now nearing completion, the file is loaded under STAIRS and available to staff members who may sign directly on to the file or who may access it during a search session with other BIOSIS on-line files. Since no portion of the file may be altered during these sessions, the file may be used by many staff members rather than a restricted few and the users need not learn any command language other than STAIRS to use it.

A given substance is retrieved through constructing the known name and displaying the resultant hits. The logical operators AND, OR, NOT, and ADJ may be used to construct searches based upon name fragments. Search results may be sorted upon Registry Number if desired, and results may be more selectively chosen by specifying Accession Number or Registry Number ranges. The record displayed lists the information in file about a substance. Verbal descriptions of the data type head each portion of the display and each synonym is listed on a separate line. STAIRS also offers the user immediate copies of a retrieved document from a remote printer, or overnight printing of entire search results.

Database Development in a Regulatory Agency†

MARILYN C. BRACKEN*
The MITRE Corporation, McLean, Virginia 22101

IRVIN J. WEISS
U.S. Consumer Product Safety Commission, Bethesda, Maryland 20207

Received April 5, 1977

A general discussion of the history and problems associated with the collection of information and the development of a database in a regulatory agency is presented. The proceedings of the U.S. Consumer Product Safety Commission for obtaining chemical formulation information for specified consumer products is discussed. Guidelines for database administrators faced with data collection activities in a regulatory agency are provided.

INTRODUCTION

The Consumer Product Safety Commission (CPSC) is responsible for programs that reduce the hazard of human injury from chemical consumer products. To make scientific conclusions and value judgments about the safety of a product, complete information on ingredients in the product is required. This paper describes the nature of proceedings of the U.S. Consumer Product Safety Commission to obtain chemical formulation information for specified consumer products, including (1) the issuance of a Special Order to obtain such information and (2) the defense against legal actions brought by a trade association to obtain a preliminary injunction to prevent the Commission from collecting this information. Recommendations for others involved in data collection and database development are presented.

BACKGROUND

Unlike the Food and Drug Administration and the Environmental Protection Agency, the Commission has no preclearance requirements in its laws which obligate manufacturers to seek approval for marketing their products. However, the Commission does have the authority (which it has not yet implemented) to prescribe procedures for the purpose of ensuring that the manufacturer of any new consumer product furnish notice and a description of such product to the Commission before its distribution in commerce. The Commission may also, by rule, require any manufacturer of a consumer product to provide to the Commission performance and technical data that relate to performance and safety as the Commission determines necessary.

In 1973, the Commission decided that in order to respond to its Congressional mandate to protect the public from unreasonable risk of injury, and for effective enforcement, product ingredient information for chemical consumer products was required. A project was initiated to collect formulation information for selected consumer products.

USES TO BE MADE OF THE DATA

The purpose of collecting formulation information on chemicals in consumer products is to provide the Commission with means of ascertaining the potential or real hazards to consumers due to exposure. The Commission receives petitions, consumer complaints, notices of product defect, and Congressional inquires. It is asked, for example, to consider the banning of a class of product(s) containing a certain chemical. With no detailed knowledge about the ingredient data of such products, it is extremely difficult to predict or evaluate the hazards involved.

There is a need to be able to characterize types of consumer products. For example, what ingredients would one find in tile adhesives or in dishwasher detergents or in any particular category of chemicals. With knowledge of formulation information the Commission can: (1) identify chemicals that are toxic, carcinogenic, mutagenic and teratogenic, (2) look at particular classes of compounds and develop a standard, or (3) address generic standards.

The Commission through its interagency committee activities receives notification of chemicals suspected of being carcinogenic and/or documented to be cancer-causing agents in animals and man. Knowledge of the existence of such chemicals in products that are under the Commission's jurisdiction is critical, particularly if a ban is under consideration. Plans for recall must be developed, as well as economic analyses regarding impact on the affected industry of generating a substitute for the chemical.

In addition, the formulation information will assist the Commission in implementing Section 13 of the Consumer Product Safety Act. This section is designed to provide the

† Presented at the 39th Annual Meeting of the American Society of Information Science, San Francisco, Calif., Oct 4–9, 1976.

