

2. Search Methodology

2.1 Searching the RaDiUS Database

RaDiUS consists of a graphical web-based user interface, a search engine, and a database of federal R&D information. Rather than coding the database for keywords and then searching the database, RaDiUS searches its database by looking at project descriptions and mission statements of the hierarchical organizations located within the database.

This means that effectively searching the database requires one to develop search terms that reflect the full breadth and depth of the topic of interest. This can be challenging since some project descriptions are written in general terms while others are written using highly specialized and precise technical language without discussing the larger context of the research.

For this reason, comprehensive RaDiUS searches can require numerous searches using different sets of terminology. Furthermore, since some of these terms may be used by other scientific and technical communities, additional efforts are often needed to screen out and remove projects that are found by RaDiUS but that are not relevant to the topic of interest. For these reasons, the first step of any RaDiUS search is to develop a search strategy that effectively targets the topics of interest.

2.2 Developing a Search Strategy

A successful search typically consists of combining two different search strategies.

The first strategy consists of identifying initial search terms and testing them by searching the database. One then looks at the results and evaluates each search term's effectiveness. Based on the results, additional search terms can be identified by reading matching project descriptions or by selecting search terms that are suggested by RaDiUS. At the same time, terms uniquely associated with irrelevant projects can be identified and used to screen out irrelevant projects. Finally, ineffective search terms can be excluded all together.

The second strategy which can be pursued roughly in parallel with the first is to identify those agencies, programs, etc. that fund relevant research and to look at the R&D performed by those organizations. In many cases, the second strategy can be done most effectively by looking at the funding sources that are identified in the initial and subsequent phases of the first strategy.

By using these two strategies iteratively, one can search broadly while excluding R&D which is clearly irrelevant.

For this project, the process of identifying initial search terms began by reviewing a number of reports to identify key terms, technologies, and interest areas. While the specific reports are listed in the References chapter, they included NSTC and member agency documents as well as the results of a RaDiUS search of FY94 R&D data (McGaraghan, 1996). Since the FY94 search focused predominantly on buildings and their lifecycle, specific efforts were made to develop a broad list of search terms in the non-building area. In addition, a special effort was made to ensure that all housing-related R&D topics were included.

The specific search terms and sponsoring organizations examined for this project can be found in Appendices A and B of this report.

2.3 Confirmation of Project Relevance

Once the iterative dual strategy process described above was completed, more than 3,500 projects had been identified. The resulting project descriptions were then read by the project team to confirm that the projects were indeed relevant. In those cases where RAND was unsure whether a topic was relevant or not relevant, the C&B Secretariat and/or federal agencies were contacted to seek guidance. (For example, fundamental/basic research on materials was considered to be outside the scope of construction and building, as was large scale power generation). As a result of the confirmation process, the total number of relevant projects was reduced to roughly 1,600.

2.4 Categorizing Search Results

Once the results were complete, the records were skimmed to determine how the results could be organized into meaningful and useful categories that could then be analyzed on an aggregate basis. While the research team had initially hoped to categorize the results according to the National Construction Goals, this format was found to be problematic because some projects could be categorized

under multiple goals while other projects could not be categorized under any of the National Construction Goals. For this reason, the project team used three categories derived from the National Construction Goals as well as new categories developed for this search.³

Just as determining search terms was an iterative process, so too was developing the categories. This is because having too many categories was judged to be as unhelpful as too few. In addition, some R&D areas can be easily combined, such as highway and bridge R&D. That said, it is important to realize that any aggregation will disappoint some stakeholders. In developing the categories, the project team sought to devise categories that would be useful to most people most of the time. (Those seeking to disaggregate categories or to create different aggregations should obtain the final electronic data and develop categories that meet their own needs). In addition, it should be stated that some projects could be placed into more than one category. For this search, however, the project team placed projects in the category that seemed most appropriate given the information in the project description.

The final categories are listed in Table 1. To better understand the specific types of research addressed within each category, the reader is referred the category sections of the results chapters where the R&D conducted within each category is explained briefly.

³ The categories derived from the National Construction Goals were "Pollution & Waste Reduction", "Reduction in Construction Work Illness & Injuries", and "Reduction in Occupant-related Illness & Injury."

R&D Categories
Building Design Improvements
Building Process Improvements & Automation
Building Product Improvement
Concrete, Cement, Pavement, & Asphalt
Energy Efficiency
Energy Supply
Forestry
Geotechnical Engineering & Soil/Groundwater Remediation
Land-use Design Improvements
Metals, Composites, & Advanced Materials (Non-wood, Non-concrete)
Other
Pollution & Waste Reduction
Reduction in Construction Work Illness & Injuries
Reduction in Occupant-related Illness & Injury
Structural Engineering & Natural Hazards
Transportation Infrastructure
Unknown
Wood Products & Quality

Table 1. Categories Used to Organize R&D Projects

2.5 Addressing Agency Data Limitations

As described earlier, RaDiUS' data is based on the Presidential budget and information given to RAND by the individual agencies. For many agencies, budget information may come from one office while project information may come from several other offices. Experience has shown that combining these data sources can reveal inconsistencies and errors for several reasons including:

- The data submitted by agencies can be incomplete or inaccurate thereby affecting project descriptions and/or funding-related information.
- RaDiUS tracks federal R&D as defined by OMB Circular No. A-11. This circular defines "Research and Development" as several budget categories with specific R&D code numbers. If agency budget records do

not reflect these codes, those projects are not included in the RaDiUS database.

- While RaDiUS does not include classified R&D, the Departments of Defense and Energy have decided that some aspects of their unclassified R&D portfolio are too sensitive to be completely and readily accessible to all users. Accordingly, access to some project information is restricted by user type.

Each of the above limits RaDiUS' ability to provide a full and complete view of the federal R&D portfolio. This is because awards without project descriptions are essentially impossible to find since the RaDiUS search engine principally matches search terms with project descriptions. Second, projects bearing incorrect budget codes are completely missed. Third, projects that lack funding information underestimate the total federal investment in a given type of R&D.

Fortunately, some of the data limitations can be partially overcome through reasonable approximations. The two main approximation approaches are described below. In addition, agencies can be contacted to request additional information where desired. For further information on the data limitations encountered during these RaDiUS searches and how they were resolved, please see the agency-by-agency breakdown in Appendix C.

2.5.1 Approximating Missing Project Funding Levels

In cases where project funding information was lacking, the funding level was roughly estimated by dividing the sponsor's "Budget Authority" by the total number of projects awarded by that sponsor. ("Budget Authority" refers to the amount of funding the sponsor has been authorized to spend in the Congressionally-approved budget). It should be noted that this form of estimation only works if RaDiUS contains information on the sponsor's Budget Authority.

2.5.2 Accounting for Missing Project Descriptions

As previously stated, the RaDiUS search engine cannot find projects that lack project descriptions. However, a second type of approximation can be used when one is confident that all or most of a sponsor's R&D is relevant. In these cases, it may be better to include the sponsors entire portfolio even though there are no project descriptions. During this search, this approximation was used for the US Army Corps of Engineers as described in Appendix C.