SKB SV-613

Case Based Reasoning: hidden soil knowledge unveiled
Learning from finished in-situ remediation projects
Final Report

A. Leijnse, N. Kukuric, L. Wipfler (TNO-NITG)
J.F. de Kreuk, C.M. Kamermans (Biosoil R&D)
C. Pijls, T. Keijzer (TAUW)

July, 2004

Gouda, CUR/SKB

Titel rapport	CUR/SKB
Case Based Reasoning: hidden soil knowledge unveiled Learning from finished in-situ remediation projects	**rapportnummer** Project rapportnummer

Auteurs	Aantal bladzijden
TNO-NITG A. Leijnse N. Kukuric L. Wipfler BioSoil J.F. de Kreuk C.M. Kamermans TAUW C. Pijls T. Keijzer	**Rapport: 44** **Bijlagen: 0**

Uitvoerende organisaties
TNO-NITG , A. Leijnse, 030-2564706 (penvoerder)
BioSoil, J.F de Kreuk,
TAUW, C.M. Pijls

Uitgever
CUR/SKB, Gouda

Samenvatting
In de afgelopen jaren is veel ervaring opgedaan met in-situ bodemsaneringen. Deze ervaring is meestal opgeslagen in de hoofden van uitvoerders en consultants of in slecht toegankelijke rapporten. Zij is derhalve niet eenvoudig beschikbaar voor derden ter beoordeling van nieuwe bodemsaneringgevallen. Het project "Case Based Reasoning, hidden soil knowledge unveiled; Learning from finished in-situ remediation projects" is opgezet om een systeem te creëren met behulp waarvan bestaande kennis omtrent in-situ bodemsaneringen eenvoudig toegankelijk wordt. Het systeem bestaat uit een database (vooralsnog met ca. 100 afgeronde bodemsaneringgevallen) en een "reasoner", met behulp waarvan in de database kan worden gezocht naar die afgeronde bodemsaneringgevallen die een zo groot mogelijke overeenkomst vertonen met nieuwe bodemsaneringen. Op deze wijze kan optimaal worden geprofiteerd van reeds bestaande kennis omtrent in-situ bodemsaneringen. Op basis van een zeer beperkt aantal karakteristieken kan tevens een schatting worden verkregen van de saneringsduur en de saneringskosten van nieuwe bodemsaneringgevallen.

Trefwoorden

Gecontroleerde termen:	Vrije trefwoorden:
Biologische afbraak, Bodemsanering, Modellen, Toegankelijkheid	Kennisbank, Expertkennis,
Informatie	informatie

Titel project	Projectleiding
Case Based Reasoning: hidden soil knowledge unveiled Learning from finished in-situ remediation projects	TNO-NITG

Dit rapport is verkrijgbaar bij:
CUR/SKB, Postbus 420, 2800 AK Gouda

Index

LIST OF FIGURES IV

SUMMARY V

1 INTRODUCTION 1

1.1 PROBLEM DEFINITION 1
1.2 GOALS OF THE PROJECT AND THE PROJECT APPROACH 1
1.3 COMPOSITION OF THE CONSORTIUM 2
1.4 CONTENT OF THIS REPORT 2

2 CONCEPT OF THE CASE BASED REASONING APPROACH 3

3 SUMMARY OF ACTIVITIES 5

3.1 EVALUATION OF THE WORK-PROCESS 5
3.2 STRUCTURE OF THE CBR DATABASE AND STORING IN-SITU REMEDIATION PROJECTS 7
3.2.1 Structure of the database 7
3.2.2 Data flow and quality control for new cases in CBR 8
3.3 GEOLOGY IN THE DATABASE 9
3.4 VALIDATION OF THE CBR DATABASE 9

4 ESTIMATES OF REMEDIATION COSTS AND TIME 12

4.1 INTRODUCTION 12
4.2 COVERAGE OF THE DATABASE 12
4.3 MODELING APPROACHES TO FILL THE GAPS IN THE DATABASE 15
4.4 DISCUSSION 15
4.5 NEURAL NETWORKS 16
4.5.1 Relationships between parameters 16

5 EVALUATION OF RESULTS 22

5.1 RESULTS OF TESTING BY EXTERNAL EXPERTS 22
5.2 RESULTS OF THE MODELLING 22

6 CONCLUSIONS AND RECOMMENDATIONS 25

6.1 CONCLUSIONS 25
6.2 RECOMMENDATIONS 25

7 REFERENCES 27

APPENDIX A SYNOPSIS OF THE CBR SEARCH PARAMETERS

APPENDIX B (CONCEPT) BEHEERSPLAN CBR-BODEM

List of figures

Figure 4-1Remediation costs related to volume of contaminated groundwater volume 15
Figure 4-2 Remediation costs related to volume of contaminated groundwater. 15
Figure 4-3 Remediation costs related to geology type. 16
Figure 4-4 Remediation costs related to heterogeneity. 16
Figure 4-5 Remediation costs related to depth of contaminated groundwater. 17
Figure 4-6 Remediation time related to geology type. 18
Figure 4-7 Remediation time relative to volume of contaminated soil. 18
Figure 5-1 Remediation time predicted by the Neural Network versus the real remediation time. Open dots are training cases and solid dots are test cases 20

Summary

Over the years contractors and consultants have obtained a large experience and knowledge with the performance of in-situ remediation projects. This experience and knowledge is stored in people's minds and in reports that are not easily accessible. It is therefore not readily available to others interested in the possibilities and limitations of in-situ remediation. The current project (Case Based Reasoning) was set up to create a system in which information about biological in situ remediation projects is condensed, and which can easily be consulted by people who want to design a new in-situ remediation project. In this way, projects with comparable characteristics will be used as a reference for future activities, which will improve the remediation approach.

In the Case Based Reasoning (CBR)-project to the following objects were created:
A database in which information about in-situ remediation projects is stored;
A reasoner, which enables the comparison of new projects with existing ones, and which determines the level of similarity between the different projects.

The project was divided into a number of sub-projects: definition of the work-process, creating the database, drafting and testing of the reasoner was, and development of a stand-alone application, which may be distributed and used as such. Using a specially developed tool, based on a neural network approach, an estimate of the expected remediation time and costs of a new contamination problem can be given, based on a limited number of characteristics.

1 Introduction

1.1 Problem definition

Over the years contractors and consultants have obtained a large experience and knowledge with the performance of in-situ remediation projects. Although most of the experience and knowledge is made public, it appears that the experience with and knowledge about the actual performance of in-situ remediation projects and their success rates are not readily available. It is stored in people's minds and in reports that are not easily accessible, and is therefore not readily available to others interested in the possibilities and limitations of in-situ remediation methods. In addition, the knowledge transfer within a project is usually not optimal. It appears that certain elements in the work process obstruct the information transfer between participants resulting in sub-optimal performance of remediation projects. Improvement of the knowledge gathering and transfer process would therefore improve the overall performance of remediation projects.

1.2 Goals of the project and the project approach

The main goal of the project is to build an application that will allow a comparison of possible new in-situ remediation projects with existing (finalized) projects. Parts of this application are:

1. A database containing information on (finalized) in-situ remediation projects in condensed form. This database should contain information which is relevant for the performance of the remediation projects, e.g. the nature and extent of the contaminants present. In addition, based on experiences in the past on soil remediation cases, the elements in the work-process that obstruct the use of the available knowledge have been identified. These elements are included in the database in order to be able to judge the influence of the work-process on the remediation results. The database should grow by adding new projects.
2. A reasoner that will allow for a comparison of new remediation projects with the existing remediation projects in the database. Since remediation projects are very site-specific, i.e. two remediation projects may be very similar but they will never be identical, the reasoner should search for similar cases. This is done on the basis of the Case Based Reasoning approach, using specific information (parameters) and specific rules, where different types of information will be weighted differently in the comparison. The reasoner is based on existing software.

The application should be easy to use. It can be used by people who want to perform an in-situ remediation or have to judge remediation plans in such a manner, that projects with comparable characteristics can be selected and be used as a reference for future activities which will improve the remediation approach.

1.3 Composition of the consortium

The consortium of the SKB project CBR consists of:

Party	Contact
TNO-NITG	Mr. A. Leijnse
BioSoil R&D	Mr. J.F. de Kreuk
TAUW	Mr. C. Pijls
BAM	Mr. A. Slagmolen
Port of Rotterdam	Mr. W. van Hattem
Handbook Soil remediation	Mr. J. van der Gun
HMVT	Mr. M. van der Brand
Province of North-Holland	Mr. F. de Graaf
Shell Global Solutions	Mr. W. Vrieling
Shell Nederland Refinery BV	Mr. R.A.A. Hetterschijt
SKB	Mr. J. Verheul
Public Waste Agency of Flanders (OVAM)	Mr. V. Dries

1.4 Content of this report

The project consisted of the following activities:
Evaluation of the work process.
Design of the database for in-situ remediation projects, filling the database and design of the reasoner.
Using additional geological information.
Testing and validation of the Case Based Reasoner application.
Estimation of remediation time and costs using modelling techniques.

In this report the concept of Case Based Reasoning will be described shortly. Next, the activities as mentioned above will be discussed. Special attention will be paid to the modelling tool for estimating the remediation time and costs, since it has not been described in other preliminary reports. Finally, conclusions and recommendations for future maintenance of the CBR system will be given.

2 Concept of the Case Based Reasoning approach

The application developed in the framework of this project is based on the Case Based Reasoning approach as described by Aamodt and Plaza (1994). In this approach, each remediation project stored in a database is described by a finite number of search parameters (referred to as attributes in the CBR software). Based on predefined **weights** for the search parameters and **similarities** between the allowed **values** of each **search parameter**, the level of similarity between a new project and the projects stored in the database can be determined. Search parameters are specially selected to cover the entire range of characteristics that users want to search for. The number of search parameters is much smaller than the total number of parameters that describe an in-situ remediation project. This complete information about the projects stored in the database is made available through Excel spreadsheets.
Searching for similar projects can be done on the basis of any number of search parameters. Increasing the number of search parameters obviously increases the relevance of the projects retrieved.
Software (CBR-Works 4) developed by the University of Kaiserslautern was used in this project. Below, the main definitions used by the CBR software that are relevant for the user, are explained shortly. **Weights** and **similarities** are fixed. They are defined by the expert team. **Filter** and **importance** can be changed by the user to enable optimal selection of similar projects.

Weights
The search parameters have a certain number/weight assigned. The weight reflects the relative importance of the different search parameters. These weights are based on technical considerations and cannot be changed by the user. However, they can be overruled by the user by using either filters or assigning a user-defined importance to the search parameter.

Similarities
The values that are assigned to the search parameters are often related to each other. For example, the technique and principle to remediate a soil contamination with diesel fuel may be comparable (similar) to the approach chosen for BTEX contamination. Hence, there is a similarity between the two contaminants. These similarities between values are laid down in a similarity table. Similarities vary between 0 and 1. They cannot be changed by the user.

Filters
By specifying a filter for a search parameter the search is restricted to those projects which satisfy the filter condition (equal to, not equal to, less than, greater than, etc.). Filter conditions are specified by the user.

Importance
The user defined importance partly overrules the weights that are assigned to the search parameters. Five levels of importance are defined.

Case Based Reasoning is a learning system. Adding new cases to the database will increase the relevance of the search for similar projects. It should be noted that also negative experiences can be added to the database.