Commission with a means of keeping abreast of new products entering the marketplace so that it can restrain imminently hazardous products in the courts or promptly institute a proceeding to ban or develop standards for products which it determines are unreasonably hazardous.

SPECIFICITY OF THE INFORMATION

A decision was made early in the project requiring submission of the formulation information to the 0.1% level for all products which were identified. For effective hazard analysis, it is essential to obtain specific information on all of the ingredients of a product so that evaluation of consumer exposure to various chemicals is as complete as possible. The formulations of many consumer products contain active ingredients at or below the 0.1% level. These chemicals are often used as deodorants, preservatives, surfactants, additives, drying agents, dyes, flame retardants, etc. The chemicals may enhance the ease of the production process, provide a unique physical or chemical quality, and/or provide a consumer-desired product. Certain chemical ingredients may be toxic if ingested or inhaled, cause contact dermatitis or delayed hypersensitivity reactions, or release toxic fumes when burned in a household fire.

Initially, when considering the toxicity of an individual product, the concentration of a chemical may appear to be innocuous. However, when evaluated from a global viewpoint with knowledge of total exposure to the chemicals, there may be indications that the safe threshold level for that chemical has been surpassed. The consumer is exposed to cumulative toxic effects of chemicals or products and to unknown synergistic or antagonistic effects resulting from constant or frequent exposure to chemicals presently unidentified in consumer products.

It was realized that in some cases specifications for a product might vary from batch to batch. It was not suggested that manufacturers have products analyzed for the purpose of this survey, since it was felt that responsible manufacturers already know the chemical composition of their products for purposes of quality control, product performance, as well as product liability.

DATA COLLECTION

Meetings were held in the project development stage with other government agencies to ascertain the existence of similar data collection activities and to ensure against duplicate solicitation of identical data.

The Commission was aware of databases containing production formulation information that exist outside CPSC, for example, the Clinical Toxicology of Commercial Products (CTCP) file at the University of Rochester, the National Clearinghouse for Poison Control Centers database, and the Poisindex system. These data typically reflect ingested products that involve poisonings. Component information is incomplete, not very specific, and/or presented in terms of ranges.

In addition, the Commission learned that the National Institute of Occupational Safety and Health (NIOSH) has a similar data collection activity for formulation information of industrial products. Meetings were held to arrange an interagency agreement to allow sharing of common nontrade secret data.

Before implementation of the project, CPSC submitted its request to the General Accounting Office (GAO), pursuant to the requirement stated in the Federal Reports Act, for review and clearance of its product and chemical names questionnaire. The Commission provided a justification of the need for the data, the data collection parameters, a survey of its efforts to identify like kinds of data already available, an estimation of the burden on industry, and the response to the questionnaire. The GAO approved the questionnaire on August 5, 1974.

Since the project funds were limited, data collection was restricted to selected product categories of information. The selection of the product categories was generated by the National Electronic Injury Surveillance System (NEISS) and Commission background information. NEISS is a data processing system which monitors injury data collected from the emergency rooms in 119 statistically selected hospitals. The product categories used for this project were obtained from the NEISS coding manual. Originally 39 categories were selected and ranked by the Commission. A decision was subsequently made to exclude certain categories under overlapping jurisdiction of CPSC and other Federal agencies. Also, a decision was made early in the project to emphasize the development of the top-ranked 15 categories since funds were not available to obtain sufficient information on all 39 categories.

CONTRACT IMPLEMENTATION

In June of 1974, a contract was entered into with Auerbach Associates, Inc., a private corporation, under which Auerbach would mail product ingredient questionnaires to manufacturers they identified, receive their responses, convert the questionnaire contents to computerized form, and deliver the questionnaires and computer tapes to the Commission. As noted, identification of manufacturers and products for preselected product categories was conducted by the contractor. The contractor consulted numerous directories, contacted trade, government and consumer organizations, obtained trade catalogs from various manufacturers, and used relevant data from the Clinical Toxicology of Commercial Products file.

