

KNOWLEDGE MANAGEMENT AND ENGINEERING AT A RISK-INFORMED REGULATORY AGENCY: CHALLENGES AND SUGGESTIONS

Nathan Siu, Kevin Coyne, and Felix Gonzalez

Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, DC, USA 20555-0001

Abstract

The U.S. Nuclear Regulatory Commission, as a risk-informed agency, is increasingly using multidisciplinary, multifaceted, and technically specialized information to support regulatory decision making. Ongoing knowledge engineering developments that may help agency staff identify, access, and assimilate relevant information in an increasingly voluminous, broad, and deep information base are promising. Further developments aimed at improving day-to-day tools used by the staff should consider short-term activities to improve database infrastructure and current search tools, as well as longer-term activities aimed at developing improved technologies to extract information from documents.

This is an extended version of a paper submitted for publication in:

Knowledge in Risk Assessment and Management, T. Aven and E. Zio, eds., John Wiley & Sons, Hoboken, NJ, 2017.

KNOWLEDGE MANAGEMENT AND KNOWLEDGE ENGINEERING AT A RISK-INFORMED REGULATORY AGENCY: CHALLENGES AND SUGGESTIONS

Nathan Siu, Kevin Coyne, and Felix Gonzalez

Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
Washington, DC, USA 20555-0001

Abstract

The U.S. Nuclear Regulatory Commission, as a risk-informed agency, is increasingly using multidisciplinary, multifaceted, and technically specialized information to support regulatory decision making. Ongoing knowledge engineering developments that may help agency staff identify, access, and assimilate relevant information in an increasingly voluminous, broad, and deep information base are promising. Further developments aimed at improving day-to-day tools used by the staff should consider short-term activities to improve database infrastructure and current search tools, as well as longer-term activities aimed at developing improved technologies to extract information from documents.

1 A LEARNING OPPORTUNITY MISSED

On the evening of December 27, 1999, storm winds caused a loss of offsite power for Units 2 and 4 of the French nuclear power plant Le Blayais (IPSN, 2000; IAEA, 2000; Gorbachev et al., 2000; Vial, Rebour, and Perrin, 2005).¹ Shortly afterwards, a combination of rising tide, storm surge, and wind-driven waves led to the overtopping of a protective dyke and flooding of the site. Floodwaters entered a number of plant buildings through unsealed cable and pipe penetrations and led to, among other things, the failure of one train of essential service water for Unit 1 and the inoperability of low head safety injection (LHSI) pumps and containment spray pumps for Units 1 and 2.² The flooding also led to falling trees and blocked roads, thereby preventing access to the site for a number of hours.

¹ Offsite power was lost completely at Units 2 and 4. There was also a partial loss of power at Units 1 and 3.

² Each reactor unit at the Blayais plant has four essential service water pumps to transport reactor decay heat to the plant's ultimate heat sink (the Gironde River estuary). The pumps are arranged in two independent trains. One pump (out of the four) is sufficient to ensure adequate heat removal. The LHSI pumps and containment spray system pumps are backup safety components provided to support safe shutdown should a loss of cooling accident occur.

Although the situation was challenging, electric power³ and decay heat removal were available to all four units, and all three operating units⁴ eventually reached a safe shutdown state. The event was classified as an “incident” (Level 2) on the International Atomic Energy Agency’s (IAEA) International Nuclear and Radiological Event Scale (INES), i.e., an event involving “significant failures in safety provisions but with no actual consequences.” However, given the nature of the hazard, the unsealed building penetrations, and the failure of plant structures and components (the dyke and various internal doors) to prevent water movement, it appears that the event could have been much worse.⁵

With the advantage of 20-20 hindsight following the Great East Japan Earthquake of March 11, 2011 and the subsequent tsunami-induced reactor accidents at the Fukushima Daiichi site, it is now clear that the Blayais flood should have been a warning to the probabilistic risk assessment (PRA⁶) community, whose job is to identify and analyze potentially risk-significant scenarios, including scenarios that lie outside a plant’s design basis.

External flooding hazards have been analyzed in a number of early PRA studies, but their risk significance was generally considered to be relatively low. Indeed, NUREG-1407 (Chen et al., 1991), which provided the U.S. Nuclear Regulatory Commission’s (NRC) guidance to licensees performing vulnerability assessments under the Individual Plant Examination of External Events (IPEEE) program, stated that “*core damage frequencies are estimated to be less than 10^{-6} per [reactor] year for a plant designed against NRC’s current [external flooding] criteria.*” Based on this perspective, NUREG-1407 indicated that licensees could avoid detailed analysis of external floods as long as the plant’s design basis for such events met the deterministic criteria provided in the NRC’s 1975 Standard Review Plan (SRP) (NRC, 1975).⁷ Using this and other criteria (e.g., regarding the estimated frequency of severe floods), the large majority of licensee submittals (58 of the 70 provided to the NRC) determined that external floods did not require detailed analysis (NRC, 2002). In PRA parlance, these events were “screened out.” Of the remaining 12 submittals, the reported flooding CDFs ranged from insignificant (around 2×10^{-8} /ry) to potentially significant but not dominant (around 7×10^{-6} /ry).

Blayais provided an empirical indicator of the potential risk significance of beyond design basis external floods. (The plant’s original design basis did not consider the combination of

³ Units 1 and 3 had power from offsite sources; Units 2 and 4 were powered by emergency diesel generators (EDGs). Unlike many plants in other countries, the EDGs at French plants are air-cooled, and are therefore not dependent on service water for cooling.

⁴ Unit 3 was in a refueling outage and had already shut down.

⁵ Vial, Rebour, and Perrin (2005) indicate that the conditional core damage frequency (CDF) during this event was about 3×10^{-3} /reactor year (ry), a value significantly larger than the baseline CDF for the plant. Conversation with a regulatory staff member monitoring the event indicates that there was considerable worry at the time because it wasn’t known where and when the rising water level within the plant buildings would peak.

⁶ In general, the terms “probabilistic risk assessment (PRA)” and “probabilistic safety assessment (PSA)” are synonymous.

⁷ NUREG-1407 provided similar guidance for non-seismic and non-fire events. These events were collectively addressed under the category “high winds, floods, transportation and nearby facility accidents.” The expectation was that such events were, for most plants, unlikely sources of vulnerabilities (as compared with seismic or fire events).

storm-driven waves with a high water level in the estuary.) Moreover, the event involved multiple concurrent hazards, affected multiple units, led to complications in plant operator response, and blocked access to the site. These are characteristics that are now widely recognized as important contributors to the Fukushima accidents, and also as challenges to existing PRA methods, models, tools, and data (e.g., Siu et al., 2013). However, neither Blayais, nor the tsunami-induced flooding of the Madras plant in 2004 (AERB, 2005), nor the flood-debris clogging of service water at the Cruas plant in 2009 (Dupuy, Georgescu, and Corenwinder, 2014), initiated much activity in the general PRA community. A review of the conference programs for the Probabilistic Safety Assessment and Management (PSAM) and PSA conferences held after Blayais and prior to Fukushima shows considerable interest in the treatment of internal flooding (especially in the 2008–2011 time period), but little activity regarding external flooding. It took the Fukushima accidents, bolstered by compelling images of the U.S. Fort Calhoun nuclear power plant site surrounded by floodwaters in June 2011,⁸ to spur serious reconsideration of external flooding in particular and extreme non-seismic hazards in general.⁹

The failure of Blayais and other pre-Fukushima flooding events to affect PRA community views on the potential importance of flooding is troubling from a very practical perspective. As indicated in Regulatory Guide (RG) 1.200 (NRC, 2009), PRA standards are important tools for ensuring that a PRA provides appropriate risk information to decision makers. However, the 2009 version of the American Society of Mechanical Engineers/American Nuclear Society PRA standard (ASME/ANS, 2009), which has been endorsed by the NRC, follows the lead of NUREG-1407 (a 1991, pre-Blayais document) and allows screening using the previously mentioned SRP criterion. Furthermore, the standard states, *“these [external flooding PRA] approaches, based on a combination of using of the recurrence intervals for the design-basis floods and analyzing the effectiveness of mitigation measures to prevent core damage, have usually shown that the contribution to CDF is insignificant.”* The ASME/ANS standard is currently being revised and may better address the potential for premature screening of external events, including external floods. Nevertheless, in the years since Blayais, whether it was because the PRA community was unaware of published reports, or the actual event did not extend beyond the “incident” stage, or readers did not consider the specifics of the event to be relevant to their facilities, or some other reason or combination of reasons, the message of Blayais and the opportunity for changes in PRA practice were missed.

The Blayais experience highlights the challenge of making significant operating experience salient and relevant to the technical community. Key information indicating the potential risk significance of external flooding was published soon after the event. However, this information was either not readily available to the proper technical communities or simply overwhelmed by

⁸ For example, see “Flooding Brings Worries Over Two Nuclear Plants,” The New York Time, June 20, 2011 (<http://www.nytimes.com/2011/06/21/us/21flood.html>).

⁹ As late as 2014, discussions at an international workshop on multi-unit risk revealed that even some prominent PRA experts were not aware of the Blayais event or its significance.

the routine deluge of information of the modern world. Thus, making user-friendly information management tools accessible to the larger technical community and helping to strengthen the signal-to-noise ratio of pertinent information sources is an important knowledge management objective.

2 OBJECTIVE AND SCOPE

The preceding discussion illustrates one risk-related knowledge management (KM) challenge for the NRC as well as the broader PRA community: ensuring that information users (e.g., researchers, analysts, and decision makers) are cognizant of relevant international operational events and their potential risk significance. Risk and KM are broad topic areas, of course, and there are many other challenges.

The objective of this paper is to provide an overview of:

- How, why, and where the NRC uses risk information in its decision making;
- The multi-faceted nature of this information and the associated knowledge engineering (KE)¹⁰ challenges for information systems supporting KM; and
- Potential technical approaches to improving the management of risk information.

The paper's discussion of KE is intended to highlight important engineering aspects of KM.¹¹ The authors' perspective is that of PRA "user" staff who (1) are charged with providing risk information to others, including decision makers; and (2) rely heavily on the NRC and public information systems to perform their work.

3 RISK INFORMATION AT THE NRC: USES AND PERSPECTIVES

This section provides background information regarding the NRC's approach to KM, its uses of risk-related knowledge and information, and different ways of viewing risk-related knowledge and information that can affect KM and pose KE challenges.

3.1 KM at the NRC

KM involves the capturing, preservation, sharing, and use of organizational knowledge (Hudson et al., 2014; Reyes, 2006). The NRC has, since its inception, conducted knowledge management activities such as staff development programs, training, and mentoring. A formal KM program was established in 2006, spurred by (1) the changing demographics among NRC

¹⁰ There appears to be no commonly accepted definition for the term "knowledge engineering." In this paper, we use it to refer to engineering activities associated with the development and maintenance of information systems.

¹¹ KE and information systems are, of course, only part of the KM toolbox. For example, in the case of Blayais, a non-technical KM solution would be to improve cross-disciplinary and cross-organizational communication on topics relevant to risk.

staff and (2) the expectation of increased work with different knowledge requirements (e.g., regarding the licensing of new reactors with potentially novel technologies and operational concepts) for the agency. The initial focus of the program was on capturing and preserving the knowledge of aging staff, then transferring that knowledge to others.

KM is currently an agency priority, as recognized by the NRC's Strategic Plan for 2014–2018 (NRC, 2014). Figure 1 shows that the scope of the program goes beyond the original vision. KM, of course, is a driving force for the NRC's standard documents (e.g., regulatory guides), information systems (notably, the NRC's Agencywide Document Access and Management System [ADAMS]¹²), and staff training programs. The NRC's KM program also covers a number of KM-dedicated activities, including the production of special KM reports (e.g., on the 1979 Three Mile Island accident [NRC, 2016b]); the organization of KM workshops, panel discussions, and seminars (e.g., on the origin of the WASH-1400 study [NRC, 2016c]); and the formation and coordination of “communities of practice” for selected topic areas.

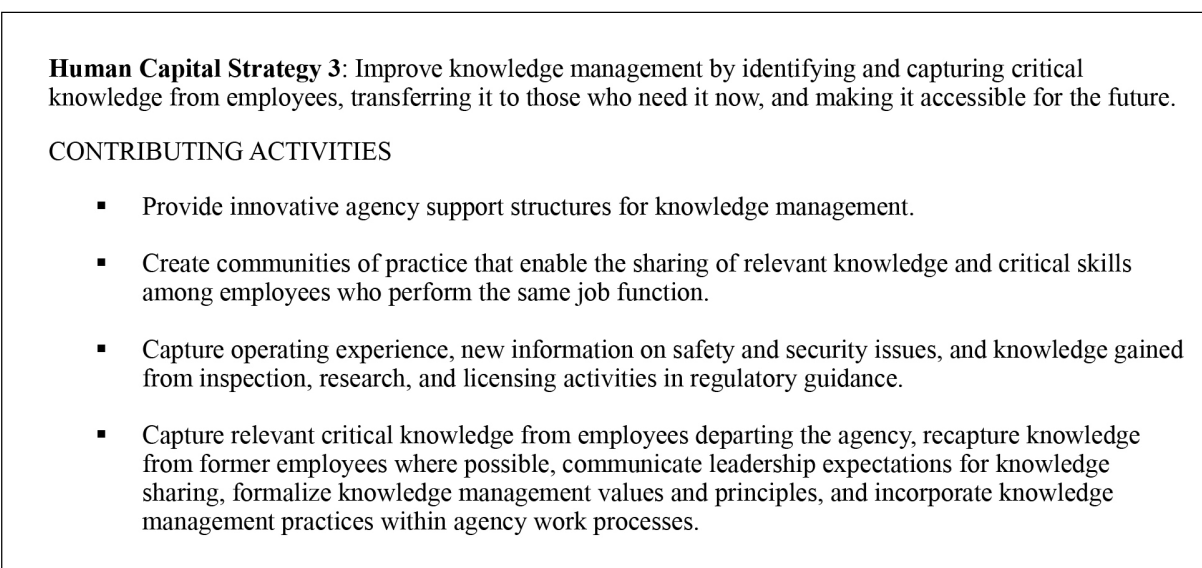


Figure 1. NRC KM-related strategic planning (NRC, 2014)

Our experience indicates that the agency's KM champions and activity leads described by Hudson (2014) have been critical for the promotion, coordination, and implementation of successful KM program activities. The communities of practice provide useful informal settings for the transfer and sharing of knowledge. (Discussion summaries and presentations can be captured in staff-created websites and reports stored in ADAMS.) More formal KM seminars,

¹² For a description of and access to ADAMS, interested readers can go to <http://www.nrc.gov/reading-rm/adams.html>.

panel discussions and workshops are recorded, transcribed and adapted to support post-activity training. In some cases, such training is required as part of a formal qualification program.

Recently, the NRC has been implementing the concept of “Centers of Expertise” (NRC, 2015a). These are defined as organizations that provide agencywide centralized services, leadership, best practices, processes, support, mentoring, training, and knowledge management for specific focus areas in accordance with established priorities.

Regarding risk-related KM, Tobin, Coyne, and Siu (2011) observe that a wide range of resources are available to NRC staff. Some of the more important resources include formal risk assessment and risk management training courses, informal training (including on-the-job training, seminars, and conferences), communities of practice, technical libraries, and electronic databases (including document collections).

3.2 Data, Information, and Knowledge

Discussions of KM systems and programs frequently use the related terms “data,” “information,” and “knowledge.” In this paper, we follow Reyes (2006):

- *Data*: Structured, factual records of discrete transactions or events.
- *Information*: Data that is structured or arranged to inform or influence by communicating a message to a recipient (user).
- *Knowledge*: Information that is influenced by and combined with the user’s own experiences.

The principal focus of this paper is on risk information. However, it should not be forgotten that the ultimate purpose of a KM program is to manage knowledge, which implies organizational activities supporting user recognition and understanding, such as those discussed previously.

It should also be recognized that knowledge can be explicit, implicit, or tacit. *Explicit knowledge*, i.e., knowledge that can be declared (e.g., the key conclusions of a report), is the most easily identified, codified, and transmitted. *Implicit knowledge*, i.e., knowledge that is implied but not directly stated (e.g., additional insights resulting from the user’s synthesis of a report’s key conclusions), is both more difficult to elicit and more user-dependent. *Tacit knowledge*, i.e., knowledge expressed without words or speech (e.g., how to abstract a complex system in developing a model, how to persuade others to take specific actions based on a report’s key messages), is generally based on skills obtained through practice and experience. By its nature, such information is the most difficult to explain and communicate. It can be seen from the Blayais example that all three forms of knowledge are relevant to risk-related KM.