3 Summary of activities

3.1 Evaluation of the work-process

This subproject aimed at the definition of the work-process and the identification of those elements in the work process that needed to be included in the database in order to judge the influence of the work-process on the quality of the remediation and the results obtained. For a detailed description of this sub-project the reader is referred to Kamermans and de Kreuk (2003).

The following activities were carried out in this subproject:
1. A description of a work-process was proposed;
2. Representatives of the authorities dealing with remediation permissions were interviewed about their experience;
3. A workshop was organized with participation of specialists from regulatory authorities, consultants and contractors;
4. Results of an internal BioSoil workshop on success factors in in-situ remediation projects were included in the study;
5. Twelve completed in-situ remediation projects were evaluated;
6. Information from ongoing SKB-projects (SV-039: a new remediation evaluation system and ROSA) was gathered and evaluated;
7. A number of elements, which were considered relevant for the system, were selected and included in the database.

Ad 1. The complete work-process consists of: identifying the need for a contamination inventory, performing the required studies, drafting a remediation plan, obtaining permission for the remediation, carrying out the actual remediation and evaluating the final results.

Ad 2. The interviews with representatives of the competent authorities revealed the following:
1) the contamination should be mapped out more correctly. The higher costs involved are considered acceptable when compared to the total cost of the remediation itself. A better design of the remediation systems will result in possible cost savings.
2) the processes which form the basis of an in-situ remediation, should be better understood. This would lead to better results and, if properly explained, facilitate the decisions of the competent authorities.
3) remediation targets must be better defined, in particular when the so called Stable Final Situation (SFS) is identified as the target. This will enable the competent authorities to make a better decision regarding the remediation plan and to define the elbowroom for the stakeholders.

Ad 3. In the workshop, both the work-process and a number of statements derived from the interviews were discussed It was concluded that:
1) At present basic information which is important for performing the remediation, such as the size and limits of the source area, detailed geological data, data on the expected degradation rates, etc. is often missing.

2) Remediation techniques should not be chosen at the start of the inventories just aiming at a reduction of the diversity of the information to be obtained.
3) SFS-criteria are now accepted as a remediation target. They should be defined for each project separately. Milestones should be indicated as well, because these are needed to be able to judge the progress of a project leading to a SFS during the 30 years timeframe over which a project may be carried out.
4) Authorities tend to judge a plan according to procedures rather than on the basis of knowledge of techniques and the merits of a plan.
5) It is important to indicate whether a remediation is target (concentration) or effort based, whether a risk analyses was made in relation to the SFS and whether the contractor was involved in drafting the remediation plan and in the design of the in-situ remediation systems.

Ad 4. In BioSoil's workshop on success factors for in-situ remediation projects, emphasis was put on the monitoring of the progress of decontamination in relation to a proper "snap shot" of the contamination at the start of the remediation.

Ad 5. An analysis of twelve finished in-situ projects showed the relevance of properly defined remediation targets. Success rates increased when the source areas were excavated (for mineral oil mainly) and when a repeated design process was applied. Proper communication between stakeholders appeared to be very important as prevention against different expectations of the sponsor, competent authority and contractor.

Ad 6. Information from the SKB projects Rosa and SV-039 was evaluated in order to identify elements that are relevant for the remediation projects, and which should be included in the CBR-database. In SV-039 a database was foreseen, which role may be fulfilled by the CBR-database.
As far as the work-process was concerned, the interviews held in the ROSA-project did yield about the same results as the CBR-workshop. Between the competent authorities, consultants and contractors both a lack of uniformity regarding the implementation of legislation and a growing gap in knowledge about techniques used in (in-situ) remediation existed.
The formats defined in the SV-039 project aiming at the evaluation of finished remediation projects were studied and elements such as Globis number, level and nature of contamination, quantities of excavated soil, extracted groundwater, the presence and extent of (D)NAPL's, treatment systems applied, etc. were included in the CBR-database.

Ad 7. From the various activities indicated above, the following list of elements to be included in the CBR-database was made:
Who took the initiative for the remediation?
Was a risk evaluation made?
Was the project effort or target based?
Was the contractor involved in the design of the plan?
What was the quality of the communication between stakeholders?
Were the remediation targets properly defined?
Have the remediation targets been met and if not, why?
What was the attention framework for remediation aspects?

Was a pilot test performed?
Which treatment techniques for groundwater (biological, activated carbon, stripping, chemical) and for extracted soil vapour (biological, activated carbon, chemical) were applied?
A possibility should be created to enter special, important remarks and information involving the project. These can be consulted when a project is selected from the database.

3.2 Structure of the CBR database and storing in-situ remediation projects

3.2.1 Structure of the database

Based on the results of the work process evaluation, a preliminary study of BioSoil and several sessions with TNO, BioSoil and Tauw experts, a set of parameters and corresponding values that describe the entire remediation project have been selected. These parameters are collected in an Excel spreadsheet. Each project is described by more than 250 parameters. A detailed description of these parameters is given in a report by Pijls and Keijzer (2003). This report also serves as a guide for the user to fill in the Excel spreadsheet in defining new remediation cases.

The Excel spreadsheet is structured in worksheets according to the following sequence:

General information
Work process
Soil and geohydrology
Contamination
In situ parameters
Remediation system
Remediation results
Remarks
CBR data, which are included in the Reasoner

Each worksheet contains a number of parameters. The final worksheet, 'CBR data', summarizes the selected search parameters for the Case Based Reasoner and is generated automatically based on the contents of the previous results. The 25 search parameters are:

Contaminant 1
Contaminant 2
Free phase present
Geology
Source treatment- principle remediation technology
Plume treatment-principle remediation technology
Status
Contamination cause
Vertical distribution
Maximum depth soil contamination
Maximum depth groundwater contamination
Volume of contamination (soil)

Volume of contamination (groundwater)
Maximum concentration (soil)
Maximum concentration (groundwater)
Remediation target
Objectives achieved
Approval of competent authorities
Special circumstances
Land-use after remediation
Municipality
Country
Longitude
Latitude
Globis number

These search parameters are used by the CBR search engine. In addition to the selection of the parameters and search parameters (called attributes for the search engine), the expert team defined weights and similarities for these attributes. These weights and similarities are given in App. A.

3.2.2 Data flow and quality control for new cases in CBR

Tauw supplied 19 in-situ remediation cases, BioSoil 50 cases, HMVT 10 cases and HBG-BAM 5 cases to the CBR database. More than 250 parameters have been introduced per case. The gathering of the data took about 1/2 an hour to 3 hours per case, depending on the familiarity of the case deliverer to the case.

The cases were introduced by ways of a specially designed Excel spreadsheet. Each spreadsheet contains one case only and consists of more than 250 data entries. The input of the spreadsheet is subject to a number of restrictions. These restrictions are required to maintain uniformity between the cases and hence obtain consistent search parameters for the CBR software.

The entries can be either (1) a description or type, (2) a number or (3) a range. Only one description, value or range can be chosen for each parameter,

There is a restriction to the number of possible descriptions, the range of values or the number of ranges. For example, the "value" of the parameter 'Free phase present' can only be yes or no, the remediation costs are divided into 5 categories (ranges) and the depth of the contaminated soil is a number which has a certain maximum value.
Only for a restricted number of fields, i.e., the comment fields, it is possible to enter free formatted text.
The guidance document provided with the Excel sheet gives a detailed description for each parameter (Pijls and Keijzer, 2003). This avoids misinterpretation of the case parameters.

Of the more than 250 entries in the spreadsheet, 25 parameters were selected to be used as search parameters in the CBR system. They are automatically generated in the spreadsheet and collected on a separate worksheet. These search parameters are introduced into the CBR software (CBR works). The following steps are needed to introduce the search parameters in the CBR program:

Translate the data to a database such as Access,
Translate the data into the format of CBR-works,
Introduce extra information needed in CBR, such as weights of the search parameters and similarities, installation of the application.

3.3 Geology in the database

The geology of a remediation site is described on a general level by one of the 8 pre-defined geology types i.e., (1) homogeneous and high permeability, (2) heterogeneous and moderate permeability, (3) sand on a low permeable layer, (4) low permeable cover, (5) homogeneous and low permeability, (6) sand on bedrock, (7) weathered and original bedrock and (8) bedrock (Pijls and Keijzer, 2003).

In addition to the general geological data in the database, extra geological information should be added through descriptions of boreholes close to the remediation site. These borehole descriptions should be supplied by the case owners. The borehole descriptions not only adds extra information to one case (it gives more body to the case), but it can also be useful to provide geological information for new cases based on coordinates. For new cases, nearby situated cases can be found relatively easily, and therefore, nearby geological information is easily accessible. The borehole descriptions must satisfy the NEN 5104 norm as is also required for ' standard soil research ' for contaminated sites in the Netherlands. It is recommended to include at least 2 shallow borehole descriptions (around 5 m deep) and one deep borehole description (around 20 m deep).

The project plan required addition of a geological map should be added to the database. This map should provide on a detailed level information about the sedimentation processes characteristics for the remediation site. It should be a cut-out of around 5-km by 5-km of the area around the remediation site. This additional geological information could provide an indication of the expected heterogeneity on a scale that might be not detected by the borehole descriptions alone. Soil heterogeneity directly affects the geohydrological accessibility of the contamination, and therefore has considerable effect on the in situ remediation efficiency.
This type of information is available for the Netherlands in a geological map of scale 1:500.000 . However, the scale of the 1:500.000 map is too coarse to give satisfactory information on a remediation site scale. TNO NITG is working on a map on a scale of 1:50.000, but this map is still not complete. It is recommended including the more detailed geological information when the 1:50.000 map has been completed.

The geological information as given by the borehole descriptions and the geological map are as yet not part of the data stored in the Excel spreadsheets. They should be included in a later stage.

3.4 Validation of the CBR database

The setup of the CBR database, defined by the search parameters, the weights for the search parameters and the similarities between the different values for the search parameters, as presented in appendix. A, are the result of expert opinion and thorough validation.
Criteria for the validation were defined as:

- a user with a general knowledge level should be able to use the search engine;
- results of a query for the CBR search engine should lead to a case which is fairly similar to the case that was given as query;
- experts should be able to verify the search results;
- the Excel spreadsheet should be able to describe an in-situ remediation case satisfactorily.

The user interface has not been evaluated, since in the final version of the application, a web based version will be used with a user interface different from the current one.

The validation was carried out by the project team and by external experts. At the time of testing, the database contained 60 remediation cases.

Project team
In the first step, the search engine and database were tested by the project team itself. One case, that was not included in the database, was used as a search case. It was concluded that the differences in weights for the different search parameters had to be increased in order to increase the distinguishing effect of the different parameters. Originally the ratio of the weights for the search parameter with the highest importance and the search parameter with the lowest importance was 5:1. This ratio had to be increased to 125:1, where the original individual weights were raised to the power 3. This resulted in the new weights as presented in appendix A. These weights gave a satisfactory response.

