secondly an overview of the footprint datasets in section B.2.

B.1 Geotechnical data

This section gives an overview of the geotechnical datasets. Two of the datasets shown are from Shell Expro in the UK (figures B.1.1 and B.1.3). One is coming from NAM, in the Netherlands (figure B.1.2). The figures B.1.4 – B.1.7 hold examples of the documents that are linked to the geotechnical data-points at NAM.

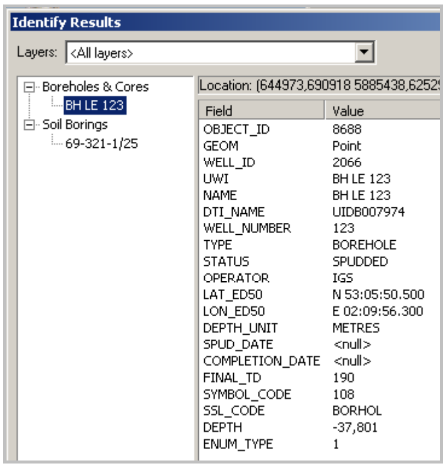


Figure B.1.1: Shell Expro geotechnical dataset, stored in Oracle, visualised via ArcGIS

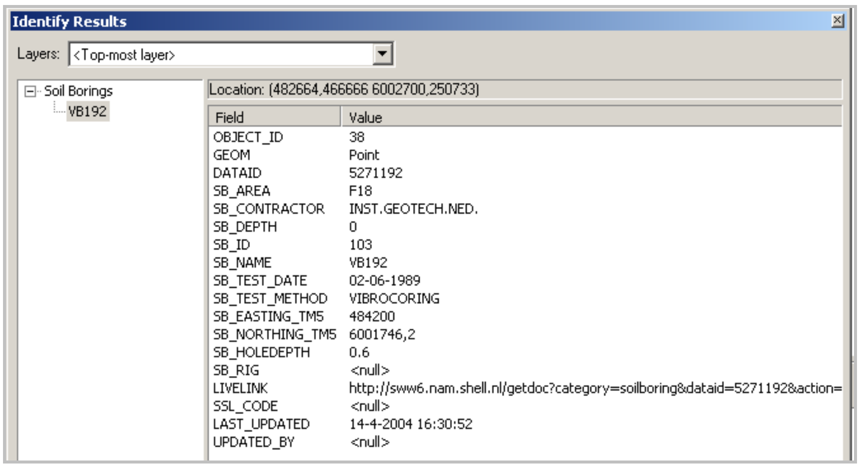
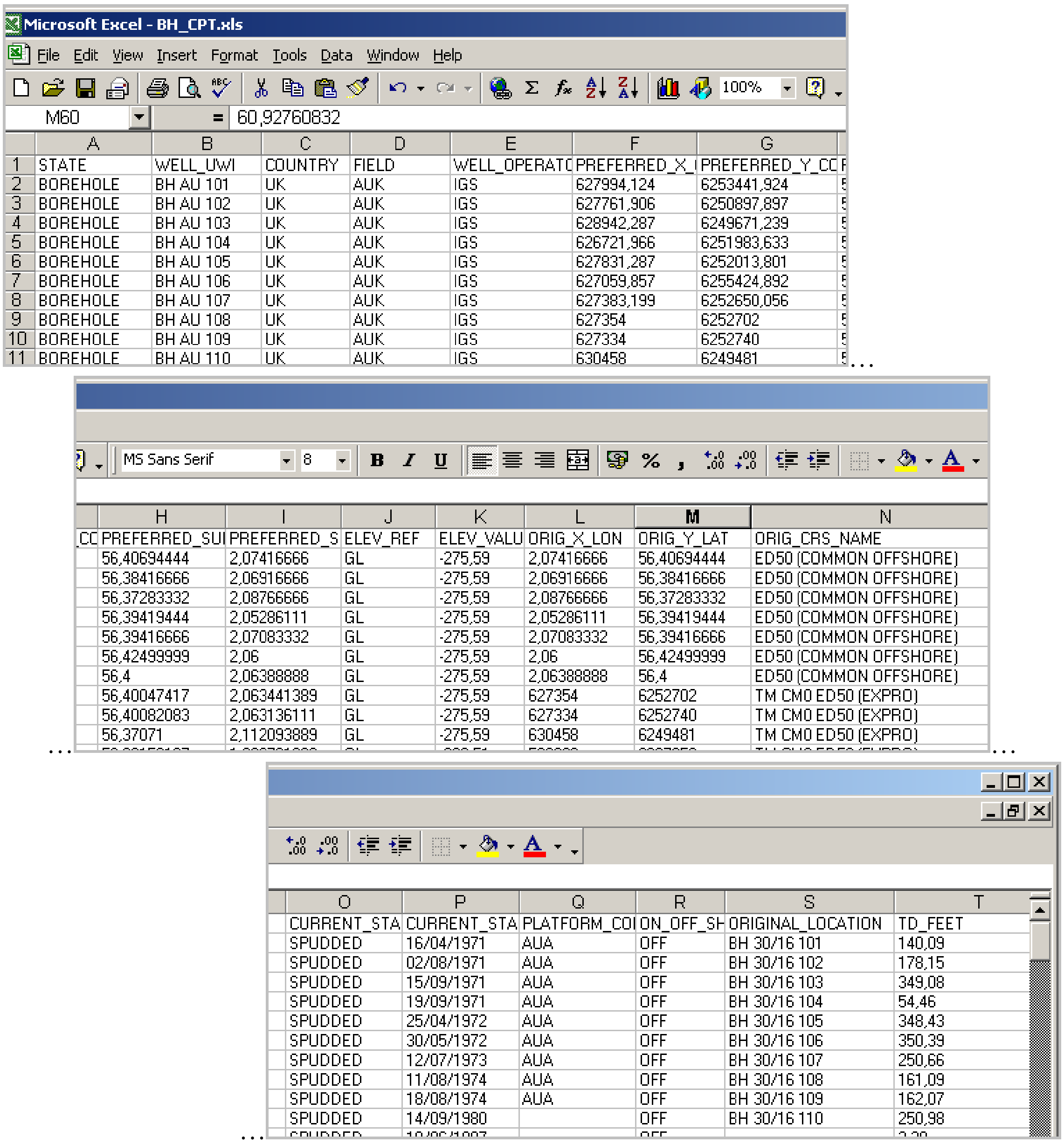


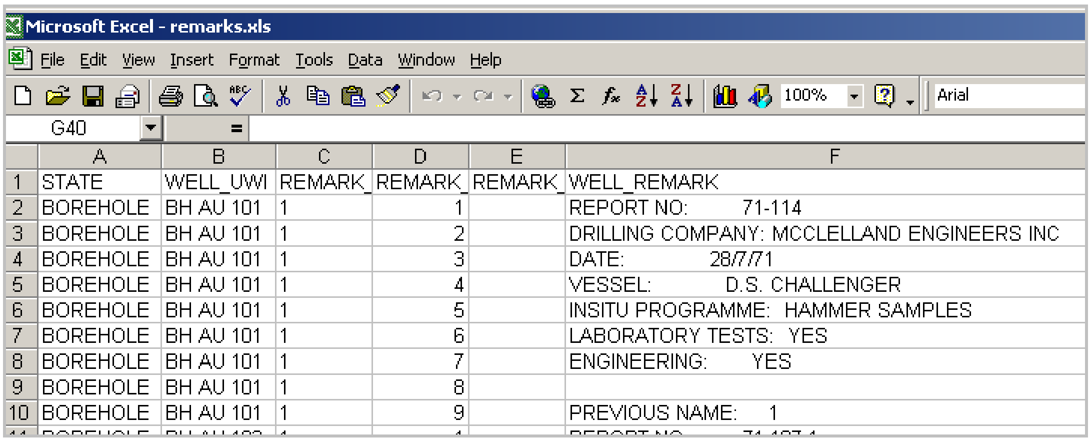
Figure B.1.2: NAM geotechnical dataset, stored in Oracle, visualised via ArcGIS



	A	B	C	D	E	F	G
1	STATE	WELL_UWI	COUNTRY	FIELD	WELL_OPERATC	PREFERRED_X_I	PREFERRED_Y_CC
2	BOREHOLE	BH AU 101	UK	AUK	IGS	627994,124	6253441,924
3	BOREHOLE	BH AU 102	UK	AUK	IGS	627761,906	6250897,897
4	BOREHOLE	BH AU 103	UK	AUK	IGS	628942,287	6249671,239
5	BOREHOLE	BH AU 104	UK	AUK	IGS	626721,966	6251983,633
6	BOREHOLE	BH AU 105	UK	AUK	IGS	627831,287	6252013,801
7	BOREHOLE	BH AU 106	UK	AUK	IGS	627059,857	6255424,892
8	BOREHOLE	BH AU 107	UK	AUK	IGS	627383,199	6252650,056
9	BOREHOLE	BH AU 108	UK	AUK	IGS	627354	6252702
10	BOREHOLE	BH AU 109	UK	AUK	IGS	627334	6252740
11	BOREHOLE	BH AU 110	UK	AUK	IGS	630458	6249481

	H	I	J	K	L	M	N
1	PREFERRED_SUI	PREFERRED_S	ELEV_REF	ELEV_VALU	ORIG_X_LON	ORIG_Y_LAT	ORIG_CRS_NAME
2	56,40694444	2,07416666	GL	-275,59	2,07416666	56,40694444	ED50 (COMMON OFFSHORE)
3	56,38416666	2,06916666	GL	-275,59	2,06916666	56,38416666	ED50 (COMMON OFFSHORE)
4	56,37283332	2,08766666	GL	-275,59	2,08766666	56,37283332	ED50 (COMMON OFFSHORE)
5	56,39419444	2,05286111	GL	-275,59	2,05286111	56,39419444	ED50 (COMMON OFFSHORE)
6	56,39416666	2,07083332	GL	-275,59	2,07083332	56,39416666	ED50 (COMMON OFFSHORE)
7	56,42499999	2,06	GL	-275,59	2,06	56,42499999	ED50 (COMMON OFFSHORE)
8	56,4	2,06388888	GL	-275,59	2,06388888	56,4	ED50 (COMMON OFFSHORE)
9	56,40047417	2,063441389	GL	-275,59	627354	6252702	TM CM0 ED50 (EXPRO)
10	56,40082083	2,063136111	GL	-275,59	627334	6252740	TM CM0 ED50 (EXPRO)
11	56,37071	2,112093889	GL	-275,59	630458	6249481	TM CM0 ED50 (EXPRO)

	O	P	Q	R	S	T
1	CURRENT_STA	CURRENT_STA	PLATFORM_COI	ON_OFF_SH	ORIGINAL_LOCATION	TD_FEET
2	SPUDDED	16/04/1971	AUA	OFF	BH 30/16 101	140,09
3	SPUDDED	02/08/1971	AUA	OFF	BH 30/16 102	178,15
4	SPUDDED	15/09/1971	AUA	OFF	BH 30/16 103	349,08
5	SPUDDED	19/09/1971	AUA	OFF	BH 30/16 104	54,46
6	SPUDDED	25/04/1972	AUA	OFF	BH 30/16 105	348,43
7	SPUDDED	30/05/1972	AUA	OFF	BH 30/16 106	350,39
8	SPUDDED	12/07/1973	AUA	OFF	BH 30/16 107	250,66
9	SPUDDED	11/08/1974	AUA	OFF	BH 30/16 108	161,09
10	SPUDDED	18/08/1974	AUA	OFF	BH 30/16 109	162,07
11	SPUDDED	14/09/1980		OFF	BH 30/16 110	250,98



	A	B	C	D	E	F
1	STATE	WELL_UWI	REMARK_	REMARK_	REMARK_	WELL_REMARK
2	BOREHOLE	BH AU 101	1	1		REPORT NO: 71-114
3	BOREHOLE	BH AU 101	1	2		DRILLING COMPANY: MCCLELLAND ENGINEERS INC
4	BOREHOLE	BH AU 101	1	3		DATE: 28/7/71
5	BOREHOLE	BH AU 101	1	4		VESSEL: D.S. CHALLENGER
6	BOREHOLE	BH AU 101	1	5		INSITU PROGRAMME: HAMMER SAMPLES
7	BOREHOLE	BH AU 101	1	6		LABORATORY TESTS: YES
8	BOREHOLE	BH AU 101	1	7		ENGINEERING: YES
9	BOREHOLE	BH AU 101	1	8		
10	BOREHOLE	BH AU 101	1	9		PREVIOUS NAME: 1

Figure B.1.3: Shell Expro geotechnical dataset, stored in OpenWorks, visualised via MS Excel

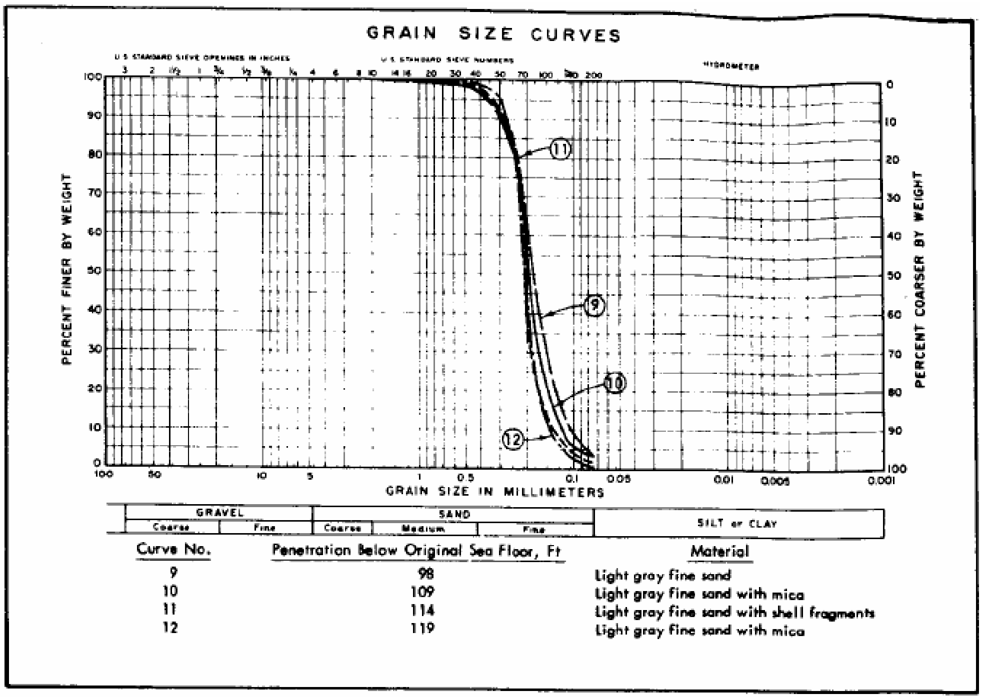


Figure B.1.4: Geotechnical site survey report: Grain size curve. Example of document, as attached to the NAM geotechnical dataset

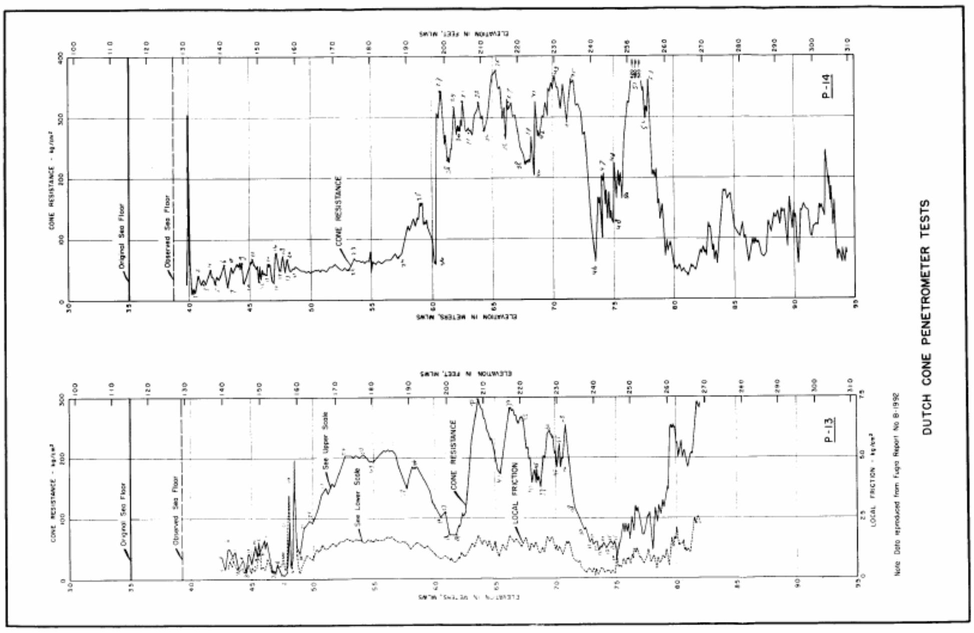


Figure B.1.5: Geotechnical site survey report: Cone penetrometer test (CPT). Example of document, as attached to the NAM geotechnical dataset

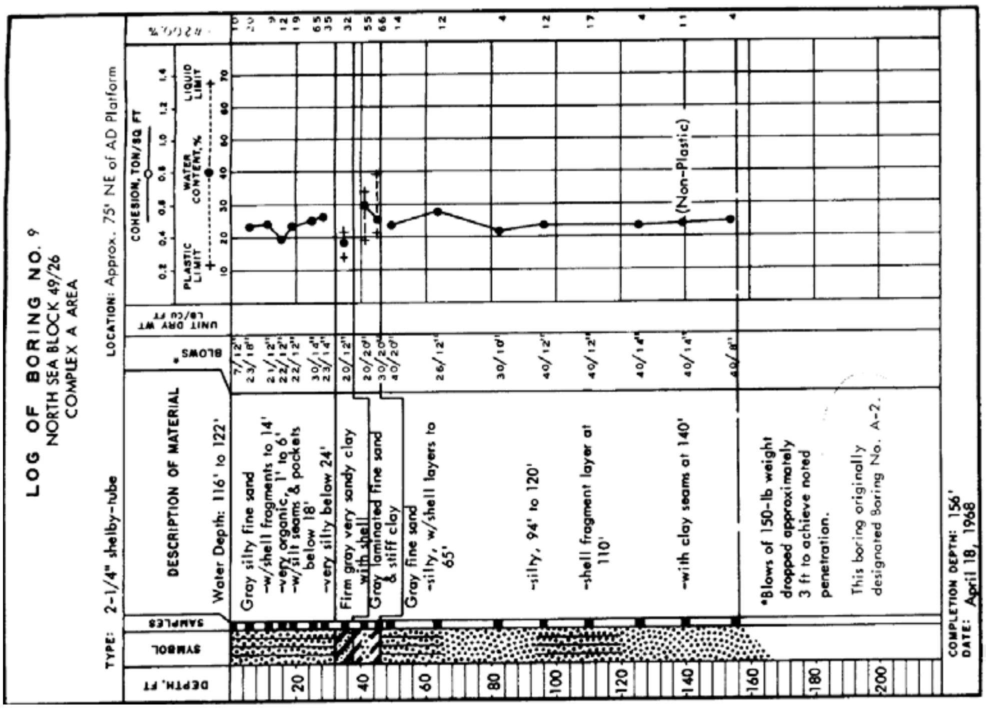


Figure B.1.6: Geotechnical site survey report: Boring log. Example of document, as attached to the NAM geotechnical dataset