3.3 NRC Uses of Risk Information

3.3.1 PRA Policy Statement

In 1995, based upon its experience with PRAs to date, the NRC issued a policy statement (NRC, 1995) to encourage increased use of PRA methods, and to help ensure that potential applications of PRA could be implemented in a consistent and predictable manner that promoted regulatory stability and efficiency. Besides setting the NRC on its current path of increasing and improving its use of risk information, this document provided some important principles for PRA, which can be paraphrased as follows:

- PRA is useful because it (1) considers a broader set of potential challenges to safety than traditional, deterministic analyses; (2) supports an improved prioritization of these challenges; and (3) can treat a broader set of resources to defend against these challenges.
- PRAs used to support regulatory decisions should be realistic.
- PRAs should be capable of supporting the relaxation as well as the addition of regulatory requirements, and should be of sufficient quality to support decision making.

Clearly, these principles have implications for the information sought from PRA studies, and for the information used to support the performance of such studies.

3.3.2 Risk-Informed Regulation

In 1998, NRC staff issued a tone-setting white paper on risk-informed, performance-based regulation that (1) defines risk as a triplet answer to the three fundamental questions posed by Kaplan and Garrick (1981): “What can go wrong?” “How likely is it?” and “What are the consequences?” and (2) describes how the NRC would use risk information in regulatory decision making (NRC, 1999);¹³ and a landmark regulatory guide, RG 1.174 (NRC, 1998), which provides an acceptable approach by which a licensee can use risk information to support a voluntary change from a plant’s current (and accepted) licensing basis to a new licensing basis.^{14,15}

¹³ The NRC Commissioners approved the white paper in 1999 via a memorandum to the staff. The memorandum (NRC, 1999) includes a slightly edited version of the original paper.

¹⁴ In the U.S., the “licensing basis” for a nuclear power plant is a legal prescription specifying the conditions under which the plant is allowed to operate.

¹⁵ Though developed for a specific application (risk-informed changes to a plant’s licensing basis), the RG is written quite broadly, and its general philosophy and some specific criteria have been adopted in many other risk-informed regulatory applications.

Both the white paper and RG 1.174 emphasize that the NRC’s approach is “risk-informed” (i.e., the results and findings of risk assessments are considered together with other factors) rather than “risk-based” (i.e., decisions are based solely on the results of the risk assessments). RG 1.174 provides specific guidance regarding the elements of a risk-informed decision. In particular, the decision needs to consider current requirements, considerations of defense-in-depth and safety margins, and the use of monitoring, as well as risk (see Figure 2).

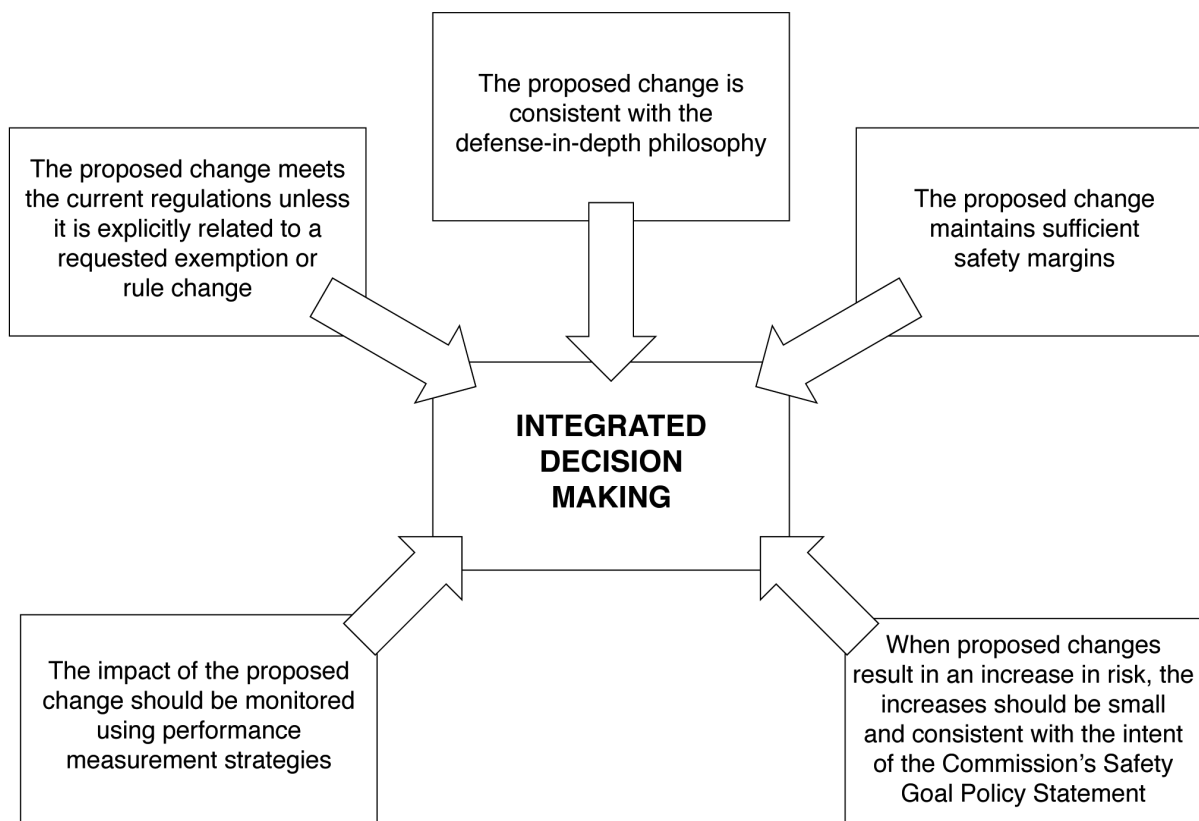


Figure 2. Risk-informed decision making (NRC, 1998)

3.4 Information to Support Risk-Informed Decision Making and “Risk Information”

Figure 2 shows that the information needed to support a risk-informed decision is not limited to “risk information,” i.e., the information indicating the risk associated with a particular situation or decision. Furthermore, it should be recognized that the risk information is more extensive than that provided by a PRA. For example, a licensee’s PRA team may submit a set of results and insights it considers suitably realistic for the decision problem at hand. In many instances, the regulatory staff will need to review the licensee’s submittal and make an independent determination of its suitability. In non-routine cases, including cases in which new

information (e.g., a major event) has come to light, the staff may need to evaluate the methods, models, tools, and data used by the PRA. An effective KM support system will help the staff efficiently locate, obtain, review, and understand information (e.g., experiments, analyses, operational experience) needed to support this evaluation. The agency has developed a series of reports (e.g., NRC, 2016d) that provide historical summaries of selected topics to help the staff understand the development of knowledge and agency positions over time.

To further expand on potential complexities affecting risk-related KM, it is useful to consider the topic of risk information from the perspectives of intended use (“form follows function”), information characteristics, and current trends.

3.4.1 Intended Use

Regulatory Applications. As discussed in NUREG-2150 (Apostolakis et al., 2012), the NRC currently uses risk information in all areas of regulatory purview, including materials (e.g., medical radioisotope sources), waste (both low-level and high-level), uranium recovery, nuclear fuel cycle facilities, interim spent fuel storage, and transportation, as well as reactors (both power and non-power).¹⁶ The extent and formality of usage varies across the program areas, and the area-specific recommendations provided by NUREG-2150 vary accordingly.

Within the reactor safety area, as discussed by NUREG-2201 (Siu et al., 2016a), risk information is being used to support all regulatory functions (see Figure 3): regulation, licensing, oversight, and operational experience (evaluation and response). Furthermore, risk information is being used at all organizational levels within the NRC. Thus, for example, risk information is used in relatively routine staff decisions (e.g., the acceptance of a proposed licensing action), in Commission policy and rulemaking decisions (e.g., regarding the imposition of broad requirements for major plant changes across the industry), and situations in between (e.g., deciding whether to allow a plant to continue operation under degraded conditions).

Decision Support. Figure 4, which is taken from NUREG-2150, provides a view of the regulatory decision making process. Although not unique to risk-informed decision making, it can be seen that different steps in the process can involve different types of risk information (e.g., information supporting the identification of potentially viable options vs. information supporting detailed analyses of the options) and different modes of support for different phases (e.g., staff development and transmission of initial technical results and insights during the analysis phase vs. staff interactions with decision makers during the deliberation phase).

¹⁶ An overview of these regulatory areas can be found on the NRC’s website: <http://www.nrc.gov/>.

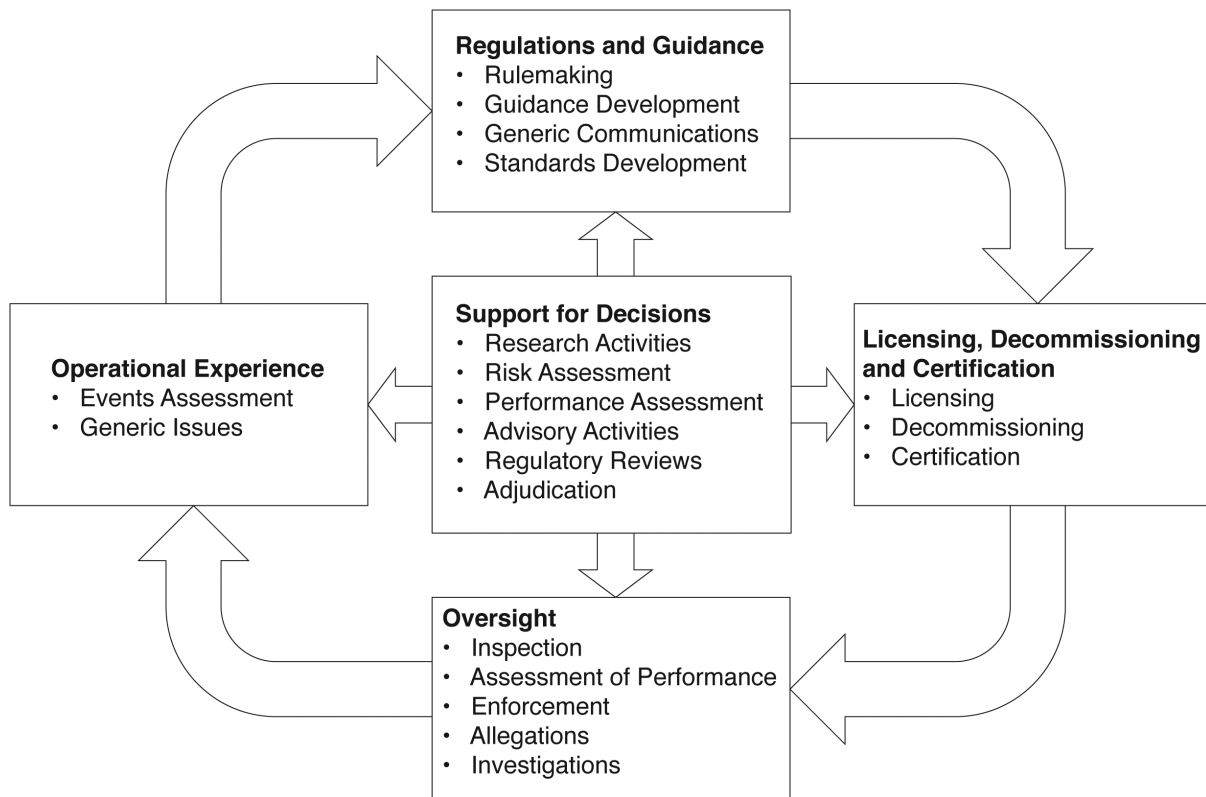


Figure 3. NRC regulatory functions

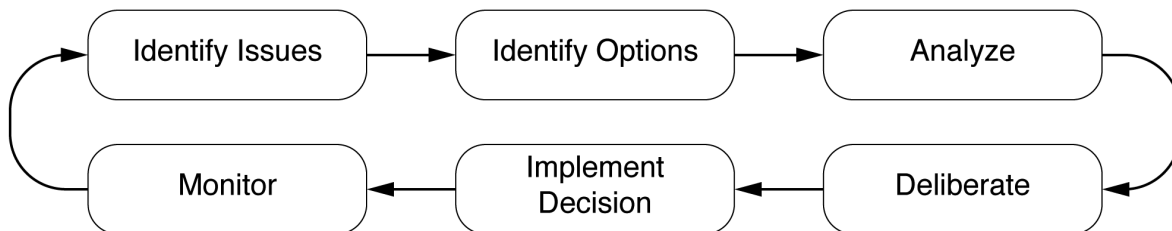


Figure 4. The regulatory decision making process (Apostolakis et al., 2012)

3.4.2 Information Characteristics

Volume. Due to the broad nature of risk information (discussed further below), the volume of information that may need to be considered for a particular decision or process can be enormous. For example, when attempting to review the Fukushima Daiichi accidents (as well as the successful earthquake/tsunami responses at other Japanese plants) for lessons leading to improved PRA methods and models, there are currently over 20 official Japanese investigation and lessons-learned reports and numerous official reports from other countries and international bodies. All are voluminous but potentially valuable (despite some duplication of material), as they are written for different reasons and from a range of viewpoints.

Form. The specific form in which information is provided can be important when developing automated tools supporting KM.

The information used in agency decision making is usually formal (in the form of documents,¹⁷ both digital and print), but can also be informal (e.g., internal knowledge of individual staff, corporate knowledge of the organization and sub-organizations). Documents can contain information in many formats, including video, audio, and images, as well as text. The information can be structured (e.g., databases, PRA models) or unstructured (e.g., free text).

Note that documents typically include both structured and unstructured information: databases usually have freeform comment fields, and text documents have internal structures (e.g., different chapters and sections, tables) that affect the meaning of text. Also note that the meaning of some documents can rely heavily on implicit information: for example, an audio recording of a meeting and the transcript of that meeting will often contain incomplete sentences whose meaning can only be inferred from other parts of the record.

Different information forms also arise from the differing purposes of documents. For example, a research paper assessing the pros and cons of specific PRA methods will have a different format (and possibly even a different intended audience, with associated differences in interests, knowledge, and preferred communication style) than a report summarizing the results of a PRA for a particular facility or an inspection report identifying risk-relevant findings for that facility. (When considering the information needs of risk-informed decision making, the variety of potential document types increases significantly, as shown in Figure 5.)

¹⁷ In this paper, the term “document” is used broadly to encompass records (e.g., computer files) that include but are not limited to traditional text documents.

