SE0000213

SKI Report 00:4

Expert Judgement in Performance Assessment

Roger D. Wilmot
Daniel A. Galson

January 2000

ISSN 1104-1374
ISRN SKI-R--00/4--SE

SKi

31/21

SKI Report 00:4

Expert Judgement in Performance Assessment

Roger D. Wilmot
Daniel A. Galson

Galson Sciences Ltd
5 Grosvenor House, Melton Road, Oakham, Rutland LE15 6AX, UK

January 2000

SKI Project Number 99132

Sammanfattning

Förslag om lokalisering, byggnad och drift av ett slutförvar för radioaktivt avfall i Sverige kommer att stödjas av säkerhetsanalyser. Sådana säkerhetsanalyser innebär att ett antal expertbedömningar måste göras. SKI förbereder sig på att granska SKB:s säkerhetsanalyser och för att genomföra egna oberoende säkerhetsanalyser. I samband med detta har SKI konstaterat att det behövs mera forskning kring hanteringen av expertbedömningar i säkerhetsanalyser.

En god förståelse för expertbedömningar är viktig av tre orsaker:

- SKI måste kunna ge riktlinjer om vad SKI förväntar sig av SKB:s säkerhetsanalyser och av arbetet med att utveckla sådana funktionsanalyser.

- SKI måste utveckla ett systematiskt angreppssätt för utvärdering av expertbedömningar i sina granskningar av SKB:s säkerhetsanalyser.

- Allmänhetens ökade engagemang i lokaliserings- och beslutsprocesserna ställer höga krav på såväl SKI:s som SKB:s förmåga att kommunicera frågor kring säkerhetsanalyser, inklusive hantering av expertbedömningar.

Denna rapport är en pilotstudie som systematiskt beskriver de olika typer av expertbedömningar som görs i samband med utvecklingen av en säkerhetsanalys och sammanfattar existerande verktyg och procedurer för hantering av expertbedömningar. Rapporten innehåller även rekommendationer för vidare arbete inom området expertbedömningar.

Expertbedömningar kan klassificeras på ett antal olika sätt. De kan klassificeras efter *varför* och efter *hur* bedömningarna görs. Vad beträffar varför bedömningar görs kan en generell distinktion göras mellan:

- Bedömningar av data, som görs därför att det inte finns några rimliga alternativ.

- Bedömningar om genomförandet av en säkerhetsanalys, som görs därför att det inte finns några alternativa sätt att fatta beslutet.

Vad beträffar hur bedömningar görs skiljer rapporten mellan icke-begärda bedömningar som görs av enskilda personer, icke-begärda bedömningar som görs av grupper av personer och begärda bedömningar från enskilda eller grupper. Dessa typer av bedömningar skiljer sig i allmänhet åt med hänsyn till omfattningen på den tillhörande dokumentationen och därmed också möjligheten att kunna spåra hur bedömningarna har gjorts.

Verktyg för värdering av expertbedömningar varierar efter typen av bedömning som undersöks. Nyckelverktyg i sammanhanget är granskning av typen ”peer review” (ungefär jämlikars granskning), system för kvalitetssäkring, dokumentation och utformningen av begärda expertbedömningar (t.ex. formellt expertförhör). Dialog med intressenter är också viktigt för att kunna fastställa om bedömningarna är motiverade i det sammanhang där de används.

Framtagandet av en säkerhetsanalys innefattar ett antal steg, från att etablera bedömningssammanhanget, via lokalisering och slutförvarkonstruktion till utveckling av scenarier och modeller samt val av indata. Rapporten diskuterar hur bedömningar används i vart och ett av dessa steg och identifierar vilka verktyg och procedurer för värdering av bedömningar som är de lämpligaste i varje enskilt steg.

Rekommendationerna för fortsatt arbete inkluderar genomförande av en praktisk övning i begärd expertbedömning (expertförhör) för att skaffa erfarenhet om fördelarna och nackdelarna med denna teknik, utveckling av riktlinjer för "peer review" och kvalitetssäkring (QA). Vidare görs en utvärdering av aktuella angreppssätt för dokumentation, som gör det möjligt att spåra de beslut och bedömningar som görs i de olika stegen av en säkerhetsanalys.

Denna rapport kompletteras av två bilagor som beskriver "peer review"-processen och procedurer för expertförhör.

Summary

Proposals to site, construct and operate a radioactive waste disposal facility in Sweden will be supported by performance assessments (PAs). Such PAs will require a range of expert judgements to be made. As part of SKI's preparation for reviewing SKB's PAs and for conducting independent PAs, SKI has identified a need for further research on the treatment of expert judgement in PA.

An understanding of expert judgement is important for three reasons:

- SKI should be able to provide guidance on what it expects from SKB's PAs and from the process of developing these PAs;
- SKI needs to develop a systematic approach for the evaluation of expert judgements in its reviews of SKB's PAs;
- The enhanced public involvement in the siting and decision-making processes puts high demands on both SKI's and SKB's ability to communicate issues related to PA, including the treatment of expert judgements.

This report is a pilot study that systematically describes the various types of expert judgement that are made throughout the development of a PA, and summarizes existing tools and practices for dealing with expert judgements. The report also includes recommendations for further work in the area of expert judgement.

Expert judgements can be classified in a number of ways, including classification according to *why* the judgements are made and according to *how* the judgements are made. In terms of why judgements are made, there is a broad distinction between:

- Judgements concerning data that are made because alternatives are not feasible; and
- Judgements about the conduct of a PA that are made because there are no alternative approaches for making the decision.

In the case of how judgements are made, the report distinguishes between non-elicited judgements made by individuals, non-elicited judgements made by groups, and elicited judgements made by individuals or groups. These types of judgement can generally be distinguished by the extent of the associated documentation, and hence their traceability.

Tools for assessing judgements vary depending on the type of judgements being examined. Key tools are peer review, an appropriate QA regime, documentation, and elicitation. Dialogue with stakeholders is also identified as important in establishing whether judgements are justified in the context in which they are used.

The PA process comprises a number of stages, from establishing the assessment context, through site selection and repository design, to scenario and model development and parameterisation. The report discusses how judgements are used in

each of these stages, and identifies which of the tools and procedures for assessing judgements are most appropriate at each stage.

Recommendations for further work include the conduct of a trial expert elicitation to gain experience in the advantages and disadvantages of this technique, the develoment of guidance for peer review and QA, and an assessment of state-of-the art approaches to documentation for tracing decisions and judgements throughout a PA.

The report is supplemented by two Appendices that outline peer review and expert elicitation procedures.

Contents

1 Introduction

The Swedish nuclear waste management programme is approaching a critical point with respect to the siting of a repository for spent nuclear fuel. The Swedish Nuclear Fuel and Waste Management Company (SKB) is expected to select two or more candidate sites for surface-based site investigations within the next two or three years. Based on the results of these investigations, SKB plans to apply for a licence to construct a repository at one of the studied sites. The earliest anticipated date for a licence application is 2008.

In preparation for the reviews of SKB's forthcoming licence applications, the Swedish Nuclear Power Inspectorate (SKI) has undertaken a wide range of research, including the development of an independent performance assessment (PA) capability. Two independent PAs have been performed (Project-90 and SITE-94). Subsequently, SKI has identified a need for further research on the treatment of expert judgement in PA. An understanding of expert judgement is important for three reasons:

- SKI should be able to provide guidance on what it expects from SKB's PAs and from the process of developing these PAs;
- SKI needs to develop a systematic approach for the evaluation of expert judgements in its reviews of SKB's PAs;
- The enhanced public involvement in the siting and decision-making processes puts high demands on both SKI's and SKB's ability to communicate issues related to PA, including the treatment of expert judgements.

As an initial step in developing this understanding of expert judgements, SKI has contracted Galson Sciences Ltd. to consider the classification and treatment of expert judgements in PA. The objectives of this pilot study are:

- To systematically describe the various types of expert judgements that are made throughout the development of a PA;
- To outline and describe existing tools and practices for dealing with expert judgement in PAs;
- To develop recommendations for further work in this area.

These objectives are addressed in the following four sections of this report:

- Section 2 provides classifications of expert judgements according to how and why they are made;
- Section 3 describes several tools for assessing expert judgements;
- Section 4 presents an outline of the PA process, describes how judgements are used at each stage and which of the assessment tools are most appropriate for each stage;

- Section 5 presents an outline research strategy.

Two Appendices present further details on the development of guidance for peer review and expert elicitaion.