External experts
Next, the search engine and the database were tested by a team consisting of members of the consortium and external experts. The team consisted of 10 experts with different backgrounds, such as the representatives of the authorities, the national department of the environmental, contractors and consultants. The testing took 2 days and each tester brought his or her own real case and asked for similar cases.
The goal of this session was to get answers pertaining to both the complete database (Excel spreadsheets) and the search engine.
With respect to the complete database, the following questions were posed:

1. Are the parameters defining a case adequate, or should parameters describing the case be added or deleted?
2. Is the structure of the excel spreadsheet, with 8 different worksheets describing each a part of the case, logical and easy to work with?
3. Is the range of possible values that a parameter is allowed to have, satisfactory or should the range be extended?
4. Is the definition of the parameters clear? Does every (relative) expert understand the explanation given in the user's guide or can the parameter be interpreted differently by different experts?

Regarding the search engine the following questions had to be answered by the participants:

1. Are there attributes (search parameters) that should be added or deleted?
2. Are the questions posed for each attribute understandable and logic?
3. Are you satisfied by the results that the CBR search engine has given for your query? What would you have expected to be different?
4. What questions did you want to be answered by the search engine?
5. Have the options filter and importance been used and was it clear how and when to use these options?

The testers were also asked to give their opinion about the similarities and weights used. Results of the tests are given in Chapter 5.

4 Estimates of remediation costs and time

4.1 Introduction

In the project plan it was foreseen that the database should be supported (at least temporarily) by adding fictive generalized cases. These fictive generalized cases will be generated by using modelling techniques. Gaps in the database should be filled in with these generalized cases. Especially, the relationship between geology and the effectiveness of the remediation techniques should have been added to the database by using modelling techniques. However, the focus of CBR has changed during the project. A detailed description of the geology that was foreseen, together with a detailed study of the effects of geological aspects on the effectiveness of a remediation method, was replaced by a more general description considering 8 possible types of geology only. Therefore, the goal of the modelling support has been changed as well, i.e., the modelling should give some basic information (such as remediation time and costs) about cases not yet available in the CBR database.

In the following section the gaps in the database are discussed for the initial 60 cases. Next, two modelling approaches that can be applied to fill these gaps will be discussed. Modelling results will be presented in Chapter 5.

4.2 Coverage of the database

To get an impression about the coverage of the database, the number of cases available for the most relevant parameters was evaluated for the initial 60 cases. The parameters chosen are the search parameter that had been given the highest weights in CBR. These parameters are:
Geology type (weight 125),
Type of contaminant (weight 125),
Source treatment (weight 64),
Plume treatment (weight 64),
Free phase presence (weight 64),
Heterogeneity level (weight 27),
Volume of contaminated soil (weight 27),
Volume of contaminated groundwater (weight 27),
Maximum depth contamination in soil (weight 27),
Maximum depth contamination in groundwater (weight 27).

The number of cases for each parameter type is given in the tables below. Obviously, there are no cases in the CBR database with bedrock geology. This is due to the fact that mainly Dutch cases have been introduced until now. Furthermore, there are few cases with BTEX contamination, heavy oil or Polyaromatic hydrocarbons. The range of possible remediation techniques is sparsely covered by the CBR database because in the project only biological in situ techniques were introduced in the database. Mostly biological treatment of the source area has been applied. Obviously, effort should be dedicated to include projects in the database that use other techniques. The contaminated volumes of groundwater and soil are fairly well covered and the depth of the contamination is relatively well covered given that fact that only Dutch soils have been introduced (very deep contaminated sites were not included in the database).

Geology types	nr. of cases
Homogeneous and high permeable	24
Heterogeneous and moderately permeable	27
Sand on low permeable layer	5
Low permeable cover	2
Homogeneous and low permeable	2
Sand on bedrock	0
Weathered on original bedrock	0
Bedrock	0

Table 1 Number of cases in the database for each geology type

Contaminant type	nr. of cases
BTEX	6
Chlorinated Hydrocarbons	4
Hydrocarbons; diesel/kerosene	25
Hydrocarbons; heavy oil	4
Hydrocarbons; petrol	16
Polyaromatic Hydrocarbons	3
Other	2

Table 2 Number of cases for each contaminant type

Combinations of treatment		nr. of cases
Source	**Plume**	
Biological	None	42
Sparging	Sparging	2
Sparging	Pump & treat	4
Sparging	None	3
Soil vapor extraction	Pump & treat	2
Soil vapor extraction	Unknown	4
Soil vapor extraction	Biological	1
Pump & treat	none	1
None	Sparging	1

Table 3 Number of cases for each combination of source and plume treatment

Free phase present?	nr. of cases
Yes	14
No	46

Table 4 Number of cases with and without free phase product

Heterogeneity level	nr. of cases
Low	27
Medium	15
High	18

Table 5 Number of cases for each heterogeneity level

Volume in soil (m3)	nr. of cases
0-200	11
200-7500	12
500-2000	12
2000-5000	12
5000-20.000	5
Unknown	8

Table 6 Number of cases for each range of volume of contaminated soil

Volume in groundwater (m3)	nr. of cases
0-500	5
500-2000	15
2000-5000	9
5000-20.000	13
20.000-50.000	6
>50.000	2
Unknown	10

Table 7 Number of cases for each range of volume of contaminated groundwater

Soil (m) / Groundwater (m)	<5	5-10	unknown
<5	10	-	2
5-10	19	11	3
>10	6	5	0
Unknown	-	3	1

Table 8 Cases of each combination of contaminated depth of soil and groundwater (m)

4.3 Modeling approaches to fill the gaps in the database

Exploring the use of modelling to fill gaps in the database, did lead to two possible modelling approaches; a neural network approach (Dowla and Rogers, 1995) and a physically based, numerical flow and transport model such as RT3D (Clement, 1997) or STOMP (White and Oostrom, 1996) .

A neural network appears to be the more appropriate approach to “add” new cases to the CBR database. Complex relationships between parameters are not assumed to be known a-priori, but are determined on the basis of existing cases. To that end, the neural network uses a large amount of input and associated output data. In fact, the model trains itself on this given data and, based on the training results; it provides output, c.q. predictions, for new combinations of input data. Similar to CBR, it is able to handle soft data, such as the data describing the quality of the work process. Neural networks can only be used for interpolation between cases. Extrapolation will result in the average value of all outputs supplied for training.

The application of a neural network is restricted due to the fact that a considerable amount of data (projects) is needed for the network to learn, especially if a lot of different types of input and output data are involved. This implies that the neural network should be designed such that a minimal number of input and output data is required.
The application of physical/numerical flow and transport models is also restricted. These models require a lot of information, while only hard data can be used. Physical/numerical models are able to extrapolate (although they need a lot of time and effort). Since a lot of assumptions are made to build these types of models, the best way to use them is to evaluate the differences between two modelling results, rather than to use the quantitative calculation results itself. Still, a lot of effort will be required to model a new case in order to get the right answers. Trained neural networks can generate cases much faster.

4.4 Discussion

For both approaches it is important to formulate very clearly the required output from the CBR database. The projects added by the modelling have to be stripped until only the basics of remediation projects remain. The carefully gathered 250 parameters that describe a case must be decreased to a small number of parameters that maintain sufficient information to generate the most relevant output.

During discussions in the CBR project team and during the testing of the CBR application and database 3 questions that had to be answered by CBR came up most frequently:

What is the most appropriate remediation technique for a contaminated site?
What are the expected remediation costs?
What is the expected remediation time?

The decision was made not to add so much imaginary cases that all parameters are equally covered in the database. This would require thousands of cases to be added. Rather, modelling will provide a tool that gives (limited) outputs based on the data available in CBR and on the input of a (limited) number of search parameters for the new case. Trained neural networks are ideal to provide such a tool, since they are able to generate output very fast.

4.5 Neural Networks

Following the questions as came forward in the discussions in the project team as indicated above, two Neural Networks were designed to predict remediation time and costs. The applied remediation techniques were not considered, since only few cases were available that were not biological treatment. However, remediation technique can be added in future when more cases are added to the CBR database. The input parameters used were the parameters that were selected also for the "gap-evaluation" in Section 4.2. As a first step, before designing the Neural Network an impression of the relationships between parameters was obtained , by evaluating the relationships between some of these parameters and the output parameters, costs and remediation time, for the cases that are in the database right now. These relationships will be discussed below.

4.5.1 Relationships between parameters

In Figures 4.1 to 4.5 the relationships between remediation costs and contaminated volume of soil, contaminated volume of groundwater, free phase present, heterogeneity, depth of the contamination in groundwater and geology type, respectively, is given. As can be observed in the figures, the relations between these parameters are not very clear. Even expected correlations between e.g., contaminated soil volume and remediation costs or between heterogeneity and remediation costs are not obvious. Focussing on the remediation time, no clear correlations can be observed. In Figures 4.6 to 4.8, the relationships between contaminated volume of soil, contaminated volume of groundwater and geology, respectively, and remediation time are shown. The low correlation between parameters can be explained by the fact that only 60 cases are available in the CBR database, and that for some parameters the coverage is very poor. Also, the relationships might not be as straight forward as expected. In addition, some specific aspects not taken into account may be relevant. Fortunately, neural networks are able to detect the not-so-obvious relationships between data.

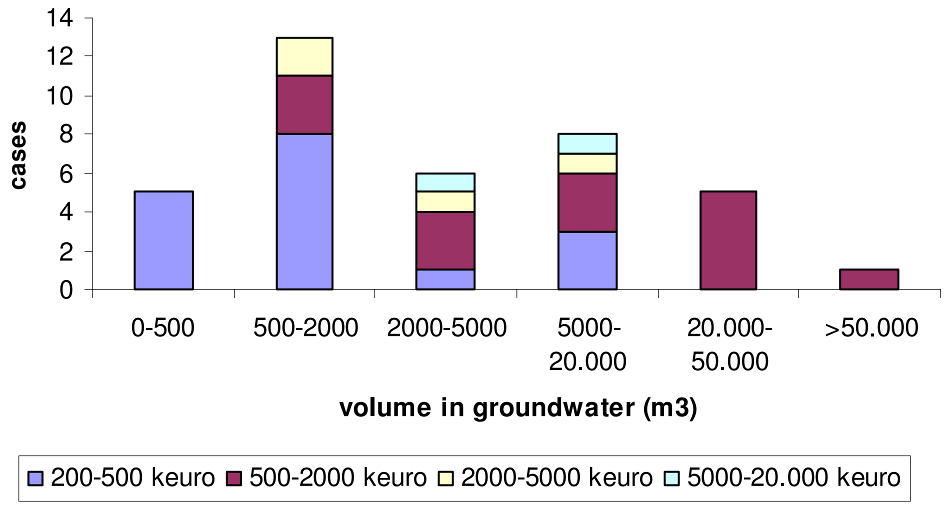

Figure 4.1Remediation costs related to volume of contaminated groundwater volume

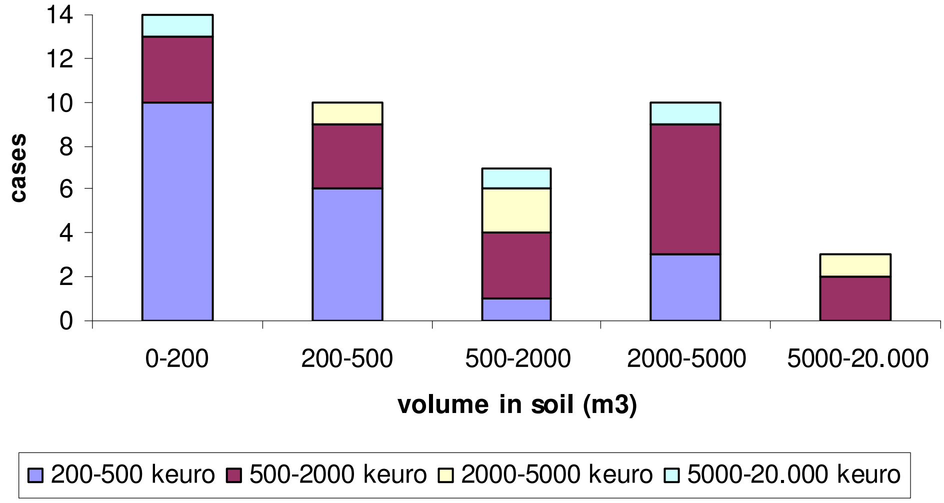

Figure 4.2 Remediation costs related to volume of contaminated groundwater.