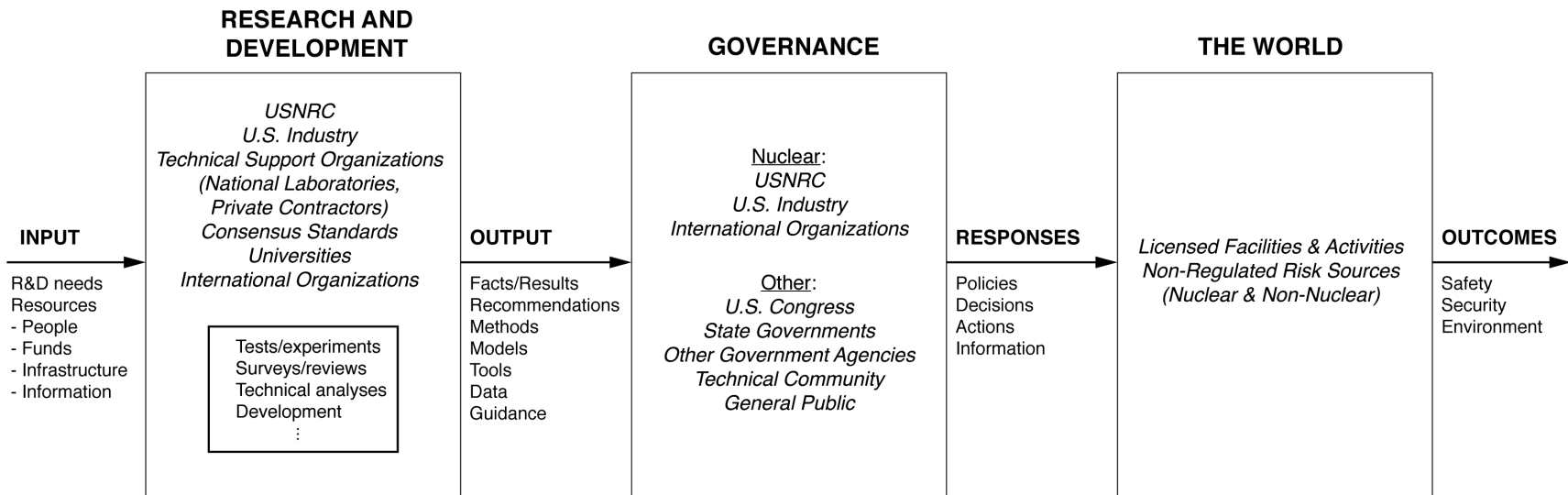


Figure 5. From R&D to outcomes: generation and transmission of information (adapted from NAP, 1996)

Risk Problems. Many risk problems of concern to the NRC have characteristics that distinguish them from other regulatory decision problems. These characteristics include:

- *Need for systems viewpoint.* A risk-informed decision generally needs to consider the performance of the system as a whole, and should not focus exclusively on one aspect of the problem. In some cases, analysts and decision makers may need to cope with situations in which potentially important parts of the system are poorly understood.
- *Diverse and implicit sources of information.* The basis for a PRA model may reside in a wide range of sources (e.g., licensing basis information, operating experience, licensee submittals) that may or may not be explicitly referenced in the PRA model's documentation. Understanding this basis can be key to the appropriate use of the model's results and insights.
- *Involvement of multiple disciplines.* Dealing with the system as a whole typically requires input from a wide range of technical disciplines. These disciplines have, in addition to their unique bodies of knowledge, their own technical cultures, which affect how they use and create information (including how they view and deal with uncertainties).
- *Problem complexity.* A risk management decision may require consideration of a large number of disparate scenarios. For example, both scenarios triggered by low-probability/high-consequence natural disasters that immediately overwhelm facility defenses and scenarios involving chains of less severe but more likely failure events could be important to a facility's risk profile.
- *Rare events.* Risk assessments and risk-informed decision making often deal with unlikely, beyond design basis events. In some situations, analysts and decision makers need to deal with novel designs and even design principles. In situations where direct experiential data are sparse, modeling (including modeling assumptions) plays a fundamental role, and it is critical that modeling details be adequately understood.
- *Legacy documents.* In addition to the long return periods of potentially key but rare phenomena (e.g., extreme earthquakes, floods), risk-informed decisions may need to consider old but still relevant information (e.g., analyses of and decisions regarding previously considered reactor concepts which have become the subject of renewed interest). Some of this information may reside in legacy, non-digitized documents.
- *Important "details."* Risk-significant scenarios can arise from unique, plant-specific design and operational features that lead to subtle dependencies between potential failure events. Changes in relatively small details (e.g., the location of a particular set of electrical cables) sometimes dramatically impact a risk study's results and insights.
- *Large uncertainties.* Due to preceding factors, the results of a PRA are often subject to large (sometimes multiple orders of magnitude) uncertainties. Determining and communicating the meaning of such results, especially in situations where the events treated are rare and the overall risks are small, is a continuing challenge.

- *Broad user base.* Within the NRC, risk information is being generated and used by staff with a broad range of technical backgrounds and exposure to risk concepts. For example, PRA models can be developed, and their results used, by inspectors with limited exposure to PRA, as well as by PRA experts. Similarly, some NRC decision makers have only limited formal training in risk-informed decision making, while others formerly worked as PRA technical experts.

3.4.3 Current Trends

A number of trends relevant to risk information needs have been underway for some time at the NRC. These include:

- *Increasing use of currently available PRA models.* The number of risk-informed licensee requests requiring NRC review and approval is increasing (see Figure 6). Many of these requests (e.g., to increase allowed outage times for safety equipment) involve relatively routine use of PRA models. The NRC's Standardized Plant Analysis Risk (SPAR) models (NRC, 2015b) are also being increasingly used (e.g., to support evaluations of the risk significance of inspection findings).

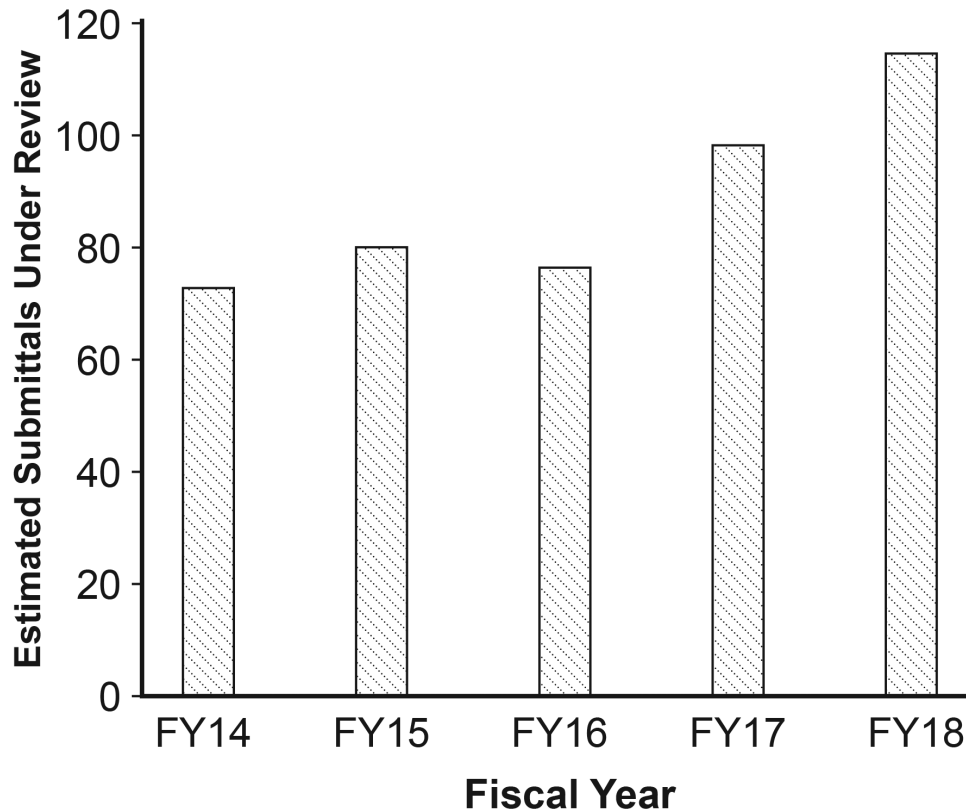


Figure 6. Projected risk-informed licensing actions (adapted from Rosenberg, 2016)

- *Increasing number of applications requiring extensions of the PRA models.* Such applications (e.g., risk-informed criteria for reactor pressure vessel embrittlement, safety significance of consequential steam generator tube ruptures) impose an associated burden on decision makers to understand model assumptions and limitations in these extended applications. Notable PRA modeling extensions in support of these applications include severe accident, seismic, and fire modeling.
- *Increasingly demanding use of the PRA results and insights.* Situations in which PRAs are used to support decisions where the absolute results play a significant role (e.g., when a recommended course of action relies in part on the “bottom line” results of the PRA) impose a greater burden on the PRA than situations in which relative results are sufficiently informative. Decisions requiring the assessment of the value of a new technology (e.g., a new approach to fire detection) can also require more precision from the PRA than decisions requiring only the assessment of a general class of risk contributors (e.g., fires).
- *Increasing need to review reactor design applications with new features.* Reactor designs involving extensions of current light water reactor designs (e.g., multiple reactor modules, new concepts of operation) or major departures (e.g., high-temperature gas or molten salt designs) can require PRAs to address phenomena, scenarios, and dependencies not addressed in PRAs for the operating fleet.
- *Changing demographics of NRC staff as more experienced members approach retirement.* These changes affect the average level of risk-related experience among the staff (many new staff members have not had the chance to develop hands-on experience with practical PRA modeling prior to joining the NRC).¹⁸

The 2011 Fukushima Daiichi reactor accidents and concurrent challenges at other Japanese nuclear power plants have highlighted a number of additional issues relevant to the performance and use of PRA. The resolution of these issues may further affect the agency’s changing risk information needs. As discussed by Siu et al. (2013), these issues include:

- the scope of current PRAs for operating U.S. plants, many of which, according to a post-Fukushima U.S. Government Accountability Office report (GAO, 2012), have not updated their treatment of external events since the IPEEE program in the 1990s;
- the risk from events involving multiple units at a site and even multiple sites;
- the assessment and treatment of uncertainty for extreme natural events (and, more generally, low-probability/high-consequence events); and

¹⁸ Note that the increasing use of digital devices and media has not only dramatically improved access to potentially relevant information, it has also led to changes in expectations (e.g., that answers to factual questions should be readily available), in increased use of non-text forms (e.g., images, video, audio) to convey information, and even changes in reading habits (e.g., skimming versus deep reading). For example, Rosenwald (2014) provides an overview of changes in reading habits spurred by digital devices and media and associated effects on comprehension.

- the appropriate balance of deterministic and probabilistic information in regulatory decision making.

Regarding the last point, following the Fukushima Daiichi accident, there have been calls to reduce the emphasis on, or even to entirely abandon, the explicit use of risk information. Such calls include recommendations to perform “worst-case analyses” and to develop mitigation strategies that are independent of accident cause. Even within the PRA community, there have been proposals to increase the emphasis on “conditional analyses” and “resilience” (e.g., Lanore et al., 2013; Takada, 2013). The ongoing discussions in this area are beyond the scope of this paper. However, it is important to recognize that the outcome of these discussions could affect the agency’s risk information needs.

4 RISK-RELATED KE CHALLENGES FROM A USER’S PERSPECTIVE

The NRC, as with any organization that deals with large volumes of information, has a number of information technology systems and associated activities aimed at (1) electronically capturing information important for the agency’s decision making efforts and (2) making the captured information accessible to the staff. In addition to the NRC’s official recordkeeping system (ADAMS), staff can access information through a variety of tools, including the agency’s website and staff-created sites used to share information. Staff can employ a variety of standard and NRC-specific search tools (e.g., those included in ADAMS) and other aids (e.g., hyperlinks, file structures, citations and reference lists, tables of content and indices) to find relevant documents (e.g., text files, spreadsheets, databases, images, computer codes and models) and specific pieces of information in these documents.

As illustrated in Section 6, current databases and tools are quite effective and efficient. However, users are naturally interested in improvements that will enhance their ability to find, access, review, and assess potentially relevant information. The KE challenges in making improvements range from the simple to complex. There are three general challenges involved:

- I. Expanding and improving the electronic database (e.g., digitizing more legacy documents, improving the accuracy of digitization);
- II. Improving search tools and aids, including guidance for users, to increase the likelihood that the search process finds desired information while reducing the number of undesired (“false positive”) results; and
- III. Providing capabilities to automatically derive explicit information from implicit information (e.g., by mimicking the ability of subject matter experts [SMEs] to develop insights from a number of documents¹⁹).

¹⁹ In the case of Blayais, for example, inferring from the event description and from siting documents (i.e., documents that provide the technical basis for the geographic positioning of a nuclear power plant) for other nuclear power plants that beyond design basis external floods can be important for other nuclear power plants.

These general challenges are clearly not unique to the risk arena; significant development efforts are underway in the commercial information technology sector, and a number of products are already available, as discussed in the following section. However, from the perspective of a risk information user, the challenges are modulated (and probably amplified) by the risk information characteristics described in the preceding section. For example, the interdisciplinary nature of a PRA implies that KE solution development may require a wide range of SMEs (e.g., to provide word/phrase associations and search heuristics). This, in turn, implies a special challenge in ensuring the efficient use of numerous and diverse experts.

It should also be noted that some of the special characteristics of risk information may affect the effectiveness and efficiency of KE solutions sought by other communities. For example, analytics-based approaches that rely solely upon the number of times a search query finds a document with matching text (i.e., the number of search “hits”) may miss situations in which a single document contains information on a rare event of interest, and may place undue emphasis on facts provided by multiple documents that are actually based on the same underlying information, as in the case of reports on the Fukushima accidents.

5 THE PROMISE OF ADVANCING TECHNOLOGY

One useful approach for meeting the general challenges discussed in the preceding section involves enlisting additional knowledge from SMEs to organize and make sense of the risk information stored in current databases. Three core technologies for automating this process are natural language processing, content analytics, and formal methods.

5.1 Natural Language Processing

Figure 7 includes an excerpt from Gorbachev et al. (2000), which provides a portion of the description of the 1999 Blayais flooding event. The meaning of the excerpt is clear to human readers familiar with the underlying terminology and concepts. However, with an automated system, challenges arise due to ambiguity, context dependence, implicitness, and non-uniqueness. For example, the text indicates that the Train A service water pumps were “lost,” while other (LHSI and containment spray) pumps were “considered completely unavailable.” Should an automated system supporting the use of risk information consider these effects synonymous? If so, under what circumstances? Table 1 provides additional examples of natural language challenges associated with the Blayais excerpt.

“Of the facilities which were flooded in Units 1 and 2 (Illustrations 10 and 11), the following should be noted:

- the rooms containing the essential service water pumps. The essential service water system of each unit comprises four pumps on two independent trains (A and B); each pump is capable of providing the entire throughput required. In Unit 1, the essential service water system pumps of Train A were lost as a result of immersion of their motors;*
- some utility galleries, particularly those running in the vicinity of the fuel building linking the pump house to the platform;*
- some rooms containing outgoing electrical feeders. The presence of water in these rooms indirectly led to the unavailability of certain electrical switchboards;*
- the bottom of the fuel building of Units 1 and 2 containing the cells of the two LHSI pumps and the two containment spray system pumps. The nuclear operator considered that the pumps were completely unavailable. The systems to which these pumps belong are the engineered safety systems of the installation which are designed mainly to compensate for breaks in the primary system.*

Illustration 12 illustrates the roles of the aforementioned systems. The essential service water system operates during normal operation of the units to cool the reactor auxiliaries and when the reactor is shut down to cool the decay heat removal system during accident situations in order to remove the decay heat by heat exchange in the containment spray system heat exchangers.”

Figure 7. Summary Description of 1999 Blayais flood (Gorbatchev et al., 2000)

Table 1. Examples of natural language challenges for automated processing arising in Blayais excerpt (Figure 7)

Challenge Type	Example Phrase	Challenge for KE Tool
Ambiguity (multiple meanings for the same word or phrase)	<i>“pumps of Train A were lost”</i>	Determining that “lost” means “failed” (as opposed to its many other possible meanings, e.g., “missing” or “bewildered”).
Context dependence (meaning depends on other factors, including document type, purpose, structure, and surrounding text)	<i>“essential service water pumps”</i>	Recognizing that “essential” is part of the name of the system (as opposed to being a descriptor; consider the possibility of narrative references to “non-essential service water pumps”).
Implicitness (meaning is not stated directly, and must be inferred from other facts in document)	<i>“the rooms containing the essential service water pumps”</i>	Recognizing that these rooms were flooded, as implied by the preceding “the following should be noted.”
Non-uniqueness (multiple ways of making a statement with the same meaning)	<i>“The essential service water system of each unit comprises four pumps on two independent trains (A and B)”</i>	Extracting information on system configuration from multiple possible alternatives (e.g., using different words, such as “...has four pumps in two separate trains...” or different grammatical constructions, such as “There are four pumps in the essential service water system, arranged in independent trains A and B”).

Figure 7 also illustrates another challenge to natural language processing algorithms: widely separated text. In this figure, the text referring to the “cells” containing the LHSI and containment spray pumps is several text passages away from the introductory text indicating that the discussion concerns flooding. It is easy for human readers to infer that the cells were flooded, but it is not simple for algorithms relying on proximity measures (e.g., the number of words separating phrases). Figure 8, which provides highlighted excerpts from NUREG/CR-6738

(Nowlen, Kazarians, and Wyant, 2001) addressing a 1975 fire at the Browns Ferry nuclear power plant, provides another example. In this case, the Page A3-1 statement that the following event descriptions generally apply to Unit 1 is necessary to determine that the Page A3-5 statement (that the reactor was scrammed at 00:31) applies to Unit 1.

The highlighted text in Figure 8 shows a further challenge: the potential need to deal with flawed data. Due to the particular optical character recognition (OCR) software used to create the digitized version of NUREG/CR-6738, the digitized text available to search tools does not recognize the structure implied by the document graphics. Thus, for example, the highlighted text for the Page A3-6 entry is literally stored as “*At 1:00pm Unit 2 control room operators demand for multi-unit shutdown may introduce [line break] observed several annunciations...*” Aside from being nonsensical, the fragmentation of the actual phrase “*may introduce unique equipment demands*” potentially masks a key message from the document.