2 Classification of Judgements

2.1 Introduction

A regulatory review of a PA submitted by a developer as part of a license application, and the regulatory decision-making subsequent to such a review, requires a thorough understanding of the way in which the assessment has been conducted. An inherent part of any assessment is the use of judgements, and it is useful to consider why and how judgements are made. It is also useful to subdivide judgements according to:

- The purpose of the judgement (why the judgement is made);
- The method of the judgement process (how the judgement is made).

The reasons why expert judgements are made fall into two principal categories:

- Alternative means of obtaining the information are too costly, too time-consuming, or otherwise not feasible;
- There are no alternative means to deciding on a course of action, choosing between alternatives, or making an assumption.

Further discussion of these two categories is presented in the Section 2.2.

Expert judgements may be made by individuals or by groups, and the judgements can be made under formal, controlled, recorded and documented circumstances, or informally, with less or no formal documentation. There are a number of possible combinations of these categories, and of intermediates between the end-members. In practice, however, there are three methods for making expert judgements that express the range of possibilities:

- Individual judgements;
- Group judgements;
- Elicited judgements.

These categories are discussed in Section 2.3.

An alternative classification that has been used is *implicit* and *explicit* judgements. This is useful in highlighting the extent to which judgements can be made without fully recognising them as judgements. Bonano *et al.* (1990) assumed that explicit judgements corresponded to elicited judgements and that all others were implicit. One aim of the current project is to discuss ways in which all expert judgements can be assessed and, if the ideal end-point is that all judgements are acknowledged, this classification would become redundant.

2.2 Why Judgements are Made

In the introduction above, two reasons are given for making expert judgements. In practice, there is a continuum between these end-members. For example, alternatives to expert judgement for defining the extent of variability in hydrogeological properties could be regarded as either too costly or as non-existent depending on the number and scale of measurements necessary. Similarly, the reason for making a judgement about the use of a network model or a continuum model for groundwater flow calculations lies between there being no means of comparing dissimilar conceptualisations, and it being too costly to acquire sufficient data to determine which model would be most appropriate under different boundary conditions.

Essentially, however, judgements concerning data are made because the alternatives are not feasible, and judgements about the conduct of a PA (including choice of models, and scope of the calculations) are made because there is no alternative approach.

This distinction between judgements about data and other judgements is useful in considering how the use of judgements can be assessed. This is because, at least in the context of data used by PA codes, there is a limit on the number of data values that may involve judgements. Data for such codes are commonly held in databases and the source of the data can be included or referenced in the PA documentation. Traceability to the source can thus be established and the role of any judgements can be assessed. The same constraint does not apply to other areas where judgements may be used. The key assumptions involved in developing an assessment code, for example, may be apparent to a reviewer, but there are likely to be many more that rely on the developer for identification and description. If such assumptions and judgements are not explicitly identified at the time they are made, traceability will be difficult to re-establish at a later time.

The distinction between the reasons why expert judgements are made is important when the overall programme of work is being reviewed or assessed. In this context, it is necessary to consider not only the adequacy of the judgements themselves, but also the adequacy of the reasons they were made. For example, a developer may use expert judgement to determine the extent of sorption of radionuclides in part of the disposal system because experiments would be costly or time-consuming. Not only should the regulator assess the judgement on the value for sorption (e.g., does it account for the full range of available data, has adequate account been taken of scaling effects, etc.), but the regulator should also consider the cost and time of the experiments to establish if the proponent's reliance on judgement in place of data is reasonable. The criteria used to determine whether experiments or other alternatives to judgements are not feasible will vary with the sensitivity of the PA calculations to the data in question. The key question is whether or not the conduct of the experiments would be cost-effective in terms of a reduction in uncertainty.

One topic that has received particular attention in terms of expert judgement, including several large expert elicitation exercises, is human intrusion. The aim of these exercises is to establish data for use in PAs, such as the effectiveness of institutional controls or the frequency of intrusive events such as drilling (Hora *et al.*, 1991). These data cannot be determined other than through judgement, because there are no experiments or measurements that could be conducted to obtain them. There may be other data that cannot be derived directly by measurement or experiment, particularly derived or non-physical parameters, but these should generally be traceable to such data. Although the reason for using expert judgement for non-physical parameters differs from that for other data, it is still useful to classify them with data because of the way in which they should be referenced in parameter databases.

In summary, expert judgements are used where there are no alternative approaches to making decisions, or where alternative approaches are too costly or otherwise regarded as not feasible. Expert judgements used in the determination of parameter values for PA codes should be readily identifiable through the documentation included as part of the PA or referenced by parameter databases. The identification of other judgements is strongly reliant on record-keeping during the development stages of a PA project.

2.3 How Judgements are Made

The terms expert judgement and expert elicitation have often been treated as synonymous, thereby implying that judgements are only made through elicitation. This can be confusing because there are many judgements made by personnel who would satisfy the criteria of expert but are not elicited. There is also potential for confusion in the minds of stakeholders if imprecise terms are used or specific meanings are attached to everyday phrases. In other contexts, judgements made without being elicited are commonly termed scientific, technical or engineering judgements. In discussing how judgements can be justified, documented, traced and assessed, it is useful to regard all judgements made by qualified personnel as expert judgements, and to distinguish between individual and group judgements for non-elicited judgements.

The classification used here is therefore:

- Individual judgements (non-elicited);
- Group judgements (non-elicited);
- Expert elicitation of individuals or of groups.

Each of these is discussed below in terms of who makes the judgements, their qualification for making judgements, and documentation of the judgements.

2.3.1 Individual judgements

- Individual judgements will be influenced by the experience and activities of the individual making them, including conversations and participation in meetings. There is, however, no collective responsibility for this type of judgement.
- Individual judgements are made by personnel working within the project, generally as part of their daily work activities.
- In a well-managed and properly resourced assessment, staff will be trained and have the knowledge and experience appropriate to the tasks they are conducting.

Experience may, however, lead to over-familiarisation with the subject area so that judgements are not explicitly identified. Inexperienced staff may also fail to recognise when they are making key judgements.

- Documentation of individual judgements can be highly varied. Judgements made in response to specific requests are likely to be documented in memoranda or reports, which should then form part of the project records. Judgements made as part of daily work activities, if recognised, should be recorded in notebooks or similar records. Such notebooks should form part of the overall project record, but the traceability of particular judgements recorded in this manner may be poor. Many individual judgements may also be made implicitly and not recorded.

2.3.2 Group judgements

- Group judgements cover a wide range from those made by two individuals in conversation to highly structured elicitation exercises involving independent specialists. The latter are described separately below, and the former share most of the attributes, at least in terms of documentation, of individual judgements.

- Groups may comprise a permanent or semi-permanent part of the overall project team (e.g., a design team for a particular code), or may be convened on an *ad hoc* basis for a particular purpose (e.g., working groups established to solve particular problems).

- One specific example of group judgements are those made by the clearing houses in SITE-94 (Chapman *et al.*, 1995). These clearing houses, comprising groups of experts with specific instructions and tasks to perform, were used to make judgements in circumstances where information processing could not be performed using computer models.

- Documentation is, to some extent, a requirement for a judgement being classified as a group judgement. If a meeting or group discussion simply serves to inform one or more of the participating individuals, without documentation of a group decision and acknowledgement of some form of collective responsibility, then subsequent judgements by those individuals would not be group judgements in the sense used here. Minutes of meetings and the decisions and judgements made should form part of the project records, but traceability may be difficult to establish for judgements made by *ad hoc* groups.

2.3.3 Elicited judgements

- Expert elicitation is a particular type of individual or group judgement in which a facilitator or normative expert questions specialists and elicits decisions, assumptions and data values. Elicitation can be conducted in meetings, either with individual specialists or with a group, or by correspondence.

- The number of specialists involved in group elicitation exercises varies, but three specialists is often found to provide a balance between available resources and breadth of knowledge (Bonano *et al.*, 1990).

  - Specialists may be external or internal to the organisation conducting the assessment. In the majority of reported formal elicitation exercises, the specialists have been external, because independence was considered an important attribute. Where scientific credibility is particularly important, selection of independent experts may involve nomination by learned societies or similar organisations.

  - In some circumstances, there may be insufficient external expertise to establish a fully independent expert panel. Internal specialists may then be used to complement external experts.

  - Where a structured approach to making judgements is considered important, the techniques of elicitation may be applied to judgements by internal specialists.

  - Members of the assessment team may participate in elicitations with external specialists to provide training and information to clarify the nature or scope of the decision or parameter being elicited.

  - The facilitator, whose role is to conduct the elicitation and not be involved at a technical level, may be independent or part of the organisation conducting the assessment.

- Specialists are specifically selected based on their experience and knowledge of a particular subject area. Independent specialists will generally require some training with respect to why the judgements they are making are required. Training in the elicitation process may also be needed, especially if parameter values are being elicited.