As the products were identified they were classified according to the NEISS code. A list of identified manufacturers and products was compiled, and it became the basis for the solicitation of the data. Since the total universe of products is not known, a random sample in the statistical sense was impossible. There was no attempt to sample manufacturers and/or products relative to size of the firm, number of products, total sales volume, or past history of the firm's regulatory compliance activities. The Commission attempted to obtain as comprehensive and representative a composition of products as was possible within the terms of the contract.

In its initial request for information the Commission asked for a voluntary contribution of data by the manufacturers. Reluctance on the part of the manufacturers to provide requested information was observed, and in some cases manufacturers who originally had responded asked for the return of data submitted. As a result, the Commission published a General Order on April 14, 1975, ordering manufacturers of certain categories of consumer products to complete and return the product ingredient questionnaires. The intended effect of the General Order was to transform the voluntary solicitation of information into a mandatory requirement.

The GAO, responding to protests by manufacturers and trade associations, required CPSC to rescind the order because GAO regarded the change from a voluntary to a mandatory order to be material, thus requiring a new clearance for the questionnaire and therefore more detailed review by GAO. In addition, GAO was concerned about possible duplication of data collection by the Commission and NIOSH. The order was rescinded on May 6, 1975. After a series of meetings between GAO and CPSC, submission of further information by CPSC including documents justifying collection of data to the 0.1% level and description of security procedures for protection of trade secrets, and GAO consideration of comments received from manufacturers and trade associations, the

GAO determined that the burden to industry was not excessive and gave its clearance for use of the questionnaire on a mandatory basis.

The Commission issued a Special Order on August 21, 1975.[1] The following provisions were made in the Special Order as a result of the GAO review: (1) time for responding was lengthened; (2) an agreement was made between CPSC and NIOSH whereby nontrade secret data collected by NIOSH on products under the overlapping jurisdiction of the two agencies would be shared with CPSC so that manufacturers need not duplicate their reporting; and (3) provision was made for manufacturers to have their jurisdictional objections considered before returning completed questionnaires.

PROCESSING THE DATA

The contractor performed the following processing functions: (1) request generation, (2) response receipt and monitoring, and (3) file generation.[2] The request generation mechanism included data input, clerical and chemical editing, and request form and mailing label generation. A manufacturers' address file and manufacturers' product file were maintained. Using reports generated after the data input, a clerk performed a trade name status review of the products to identify and resolve any synonymous product entries. A chemist reviewed the products to identify those whose chemical components could be resolved in-house, in order to reduce generation of unnecessary data requests to manufacturers.

The computer generated a label for each product containing a unique request number for the product name and for each packet two mailing labels containing the manufacturer's name and address and the total number of product request forms. The packets were than assembled for mailout. When a response was received from a manufacturer, the data were reviewed for completeness and accuracy of information and the action required was coded. In addition, the contractor generated the product formulation files from the data collected for delivery to the Commission.

SECURITY OF TRADE SECRET DATA

If the manufacturer decided that the ingredient data for his product(s) were trade secret, he checked the appropriate box on the questionnaire and explained on a separate sheet of paper what specifically made the data a trade secret. If he omitted his letter of trade secret justification, the contractor followed up this omission with a form letter that requested a written explanation of why ingredient data should be accorded trade secret processing.

The contractor was responsible for filing all the returned questionnaires and for maintaining the security of the questionnaires, punch cards, magnetic tapes, and hard copy listings generated in the course of the contract until they were delivered to the Commission.

The security procedures used by the contractor with respect to ensuring the protection of trade secret information collected under the CPSC contract involved signing affidavits of nondisclosure. All individuals having access to the areas where trade secret information was stored were required to sign the affidavit. The contractor permitted only authorized project personnel to have access to the locked storage cabinets, and all trade secret data were returned to the locked cabinets immediately after any authorized use. Access to the rooms where trade secret data were maintained was limited to authorized personnel, and these rooms were locked except when occupied.