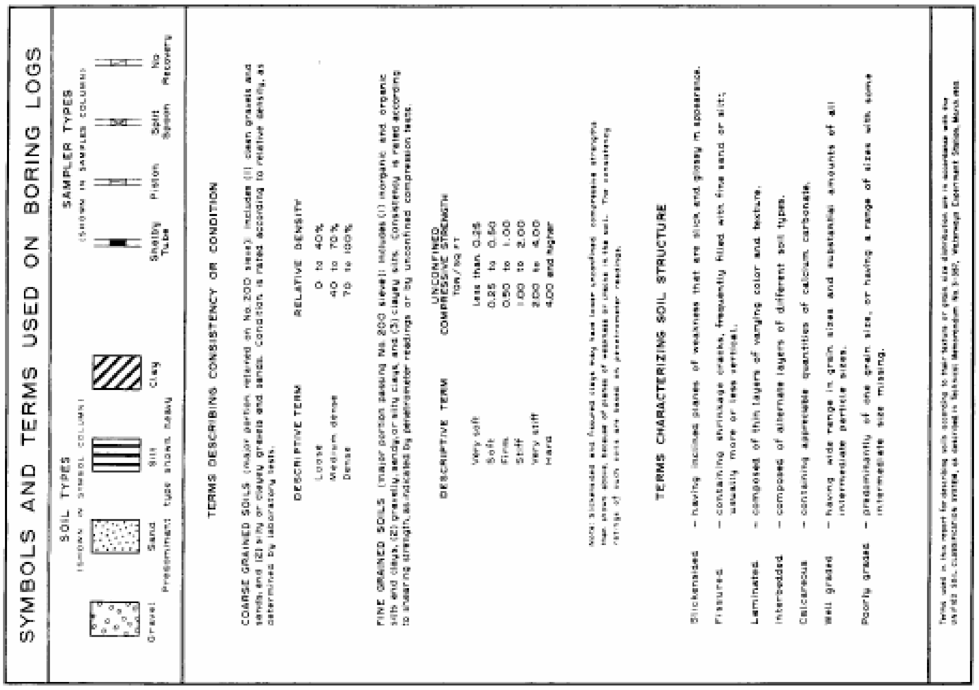


Figure B.1.7: Geotechnical site survey report: Legend for boring log (figure B.1.6). Example of document, as attached to the NAM geotechnical dataset

B.2 Footprint data

This section holds examples of the footprint datasets stored in NAM (figure B.2.4) and Shell Expro (figures B.2.5 – B.2.7). First the figures B.2.1 – B.2.3 try to give an impression of the shape and impacts the footprints have on the seabed. To conclude in table B.2.1 an overview is given of the data-attribute preferences that came out of the stakeholder engagement sessions conducted in Q4 2002.

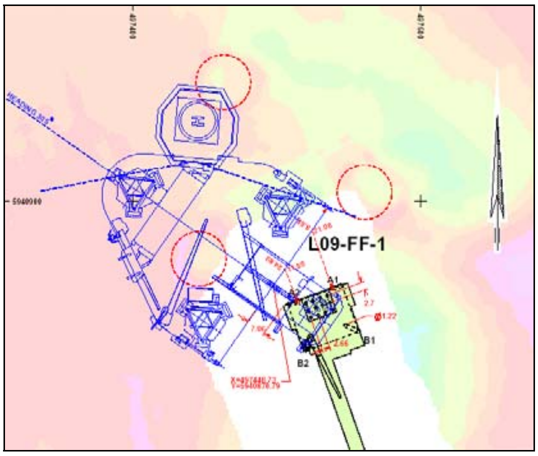


Figure B.2.2: Footprint imagery: Rig positioning map, with seabed surface contours, platform, rig and (former) footprints (as red circles)

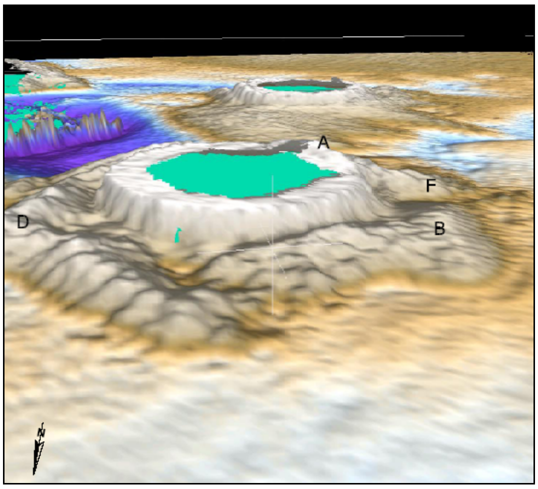


Figure B.2.3: Footprint imagery: 3-D impression of seabed surface around footprints

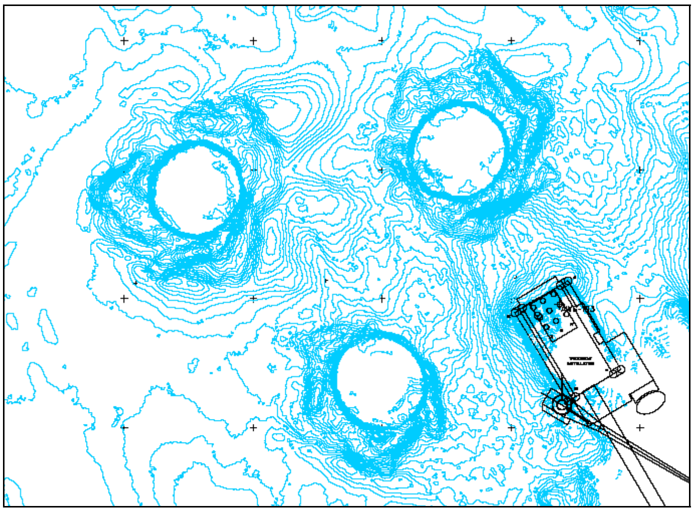


Figure B.2.3: Footprint imagery: Contour lines of seabed surface around footprints

BLOCK = A08			*Scour penetration ID= 11*		
Rig Name	***ENSCO 71***		Info	Scourpen Date	6/27/96
Block	A08			Scour Protection	NO
Well Name	A08- 1		Predicted Penetration	Minimum	1.5
Platform Name				Maximum	1.5
			Penetration	Minimum	1.3
				Maximum	1.3
Remarks	Very hard seabed at location,				

Figure B.2.4: NAM dataset, stored in MS Access, visualised via ASP (web)

Identify Results

Layers: <Top-most layer>

- Spudcan Footprints
  - SEAFOX 2

Location: (642691,106268 5884739,351976)

Field	Value
OBJ_ID	735
GEOM	Polygon
POINT_COORD_ID	735
NAME	SEAFOX 2
NAME_DATE	SEAFOX 2 06-JUL-94
SURVEY_DATE	6-7-1994
LEG_RADIUS	5,5
NO_LEGS	4
LATITUDE	N 53 05 30.965
LONGITUDE	E 002 07 50.589
LOCATION_DESCRIPTION	FINAL
START_DATE	6-7-1994 2:00:00
END_DATE	22-9-1994
GEOM.AREA	0
GEOM.LEN	0

Figure B.2.5: Shell Expro footprint dataset, stored in Oracle, visualised via ArcGIS

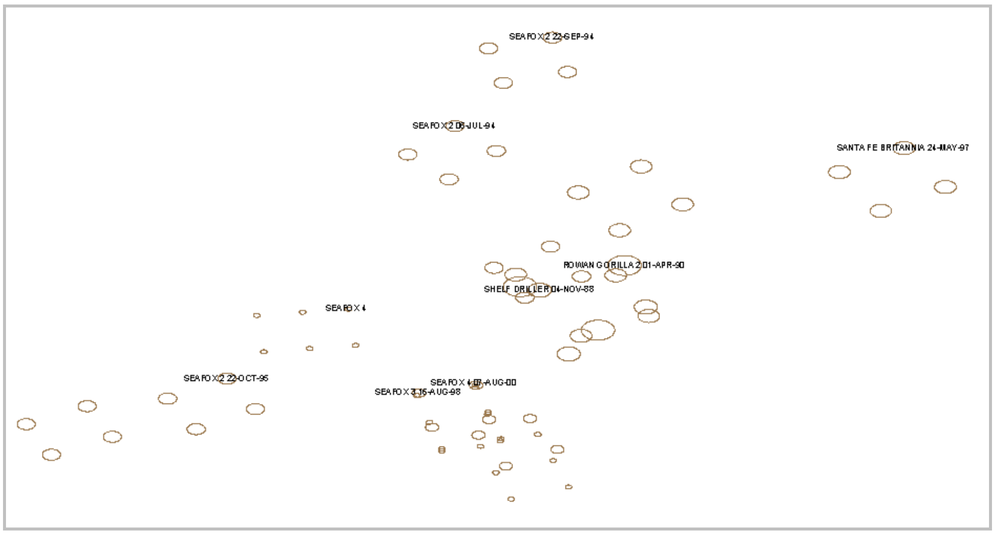


Figure B.2.7: Shell Expro footprint dataset visualised as map via ArcGIS

Table B.2.1: Stakeholders preferred footprint data-attribute and availability, created app. Q4 2002, based in [1]

Data-attribute	Availability of attribute in Shell Expro footprint dataset (Oracle)	Availability of attribute in NAM footprint dataset (MS Access)	Requirement as given by stakeholders
Drilling rig name	Yes	Yes	Required
Spudcan type	No	No	Required
Side view	No - Available in design files and PDFs	No - Available in design files and PDFs	Required
General layout of spudcans	No - Available in design files and PDFs	No - Available in design files and PDFs	Required
Dimensions (diameter)	Yes - Given as radius	No	Required
Spudcan shape (geometry)	No	No	Required
Block	Yes	Yes	Required
Platform name	Yes	Yes	Required
Well name	Yes	Yes	Required
No of legs	Yes	No	Required
Predicted spudcan penetration per leg (m)	No	No - Single figure quoted for rig	Required
Spudcan penetration per leg (m) (Leg naming convention: port 1, starboard 1, bow, etc.)	No	No - Single figure quoted for rig	Required
Pre-loads	No	No	Required
Spudcan penetration surveys (inspection data) performed by underwater engineering	No	No - Currently reports with Underwater dept. Started receiving copies	Required
Scour protection (gravel dump) per rig leg (in tonnage)	No	No - General statement Yes or No without qualifications	Required
Date of scour protection	No	No	Required
Period of rig on location (from [date] to [date])	Yes - Given as start and end date	No	Required
Recent bathymetry data of platform location	No - Available in design files or PDFs	No - Available in design files or PDFs	Required
Remarks	No	Yes	Required

Hyperlinks to:

Anchor / jack-up charts	No	No	Nice to have
Rig move reports	No	No	Nice to have
Water depth	No	No	Nice to have
JIM reports	No	No	Nice to have
Rock / gravel dump reports	No	No	Nice to have

Appendix C: Interviews and questionnaire

This section states the outcomes of the interviews and questionnaire held to aid in the analysis and inventory of the geotechnical & footprint datasets, data- and workflow, and user preferences. Section C.1 concerns the analysis and outcome of the interviews, while section C.2 describe the results of the questionnaire held amongst the stakeholders.

C.1 Interviews

In this section the outcomes of the interviews are presented. Table C.1.1 categorises the interviewees and state whether they are included in the questionnaire. In table C.1.2 the non-structured in-depth interviews are analyses and structured to reflect the issues concerning the data inventory and workflow, and the dataflow and CGII.

Table C.1.1: Interviewee characteristics

(Each number corresponds to one stake-holder)	Interviewee categories								Questionnaire
	Location		Discipline / department						Send (S) – Responded (R)
	NAM (Assen) / Shell Expro (Aberdeen) / Norske Shell (Stavanger)	Shell / Contractor	Offshore Surveys / Rig moves	Geotechnical Engineering	Pipeline Engineering	Geomatics	Marine Logistics	Onshore Surveys	
# 1	Norske Shell (Stavanger)	Shell			X				S
# 2	Shell Expro (Aberdeen)	Shell			X				S
# 3	NAM (Assen)	Shell					X		S – R
# 4	Shell Expro (Aberdeen)	Shell				X			
# 5	Shell Expro (Aberdeen)	Shell		X					S – R
# 6	Shell Expro (Aberdeen)	Shell	X						S – R
# 7	Shell Expro (Aberdeen)	Shell	X						
# 8	Shell Expro (Aberdeen)	Shell	X						
# 9	Shell Expro (Aberdeen)	Shell	X	X		X			
# 10	NAM (Assen)	Shell			X				S
# 11	NAM (Assen)	Shell	X						S – R
# 12	NAM (Assen)	Shell	X						
# 13	NAM (Assen)	Contractor	X	X					S
# 14	NAM (Assen)	Shell	X						S – R
# 15	Norske Shell (Stavanger)	Shell				X			S – R
# 16	NAM (Assen)	Shell						X	
# 17	Shell Expro (Aberdeen)	Shell	X						S
# 18	Shell Expro (Aberdeen)	Shell			X				S - R

Table C.1.2: Interview answers

Interviewee (Each number corresponds to one interviewee)	**Questions concerning the data inventory and workflow**			
	Which geotechnical and footprint data you currently use?	What purpose do you use the data for?	In what format you currently receive those data?	Which data you would like to have / use? (Which data are missing, which data are redundant?)
Person #1	CPTs + seabed slopes, debris and pipeline routing surveys. The last three items are also produced by contractors, who use in-house or subcontracted survey companies. They use side-scan sonar/EDM surveys to produce raw data, which is processed and issued in paper charts/alignment sheets.		Paper based reports used, these are usually based on seabed surveys (mainly CPTs) carried out by the construction company contracted to carry out the work (either directly or via a hydrographic survey subcontractor) the raw data is then processed by a third party (e.g. NGI) to provide the soils data used as the basis for design.	
Person #2	For a particular pipeline routing we generally run a number of sonar lines with the survey vessel to get a picture of the seabed bathymetry and sub-seabed conditions including any objects such as boulders, third party activity etc. over a 500 meter wide corridor. Based on the sonar results the vessel will then take a series of Vibrocores and CPT's along the route centreline to acquire soils data for engineering analysis as in most cases in the UK N.Sea we trench the pipelines below the seabed after installation			
Person #3	Except the cone penetrometer test I'm interested in all the listed items.		Have access via emails with attachments and to the LiveLink environment. Contacts in this are the Assen Offshore Survey team	

Table C.1.2: Interview answers … continued

…	Questions concerning the dataflow and CGII			Comments
In which format you would like?	What do you think of the idea to store the geotechnical and footprint data of the whole of EP Europe in one system? And to offer these data through the Intranet? Would you make use of it?	Who do you get the data from?	Who too uses the data, according to you?	
Access via the web is most convenient. We have been using LiveLink for our last project with some success. Documents have been uploaded to a central site, which can be accessed by selected contractors and the shell project team. This generally works fine but lacks some of the controls of the old paper based DCC system. The shortfalls can be made up and this will be a very good system in the future.	Seabed data is **very** site specific, the soil strength and type can vary a lot over a small area, so although it will be very useful to be able to consult a database for soils information in the field you are about to work in, it is likely that fresh data will be needed for each new location.			
		Have access via emails with attachments and to the LiveLink environment. Contacts in this are the Assen Offshore Survey team		

Table C.1.2: Interview answers … continued

Interviewee	**Questions concerning the data inventory and workflow**			
	Which geotechnical and footprint data you currently use?	What purpose do you use the data for?	In what format you currently receive those data?	Which data you would like to have / use? (Which data are missing, which data are redundant?)
Person #4				
Person #5			EnQuire is the predecessor of LiveLink and also contains geotechnical reports	
			The reports in the boxes at the Assen Offshore Survey team are (unofficial) extra made copies.	
			The predecessor of OpenWorks was Epigen. When data were copied from Epigen to OpenWorks, some data were lost.	
			Vibrocores reports are copied and stored at the Aberdeen geotechnical engineering team.	

Table C.1.2: Interview answers … continued

…	**Questions concerning the dataflow and CGII**			Comments
In which format you would like?	What do you think of the idea to store the geotechnical and footprint data of the whole of EP Europe in one system? And to offer these data through the Intranet? Would you make use of it?	Who do you get the data from?	Who too uses the data, according to you?	
			End user of geotechnical data is the Aberdeen Offshore Survey team, using them trough the hazard notifications. They also collect data for the Hazard Notifications.	Oracle was the corporate database structure, focussed on Well data, but was too complicated and wasn't working properly and is thus replaced.
			The main data users: are geophysical engineers, surveyors and the EPE Offshore Survey team, but project contractors as well	
			Concerning Norwegian data contact geophysics (EPE-T-PM or EPE-T-D)	

Table C.1.2: Interview answers … continued

Interviewee	**Questions concerning the data inventory and workflow**			
	Which geotechnical and footprint data you currently use?	What purpose do you use the data for?	In what format you currently receive those data?	Which data you would like to have / use? (Which data are missing, which data are redundant?)
Person #6	Mainly geophysical information	Geo-hazard detection; conjunctions; anchoring rigs; foundations for rigs and platforms, both shallow and deep; pipeline placements	The data is delivered as hard copy reports to the Aberdeen geotechnical engineering team. These are scanned and imported to LiveLink as whole documents, instead of only the needed images / graphs from the boreholes stored at the right position.	Mainly geophysical information
		Also responsible for surveying and data gathering both by Shell and contractors; could be convenient to have standardised listing for data gathering to put in contract	There are two data search bases for seabed surveys: ArcGIS (as background data) and hazard notification.	Most important is the location; where in space is it, and the visualization through ArcGIS. Added to it a preference for digital access to other data.
				The preferred data format is ArcGIS interface with links as attributes to the log files (graphs in pdf-format). No extra data is needed, because that is retrieved from the Hazard Notifications.
Person #7		Producing maps in ArcGIS of MicroStation (if wanted by contractor)	Input are shape-files from [the Aberdeen geotechnical engineering team]	Data is sufficient
		The location data is only used as background image	Access to the data is the main issue	

Table C.1.2: Interview answers … continued

…	Questions concerning the dataflow and CGII			Comments
In which format you would like?	What do you think of the idea to store the geotechnical and footprint data of the whole of EP Europe in one system? And to offer these data through the Intranet? Would you make use of it?	Who do you get the data from?	Who too uses the data, according to you?	
Most important is the location; where in space is it, and the visualization through ArcGIS. Added to it a preference for digital access to other data.				
The preferred data format is ArcGIS interface with links as attributes to the log files (graphs in pdf-format). No extra data is needed, because that is retrieved from the Hazard Notifications				

Table C.1.2: Interview answers … continued

Interviewee	**Questions concerning the data inventory and workflow**			
	Which geotechnical and footprint data you currently use?	What purpose do you use the data for?	In what format you currently receive those data?	Which data you would like to have / use? (Which data are missing, which data are redundant?)
# 7 … continued		Geo Hazard Detection: is a querying add-on for ArcGIS, it makes a buffer-query: every data from all layers within the variable radius of the buffer is selected and from these a preset selection of tables is written to a word-document, including a map and an image		
Person #8	The information in the OpenWorks database contains few data, a.o. Coordinate data, reference system, name, and depth.		OpenWorks is a database management system for all Well data. Initiated app. 4 years ago. At initiation all Borehole and Cone Penetration data were put in as well, as if it were a kind of Well data. But these data are not used nor maintained since, nor updated.	
	The coordinate data in the OpenWorks database are in UM0.			
Person #9				

Table C.1.2: Interview answers … continued

…	**Questions concerning the dataflow and CGII**			Comments
In which format you would like?	What do you think of the idea to store the geotechnical and footprint data of the whole of EP Europe in one system? And to offer these data through the Intranet? Would you make use of it?	Who do you get the data from?	Who too uses the data, according to you?	
	Probably there are no official Shell-conventions for naming soil sample data.			
Users have no clear idea which data they really need or miss. Haven't thought about whether they miss or require certain data.	Current situation works, although not ideally. The demand for one central database is clearly there. But any improvement is already considered as good.	Users have hardly an idea where data is coming from; rely on data to be correct.		