Within the information technology industry, advances continue in improving the access to and use of information. One of the most widely publicized activities was highlighted on January 14, 2011, when a computer system called “Watson,” developed by International Business Machines (IBM®), defeated two human experts on the television quiz show “Jeopardy!”.²⁰ In addition to addressing complexities associated with natural language processing, the Watson project demonstrated the ability of computer technology, including the technology currently available and the technology developed specifically for the project, to address challenges associated with the volume, breadth, form, and trustworthiness of potentially relevant information. However, the Watson project, which was large and sustained (the project started in 2005 and involved a core team of about 20 researchers [Ferrucci et al., 2010]), was a focused research activity with a relatively narrow problem domain.

Work is ongoing within many organizations to apply the technologies demonstrated by Watson, including the content analytics technology discussed in Section 6, to a variety of practical problems (Keim, 2015). The relatively free-form query interface supported by common search engines and the widespread deployment and use of voice-activated virtual assistants provide additional demonstrations of the significant progress that has been made in understanding natural language queries and responding in kind.

²⁰ For a popular account, see: “Computer Wins on ‘Jeopardy!’: Trivial, It’s Not,” The New York Times, February 16, 2011 (http://www.nytimes.com/2011/02/17/science/17jeopardy-watson.html?pagewanted=all&_r=0).

From Page A3-1:

Whether an event from the chain of events is typically included in a fire PRA is discussed where deemed appropriate. Lessons that may be gleaned from a specific event in the context of fire PRA are also provided. **Unless otherwise noted, the event descriptions refer to events impacting Unit 1.**

Time (rel. to ignition) (hr:min)	Event Description (Note 1)	Fire PRA Implications
Prior to the incident	The power cables for two 480 VAC boards from opposite safety trains were routed during construction, erroneously, inside the same cable tray. (Regulatory Guide 1.75 which was in effect at the time disallows this practice.)	In a fire PRA, error in routing of cables is not taken into consideration. The actual discovery of such a construction error is rare. No other such incidents are known to the authors. Therefore, the assumption used in fire PRAs should generally be considered as acceptable.

From Page A3-5:

00:31	At 12:51pm, operators manually scrammed the reactor from 704 MWe power level.	It is not entirely clear why operators delayed the scram for 15 minutes after learning of the fire. In a fire PRA a scram immediately upon a report of an unsuppressed CSR fire would typically be assumed.
-------	--	---

From Page A3-6:

--	On the Unit 2 control panel, operators noticed malfunctions on ECCS panel 9-3 and feedwater panel. Unit 2 RB fans were switched to low by the operators.	Typical fire PRAs consider the impact of a fire only on a single unit, even if that fire occurs in a common or shared plant area. In this case, the second unit also experienced some difficulties and was shut down. Simultaneous demand for multi-unit shutdown may introduce unique equipment demands that may not be covered by current fire PRAs.
00:40	At 1:00pm Unit 2 control room operators observed several annunciators regarding DC power and that one reactor protection M-G set had tripped. They proceeded to scram the Unit 2 reactor and initiate shutdown cooling. Unit 2 operator confirmed that all rods inserted.	

Figure 8. Excerpts from analysis of 1975 Browns Ferry fire (Nowlen, Kazarians, and Wyant, 2001)

5.2 Content Analytics

In the information technology world, where increasing amounts of resources are being spent to make better use of large (and ever-increasing) amounts of unstructured information, “content analytics” tools are being developed to, among other things, help users improve their searches and enhance their “discovery” activities (i.e., activities to develop insights through exploration of databases).^{21,22} As further discussed in Section 6, such tools use software routines to convert unstructured data (typically free text) into structured data (e.g., terms with assigned characteristics), make that data readily accessible to user queries, and provide various means (quantitative and qualitative) of characterizing query results.

5.3 Formal Methods

A third line of technology development that may be helpful to the NRC’s risk-informed activities concerns the use of so-called “formal methods.” Formal methods, which are well known in the computer science field, involve the development of mathematical specifications for hardware and software systems, and are intended to support the development and verification of such systems.

The PRA community has long recognized that logically equivalent (or nearly equivalent) models can take many different visual forms. (This recognition is exemplified by the resolution of the “large event tree/small fault tree” versus “small event tree/large fault tree” debate in the early days of PRA, later discussions on the merits of event sequence diagrams versus event trees, and current work on the automated development of binary decision diagram models.) A formal modeling approach could help suitably trained reviewers understand the essential aspects of the model despite these different forms.

A second potential benefit is that a formal modeling approach can put the PRA model being reviewed and external benchmarks (e.g., models of similar systems in other PRAs, relevant operational experience) into a common format, thereby helping a reviewer to identify key similarities and differences.

²¹ The term “content analytics,” although widely used in an information technology context, does not appear to have a standard definition. In this paper, it is used to refer to a broad class of software tools that use a variety of approaches (e.g., natural language queries, trends analysis, contextual discovery, and predictive analytics) to identify patterns and trends across an unstructured database (e.g., text). Note that this usage is somewhat more limited than that implied from dictionary definitions of “content analysis,” which refer to latent as well as manifest content (Webster, 1969).

²² In content analytics literature, the terms “discover” and “explore” are used to indicate a more open-ended use of a database than is implied by the term “search,” which involves looking for the specific answers to a particular question. Thus, a tool aimed at supporting “discovery” can, in addition to responding to a particular search, provide information suggesting further searches that provide alternate perspectives on the search topic.

The Open PSA initiative discussed by Epstein and Rauzy (2013), which is aimed at providing a standardized modeling language for PRAs, provides a promising technology for NRC staff, who often need to function as model reviewers rather than developers.²³ Friedlhuber, Hibti, and Rauzy (2015) present a model comparison methodology, and Meléndez Asensio and Santos (2015) present review-oriented applications developed and considered for Consejo de Seguridad Nuclear (CSN, the Spanish regulatory authority). Both works are based on this technology. Although not further discussed in our paper, it seems clear that formal methods could be very useful for the NRC’s risk-related applications.

6 A RECENT EXPLORATION

In order to provide an indication of the status and potential value of commercially available tools benefitting from advances in natural language processing and document content analysis, the NRC’s Office of Nuclear Regulatory Research (RES) has performed a feasibility study to explore the application of advanced KE tools and techniques to support PRA activities. The study, which was initiated in 2014 and completed in 2016, was conducted under the auspices of the NRC’s Long-Term Research Program (LTRP), which is used to investigate topics expected to meet critical mission needs in five to ten years (NRC, 2012).²⁴ Following LTRP guidelines, the project was conducted as a scoping study aimed at the planning of future KE-related activities.

The following discussion provides a summary of the project. Additional details can be found in Siu et al. (2016b).

6.1 Project Objectives and Scope

The overall objective of the project was to determine whether additional agency effort to develop production-level KE tools aimed at supporting risk-informed applications could be worthwhile.

As a scoping study, the project employed the following limitations:

- The evaluation was limited to the consideration of content analytics tools.
- The evaluation was performed using IBM Content Analytics Version 2.2 (ICA 2.2), which was already licensed to the NRC. Based on responses to a staff request for information on capabilities from potential contractors (a “sources sought” notice [NRC, 2013a]), this tool was judged by the project team to be representative of the broad set of commercially available content analytics tools.

²³ More information on this initiative is available at www.open-psa.org.

²⁴ Per NRC (2016), the NRC is re-evaluating the LTRP as part of its ongoing activities to improve agency efficiency.

- The evaluation considered three applications (“use cases”) described in the following section.
- The documents selected to provide the search space for ICA 2.2 were limited to the document types shown in Table 2. This document set, called a “corpus,” was finalized in late 2015. It included over 330,000 documents, representing a combination of selected documents from the ADAMS library (which, at the time of the project, contained around 2 million documents, of which roughly half are publicly available) and a number of other documents.

Table 2. Project corpus contents

Description	Notes
Publicly available documents from the NRC’s ADAMS Main Library	Includes NRC staff (NUREG) and contractor (NUREG/CR) reports, staff papers to the Commission (SECY papers) and Commission Staff Requirements Memoranda (SRMs), License Amendment Requests (LARs), and New Reactor Design Control Documents.
Final Safety Analysis Reports (FSARs)	Provide terminology and design-related information useful for event analysis.
Documentation for NRC Standardized Plant Analysis Risk (SPAR) models	Provides design-related information useful for event analysis (e.g., the size of the system involved) and PRA results that can be compared with licensee/applicant results.
Immediate notifications	Documents notifying the NRC of events submitted per the requirements of 10 CFR 50.72.
Licensee Event Reports (LERs)	Documents notifying the NRC of events submitted per the requirements of 10 CFR 50.73.
Inspection reports	Staff reports from the NRC’s Reactor Oversight Process (1999–present).
Individual Plant Examinations (IPEs)	Licensee submittals in response to Generic Letter 88-20.
Individual Plant Examination of External Events (IPEEEs)	Licensee submittals in response to Generic Letter 88-20, Supplement 4.
Advisory Committee on Reactor Safeguards (ACRS) letter reports	1985–present.
ACRS Meeting Transcripts	1999–present (subcommittee as well as full committee).

ICA 2.2 consists of a number of major software components, including the following (Zhu et al., 2011):

- “Crawlers,” which go through the documents in the corpus and extract document content;
- Document processors, which convert the unstructured text data generated by the crawlers into structured data using rules provided by text analytic “annotators” (including standard annotators to do such things as identify the document language, perform a linguistics analysis, and identify text patterns using user-supplied rules, as well as any additional custom annotators);

- An indexer, which prepares an optimized index of the processed document content (called a “text analytics collection,” or “collection” for short) suitable for high-speed text mining and analysis; and
- A text mining application, which provides the user interface that enables an analyst to search the corpus.

ICA 2.2 is a general product that can be customized to address the needs of specific problems. This customization process requires (1) that software engineers configure the tool (e.g., to control how a crawler uses system resources and when it should be run) and develop desired annotators, and (2) that SMEs work with the software engineers to collaboratively define the search problem of interest and ensure efficient tool development.

From an end-user perspective, most of the work performed by the software engineers is “behind the scenes.” For example, the SME generally does not construct or perform a detailed review of the annotators produced by the software engineers, but uses a customized text mining application, also produced by the software engineers, which provides a number of tools supporting user searches and discovery. The principal tools are “facets,” different subject-oriented collections of keywords that provide different views of the corpus data, and their associated searches.²⁵ A significant portion of the SME effort in developing the customized ICA 2.2 tool is involved in developing facets for a particular use case that help identify relevant documents without an excessive number of false positives.

Other tools can filter search results and support the development of statistics (e.g., matching document counts, frequencies of and trends in search phrase occurrences, and correlations between pairs of search phrase occurrences) and the visual identification of relationships between facets.

6.2 Overall Approach

The work involved the performance of three case studies (“use cases”) summarized in Table 3: the identification and characterization of operational events involving multiple reactor units (referred to as “Use Case 1”); the determination of current CDF estimates developed in licensee PRAs (“Use Case 2”); and a general exploration of a wide set of documents to identify potentially interesting risk-relevant topics for more detailed investigation (“Use Case E”). The first two use cases employ the ICA 2.2 tool in a traditional search mode to address the typical staff task of searching for answers to highly specific questions. The last employs the ICA 2.2 tool in a more general, discovery-oriented mode.

²⁵ For example, a facet intended to provide a view of operational events involving multiple units could be constructed from sets of keywords capturing important aspects of such events (e.g., extent of effect, causes, coupling mechanisms, near misses). A search hit involving one of these keywords would indicate that the identified document addresses one of these aspects and is therefore potentially relevant to an analysis of multi-unit events.

Table 3. Project use cases

ID	Description	Notes
1	Search for multi-unit events	Supports characterization of past events involving multiple units at a site. This characterization could identify events that may need to be addressed in a site-wide PRA model.
2	Characterization of current licensee PRA results	Supports decision makers' understanding of current risk levels and contributors. This activity addresses a common question raised by managers and external stakeholders.
E	Exploration of corpus	Uses ICA 2.2 in a discovery/exploration mode. This use case supports the project's evaluation of the tool when used in a non-direct search mode.

For each use case, the ability of ICA 2.2 to effectively and efficiently meet staff needs was assessed and compared with the capabilities of other tools currently available to the staff. Use Cases 1 and 2 involved a team of SMEs and software engineers performing four steps:

1. Specify the search problem.
2. Develop a project-specific, customized search application using ICA 2.2²⁶.
3. Test and refine the customized application.
4. Exercise the application to identify and retrieve documents containing the information sought, and compare with alternate approaches.

Use Case E involved a single SME exercising the application “as-is” (i.e., without further modification for the needs of Use Case E).

It is important to recognize that for all use cases, exercising the application involved an iterative search process in which the user provided an initial search query, reviewed results, refined the query, and so forth until either the desired results were achieved or the effort was terminated. Thus, both the application and ICA 2.2 should be viewed as human-in-the-loop tools, rather than fully automatic answer generators, as is the case with IBM's Watson.

6.3 Use Case 1

As argued by Fleming (2005) and illustrated by the March 2011 reactor accidents at the Fukushima Daiichi nuclear power plant, events involving multiple reactor units at a single site can be important contributors to site risk. There are numerous technical challenges in assessing these contributions. NRC/RES is currently engaged in a full-scope Level 3 PRA study intended to address all relevant site radiological sources (including the spent fuel pool and dry cask storage), internal and external initiating event hazards, and modes of operation for a two-unit Westinghouse four-loop pressurized water reactor station with a large, dry containment (NRC,

²⁶ For expediency, a single tool (with multiple use case-specific facets) was developed for the project.

2011; Kuritzky et al., 2013). The technical approach for addressing multi-unit (and, more generally, multisource) events is described in broad terms in the project's Technical Analysis Approach Plan (NRC, 2013b). To inform the modeling of such events, it is a good idea to review past operational events to provide an indication of the likelihood and impact of these events, and of their salient features.

However, such a review, although straightforward in principle, can be extremely labor-intensive. The NRC receives thousands of Licensee Event Reports (LERs) each year, containing both structured and unstructured data. (Figure 9 provides the first page of an example LER.) Aids such as LERSearch,²⁷ search tools provided by the ADAMS system, and general search aids (e.g., indices for pdf files created using programs such as Adobe Acrobat) are helpful but are not tailored to address the multi-unit problem, and (in the case of LERSearch) do not provide access to a number of non-LER related documents that might be useful.

6.3.1 Use Case 1 Objective and Scope

The specific objective of this use case was to evaluate the effectiveness and efficiency of ICA 2.2 in helping users identify and characterize past U.S. operational events involving multiple reactors. The use case scope limitations were as follows:

- The project corpus was limited to the document types shown in Table 2.
- The focus was on events involving an “initiating event” (i.e., an event that perturbs the steady-state operation of a nuclear power plant and could lead to an undesired plant condition²⁸) at one or more units at a single site. The search did not exclude but was not aimed at identifying degraded conditions that could affect the response of multiple units at a site during an accident, or at identifying events/conditions affecting multiple sites.
- The events were characterized in terms of the event date, site involved, event extent, and event cause.

²⁷ LERSearch (<https://lersearch.inl.gov/LERSearchCriteria.aspx>) provides public access to the LERs and enables various searches of these records.

²⁸ See NUREG-2122 (Drouin et al., 2013c) for definitions of PRA terms.