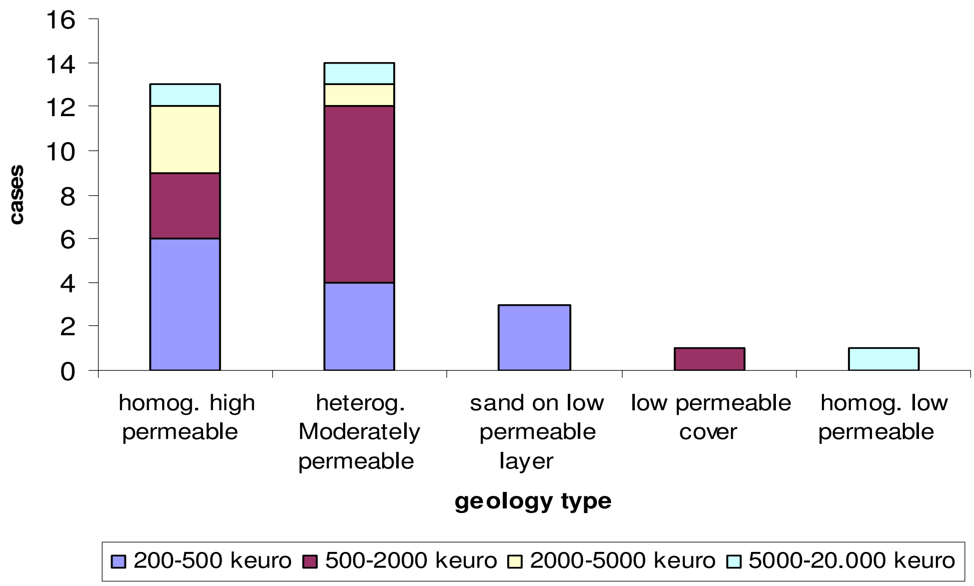

Figure 4.3 Remediation costs related to geology type.

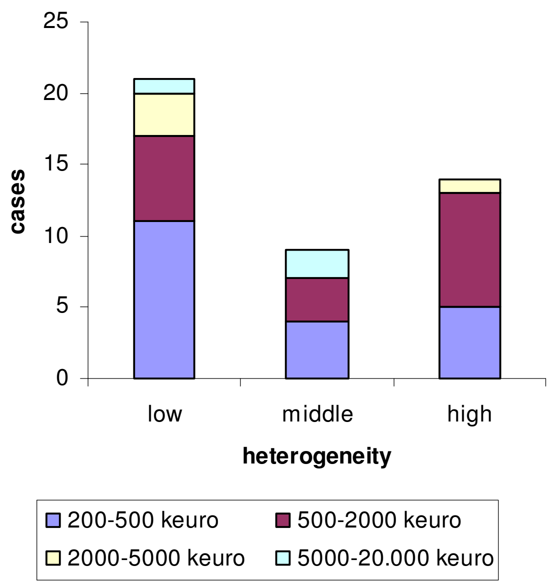

Figure 4.4 Remediation costs related to heterogeneity.

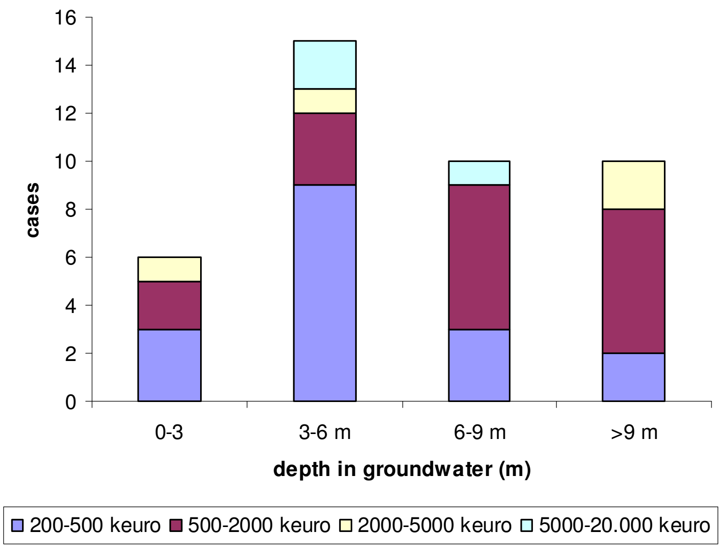

Figure 4.5 Remediation costs related to depth of contaminated groundwater.

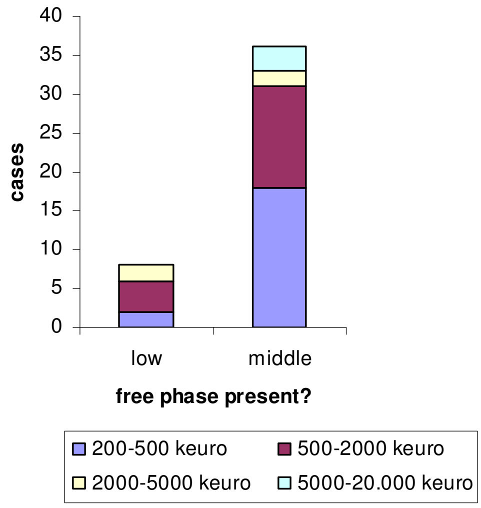

Figure 4.6 Remediation costs related to existance of the free phase

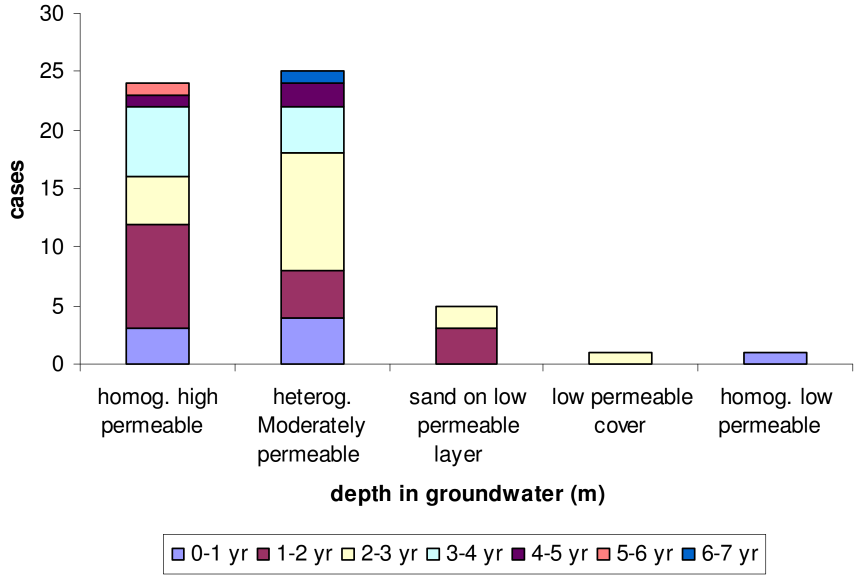

Figure 4.7 Remediation time related to geology type.

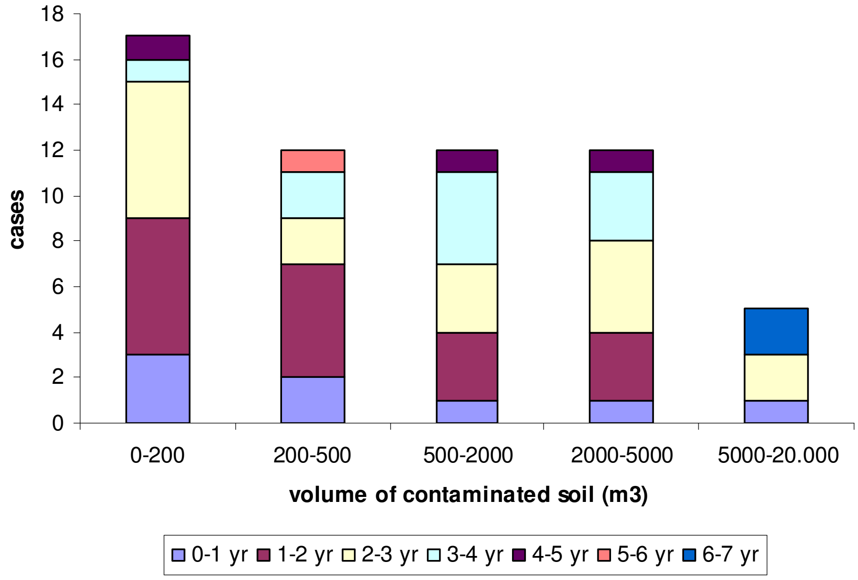

Figure 4.8 Remediation time relative to volume of contaminated soil.

5 Evaluation of results

5.1 Results of testing by external experts

The CBR system for in-situ remediation projects was evaluated very positively. The testers learned to use the Reasoner easily and the cases that were produced were generally satisfactory. It was especially valued that due to the development of the database the description of in-situ remediation projects became very well structured.

One of the main issues that became apparent during the testing was that it was not fully clear to the testers what the possibilities and restrictions of the Reasoner were. The testers supplied their own cases and looked for a similar case in the CBR database. They were, however, not always aware of the impact of their query. For example, when asked if the (new) case was finished or in intermediate stage, the testers choose intermediate, since their case was not finished yet. However, this choice resulted into a series of cases that were not finished either. Since the tester really wants to know what remediation technique to use, it would have been better to search for finished cases. At the start testers tended to try to fill in all search parameters, which is not really necessary. Later on, they understood that in some cases, to get a better answer to the question asked it is better to be selective.
In order to get a better understanding of the CBR Reasoner, it was decided that an additional user's guide of around 3 pages should be made available (App. A). Furthermore, it was suggested to ask for the search parameters in order of importance (weights, see Appendix A). This would help the user to understand which parameters are the most decisive to the result.

The testers were not always satisfied with the cases that were found by CBR. This was mainly due to the fact that only 60 cases were available and that therefore not all type of cases could be found. One tester did not find his own case, although it was included in the database. That was, however, due to a misinterpretation about the description of the geology. It was decided that definitions needed to be clarified and that parameters needed to be redefined. Furthermore, the testing showed that there was a need to add a search parameter called special characteristic. This special characteristic could be circumstances such as the existence of sand poles or a dike etc.

The testing further resulted in adding a few search parameters and also in the removal of 2 of them. Some parameters were redefined. Weights and similarities were not changed.

5.2 Results of the modelling

Two neural networks were developed to predict the remediation costs and time of new unknown cases, based on the data from the available cases in CBR.