- Documentation is a key part of the elicitation process, particularly if different experts express different views that require reconciliation or combining. The way in which the elicitation is conducted (group or individual meetings or correspondence) will affect how differences are resolved, but in all cases the specialists should review and approve the final outcome. Documentation responsibilities may be divided between the facilitator and members of the assessment team participating in the elicitation.

3 Tools for Assessing Judgements

3.1 Introduction

Previous sections have outlined one possible classification of expert judgements that is intended to assist SKI both in planning any future assessments and in understanding assessments submitted by developers. The wide range of subject areas in which judgements are made in assessments means, however, that this simple classification of why and how judgements are made cannot be used alone to determine the best means of assessing judgements. In this section, we summarise five different approaches or tools for assessing judgements.

- Peer review;
- Quality assurance;
- Dialogue;
- Documentation;
- Elicitation.

Not all of these tools will necessarily be applied by the regulator in a review. Instead, the regulator will expect evidence that they have been used by the developer to ensure that judgements have been made and used appropriately. By examining the developer's use of these tools as well as using some directly, the regulator will obtain an overall picture of the adequacy and appropriateness of judgements.

In the Section 4, we provide a further means of classifying expert judgements, by the types of judgements made in different stages of the assessment process, and we identify which of the tools described here is most appropriate for a regulatory understanding of each type of judgement.

3.2 Peer Review

Peer review is one of the most useful methods for assessing judgements in the PA process because, if properly conducted, it allows for changes to be made in response to comments before PA calculations are completed or a licence application submitted. Proper documentation of the peer review process also provides assurance in subsequent assessments of the PA.

The most complete description of the peer review process is that presented in a US Nuclear Regulatory Commission report entitled "Peer Review for High-Level Nuclear Waste Repositories" (NUREG-1297; NRC, 1988). Peer reviews conducted by the US Department of Energy as part of its demonstration of compliance with 40 CFR Part 191 (EPA, 1993) were required to be compatible with the NRC methodology (40 CFR § 194.27: EPA, 1996).

NUREG-1297 defines peer review as "... a documented, critical review performed by peers who are independent of the work being reviewed" and states that a "... peer review is an in-depth critique of the assumptions, calculations, extrapolations, alternate interpretations, methodology, and acceptance criteria employed, and of conclusions drawn from the original work." The use of peer review is "... as a management tool to achieve confidence in the validity of certain technical and programmatic judgments. The intent of a peer review is to pass judgment on the technical adequacy of the work or data submitted for review, to identify aspects of the work on which technical consensus exists, to identify aspects on which technical consensus does not exist, and to identify aspects of the reviewed work which the reviewers believe to be incorrect or which need amplification. A peer review provides assurance in cases where scientific uncertainties and ambiguities exist but in which technical and programmatic judgments and decisions still must be made."

NUREG-1297 is a generic model for peer review that provides useful guidance for the development of a project-specific peer review procedure. Key areas of such a procedure include:

- What to evaluate. The procedure should specify broad areas that are appropriate for peer review. The procedure should be flexible enough to account for future changes in the structure or conduct of the assessment, and not restrict peer review to specific areas of work. The procedure should also specify what aspects of the reports or programme should be reviewed.

- The peer review group. The procedure should not identify specific individuals or organisations to undertake peer review, but should describe how reviewers can be identified, the level of expertise considered appropriate, and the degree of independence expected of a peer review team.

- Documentation. The procedure should outline a structure for the documentation of the peer review, including the initial findings of the review team, responses from the organisation undertaking the assessment, and reviewers' assessment of those responses. The procedure should ensure that unresolved concerns will be fully documented, and also that minority opinions from reviewers are retained.

Further details on these topics, in the form of extracts from NUREG-1297 are provided in Appendix A.

3.3 QA Programme

Quality assurance (QA) shares some attributes with peer review as a tool for assessing judgements. Significant differences between the two approaches arise, however, because QA is an integral, day-to-day process within any assessment programme, and is primarily conducted by members of the assessment organisation. In contrast, peer review is best conducted in a staged manner and at particular milestones within the assessment project. Peer review is also conducted by independent experts who do not otherwise have an ongoing involvement in the programme.

QA applies to all aspects of an assessment programme, including contractual, financial and management aspects. In terms of QA as a tool for assessing judgements, however, there are three key areas where the application of QA is of most interest to the regulator:

- Reporting;
- Code development and calculations;
- Observations and measurements (including site characterisation and experimental programmes).

Reporting

Judgements made through expert elicitation are made explicitly, and the documentation recording these describes the judgements, but is not the principal acknowledgement that a judgement has been made. In other words, there must be a prior decision process to undertake elicitation and expend significant resources on particular judgements. Other judgements are commonly made implicitly, and reporting is the first opportunity for an examination of both why a judgement has been made and the value selected or assumption adopted.

Because reports are the principal source of information on many judgements made by individuals, technical review of all reports is an important element of QA. Equally important are management and quality regimes within the organisation that ensure adequate resources (time and budget) are available for review, and that signing and approving reports is only done with evidence that review has been completed. Review comments, resolution and records of outstanding issues are important parts of the QA record. Without these records, the regulator will have no evidence that the judgements in the reported work have been reviewed and considered by more than one individual.

Peer review is a tool that is available to both the developer and to the regulator. Technical review within the overall QA regime of the assessment programme is only available within the assessment programme, be it the developer's or the regulator's. As with peer reviews, the developer may consider that, if all outstanding issues and comments have been addressed in the final reports submitted to the regulator, these reports will be sufficient. However, the regulator must have access to all the QA records if it is to have confidence that the developer has properly assessed the use of judgements.

Code development and calculations

Many of the same issues apply to the QA of codes and calculations as discussed above for reports. Because of the team approach often used in code design, a number of the judgements involved in code development are likely to be group judgements, which provides some assurance that they have been made explicitly and considered at some level of detail. QA procedures should ensure the thorough documentation of the code development cycle, including specification, design, testing, verification and evaluation. A traceable history of development through different versions of codes is also important if the regulator is to understand how judgements have evolved into those that underlie the final assessment calculations.

QA of calculations requires a rigorous approach to configuration management so that all of the information contributing to results can be traced to source. This information includes the set of parameter values used for particular code runs and the versions of the codes used, including run-time control information. Archiving of codes is crucial if results are to be repeatable. Sensitivity studies conducted at a later stage by the developer or regulator will require access to the same codes and versions if the results are to be useful.

Observations and measurements

QA of observations and measurements will help to ensure that any mistakes or errors are detected quickly enough for repeat measurements to be made. This is critical because observations and measurements often cannot be repeated at a later stage if errors are not detected until review. Although in no sense recommended, mistakes and errors in reporting and code development can, if necessary, be corrected following regulatory review, along with the repetition of any affected calculations.

Observations and measurements should be made in accordance with a test plan or similar procedure, and many of the judgements involved will be made in establishing such a plan. QA procedures should ensure adequate review of these judgements prior to the conduct of the experiments. For certain key experiments and observations, the test plan would be a milestone subject to peer review. Additional QA procedures may be required to ensure that data from earlier stages of site characterisation, or from sources not under the control of the developer, are qualified for use in the PA[1].

As with other aspects of QA, it is important the regulator has access to all of the records so as to build confidence. Because of the non-repeatable nature of some of this work, however, it is also necessary for this confidence to be built at an early stage. This is one topic where early dialogue between regulator and developer would be of particular value.

3.4 Elicitation Exercise

Section 2.3.3 summarises how judgements are made using expert elicitation. A key feature of this type of judgement is that the need for the judgement and a precise definition of the judgement required are necessary prior to assembling a group of experts. In this sense, these judgements differ from those made by individuals or groups that are not recognised in advance and are imprecisely defined. Precise definitions and the use of independent specialists mean that judgements made using elicitation are

---

[1] 40 CFR Part 194 (EPA, 1996) lists several methods for the qualification of existing data (§194.22(b)): peer review; corroborating data; confirmatory testing; or a demonstration of an equivalent QA programme.

relatively easy to assess, although this assessment is still dependent upon good documentation of the elicitation exercise.

The majority of expert elicitations that have been conducted as part of assessment programmes have been concerned with the elicitation of quantitative information such as parameter distributions for use in assessment calculations. However, there is also some experience with the elicitation of qualitative information. For example, elicitations conducted during the development of the PA for the WIPP considered (i) aspects of future human behaviour needed to be quantified using elicitation, and (ii) the types of information to be included on site markers (Trauth *et al.*, 1993).

There is no particular reason why elicitation should be used principally for quantitative information or modelling decisions. A recent report for the Health and Safety Executive identifies a number of case studies in which external expertise has been used to make qualitative judgements on a variety of issues (Taig *et al.*, 1997). Judgements regarding key assumptions and decisions, such as those related to the assessment context (Section 4.1), could benefit from close questioning by a facilitator and documentation of the process by which the judgements are reached.