NRC FORM 366 (1-2001)		U.S. NUCLEAR REGULATORY COMMISSION		APPROVED BY OMB NO. 3150-0104		EXPIRES 6-30-2001																																									
<p style="text-align: center;">LICENSEE EVENT REPORT (LER)</p> <p style="text-align: center; font-size: small;">(See reverse for required number of digits/characters for each block)</p>																																															
FACILITY NAME (1) Peach Bottom Atomic Power Station, Unit 2				DOCKET NUMBER (2) 05000 277		PAGE (3) 1 OF 4																																									
TITLE (4) Loss of Offsite Power Source Results in Specified System Actuation and Safety System Functional Failure.																																															
EVENT DATE (5)			LER NUMBER (6)			REPORT DATE (7)																																									
MO	DAY	YEAR	YEAR	SEQUENTIAL NUMBER	REV NO	MO	DAY	YEAR																																							
06	18	01	01	001	00	08	17	01																																							
OPERATING MODE (9) 1			THIS REPORT IS SUBMITTED PURSUANT TO THE REQUIREMENTS OF 10 CFR §: (Check all that apply) (11)																																												
POWER LEVEL (10) 100			<table border="1" style="width:100%; border-collapse: collapse;"> <tr> <td>20.2201(b)</td> <td>20.2203(a)(3)(ii)</td> <td>50.73(a)(2)(ii)(B)</td> <td>50.73(a)(2)(ix)(A)</td> </tr> <tr> <td>20.2201(d)</td> <td>20.2203(a)(4)</td> <td>50.73(a)(2)(iii)</td> <td>50.73(a)(2)(x)</td> </tr> <tr> <td>20.2203(a)(1)</td> <td>50.36(c)(1)(i)(A)</td> <td>x 50.73(a)(2)(iv)(A)</td> <td>73.71(a)(4)</td> </tr> <tr> <td>20.2203(a)(2)(i)</td> <td>50.36(c)(1)(ii)(A)</td> <td>50.73(a)(2)(v)(A)</td> <td>73.71(a)(5)</td> </tr> <tr> <td>20.2203(a)(2)(ii)</td> <td>50.36(c)(2)</td> <td>50.73(a)(2)(v)(B)</td> <td>OTHER</td> </tr> <tr> <td>20.2203(a)(2)(iii)</td> <td>50.46(a)(3)(ii)</td> <td>50.73(a)(2)(v)(C)</td> <td>Specify in Abstract below or in NRC Form 366A</td> </tr> <tr> <td>20.2203(a)(2)(iv)</td> <td>50.73(a)(2)(i)(A)</td> <td>x 50.73(a)(2)(v)(D)</td> <td></td> </tr> <tr> <td>20.2203(a)(2)(v)</td> <td>50.73(a)(2)(i)(B)</td> <td>x 50.73(a)(2)(vii)</td> <td></td> </tr> <tr> <td>20.2203(a)(2)(vi)</td> <td>50.73(a)(2)(i)(C)</td> <td>50.73(a)(2)(viii)(A)</td> <td></td> </tr> <tr> <td>20.2203(a)(3)(i)</td> <td>50.73(a)(2)(ii)(A)</td> <td>50.73(a)(2)(viii)(B)</td> <td></td> </tr> </table>					20.2201(b)	20.2203(a)(3)(ii)	50.73(a)(2)(ii)(B)	50.73(a)(2)(ix)(A)	20.2201(d)	20.2203(a)(4)	50.73(a)(2)(iii)	50.73(a)(2)(x)	20.2203(a)(1)	50.36(c)(1)(i)(A)	x 50.73(a)(2)(iv)(A)	73.71(a)(4)	20.2203(a)(2)(i)	50.36(c)(1)(ii)(A)	50.73(a)(2)(v)(A)	73.71(a)(5)	20.2203(a)(2)(ii)	50.36(c)(2)	50.73(a)(2)(v)(B)	OTHER	20.2203(a)(2)(iii)	50.46(a)(3)(ii)	50.73(a)(2)(v)(C)	Specify in Abstract below or in NRC Form 366A	20.2203(a)(2)(iv)	50.73(a)(2)(i)(A)	x 50.73(a)(2)(v)(D)		20.2203(a)(2)(v)	50.73(a)(2)(i)(B)	x 50.73(a)(2)(vii)		20.2203(a)(2)(vi)	50.73(a)(2)(i)(C)	50.73(a)(2)(viii)(A)		20.2203(a)(3)(i)	50.73(a)(2)(ii)(A)	50.73(a)(2)(viii)(B)	
20.2201(b)	20.2203(a)(3)(ii)	50.73(a)(2)(ii)(B)	50.73(a)(2)(ix)(A)																																												
20.2201(d)	20.2203(a)(4)	50.73(a)(2)(iii)	50.73(a)(2)(x)																																												
20.2203(a)(1)	50.36(c)(1)(i)(A)	x 50.73(a)(2)(iv)(A)	73.71(a)(4)																																												
20.2203(a)(2)(i)	50.36(c)(1)(ii)(A)	50.73(a)(2)(v)(A)	73.71(a)(5)																																												
20.2203(a)(2)(ii)	50.36(c)(2)	50.73(a)(2)(v)(B)	OTHER																																												
20.2203(a)(2)(iii)	50.46(a)(3)(ii)	50.73(a)(2)(v)(C)	Specify in Abstract below or in NRC Form 366A																																												
20.2203(a)(2)(iv)	50.73(a)(2)(i)(A)	x 50.73(a)(2)(v)(D)																																													
20.2203(a)(2)(v)	50.73(a)(2)(i)(B)	x 50.73(a)(2)(vii)																																													
20.2203(a)(2)(vi)	50.73(a)(2)(i)(C)	50.73(a)(2)(viii)(A)																																													
20.2203(a)(3)(i)	50.73(a)(2)(ii)(A)	50.73(a)(2)(viii)(B)																																													
LICENSEE CONTACT FOR THIS LER (12)																																															
NAME Steven C. Beck - Regulatory Assurance				TELEPHONE NUMBER (include Area Code) (717) 456-3243																																											
COMPLETE ONE LINE FOR EACH COMPONENT FAILURE DESCRIBED IN THIS REPORT (13)																																															
CAUSE	SYSTEM	COMPONENT	MANUFACTURER	REPORTABLE TO EPIX	CAUSE	SYSTEM	COMPONENT	MANUFACTURER	REPORTABLE TO EPIX																																						
SUPPLEMENTAL REPORT EXPECTED (14)					EXPECTED SUBMISSION DATE (15)																																										
YES (If yes, complete EXPECTED SUBMISSION DATE).					MONTH DAY YEAR																																										
X NO																																															
ABSTRACT (Limit to 1400 spaces, i.e., approximately 15 single-spaced typewritten lines) (16)																																															
<p>On June 18, 2001 at approximately 0315 hours, the 343SU-E offsite power source experienced an electrical fault causing it to trip. Two emergency 4160 VAC buses on both Units 2 and 3 transferred to their respective alternate power supply (2SU-E offsite power source). As a result, an invalid actuation of primary containment isolation valves in multiple systems occurred as expected for the loss of power.</p> <p>After the affected emergency buses transferred to the 2SU-E emergency power source, the operating crew entered procedure "Response to a Loss of #343 Off-site Startup Source." The procedure directed the crew to "flag" all "tripped" breaker control switches to their actual position; however, it did not address the alternate breaker control switches. This resulted in the normal power supply (tripped) breaker control switches being taken to open and the alternate power supply breaker control switch being left in the open position (with actual breaker position being closed). With both the normal and alternate breaker control switches in the open position, an interlock would have prevented the emergency diesel generator associated with two emergency 4160 VAC buses on each unit from automatically supplying power to their respective bus in the event of a loss of offsite power/LOCA. This condition existed for approximately three hours, when the operating crew took the control switch for the alternate power supply breaker to the closed position.</p>																																															

NRC FORM 366 (1-2001)

**Figure 9. Excerpt from example LER for a multi-unit event
(multi-unit fields highlighted)**

6.3.2 Use Case 1 Technical Challenges

On the surface, it might seem that a search for multi-unit initiating events should be straightforward. After all, surely a human analyst, upon reading an event summary, can readily determine whether that event involved initiating events at multiple units or not. However, there are an enormous number of event reports to review. (For example, the project's corpus contains nearly 55,000 LERs covering the period 1980–2014.) Furthermore, although determining whether an event involved multiple units is straightforward (see the highlighted text in Figure 9), the event descriptions often must be read carefully to determine whether the event involved an initiating event or a degraded condition (i.e., a situation that weakened the plant but did not actually involve an accident).

Computer tools, at least in principle, are well suited for addressing large numbers of documents. However, at least for text-based tools such as ICA 2.2, there are significant challenges in recognizing the significance of graphical elements (e.g., box lines around text to provide special emphasis) in an arbitrary document; taking advantage of the highlighted field in Figure 9 requires a non-trivial custom programming effort. Another challenge arises from the natural language used to describe events. As an example, Table 4 provides a sample of multi-unit events identified as precursors by the NRC's Accident Sequence Precursor (ASP) Program.²⁹ The last column in the table contains key phrases from the associated LERs indicating that the event involved initiating events at multiple units. Not only are the phrases non-standardized, sometimes the effects on different units are described in different places in the LER. Additional challenges to software tools include those discussed in Section 5.1 (e.g., flawed digitized data).

6.3.3 Use Case 1 Approach

The general approach for this use case followed the process described in Section 6.2. The use case team comprised three SMEs and two software engineers. Two of the SMEs had pre-project experience performing manual searches of LERs for multi-unit events and conditions, and helped the software engineers develop the use case-specific facets for the customized search application (constructed using ICA 2.2). The third SME, who had no formal experience searching for multi-unit events, conducted the final demonstration (Step 4, as described in Section 6.2).

²⁹ Per SECY-15-0124 (NRC, 2015b), which provides the status and results of the ASP program as of 2015, a nuclear power plant accident precursor is defined as an event with a conditional core damage probability (CCDP) or a change in core damage probability ("delta CDP" or ΔCDP) greater than or equal to 1×10^{-6} .

Table 4. Example multi-unit precursor events with indicative phrases

Date	Site	Type	LER(s)	Indicative Phrase(s)
6/22/82	Quad Cities	LOOP	254/82-012	Separated text, requires inference: “Unit Two reactor tripped” AND “Due to the degraded mode of the Unit One emergency AC power system, a Generating Station Emergency Plan Unusual Event was declared.” Could also infer from: “Unit 1/2 Diesel Generator tripped.”
8/11/83	Salem	LOOP	272/83-033, 272/83-034	Direct statements: “Both Salem units tripped,” “Salem Units 1 and 2 Reactor Trips.”
7/26/84	Susquehanna	SBO during test	388/84-013	Separated text, requires inference: “Unit 2 operating” AND “This resulted in a scram” AND “Unit 1 entered an LCO.”
5/17/85	Turkey Point	LOOP	251/85-011	Direct statement: “An Unusual Event was declared for both Units 3 and 4.”
7/23/87	Calvert Cliffs	LOOP	317/87-012	Direct statement: “Resulting in both reactors tripping on loss of load.”
3/20/90	Vogtle	LOOP	424/90-006, 425/90-002	Direct statement: “Tripped Unit 1 RAT A and Unit 2 RAT B.” Could also infer from: Unit 1 LER (424/90-006) “Further description of the Unit 2 response to this event is provided in LER 50-425/1990-002” OR Unit 2 LER (425/90-002) “See Licensee Event Report 50-424/1990-006 for a discussion of the resulting effect on Unit 1.”

The analysis was performed in two phases to exercise the customized search application in two different usage modes: informed search (in which very specific information is known about the target documents) and basic search (in which only general information is known about the target documents). In all cases, the demonstration was limited to events involving initiating events. This greatly reduced the number of LERs to be reviewed. (For example, of the 392 multi-unit LERs identified by Schroer and Modarres (2013) for the period 2000–2011, the large majority do not involve initiating events.)

Phase 1 – Informed Search. This phase, which involved searches of the corpus to find specific LERs for multi-unit events, was performed in two stages.

The first stage, which helped the SME conducting the final demonstration to become better acquainted with the use case-specific facets of the customized search application, was aimed at finding the LERs for a 2011 dual-unit loss of offsite power (LOOP) at the North Anna nuclear power plant (caused by an earthquake) and a 2011 three-unit LOOP at the Browns Ferry plant (caused by a tornado). The search process involved performing an initial search using selected facets and individual keywords. Progressive refinements of the search query, sometimes using additional user-supplied keywords to supplement the built-in keywords, eventually resulted in a manageable number of hits. At this point, a quick review of contextual text accompanying a

search hit (see Figure 10 for an example screen shot), or of the target documents, was usually sufficient to determine whether the hits represented desired search results.

The second stage involved a search of the corpus for the LERs for all of the multi-unit initiating events judged to be accident precursors by the NRC's ASP program (NRC, 2015b).³⁰ This stage used a user-constructed search query building on the keywords included in the customized search application, and taking advantage of the ICA 2.2 interface compatibility with standard word processors, which facilitated the construction of complex queries.

Phase 2 – Basic Search. This phase involved two separate searches for multi-unit initiating events that exercised the customized search application in a more exploratory mode, i.e., without prior knowledge regarding which specific events involved multiple units. The first search focused on the LERs in the project corpus. The second search focused on finding ASP-related SECY papers referring to multi-unit initiating events.³¹ The search was only aimed at identifying relevant SECY papers; the papers themselves typically provide the LER numbers for the events.

6.3.4 Use Case 1 Results

The general results of Use Case 1 are provided below. Additional details are provided by Siu et al. (2016b).

With respect to multi-unit event identification, when provided with highly discriminating information (e.g., unique characteristics, such as the occurrence of an earthquake or tornado, or specific event identifiers, such as LER number), the customized search application enabled effective and efficient searches. The search results were as complete as could be expected (search misses were caused by missing documents in the corpus rather than application deficiencies) and resulted in very few false positives. The application was easy to use and provided rapid responses (often within a few seconds) to queries.

When provided with less specific information, the searches were less successful; they only identified a small number of relevant events and also identified a fair number of false positives. Improved keyword lists better reflecting the variety of key terms used in the LERs would probably help, but more advanced programming (e.g., to draw inferences across widely separated text) is likely necessary to ensure that the searches are effective and efficient. Such additional effort was not judged to be necessary for the purposes of this feasibility study.

³⁰ Per Siu et al. (2016b), there were 27 such events from 1969–2015.

³¹ SECY papers are staff papers submitted to the NRC Commissioners to inform them about policy, rulemaking, and adjudicatory matters. More information can be found at <http://www.nrc.gov/reading-rm/doc-collections/commission/secys/>.

IBM Content Analytics with Enterprise Search

Collection: LTRP (change) | Preferences | My Profile | Help | About | Log Out

Help for query syntax

Query window. Currently shows search query based on phrases contained within subfacet "Facility by Name/Multi-Unit"

Query: ** AND subfacet: Facility by Name/Multi-Unit

Saved Searches (4) | Advanced Search | Query Tree

Documents | Facets | Time Series | Deviations | Trends | Facet Pairs | Connections

Results 1-100 of 3000 (41776/333512 results matched)

Facet Navigation | Default order

Filter: [] Clear

Facets

- Part of Speech
- Phrase Constituent
- Multi-Unit Events
- Event Date
- Facility by Name
 - Single Unit
 - Multi-Unit
- Cause
- Extent
- Core Damage Frequency
- Large Early Release Frequency
- Probabilistic Risk Assessment
- Corrective Actions
- ADAMS Docket Number
- ADAMS Author Affiliation
- Document Source
- ADAMS Document Type
- Flags

Search type: Subfacet search

Facet Path: /Facility by Name/Multi-Unit

Value: []

New search | Add to search | Search

Link to document "hit"

Instance of text "hit" within document

Contextual text for document "hit"

Document 1: ML13141A353.pdf

Document 2: ML13218A300.pdf

Document 3: 022200478.pdf

Document 4: XT101949.pdf

Document 5: C111347.pdf

Document 6: XT104362.pdf

Figure 10. Example ICA 2.2 screenshot showing customized search application interface

With respect to multi-unit event characterization, the customized search application provides a number of aids (principally highlighted contextual text) that help users identify event characteristics of interest (e.g., event date, facility name, and event extent). However, these aids were not helpful for all LERs; document download and review remained the surest approach to collecting the desired information. In this light, the primary value of the application was in identifying the best documents to download and review.

Two other tools available to NRC staff to identify and characterize multi-unit LERs are LERSearch and the pdf library search capabilities provided by Adobe Acrobat. LERSearch also proved extremely effective and efficient for simple searches. However, as compared with the customized search application, its advanced query capabilities are somewhat less powerful, its search space is restricted to LERs, and it lacks the ability to save searches. (This last point becomes especially important when refining a search query, and when performing multiple searches.) Adobe Acrobat searches of the library of LERs used in this project were slower than those of the customized search application or LERSearch, less flexible, and less helpful. (Even though contextual text is provided, users typically will need to download and review documents to identify targeted information.)

Overall, the customized search application appears to have potential for future use as an event-search tool. Even in its current feasibility demonstration state, it can support more efficient LER searches than those currently performed through LERSearch or general purpose search tools provided by ADAMS.³² With further development of its facets and keywords (see Siu et al., 2016b) and perhaps some custom programming (e.g., to take advantage of structured data, such as report tables), it might provide NRC users with an even more powerful search tool to address PRA-related information needs. However, such developments are likely to require non-trivial levels of effort; decisions regarding further work will need to consider overall agency needs.