Table C.1.2: Interview answers … continued

Interviewee	Questions concerning the data inventory and workflow			
	Which geotechnical and footprint data you currently use?	What purpose do you use the data for?	In what format you currently receive those data?	Which data you would like to have / use? (Which data are missing, which data are redundant?)
Person #10	Jack-up rig spud can dimensions Footprints on location Rig plan Borehole logs CPT locations Scour patterns Seabed slopes Debris data Pipeline routing surveys (if linked)		Contractors are always asked to take along the core information on the route maps, to visualise the objects at the location.	Soil constellation (sand, clay, etc) Classification of soil Soil density Grain distribution Vibrocores CPTs
	We are mainly interested in ground/soil mechanic parameters, however these parameters are almost always present in the pipeline route survey reports.		We are mainly interested in ground/soil mechanic parameters, however these parameters are almost always present in the pipeline route survey reports.	
Person #11	In UK there are templates for the HN. In NL the layers should be selected by hand. Disadvantage of templates is that there are no fast updates possibles if for example one layer has to be renewed. Disadvantage of selection by hand is that if a scanned background layer is selected, the HN stocks		In UK there are templates for the HN. In NL the layers should be selected by hand. Disadvantage of templates is that there are no fast updates possibles if for example one layer has to be renewed. Disadvantage of selection by hand is that if a scanned background layer is selected, the HN stocks	

Table C.1.2: Interview answers … continued

…	**Questions concerning the dataflow and CGII**			Comments
In which format you would like?	What do you think of the idea to store the geotechnical and footprint data of the whole of EP Europe in one system? And to offer these data through the Intranet? Would you make use of it?	Who do you get the data from?	Who too uses the data, according to you?	
			Contractors are always asked to take along the core information on the route maps, to visualise the objects at the location.	
				First an outline is created. This can be a selected entity or a custom drew. To generate input, there should be officially logged on, to make sure that a QC can be done. This does not happen in NL.

Table C.1.2: Interview answers … continued

Interviewee	**Questions concerning the data inventory and workflow**			
	Which geotechnical and footprint data you currently use?	What purpose do you use the data for?	In what format you currently receive those data?	Which data you would like to have / use? (Which data are missing, which data are redundant?)
# 11 … continued			Except for this document there are shape- files and a geo-database are generated.	
Person #12	If any rigs have been placed there before. Spudcan debris Expected and actual rig leg penetration Structural assessments Survey report Rig anchor positions Water depths Bathymetry Gravel dumps	For an open location, a site-survey is done. For a location with already a platform a debris-survey is done.	The rig overviews and side-views are maintained in Mercator by the Offshore Surveys team.	
			In GIS the survey areas are stored with corner coordinates and a few attribute data.	

Table C.1.2: Interview answers … continued

…	**Questions concerning the dataflow and CGII**			Comments
In which format you would like?	What do you think of the idea to store the geotechnical and footprint data of the whole of EP Europe in one system? And to offer these data through the Intranet? Would you make use of it?	Who do you get the data from?	Who too uses the data, according to you?	
				From a custom defined radius a buffer is created around the entity. With this buffer an overlay is done on the selected map layers, from this automatically a report is generated in which a map (screen dump) is taken in together with a data dumping of all the attribute tables of the selected element of the layers that lay within the buffer.
Desired is to receive the data from dump-reports and to store them in a database, to get a bigger insight in them. These reports currently are created by others and perhaps also already maintained.	There are no set conventions for naming. Except for wells data, these come from the asset-register			In general the rig legs are geometrically invariant., therefore constraints are possible on the geometrical relation between the legs themselves. Furthermore it is desired to store the model of the foot
	There are no actual standard listings, but there exist workflow procedures, in which one and other is described (also known as A18 or EP 18 procedures)			

Table C.1.2: Interview answers … continued

Interviewee	**Questions concerning the data inventory and workflow**			
	Which geotechnical and footprint data you currently use?	What purpose do you use the data for?	In what format you currently receive those data?	Which data you would like to have / use? (Which data are missing, which data are redundant?)
Person #13		As contractor he is responsible for rig moves. His colleague is responsible for the positioning of production platforms. Both depend on the data they receive from the Shell EPE Offshore Surveys team, the Aberdeen geotechnical engineering team and the Shell Intranet.		Scour penetration data; footprint locations; soil data and structure
Person #14				The engagement session with the stakeholders, as described in appendix D of the FAME-project plan, is held approx. 1.5 years ago. The answers are given by the Aberdeen geotechnical engineering team & a rig move contractor. It is important that the given answers are validated and rechecked with the users, to be sure that these attributes are really necessary.

Table C.1.2: Interview answers … continued

…	**Questions concerning the dataflow and CGII**			Comments
In which format you would like?	What do you think of the idea to store the geotechnical and footprint data of the whole of EP Europe in one system? And to offer these data through the Intranet? Would you make use of it?	Who do you get the data from?	Who too uses the data, according to you?	
	There is no standard listing for the data. The delivered data are each time again appointed by expertise and consultancy. However a standardised listing should be convenient to avoid unclear situations..	As contractor he is responsible for rig moves. His colleague is responsible for the positioning of production platforms. Both depend on the data they receive from the Shell EPE Offshore Surveys team, the Aberdeen geotechnical engineering team and the Shell Intranet.		
	Important is a central coordination of the stored data; stored in Oracle flat tables and accessible via both ArcGIS and a Web-interface.		The engagement session with the stakeholders, as described in appendix D of the FAME-project plan, is held approx. 1.5 years ago. The answers are given by the Aberdeen geotechnical engineering team & a rig move contractor. It is important that the given answers are validated and rechecked with the users, to be sure that these attributes are	FAME is originally initiated for offshore infrastructure data, which are stored as point data. These will possibly be extended with all point data and in the future with line and area data.

Table C.1.2: Interview answers … continued

Interviewee	**Questions concerning the data inventory and workflow**			
	Which geotechnical and footprint data you currently use?	What purpose do you use the data for?	In what format you currently receive those data?	Which data you would like to have / use? (Which data are missing, which data are redundant?)
# 14 … continued				The European website for seabed soil data can be set up as a background or reference layer, comparable to option [1] KLIC of option [2] GBKN: Option [1] is using the data as a background layer / image that states if there are any objects in the vicinity of the event planned. If so, then the data owner can be contacted to acquire the data. Option [2] is to acquire the for EP Europe interesting data, strip them from all unneeded attributes and make them available as reference layer. In both options, the data layer should be updated regularly, either when the data change or following a set time interval.

Table C.1.2: Interview answers … continued

…	**Questions concerning the dataflow and CGII**			Comments
In which format you would like?	What do you think of the idea to store the geotechnical and footprint data of the whole of EP Europe in one system? And to offer these data through the Intranet? Would you make use of it?	Who do you get the data from?	Who too uses the data, according to you?	
			Involved parties are: Seabed survey team; Marine department Structural engineers Geotechnical engineers Underwater inspection	

Table C.1.2: Interview answers … continued

Interviewee	Questions concerning the data inventory and workflow			
	Which geotechnical and footprint data you currently use?	What purpose do you use the data for?	In what format you currently receive those data?	Which data you would like to have / use? (Which data are missing, which data are redundant?)
Person #15	As far as I know we have the following geotechnical data available: Draugen field: Site investigations from 1987. Boreholes and CPTs. Fugro. Pipeline route survey of Draugen Gas Export, 1998. CPT, samples, boreholes and deep water corals. Fugro. Pipeline route survey of Garn West and Rogn South, 1999. CPT, samples, boreholes and deep water corals. NGI reports. Ormen Lange: Transfer of data from Norsk Hydro to Norske Shell. Garn West & Rogn south: AutoCad files		As far as I know we have the following geotechnical data available: Draugen field: Site investigations from 1987. Boreholes and CPTs. Fugro. Pipeline route survey of Draugen Gas Export, 1998. CPT, samples, boreholes and deep water corals. Fugro. Pipeline route survey of Garn West and Rogn South, 1999. CPT, samples, boreholes and deep water corals. NGI reports. Ormen Lange: Transfer of data from Norsk Hydro to Norske Shell. Garn West & Rogn south: AutoCad files	
Person #16	The Onshore Surveys team does not make use of the stored geotechnical data. There is a government-owned database having more and more up-to-date data. According to the Onshore Surveys team, the onshore geotechnical data can be deleted or not migrated to a new system			The Onshore Surveys team does not make use of the stored geotechnical data. There is a government-owned database having more and more up-to-date data. According to the Onshore Surveys team, the onshore geotechnical data can be deleted or not migrated to a new system

Table C.1.2: Interview answers … continued

…	**Questions concerning the dataflow and CGII**			Comments
In which format you would like?	What do you think of the idea to store the geotechnical and footprint data of the whole of EP Europe in one system? And to offer these data through the Intranet? Would you make use of it?	Who do you get the data from?	Who too uses the data, according to you?	
		As far as I know we have the following geotechnical data available: Draugen field: Site investigations from 1987. Boreholes and CPTs. Fugro. Pipeline route survey of Draugen Gas Export, 1998. CPT, samples, boreholes and deep water corals. Fugro. Pipeline route survey of Garn West and Rogn South, 1999. CPT, samples, boreholes and deep water corals. NGI reports. Ormen Lange: Transfer of data from Norsk Hydro to Norske Shell. Garn West & Rogn		
		The Onshore Surveys team does not make use of the stored geotechnical data. There is a government-owned database having more and more up-to-date data. According to the Onshore Surveys team, the onshore geotechnical data can be deleted or not migrated to a new system		

C.2 Questionnaire

This section describes the outcomes of the questionnaire held amongst the stakeholders listed in table C.1.1. This questionnaire is held to gain knowledge about the users' preferences concerning the data-attributes to be stored in the geotechnical & footprint datasets. For this they were asked to scale the importance of the data-attributes on a scale of 1-4. This scale is derived from the MoSCoW-principle (see section 3.3), where 1 reflects the Must have's or being most important, and 4 stands for Want to have, or least important. Before the outcomes given in table C.2.2 and visualised by average in figures C.2.1 – C.2.3, these data-attributes are explained in table C.2.1.

Table C.2.1: Geotechnical and footprint data-attributes and explanation as used in the questionnaire

Geotechnical data-attribute	Notes & terminology
UID	Unique Identification number; for internal database referencing
Borehole name	
X coordinate	
Y coordinate	
EPSG code*	Standardised coordinate reference system coding, used by oil industry
Boring type/ method	Type of boring: borehole, CPT or vibrocore
Contractor	Name of contractor
Block code	Block number
Date	Date of boring
Depth (m)	Depth of borehole
SSL	Shell Standard Legend: standardised cartographic visualisation code
Report-link	Hyperlink to the LiveLink location of the borehole report. If there are multiple reports or links needed, please state it under remarks
Last updated	Date at which the data is last updated /altered
Updated by	Person who last updated / altered the data
Status	Status of the borehole, please state in remarks what possibilities there are and what they are needed for
EPIGEN code	Borehole code from EPIGEN, to be able to search by old data names
Water depth (m)	Water depth in metres
Remarks	
Country name	
Platform name	
Well name	

Table C.2.2: Footprint set data-attributes and explanation as used in the questionnaire

Footprint set data-attribute	Notes & terminology
Data item	
Footprint Set UID	Unique Identification number; for internal database referencing
Rig name	Name of the rig that
Start date	Start date of rig on position
Block name	Block number
End date	End date of rig on position
Spudcan type	Type of rig foot
SSL	Shell standard legend: standardised cartographic visualisation code
Well name	
Platform name	
Last updated	Date at which the data is last updated /altered
Updated by	Person who last updated / altered the data
Layout of spudcans	Report / drawing of rig foots layout
Spudcan penetration survey	Report of penetration survey
Bathymetry data	Report of bathymetry
Remarks	
Anchor / jack-up charts	Report anchor / jack-up charts
Rig move reports	Report of rig move
Water depth	Report of tide
JIM reports	Report of jack-up integrity
Rock gravel dump reports	Report of gravel and rock dump

Table C.2.3: Footprint data-attributes and explanation as used in the questionnaire

Footprint data-attribute	Notes & terminology
Data item	A rig can have 3, 4 or 6 legs, thus a minimum of 3 leg IDs is required and a maximum of 6 possible
Footprint UID	Unique Identification number; for internal database referencing
X coordinate	
Y coordinate	
EPSG code	Standardised coordinate reference system coding, used by oil industry
Footprint set UID	Spudcan identification number, to be able to link footprint data to each other via footprint set data
Leg name	Rig leg name, according to naming convention
Pred. penetration p. leg (m)	Predicted penetration per leg in metres
Penetration p. leg (m)	Actual penetration per leg in metres
Spudcan shape (radius)	Radius of the spudcan and the footprint
Pre-loads	
Scour protection (tonnage)	Total tonnage of scour protection
Date of scour protection	Date of scour protection
Scour protection type	Type of scour protection
Last updated	Date at which the data is last updated /altered
Updated by	Person who last updated / altered the data
Remarks	

Table C.2.4: Results of questionnaire: Geotechnical data-attribute importance scaling and percentage

Interviewee #	# 3		# 5		# 14		# 11		# 6		# 15		Average	
Geotechnical data	Scale	%	Scale	%	Scale	%	Scale	%	Scale	%	Scale	%	Scale	%
UID	2	0,75	1	1	4	0,25	4	0,25	1	1	1	1	2,17	0,71
Borehole name	4	0,25	1	1	1	1	1	1	2	0,75	1	1	1,67	0,83
X coordinate	2	0,75	1	1	1	1	1	1	1	1	1	1	1,17	0,96
Y coordinate	2	0,75	1	1	1	1	1	1	1	1	1	1	1,17	0,96
EPSG code	4	0,25	1	1	3	0,5	3	0,5	1	1	3	0,5	2,5	0,63
Boring type/ method	4	0,25	1	1	1	1	1	1	1	1	1	1	1,5	0,88
Contractor	4	0,25	2	0,75	3	0,5	2	0,75	2	0,75	2	0,75	2,5	0,63
Block code	4	0,25	1	1	3	0,5	2	0,75	2	0,75	2	0,75	2,33	0,67
Date	4	0,25	1	1	2	0,75	2	0,75	3	0,5	2	0,75	2,33	0,67
Depth	4	0,25	1	1	1	1	1	1	1	1	1	1	1,5	0,88
SSL	4	0,25	4	0,25	4	0,25	3	0,5	3	0,5	3	0,5	3,5	0,38
Report-link	4	0,25	1	1	1	1	1	1	3	0,5	3	0,5	2,17	0,71
Last updated	4	0,25	3	0,5	3	0,5	3	0,5	3	0,5	3	0,5	3,17	0,46
Updated by	4	0,25	3	0,5	3	0,5	4	0,25	4	0,25	4	0,25	3,67	0,33
Status	4	0,25	4	0,25	4	0,25	3	0,5	3	0,5	3	0,5	3,5	0,38
EPIGEN code	4	0,25	1	1	4	0,25	3	0,5	3	0,5	4	0,25	3,17	0,46
Water depth	2	0,75	1	1	1	1	2	0,75	3	0,5	3	0,5	2	0,75
Remarks	2	0,75	2	0,75	2	0,75	3	0,5	4	0,25	4	0,25	2,83	0,54
Country name	4	0,25	2	0,75	2	0,75	2	0,75	4	0,25	4	0,25	3	0,5
Platform name	2	0,75	2	0,75	1	1	1	1	2	0,75	4	0,25	2	0,75
Well name	2	0,75	2	0,75	2	0,75	1	1	2	0,75	4	0,25	2,17	0,71

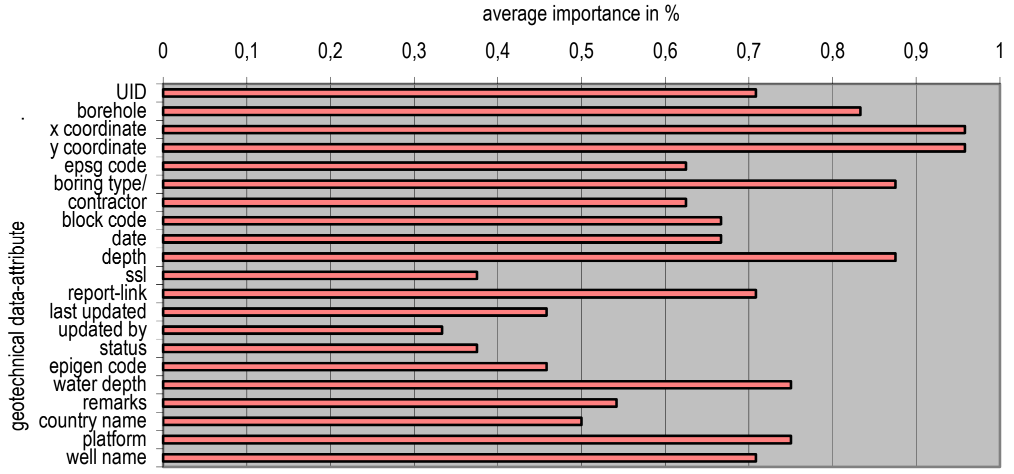
Table C.2.5: Results of questionnaire: Footprint set data-attribute importance scaling and percentage