Remediation time

The first network had as output parameter the remediation time and as input parameters:

geology type,
contaminant type,
source treatment,
free phase present,
heterogeneity,
source treatment,
volume of contaminated soil and
maximum depth of the contamination.

The latter is translated as a number that weights the depth in groundwater and the depth in soil. The network was trained using 54 cases and it was tested with the other 6 cases. Results are given in figure 5.1:

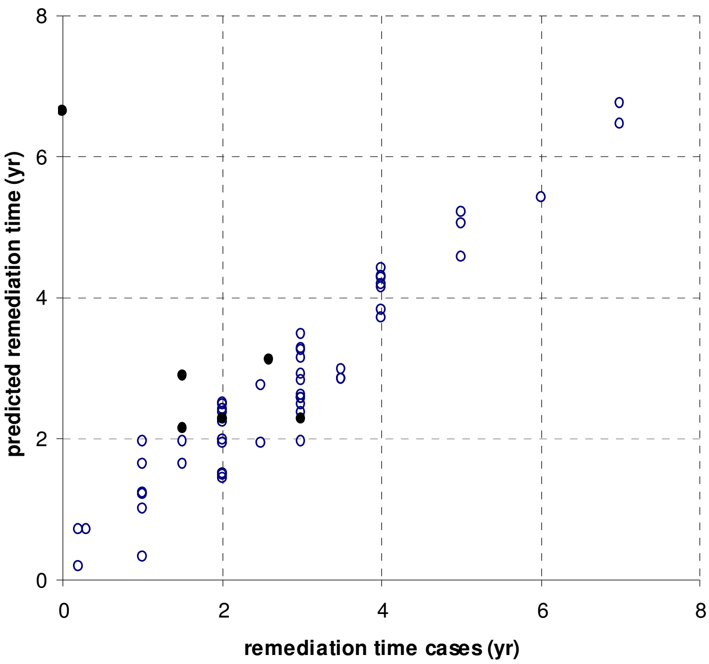

Figure 5.1 Remediation time predicted by the Neural Network versus the real remediation time. Open dots are training cases and solid dots are test cases

The figure shows the relationship between the real remediation times and the remediation times predicted with the neural network for the cases used for the training (open dots) and for the cases used as test cases (solid dots). The figure shows that in general the prediction of the remediation time is satisfactory. The one test that was not predicted well is a case where the remediation time was unknown, which is interpreted as zero remediation time. It is expected that when more cases are added as training cases, the results will improve.

Remediation costs

For the prediction of the remediation costs the same input data as for the prediction of the remediation time was used. It can not be represented by a figure since *ranges* of costs are given. In the table below the predictions of the neural network are presented for the training data and the test data:

Remediation costs	Training cases (54)		Test cases (6)	
Prediction	Within range	Outside range	Within range	Outside range
< € 100.000	15	0	2	1
€ 100.000 - € 300.000	14	2	0	1
€ 300.000 - € 500.000	3	0	0	1
€ 500.000 - € 1.000.000	2	0	0	0
unknown	14	0	0	1

The test cases are not predicted well, which is probably partly due to the fact that in 14 cases the costs are unknown. Improvement is required to get more satisfactory results. It is expected that when 100 cases or more are available for the training the prediction of the remediation costs by the neural network will improve.

In general, to obtain better results with the neural network, case owners should be pressed to include remediation costs and time to their new cases.

6 Conclusions and recommendations

6.1 Conclusions

In this project a Case Based Reasoner system has been developed that enables the information transfer between remediation project participants in and outside of the Netherlands. The system consists of a database and an intelligent search engine referred to as reasoner. It can be consulted by people who want to perform an in-situ remediation in such a manner, that projects with comparable characteristics can be used as a reference for future activities. It can also be used for improvement of remediation approaches.

In addition, elements in the work-process that obstruct the use of the available knowledge, are identified and included in the database in order to be able to judge the influence of the work-process on the remediation results. The reader is referred to Kamermans and de Kreuk (2003) for a detailed evaluation considering the improvement of the work process.
Based on a limited number of characteristics an estimate of the expected remediation time and costs of new soil contamination cases can be given.

Users working in the field of soil remediation have evaluated the system positively. However, the selection of cases by the CBR system will improve if more projects are added to the database. Also the tool that estimates remediation time and costs will improve if more projects are used and, if for each remediation project the database has been filled in as complete as possible (including e.g. remediation costs). To new cases geological information should be added by way of 2 shallow borehole descriptions and one deep borehole description. Cut out of geological maps should be added as soon as 1:50.000 geological maps become available. These maps are currently developed by TNO NITG.

6.2 Recommendations

It is recommended that new cases are added to the database. Especially, extra effort should be dedicated to include projects that improve the coverage of the database with emphasis on remediation techniques other than biological treatment. Although the database is operational, it is recommended to enhance the accessibility in the future, but also to make people aware of the advantages of using the CBR database and to deliver new cases to the database. Therefore a maintenance plan should be developed which can guarantee the accessibility of the database and its quality level. This maintenance plan also guarantees regular updates in cooperation with the users, to keep track on the demands of the users of the system.

Two types of maintenance activities can be distinguished; the qualitative maintenance and the database maintenance. The reader is referred to the maintenance document (Appendix B, in Dutch) for a detailed description of the plan and the related costs:
For the qualitative maintenance, the following activities should be carried out (on regular bases):

- Stimulating the soil remediation market to supply projects to the system.

- Quality control of the new cases. An expert team has to evaluate the data of new cases based on the following reports that must be delivered to the new case; the Soil Assessment, Remediation Plan and Remediation Evaluation. The cases will have to be judged on (1) the completeness of the required data (2) reliability and consistency, (3) the correct format.
- Consider in a rigorous way adaptation or extension of the system such as (search) parameters, weights and similarities.
- Organization of training sessions.
- Inventory of the requirements of the users.

For the database maintenance, the following activities should be carried out:
- Development of the web application and a user friendly interface.

On regular bases the following activities should be carried out:
- Support of the users. Users should be able to direct their questions to a helpdesk
- Adding of new cases to the database. Since the database and the search engine are coupled the application has to be installed each time new cases are added.
- Improvement of the database and the application to the new requirements of the users. This should be done for new cases only.

It is recommended to evaluate and reconsider the maintenance activities after 2 years and reconsider follow-up activities.

7 References

Aamodt, A. and E. Plaza, "Case-Based Reasoning: Foundational Issues, Methodological Variations and System Approaches", *AICOM*, Vol. 7, No. 1, pp. 39-59, 1994.

Clement, T.P., "RT3D, A Modular Computer Code for Simulating Reactive Multispecies Transport in 3-Dimensional Groundwater Systems", PNNL-11720, Pacific Northwest National Laboratory, Richland, WA, 1997.

Dowla, F.U. and Rogers, L.L. '' Solving problems in Environmental Engineering and Geosciences with Artificial Neural Networks'' The MIT press, Massachusetts, 1995.

Kamermans, C.M. and J.F. de Kreuk, " Case Based Reasoning: hidden knowledge of soil in-situ remediation unveiled'' Stichting Kennisontwikkeling Kennisoverdracht Bodem, projectnr SV613, Breda 2003.

Pijls, C. and Th. Keijzer, " Case Based Reasoning for in-situ remediation: Structure of the CBR database", Stichting Kennisontwikkeling Kennisoverdracht Bodem, projectnr SV613, Breda 2003.

White, M.D. and M. Oostrom, STOMP, "Subsurface Transport Over Multiple Phases. Theory Guide." PNNL-11217. Pacific Northwest National Laboratory, Richland, WA, 1996.

Appendix A Synopsis of the CBR search parameters

Case Based Reasoning

Synopsis of the CBR search parameters

Draft

5 December 2003

1 Introduction

1.1 General introduction

The SKB Case Based Reasoning project aims at improving soil remediation projects by identifying success and problem areas on the basis of practice over the complete duration of the project (preliminary field studies, remediation and its results concerning decontamination, time and costs). The CBR system is set up with the data of evaluated biological in situ remediation projects.

End users can asses the database in the CBR system by introducing **search parameters** for a similar case. The CBR program will select projects from the database based on the relative **weights** of the search parameters and the **similarities** between the allowed **values** for each search parameter (see also definitions).
In total 22 search parameters can be entered by the end user in the CBR system to retrieve similar cases. These search parameters cover the entire range of a case from type of contaminant, its distribution in the soil and groundwater, the technique applied etcetera. In the CBR system each search parameter is given a specific weight indicating its importance relative to the other search parameters. Furthermore, cases are also retrieved by the similarities between the allowed values for a given search parameter.

In this document the search parameters are briefly described for the user of the CBR application, together with the weight that is assigned to the parameter relative to the other parameters. Knowledge of the search parameter, the topic and scope it covers and intentions of the builders of the CBR system, helps the user of the CBR system in understanding the order and match of the retrieved cases. Similarities between the values for a search parameter are also given.

1.2 Definitions

Search parameter
The CBR database contains 22 search parameters that can be used to search the database. These parameters are key parameters in the description of the cases within the CBR database and cover all topics related to the remediations. The database can be searched by entering one, a number or all search parameters. Increasing the number of search parameters will increase the relevance of the cases retrieved. CBR Woks refers to the search parameters as **attributes**.

Weight
The search parameters have a certain weight assigned. This weight reflects the relative importance of the different search parameters. The weight is assigned to the search parameters based on technical considerations for the application of an in situ technique. For example, the soils geology and type of contaminant are regarded as important parameters for the type and design of the in situ measures, thus both search parameters have a high weight factor. The municipality or country is of lesser importance. If the user searches for the cases with a BTEX contamination it doesn’t really matter where it was.

Assigned weights in the CBR database can be overruled by the end-user by either using **filters** or assigning a user defined **importance** to the search parameter.

Similarities
For most search parameters values are assigned that are common or even obvious for the search parameter. These values are often related to each other. For example, the technique and principles to remediate a soil contamination with diesel fuel may be comparable (similar) to the approach chosen for a BTEX contamination. Both contaminants also show similar transport and fate in soil and groundwater. Thus, there is a similarity between the two contaminants. These similarities between the different values for a search parameter are laid down in a similarities table.

Filters
For queries, the input for attributes allows to change the filter properties of attributes additional to its value. Filter is a property that influences the behaviour of the retrieval engine. By specifying a filter for an attribute you restrict the search to all those cases which fulfil the filter condition.

Importance
For queries, the input for attributes allows to change the importance properties of attributes additional to its value. Importance is a property that influences the behaviour of the retrieval engine. It influences the calculation of the similarity. The higher an attribute's importance the more it will count in the calculation of concept's similarity. One can choose of five levels of importance. Important: changing importance (default is 'Medium') can have a substantial impact on results!

2 Search parameters

For each parameter that can be used in the CBR system to search the database a short description is given together with the allowed values, its weight and the similarities between the pre-set values of the search parameter. The search parameters are listed based on the weight assigned to the parameter.