6.4 Use Case 2

In the U.S., the adoption of a risk-informed approach is generally voluntary for regulatory applications involving operating reactors; there is no legal requirement for an operating plant licensee to have a PRA for its plants,³³ or to submit such a PRA (or its results) to the NRC for review. However, if a licensee chooses to adopt a risk-informed approach, then PRA results must be included as part of the submittal for regulatory approval. For example, if a licensee wishes to transition a plant's deterministic fire protection program to a risk-informed, performance-based

³² For this purpose, the restricted corpus of the project (with about 330,000 documents) is actually a benefit, as it reduces the search time required by the more generic ADAMS tools (which need to consider the entire ADAMS collection – over 2 million documents at the time of the project).

³³ As one partial exception, the calculation of the Mitigating Systems Performance Index (MSPI)—a risk-informed element of the NRC's Reactor Oversight Program—relies on licensee plant-specific, limited-scope PRA addressing events occurring during power operation. More information on the MSPI can be found at <http://www.nrc.gov/reading-rm/doc-collections/nuregs/staff/sr1816/>. Also note that plants licensed under 10 CFR 52 are required to have PRAs, although they are not required to submit these PRAs to the NRC.

program per the requirements of 10 CFR 50.48(c),³⁴ the licensee's license amendment request (LAR) must, among other things, provide results from its PRA supporting the acceptability of the transition request. Licensees applying for plant license renewals also typically submit PRA results in support of evaluations required for environmental assessments. The NRC staff's reviews of these evaluations, which include the PRA results, are documented in plant-specific supplements to NUREG-1437 (NRC, 2013c).

It can be seen that this voluntary approach to risk-informed applications causes the NRC to receive plant-specific risk information for operating plants on an irregular basis. Moreover, the information for the overall operating fleet is distributed across a variety of documents (including risk-informed LARs and license renewal requests), and because the plants and their PRAs typically change over time, the risk information for a given plant can vary from submittal to submittal.

To address these challenges, an analyst tasked with the development of a summary set of current PRA results for all plants must first identify the document containing the latest set of PRA results for each plant, and then find those results within the document.³⁵ In simple cases, the results are contained in a summary table somewhere within the document. In more difficult cases, the results are embedded in the document text. Thus, the analyst's task, while not conceptually difficult, can be quite labor intensive. (Recent performances have required several staff-days of effort.)

6.4.1 Use Case 2 Objective and Scope

The objective of this use case was to evaluate the ability of ICA 2.2 to help analysts efficiently identify documents containing the most recent risk information for operating plants. To limit staff and contractor resource requirements, and in keeping with the exploratory nature of LTRP projects, the following scope limitations were employed:

- The project corpus was limited to the document types shown in Table 2.
- The task was limited to the consideration of CDF (at-power operation, consideration of all hazards).
- The task was limited to information from four representative plants: Brunswick 1, Calvert Cliffs 1, Wolf Creek 1, and LaSalle 2.

³⁴ This rule is commonly cited as NFPA 805, the National Fire Protection Association (NFPA) standard endorsed by the rule (NFPA, 2001).

³⁵ For plants that have not undertaken any risk-informed application or requested license renewal, the most recent information available to the NRC may be from the Individual Plant Examination (IPE) and Individual Plant Examination of External Events (IPEEE) programs of the mid-1990s. Even for plants participating in risk-informed applications or license renewal, the plant PRAs may be limited to the treatment of internal events, so the most recent information on the risk from other hazards may be that developed for the IPEEE program.

The task focused on CDF because it is a widely used metric in current risk-informed applications, and because it is expected that lessons learned from a search for CDF will be relevant in searches for other risk metrics (e.g., large early release frequency [LERF]).

Brunswick and Calvert Cliffs were selected because they (1) have recent LARs to transition the plant’s deterministic fire protection program to a risk-informed, performance-based program per the requirements of 10 CFR 50.48(c); and (2) have also been approved for license renewal, as documented in appropriate plant-specific supplements to NUREG-1437. Both the NFPA 805 LARs and the NUREG-1437 supplements (e.g., NRC, 2006) contain relevant CDF information; the latter, which discuss the environmental impact of the license renewal, provide the CDF information in a discussion of potential severe accident mitigation alternatives (SAMAs). Wolf Creek and LaSalle were selected to test ICA 2.2 under more information-limited conditions. The former has a SAMA analysis but not an NFPA 805 analysis, and the latter has neither.

6.4.2 Use Case 2 Technical Challenges

For an automated search tool, a key challenge for this use case is the identification of documents containing pertinent information (e.g., the most recently estimated CDF for a given plant). The tool needs to:

- Recognize the variety of non-standardized phrases that refer to numerical estimates of the plant CDF (e.g., see Table 5) and
- Recognize that CDF estimates often appear in tables (e.g., see Figure 11).

Table 5. Examples of indicative text for plant CDF

Source	Indicative Text
NUREG-1437, Supplement 25 (NRC, 2006)	“The baseline core damage frequency (CDF) for the purpose of the SAMA evaluation is approximately 4.19×10^{-5} per year.”
LAR for Technical Specification Change (Exelon, 2005)	“The base CDF for the LSCS Unit 2 PRA is $6.64\text{E-}6/\text{yr} \dots$ ”
Evaluation of Integrated Leak Rate Test Extension (Johnson, 2015)	“...the total Internal Events Core Damage Frequency (CDF) = $1.61 \text{ E-}5/\text{year}$ for Unit 1 and CDF = $1.41 \text{ E-}5/\text{year}$ for Unit 2.”

Table 5-3. BSEP Core Damage Frequency for Internal Events

Initiating Event	CDF (per year)	Percent Contribution to CDF
Loss of offsite power (LOOP) – dual unit	1.47×10^{-5}	35.1
Turbine trip	1.14×10^{-5}	27.3
Main steam isolation valve (MSIV) closure/loss of condenser vacuum	4.78×10^{-6}	11.4
Loss of direct current (DC) panel	3.18×10^{-6}	7.6
Loss of alternating current (AC) emergency bus	2.39×10^{-6}	5.7
Loss of control rod drive (CRD)	1.72×10^{-6}	4.1
LOOP – single unit	1.01×10^{-6}	2.4
Other	1.01×10^{-6}	2.4
Internal floods	8.80×10^{-7}	2.1
Loss of reactor building closed cooling water (RBCCW)	4.60×10^{-7}	1.1
Interfacing systems loss of coolant accident (ISLOCA)/excessive LOCA	3.40×10^{-7}	0.8
Total CDF (internal events)	4.19×10^{-5}	100

Figure 11. Example table identifying plant CDF (NRC, 2006)

Regarding the appearance of data in tables, the tool needs to determine the table structure (which may be obvious visually but not obvious to a text-oriented tool), understand the meaning of the table structure (e.g., that the middle column of Figure 11 contains the CDF estimates), and understand the meaning of qualifiers (such as the “internal events” parenthetical in the last line of Figure 11).

For a semi-automated, human-in-the-loop tool such as ICA 2.2, the need to meet the above challenges is significantly reduced. However, to be efficient and practical, the tool needs to produce a relatively small number of hits (both documents and hits within a document) requiring manual review.

6.4.3 Use Case 2 Approach

The general approach followed a process similar to that for Use Case 1. The task team comprised one SME and two software engineers. The SME was a PRA analyst who had, prior to the project, performed manual searches of the ADAMS Main Library for information of interest

to the use case. The software engineers were the same individuals who had worked on Use Case 1.

Similar to Use Case 1, the SME used the customized search application (developed using ICA 2.2) to search the corpus and identify potential problems. Following discussions with the software engineers, the latter developed refinements for agreed-upon issues. The potential problems generally involved either a failure to identify corpus documents known to contain the desired information or an excessive number of “false positives” (i.e., documents that were identified by the tool but did not contain the desired information). The refinements ranged from complete changes in the search strategy³⁶ to the development of new facets, to modifications of the list of search phrases in a given facet. In some cases, it was determined that the corpus did not contain key documents, and the corpus was updated. Once the customized search application was finalized, it was exercised in two modes, informed and basic.

Phase 1 – Informed Search. In the informed mode, it was assumed that the user knows that:

- CDF information (for all hazards) is often provided in a plant’s risk-informed LAR (or associated documents) if the plant has applied for approval of a risk-informed application;
- The NFPA 805 LARs are fairly recent and should represent up-to-date information; and
- If a plant has not made a risk-informed submittal, or if the submittal does not address the CDF from all hazards, the SAMA analyses typically provide this information. (Although many of the analyses are dated, they are more up-to-date than the IPE and IPEEE analyses.)

Phase 2 – Basic Search. In the basic mode, it was assumed that the user does not know about the above sources of information and starts with a “blind search” of the database. (It was assumed that the user knows that documents containing information on total CDF are likely to have information on the contributions from specific hazards, including fire.)

As performed during the project, both search modes involved fairly complex queries regarding the term “core damage frequency” and its variants. In hindsight, such complexity was probably not necessary because corpus documents referring to core damage frequency almost invariably employ the acronym “CDF.” Thus, both the informed and basic searches for this use

³⁶ An early approach involved focusing on the exponential notation typically used in reporting CDFs. For example, recognizing that these CDFs are typically very small numbers, indicative tokens for a reported CDF of $1 \times 10^{-4}/\text{ry}$ could be the character strings “ 1×10^{-4} ,” “1E-4,” “1E-04,” “ 1.0×10^{-4} ,” “1.0E-4,” and so forth. However, since exponential notation is also widely used in non-PRA contexts, this approach yielded an excessive number of false positives and was not further pursued.

case could have been performed with very simple queries, and would likely have yielded the same results.³⁷

6.4.4 Use Case 2 Results

For both the informed and basic searches, the customized search application proved to be effective and efficient in identifying target documents containing the desired information (recent CDFs for a given set of plants). The application was easy to use and generally helped the user find the target documents in a short amount of time. With minor revisions to the facets (see Siu et al., 2016b), the tool should be useful to a broader range of staff.

Other general purpose search tools provided by the ADAMS system also proved effective and efficient in identifying the target documents. As compared with one internal NRC tool (ADAMS P8, scheduled to be phased out in the near future) and the ADAMS search tool available on the NRC's public website (<http://www.nrc.gov/>), the customized search application provided significant advantages through its advanced interface, which, among other things, facilitated the construction and saving of complex queries and provided highlighted contextual text that helped the user more rapidly determine the relevance of a particular hit. As compared with a more advanced internal NRC tool (ADAMS Enterprise Search), which employs an interface similar to that of ICA 2.2, the use case-specific facets developed for the customized search application were of some use, but may not have been necessary for the simple searches involved in this use case. (Customized facets were of greater use in Use Cases 1 and E.)

It should be emphasized that the Use Case 2 results are based on a search process that takes advantage of the recency of CDF results developed for NFPA 805 applications and the standardized reporting of these results. (Later NFPA 805 LARs report CDF contributions from various hazards in a standard table in a standard section of the LAR. This consistent approach makes it easy for a user to rapidly review a search-identified document to see if it contains the desired CDF information.) More general CDF searches may need to consider a wider range of phrases and reporting formats.

6.5 Use Case E

The questions behind Use Cases 1 and 2 ("What are some key multi-unit events worth further examination?" "What are the current CDFs for U.S. plants?") are fairly specific and can be answered using direct (and often quite simple) search queries. ICA 2.2 is a useful tool for developing applications to perform such searches, but it is primarily designed to support more

³⁷ In some PRA-relevant documents, the acronym "CDF" can also stand for "cumulative distribution function." Thus, it is possible that the search strategy used was actually helpful in avoiding related false positives. However, we did not investigate this matter.

complex, open-ended explorations of available data.³⁸ A limited-scope use case was performed to provide a quick look at the capabilities of ICA 2.2 in this latter mode.

6.5.1 Use Case E Objective and Scope

The objective of this use case was to provide insights regarding the discovery/exploration capabilities of ICA 2.2. To limit the time and resource requirements, the following scope limitations were employed:

- The use case employed the same project corpus and customized search application developed for Use Cases 1 and 2; no additional modifications were made to the corpus or the application to support this use case, even if a particular search led to inconclusive results.
- The use case involved the exploration of a small number of topics (described below).

6.5.2 Use Case E Challenges

The principal challenge for this use case was user- rather than technology-oriented. It required that the application user change his mindset from “searching” to “exploring.” Since the latter mindset is not strongly aligned with typical staff tasks and, by extension, with typical staff uses of available data, this challenge proved more difficult than might be expected.³⁹ The approach described in the following section can be viewed as a compromise: it addresses a broader question than is typically addressed in staff activities, but it isn’t completely open-ended.

A related secondary challenge involved determining how to use the analytics tools provided by ICA 2.2 to facilitate the exploration process. For example, a key question is how to generate information that suggests where to look next (as opposed to information that helps narrow a search for a particular document or particular document content).

6.5.3 Use Case E Approach

This use case involved a single SME. The broad question addressed was “What do the data in the corpus tell us about the following topics?”

³⁸ Similar to browsing the stacks of a technical library or “surfing” the website of an organization, the specific endpoint of a content analytics-guided exploration of a database may not be known at the beginning of the activity. The goal is to extract useful insights from the mass of available information, rather than to obtain the answer to a specific predefined question.

³⁹ To some extent, this challenge played a role throughout the project, as the SMEs did not fully appreciate the principal focus of ICA 2.2 or how it worked for quite a while. See Section 6.6 for further discussion.

- External events
 - Reported events
 - Analyses
- Ice storms at or near the Vogtle Nuclear Power Plant⁴⁰
- PRA-relevant content in NRC Inspection Reports

The topics were not developed through any systematic process, but reflect questions of potential interest to PRA analysts.

For each topic, the SME used the facets in the customized search application, sometimes in combination with additional keywords, to perform an initial search. Using the results of this search (principally hits and facet keyword frequency counts, but also more advanced analytics, such as correlations between facets), follow-up search questions (perhaps exploration of sub-topics) were identified and pursued. As might be expected, the queries sometimes led to a large number of hits whose relevance could only be determined through document download and review. Given the scoping nature of this use case, the exploration was generally terminated at this point.

For some topics, upon completion of the exploration process, a number of the analytics features of ICA 2.2 (e.g., providing trending or strength-of-relationship information) were applied to see what insights (e.g., confirmation of current understanding, surprises) or suggestions for further exploration they might provide.

6.5.4 Use Case E Results

Each of the topic explorations led to observations that supported prior understandings (i.e., were not surprising); provided interesting information and even surprises, suggesting areas for follow-up; provided seemingly spurious correlations requiring further exploration; or provided indications suggesting potential areas for improvement in the customized search application.

Some of the confirmatory observations included the following:

- LERs generally do not include quantitative CDF estimates. (LER references to CDF are generally high-level and qualitative, typically indicating that an analysis of CDF was performed to support assessments of the significance of the event.)
- Only a small number of LERs (10) involve ice storms.
- A number of inspection reports indicate that PRA has been used to determine the importance of inspection findings.
- There is a strong analytic relationship between references to the Turkey Point plant,

⁴⁰ This subject was picked as an exploratory topic that may be of interest to the NRC's Level 3 PRA project (NRC, 2011), recognizing that the project involves the Vogtle plant, that severe ice storms can occur in the Southeast (where the plant is located), and that such storms are not typically addressed in detail in current PRAs.

which is located in Florida, and references to hurricanes.

The more interesting observations included the following:

- The search revealed an unexpected path to documented staff analyses of some operational events.⁴¹
- There are more LERs with seismic- and flooding-related keywords than LERs with keywords related to high winds.
- The annual number of external hazard-related LERs exhibits a major discontinuity in 1988 (see Figure 12).⁴²
- The corpus includes a number of references to several licensee full-scope Level 3 PRAs. (These studies might provide useful suggestions and context for the ongoing NRC study.)
- A number of inspection report references to ice storms (for plants in the Southeast) have not been captured by LERs. Although these storms presumably did not meet mandatory reporting criteria, some of them caused the loss of emergency sirens, a potentially important event for a Level 3 PRA.
- Some of the inspection reports identify manual actions determined to be important in the plant PRA.