Interviewee #	# 3		# 18		# 5		# 14		# 11		Average	
Footprint set data	Scale	%	Scale	%	Scale	%	Scale	%	Scale	%	Scale	%
Footprint set UID	3	0,5	1	1	1	1	4	0,25	4	0,25	2,6	0,6
Rig name	1	1	1	1	2	0,75	1	1	1	1	1,2	0,95
Start date	2	0,75	1	1	1	1	3	0,5	2	0,75	1,8	0,8
Block name	3	0,5	1	1	1	1	2	0,75	1	1	1,6	0,85
End date	2	0,75	2	0,8	1	1	3	0,5	2	0,75	2	0,75
Spudcan type	2	0,75	3	0,5	1	1	1	1	3	0,5	2	0,75
SSL	4	0,25	1	1	1	1	3	0,5	3	0,5	2,4	0,65
Well name	2	0,75	2	0,8	1	1	3	0,5	3	0,5	2,2	0,7
Platform name	2	0,75	1	1	1	1	2	0,75	1	1	1,4	0,9
Last updated	4	0,25	1	1	3	0,5	3	0,5	1	1	2,4	0,65
Updated by	4	0,25	4	0,3	3	0,5	3	0,5	4	0,25	3,6	0,35
Layout of spudcans	2	0,75	1	1	2	0,75	1	1	1	1	1,4	0,9
Spudcan penetration survey	2	0,75	1	1	1	1	1	1	2	0,75	1,4	0,9
Bathymetry data	2	0,75	2	0,8	2	0,75	1	1	2	0,75	1,8	0,8
Remarks	2	0,75	1	1	4	0,25	2	0,75	3	0,5	2,4	0,65
Anchor / jack-up charts	1	1	2	0,8	3	0,5	2	0,75	3	0,5	2,2	0,7
Rig move reports	2	0,75	1	1	2	0,75	1	1	3	0,5	1,8	0,8
Water depth	2	0,75	1	1	2	0,75	1	1	2	0,75	1,6	0,85
JIM reports	2	0,75	1	1	2	0,75	1	1	3	0,5	1,8	0,8
Rock gravel dump reports	1	1	1	1	2	0,75	2	0,75	4	0,25	2	0,75
Rig type *					1	1						
Previous rig visits *					1	1						
Site Investigation Report *					1	1						
Geophysical Site Survey report *					1	1						

Table C.2.6: Results of questionnaire: Footprint data-attribute importance scaling and percentage

Interviewee #	# 3		# 18		# 5		# 14		# 11		Average	
Footprint data	Scale	%	Scale	%	Scale	%	Scale	%	Scale	%	Scale	%
Footprint UID	2	0,75	2	0,8	1	1	3	0,5	4	0,25	2,4	0,65
X coordinate	3	0,5	2	0,8	1	1	1	1	1	1	1,6	0,85
Y coordinate	3	0,5	2	0,8	1	1	1	1	1	1	1,6	0,85
EPSG code	4	0,25	2	0,8	1	1	3	0,5	3	0,5	2,6	0,6
Spudcan UID	3	0,5	2	0,8	1	1	4	0,25	2	0,75	2,4	0,65
leg name	3	0,5	2	0,8	1	1	1	1	3	0,5	2	0,75
Pred. penetration p. leg	1	1	2	0,8	1	1	1	1	2	0,75	1,4	0,9
Penetration p. leg	1	1	1	1	1	1	1	1	1	1	1	1
Spudcan radius	2	0,75	2	0,8	1	1	1	1	1	1	1,4	0,9
Pre-loads	2	0,75	1	1	1	1	1	1	2	0,75	1,4	0,9
Scour protection tonnage	2	0,75	1	1	2	0,75	2	0,75	2	0,75	1,8	0,8
Date of scour protection	2	0,75	1	1	2	0,75	2	0,75	1	1	1,6	0,85
Scour protection type	2	0,75	1	1	2	0,75	2	0,75	1	1	1,6	0,85
Last updated	2	0,75	2	0,8	3	0,5	4	0,25	3	0,5	2,8	0,55
Updated by			2	0,8	3	0,5	4	0,25	3	0,5	3	0,5
Remarks			1	1	4	0,25	4	0,25	3	0,5	3	0,5

Figure C.2.1: Importance of geotechnical data-attributes



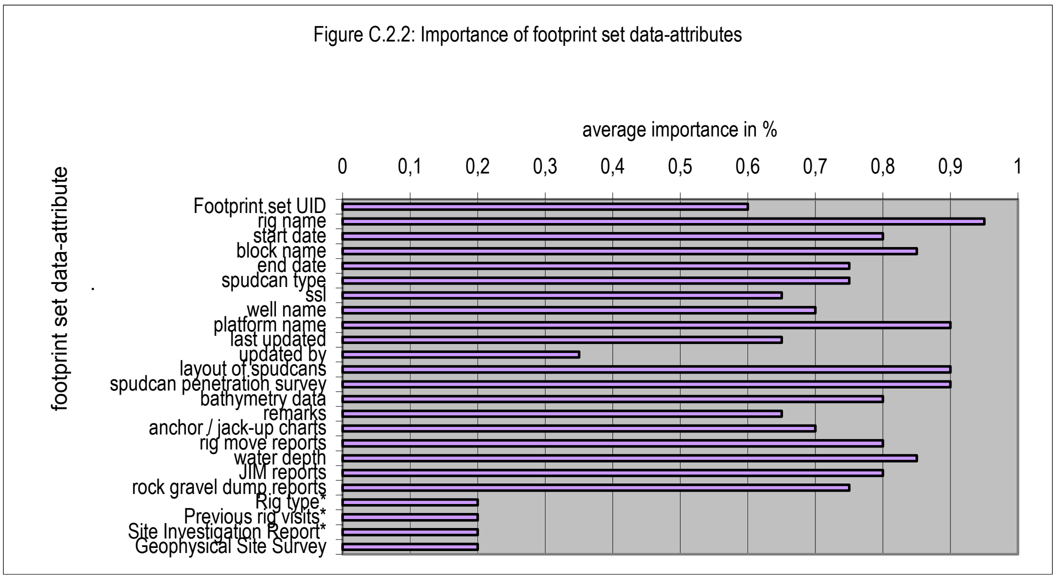


Figure C.2.2: Importance of footprint set data-attributes
average importance in %
0
0,1
0,2
0,3
0,4
0,5
0,6
0,7
0,8
0,9
1
footprint set data-attribute
Footprint set UID
rig name
start date
block name
end date
spudcan type
ssl
well name
platform name
last updated
updated by
layout of spudcans
spudcan penetration survey
bathymetry data
remarks
anchor / jack-up charts
rig move reports
water depth
JIM reports
rock gravel dump reports
Rig type*
Previous rig visits*
Site Investigation Report*
Geophysical Site Survey

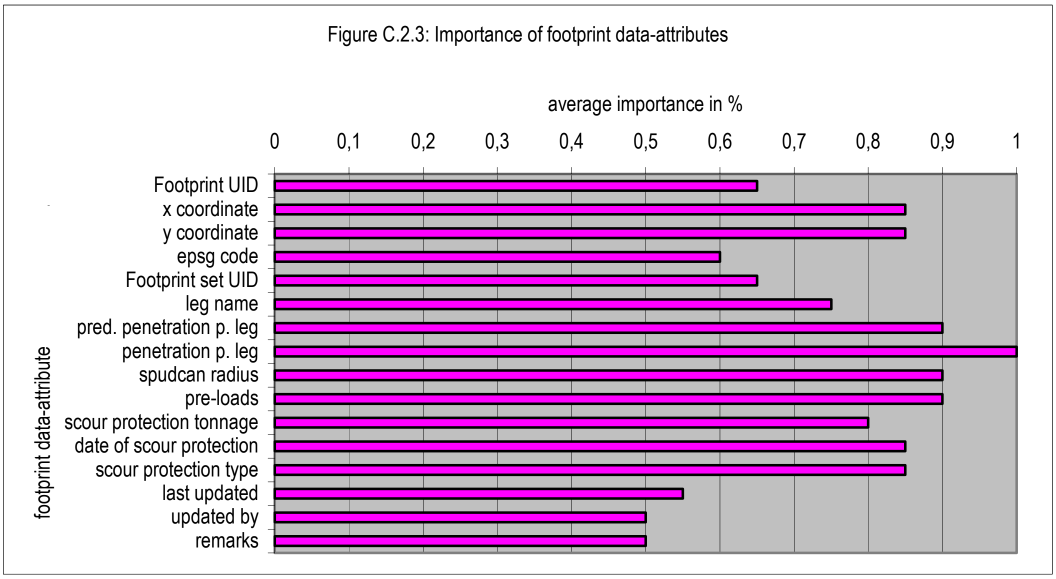


Figure C.2.3: Importance of footprint data-attributes
average importance in %
0
0,1
0,2
0,3
0,4
0,5
0,6
0,7
0,8
0,9
1
footprint data-attribute
Footprint UID
x coordinate
y coordinate
epsg code
Footprint set UID
leg name
pred. penetration p. leg
penetration p. leg
spudcan radius
pre-loads
scour protection tonnage
date of scour protection
scour protection type
last updated
updated by
remarks

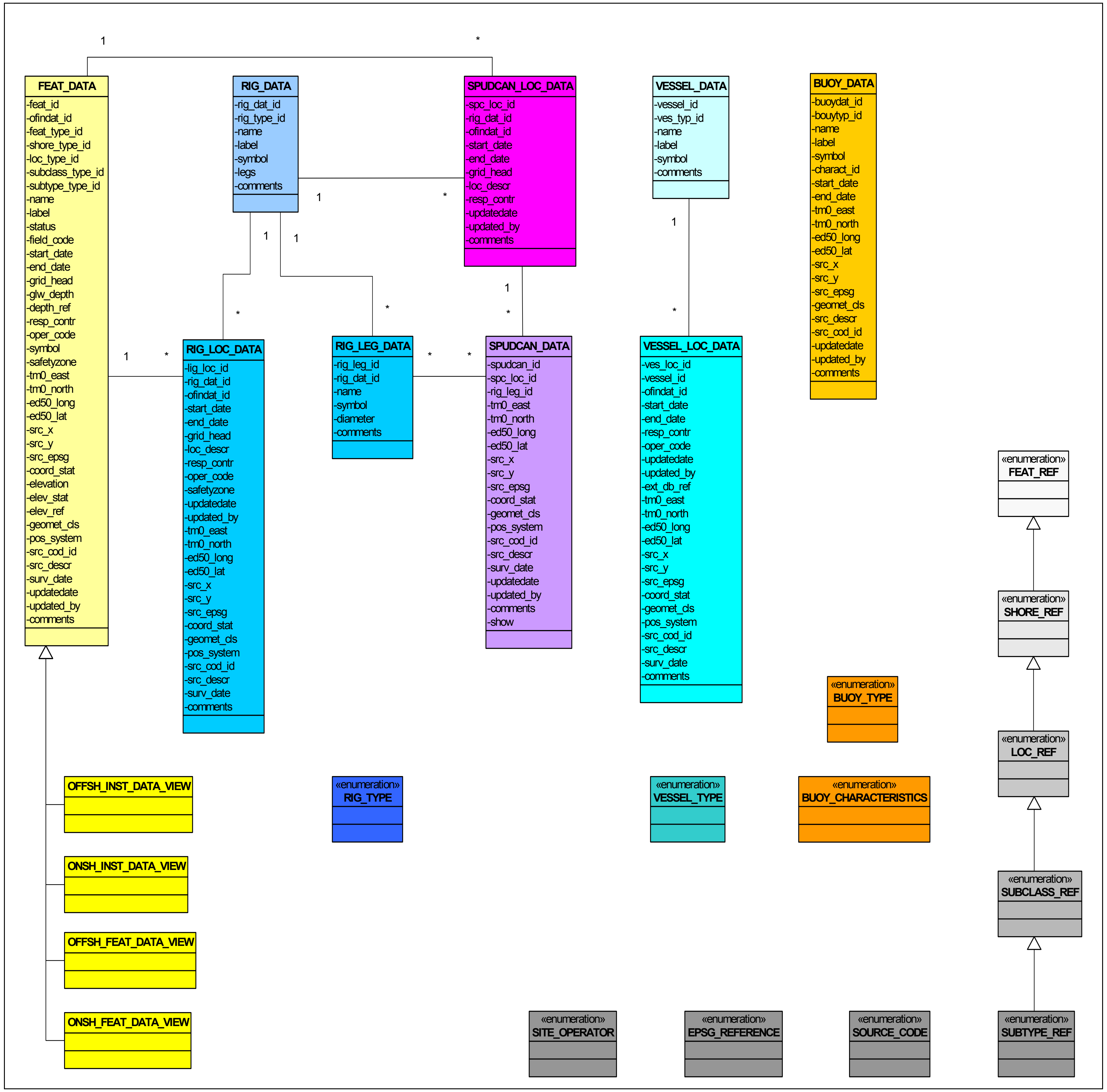
Appendix D: Data-models

Appendix D holds the data-models that are of influence on the geotechnical & footprint data. It is split into the FAME-1 data model in section D.1, the functional data-model of FAME-2 in section D.2, and the technical data-model of the geotechnical & footprint entities as they occur in the FAME-2 data-model.

D.1 Original FAME data model

This section shows the (original) FAME-1 data model, as it was used in Shell Expro. Although originally it was visualised in a different type of schema, here it is converted to and visualised as UML-class diagram. Due to the incomplete description of the original visualisation, the UML-diagram might show [onvolkomenheden]. This FAME-1 data-model is originally used in Shell Expro in the UK. This visualisation is converted from [1].

Figure D.1.1: Original FAME technical data-model, as used in Shell Expro, Aberdeen, visualised in UML class diagram, based on [1]



D.2 Conceptual models

This section shows different functional models of the FAME-2 data-model. Firstly there is the FAME-2 data-model as it is visualised originally in the project team (figure D.2.1). Figure D.2.2 on the other hand shows the same data-model, but now as UML-schema, figure D.2.3 gives a variation on the UML-schema given in figure D.2.2. This variation is based on the ambiguity described in section 4.4.

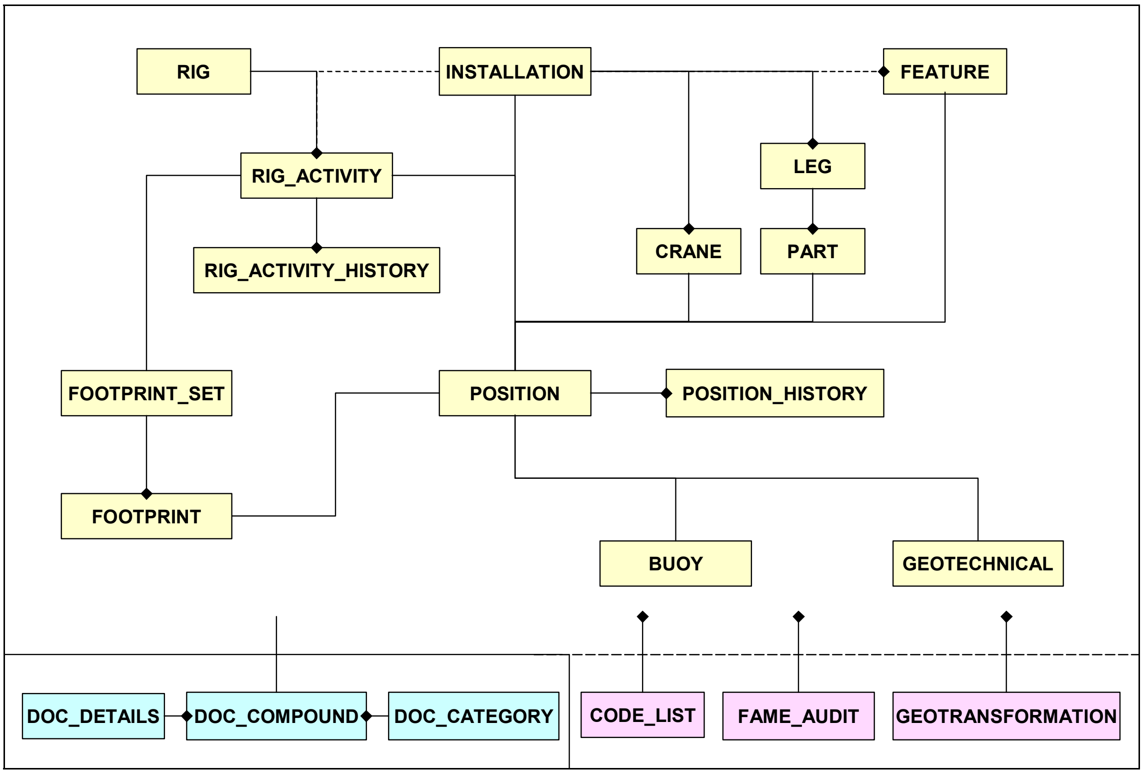


Figure D.2.1: FAME-2 conceptual model [11]

Figure D.2.2: FAME-2 conceptual model, visualised in UML, based on [11]