The characteristics that apply to elicitation of judgements by the developer in preparing a PA also apply to the use of elicitation as a tool by the regulator during review. Elicitation is resource intensive, and there may be some problems at the time of regulatory review in establishing a group of experts with sufficient specialist knowledge who are not subject to some unintentional bias through knowledge of the assumptions and decisions already made in the PA under review. Nevertheless, for key areas of a PA where the regulator has concerns that judgements are inadequately justified, a regulatory elicitation exercise could be of value in establishing that the basis of a PA is reasonable.

Further details of the procedures used to undertake expert elicitation, based on those established for the WIPP project (Trauth *et al.*, 1993) are presented in Appendix B.

3.5 Dialogue

A traditional view of PA has been that it is a scientific / technical approach to demonstrating safety, and that regulatory decision-making will be based in large part on the technical merits of the PA. This view recognised that the full technical detail of a PA would not be accessible to all stakeholders and that less technical presentations would be required to explain the analyses and the decisions based on them. Judgements made in this sense were regarded essentially as substitutes for further analyses, so that, as more information was gathered, PA results would converge upon a "true" answer.

Two things have changed in this traditional view. First, it is now widely recognised that the results of PAs are illustrative rather than some approximation to a "true" performance. Second, it is also recognised that regulatory decision-making is not based solely on scientific and technical merits but that many socio-political issues are also involved. Stakeholders have a real role to play in decision-making and two-way dialogue with stakeholders, rather than one-way presentations, is a key part of the process.

Dialogue is an ongoing process and, as with QA, it cannot be "added later" if it is to be effective. Dialogue is therefore not an effective tool for assessing the judgements made in conducting a PA once the assessment has been submitted as part of a licence application. Dialogue does have a role, however, in making and assessing judgements prior to and during the PA, so that a clear understanding of what the PA is and how it helps decision-making can be developed, and so that stakeholders do not consider themselves faced with a *fait accompli*.

Dialogue with stakeholders is more commonly associated with the environmental impact assessment (EIA) process than the PA process. Examples include public consultation by Posiva (Posiva, 1998), and the requirement in the Swedish Environmental Code for SKB to undertake wide consultation as part of its EIA. An example of how dialogue may affect the conduct of a PA is provided by the review undertaken by an Environmental Assessment Panel of the EIA of a disposal concept submitted by Atomic Energy of Canada Limited (Environmental Assessment Panel, 1998). This panel called for a public review of the acceptability of a risk criteria, or whether a dose-based regulation would make decisions more explicit and visible. Without dialogue and the establishment of a consensus on such a key judgement, the assessment process cannot sensibly continue if it is to have credibility.

Further discussion on the use of dialogue within the risk assessment process is presented in Andersson *et al.* (1998).

3.6 Documentation

Documentation is a key element in allowing an assessment of judgements, because if judgements are not recorded they cannot be reviewed. If elicitation conducted as part of a review gives rise to similar parameter values, assumptions and decisions as used in an assessment, this will provide some assurance that any implicit judgements are reasonable. Similarly, peer review can, to an extent, establish whether implicit judgements that are not documented are reasonable. However, elicitation and peer review are resource intensive and generally only available as tools for review to the regulator. For other stakeholders, and for the majority of the regulator's review, it is important that the judgements made by the developer in an assessment are properly documented.

PA documentation is extensive, and the inclusion of a detailed record of every judgement in the PA within conventional paper-based reports would make it unmanageable. Alternative approaches are available:

- Paper-based records systems. All project records, including reports, memoranda, meeting minutes and notebooks are filed centrally, and assessment documents make reference to these files. This system requires on-going resources in terms of development and staffing of a records centre, and providing copies of material to stakeholders and the regulator as required. Paper-based records systems are easily updated as new information is produced, but considerable effort is required to include appropriate references in higher-level documents and to update these as the PA evolves.

- Electronic records systems. Both assessment documents and supporting material is supplied in electronic form to reviewers (e.g., on CD-ROM). Electronic links allow easy access to referenced material. Once distributed, no further support is required for reviewers. Since PAs will evolve as repositories are planned, built, used and decommissioned, however, it will be necessary to maintain an additional electronic records system that can respond to changes, and that can be used to compile future sets of documentation. The same considerable effort is required to include references as is the case with paper-based systems, but there will generally be fewer incorrect references, recursive relationships or other errors.

- Electronic databases. Electronic databases are widely used in PA programmes. Examples include the International FEP Database published by the NEA, a database of process influences developed by SKI (Chapman *et al.*, 1995), and a database of FEPs, models, assumptions and parameters designed to track developments within the WIPP PA (Crawford *et al.*, 1998). Assessment codes are also commonly supported by parameter databases. Configuration management requires traceable links to the input data used for any calculation, and the easiest means of providing this is through databases, from which parameter values can be selected according to the analysts' criteria. Databases may include fields for recording the source of the data and a commentary on the level of uncertainty, range of applicability, etc. that can be used by the analyst to select appropriate values or by reviewers to trace judgements.

The electronic databases outlined above are used to document parts of the PA process. Sandia National Laboratories, supported by Galson Sciences Ltd., have investigated and begun to develop a data model for the whole of the assessment process. The corresponding database records all activities, such as performing experiments, screening FEP lists, developing mathematical models, and performing calculations, as "transactions". Each transaction would require resources, such as staff, equipment, or samples, all of which would have attributes (e.g., staff qualifications, equipment calibration, sample location). Any transaction requiring judgement would require completion of an appropriate field before the transaction was validated and the output or result made available to other transactions. Such a system provides traceability, but also requires the same commitment to quality, in terms of properly recording the reasons for judgements, as any other documentation system.

NEXT PAGE(S)
left BLANK

4 Expert Judgements in Performance Assessment

In this section, the PA process has been divided into six key elements:

- Assessment context.
- Site selection / site characterisation;
- Repository design / optimisation;
- Scenario development;
- Model development;
- Parameterisation;

Each of these elements requires the application of judgements, but the nature of the judgements differs between elements and the most appropriate means for assessing these judgements differs as a consequence. The descriptions of these elements are not intended to constitute a complete description of the PA process, partly because the conduct of a PA is not in itself a defined process (see "Assessment Context" below). Sufficient information is provided to demonstrate the way in which expert judgements are required and the basis for determining how the regulator can best assess these judgements in their own programme and in the developer's programme.

A summary of the tools considered most appropriate for assessing judgements made within each element is presented in Table 1.

	Peer Review	QA	Dialogue	Documentation	Elicitation
Assessment Context			✔	✓	✓
Site Selection			✔	✓	
Site Characterisation		✔		✔	
Repository Design		✓		✔	
Optimisation			✔	✓	
Scenario Development					
FEP list	✔		✓	✓	✔
Screening Criteria			✔	✓	
Screening	✔			✔	
Scenario Identification			✔	✓	
Model Development					
Conceptualisation	✔	✓		✓	
Implementation		✔		✔	
Parameterisation		✓		✔	✔

Table 1. Summary of the tools considered most appropriate for assessing judgements made within each element of a performance assessment.

✔ - Key approach: ✓ - Supporting approach

4.1 Assessment Context

Section 3.5 outlines how attitudes to the meaning and use of PAs have changed, and notes that PAs are now recognised as indications of system performance to inform decision-making. Other factors influence decision-making and there are non-technical issues that affect how an assessment is performed, what is considered in the analyses, and how results are presented and used. The assemblage of factors that influence the conduct of an assessment is commonly termed the "assessment context" (IAEA, 1999). This corresponds in part to the "strategic repository decisions" of Bonano *et al.* (1990).

By definition, all of the decisions made in defining the assessment context are judgements because they relate to issues that cannot be quantified by any observations or analyses, although ranking methods can be used as a means of comparing values held by different stakeholders. Stakeholders should have a role in defining the assessment context because much of the public debate will focus on overall issues and approaches rather than on the details of the technical analysis.

An example of a judgement to be made in the assessment context is the treatment of future human activities. The evolution of societies and technology cannot be modelled or predicted, and so judgement must be used to determine what assumptions are made in the assessment calculations. Recent work for SKI addresses regulatory strategies regarding the treatment of human activities and concludes that dialogue with stakeholders must be a part of the strategy (Galson *et al.*, 1999). Similarly, the Environment Agency of England and Wales is considering how estimates of risk are best calculated and presented. Preliminary conclusions (Wilmot *et al.*, 1999) are that highly uncertain events should be the subject of supporting calculations, and that the scope of these can best be determined by dialogue between regulator, developer and stakeholders.