CPSC security procedures for handling trade secret information discuss the requirements placed on the contractor by the Commission and the internal procedures for handling such data.[3] The procedures refer to a DHEW publication (IPS-PUB-3) which establishes the procedures that a computer center must follow for additional processing of such data.[4]

If the Commission receives a request for disclosure of information, or believes it desirable to disclose the information to carry out its legal responsibilities, the Commission contacts those manufacturers whose products relate to the request and provides him the opportunity to present additional arguments and views, if need be, regarding the status of the materials. The determination with respect to release of the information is based on the applicable provision of: (1) CPSA, (2) the Freedom of Information Act, (3) 18 USC 1905, (4) the Commission's proposed and interim regulations on the procedures for disclosure or production of information under the Freedom of Information Act, 16 CFT Part 1015, and (5) the most recent judicial interpretation of these provisions. In the event that there is a disagreement as to trade secret status, a manufacturer will be given 10 days notice before the information is released. The Commission's policy, however, is not to disclose a trade secret formula.

LAW SUIT BY TRADE ASSOCIATION

On October 17, 1975, the Chemical Specialities Manufacturers' Association (CSMA), a trade association representing its members affected by the Commission's order, sought a preliminary injunction in the United States District Court for the District of Columbia to enjoin the Commission from collecting the chemical formula. CSMA arguments questioned the Commission's authority to issue the Special Order to collect the information sought, and stated that the contractor and the Commission had not provided for adequate security precautions to prevent the disclosure of any trade secret obtained as a result of that data collection.

On November 3, 1975, the District Court ruled in favor of the Commission on all points raised. The Order was immediately appealed by CSMA, but on November 18, 1975, the United States Court of Appeals for the District of Columbia upheld the District Court decision in a *per curiam* opinion. As the Commission pointed out in the various papers which it filed with the courts, the Commission believed the product survey was essential, that adequate security measures had been taken, that the burden on industry would be reasonable, and the CPSC was authorized to employ a contractor to process the data. The suit is at this time still pending because the plaintiffs want a permanent injunction. There has been no decision as to permanent injunction.

RECOMMENDATIONS FOR DATA COLLECTION AND DATABASE DEVELOPMENT

Based on the experiences over the last two years in data collection activities and database development resulting from the data collection, certain recommendations can be made to others attempting similar efforts. Figure 1 portrays the overall decision process in a generalized form. The following comments address selected key elements of the process.

Any database administrator or database manager in a regulatory agency or any government agency should be cognizant of statutory authority that the agency has with respect to data collection, storage, and dissemination of information. It is recommended that the agency General Counsel be consulted. A database administrator should have knowledge of other legislation (and agency regulations) which have an impact on database management. These are, for example, the Federal Reports Act, the Freedom of Information Act, and the Privacy Act of 1974.

Initially, in the project development stage the database administrator should clearly and explicitly identify the data to be collected, the need for such data, the uses to be made of the data, and the authority to collect the data. All of this