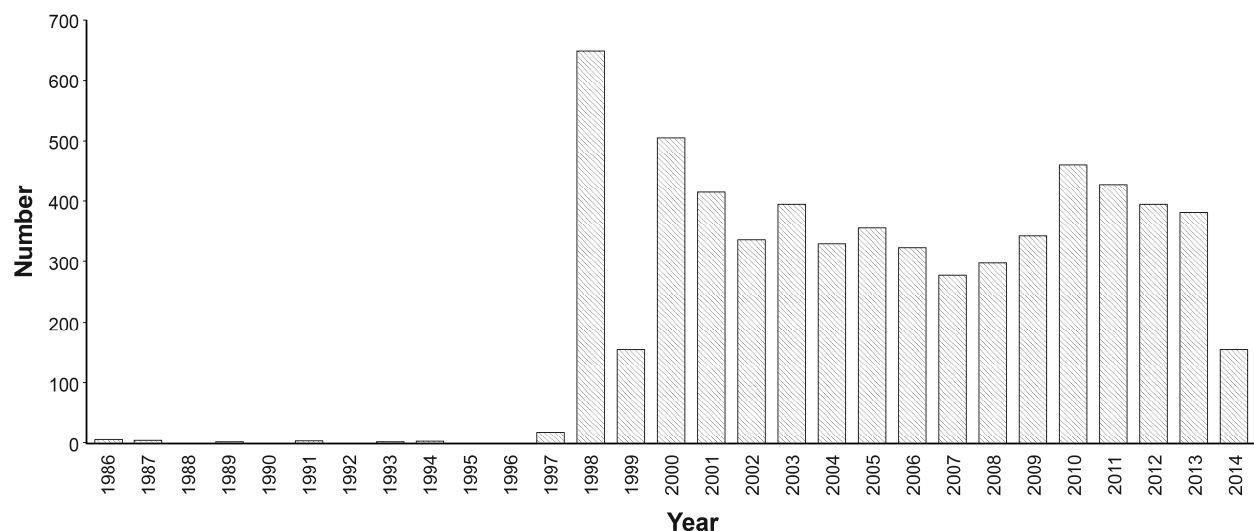


Figure 12. LERs over time

⁴¹ As a matter of standard NRC procedure, these analyses are uploaded into ADAMS upon finalization. However, staff members not involved in performing the analyses may not know where to look to find the documents.

⁴² A quick check indicates that this observation applies to all LERs, not just those associated with external hazards. Further exploration indicates that the discontinuity might be at least partially explained by the January 1988 publication of NUREG-1022, Rev. 1, which clarified the 30-day reporting requirement for LERs (NRC, 1988). Note that the plot only includes LERs for which the event date could be determined from the associated pdf file. The discontinuity in 1988 may be due to changes in LER format, which might have affected the performance of the routine, as well as changes in actual reporting.

The seemingly spurious correlations included the following:

- A large number of hits involving documents referring to storm surge and fire protection. Further investigation (involving document download and review) showed these largely arose from one plant's Final Safety Analysis Report (FSAR).
- A number of hits involving documents referring to ice storms and the Vogtle plant. Further investigation showed that in many of the documents, the references to ice storms and to Vogtle appeared in independent discussions.

The indications for potential areas of improvement included the following:

- A large number of hits associated with surge tanks and surge lines when searching for operating events involving external hazards. This suggests a refinement of the keywords used in the customized tool, where "surge" could also be associated with a flood.
- Separate analytics for keywords that differed only in capitalization (e.g., "seismic" vs. "Seismic"). Grouping the results for such keywords would lead to a more accurate understanding of the data.
- Many hits referring to the summer season when searching for events involving the Virgil C. Summer plant. Minor improvements in the keywords identifying plant names should easily address this problem.

With respect to the ability of the customized search application (and ICA 2.2) to support database explorations, the interface was easy to use, the response to queries was suitably quick, a number of the facets were helpful (even though they were developed specifically to support Use Cases 1 and 2), and, as in the previous use cases, the contextual text provided with search results was quite useful. Of the content analytics provided, keyword frequency counts and time series data plots were helpful. Other analytics (e.g., for indicating the strength of relationships between keywords) might be useful with further development.

6.6 Scoping Study Conclusions and Commentary

6.6.1 Conclusions

The LTRP scoping study employed three case studies ("use cases") to investigate the ability of a particular content analytics tool, ICA 2.2 (customized with problem-specific facets), to support searches and database explorations of interest to PRA and risk-informed decision making. The following conclusions are based on the results and observations from the use cases.

- The customized search application is generally effective and efficient in identifying target documents of interest to the use cases. In the one test situation in which the tool was not effective (a basic, uninformed search for LERs involving multi-unit events), additional refinements (particularly updating the tool facets) would likely improve its performance.
- The application is capable of supporting more open-ended explorations of the database that lead to potentially interesting insights and suggest avenues for further exploration.
- The human-in-the-loop, stepwise search approach underlying ICA 2.2 is comfortable to use, at least for the corpus and use cases tested. Feedback from queries is quick (typically on the order of a few seconds) and informative, and document downloads (when more detailed information is needed) are also quick.
- The initial development and subsequent refinement of a useful application requires extensive interactions between the SMEs and the software engineers to ensure mutual understanding of (1) the technical problem(s) targeted by the tool, (2) examples of a successful search, and (3) the objectives and capabilities of the tool.
- Although the customized application was developed only to support this LTRP project's technology evaluation, it appears to be capable of assisting staff interested in extracting PRA-relevant lessons from operational experience documents.
 - As compared with LERSearch (the current staff tool of choice), the ICA 2.2 interface provides additional capabilities (e.g., supporting the development of complex searches, the saving of these searches, and the rapid screening of search results through contextual text). The ICA 2.2 tool also provides access to potentially useful documents beyond LERs.
 - As compared with more general ADAMS-based tools (e.g., ADAMS Enterprise Search), the reduced size and pre-indexing of the project corpus leads to significantly more rapid searches.
- Further work, perhaps requiring major programming effort or even technology development, could significantly increase the application's power and ease of use. This includes work to take advantage of data structures in technical documents, including document sections, structures within text passages (e.g., subordinate clauses), and tables.

6.6.2 Additional Observations

The following observations, derived from the experiences of the scoping project, should be useful when developing KE solutions:

- In general, problems with database documents (e.g., due to errors in the documents, OCR faults, or faulty document profiling) can hinder text-based searches by any tool. In many cases, the keywords of interest occur multiple times within a document, so database problems may not significantly affect search results. However, cases can arise (e.g., when searching for a document with a specific identifier) in which such problems are critical. If

it is important that the search identify all documents matching a specific query, considerable effort may be needed to ensure that potential errors in the documents are identified and handled by the tool.

- The willingness of users to pursue searches (or explorations) using any tool depends on, among other things, the time required to obtain informative feedback for each query. To help ensure rapid yet helpful feedback, it may be useful to:
 - Focus applications on problems that can be addressed with a smaller corpus; and
 - Provide users with tips for developing queries that generate quicker responses.
- For ICA 2.2 and similar tools, document download and review is an integral part of the search process. Download by hyperlink is straightforward. However, the review portion can be resource-intensive. For Use Case 1, the review was aided by the title and summary sections of LERs. For Use Case 2, the review was aided by LARs that provided standardized tables of CDF information in standard document sections. Thus, although ICA 2.2 has been developed to deal with unstructured data, the overall search process benefits from structured data.

6.6.3 Commentary – On Oracles Versus Aides

At the beginning of the LTRP project, encouraged by the implications of the IBM Watson “Jeopardy!” demonstration and the natural language capabilities of personal assistant software, the project SMEs hoped that ICA 2.2 would be able to provide direct answers to such natural language questions as “What are some key multi-unit events worth further examination?” (Use Case 1) or “What is the CDF for Plant X?” (Use Case 2). As the project progressed, it became clear that ICA 2.2 is not targeted at this kind of problem.

First, as discussed in Section 6, ICA 2.2 is largely intended to support database exploration. When employed in a direct question/answer mode, it can generate applications that produce informative intermediate results (e.g., which LERs involving multiple units are referenced in ASP SECY papers) and potentially useful statistics (e.g., how many documents include references to total CDF). However, in general, the user must review contextual text or review linked documents to answer a posed question. Furthermore, given the natural language variations in source documents (e.g., see Tables 4 and 5 and Figure 9), significant effort (well beyond that employed in this technology evaluation project) is necessary to ensure that the search results are reasonably complete (without including an excessive number of false positives).

Second, and related to the point above, ICA 2.2 is designed as a human-in-the-loop tool. Thus, in search mode, the tool does not function as an oracle that provides final answers to a user’s questions. Rather, it acts as an aide, providing (1) information that suggests, as the search progresses, the next steps a user might take to refine a search, and then (2) hyperlinks that help the user download and review documents that might contain the answers.

Due to the limited scope of this project, we do not have any empirical data relevant to the current effectiveness and efficiency of commercial, off-the-shelf software to (after appropriate

customization) directly answer questions of the sort underlying Use Cases 1 and 2.⁴³ However, given the complexities revealed in the two use cases, it appears likely that the development of an industrial grade, fully automated solution will require considerable SME and software engineer involvement. Moreover, by not involving the SME as an integral part of the actual search process, such a solution:

- May not take full advantage of SME skills (e.g., recognizing words and numbers despite faulty OCR or faulty entry of metadata—titles, authors, dates, etc.—characterizing documents stored in databases, recognizing the data relationships implied by a tabular structure) and knowledge (e.g., to recognize apparent conflicts between documents);
- May generate results not fully trusted by the SME; and
- Will minimize the learning benefits associated with formulating and refining a search (including learning from efforts to develop a search strategy, lessons from “failed searches,” and useful information and insights from intermediate search results).

In addition to the fully automated (“oracle”) versus human-in-the-loop (“aide”) issue, KE solution developers need to consider whether the emphasis is on:

- Providing a partner (which collaborates with the user to build knowledge, and may even alert the user when items of interest, such as the Blayais event, arise⁴⁴) or a servant (which only responds to requests);
- Supporting open-ended exploration or answering specific factual questions; and
- Developing broad base understanding by encouraging user “play,” or answering immediate, task-oriented needs.

A notional representation of how current technologies appear to be approaching these considerations is shown in Figure 13.

⁴³ Keim (2015) provides an interesting discussion of the status of and challenges faced by Watson (and other artificial intelligence tools) in the medical field.

⁴⁴ Note that such a “watchdog” application would need to notice and prioritize the event prior to alerting the user. Furthermore, it might need to ensure it collects and organizes enough information to overcome user preconceptions.

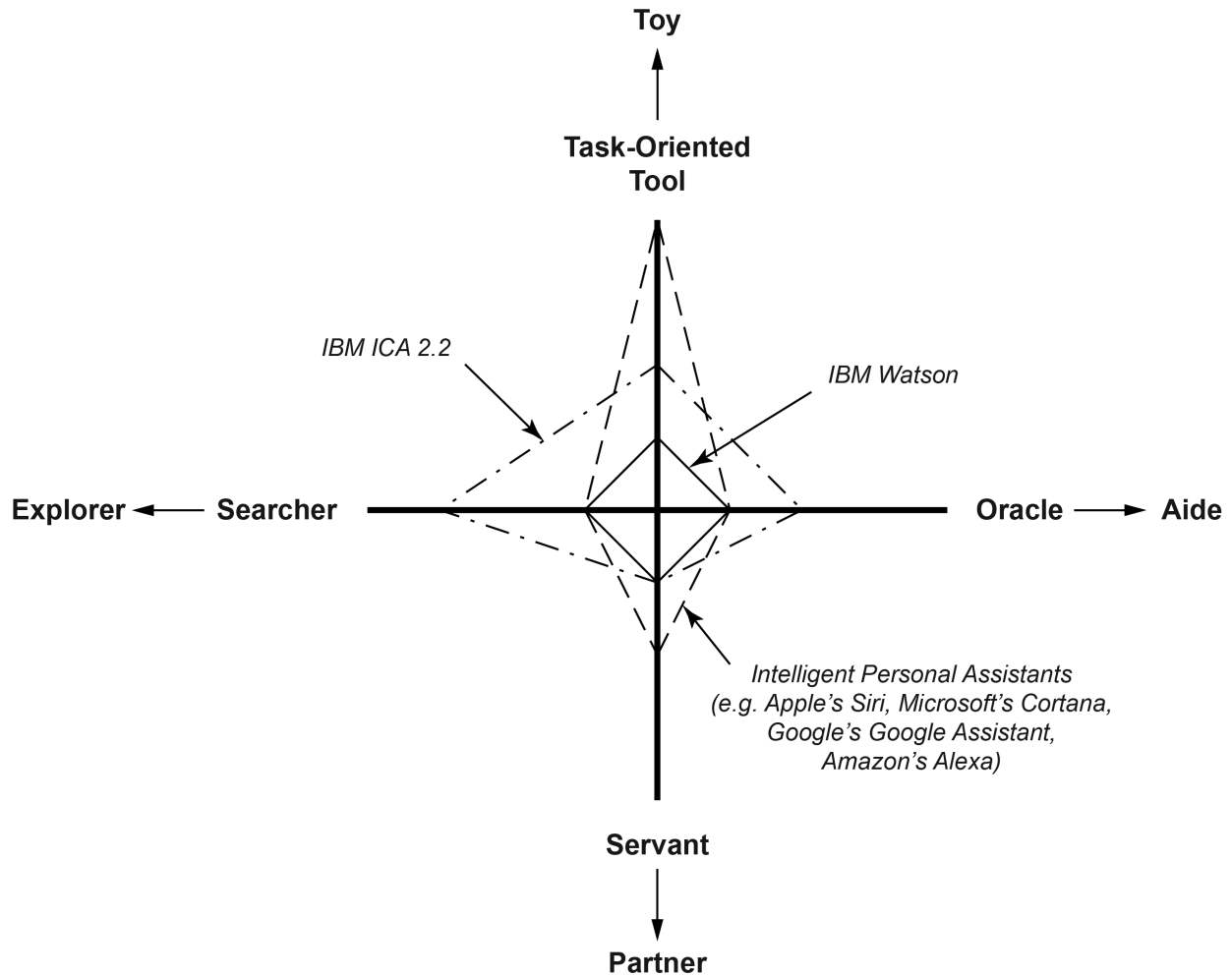


Figure 13. Notional representation of different KE solutions

For such organizations as the NRC, near-term efforts are likely to be aimed at near-term, highly focused and pragmatic developments. However, it is important to recognize that a broad staff knowledge base is important for flexible and agile agency operations, and non-traditional KM approaches (with associated KE solutions) may be helpful in developing it.

7 CONCLUSIONS AND SUGGESTIONS FOR FUTURE DEVELOPMENTS

In this paper, we have shown the following:

- The NRC uses information to support risk-informed decision making in a wide variety of applications. The breadth of applications and the inherent breadth of considerations involved in risk-informed decision making imply a wide variety of informational needs.

- The special characteristics of risk information, including the information supporting PRAs and the information resulting from PRAs, pose special challenges to KE activities supporting the creation, management, retrieval, and use of risk information.
- Advanced KE technologies are evolving to support the growing needs of organizations relying on massive amounts of unstructured information. Currently available commercial tools based on these technologies are sufficiently capable of supplementing tools used by NRC staff in risk-relevant activities.
- Additional efforts from the user and developer communities are likely to result in improved tools for the staff.
 - In the near term, useful work could involve improving the electronic database (e.g., through the digitization of legacy documents and the correction of faulty digitized records) and the development of more efficient and effective query structures (e.g., facets in the case of ICA 2.2) for targeted staff tasks (e.g., those represented by Use Cases 1 and 2). The organizational resources and commitment required for such work should not be underestimated.
 - Somewhat longer-term work, which would require significant programming effort at least in the case of ICA 2.2,⁴⁵ could involve the development of software tools that can take advantage of internal document structures (e.g., document sections, structures within text passages, and tables).
 - In the long term, work to develop (1) tools that create explicit risk information from implicit information (i.e., tools that “connect the dots” between declarations) and (2) a “watchdog” application that alerts users when noteworthy information (e.g., regarding a potential accident precursor, per the Blayais example) becomes available could be valuable⁴⁶.

It is important to recognize that there are many communities actively engaged in improving the access to and use of information. These include communities concerned with artificial intelligence and expert systems, natural language processing, analytics, “big data,” library science, and education, as well as KM. Involving these communities in future discussions regarding risk information will not only help avoid unnecessary duplication of efforts, it will also add a breadth of views that could improve the tools provided to users.