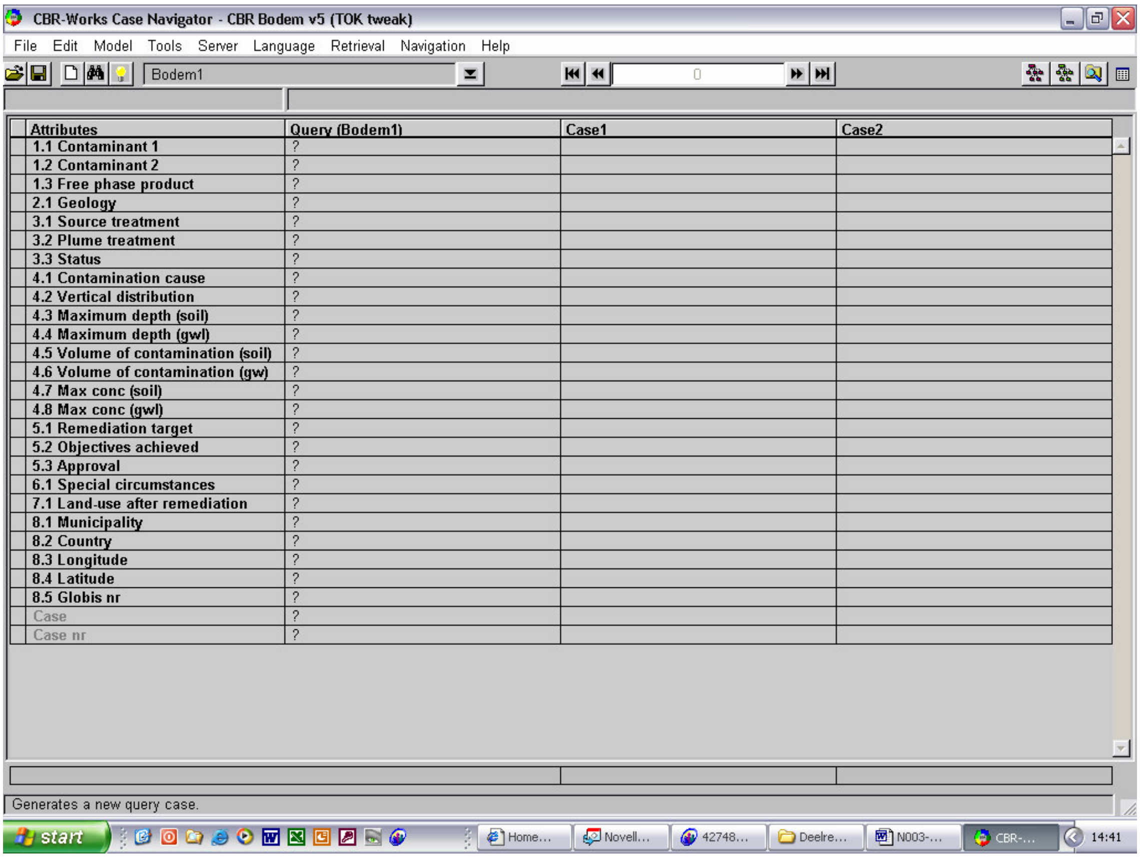

Figure 2.1 Lay-out of the case navigator showing the search parameters (Attributes)

2.1 Contaminant 1

This is the main contamination present in the soil. The remediation system is designed and objectives are defined based on this contaminant. Most common contaminants are available in the database as pre-set values. Rare or specific contaminants can be entered as Other. Specific contaminants that are available in the database can be search via the Specific Circumstances. For example, the database contains several cases with phthalates which can be directly accessed by selecting: Contaminant 1=Others, Special Circumstances=Phthalates.

Weight
125

Values
Hydrocarbons (TPH): petrol
Hydrocarbons (TPH): diesel / kerosine
Hydrocarbons (TPH): heavy oil
Volatile Chlorinated Hydrocarbons (VOC)
Aromates (BTEXN)
Polycyclic aromatic Hydrocarbons (PAH)
Pesticides
Other (see remarks)

Similarities

	HC petrol	HC diesel / kerosine	HC heavy oil	CHC	BTEX	PAH	Pesticides	Other
Hydrocarbons (HC) petrol	1							
Hydrocarbons (HC) diesel / kerosene	0.7	1						
Hydrocarbons (HC) heavy oil	0.3	0.7	1					
Volatile Chlorinated Hydrocarbons (CHC)	0	0	0	1				
Aromates (BTEX)	0.9	0.5	0.1	0.2	1			
Polycyclic aromatic Hydrocarbons (PAH)	0.1	0.2	0.8	0	0.2	1		
Pesticides	0	0	0	0.3	0	0.2	1	
Other (see remarks)	0	0	0	0	0	0	0	1

2.2 Contaminant 2

Besides the main contaminant a related compound can be present which may influence, positively or negatively the remediation. For example BTEX as a co-contaminant when light mineral oil such as petrol is present.

Weight
100

Values
Hydrocarbons (TPH): petrol
Hydrocarbons (TPH): diesel / kerosine
Hydrocarbons (TPH): heavy oil

Volatile Chlorinated Hydrocarbons (VOC)
Aromates (BTEXN)
Polycyclic aromatic Hydrocarbons (PAH)
Pesticides
Other (see remarks)

Similarities
None

2.3 Free phase product present?

At the start of the in situ remediation was there free phase present. Free phase covers both the light and dense non-aqueous phase liquids (LNAPL and DNAPL). If a pretreatment of the source area is performed, there will likely be no more free phase NAPL.

Weight
64

Values
Yes
No

Similarities
No comment

2.4 Geology

Geology refers to the geological and hydrological properties of the soil layer in which the contaminant was present and thus in which the remediation was performed. These properties largely determine the remediation technique applied. It does not refer to the entire geological profile of the soil at the given site.

Weight
125

Values
Homogeneous/high permeability (> 5 m/day)
Homogeneous/moderate permeability (0.5 - 5 m/day)
Homogeneous/low permeability (< 0.5 m/day)
Heterogeneous/moderate permeability (0.5 - 5 m/day)
Heterogeneous/low permeability (< 0.5 m/day)
Bedrock

Similarities

	Homogen. high perm.	Homogen. moderate perm.	Homogen. low perm.	Hetereog. moderate perm.	Hetereog. low perm.	Bedrock
Homogeneous/high permeability (> 5 m/day)	1					
Homogeneous/moderate permeability (0.5 - 5 m/day)	0.8	1				
Homogeneous/low permeability (< 0.5 m/day)	0.5	0.6	1			
Heterogeneous/moderate permeability (0.5 - 5 m/day)	0.3	0.4	0.4	1		
Heterogeneous/low permeability (< 0.5 m/day)	0.1	0.2	0.6	0.4	1	
Bedrock	0.3	0.1	0	0.2	0.4	1

2.5 Source treatment - principle remediation technology

The basic principle of the remediation process applied for the remediation of the source area of the contaminant.

Weight
64

Values
Biological aerobic
Biological anaerobic
Biological combined
Sparging
Soil vapour extraction (SVE)
Pump and treat (P&T)
Chemical oxidation (ISCO)
Thermal treatment
Natural attenuation (NA)

Similarities

	Biological aerobic	Biological anaerobic	Biological combined	Sparging	SVE	P&T	ISCO	Thermal	NA
Biological aerobic	1								
Biological anaerobic	0,2	1							
Biological combined	0,5	0,5	1						
Sparging	0,8	0	0,4	1					
SVE	0,2	0	0,2	0,2	1				
Pump and treat	0	0	0	0	0	1			

ISCO	0	0	0	0	0	0	1		
Thermal treatment	0	0	0	0	0,5	0,5	0	1	
Natural attenuation	0,2	0,4	0,3	0	0	0	0	0	1

2.6 Plume treatment - principle remediation technology

The basic principle of the remediation process applied for the remediation of the plume of the contaminant.

Weight
64

Values
Biological aerobic
Biological anaerobic
Biological combined
Sparging
Soil vapour extraction (SVE)
Pump and treat (P&T)
Chemical oxidation (ISCO)
Thermal treatment
Natural attenuation (NA)

Similarities

	Biological aerobic	**Biological anaerobic**	**Biological combined**	**Sparging**	**SVE**	**P&T**	**ISCO**	**Thermal**	**NA**
Biological aerobic	1								
Biological anaerobic	0,2	1							
Biological combined	0,5	0,5	1						
Sparging	0,8	0	0,4	1					
SVE	0,2	0	0,2	0,2	1				
Pump and treat	0	0	0	0	0	1			
ISCO	0	0	0	0	0	0	1		
Thermal treatment	0	0	0	0	0,5	0,5	0	1	
Natural attenuation	0,2	0,4	0,3	0	0	0	0	0	1

2.7 Status

The CBR system allows for the introduction of completed remediations as remediation projects that are still in progress. Pilots to determine if a specific technique is applicable at a given site are also entered in the database as cases.

Weight

27

Values
Pilot
Intermediate
Final

Similarities

	Pilot	Intermediate	Final
Pilot	1	0.2	0
Intermediate	0.2	1	0.5
Final	0	0.5	1

2.8 Contamination cause

Describes the cause of the contamination. As pre-set values the most common causes are listed which explain themselves. A soak away refers to a sump or in Dutch a *zinkput*.

Weight
27

Values
Spill
Rupture of piping
Storage tank, above ground
Storage tank, underground
Soak away
Sewer
Dumping
Unknown

Similarities

	Spill	Rupture of piping	Storage tank, above ground	Storage tank, underground	Soak away	Sewer	Dumping
Spill	1						
Rupture of piping	0.8	1					
Storage tank, above ground	0.5	0.5	1				
Storage tank, underground	0.2	0.6	0.3	1			
Soak away	0	0	0	0.2	1		
Sewer	0.2	0.2	0	0.5	0.6	1	
Dumping	0.8	0	0	0.1	0.2	0.1	1

2.9 Vertical distribution

Refers to the distribution of the contaminant in the soil and groundwater. Was the contaminant solely present in the unsaturated zone or saturated zone?

Weight
27

Values
Unsaturated (vadose) zone
Saturated zone
Vadose and saturated zone

Similarities

	Unsaturated (vadoze) zone	Saturated zone	Vadose and saturated zone
Unsaturated (vadoze) zone	1		
Saturated zone	0	1	
Vadose and saturated zone	0.5	0.5	1

2.10 Maximum depth soil contamination

The maximum depth at which the main contaminant was found in the soil in meters below ground level.

Weight
27

Values
None

Similarities
Similarities are calculated based on the relative distance between the values.

2.11 Maximum depth groundwater contamination

The maximum depth at which the main contaminant was found in the groundwater in meters below ground level.

Weight

27

Values
None

Similarities
Similarities are calculated based on the relative distance between the values.

2.12 Volume of contamination soil

Volume of the contamination (contaminant 1) in the soil in m^3.

Weight
27

Values
0-200
200-500
500-2000
2000-5000
5000-20000
>20000

Similarities
Similarities are calculated based on the relative distance between the values.

2.13 Volume of contamination groundwater

Volume of the contamination (contaminant 1) in the groundwater in m^3.

Weight
27

Values
0-500
500-2000
2000-5000
5000-20000
20000-50000
>50000

Similarities
Similarities are calculated based on the relative distance between the values.

2.14 Maximum concentration soil

The highest concentration of the main contaminant at the start of the in situ remediation actions in the soil. This concentration can be lower than determined in the soil survey as the source of the contamination may have been removed via excavation.