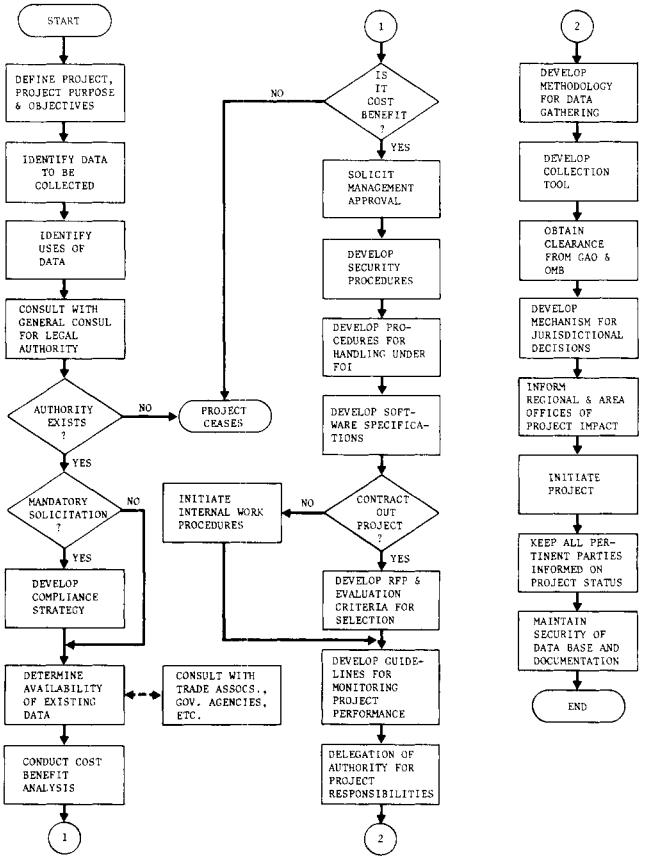

Figure 1. Flow chart of database development in a regulatory agency.

information should be documented and approved by the project and agency management to demonstrate a commitment by management to the activity. In addition, this information will be required by GAO if a clearance of a collection tool is necessary, by Members of Congress inquiring on behalf of their constituents, by trade associations, and possibly (as in the CPSC case) for use in defense of a legal action.

The database administrator will need to demonstrate that the agency has attempted to ascertain the existence of similar databases in order to avoid duplicate and/or redundant data collection activities. A survey of government agencies, industries, and universities for useable data should be conducted. A check can be made of the National Referral Center of the Library of Congress, the National Technical Information Services, and Smithsonian Science Information Exchange. Review articles in publications of the Association for Computing Machinery (ACM), American Society for Information Science (ASIS), etc., can be scanned for leads to the availability of like kinds of data. Trade Associations can be contacted and meetings held with industry to inquire about the availability of such data and to open dialogues regarding minimizing the burden to the affected industry.

Decisions needed to be made early in the project development stage include the following:

- What are the cost and benefits of building the database?
- What data will be collected by contract or in-house?
- Will data be processed by contractor or internally?
- What is the half-life of the data?
- Is the effort continuous or one-time?
- What kinds of reports are needed?
- What kinds of security procedures are necessary?
- What are the software requirements?
- How will the results of the activity be disseminated, if at all?
- Who will collaborate on procedures to be developed and decisions to be made?

After these decisions are made, the following procedures and specifications will need to be developed in this kind of a database activity:

- a database dictionary which defines and documents the need for all data elements
- security procedures for handling proprietary or personal data (internal and for any contractors)
- methodology for data collection including sampling procedures
- mechanisms to consider jurisdiction decisions
- briefing materials for informing agency regional and area offices if collection activity will impact them
- development of questionnaire or collection tool
- preparation of clearance requests to be submitted to the proper government agency—General Accounting Office (GAO) or the Office of Management and Budget (OMB)
- development of software specifications
- development of Request for Proposal (RFP)
- development of evaluation criteria for the RFP
- specification of machine input mechanism

It is recommended that a project management system be developed which includes further procedures for (1) monitoring the contractor's performance, (2) delegation of authority for certain responsibilities among project staff, (3) maintenance of an adequate filing system for all correspondence, procedures, memoranda, etc., to assure adequate documentation, and (4) maintenance of the integrity and security of the database.

LITERATURE CITED

(1) U.S. Consumer Product Safety Commission, "Chemical Formulations for Specified Consumer Products", Federal Register 40 36617, Aug 21, 1975.
(2) W. L. Byer, H. B. Landau, M. Lynne Neufeld, and H. Rosenthal, "Building a Chemical Ingredient Data Base for Industrial and Consumer Products", *J. Chem. Inf. Comput. Sci.*, **16**, 137–141 (1976).
(3) U.S. Consumer Product Safety Commission, "Procedures for Safeguarding Confidential Information", 16 CFR Part 1017, Federal Register 41 36648, Aug 31, 1976.
(4) U.S. Department of Health, Education and Welfare. "ADP System Security Required by the Privacy Act of 1974", Department of Health, Education and Welfare Information Processing Standards Publication, 1975.