Finally, it is also important to recognize that KE solutions are only part of the KM toolbox. There are, as discussed in this paper, many non-technical approaches for enhancing staff awareness of and access to important information. The prioritization of activities to develop and implement improved KE technology will need to consider the full range of potentially viable approaches for addressing the staff’s needs.

⁴⁵ Note that our work has been limited to the use of ICA 2.2. Given the current state of KE technology, it is possible that the desired tools are available in other software products.

⁴⁶ Such a watchdog function could be viewed as an extension of existing commercial products (e.g., Google Alerts (<https://www.google.com/alerts>) that leverages knowledge management technology.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the LTRP project support provided by M. Tobin, S. Dennis, P. Appignani, G. Young, S. Raimist, and K. Bojja; the information provided by G. Georgescu, C. Pfefferkorn, and A. D’Agostino; and the editing support provided by C. Siu.

ACRONYMS AND ABBREVIATIONS

ACRS	Advisory Committee on Reactor Safeguards (NRC)
ADAMS	Agencywide Document Access and Management System (NRC)
AERB	Atomic Energy Regulatory Board (India)
ANS	American Nuclear Society
ASME	American Society of Mechanical Engineers
ASP	accident sequence precursor
CDF	core damage frequency
CFR	U.S. Code of Federal Regulations
CSN	Consejo de Seguridad Nuclear (Spain)
EDG	emergency diesel generator
FSAR	Final Safety Analysis Report
FY	fiscal year
GAO	U.S. Government Accountability Office
IAEA	International Atomic Energy Agency
ICA 2.2	IBM Content Analytics Version 2.2
IBM	International Business Machines
IPE	Individual Plant Examination
IPEEE	Individual Plant Examination of External Events
IPSN	Institut de Protection et de Sûreté Nucléaire (France)
KM	knowledge management
LAR	license amendment request
LER	licensee event report
LHSI	low head safety injection
LOOP	loss of offsite power
LTRP	Long-Term Research Program (NRC)
MSPI	Mitigating Systems Performance Index
NAP	National Academy Press
NFPA	National Fire Protection Association
NRC	U.S. Nuclear Regulatory Commission
NUREG	designator for reports issued by the NRC

NUREG/CR	designator for contractor-developed reports issued by the NRC
OCR	optical character recognition
PRA	probabilistic risk assessment
PSA	probabilistic safety assessment
PSAM	probabilistic safety assessment and management
RG	Regulatory Guide (NRC)
ry	reactor year
SAMA	severe accident mitigation alternative
SECY	designator for NRC staff papers addressed to the Commission
SME	subject matter expert
SPAR	Standardized Plant Analysis Risk
SRM	Staff Requirements Memorandum (NRC)
SRP	Standard Review Plan (NRC)

REFERENCES

Note: ADAMS document accession numbers (starting with the designator “ML”) can be used to obtain the associated documents from the NRC ADAMS Public Documents System (see <http://www.nrc.gov/reading-rm/adams.html>). Similarly, NRC reports (with the NUREG designator) can be obtained from the NRC website (see <http://www.nrc.gov/reading-rm/doc-collections/>).

- AERB (2005) “AERB Annual Report for the Year 2004-2005,” Atomic Energy Regulatory Board, Mumbai, India.
- Apostolakis, G., et al. (2012) “A Proposed Risk Management Regulatory Framework,” *NUREG-2150*, U.S. Nuclear Regulatory Commission, Washington, DC, USA.
- Meléndez Asensio, E. and R. H. Santos (2015) “Use of PSA model XML Standard Formats for V&V,” *Proceedings of ANS PSA 2015 International Topical Meeting on Probabilistic Safety Assessment and Analysis, Sun Valley, ID, April 26-30*.
- ASME/ANS (2009) “Standard for Level 1/Large Early Release Frequency Probabilistic Risk Assessment for Nuclear Power Plant Applications,” *ASME/ANS RA-Sa-2009, Addendum A to RA-S-2008*, ASME, New York, NY, American Nuclear Society, La Grange Park, Illinois.
- Chen, J. T., et al. (1991) “Procedural and Submittal Guidance for the Individual Plant Examination of External Events (IPEEE) for Severe Accident Vulnerabilities,” *NUREG-1407*, U.S. Nuclear Regulatory Commission, Washington, DC, USA.
- Drouin, M., et al. (2013) “Glossary of Risk-Related Terms in Support of Risk-Informed Decisionmaking,” *NUREG-2122*, U.S. Nuclear Regulatory Commission, Washington, DC, USA.

- Dupuy, P., G. Georgescu, and F. Corenwinder (2014) “Treatment of the loss of ultimate heat sink initiating events in the IRSN Level 1 PSA,” *NEA/CSNI/R(2014)9, Probabilistic Safety Assessment (PSA) of Natural External Hazards Including Earthquakes: Workshop Proceedings, Prague, Czech Republic, June 17-20, 2013*, Nuclear Energy Agency, Boulogne-Billancourt, France.
- Epstein, W. and A. Rauzy (2013) “New Developments in Open PSA,” *Proceedings of ANS PSA 2013 International Topical Meeting on Probabilistic Safety Assessment and Analysis, Columbia, SC, September 22-26*.
- Exelon (2005) “Request for a License Amendment to Extend the Completion Times Related to Technical Specifications associated with Residual Heat Removal Service Water, Diesel Generator Cooling Water and the Opposite Unit Division 2 Diesel Generator,” J.A. Bauer, Exelon Generation, letter to U.S. Nuclear Regulatory Commission, April 13, 2005. (ADAMS ML051040149)
- Ferrucci, D., et al. (2010) “Building Watson: an overview of the DeepQA Project,” *AI Magazine*, 31, No. 3, pp. 59-79.
- Fleming, K. (2005) “On the issue of integrated risk – a PRA practitioners perspective,” *Proceedings of ANS. International Topical Meeting on Probabilistic Safety Analysis (PSA ’05), San Francisco, CA, September 11-15*.
- Friedlhuber, T., M. Hibti, and A. Rauzy (2015) “A Method to Compare PSA Models in a Modular PSA,” *Proceedings of ANS PSA 2015 International Topical Meeting on Probabilistic Safety Assessment and Analysis, Sun Valley, ID, April 26-30*.
- GAO (2012) “Nuclear Regulatory Commission: Natural Hazard Assessments Could Be More Risk-Informed,” *GAO-12-465*, U.S. Government Accountability Office, Washington, DC, USA.
- Gorbachev, A., et al. (2000) “Report on flooding of Le Blayais power plant on 27 December 1999,” *Proceedings of EUROSAFE 2000, Cologne, Germany, November 6-7*, Gesellschaft für Anlagen- und Reaktorsicherheit (GRS) GmbH, Cologne, Germany.
- Hudson, J. L., et al. (2014) “A Model of Effective Governance for Knowledge Management: A Case Study at the U.S. Nuclear Regulatory Commission,” IAEA CN-220-343, *Proceedings of International Conference on Human Resource Development for Nuclear Power Programmes: Building and Sustaining Capacity, Vienna, Austria, May 12-16*, International Atomic Energy Agency, Vienna, Austria. (ADAMS ML14300A476)
- Johnson, M. (2015) “Calvert Cliffs Nuclear Power Plant: Evaluation of Risk Significance of Permanent ILRT Extension,” Jensen Hughes, February 17, 2015. (ADAMS ML15051A410)
- Keim, B. (2015) “IBM’s Dr. Watson Will See You ... Someday,” *IEEE Spectrum*, May 29, 2015.
- Kuritzky, A., et al. (2013) “L3PRA: Updating NRC’s Level 3 PRA insights and capabilities,” *Proceedings IAEA Technical Meeting on Level 3 Probabilistic Safety Assessment, Vienna, Austria, July 2-6, 2012*, International Atomic Energy Agency, Vienna, Austria. (ADAMS ML12173A092)

- IAEA (2000) “Measures to Strengthen International Co-Operation in Nuclear, Radiation and Waste Safety including Nuclear Safety Review for the Year 1999,” *IAEA General Conference*, International Atomic Energy Agency, Vienna, Austria.
- IPSN (2000) “Rapport Sur L’Inondation Du Site Du Blayais,” Institut de Protection et de Sûreté Nucléaire, Fontenay-aux-Roses, France.
- Kapan and Garrick (1981) “On the quantitative definition of risk,” *Risk Analysis*, **1**, pp. 11-37.
- Lanore, J.-M., et al. (2013) “IRSN analysis of post-Fukushima ‘Hardened Safety Core’: use of PSA insights,” *Proceedings of PSAM Topical Conference in Light of the Fukushima Dai-Ichi Accident, Tokyo, Japan, April 15-17*.
- NAP (1996) “World-Class Research and Development Characteristics for an Army Research, Development and Engineering Organization,” National Academy Press, Washington, DC, USA.
- NFPA (2001) “Performance-Based Standard for Fire Protection for Light Water Reactor Electric Generating Plants, 2001 Edition” *NFPA 805*, National Fire Protection Association, Quincy, MA, USA.
- Nowlen, S. P., M. Kazarians, and F. Wyant (2001), “Risk Methods Insights Gained from Fire Incidents,” *NUREG/CR-6738*, U.S. Nuclear Regulatory Commission, Washington, DC, USA.
- NRC (1988) “Event Reporting Guidelines: 10 CFR 50.72 and 50.73,” *NUREG-1022, Rev. 1*, U.S. Nuclear Regulatory Commission, Washington, DC, USA.
- NRC (1995) “Use of Probabilistic Risk Assessment Methods in Nuclear Activities: Final Policy Statement,” *Federal Register*, **60**, pp. 42622-42629, August 16.
- NRC (1998) “An Approach for Using Probabilistic Risk Assessment in Risk-Informed Decisions on Plant-Specific Changes to the Licensing Basis,” *Regulatory Guide RG 1.174*, U.S. Nuclear Regulatory Commission, Washington, DC, USA.
- NRC (1999) “Staff Requirements - SECY-98-144 - White Paper on Risk-Informed and Performance-Based Regulation,” *Memorandum from A. L. Cook to W. D. Travers, Revised March 1, 1999*, U.S. Nuclear Regulatory Commission, Washington, DC, USA. (ADAMS ML003753601)
- NRC (2002) “Perspectives Gained From the Individual Plant Examination of External Events (IPEEE) Program,” *NUREG-1742*, U.S. Nuclear Regulatory Commission, Washington, DC, USA.
- NRC (2006) “Generic Environmental Impact Statement for License Renewal of Nuclear Plants, Supplement 25 Regarding Brunswick Steam Electric Plant, Units 1 and 2, Final Report,” *NUREG-1437, Supplement 25*, U.S. Nuclear Regulatory Commission, Washington, DC, USA.
- NRC (2009) “An Approach for Determining the Technical Adequacy of Probabilistic Risk Assessment Results for Risk-Informed Activities,” *Regulatory Guide RG 1.200 Rev. 2*, U.S. Nuclear Regulatory Commission, Washington, DC, USA.

- NRC (2011) “Options for Proceeding with Future Level 3 Probabilistic Risk Assessment (PRA) Activities,” *SECY-11-0089*, U.S. Nuclear Regulatory Commission, Washington, DC, USA. (ADAMS ML11090A039)
- NRC (2012) “Office of Nuclear Regulatory Research Long-Term Research Program,” *NUREG/BR-0506*, U.S. Nuclear Regulatory Commission, Washington, DC, USA.
- NRC (2013a) “Advanced Knowledge Engineering Tools to Support Risk-Informed Decision Making,” *Solicitation RES-13, July 1*, U.S. Nuclear Regulatory Commission, Washington, DC, USA. (Available from www.fbo.gov)
- NRC (2013b) “Technical Analysis Approach Plan for Level 3 PRA Project, Rev 0b,” U.S. Nuclear Regulatory Commission, Washington, DC, USA. (ADAMS ML13296A064)
- NRC (2013c) “Generic Environmental Impact Statement for License Renewal of Nuclear Plants,” *NUREG-1437*, Rev. 1, U.S. Nuclear Regulatory Commission, Washington, DC, USA.
- NRC (2014) “Strategic Plan, Fiscal Years 2014-2018,” *NUREG-1614 Vol. 6*, U.S. Nuclear Regulatory Commission, Washington, DC, USA.
- NRC (2015a) “Project Aim and Centers of Expertise,” *SECY-15-0143*, U.S. Nuclear Regulatory Commission, Washington, DC, USA. (ADAMS ML15292A249)
- NRC (2015b) “Status of the Accident Sequence Precursor Program and the Standardized Plant Analysis Risk Models,” *SECY-15-0124*, U.S. Nuclear Regulatory Commission, Washington, DC, USA. (ADAMS ML15187A434)
- NRC (2016a) “Staff Requirements – COMSECY-16-0006 – Revision to the Agency’s Long-Term Research Program and Related Reporting to the Commission,” *Memorandum from A. L. Vietti-Cook to V. M. McCree*, U.S. Nuclear Regulatory Commission, Washington, DC, USA. (ADAMS ML16104A299)
- NRC (2016b) “Three Mile Island Accident of 1979 Knowledge Management Digest,” *NUREG/KM-0001*, Rev. 1, U.S. Nuclear Regulatory Commission, Washington DC, USA.
- NRC (2016c) “WASH-1400, The Reactor Safety Study: The Introduction of Risk Assessment to the Regulation of Nuclear Reactors,” *NUREG/KM-0010*, U.S. Nuclear Regulatory Commission, Washington DC, USA.
- NRC (2016d) “A Compendium of Spent Fuel Transportation Package Response Analyses to Severe Fire Accident Scenarios, Draft Report for Comment,” *NUREG/CR-7209*, U.S. Nuclear Regulatory Commission, Washington DC, USA.
- Reyes, L. A. (2006) “The NRC Knowledge Management Program,” *SECY-06-0164*, U.S. Nuclear Regulatory Commission, Washington, DC, USA. (ADAMS ML061550002)
- Rosenberg, S. (2016) “Risk-Informed Licensing Activities,” *Transcript for Meeting of ACRS Reliability and PRA Subcommittee, September 7*, U.S. Nuclear Regulatory Commission, Washington, DC, USA. (ADAMS ML16270A602)
- Schroer, S. and M. Modarres (2013) “An event classification schema for evaluating site risk in a multi-unit nuclear power plant probabilistic risk assessment,” *Reliability Engineering and System Safety*, 117, pp. 40-51.

- Siu, N., et al. (2013) “PSA technology challenges revealed by the Great East Japan Earthquake,” *Proceedings of PSAM Topical Conference in Light of the Fukushima Dai-Ichi Accident*, Tokyo, Japan, April 15-17. (ADAMS ML13038A203)
- Siu, N., et al. (2016a) “Probabilistic Risk Assessment and Regulatory Decisionmaking: Some Frequently Asked Questions,” *NUREG-2201*, U.S. Nuclear Regulatory Commission, Washington, DC, USA.
- Siu, N., et al. (2016b) “Advanced Knowledge Engineering Tools to Support Risk-Informed Decision Making: Final Report,” U.S. Nuclear Regulatory Commission, Washington, DC, USA. (ADAMS ML16355A373)
- Takada, T. (2013) “Robustness and resilience as extensions [sic] of risk concept after Fukushima event,” *Proceedings of PSAM Topical Conference in Light of the Fukushima Dai-Ichi Accident*, Tokyo, Japan, April 15-17.
- Tobin, M., K. Coyne, and N. Siu (2011) “Current PRA Knowledge Management Activities at the NRC,” *Proceedings of ANS PSA 2011 International Topical Meeting on Probabilistic Safety Assessment and Analysis*, Wilmington, NC, March 13-17, American Nuclear Society, LaGrange Park, IL, USA. (ADAMS ML110210666)
- Vial, E., V. Rebour, and B. Perrin (2005) “Severe Storm Resulting in Partial Plant Flooding in ‘Le Blayais’ Nuclear Power Plant,” *Proceedings of International Workshop on External Flooding Hazards at Nuclear Power Plant Sites (jointly organized by Atomic Energy Regulatory Board of India, Nuclear Power Corporation of India, Ltd., and International Atomic Energy Agency)*, Kalpakkam, Tamil Nadu, India, August 29 – September 2.
- Webster (1969) *Webster’s Third New International Dictionary, Unabridged*, P. B. Gove, Ed., G. & C. Merriam Co., Springfield, MA, USA.
- Zhu, W.-D., et al. (2011) “IBM Content Analytics Version 2.2: Discovering Actionable Insight from Your Content, Second Edition,” International Business Machines Corporation, Armonk, NY, USA.