A clear description of the assessment context is important so that reviewers have a clear understanding of the extent and purpose of the PA. This is particularly important for preliminary PAs and for those addressing issues such as optimisation. Because the assumptions and decisions that contribute to the assessment context have a major influence on all of the subsequent elements of the PA process, their justification will be a focus of the regulatory review. Elicitation to determine the judgements involved (by the developer) or their reasonableness (by the regulator) could be of benefit.

4.2 Site Selection / Characterisation

Site characterisation is aimed at establishing the properties of the region around a proposed repository. The properties of interest are those that could affect the transport of radionuclides away from the repository and thereby affect radionuclide concentrations in the biosphere and any subsequent doses received by individuals or populations. Key topics within any site characterisation programme will be the overall geological structure, and the determination of hydrogeological and geochemical conditions.

The fundamental sources of information for site characterisation are observations made at the surface, in underground excavations, or from within boreholes, and tests and analyses on material extracted at the surface, in underground excavations, or from within boreholes. Direct observations are supplemented by geophysical measurements made at the surface, in underground excavations, and within and between boreholes. Only a finite number of boreholes can be drilled, both because of the resources available and because too many boreholes could adversely affect the properties of the geosphere and provide short-circuits for radionuclide transport to the biosphere. Although some continuous or pseudo-continuous observations can be made (e.g., core logging, wireline logs and seismic profiles), the majority of observations are made on samples from specific points or over specific intervals. Similarly, many hydrogeological tests measure properties and responses over specific intervals. Measurements in and around boreholes need to be extrapolated for the regions between boreholes, and upscaling of measured properties may be required to extend measurements on a local scale to the scales of interest to PAs. A useful survey of the types of information required for the hydrogeological aspects of site characterisation, and the documentation of this information, is presented by Geier (1998).

Site characterisation requires judgements concerning the number and location of boreholes, the location and amount of underground excavations, the analyses to be performed and the techniques to be used, the number and location of samples and the techniques used to obtain them, and the interpretation, synthesis and presentation of the results. Judgements that lead to large costs and/or long timescales will be made at a strategic or planning scale, but many of the judgements involved in site characterisation will be pragmatic judgements made by individuals as sampling and testing is conducted. For example, some samples may be taken on a regular spacing or timing, but others require judgements about the most appropriate location (e.g., at boundaries, avoiding damaged material, or during certain events).

The outline above is broadly applicable to a site characterisation programme for any type of facility, although the scale of the programme is generally much greater for a radioactive waste repository site. There are key similarities between investigations of contaminated sites and site characterisation for a repository, but in the former case the area under investigation is already determined. In the case of prospective sites for radioactive waste repositories, however, key judgements are also involved in the site selection procedure. Final site selection will depend upon many factors, including socio-economic factors and community acceptance, as well as geological criteria. A decision-making process needs to be established in order to reduce the number of prospective sites identified by desk studies to a number that can be investigated more thoroughly. The criteria used in this process are not absolute criteria but must be a matter for judgement.

Site selection

Considerable resources will be expended in the site characterisation programme for a prospective radioactive waste facility. Regulatory decision-making can only be based on an application incorporating information from such a programme, but dialogue between regulator, developer and stakeholders prior to significant expenditure at a particular site will ensure that the judgements made in site selection are appropriate and broadly acceptable. Clear and comprehensive documentation of the site selection process will help at all stages of decision-making.

Site characterisation

Section 3.3 discusses quality assurance and its importance in several aspects of an assessment programme. The point is made there that many observations cannot be easily repeated, so that QA of site characterisation programmes is important in establishing confidence in the observations and measurements that underlie descriptions, syntheses and conceptualisations of the site. Descriptions of site characteristics will be a synthesis of large amounts of raw and interpreted data, and incorporate large numbers of assumptions and judgements about variability, uncertainty and site evolution. The regulator generally has no remit or resources for independent site characterisation, and so the assessment of this synthesis and the judgements involved will rely heavily on well-structured, referenced and supported documentation from the developer.

4.3 Repository Design / Optimisation

Repository design is an iterative process that is closely linked to site selection and characterisation, and to the results of preliminary PAs. Repository design may also evolve during the lifetime of the repository in response to lessons learned during the constructional and operational phases. Several conceptual designs have been developed for repositories, depending on the rock type and waste type involved. All such designs are based on the multi-barrier approach, which ensures that safety will not be unduly compromised by failure of one element of the design. The actual design selected will of necessity be a compromise between cost and the extent of conservatism incorporated into both the individual barriers (e.g., thickness of copper canisters) and the overall design (e.g., whether seals are emplaced between sections of the repository). Optimisation studies are used to determine that the compromise is appropriate, and that the impacts of the selected design satisfy a criterion such as As Low As Reasonably Practicable (ALARP).

Design

The design process, particularly in the early conceptual stage, will require many judgements and assumptions to be made in order to provide sufficient information for the conduct of a meaningful PA. Some of the judgements, for example the detailed design of repository elements such as seals and waste canisters, may be revised as more information becomes available. Significant changes would, however, require a re-examination of many other judgements and calculations. There is, therefore, an onus on the repository design team to make robust judgements from the outset. The assessment of these judgements will rely heavily on well-structured, referenced and supported documentation from the developer. Supporting calculations will be conducted as part of the design process, and QA of these will provide confidence in their use.

Optimisation

Even in assessments where optimisation is not identified as a separate element, there must be underlying judgements on the degree of conservatism incorporated into the design. Where optimisation is explicitly acknowledged, judgements are required on the approach to optimisation. For example, optimisation studies could focus on one key element of a selected repository design[2]. Alternatively, several radically different designs could be examined. The second approach is more usually included in environmental impact assessments (EIAs) than in PAs. Because costs are an important aspect of optimisation studies, the judgements involved include many related to external, socio-political factors, such as societal attitudes to risk and the economic model used in determining national energy policy. Dialogue between regulator, developer and stakeholders to determine acceptable approaches to optimisation will be important. Clear and comprehensive documentation of the optimisation studies will help at all stages of decision-making.

4.4 Scenario Development

Scenarios are variously defined in different PA programmes, but the essence of scenario development in all PAs is the decision on what is to be included in the quantitative system performance assessment. The initial phase of scenario development is the derivation of a list of features, events and processes (FEPs) that may be relevant to system performance. Using predefined selection criteria, this list is screened to provide a list of FEPs for inclusion in the quantitative system assessment. In the majority of PAs, a final stage involves the assignment of different groups of FEPs to different scenarios (e.g., a base case scenario, a glaciation scenario, a human intrusion scenario). The alternative approach of including all screened-in FEPs into a single set of analyses has also been explored.

Judgements are required at each stage of the scenario development process. FEP lists have been developed by a number of assessment programmes using various group approaches in an effort to ensure comprehensiveness, but these approaches have not always used formal expert elicitation, in the sense of using a facilitator, specialists (subject experts) and generalists (PA staff), as described in Section 2.3.3. An international FEP list has been published by the NEA, and this incorporates many of the individual FEP lists published by assessment programmes. To a certain extent, the existence of these published lists reduces the need for assessment programmes to develop new, independent FEP lists. For example, the FEP list developed by the US Department of Energy for its Compliance Certification Application (CCA) for the WIPP site was based in part on the compilation published by SKI (Stenhouse *et al.*, 1993). This compilation, however, still required the use of expert judgements to make it specific to the WIPP disposal concept.

---

[2] In the Compliance Certification Application for the WIPP, the Department of Energy examined the impacts of a range of different backfill compositions, but did not document the consequences of other design changes (DOE, 1996).

The selection or screening criteria used to determine which FEPs should be included in PA calculations may be specified in the regulations applicable to a specific site, or they may be derived on an assessment by assessment basis. For example, the FEPs included in calculations undertaken for optimisation may differ from those used to demonstrate performance of the selected design, or for the purpose of regulatory decision-making. Whatever the origin of the criteria, judgements are required in their application. Commonly used criteria include regulatory exclusion or assessment context, low probability, and low consequence. While side calculations, sometimes using PA codes, may support some of the screening decisions, many of these decisions will be based on qualitative screening arguments.

The final stage of scenario development, the definition of scenarios and the assignment of FEPs to them, may be governed by regulatory requirements (e.g., the US EPA's regulations that require analysis of system performance under both undisturbed and disturbed [human intrusion] conditions (EPA, 1993)). Judgements are required as to whether the regulator or the developer determines the scenarios, as there are no absolute criteria that can be adopted. The basis for scenario selection, apart from regulatory requirements, may include transparency and ease of understanding, comparisons with assessments of other sites or designs, and ease of computation.

