Harmonisation of distributed geographic datasets - A model driven approach for geotechnical & footprint data.

Bart van den Berg (1,2), Bart Hulshof (1),
Theo Tijssen (2) en Peter van Oosterom (2)

(1) Shell/NAM, Geo-Information Management,
PO Box 28000, 9400 HH Assen, The Netherlands
(tel. +31 592 3 63 295)

(2) Delft University of Technology, GIS technology,
P.O. Box 5030, 2600 GA Delft, The Netherlands
(tel. +31 15 2786950, fax +31 15 2782745)

A.H.H.J.vandenBerg@student.tudelft.nl, Bart.Hulshof@Shell.com,
T.Tijssen@otb.tudelft.nl, oosterom@otb.tudelft.nl

SUMMARY

This document describes the efforts to harmonise the different and distributed geographic datasets within the new corporate Geo-Information Infrastructure (CGII) of Shell. This research focuses on geotechnical and jack-up rig footprint data. In a related project a semantic model for other related fixed and mobile (offshore infrastructure) entities has been developed (FAME-2). Both models are harmonized to form a meaningful, semantic foundation of the CGII. Following the model driven architecture (MDA) parts of the system have actually been realized based on the conceptual models, which are in the form of UML class diagrams. It is concluded that this approach has proven to be applicable and will be part of a sustainable, standards based, CGII.

KEYWORDS: *Harmonisation, Model Driven Architecture, distributed geographic data, UML, Geo-Information Infrastructure, Spatial Data Infrastructure, Spatial Information Management*

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) (Bishr et al, 1998, Groot&McLaughlin, 2000, Nebert, 2004, van Oosterom et al, 2000).

DATA MANAGEMENT SHELL EP EUROPE

Shell Exploration & Production in Europe (Shell EP Europe), the operating company of Shell for European countries with main offices in Norway (Norske Shell), Scotland (Shell Expro) and The Netherlands (NAM), has made a start to harmonise the diverse and distributed offshore point datasets. This is organised in the project “aligning offshore infrastructure data-models”. At the departments involved this project is commonly referred to by the name of its data-model: Fixed And Mobile Entities version 2 (FAME-2) (Al-Harty et al, 2004).

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 perspective of 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 (figure 1). These data are required because old imprints should be avoided when an installation is positioned near the same platform at a later date.

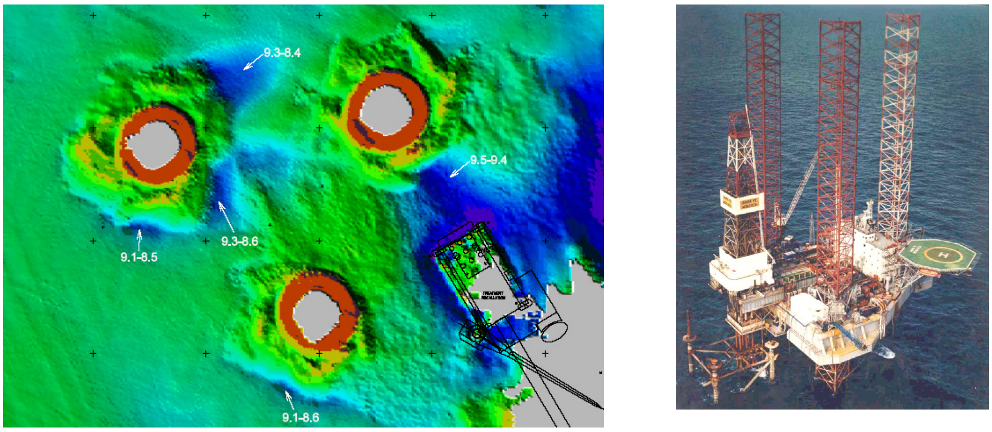

Figure 1: Bathymetry impression of jack-up rig footprints (up to 15 metres deep).

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 be implemented in the corporate GII. A harmonised data-model implies that there should be concordance between the different definitions/contents/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.

CORPORATE GII DEFINITION AND CONTENT

In this research the following definition of a GII is used: a facility to improve 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 used.

An inventory of the current (geotechnical and footprint) 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 a Shell EP Europe-wide data-model. The absence of one single data-model for the combination of geotechnical & footprint data defines the need for a new data-model.

MODEL DRIVEN ARCHITECTURE

The reason for developing data-models at all is based on the concept of the Model Driven Architecture (MDA), set up by the Object Management Group (OMG, 2001). In this MDA a data-model is used as representation of the real world in the computer. The 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 represented.

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" (OMG, 2001). This PIM is regarded the central axis of the system (figure 2). 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.

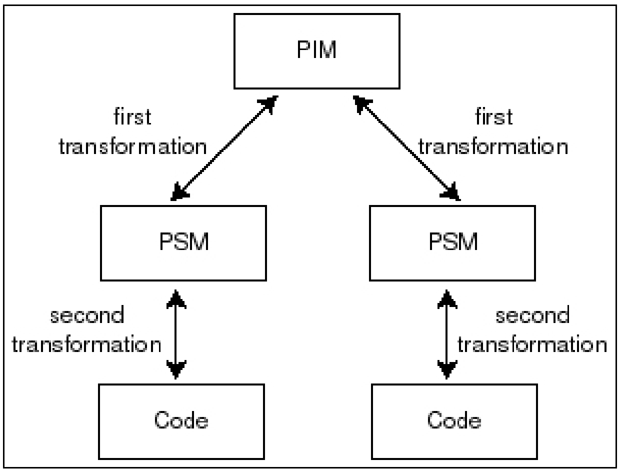

Figure 2: Model driven Architecture (MDA) process steps

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 (Marshall, 1999, Warmer&Kleppe, 2001).