Weight
27

Values
0-500 TPH (Total Petroleum Hydrocarbons i.e. mineral oil)
500-2000 TPH
2000-5000 TPH
5000-20000 TPH
>20000 TPH
0-50 VOC (Volatile Organochloro Compounds i.e. Per, Tri)
50-500 VOC
>500 VOC
0-50 BTEXN (light aromatic compounds such as benzene, toluene)
50-500 BTEXN
>500 BTEXN
0-50 PAH (Polycyclic Aromatic Hydrocarbons)
50-500 PAH
>500 PAH
0-2 Pesticides
2-20 Pesticides
>20 Pesticides
0-50 Other
50-500 Other
>500 Other

Similarities
Similarities are calculated based on the relative distance between the values.

2.15 Maximum concentration groundwater

The highest concentration of the main contaminant at the start of the in situ remediation actions in the groundwater. This concentration can be lower than determined in the soil survey as part of the source of the contamination may have been removed.

Weight

27

Values
0-500 TPH
500-5000 TPH
>5000 TPH
0-1000 VOC
1000-10000 VOC
>10000 VOC
0-500 BTEXN
500-5000 BTEXN
>5000 BTEXN
0-5 PAH
5-50 PAH
0-10 Pesticides
10-100 Pesticides
>100 Pesticides
0-500 Other
500-5000 Other
>5000 Other

Similarities
Similarities are calculated based on the relative distance between the values.

2.16 Remediation target

During the work process before the actual remediation targets are laid down to determine whether a remediation was successful and can be stopped, or was unsuccessful and needs to be prolonged or to be adjusted. The two values for this search parameter are concentration related, i.e. when end concentrations of the contaminants are agreed upon between the problem owner and he competent authority. Situation related remediation targets refer to acceptable levels for the intended future use of the location.

Weight
27

Values
Concentration related
Situation related

Similarities
None

2.17 Objectives achieved

Are the remediation objectives reached. The objectives are laid down in the remediation targets for the main contaminant, for example are the desired concentrations as agreed between the competent authorities and the problem owner reached after the in situ remediation.

Weight
27

Values
Yes
No

Similarities
No comments

2.18 Approval of competent authorities?

Did the competent authorities gave their approval to the results of the remediation. For a given remediaton the authorities may have given their approval without the remediation targets being achieved or only partly achieved. Approval can be given to end the active part of the in situ remediation but additional activities such as monitoring may still apply to the location.

Weight
1

Values
Yes
No

Similarities
No comments

2.19 Special circumstances

Not all the remediations can be covered by the search parameters and values given in the CBR database. Special circumstance related to the geology, contaminant or site specific circumstances are in a number of cases worth added to the database thus allowing direct retrieval of these cases.

Weight
27

Values

Special circumstance identified for individual cases in the CBR database.

Similarities
None

2.20 Land-use after remediation

In most countries of the EU remediation targets depend on the land-use after the remediation. Threshold values and allowable values for most contaminants in soil and groundwater are stricter for soils used for domestic purposes compared to industrial use. Domestic use refers to soils underneath housing, gardens etc., whereas public ground refers to public parks, greens shrubbery etc. Agricultural use refers to soils used for grassland, farming and woodland.

Weight
8

Values
Domestic
Public ground
Industrial
Agricultural

Similarities

	Domestic	Public ground	Industrial	Agricultural
Domestic	1			
Public ground	0.5	1		
Industrial	0	0.2	1	
Agricultural	0.2	0.2	0.1	1

2.21 Municipality

Refers to the municipality in which the site that was remediated is situated. Only municipalities that are represented in the CBR database can be selected.

Weight
1

Values
Municipalities that are represented in the CBR database.

Similarities

None

2.22 Country

Description
Country refers to the country in which the remediation was conducted. Only countries that are represented in the CBR database can be selected.

Weight
1

Values
Countries that are represented in the CBR database.

Similarities
None

2.23 Longitude & latitude

Refers to the actual location of the site based on longitudal and lattitudal coordinates. Value must be entered as a real e.g. 54°23'45'' is entered as 54.2345.

Weight
8

Values
Longitude and latitude that are represented in de CBR database.

Similarities
Similarities are calculated based on the relative distance between the values for the longitude and latitude.

2.24 Globis number

Globis is a Dutch digital information system in which authorities, starting in 2003, will introduce and manage soil contamination and remediation information. If known a specific case can be retrieved solely on its Globis number. Older cases, predating the Globis data base, and remediations in the CBR database outside the Netherlands do not have a Globis number.

Weight

1

Values
Between 0 - 100000

Similarities
Similarities are calculated based on the relative distance between the values.

Appendix B (Concept) Beheersplan CBR-Bodem

1 Inleiding

In de afgelopen decennia is door een groot aantal partijen (aannemers, consultants) veel kennis en ervaring vergaard over de opzet en voortgang van in-situ bodemsaneringen. Deze kennis is in het algemeen niet makkelijk toegankelijk, omdat zij voornamelijk is opgeslagen in de hoofden van mensen en in minder makkelijk toegankelijke rapporten.
Om de toegankelijkheid van deze kennis te vergroten, is in 2003 het SKB project SV-613: "Case Based Reasoning: verborgen bodemkennis ontsloten" uitgevoerd. In het kader van dit project is informatie omtrent een honderdtal in-situ bodemsaneringen verzameld en opgeslagen in een database. Daarnaast is een applicatie (reasoner genaamd) ontwikkeld met behulp waarvan de database kan worden geraadpleegd, en waarmee op basis van een aantal strikte regels voor nieuwe saneringen kan worden nagegaan wat de overeenkomst is met in de database opgeslagen uitgevoerde saneringen is. Met behulp van deze techniek (Case Based Reasoning) kunnen potentiële gebruikers op eenvoudige wijze nieuwe bodemsaneringen vergelijken met afgeronde bodemsaneringen om zo maximaal te kunnen profiteren van bestaande kennis en op basis van die bestaande kennis tot een zo efficiënt mogelijke keuze van te gebruiken technieken en methoden te komen.

De in het kader van het project SV-613 ontwikkelde applicatie en database zijn operationeel. Om in de toekomst de beschikbaarheid van de applicatie en de database voor zoveel mogelijk gebruikers te waarborgen, dient een beheerssysteem te worden opgezet dat zorgt voor zowel het dagelijks beheer als de kwaliteitsborging van de database, waaronder ook het invoeren van nieuwe gevallen is begrepen. De database breidt zich zo uit, waardoor ook de waarde ervan toeneemt. In principe kunnen twee soorten beheersactiviteiten worden onderscheiden: het ICT beheer en het kwalitatief beheer. Voor het ICT beheer is een éénmalige investering noodzakelijk voor het creëren van een web applicatie die door een aantal gebruikers tegelijkertijd kan worden geraadpleegd. Met zowel het ICT beheer als het kwalitatief beheer zijn jaarlijks terugkerende kosten gemoeid. De uit te voeren activiteiten en de daarmee gemoeide kosten zullen in het navolgende in meer detail worden besproken.

2 ICT beheer

Een belangrijk onderdeel van het beheer van de CBR applicatie en de database is het ICT gerelateerde beheer. Eenmalig zullen een aantal activiteiten moeten worden ontplooid om te komen tot de ontwikkeling en implementatie van een web applicatie die via internet door een aantal gebruikers tegelijkertijd kan worden benaderd. Daarnaast zullen elk jaar opnieuw een aantal activiteiten moeten worden ontplooid om de applicatie en de database up to date en "in de lucht" te houden, en gebruikers te ondersteunen bij het gebruik van de CBR applicatie.
In het navolgende zullen de activiteiten nader worden uitgediept, zullen mogelijke oplossingen voor de uitvoering daarvan worden voorgesteld, en zal een schatting van de benodigde kosten worden gegeven.

2.1 Opzet van de web applicatie (eenmalig)

Bestaande commercieel beschikbare programmatuur voor Case Based Reasoning is in principe in staat om een web applicatie te creëren. Gegeven de ervaring die is opgedaan binnen het project SV-613 zijn er in principe drie opties om tot een web applicatie te komen:

1. het gebruik van het Franse software pakket KATE;
2. het gebruik van het Duitse software pakket CBR-Works;
3. ontwikkeling in eigen beheer.

Het Franse software pakket KATE is in principe het verst ontwikkeld en biedt de meeste mogelijkheden. Echter, de kosten die gemoeid zijn met licenties voor het gebruik van dit pakket (ca. € 150.000 in het eerste jaar en vervolgens ca. € 30.000 per jaar, inclusief ondersteuning en onderhoud) zijn dermate hoog dat dit niet als een realistische optie moet worden gezien.
In het kader van het project SV-613 is voornamelijk gewerkt met het pakket CBR-Works. Binnen het project is een licentie aangeschaft (kosten € 2.600) die “niet-commerciële” web applicaties zonder extra kosten toelaat. Gedurende de ontwikkeling van de applicatie gebaseerd op CBR-Works, bleken bij het genereren van de web applicatie een aantal problemen op te treden. Nader onderzoek leerde, dat deze problemen zonder al te grote inspanning kunnen worden opgelost. Een ruwe schatting van de kosten, die daarmee zijn gemoeid, is € 12.000.
Voor de ontwikkeling in eigen beheer zullen allereerst duidelijke specificaties van de eisen die aan de te ontwikkelen software worden gesteld moeten worden geformuleerd. Geschat wordt dat de ontwikkeling van nieuwe software dat dezelfde mogelijkheden biedt als de bestaande software ruwweg tussen € 40.000 en € 50.000 gaat kosten.

Eén van de belangrijke overwegingen bij de implementatie van de web applicatie is, dat in CBR-Works de applicatie en de database niet gescheiden zijn. M.a.w. de database vormt een integraal onderdeel van de applicatie. Normaliter zijn dat gescheiden eenheden, waardoor de beveiliging van een centrale server waarop de database is geïnstalleerd kan worden gewaarborgd. Omdat dat in het geval van CBR-Works niet het geval is, ligt het voor de hand dat de beheerder van het systeem waarop de applicatie beschikbaar wordt gesteld zal eisen dat hier een aparte server wordt aangeschaft. De (éénmalige) kosten die daarmee zijn gemoeid bedragen € 6.000 aanschaf en € 2.000 installatie.
Hoewel het voor de potentiële gebruiker niet van belang is waar de web applicatie wordt geïnstalleerd, is dat voor de uitvoering van het ICT beheer wel degelijk van belang. De betreffende organisatie zal ervoor moeten zorgen dat de applicatie en database maximaal beschikbaar zijn, en dat tijdig updates van de applicatie en database worden geïnstalleerd (zie ook jaarlijks terugkerende activiteiten). Logischerwijze zijn er twee opties bij de keuze van de instelling waar de applicatie wordt geïnstalleerd:

1. SKB: de applicatie/database is ontwikkeld in het kader van het SKB project SV-613, en is contractueel eigendom van SKB, met de daarbij behorende zorgplicht.
2. TNO-NITG: NITG heeft zowel de kennis omtrent de ontwikkelde applicatie/database in huis als veel ervaring op het gebied van het beschikbaar stellen van data aan derden via web applicaties.

De kosten gemoeid met de ontwikkeling en implementatie zijn onafhankelijk van de keuze van de instelling waar de applicatie wordt geïnstalleerd.