FEP lists

The list of potentially important FEPs is an important starting point for a PA if the regulator and stakeholders are to be confident that the set of issues considered is comprehensive. Elicitation of a FEP list by the developer or by the regulator, or comparison with some other elicited list, are all useful ways of building this confidence. Because of its importance to later stages of the assessment, a FEP list that has not been elicited should probably be treated as a milestone for peer review. FEP list development should allow for the addition of FEPs as a result of dialogue with stakeholders. Clear documentation of the assessment FEP list, possibly through use of an electronic database, will help to establish traceability.

Screening criteria

Overly stringent criteria for determining which FEPs to include in scenario identification would impose a resource burden out of proportion to the increase in confidence provided. Conversely, relaxed criteria will reduce the number of FEPs included in assessment calculations but may exclude issues of concern. Dialogue involving the regulator, developer and stakeholders could be an appropriate means of establishing criteria acceptable to all parties.

Screening

Once screening criteria have been established, each FEP must be assessed against the criteria. The results of this screening, which may be based on quantitative, semi-quantitative or qualitative arguments, will rely on well-structured, referenced and supported documentation. The judgements involved in determining which FEPs remain after screening should probably be treated as a milestone for peer review.

Scenario identification

Unless an attempt is made to incorporate all screened-in FEPs into a single analysis, the set of scenarios that are analysed must be established in a way that aids an understanding of system behaviour and addresses concerns about interactions between different aspects of system performance. Dialogue involving the regulator, developer and stakeholders could be an appropriate means of establishing scenarios that are acceptable to all parties prior to the conduct of assessment calculations. A clear description of the scenarios adopted will help to establish traceability.

4.5 Model Development

Scenario development gives rise to broad-brush descriptions of the disposal system and its evolution. Before system performance calculations can be performed, conceptual models of the features, events and processes, and their interactions, must be developed. For example, groundwater flow and transport are likely to be included in most scenarios, but models of the way in which flow and transport take place must be established before any assessment of performance can be made. Conceptual models are the initial stage of model development. In the simple example above, the conceptual model would require decisions as to whether porous media or fracture flow, or a dual-permeability flow regime, was the most appropriate description of flow, and the extent to which processes such as advection, diffusion and retardation interact. A key issue in conceptual model development is the extent to which different parts of the system interact and whether such interactions require explicit inclusion in the model.

Once a conceptual model has been established that accounts for the available information and associated uncertainties (and it is possible for alternative conceptual models to be developed for the same scenario), mathematical and computational models are required that express the conceptual models in a form that can be used in assessment calculations. Except for very simple systems, computer codes are required to perform the actual computations. Simplifications, assumptions and decisions are required at every stage of model development, including judgements on the compromise between numerical accuracy and computational resources.

Conceptualisation

Establishing the appropriateness of the conceptual model(s) of a disposal system and its evolution are of key importance as they govern all the remaining stages of model development. The derivation of conceptual models should therefore be treated as a milestone for the conduct of peer review. QA and clear documentation will both support peer review by helping to establish traceability to underlying observations and data.

Implementation

Stakeholder concerns often focus on what is analysed within an assessment rather than on how. Nevertheless, the implementation of conceptual models into codes requires many judgements that may be difficult to extract and assess if they are not documented during development. Confidence in the testing, verification and evaluation of codes is best developed by the existence of appropriate levels of QA, and of a management culture that recognises the importance of timely QA rather than viewing it as a process that can be completed later to satisfy the regulator.

4.6 Parameterisation

All models used in PAs require the specification of parameter values. These values may be single-valued, representing a best estimate or "conservative but realistic" description of system performance, or they may be expressed as ranges of values that are sampled repeatedly in a probabilistic assessment to account for uncertainty. Parameters include both those corresponding to physical attributes of the disposal system (e.g., the radionuclide inventory in the repository, dimensions of waste containers, or porosity of units in the geosphere), and derived and non-physical parameters required for modelling purposes (e.g., grid size, time-step length, indices for selecting between conceptual models, tortuosity of units in the geosphere). The first of these types of parameters can, in theory, be derived directly from observations and measurements. In practice, however, judgements will be required to determine the spatial and/or temporal applicability of such measurements, and the way in which variability is treated. Results from experiments (and literature surveys) may also be used to derive this type of parameter, although again judgements will be required because experiments cannot reproduce all of the characteristics of the disposal system. Parameters of the second type must, by definition, be derived, rather than directly measured, and therefore require judgements to be made.

The derivation of parameter values is perhaps the most common area of PAs where judgements have been explicitly acknowledged and documented. As noted above, this is in part because the number of parameters is finite, and also because databases, or at least structured documentation, are commonly used for maintaining information about parameter values and their derivation. Because it is a common area for the judgements to be acknowledged, the parameterisation of PA models is also the area in which the majority of formal expert elicitation has taken place. Elicitation is a useful method for deriving – and assessing - both data values and assumptions or decisions developed using the other categories of expert judgement.

Depending on the complexity of the assessment codes used, there may be a significant number of parameters requiring documentation, including reference to supporting material. An assessment of the judgements involved will therefore rely on well-structured documentation, and a database or other form of electronic documentation would facilitate the review and assessment process. The development and maintenance of parameter databases will require a commitment to QA if confidence in their use is to be established.

NEXT PAGE(S) left BLANK

5 Outline Research Strategy

Based on the classification of expert judgements in PA, the tools available for assessing these judgements, and the most appropriate tools for assessing judgements at different stages of a PA, we have developed an outline research strategy for consideration by and discussion with SKI. This strategy does not yet take account of resource constraints or the timescales available. More detailed descriptions of the suggested topics would need to consider these aspects.

The topics included in the research strategy include both the provision of guidance to the developer on procedures for documenting and assessing judgements, and activities that could form part of SKI's independent assessment programme. Both types of activity would be of value in building SKI's capability to undertake a thorough regulatory review of PAs submitted as part of future licence applications. An understanding of the extent and impact of expert judgements in such PAs will be important for credible decision-making.

Peer review

We recommend that SKI develop guidance on the use of peer review in PA and the contexts in which it would expect SKB to use peer review. We believe that NUREG-1927 is a good basis for this guidance, but that it should be updated and consideration given to the Swedish context.

QA

We recommend that SKI develop guidance on the standards of QA that it expects to be applied to the different elements of PA. We also recommend that SKI develops QA plans and procedures for the conduct of its own independent PA and evaluations. These procedures will act as examples to SKB of SKI's requirements. Because of the key role of QA in site characterisation, we recommend that SKI review and approve SKB's test plans and QA procedures before the initiation of significant site-specific work at the selected repository site.

Dialogue

SKI has been at the forefront of researching the dialogue process, including playing a key role in the DIALOGUE (Andersson *et al.*, 1993) and RISCOM (Andersson *et al.*, 1998) projects. We are aware that SKI is also involved in putting forward proposals for further work in this area, and believe that this type of work will be of value in understanding how other stakeholders view the use of expert judgements in PAs.

Documentation

We recommend an assessment of the use of electronic documentation for the presentation of PAs, including the use of links to references and supporting material. Experience from other programmes suggests that this would be a valuable aid in the review process. Because of the discipline it imposes on the structure of the documentation, errors in references are reduced and judgements may be more readily traced to source material. It would also make the results of assessments more accessible to a wide range of stakeholders. Depending on the results of the review, and experience with their own published material, SKI could consider making this form of presentation a requirement for material submitted in licence applications.

We also recommend that SKI consider the merits of the development and use by both SKI and SKB of an integrated assessment database that records the collection of site characterisation and experimental data and traces its interpretation and incorporation into parameter values for assessment models. This type of database would be most appropriate if introduced early in an assessment programme so that all relevant material was included from the start. Nevertheless, we believe that viewing the whole of the assessment process from a database perspective would provide SKI with useful insights into the structure of a PA and how judgements could be made more explicit.

Elicitation

We recommend that SKI conducts a formal expert elicitation exercise. This would provide SKI with useful experience of the advantages and disadvantages of this approach to making judgements. Elicitation of data values has been conducted in a number of assessment programmes and there may be sufficient information in the documentation of these for SKI to develop a view on this particular aspect. More useful would be an elicitation exercise focusing on assumptions and decisions rather than on data values.

Experience from such an exercise would enable SKI to develop guidance on appropriate ways to use elicitation for decision-making in assessments. An analysis of this type of elicitation would also be of value to regulators in other industries and in other countries.