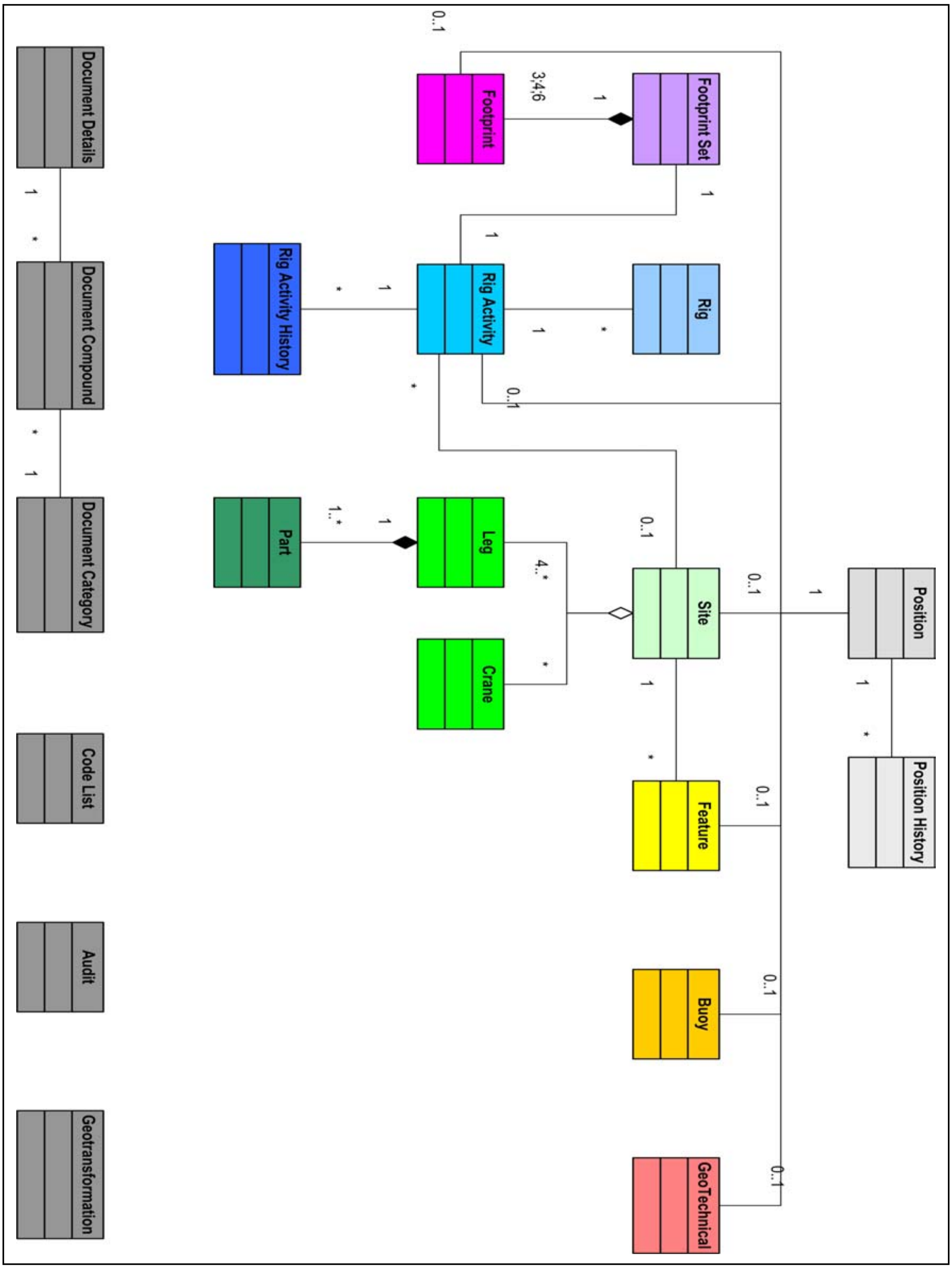
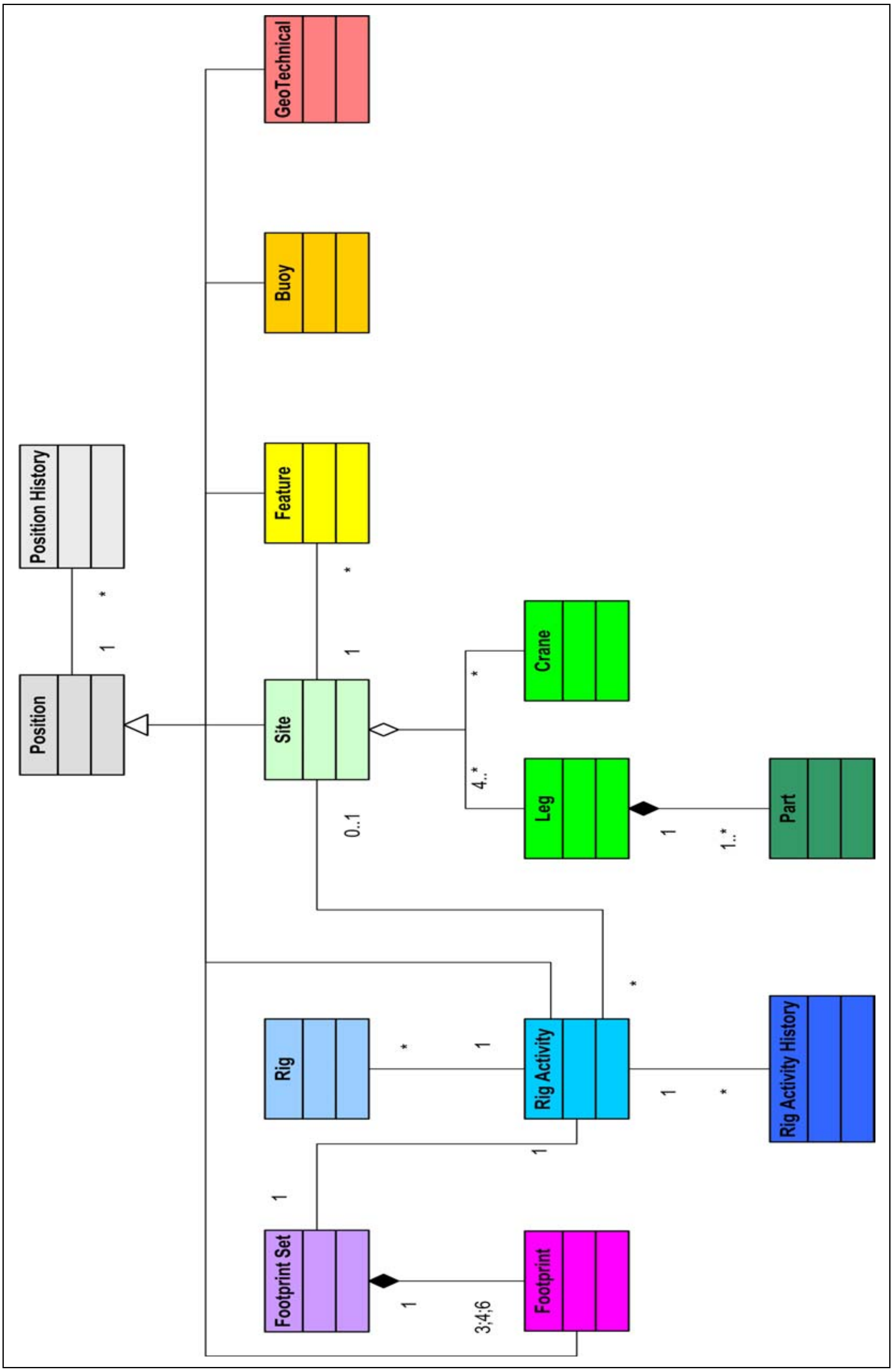


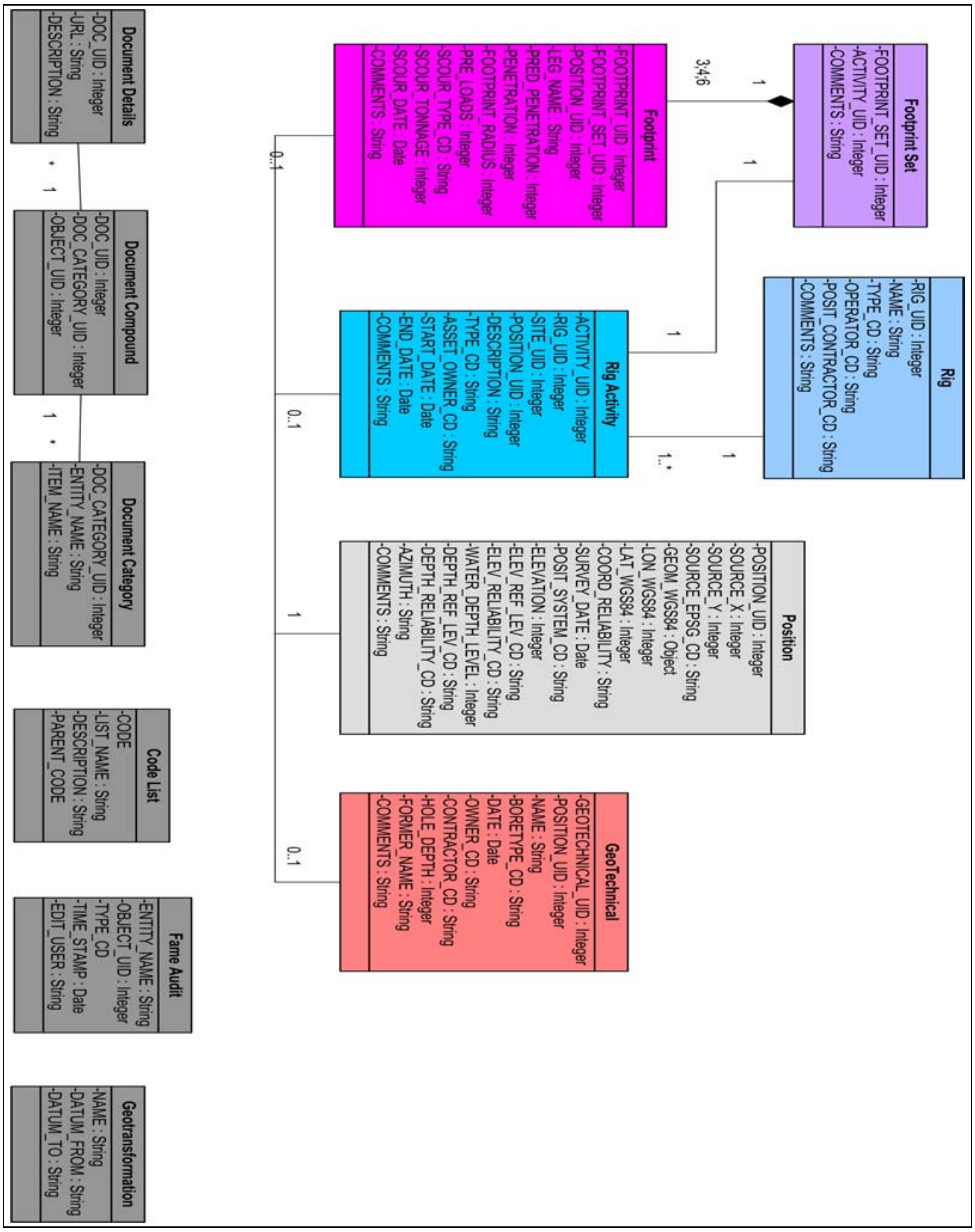
Figure D.2.3: FAME-2 conceptual model, visualised in UML, based on [11]; with additional abstract metaclass ‘Entity’



D.3 Geotechnical & footprint technical model

In this section the technical data-model of the geotechnical sites & footprints is given. In the figure D.3.1, these are elaborated with the FAME-2 entities directly related to them. The whole of figure D.3.1 can be seen as a representation of a specific selection of the FAME-2 technical data-model.

Figure D.3.1: Technical model of geotechnical and footprint data, in relation to the linked FAME-2 datasets.



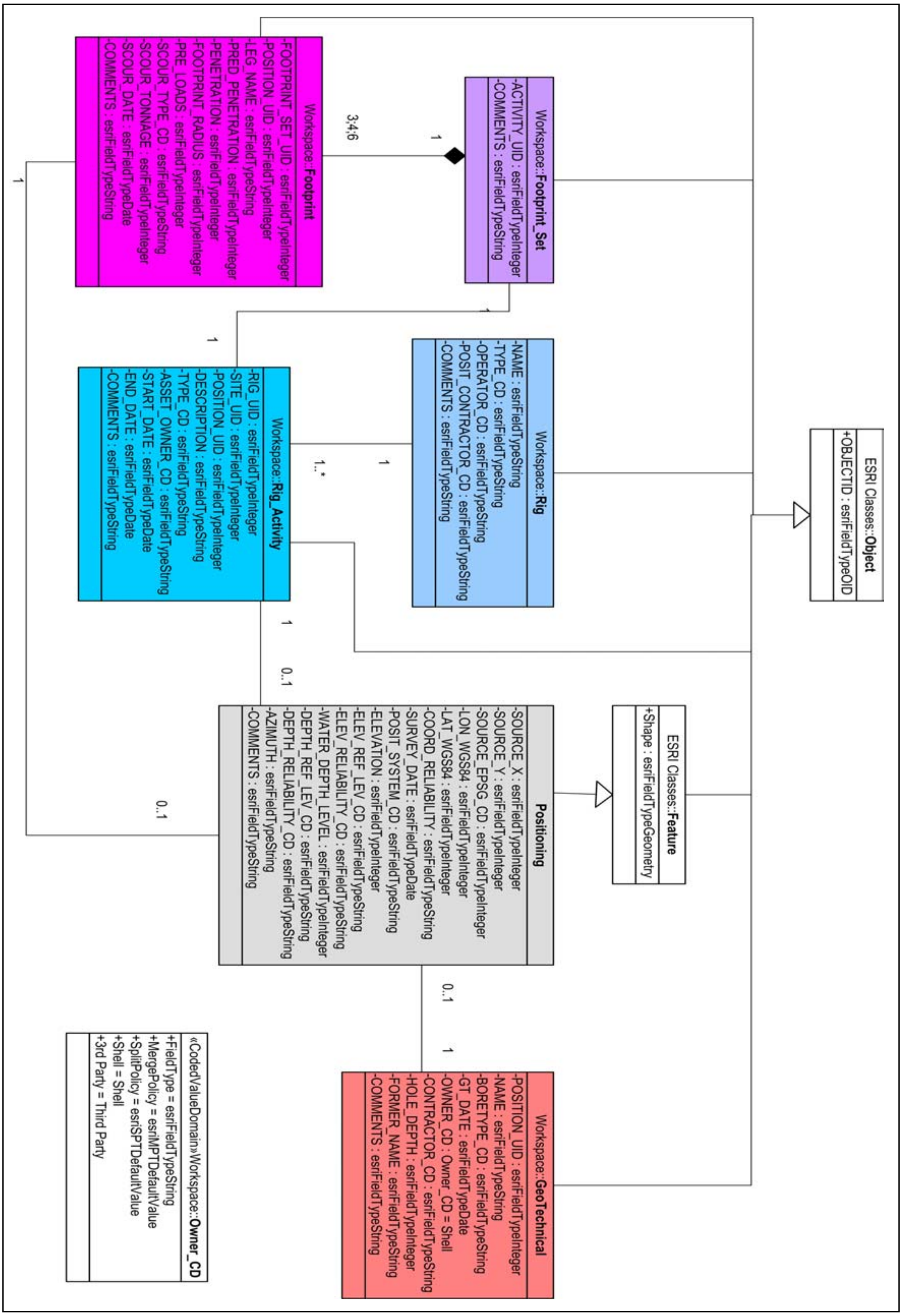
Appendix E: UML to database conversion test results

This section gives the results of and input used in the conversion testing. In this testing the possibilities for automatic database creation from UML-schemas is looked at. Firstly section E.1 gives the UML-schema, which was input for the test. Section E.2 states a specific part of the XMI-file, which was the in-between result. To conclude, section E.3 holds the log-file that came with the end result.

E.1 UML test model

This section shows the UML-schema that was used in the testing to convert this schema to a database. This schema is based on the geotechnical & footprint technical data-model that is given in figure D.3.1. It is, however, adapted to fit the rules given by the ESRI-template, which was required for the conversion functionality.

Figure E.1.1: Technical model of geotechnical and footprint data, as used in the UML to database conversion.