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 (figure 3) is therefore implemented in the FAME-2 data-model. In this model they are related to the other datasets.

REALISATION

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.

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 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).

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 (ESRI, 2003, MacDonald, 1999).

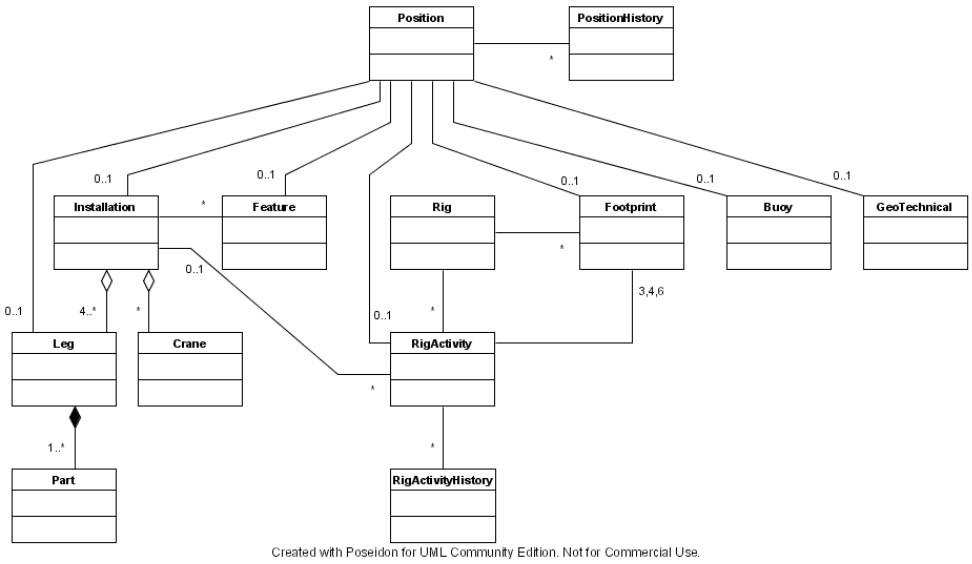

Figure 3: Conceptual model depicted in UML class diagram

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. Within ArcCatalog's '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 4. On approval, this is displayed as a log file and the geodatabase in created. Now it can be populated.

The user interface is currently custom developed for the data custodians, though one could argue that for a large part the user interface forms could also be derived from the data model. 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 detailed 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. Based on this research, several conditions and recommendations are given for proper 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.

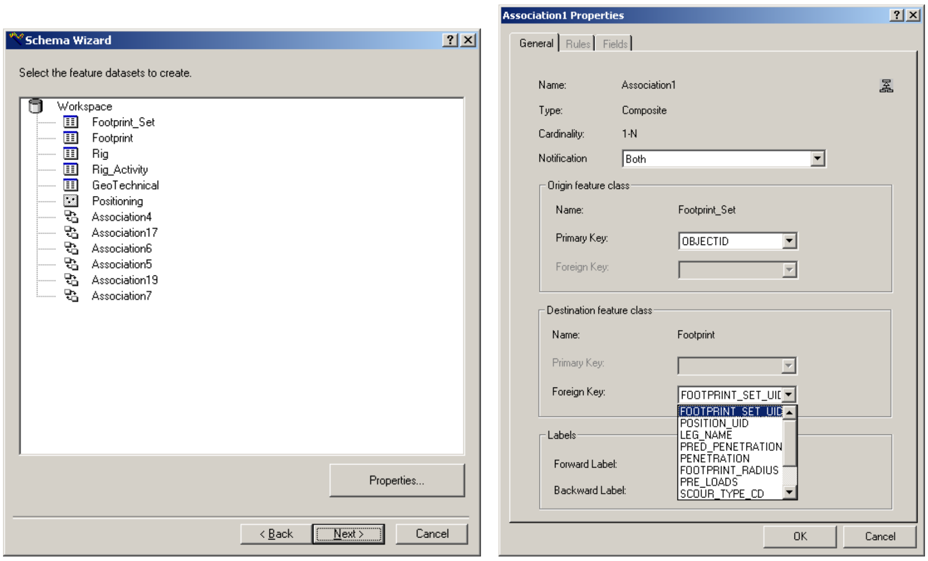

Figure 4: Conceptual model converted and created in the geodatabase

Standardisation: Standardise the workflows and data specifications; the new standardised list can be distributed to the contractors, which 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 widespread.

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 completely web-based interface, for both the data-requests, the data upkeep and the Hazard Notification.

Datasets: Clearly set out the responsibility for the data. Fill in the data gaps as soon as possible, and keep the data as accurate as possible. Create a dataset of the 3rd party data, since looking at the EU-Seased website is not part of the workflow, otherwise make it part of the workflow.

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 this 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.

The UML-modelling and automated conversion testing led to the decision, that the Shell 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.

BIBLIOGRAPHY

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

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).

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.

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

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

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

Object Management Group (OMG), 2001, Model Driven Approach: A technical perspective, OMG, http://www.omg.org/mda/presentations

Oosterom, van, 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.

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