Seminar

Little attention has been given to expert judgement issues at international level, despite the key importance of such issues in development and review of safety cases for radioactive waste disposal. We therefore recommend that SKI consider the idea of prompting or organising an international seminar on the use of expert judgement in PA. The purposes of the seminar would be to compare approaches for dealing with expert judgement in PA, and to exchange information on the latest developments on topics such as peer review, QA, dialogue to resolve judgement issues, documentation (particularly in the form of electronic systems for data management), and elicitation of parameters and concepts for PA. The seminar should involve both regulators and developers.

6 References

Andersson, J., Andersson, K. and Wene, C.-O., 1993. The Swedish DIALOGUE project. An attempt to explore how different actors may take part in the decision process related to disposal of radioactive waste, in *Proc. Fourth Annual Int. Conf. on High Level Radioactive Waste Management.* (Las Vegas, 26-30 April, 1993), American Society of Civil Engineers, New York, NY and American Nuclear Society, La Grange Park, IL.

Andersson, K., Espejo, R. and Wene, C.-O., 1998. *Building Channels for Transparent Risk Assessment: Final Report RISCOM Pilot Project.* SKI Report 98:6. Swedish Nuclear Power Inspectorate, Stockholm.

Bonano, E.J., Hora, S.C., Keeney, R.L. and von Winterfeldt, D., 1990. *Elicitation and use of expert judgment in performance assessment for high-level radioactive waste repositories.* Sandia National Laboratories Report SAND89-1821. Sandia National Laboratories, Albuquerque, NM.

Chapman, N.A., Andersson, J., Robinson, P., Skagius, K., Wene, C.-O., Wiborgh, M. and Wingefors, S., 1995. *SITE-94: Systems Analysis, Scenario Construction and Consequence Analysis Definition for SITE-94.* SKI Report 95:26. Swedish Nuclear Power Inspectorate, Stockholm.

Crawford, M.B., Wilmot, R.D., Galson, D.A., Swift, P.N. and Fewell, M.E., 1998. PASS: A component of Desktop PA for the WIPP, in *Proc. Eighth Int. High Level Radioactive Waste Management Conf.* (Las Vegas, 10-14 May 1998), American Nuclear Society, La Grange Park, IL and American Society of Civil Engineers, New York, NY.

DOE (Department of Energy). 1996. *Title 40 CFR Part 191 Compliance Certification Application for the Waste Isolation Pilot Plant, October 1996.* DOE/CAO-96-2184.

DOE (Department of Energy). 1997. *Final Report - Expert Elicitation on WIPP Waste Particle Size Distributions During the 10,000 Year Regulatory Post-Closure Period.* Department of Energy, Carlsbad Area Office, June 1997.

Environmental Assessment Panel, 1998. *Nuclear Fuel Waste Management and Disposal Concept.* Report of the Nuclear Fuel Waste Management and Disposal Concept Environmental Assessment Panel. Canadian Environmental Assessment Agency, Quebec.

EPA (Environmental Protection Agency). 1993. *40 CFR Part 191 Environmental Radiation Protection Standards for the Management and Disposal of Spent Nuclear Fuel, High-Level and Transuranic Radioactive Wastes: Final Rule.* Federal Register, Vol. 58, No. 242, pp. 66398-66416, December 20, 1993, Washington, D.C.

EPA (Environmental Protection Agency). 1996. *40 CFR Part 194: Criteria for the Certification and Recertification of the Waste Isolation Pilot Plant's Compliance with the 40 CFR Part 191 Disposal Regulations: Final Rule.* Federal Register, Vol. 61, No. 28, pp. 5224-5245, February 9, 1996, Washington, D.C.

Galson, D.A., Wilmot, R.D., Wickham, S.M., Lilja, C., Norrby, S. and Wingefors, S., 1999. Future human actions in safety assessment for radioactive waste repositories: Elements of a regulatory strategy, in: *Int. Symp. Radioactive Waste Disposal: Health and Environmental Criteria and Standards* (Stockholm, 31 August – 4 September 1998), Swedish Radiation Protection Institute, Stockholm.

Geier, J., 1998. *Site Characterization Data Needs for Hydrogeological Evaluation.* SKI Report 98:1. Swedish Nuclear Power Inspectorate, Stockholm.

Hora, S.C., von Winterfeldt, D. and Trauth, K.M., 1991. *Expert Judgment on Inadvertent Human Intrusion into the Waste Isolation Pilot Plant.* Sandia National Laboratories Report SAND90-3063. Sandia National Laboratories, Albuquerque, NM.

IAEA, 1999. *BIOMASS, The IAEA Programme on Biosphere Modelling and Assessment Methods. BIOMASS THEME 1, Radioactive Waste Disposal. Alternative Assessment Contexts: Implications for Development of Reference Biospheres and Biosphere Modeling.* BIOMASS/T1/WD02, International Atomic Energy Agency, Vienna.

Nuclear Regulatory Commission (NRC), 1988. *Peer Review for High-Level Nuclear Waste Repositories Generic Technical Position.* NUREG-1297, Division of High-Level Nuclear Waste Management, Office of Nuclear Material Safety and Safeguards. U.S. Nuclear Regulatory Commission, Washington, D.C.

Posiva, 1998. *The final disposal facility of spent nuclear fuel – Environmental Impact Assessment Programme.* Helsinki, Posiva.

Rechard, R.P., Trauth, K.M., and Guzowski, R.V. 1992. *Quality Assurance Procedures for Parameter Selection and Use of Expert Judgment Panels Supporting Performance Assessments of the Waste Isolation Pilot* Plant, SAND91-0429, Sandia National Laboratories, Albuquerque, NM.

Stenhouse, M.J., Chapman, N.A., and Sumerling, T.J. 1993. *SITE-94 Scenario Development FEP Audit List Preparation: Methodology and Presentation.* SKI Technical Report 93:27. Swedish Nuclear Power Inspectorate, Stockholm.

Taig, A., Bell, R., Brearley, S., Brown, M. and Garlick, A., 1997. *Expert judgement: An Investigation of Approaches.* AEA Technology Report AEAT-1321. Health and Safety Executive, London.

Trauth, K.M., Hora, S.C. and Guzowski, R.V., 1993. *Expert Judgment on Markers to Deter Inadvertent Human Intrusion into the Waste Isolation Pilot Plant.* Sandia National Laboratories Report SAND92-1382. Sandia National Laboratories, Albuquerque, NM.

Trauth, K.M., Guzowski, R.V., and Hora, S.C. 1993. A Formal Expert Judgment Procedure for Performance Assessments of the Waste Isolation Pilot Plant. Sandia National Laboratories Report SAND93-2430. Sandia National Laboratories, Albuquerque, NM.

Wilmot, R.D., Pollard, S.J.T., Smith, R.E., Yearsley, R.A and Galson, D.A., 1999. Transparency in risk assessments - presenting the "Expectation Value" of post-closure risks from radioactive waste repositories, in *Proceedings of the VALues in Decisions On Risk (VALDOR) Symposium* (Stockholm, 13-17 June 1999), SKI, SSI and EC, Stockholm.

NEXT PAGE(S) left BLANK

Appendix A – Peer Review

This Appendix presents relevant sections from the US Nuclear Regulatory Commission report entitled "Peer Review for High-Level Nuclear Waste Repositories" (NUREG-1297). NUREG-1297 is not a description of peer review for a particular programme, but rather a generic model that can be adapted depending on the exact requirements. It provides useful guidance in terms of:

- What to evaluate

  "In meetings and/or correspondence, the peer review group should evaluate and report on: (a) validity of assumptions; (b) alternate interpretations; (c) uncertainty of results and consequences if wrong; (d) appropriateness and limitations of methodology and procedures; (e) adequacy of application; (f) accuracy of calculations; (g) validity of conclusions; (h) adequacy of requirements and criteria. Furthermore, full and frank discussions between the peer reviewers and the performers of the work are encouraged."

- The peer review group

  "The number of peers comprising a peer group should vary with the complexity of the work to be reviewed, its importance to establishing that safety or waste isolation performance goals are met, the number of technical disciplines involved, the degree to which uncertainties in the data or technical approach exist, and the extent to which differing viewpoints are strongly held within the applicable technical and scientific community concerning the issues under review. The collective technical expertise and qualifications of peer group members should span the technical issues and areas involved in the work to be reviewed, including any differing bodies of scientific thought. Technical areas more central to the work to be reviewed should receive proportionally more representation on the peer review group."

  "The peer review group should represent major schools of scientific thought. The potential for technical or organizational partiality should be minimized by selecting peers to provide a balanced review group."

  "The technical qualifications of the peer reviewers, in their review areas, should be at least equivalent to that needed for the original work under review and should be the primary consideration in the selection of peer reviewers. Each peer reviewer should have recognized and verifiable technical credentials in the technical area he or she has been selected to cover. The technical qualifications of each peer, and hence of the peer review group as a whole, should relate to the importance of the subject matter to be reviewed."