E.2 Selection of the XMI-file as visualisation of the in-between results

This section gives a selection of the XMI-file. The creation of this file is an in-between step in the conversion from UML to database. The selection given concerns Association 7, which is the association between the footprint set and the footprint entities. This selection is chosen, since it holds a special constraint in the relation. Besides being a composite relation, it holds the constraint that there should always be either three, four or six footprints in a footprint set. This constraint however, is lost in the final database result. Since it is still stated here (see highlight) it can be concluded that it is respected in the conversion from UML to XMI, and that it is lost in the conversion from XMI to database.


```xml
<?xml version="1.0" encoding="UTF-16" ?>
  <!DOCTYPE XMI (View Source for full doctype...)>
- <XMI xmi.version="1.0">
    - <XMI.header>
        - <XMI.documentation>
              <XMI.exporter>Microsoft Visio</XMI.exporter>
              <XMI.exporterVersion>1.0</XMI.exporterVersion>
          </XMI.documentation>
          <XMI.metamodel xmi.name="UML" xmi.version="1.1" />
      </XMI.header>
    - <XMI.content>
[...]
- <Foundation.Core.Association xmi.id="UID7675F755-AB2D-466A-A69D-975AC851CB42">
    <Foundation.Core.ModelElement.name>Association7</Foundation.Core.ModelElement.name>
    <Foundation.Core.ModelElement.visibility xmi.value="public" />
     <Foundation.Core.GeneralizableElement.isRoot xmi.value="false" />
      <Foundation.Core.GeneralizableElement.isLeaf xmi.value="false" />
      <Foundation.Core.GeneralizableElement.isAbstract xmi.value="false" />
    - <Foundation.Core.ModelElement.namespace>
          <Model_Management.Package xmi.idref="UIDE2D5992F-13D6-4D21-A9A8-0B59F3302DB6" />
      </Foundation.Core.ModelElement.namespace>
    - <Foundation.Core.ModelElement.taggedValue>
        - <Foundation.Extension_Mechanisms.TaggedValue xmi.id="UIDCEA9CA97-B3EE-43B6-8E6B-1DE3D26A165E">
              <Foundation.Extension_Mechanisms.TaggedValue.tag>OriginPrimaryKey</Foundation.Extension_Mechanisms.TaggedValue.tag>
              <Foundation.Extension_Mechanisms.TaggedValue.value>OBJECTID</Foundation.Extension_Mechanisms.TaggedValue.value>
          </Foundation.Extension_Mechanisms.TaggedValue>
      </Foundation.Core.ModelElement.taggedValue>
    - <Foundation.Core.ModelElement.taggedValue>
        - <Foundation.Extension_Mechanisms.TaggedValue xmi.id="UID9109EB9B-4A36-4B29-8029-030071C592FE">
              <Foundation.Extension_Mechanisms.TaggedValue.tag>OriginForeignKey</Foundation.Extension_Mechanisms.TaggedValue.tag>
              <Foundation.Extension_Mechanisms.TaggedValue.value>FOOTPRINT_SET_UID</Foundation.Extension_Mechanisms.TaggedValue.value>
          </Foundation.Extension_Mechanisms.TaggedValue>
      </Foundation.Core.ModelElement.taggedValue>
    - <Foundation.Core.ModelElement.taggedValue>
        - <Foundation.Extension_Mechanisms.TaggedValue xmi.id="UID68852AF9-9465-4EDE-8410-61EB4C03AA7C">
              <Foundation.Extension_Mechanisms.TaggedValue.tag>Notification</Foundation.Extension_Mechanisms.TaggedValue.tag>
              <Foundation.Extension_Mechanisms.TaggedValue.value>esriRelNotificationBoth</Foundation.Extension_Mechanisms.TaggedValue.value>
          </Foundation.Extension_Mechanisms.TaggedValue>
      </Foundation.Core.ModelElement.taggedValue>
    - <Foundation.Core.ModelElement.taggedValue>
        - <Foundation.Extension_Mechanisms.TaggedValue xmi.id="UID463987FA-5EF3-40B2-AAA6-B487AA494CBD">
              <Foundation.Extension_Mechanisms.TaggedValue.tag>documentation</Foundation.Extension_Mechanisms.TaggedValue.tag>
              <Foundation.Extension_Mechanisms.TaggedValue.value />
          </Foundation.Extension_Mechanisms.TaggedValue>
      </Foundation.Core.ModelElement.taggedValue>
    - <Foundation.Core.Association.connection>
        - <Foundation.Core.AssociationEnd xmi.id="UID62369820-8E82-4382-95A0-F9BA88B392CB">
              <Foundation.Core.ModelElement.name>End13</Foundation.Core.ModelElement.name>
              <Foundation.Core.ModelElement.visibility xmi.value="private" />
              <Foundation.Core.AssociationEnd.isNavigable xmi.value="false" />
              <Foundation.Core.AssociationEnd.isOrdered xmi.value="false" />
              <Foundation.Core.AssociationEnd.aggregation xmi.value="composite" />
              <Foundation.Core.AssociationEnd.multiplicity>1</Foundation.Core.AssociationEnd.multiplicity>
              <Foundation.Core.AssociationEnd.changeable xmi.value="none" />
              <Foundation.Core.AssociationEnd.targetScope xmi.value="instance" />
            - <Foundation.Core.AssociationEnd.type>
                  <Foundation.Core.Class xmi.idref="UIDBF31C111-3F37-48C4-8F84-9494634C97A7" />
              </Foundation.Core.AssociationEnd.type>
            - <Foundation.Core.ModelElement.taggedValue>
                - <Foundation.Extension_Mechanisms.TaggedValue xmi.id="UIDCF6A7D91-2C11-4E77-8C97-E4A645860B9A">
                      <Foundation.Extension_Mechanisms.TaggedValue.tag>documentation</Foundation.Extension_Mechanisms.TaggedValue.tag>
                      <Foundation.Extension_Mechanisms.TaggedValue.value />
                  </Foundation.Extension_Mechanisms.TaggedValue>
              </Foundation.Core.ModelElement.taggedValue>
```

```
          <Foundation.Core.AssociationEnd.qualifier />
        </Foundation.Core.AssociationEnd>
      - <Foundation.Core.AssociationEnd xmi.id="UIDBC535463-27FF-4E67-9E35-6D57F74E980D">
          <Foundation.Core.ModelElement.name>End14</Foundation.Core.ModelElement.name>
          <Foundation.Core.ModelElement.visibility xmi.value="private" />
          <Foundation.Core.AssociationEnd.isNavigable xmi.value="false" />
          <Foundation.Core.AssociationEnd.isOrdered xmi.value="false" />
          <Foundation.Core.AssociationEnd.aggregation xmi.value="none" />
          <Foundation.Core.AssociationEnd.multiplicity>3;4;6</Foundation.Core.AssociationEnd.multiplicity>
          <Foundation.Core.AssociationEnd.changeable xmi.value="none" />
          <Foundation.Core.AssociationEnd.targetScope xmi.value="instance" />
        - <Foundation.Core.AssociationEnd.type>
            <Foundation.Core.Class xmi.idref="UID2729350C-598A-45AE-BAAB-C1B937E4D491" />
          </Foundation.Core.AssociationEnd.type>
        - <Foundation.Core.ModelElement.taggedValue>
          - <Foundation.Extension_Mechanisms.TaggedValue xmi.id="UID685AD313-81C7-4D01-AA05-A53C10F078F8">
              <Foundation.Extension_Mechanisms.TaggedValue.tag>documentation</Foundation.Extension_Mechanisms.TaggedValue.tag>
              <Foundation.Extension_Mechanisms.TaggedValue.value />
            </Foundation.Extension_Mechanisms.TaggedValue>
          </Foundation.Core.ModelElement.taggedValue>
          <Foundation.Core.AssociationEnd.qualifier />
        </Foundation.Core.AssociationEnd>
      </Foundation.Core.Association.connection>
    </Foundation.Core.Association>
```

E.3 UML to database conversion Log-file

Here the log file is given, that is [created] as final step at the UML to database conversion. It holds an overview of the tables, attributes and relationships. The missing 3-4-6-constraint of Association7 is highlighted. Striking as well is the change of the (0..1:1)-relationships into (1:1)-relationships. Highlighted is the association 4, the relationship of geotechnical : positioning.

```
------------------------------------------------------------------------
CASE Tools - Schema wizard log file
Creating schema for model : ArcInfo Uml Model (20-12-2004
11:31:14)
Target database : […]UMLTest.mdb
------------------------------------------------------------------------

Creating/Updating Domains

Creating   Domain :: Owner_CD
    Domain Type  :: Coded Value Domain
    Field Type   :: String
    Merge Policy :: Default Value
    Split Policy :: Default Value
      Shell = Shell
      3rd Party = Third Party


------------------------------------------------------------------------
Creating tables and stand-alone feature classes (at workspace
level)
------------------------------------------------------------------------

---
Rig
---
  Feature class: Rig
    Rig is an object class (table)
    Behavior class: esriCore.Object

    ------
    Fields
      Field: OBJECTID
        Type: Object ID
        Is Nullable: No
      Field: NAME
        Type: String
        Is Nullable: Yes
        Length: 0
      Field: TYPE_CD
        Type: String
        Is Nullable: Yes
        Length: 0
      Field: OPERATOR_CD
        Type: String
        Is Nullable: Yes
        Length: 0
      Field: POSIT_CONTRACTOR_CD
        Type: String
        Is Nullable: Yes
        Length: 0
      Field: COMMENTS
        Type: String
        Is Nullable: Yes
        Length: 0

Saving schema parameters to Repository for Rig

------------
Rig_Activity
------------
  Feature class: Rig_Activity
    Rig_Activity is an object class (table)
    Behavior class: esriCore.Object

    ------
    Fields
      Field: OBJECTID
        Type: Object ID
        Is Nullable: No
      Field: RIG_UID
        Type: Long Integer
        Is Nullable: Yes
        Precision: 0
      Field: SITE_UID
        Type: Long Integer
        Is Nullable: Yes
        Precision: 0
      Field: POSITION_UID
        Type: Long Integer
        Is Nullable: Yes
        Precision: 0
      Field: DESCRIPTION
        Type: String
        Is Nullable: Yes
        Length: 0
      Field: TYPE_CD
        Type: String
        Is Nullable: Yes
        Length: 0
      Field: ASSET_OWNER_CD
        Type: String
        Is Nullable: Yes
        Length: 0
      Field: START_DATE
        Type: Date
        Is Nullable: Yes
      Field: END_DATE
        Type: Date
        Is Nullable: Yes
      Field: COMMENTS
        Type: String
        Is Nullable: Yes
        Length: 0

Saving schema parameters to Repository for Rig_Activity

---------
Footprint
---------
  Feature class: Footprint
    Footprint is an object class (table)
    Behavior class: esriCore.Object

    ------
    Fields
      Field: OBJECTID
        Type: Object ID
        Is Nullable: No
      Field: FOOTPRINT_SET_UID
        Type: Long Integer
        Is Nullable: Yes
        Precision: 0
      Field: POSITION_UID
```

```
        Type: Long Integer
        Is Nullable: Yes
        Precision: 0
      Field: LEG_NAME
        Type: String
        Is Nullable: Yes
        Length: 0
      Field: PRED_PENETRATION
        Type: Long Integer
        Is Nullable: Yes
        Precision: 0
      Field: PENETRATION
        Type: Long Integer
        Is Nullable: Yes
        Precision: 0
      Field: FOOTPRINT_RADIUS
        Type: Long Integer
        Is Nullable: Yes
        Precision: 0
      Field: PRE_LOADS
        Type: Long Integer
        Is Nullable: Yes
        Precision: 0
      Field: SCOUR_TYPE_CD
        Type: String
        Is Nullable: Yes
        Length: 0
      Field: SCOUR_TONNAGE
        Type: Long Integer
        Is Nullable: Yes
        Precision: 0
      Field: SCOUR_DATE
        Type: Date
        Is Nullable: Yes
      Field: COMMENTS
        Type: String
        Is Nullable: Yes
        Length: 0

Saving schema parameters to Repository for Footprint

------------
GeoTechnical
------------
  Feature class: GeoTechnical
    GeoTechnical is an object class (table)
    Behavior class: esriCore.Object

    ------
    Fields
      Field: OBJECTID
        Type: Object ID
        Is Nullable: No
      Field: POSITION_UID
        Type: Long Integer
        Is Nullable: Yes
        Precision: 0
      Field: NAME
        Type: String
        Is Nullable: Yes
        Length: 0
      Field: BORETYPE_CD
        Type: String
        Is Nullable: Yes
        Length: 0
      Field: GT_DATE
        Type: Date
        Is Nullable: Yes
      Field: OWNER_CD
        Type: String
        Domain: Owner_CD
        Default value: Shell
        Is Nullable: Yes
        Length: 0
      Field: CONTRACTOR_CD
        Type: String
        Is Nullable: Yes
        Length: 0
      Field: HOLE_DEPTH
        Type: Long Integer
        Is Nullable: Yes
        Precision: 0
      Field: FORMER_NAME
        Type: String
        Is Nullable: Yes
        Length: 0
      Field: COMMENTS
        Type: String
        Is Nullable: Yes
        Length: 0

Saving schema parameters to Repository for GeoTechnical

-------------
Footprint_Set
-------------
  Feature class: Footprint_Set
    Footprint_Set is an object class (table)
    Behavior class: esriCore.Object

    ------
    Fields
      Field: OBJECTID
        Type: Object ID
        Is Nullable: No
      Field: ACTIVITY_UID
        Type: Long Integer
        Is Nullable: Yes
        Precision: 0
      Field: COMMENTS
        Type: String
        Is Nullable: Yes
        Length: 0

Saving schema parameters to Repository for Footprint_Set

-----------
Positioning
-----------
  Feature class: Positioning
    Positioning  is a stand-alone feature class
    Spatial reference name: Unknown
    Feature type: FEATURE
    Behavior class: esriCore.Feature

    ------
    Fields
      Field: OBJECTID
        Type: Object ID
        Is Nullable: No
      Field: Shape
        Type: Geometry
        Geometry type: Point
        Has Measures: No
        Has Z Values: No
           Num of grids: 1
          Grid(0) = 1000

      Field: SOURCE_X
        Type: Long Integer
        Is Nullable: Yes
        Precision: 0
      Field: SOURCE_Y
        Type: Long Integer
        Is Nullable: Yes
        Precision: 0
      Field: SOURCE_EPSG_CD
        Type: Long Integer
        Is Nullable: Yes
        Precision: 0
      Field: LON_WGS84
        Type: Long Integer
        Is Nullable: Yes
        Precision: 0
      Field: LAT_WGS84
        Type: Long Integer
        Is Nullable: Yes
        Precision: 0
```

```
    Field: COORD_RELIABILITY
      Type: String
      Is Nullable: Yes
      Length: 0
    Field: SURVEY_DATE
      Type: Date
      Is Nullable: Yes
    Field: POSIT_SYSTEM_CD
      Type: String
      Is Nullable: Yes
      Length: 0
    Field: ELEVATION
      Type: Long Integer
      Is Nullable: Yes
      Precision: 0
    Field: ELEV_REF_LEV_CD
      Type: String
      Is Nullable: Yes
      Length: 0
    Field: ELEV_RELIABILITY_CD
      Type: String
      Is Nullable: Yes
      Length: 0
    Field: WATER_DEPTH_LEVEL
      Type: Long Integer
      Is Nullable: Yes
      Precision: 0
    Field: DEPTH_REF_LEV_CD
      Type: String
      Is Nullable: Yes
      Length: 0
    Field: DEPTH_RELIABILITY_CD
      Type: String
      Is Nullable: Yes
      Length: 0
    Field: AZIMUTH
      Type: String
      Is Nullable: Yes
      Length: 0
    Field: COMMENTS
      Type: String
      Is Nullable: Yes
      Length: 0

Saving schema parameters to Repository for Positioning


------------------------------------------------------------------------
Creating relationship classes at the workspace level
------------------------------------------------------------------------
Working on relationship class : Association17
Creating relationship class Association17
Relationship class:: Association17
  Origin Primary Key : OBJECTID
  Origin Foreign Key : RIG_UID
  Type          :: Simple
  Cardinality   :: 1-N
  Origin        :: Rig
  Backward Label:: End34
  Destination   :: Rig_Activity
  Forward Label :: End33

Working on relationship class : Association19
Creating relationship class Association19
Relationship class:: Association19
  Origin Primary Key : OBJECTID
  Origin Foreign Key : ACTIVITY_UID
  Type          :: Simple
  Cardinality   :: 1-1
  Origin        :: Rig_Activity
  Backward Label:: End38
  Destination   :: Footprint_Set
  Forward Label :: End37

Working on relationship class : Association5
Creating relationship class Association5
Relationship class:: Association5
  Origin Primary Key : OBJECTID
  Origin Foreign Key : POSITION_UID
  Type          :: Simple
  Cardinality   :: 1-1
  Origin        :: Positioning
  Backward Label:: End9
  Destination   :: Footprint
  Forward Label :: End10

Working on relationship class : Association4
Creating relationship class Association4
Relationship class:: Association4
  Origin Primary Key : OBJECTID
  Origin Foreign Key : POSITION_UID
  Type          :: Simple
  Cardinality   :: 1-1
  Origin        :: Positioning
  Backward Label:: End8
  Destination   :: GeoTechnical
  Forward Label :: End7

Working on relationship class : Association6
Creating relationship class Association6
Relationship class:: Association6
  Origin Primary Key : OBJECTID
  Origin Foreign Key : POSITION_UID
  Type          :: Simple
  Cardinality   :: 1-1
  Origin        :: Positioning
  Backward Label:: End11
  Destination   :: Rig_Activity
  Forward Label :: End12

Working on relationship class : Association7
Creating relationship class Association7
Relationship class:: Association7
  Origin Primary Key : OBJECTID
  Origin Foreign Key : FOOTPRINT_SET_UID
  Type          :: Composite
  Cardinality   :: 1-N
  Origin        :: Footprint_Set
  Backward Label:: End13
  Destination   :: Footprint
  Forward Label :: End14
```

Appendix F: FAME-2 user interface testing

This section describes the questions, tasks and results of the FAME-2 user interface usability testing. Section F.1 describes the users executed usability testing: the questions, tasks and results. Section F.2 states the heuristic evaluation and section F.3 the resulting actions list to improve the user interface.

F.1 Usability Testing

This section gives the questions, tasks and results of the user executed usability testing of the FAME-2 user interface. It is divided into questions preceding the user test in section F.1.1, the results of the user testing and the tasks done in relation to the tasks done, in section F.1.2. Section F.1.3 holds the answers to the questions after the user test, whereas section F.1.4 gives the results of the questionnaire concerning the usability of the interface. Section F.1.5 states the answers on the questions about the test itself and section F.1.6 concludes with the results of the questions concerning the interface layout.

F.1.1 Questions preceding the test

- What do you expect of the system interface?
  To be an easy to use, dynamic, efficient and intuitive interface with standard functionality (copy, past, sort, search) displaying only the data that is manageable.

- What do you intrinsically want to be implemented in the system interface?
  - *The use of drop down lists from reference tables, which can be edited from the interface and automatically select the belonging data from the reference table.*
  - *Basic search functionality, which allows searches by a.o. update-date and user.*
  - *Efficient publishing method.*
  - *Automatic attribute block name generation while typing coordinates.*
  - *LiveLink file loading interface (for geotechnical data).*
  - *Possibility to see the typed in coordinate on the map.*

- What do you intrinsically *NOT* want to be implemented in the system interface?

- How much will you use the system?
  It will be used differing from a daily basis to a weekly basis

- To what extent do you think of the new system interface as an improvement of the current situation?
  Everything can be maintained in one system with a smarter interface.

- Some time ago you stated that a cartographic component in the system interface could be a separate functionality (ArcMap) and did not need to be an integral part of the interface itself. Why is that?
  Because from past experience it became clear that it was not necessary, it was never used for QC at the time of data entry. And if it could not be a platform independent tool, it could cause problems

- Any other remarks
  "Can we call it something else? FAME always reminds me of the dodgy TV show!"

F.1.2 User – Interface Testing

Create and edit all instances; Fill in all the fields of the menus:

- Buoy
- Geotechnical [site]
- Installation
- Feature
- Rig
- Rig Activity
- Footprint Set
- Footprints
- Position
- Document

If there are any other functionalities you would like to test, feel free to do so!