2.2 Jaarlijks terugkerende activiteiten

Jaarlijks moeten de volgende vier activiteiten worden ontplooid:

1. Helpdesk-activiteiten. Gebruikers moeten bij moeilijkheden met het gebruik van CBR-Bodem kunnen terugvallen op een helpdesk waar de meeste problemen direkt kunnen worden opgelost. Onderdeel van deze activiteiten is o.a. 1) het opzetten van een voor de gebruiker toegankelijke rubriek met FAQ (Frequently Asked Questions), waardoor de gebruiker in veel gevallen in staat zal zijn zelf de oplossing van zijn probleem te vinden en 2) het beschikbaar stellen van een e-mail adres waar de gebruiker eventuele wensen kan kenbaar maken.
De kosten gemoeid met deze activiteit worden geraamd op € 5.000.
Met nadruk dient erop te worden gewezen dat de helpdesk alleen beschikbaar is voor het beantwoorden van vragen over het gebruik van CBR-Bodem. Vragen die te maken hebben met de interpretatie van de uitkomsten van zoekacties in CBR-Bodem zullen worden doorverwezen naar de betreffende aannemer of consultant. Beantwoording van deze vragen zal in het algemeen niet kosteloos kunnen gebeuren.
2. Het invoeren van nieuwe cases in de database. Naar verwachting zullen in de loop der tijd nieuwe cases ter opname in de database worden aangeboden. Door de koppeling van de applicatie en de database zal bij uitbreiding van de database ook de applicatie opnieuw moeten worden geïnstalleerd. Hoewel er op dit moment geen zicht is op het aantal nieuwe cases dat in de komende tijd zal worden aangeboden (mede afhankelijk van de activiteiten van de groep die het kwalitatieve beheer doet) wordt ruwweg geschat dat er met deze activiteit een bedrag van € 6.000 is gemoeid. Nieuwe cases zullen dan in driemaandelijkse sessies worden ingevoerd.
3. Het aanpassen van de database en applicatie aan veranderende gebruikerswensen. Op basis van ervaringen die in de loop der tijd met CBR-Bodem worden opgedaan zullen er wensen ontstaan met betrekking tot aanpassing (verandering, uitbreiding) van de gegevens die worden gebruikt bij het zoeken in de database (attributen) en de gewichtsfactoren die aan de verschillende attributen worden toegekend. Deze aanpassingen vereisen beduidend meer werk per uitvoering dan het toevoegen van nieuwe cases aan de database, omdat bij het uitbreiden van het aantal attributen wellicht ook alle bestaande cases moeten worden aangepast. Uitbreiding van de weergegeven informatie per case zal alleen bij de nieuwe cases worden doorgevoerd, omdat anders de actie met betrekking tot het verzamelen van gegevens voor alle cases gedeeltelijk zou moeten worden herhaald. Dit is niet haalbaar en voor de oudere cases is ook minder informatie voorhanden dan voor de nieuwere. De verwachting is dat dit soort aanpassingen veel minder frequent zullen voorkomen dan uitbreiding van het aantal cases. De kosten voor deze aanpassingen worden ruwweg geschat op € 6.000.
4. Bij vrijwel alle activiteiten zal regelmatig ondersteuning van de ICT afdeling van de instelling die het beheer doet, nodig zijn. Ruwweg worden de kosten hiervoor geschat op € 3.000.
Met name de activiteiten genoemd onder 2 en 3 worden sterk beïnvloed door de activiteiten die de kwalitatieve beheersgroep zal ontwikkelen.
De kosten die zijn gerelateerd aan het ICT beheer zijn onafhankelijk van de instelling waar de applicatie wordt beheerd.

2.3 Voorstel voor uitvoering

Mede op grond van eerder gegeven argumenten ligt het voor de hand om voor het ICT beheer de volgende oplossing te kiezen:

- Ontwikkeling van de web applicatie op basis van CBR-Works.
- Implementatie van de web applicatie en database op een server bij TNO-NITG.

• Ondersteuning van gebruikers en onderhoud van applicatie en database uit te voeren door TNO-NITG.

Belangrijkste argument voor de ontwikkeling van de web applicatie op basis van CBR-Works zijn de kosten. Gebruik van het Franse pakket KATE is te duur, terwijl de ontwikkeling in eigen beheer teveel tijd in beslag zal nemen en daarmee ook te duur wordt. Met CBR-Works is in het project SV-613 al de nodige ervaring opgedaan, zodat geen onoverkomelijke problemen worden verwacht bij de ontwikkeling van de web applicatie.

Implementatie bij TNO-NITG ligt voor de hand om twee redenen. Allereerst beschikt TNO-NITG over een ruime ervaring wat betreft het beschikbaar stellen van applicaties en data aan derden via internet. Bovendien is kennis en ervaring met betrekking tot CBR-Works en de structuur van de database aanwezig, zodat zowel het “dagelijks” beheer is gewaarborgd alsook de ondersteuning van het kwalitatief beheer (aanpassingen aan database en applicatie). Aansluiting bij Globis en SV039 kan worden verzorgd door het aanbrengen van “links”, omdat een fysieke koppeling van CBR, Globis en SV039 niet is voorzien.

Deze keuze voor TNO-NITG als de ICT beheerder heeft geen invloed op de eenmalige of jaarlijkse kosten.

3 Kwalitatief beheer

Voor het kwalitatieve beheer van de CBR applicatie en database kunnen de volgende taken worden gedefinieerd:

• Het stimuleren van de bodemsaneringsmarkt om nieuwe cases in de CBR-Bodem in te brengen. Daartoe zal de kwalitatieve beheersgroep aansluiting moeten zoeken bij bij Globis en SV039.
• Ervoor zorgen dat nieuwe cases die in de database moeten worden ingebracht van “voldoende” kwaliteit zijn, d.w.z. dat de informatie met betrekking tot die nieuwe cases correct is. In dit verband zal door de kwalitatieve beheersgroep een programma van eisen moeten worden opgesteld voor het aanleveren en invoeren van nieuwe cases.
• Ervoor zorgen dat gefundeerde beslissingen worden genomen met betrekking tot eventuele veranderingen in de applicatie (verandering van de weging van de verschillende attributen, toevoegen van nieuwe attributen). Daarbij moet rekening worden gehouden met de wensen en ervaringen van de gebruikers, zoals die ook door helpdesk zullen worden verzameld.
• Het organiseren van cursussen voor het gebruik van CBR-Bodem. Met name in het begin zal het nodig zijn dat nieuwe gebruikers worden geintroduceerd in het gebruik van de CBR-Bodem. Degelijke cursussen (een halve dag tot maximaal één dag) kunnen worden georganiseerd als onderdeel van bestaande cursussen zoals georganiseerd door de stichting Postacademisch Onderwijs (PAO) te Delft (zoals bv. de cursus “Grondwatersanering door gestimuleerde biologische afbraak”).
• Het organiseren van regelmatige gebruikers bijeenkomsten om de gebruikerswensen te inventariseren en in samenspraak met de gebruikers te komen tot specificaties van mogelijke aanpassingen van database en applicatie.

Deze taken behelzen niet éénmalige werkzaamheden, doch betreffen terugkerende activiteiten. Om voor voldoende continuiteit in de uitvoering van deze taken te zorgen, dient er een permanente beheersgroep te worden ingesteld. Logischerwijze bestaat deze kwalitatieve beheersgroep uit vertegenwoordigers van de drie instellingen die de ontwikkeling van CBR-Bodem hebben gedragen: BioSoil, TAUW en TNO-NITG, zo mogelijk onder voorzittersschap van een SKB vertegenwoordiger.

De (terugkerende) kosten die zijn gemoeid met het kwalitatief beheer betreffen de tijd die de leden van de beheersgroep moeten besteden aan de beheerstaken, aangevuld met een budget voor het eventueel inhuren van externe deskundigen om de kwaliteit van nieuw in te brengen cases te beoordelen. In totaal wordt dit geschat op een bedrag van € 20.000 per jaar.
De kosten die zijn gemoeid met de gebruikerscursussen dienen geheel door de deelnemers te worden gedragen. Voor de deelname aan gebruikersbijeenkomsten worden geen vergoedingen gegeven.

3.1 Voorstel voor uitvoering

Door SKB wordt een voorzitter van de kwalitatieve beheersgroep benoemd. Deze benadert BioSoil, TAUW en TNO-NITG om een permanente vertegenwoordiger in de kwalitatieve beheersgroep aan te wijzen. De beheersgroep dient minimaal tweemaal per jaar bijeen te komen. In de eerste bijeenkomst dienen afspraken te worden gemaakt over de benadering van de bodemsaneringsmarkt om zowel het gebruik van CBR-Bodem, als ook om het inbrengen van nieuwe cases te stimuleren. Daarnaast zal de frequentie van de gebruikersbijeenkomsten te worden vastgelegd.
De beheersgroep organiseert op korte termijn in overleg met de stichting PAO een aantal gebruikerscursussen.

4 Samenvatting van activiteiten en kosten

Samengevat worden de volgende activiteiten voorzien:
Ontwikkeling van een web applicatie op basis van CBR-Works;
    Eenmalige kosten: € 12.000
Aanschaf van een server voor de web applicatie en installatie bij TNO-NITG;
    Eenmalige kosten: € 6.000 aanschaf en € 2.000 installatie.
Inrichting van een helpdesk bij TNO-NITG;
    Jaarlijkse kosten € 5.000.
Regelmatig invoeren nieuwe cases in CBR-Bodem;
    Jaarlijkse kosten € 6.000.
Regelmatige aanpassingen/uitbreidingen applicatie/database;
    Jaarlijkse kosten € 6.000.
ICT ondersteuning bij het beheer;
    Jaarlijkse kosten € 3.000.
Instelling van een kwalitatieve beheersgroep door SKB;
    Jaarlijkse kosten € 20.000.
Organisatie van gebruikerscursussen in samenwerking met de stichting PAO;
    Kosten te dragen door deelnemers.

Totale kosten:

Eénmalig voor ICT beheer:	€ 20.000
Jaarlijks voor ICT beheer:	€ 20.000
Jaarlijks voor kwalitatief beheer:	€ 20.000

5 Financiering

Belangrijkste uitganspunt bij de finaciering van het beheer van de CBR-Bodem is de overweging dat het gebruik zo laag-drempelig mogelijk moet zijn. Als gevolg daarvan kan van de potentiële gebruikers geen bijdrage in welke vorm dan ook (contributie of per consult) worden gevraagd. De financiëring zal derhalve extern moeten geschieden. Het ligt voor de hand om, gezien de opzet van SKB II en de mogelijke gebruikers van CBR-Bodem, zowel bij SKB als bij het ministerie van VROM een verzoek tot financiëring neer te leggen. Het is logisch, dat door beide organisaties dan wel een langjarige verplichting wordt aangegaan. Vooralsnog zou kunnen worden uitgegaan van een verplichting voor 2 jaren. Na deze periode zal een evaluatie moeten plaatsvinden om te komen tot een beslissing omtrent continuering van het beheer van CBR-Bodem. Bij deze evaluatie moeten uiteraard de volgende punten worden betrokken:

Evaluatie van het gebruik van CBR-Bodem: wordt de database in voldoende mate geraadpleegd om continuering van het beheer te rechtvaardigen?

Evaluatie van de uitbreidingen van de database: worden voldoende nieuwe cases aangeleverd om te zorgen dat de database “up to date” blijft? Is dat niet het geval, op welke wijze kan dat alsnog worden gerealiseerd?

Evaluatie van de software: voldoet de gebruikte software aan de (veranderende) eisen die worden gesteld, en in het geval dat dat niet zo is, welke keuzes moeten worden gemaakt om aan die eisen te kunnen voldoen?