  "Members of the peer review group should be independent of the original work to be reviewed. Independence in this case means that the peer, a) was not involved as a participant, supervisor, technical reviewer or advisor in the work being reviewed, and b) to the extent practical, has sufficient freedom from funding considerations to assure the work is impartially reviewed."

- Documentation

  "A written report documenting the results of the peer review should be issued. It is usually prepared under the direction of the chairperson of the peer review group, and is signed by each member individually. It should clearly state the work or issue that was peer reviewed and the conclusions reached by the peer review process ... The report should include individual statements by peer review group members reflecting dissenting views or additional comments, as appropriate. The peer review report should contain a listing of the reviewers and any acceptability information (i.e., technical qualifications and independence) for each member of the peer group, including potential technical and/or organizational partiality." Furthermore, "... minutes should be prepared of meetings, deliberations, and activities of the peer review process."

- QA

  NUREG-1297 envisages that peer review will take place under the auspices of the reviewed organisation's QA programme, and that: "As a minimum, the QA organization should provide surveillance of the peer review process to ensure that the procedures conform to the guidance of this [Generic Technical Position] and that they are followed by the peer review group."

Appendix B – Expert Elicitation

This Appendix presents a description of the stages and procedures required in an expert elicitation exercise. It is based on the expert elicitation procedures established for the Waste Isolation Pilot Plant (WIPP). The WIPP studies included elicitation of information on human intrusion, the longevity of markers at a waste disposal site, retardation, and radionuclide concentrations in brine (Trauth *et al.* 1993), and a more recent effort on waste particle size (DOE, 1997). Full details of the Quality Assurance procedures established by the WIPP project for expert elicitation are provided in Rechard *et al.* (1992).

The report by Trauth *et al.* (1993), and this Appendix, provide a broad overview of the procedures and useful guidance. The main points are as follows:

Elicitation Procedure

Expert elicitation should be conducted according to a written procedure:

- A documented procedure for eliciting expert judgement is an aid both in maintaining the quality of the process and in supporting internal and external reviews of the process.

- The documented procedure allows reviewers to reconstruct the logic and the events involved in the use of expert judgement. The availability of such documentation supports a better understanding and acceptance of what was undertaken.

Issue Statement

Any elicitation requires an issue statement that describes what must be addressed by the expert(s) and the format in which the judgement(s) must be presented. The development of the issue statement also plays a major role in identifying the types of experts and/or the fields of expertise to be included for eliciting expert judgement.

Panel Selection

The selection of a panel of experts is a key element of the overall elicitation process:

- Expert-panel members should be carefully selected to ensure that the panel is made up of individuals with the appropriate qualifications and a willingness to participate fully in the expert-judgement process.

- It is also important to include individuals representing the spectrum of thought on a topic, if there may be different approaches or paradigms.

- Also required for an effective panel are members who understand and accept the constraints of the expert-judgement elicitation process. For example, an expert may not be comfortable in making judgements based on incomplete data, but the information may be necessary even if the experiments are still underway or cannot completely answer all questions. Another example of a constraint is that while an expert may not be familiar with expressing his or her judgements as probabilities, the issue statement requires the judgement to be in this format.

- The use of an independent selection advisory committee (SAC), whose function is to rank the qualifications of the nominees and suggest a final composition that covers all the required disciplines, can bring additional credibility to the panel selection process. Whenever possible, the pertinent disciplines should be represented on the SAC.

Training

The elicitation process is more than just the application of probability assessment tools to the judgements of experts. Experts may not be accustomed to expressing their beliefs in the form of probabilities, and training of the experts is therefore usually required:

- This training has multiple objectives. One is previewing the process including how the experts' judgements will be used. Other objectives are to instil confidence in the experts, not only so that they can express beliefs as probabilities and probability density functions (pdfs), but also because lack of confidence in the process may undermine the effort.

- Experts may object to the formal elicitation and encoding of judgements into probabilities and pdfs because they believe 'opinion' is being substituted for 'objective' scientific research. However, the experts' role is not creating knowledge, but is instead synthesizing disparate, incomplete, or conflicting sources of information to produce an integrated picture. Experts who appreciate their role from this perspective are likely to be cooperative.

- The fundamental objective of elicitation training is to help the experts accurately express their beliefs as probabilities and pdfs. Training introduces experts to the tasks they must perform. It also identifies their potential biases and provides practice in controlling them. Although practice may not lead to perfect elicitation, evidence shows that practice improves elicitation.

Organization of Work

There are alternative approaches to organizing a group of experts. These approaches vary with respect to the scope of the issues being addressed, the amount and type of interaction among the experts, the amount of redundancy, and the role of the experts in defining objectives. The availability of the experts and other project management constraints may also affect the way in which the elicitation is organized:

- The simplest organization is experts working in isolation from each other. When there are several experts addressing the same issues, redundancy is useful because multiple experts will have alternative viewpoints, thereby increasing the potential for describing the full range of uncertainty. Alternatively, panels may be organized in which experts work together sharing information and approaches to the issues.

- Another strategy for the analysis of complex issues involves multi-disciplinary teams of experts. This approach is relevant when the issue to be addressed is difficult to decompose into a series of smaller, independent or conditional issues. Multi-disciplinary teams were used for the WIPP human intrusion and markers panels.

- Once the panel of experts has been organized, their efforts must also be organized. Panels convened for the WIPP project participated in an initial informational meeting, where the issue statement was discussed and the training took place. The initial meeting was followed by a work period during which the experts considered the issue statement. All actual elicitation occurred during a second meeting of the panel.

Elicitation Facilitator

Whatever organizational scheme is used for the experts, elicitation is usually accomplished through the use of specialists, sometimes termed normative analysts, from the fields of probability assessment and decision analysis:

- One useful design for elicitation is an assessment team working with one expert. The team members might include a normative analyst who handles the elicitation, a staff analyst who is familiar with the subject area and assists in the communication between the normative analyst and the expert, and a person responsible for documenting the session.

- The elicitation session normally takes place under the control of the normative analyst. The expert is first questioned about the fundamental way he or she has analyzed or decomposed the issue. If the question has been decomposed by the expert, then the decomposition provides the starting point for the discussion. The normative analyst has three goals: to extract the rationales for judgements; to quantify the judgements; and to assist the expert in making the elicited information accurately reflect the expert's knowledge. Reaching these goals requires a good deal of interaction between expert and normative analyst; the expert is often challenged to justify his or her beliefs. Each aspect of the elicitation is approached from multiple viewpoints to reveal and resolve inconsistencies.

- Care must be taken that the normative analyst and staff analyst share a common understanding of the questions being addressed. In particular, if several assessment teams are working simultaneously, some supervision is needed to ensure the consistency of their efforts. In the WIPP study of future societies, some differences among the four teams can be accounted for by the differing approaches used by the normative analysts.

Consistency Check

A variety of elicitation techniques can be used in a single assessment session to facilitate consistency checks:

- When assessing continuous distributions, for example, direct elicitation of interval probabilities and interval bisection, the process of dividing successive intervals into subintervals of equal probability, can be used together. The analyst guiding the elicitation will ask questions that permit comparisons of probabilities. When inconsistencies are found, the specialist will inform the expert of the incompatibility of the probabilities and assist in modifying the assessments and reconciling the differences.

Problem Decomposition

Decomposition of a problem is also an important tool in a successful elicitation:

- The principle behind decomposition is that better quality probabilities and pdfs can be obtained when the assessment tasks are better defined. One decomposition tool is the "influence diagram". An influence diagram is a graphical representation showing the interactions of influencing factors and a decision.

- The issue may be decomposed by the sponsor, the analyst, the experts, or cooperatively. Although the need for expert judgement and its intended use must be explained to the experts, the actual recomposition of the elicited information should remain the responsibility of the sponsor. This ensures that the recomposed information is compatible with its intended use.

- The purpose of decomposition is to subdivide a complex problem into components that are easier to address. During the elicitations, the normative analyst attempts to elicit best estimates of the individual components, without considering what the recomposed result might be. With the sponsor as the party responsible for the recomposition, there is no opportunity, either intentional or unintentional, for panel members to alter individual component judgements in order to explicitly affect the eventual recomposition.

www.ski.se

STATENS KÄRNKRAFTINSPEKTION
Swedish Nuclear Power Inspectorate

POST/POSTAL ADDRESS SE-106 58 Stockholm
BESÖK/OFFICE Klarabergsviadukten 90
TELEFON/TELEPHONE +46 (0)8 698 84 00
TELEFAX +46 (0)8 661 90 86
E-POST/E-MAIL ski@ski.se
WEBBPLATS/WEB SITE www.ski.se