User testing results

General:

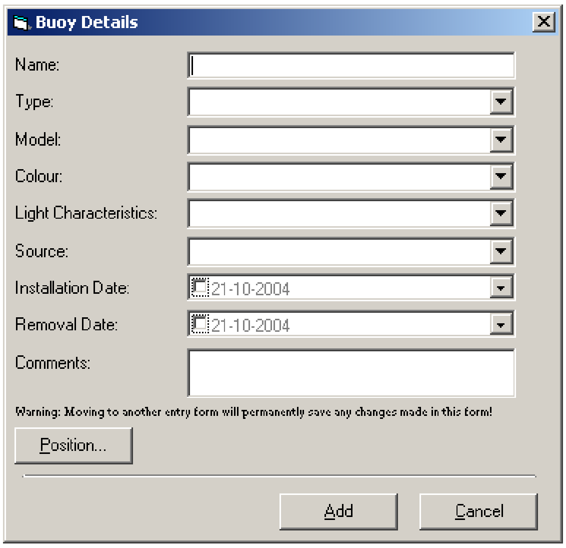
- *Put the dimension unit in the field name (e.g. [m] = metres)*
- *Drop down / code lists have to be completed and correct*
- *Add the possibility to clear a drop down list*
- *Would be nice to use the tab-button to move in order through the fields*
- *Add a star to the field name of obligatory fields*
- *Put drop down lists content in alphabetical order*
- *Reduce the number of screens*
- *Set date standard to today's date*
- *The obligatory fields should be clearly set and thought over*
- *Set the same constraints that are active in the database also active in the interface (e.g. coordinate field in database can only have 7 digits, coordinate field in interface can have more, this causes problems)*

ArcGIS

- *Would like to have a relation between the point that is updated in fame and the visualisation in ArcGIS (e.g. highlight the same point in fame and ArcGIS automatically)*
- *There are ArcGIS-problems with visualisation the data via 'QC Position': the point cannot be selected in a query and disappears spontaneously of the screen.*

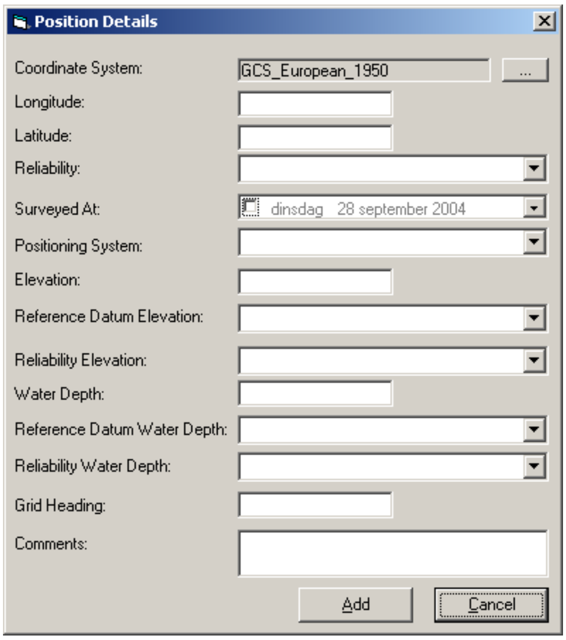
Buoy

- *There is a desire to know who the owner is. Is the field 'source' sufficient or does there need to be a field 'owner' as well?*



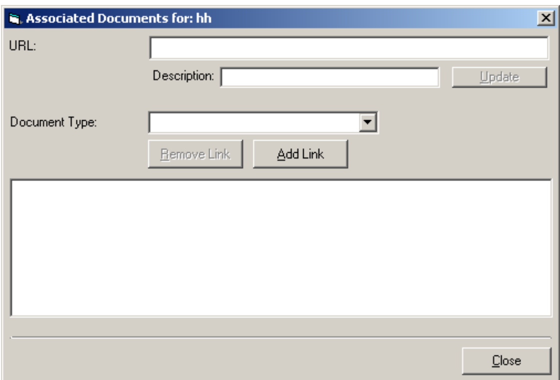
Position

- *User enters a 'reference' for 'elevation', but no 'elevation' value, perhaps a constraint or trigger is desired to prevent this?*
- *Unclear which 'reference' fields are needed for 'elevation' en which for 'depth'?*
- *For each object the 'position' menu is the same. It is not always necessary to have the same position attributes.*
- *Set coordinate reference system to personal preference for all positions*
- *It is unclear to what the height for 'elevation' is relative to.*



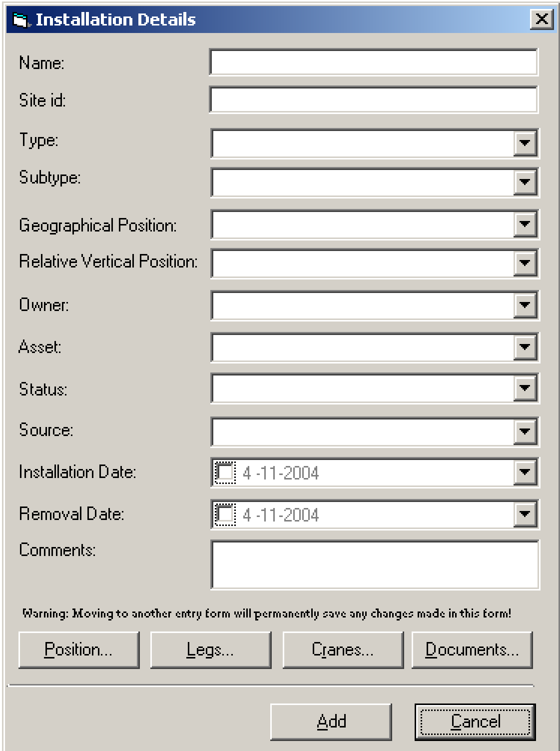
Docs & LiveLink

- *Preferred is a direct connection to LiveLink, because in this situation attribute data has to be filled in twice: both in LiveLink and in fame. Why is there not a direct entry in fame with link between LiveLink en fame*
- *Users would like LiveLink to be black box behind fame and to use fame as interface o add documents to LiveLink*
- *Layout of document menu is a bit unclear*
- *Would like to have the 'url' field cleared after having added one*
- *Would like to have the 'description' field directly available*
- *How will objects with multiple document links be dealt with in ArcGIS?*



Installation

- *What is 'site-id'? Is it coming from SAP, if so is it possible to have a drop down list? Otherwise, what is it for?*

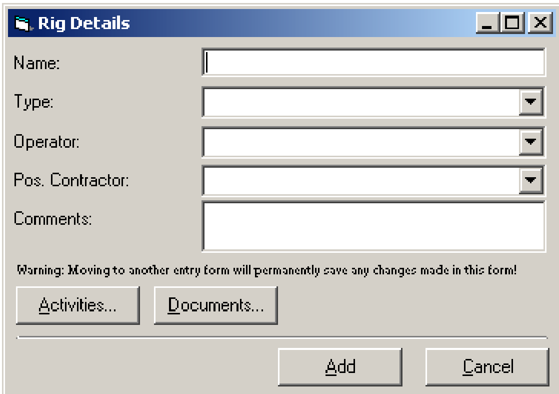


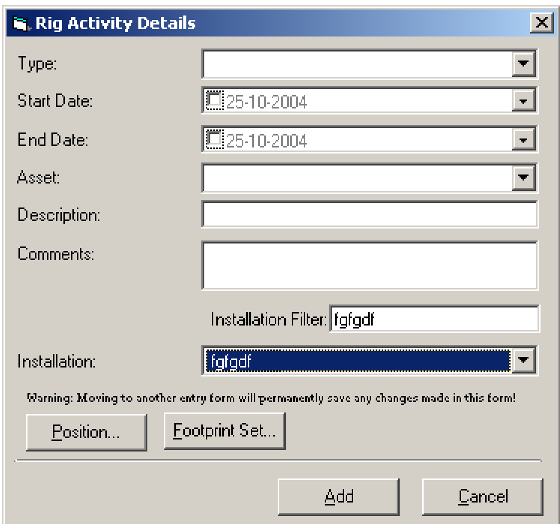
Legs - Parts

- *For a new entry: it would be convenient if 'bottom position' would be similar to previous 'top position'.*
- *Make a distinction between a vertical leg and a diagonal leg to easily fit in legs with only one position and part.*
- *But perhaps these requests are many details for a relatively not frequently used menu.*

Rigs – Rig activities

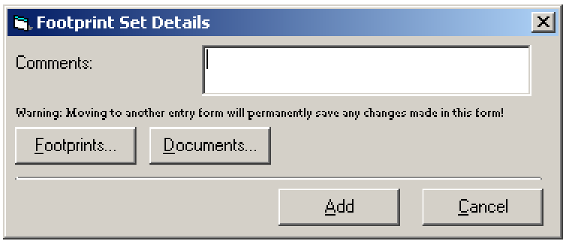
- *To add details for 'rig' is not standard workflow, suggested is to have the same structure as in old fame, where 'rig' is a sub-object of 'rig activity' and where the detail attributes from the rig are automatically entered in a menu when a specific rig is selected (e.g. 'rig name' should be populated from rig definition table).*
- *The field 'owner of rig' should be called rig_owner.*
- *Desired is to have a menu where data about the rig legs are entered (like in old fame) and that is directly related to a certain rig, having a.o. a preset number of legs and leg diameter.*
- *Desired is to see the coordinates in the 'rig activity' detail menu.*

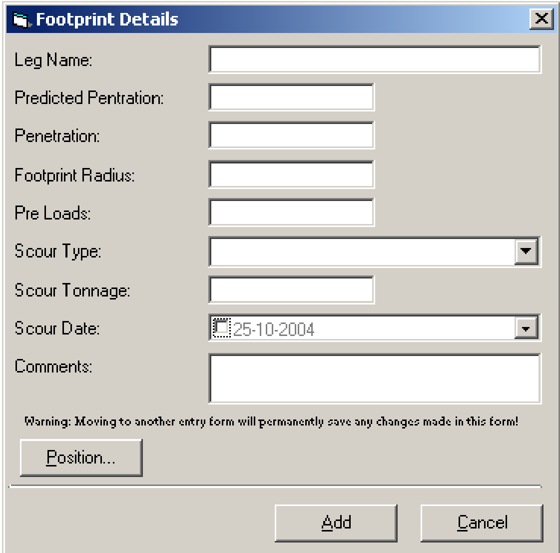


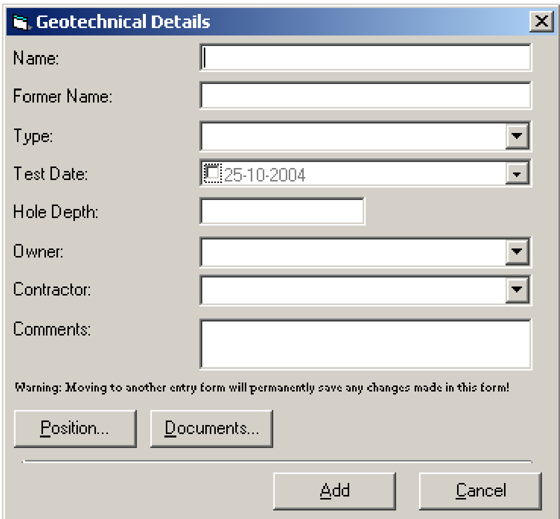


Footprints

- *Constraint on number of footprints: 3, 4 or 6*
- *Reduce number of screens*
- *Desired is to have a direct relationship between a certain rig and the footprints, where the leg names and number and the diameter is defined by the rig (as in old fame)*
- *The field 'scour type' should become 'scour protection type'*
- *Desired is to directly see the next footprint after having filled in the previous*
- *All 3/4/6 footprints should obligatory to enter*
- *Footprint data is very hard to come to, due to the many screens coming up, please reduce the number of screens*
- *Note that all footprints of a set have the same 'date', could that be triggered?*
- *Desired flow is to have an Entry at rig activity list; select Shell rigs; add spudcans to the rig at location*
- *Could the 'position' of a footprint position be picked up from the previous one?*







Any other comments

- *Would like the possibility to look up in drop down list by means of typing name, and not by jumping from 1st letter*
- *Would like to see directly the result of which object is added in result screen*
- *Would like to see source coordinates in main result list (X-Y-Src)*
- *Preferred search options:*
  - *By country*
  - *By editor*
  - *By update-date*
  - *By predefined specific complex queries*
- *A would like is to have an export functionality to export*

Search Field

- *Search ability is repeatedly used as new entry field -> change the layout by either highlighting the words 'new' and 'search' or by changing places of the 'new-edit-delete' functionalities going to the top with the 'search' functionality going to the bottom*
- *Preferred is to directly see what you have entered*
- *Likable is to enter the data for a new instance directly in the first interface that comes up, but the selection of an old feature as input for a new one is very likable*

Delete

- *Desired is to have the selected object removed and all the sub-objects belonging to it (e.g. delete along with platform the cranes and legs and parts). In addition to that a list should come up with all the related (main level) objects of the deleted in which the user can select the related objects that should be deleted as well (e.g. list of anchors related to the platform).*

F.1.3. Questions about the system test - after completion

- Did the system interface come up to your expectations?
  Yes and No

- Which of the elements that you intrinsically wanted to be implemented in the system interface were present?
  An efficient interface for entering coordinates and attributes and the use of drop down lists from reference tables.

- Did you miss the other elements that you intrinsically wanted to be implemented in the system interface and / or were they supported in another way?
  - *There was no ability to edit the drop downs from interface.*
  - *No upload module to LiveLink for documents.*
  - *No automatically displayed data from the reference tables.*
  - *No graphical viewer for QC-ing the typed in coordinates.*

- Were there any elements that you intrinsically *NOT* wanted to be implemented in the system interface? Which ones?

- To what extent do you think of the new system interface as an improvement of the current situation now you have done the test?
  The new interface is generally an improvement. With the slight improvements discussed during testing it should be a much lighter to use, straightforward interface. However without the improvements it definitely has too many boxes popping up for adding data.

- How much will you use the system?
  Same as stated before

- Some time ago you stated that a cartographic component in the system interface could be a separate functionality (ArcMap) and did not need to be an integral part of the interface itself. Is that still the case? Why is that?
  Same as stated before

- Any other remarks
  - *As discussed in the testing: there is a desire to have the ability to define a user preference for the coordinate reference system.*
  - *Publishing to external data formats was not able in the current prototype.*
  - *Footprint diameter and # of rigs should be coming from a rig reference table.*
  - *The platform leg coordinates should have a fixed top and bottom and the possibility to insert extra coordinates between the top and bottom.*

F.1.4. Questions about the usability of the system interface

- What is you overall opinion of the system interface?
  It was not clear because all of the lose screens, but when enhancements are made it will be an easy to use intuitive interface.

- Please indicate in the table below what your opinion is about the system interface:
  (Please check off the box that corresponds with your opinion)

<table>
<tr><th></th><th>1</th><th>2</th><th>3</th><th>4</th><th>5</th><th></th></tr>
<tr><td>Pleasing</td><td></td><td></td><td colspan="2">X</td><td></td><td>Irritating</td></tr>
<tr><td>Complete</td><td></td><td colspan="2">X</td><td></td><td></td><td>Incomplete</td></tr>
<tr><td>Cooperative</td><td></td><td></td><td>X</td><td></td><td></td><td>Uncooperative</td></tr>
<tr><td>Simple</td><td></td><td></td><td colspan="2">X</td><td></td><td>Complicated</td></tr>
<tr><td>Fast to use</td><td></td><td colspan="2">X</td><td></td><td></td><td>Slow to use</td></tr>
<tr><td>Safe</td><td></td><td></td><td>X</td><td></td><td></td><td>Unsafe</td></tr>
<tr><td>Easy to understand</td><td></td><td>X</td><td></td><td></td><td></td><td>Hard to understand</td></tr>
<tr><td>Efficient</td><td></td><td></td><td colspan="2">X</td><td></td><td>Inefficient</td></tr>
<tr><td>Helpful</td><td></td><td></td><td>X</td><td></td><td></td><td>Unsupportive</td></tr>
<tr><td>Controllable</td><td></td><td colspan="2">X</td><td></td><td></td><td>Uncontrollable</td></tr>
<tr><td>Recognisable</td><td colspan="2">X</td><td></td><td></td><td></td><td>Alien</td></tr>
</table>

- Please indicate your opinion about the statements below:
  (Please check off the box that corresponds with your opinion about the statement)

<table>
<tr><th></th><th>Strongly disagree</th><th>Somewhat disagree</th><th>Neutral</th><th>Somewhat agree</th><th>Strongly agree</th></tr>
<tr><td>The system interface is easy to learn</td><td></td><td></td><td></td><td>X</td><td></td></tr>
<tr><td>Once learned, the system supports a high productivity level.</td><td></td><td></td><td colspan="2">X</td><td></td></tr>
<tr><td>The system interface would be easy to remember to reuse after not having used it for some time.</td><td></td><td></td><td></td><td>X</td><td></td></tr>
<tr><td>The system interface has a low error rate.</td><td></td><td></td><td></td><td colspan="2">X</td></tr>
<tr><td>The system interface supports easy error recovery.</td><td></td><td></td><td>X</td><td></td><td></td></tr>
<tr><td>The system is pleasant to use.</td><td></td><td></td><td colspan="2">X</td><td></td></tr>
</table>

F.1.5 Questions about the user – interface test

- Please indicate about each of the tasks given in the user - interface test whether you felt that the **time** needed to complete the task was sufficient or too long.
  The time needed to complete the tasks was sufficient, however the rig parts creation and edition could improve.

- Please indicate about each of the tasks given in the user - interface test whether you felt that the **actions** that took you to complete the task were sufficient or too many.
  The actions to take were sufficient apart from the Rigs/Spudcan parts, which had too many screens.

- Please indicate whether the tasks given in the user - interface test were representative and whether you perhaps missed specific tasks appropriate for the testing.
  The tests were representative and covered most of the tasks that would happen in a production environment.

F.1.6 Questions about the interface layout

- Do you have any general comments about the hierarchy of the menus of the system interface?
  Must Have - Rig tab should show the Rig Activity, not the Rig definitions. These can be moved to a less prominent location.

- Do you have any comments about the layout of the menus of the system interface?
  Should Have – Search and Result areas maybe moved lower with Data Entry "New" "Edit" etc. moved to top.

- Please state about each menu:
  Which entry-fields you think should be obligatory entries.
  If you have any remarks about the layout.
  --- See table F.1.1 for answer ---

Table F.1.1: users' remarks about obligatory entries and layout of FAME-2 user interface

Obligatory entries	Layout remarks
Installation:	
NAME TYPE SUBTYPE GEOGRAPHICAL POSITION OWNER STATUS SOURCE	▪ Site ID should be a drop down list
Feature:	
NAME TYPE SUBTYPE GEOGRAPHICAL POSITION OWNER STATUS SOURCE	
Buoys:	
NAME TYPE	▪ Light characteristics should be an open field
Geotechnical:	
NAME TYPE HOLE DEPTH CONTRACTOR	
Rig:	
NAME TYPE	▪ Extra fields coming from footprints: No. of legs ▪ Leg diameter
Rig Activity:	
	▪ Installation – for rigs on hire by Shell only ▪ Source – i.e. Shell, NTM, 3rd party
Footprints	
LEG NAME PENETRATION FOOTPRINT RADIUS	▪ Leg name & footprint radius should come from rig library ▪ Penetration should be actual penetration ▪ Scour penetration type should be scour protection type
Documents	
	▪ Description – from drop down list
Position	
COORDINATE REFERENCE SYSTEM LONGITUDE LATITUDE POSITIONING SYSTEM	▪ Grid heading – for those instances that don't have headings, e.g. buoys, FPSOs, enter 0 ▪ If elevation is entered, then ref. dat. elevation should be obligatory ▪ If water depth is entered, then ref. dat. water depth should be obligatory ▪ Surveyed at should be survey date
Crane	
	▪ Type should be open field ▪ Manufacturer should be open field

F.2 Heuristic Evaluation

Here the conclusions of the heuristic evaluation are given. They are structured by the heuristic evaluation concepts as described by Nielsen [39].

Visibility of the system status

Is clear, except for going to footprints, then the number of screens coming up make it unclear which screen is active.

Match between system and the real world

Is clear, however for a green and fresh user a data dictionary would come in handy.

User control and freedom

Is okay.

Consistency and standards

Is at high level, only the order of the same attributes differs in some different object menus.

Error prevention

Good

Recognition rather than recall

Sufficient, but could improve:
The automatic position-attribute copier fails sometimes
Some multiple- or related objects could be improved with such an attribute copier (e.g. footprints or rigs-rig activities)

Flexibility and ease of use

Good, the selection / filtering option which prevents the loading of all data at once is a good accelerator.

Aesthetic and minimalistic design

Could improve, mainly because the huge amount of screens popping up. Also if in position not all entries are fit for that object, they could be blanked out.

Help users recognize, diagnose and recover from errors

Except for the major breakdown, these are covered well.

Help and documentation

A (small) user's guide would be a great improvement.

Conclusions:

- *Reduce the number of screens.*
- *Set up data dictionary and user's guide.*
- *Align the order of the same attributes in different object menus.*

- *Fix the automatic position-attribute copier.*
- *Set up automatic attribute copier (or alike) for other objects.*
- *If in position menu not all entries are fit for the related object, they could be blanked out.*

F.3 Resulting action list from usability testing & heuristic evaluation

This section gives the action list that is a result of the usability testing and the heuristic evaluation combined. It is structured by entity and scaled by MoSCoW-concept (as is described in section 3.3).

General:	
MM	Reduce the number of screens
M	Set the same constraints that are active in the database also active in the interface (e.g. coordinate field in database can only have 7 digits, coordinate field in interface can have more, this causes problems)
M	Add a star to the field name of obligatory fields
M	Put the dimension unit in the field name (e.g. [m] = meters)
S	Set date standard to today's date
S	Fix the automatic position-attribute copier
S	Set up user's guide with data dictionary
C	Set up automatic attribute copier (or alike) for other objects
W	Would be nice to use the tab-button to move in order through the fields
W	Align the order of the same attributes in different object menus
W	If in position menu not all entries are fit for the related object, they could be blanked out
W	Would like the possibility to look up in drop down list by means of typing name, and not by jumping from 1st letter
W	Would like to see directly the result of which object is added in result screen
W	Would like to see source coordinates in main result list (X-Y-Src)
W	Would like to have an export functionality to export

Main screen	
S	The layout should be changed, either to highlight the words 'new' and 'search' or by changing places of the 'new-edit-delete' functionalities going to the top with the 'search' functionality going to the bottom. The latter one is preferred.
C	Preferred search options: By country, editor, update-date, predefined specific complex queries
C	Preferred is to directly see what you have entered
W	Likable is to enter the data for a new instance directly in the first interface that comes up, but the selection of an old feature as input for a new one is very likable

Delete	
C	Desired is to have the selected object removed and all the sub-objects belonging to it (e.g. delete along with platform the cranes and legs and parts). In addition to that a list should come up with all the related (main level) objects of the deleted in which the user can select the related objects that should be deleted as well (e.g. list of anchors related to the platform).

Drop down lists:	
M	Drop down / code lists have to be completed and correct
M	Put drop down lists content in alphabetical order
C	Add the possibility to clear the field of a drop down list

Buoy:	
S	Light characteristics should be an open field

Position:	
M	Set coordinate reference system to personal preference for all positions
S	It is unclear to what the height for 'elevation' is relative to
S	If water depth is entered, then the reference datum should be obligatory, the same situation for the elevation. Perhaps set a constraint or trigger to prevent the fields from being active if there is no entry
S	Surveyed as should be survey date
W	For each object the 'position' menu is the same. It is not always necessary to have the same position attributes.

Documents	
C	Description could come from drop down list
W	Would like to have the 'description' field directly available
W	Would like to have the 'url' field cleared after having added one
W	Users would like LiveLink to be black box behind fame and to use fame as interface o add documents to LiveLink

Installation	
S	Site ID should be a drop down list

Parts: (perhaps these requests are many details for a relatively not frequently used menu)	
C	Desired is to have a fixed top and bottom for the platform leg coordinates and the possibility to insert extra coordinates between the top and bottom.
C	Make a distinction between a vertical leg and a diagonal leg to easily fit in legs with only one position and part
W	For a new entry: it would be convenient if 'bottom position' would be similar to previous 'top position'

Rigs – rig activities – footprints	
MM	Change the set-up of the screens from rigs till footprints: have rig activities to be the entry on the main screen. Have a button in the rig activity menu to select the rig and to enter the rig menu in which also the number of legs and the diameter are to be entered (as in old FAME).
M	Have the number of legs that is entered in the rig menu be a constraint for the number of footprints.
S	Make a constraint to have all footprints entered if entered one.
S	The field 'owner of rig' should be called rig_owner
S	Desired is to see the coordinates in the 'rig activity' detail menu
S	The field 'scour type' should become 'scour protection type'
S	Desired is to directly see the next footprint after having filled in the previous
M	Reduce the number of screens to come to footprint data
S	All footprints (of a set) have the same date
C	The 'position' of a footprint position could be picked up from the previous one

Appendix G: Data migration and transformation listings

This appendix lists the data-attributes of the geotechnical & footprint datasets in order to relate them to the new FAME-2 counterparts. These relationships are set out in section G.1 for the geotechnical dataset and in section G.2 for the footprint datasets. When these relations are known, for the new FAME-2 dataset then can be listed where the data-entries should come from. Also can then be listed which transformations are required to get the data in the required format. These listings are given in section G.3.

G.1 Current geotechnical data related to new FAME-2 counterpart

This section lists the data-attributes of the geotechnical datasets in relation to their new FAME-2 counterparts. It is stated which fields they are directly related to and which fields require a transformation. Table G.1.1 lists the NAM dataset, tables G.1.2 - G.1.5 list the Expro datasets.

Table G.1.1: NAM – Oracle geotechnical data related to new FAME-2 counterpart

NAM: GIS - Oracle	FAME-2				
	Direct conversion		New counterpart		Created in GIS layer by FME
	Table	Field	Table	Field	
GEOM			POS	GEOM_WGS84	
SB_HOLEDEPTH	GEO	HOLE_DEPTH			
SB_TEST_DATE	GEO	DATE			
OBJECT_ID			GEO	GEOTECHNICAL_UID	
DATAID			GEO	GEOTECHNICAL_UID	
SB_AREA					BLOCK NUMBER
SB_CONTRACTOR			GEO	CONTRACOTOR_CD	
SB_DEPTH			POS	WATER_DEPTH	
SB_ID			GEO	GEOTECHNICAL_UID	
SB_NAME			GEO	NAME EPIGEN_NAME	
SB_TEST_METHOD			GEO	BORETYPE_CD	
SB_EASTING	POS	SOURCE_X			
SB_NORTHING	POS	SOURCE_Y			
SB_RIG			RIG		
LIVELINK	DOC	URL			
SSL_CODE					SSL CODE
LAST_UPDATED			AUD	TIME_STAMP	
UPDATED_BY			AUD	EDIT_USER	

Table G.1.2: Expro - Oracle geotechnical data related to new FAME-2 counterpart

Expro: GIS - Oracle	FAME				
	Direct conversion		New counterpart		Created in GIS layer by FME
	Table	Field	Table	Field	
GEOM			POS	GEOM_WGS84	
UWI			GEO	NAME EPIGEN_NAME	
NAME			GEO	NAME EPIGEN_NAME	
TYPE			GEO	BORETYPE_CD	
STATUS					
OPERATOR			GEO	CONTRACTOR_CD	
LAT_ED50	POS	SOURCE_Y			
LON_ED50	POS	SOURCE_X			
FINAL_TD	GEO	HOLE_DEPTH			
WELL_NUMBER			GEO	GEOTECHNICAL_UID	
DEPTH_UNIT					
SPUD_DATE	GEO	DATE			
COMPLETION_DATE					
SYMBOL_CODE					SSL CODE
SSL_CODE					SSL CODE
ENUM_TYPE					
DTI_NAME			GEO	NAME EPIGEN_NAME	
WELL_ID			GEO	GEOTECHNICAL_UID	
OBJECT_ID			GEO	GEOTECHNICAL_UID	
DEPTH	POS	WATER_DEPTH_LEVEL			

Table G.1.3: Expro - OpenWorks geotechnical data related to new FAME-2 counterpart

Expro: OpenWorks	FAME				
	Direct conversion		New counterpart		Created in GIS layer by FME
	Table	Field	Table	Field	
WELL_UWI			GEO	FORMER_NAME	
STATE			GEO	BORETYPE_CD	
CURRENT_STATUS					
WELL_OPERATOR			GEO	CONTRACTOR_CD	
PREFERRED_SURF_LAT_ED50					
PREFERRED_SURF_LON_ED50					
TD_FEET	GEO	HOLE_DEPTH			
CURRENT_STATUS_DATE	GEO	DATE			
COUNTRY					COUNRTY
PREFERRED_X_COORD_SURF_TM0					
PREFERRED_Y_COORD_SURF_TM0					
ELEV_REF			POS	DEPTH_REF_LEV_CD	
ELEV_VALUE_FEET	POS	WATER_DEPTH_LEVEL			
ORIG_X_LON	POS	SOURCE_X			
ORIG_Y_LAT	POS	SOURCE_Y			
ORIG_CRS_NAME			POS	SOURCE_EPSG_CD	
PLATFORM_CODE					PLATFORM
ON_OFF_SHORE					
ORIGINAL_LOCATION			GEO	FORMER_NAME	
WELL_REMARK – report no	DOC	DESCRIPTION			
WELL_REMARK – drilling company			GEO	CONTRACTOR_CD	
WELL_REMARK – date	GEO	DATE			
WELL_REMARK – vessel	GEO	COMMENTS			
WELL_REMARK – insitu programme	GEO	COMMENTS			
WELL_REMARK – laboratory tests	GEO	COMMENTS			
WELL_REMARK – engineering	GEO	COMMENTS			

Table G.1.4: Expro - Excel geotechnical data related to new FAME-2 counterpart

Expro: Excel-files - NNS / CNS / SNS	FAME				
	Direct conversion		New counterpart		Created in GIS layer by FME
	Table	Field	Table	Field	
EPIGEN NUMBER			GEO	FORMER_NAME	
LATITUDE					
LONGITUDE					
BLOCK NUMBER					BLOCK NUMBER
REPORT - author			GEO	CONTRACTOR_CD	
REPORT – no			DOC	DESCRIPTION	
REPORT - year			DOC	DESCRIPTION	
EASTING	POS	SOURCE_X			
NORTHING	POS	SOURCE_Y			
BOREHOLE NUMBER			GEO	FORMER_NAME	
ORIGINAL NAME					
PLATFORM					RIG NAME
BOX NUMBER			DOC	DESCRIPTION	
EXPRO NUMBER			DOC	DESCRIPTION	
VC – NAME			GEO	FORMER_NAME	
VC – 1ST COORD	POS	SOURCE_X			
VC – 2ND COORD	POS	SOURCE_Y			

G.2 Current footprint data related to new FAME-2 counterpart

This section lists the data-attributes of the footprint datasets in relation to their new FAME-2 counterparts. It is stated which fields they are directly related to and which fields require a transformation. Table G.2.1 lists the NAM dataset; tables G.2.2 lists the Expro dataset.

Table G.2.1: NAM – Access footprint data related to new FAME-2 counterpart

NAM: Access-database	FAME				
	Direct conversion		New counterpart		Created in GIS layer by FME
	Table	Field	Table	Field	
SCOUR PENETRATION ID			FOT	FOOTPRINT_UID	
RIG NAME					RIG NAME
SCOURPEN DATE	FOT	SCOUR_DATE			
BLOCK					BLOCK NUMBER
WELL NAME					WELL NAME
PLATFORM NAME					PLATFORM NAME
SCOUR PROTECTION			FOT	SCOUR_TYPE_CD	
PREDICTED PENETRATION MIN.					
PREDICTED PENETRATION MAX.	FOT	PRED_PENETRATION			
PENETRATION MIN.					
PENETRATION MAX.	FOT	PENETRATION			
REMARKS	FOT	COMMENTS			

Table G.2.2: Expro - Oracle footprint data related to new FAME-2 counterpart

Expro: Oracle-database	FAME				
	Direct conversion		New counterpart		Created in GIS layer by FME
	Table	Field	Table	Field	
OBJ_ID			FOT	FOOTPRINT_UID	
NAME					RIG NAME
SURVEY_DATE	POS	SURVEY_DATE			
LEG_RADIUS	FOT	FOOTPRINT_RADIUS			
NO_LEGS			FTS	FOOTPRINT_SET_UID	
START_DATE	FOT	SCOUR_DATE			
END_DATE					
GEOM			POS	GEOM_WGS84	
POINT_COORD_ID			FOT	FOOTPRINT_UID	
NAME_DATE					
LATITUDE	POS	SOURCE_Y			
LONGITUDE	POS	SOURCE_X			
LOCATION_ DESCRIPTION					
GEOM.AREA					
GEOM.LEN					

G.3 Conversion transformations for geotechnical & footprint data

This section lists .the relations and transformations required to migrate the data from the different datasets to the new FAME-2 dataset. Table G.3.1 lists the geotechnical data and transformations, table G.3.2 lists the footprint data and transformations.

Table G.3.1: Transformations to come to FAME-2 geotechnical dataset

Geotechnical	Table	Field	Value	Transformation
GEOTECHNICAL_UID		*		To be created
POSITION_UID		*		To be created
NAME		*		To be created: Naming according to naming convention
BORETYPE_CD	EXL	OPENWORKS LIST.XLS ... VC.XLS		*EXCEL-FILES: VIBROCORES ARE IN SEPARATE FILES; BOREHOLES AND CPTS ARE DISTINGUISHABLE BY FIRST 2 LETTERS OF NAME AND BY FILE “OPENWORKS LIST.XLS”*
	GIS	TYPE		
	OW	STATE		
	NL	SB_TEST_METHOD		
DATE	GIS	SPUD_DATE		OW-entries: if dates are not similar, select latest date
	OW	CURRENT_STATUS_DATE WELL_REMARKS – date		
	NL	SB_TEST_DATE		
OWNER_CD			“Shell”	
CONTRACTOR_CD	EXL	*REPORT - author*		OW-entries: preferred entry = WELL_REMARK – drilling company
	GIS	OPERATOR		
	OW	well_operator *WELL_REMARK – drilling company*		
	NL	SB_CONTRACTOR		
HOLE_DEPTH	GIS	FINAL_TD		GIS & OW entries have to be converted from feet to metres.
	OW	TD_FEET		
	NL	SB_HOLEDEPTH		
FORMER_NAME	EXL	*EPIGEN NUMBER* *BOREHOLE NUMBER* *VC - NAME*		EXL-entries: preferred entry = EPIGEN NUMBER & VC - NAME GIS-entries: preferred entry = UWI = NAME OW-entries: preferred entry = WELL_UWI
	GIS	uwi *NAME* *DTI_NAME*		
	OW	*WELL_UWI* *ORIGINAL_LOCATION*		
	NL	*SB_NAME*		
COMMENTS	OW	WELL_REMARK – vessel – insitu programme – laboratory tests – engineering		

Table G.3.1: Transformations to come to FAME-2 geotechnical dataset … continued

Position	Table	Field	Value	Transformation
POSITION_UID		*		To be created
SOURCE_X	EXL	EASTING VC – 1ST COORD		
	GIS	LON_ED50		
	OW	ORIG_X_LON		
	NL	SB_EASTING		
SOURCE_Y	EXL	NORTHING VC – 2ND COORD		
	GIS	LAT_ED50		
	OW	ORIG_Y_LAT		
	NL	SB_NORTHING		
SOURCE_EPSG_CD	OW	ORIG_CRS_NAME		EXL-entries: Tm0; GIS-entries: ED50; NL-entries: TM5
GEOM_WGS84		*		To be created
LON_WGS84		*		To be created
LAT_WGS84		*		To be created
COORD_RELIABILITY			"Estimated"	Standard value for migrated data (?)
SURVEY_DATE			"Unknown"	
POSIT_SYSTEM_CD			"Unknown"	
ELEVATION			No entry	
ELEV_REF_LEV_CD			No entry	
ELEV_RELIABILITY_CD			No entry	
WATER_DEPTH_LEVEL	GIS	DEPTH		OW entries have to be converted from feet to metres.
	OW	ELEV_VALUE_FEET		
	NL	SB_DEPTH		
DEPTH_REF_LEV_CD	OW	ELEV_REF		Other migrated data: value = "Unknown"
DEPTH_RELIABILITY_CD			"Estimated"	Standard value for migrated data (?)
AZIMUTH			"Unknown"	
COMMENTS				

Document	Table	Field	Value	Transformation
DOC_UID		*		To be created
URL	NL	LIVELINK		Reports to be scanned
DESCRIPTION	EXL	REPORT - no REPORT - year BOX NUMBER EXPRO NUMBER		Reports to be scanned
	OW	WELL_REMARK – report no		

Table G.3.2: Transformations to come to FAME-2 geotechnical dataset

Footprint set	Table	Field	Value	Transformation required
FOOTPRINT_SET_UID		*		To be created
ACTIVITY_UID		*		To be created
COMMENTS	NL	REMARKS		

Footprint	Table	Field	Value	Transformation
FOOTPRINT_UID		*		To be created
FOOTPRINT_SET_UID		*		To be created
POSITION_UID		*		To be created
LEG_NAME			New entries only	Naming according to naming convention
PRED_PENETRATION	NL	PREDICTED PENETRATION MAX		Other migrated data: value = “Unknown”
PENETRATION	NL	PENETRATION MAX		Other migrated data: value = “Unknown”
FOOTPRINT_RADIUS	UK	LEG_RADIUS		Other migrated data: value = “Unknown”
PRE_LOADS			New entries only	
SCOUR_TYPE_CD	NL	SCOUR PROTECTION		Only “Yes” / “No” entries in NL-data. If “Yes” then value = “Unknown”
SCOUR_TONNAGE			New entries only	
SCOUR_DATE	UK	START_DATE		
	NL	SCOURPEN DATE		
COMMENTS	NL	REMARKS		

Table G.3.2: Transformations to come to FAME-2 footprint dataset … continued

Position	Table	Field	Value	Transformation
POSITION_UID		*		To be created
SOURCE_X	UK	LONGITUDE		Other migrated data: value = "Unknown"
SOURCE_Y	UK	LATITUDE		Other migrated data: value = "Unknown"
SOURCE_EPSG_CD				UK-entries: TM0; NL-entries: TM5
GEOM_WGS84		*		To be created
LON_WGS84		*		To be created
LAT_WGS84		*		To be created
COORD_RELIABILITY			"Estimated"	Standard value for migrated data (?)
SURVEY_DATE	UK	SURVEY_DATE		Other migrated data: value = "Unknown"
POSIT_SYSTEM_CD			"Unknown"	
ELEVATION			No entry	
ELEV_REF_LEV_CD			No entry	
ELEV_RELIABILITY_CD			No entry	
WATER_DEPTH_LEVEL			"Unknown"	
DEPTH_REF_LEV_CD			"Unknown"	
DEPTH_RELIABILITY_CD			"Unknown"	
AZIMUTH			"Unknown"	
COMMENTS			"Unknown"	

Document	Table	Field	Value	Transformation
DOC_UID		*		To be created
URL				Reports to be scanned
DESCRIPTION				Reports to be scanned

Notes

Notes