ESTABLISHING A PROGRAM/DATA LIBRARY AT IIASA

Tom W. Carroll
Michael H. Abkin

January 1978 WP-78-3

2361
Laxenburg
Austria

International Institute for Applied Systems Analysis

Preface

IIASA's *1976 Research Plan* (March 1976) lists the development of a program library and a computer data base at IIASA as one of the research activities of the Computer Science Section in the Systems and Decision Sciences Area. However, most of the effort in the past year has been focused on developing an international computer network under IIASA's leadership.

In the *Research Plan 1977* (March 1977) Task 4 in the Systems and Decision Sciences Area on "IIASA Computer Network (Including Data Bases)" specific reference to IIASA's program library (p.128) is made:

> The main goals [of Task 4] are to provide IIASA applied projects with a tool for access to programs and data in collaborating institutions, and to foster collaboration by the provision of adequate means of communication and access to IIASA's program library for collaborating institutions... Significant effort will be undertaken to prepare programs developed at IIASA in the form of well-documented packages, which would be available through the IIASA computer network.

Computer Services, a separate service division, has made considerable progress during the past several years in establishing a computer facility at IIASA using a PDP 11/45 time-sharing system [1]. It also provides access to other computer systems including a CDC CYBER 74 at the Technische Universitaet Wien (TU), IBM 370/158 at IBM Wien, and a UNIVAC 1106 Wirtschafts- und Sozialwissenschaftliches Rechenzentrum Wien (WSR). Basic program libraries are available on these computers including language processors (e.g., FORTRAN, BASIC, APL, and PL/1), mathematical programming routines (e.g., APEX-II, EISPACK, IMSL, MATH-PACK, NAG, SSP, IIASA's OURLIB), statistical routines (e.g., BMD, SELCOM, SPSS, SSP, STAT-PACK, TSP, IMSL, OURLIB), and various plotting packages for the sites that support plotters. GPSS, a discrete event simulation language is available at IBM and WSR. DYNAMO will be acquired shortly by IIASA for the PDP. It is clearly evident that IIASA researchers already have a powerful combination of computing resources available to support their research activities.

In addition, good progress is being made in developing the international computer network. Operational nodes currently exist in Budapest, Warsaw, and Bratislava, and nodes in Vienna and Pisa are planned to be in operation by the middle of 1977.

As the network becomes more fully operational for researchers, the range of available facilities, programs, and data bases will continue to increase.

Thus, with good progress already made in providing basic computer services, the time is right to devote more attention to establishing a program/data library at IIASA with the primary objective of storing and indexing those programs and data bases developed and/or utilized at IIASA in applied systems analysis. Some progress has already been made. A new IIASA publication series called IIASA Software Library Series has been started. The series currently includes three titles [2, 3, 4]. However, much more needs to be done in order to establish a working library.

The issue is timely because during this past Spring delegates from nine countries formalized the establishment of a new European organization to provide a comprehensive range of services in support of users of scientific software. Known as the European Association for Software Access and Information Transfer (EASIT), the organization's activities are likely to include the provision of software libraries, consultancy services for users of networks and computer centers, software documentation, and information services. Program distribution agencies, user groups and originators of major packages will be among the membership. The Association aims to improve services in support of European users of software by promoting exchange of information among appropriate scientific, professional, and governmental organizations. Six EASIT working groups have been established to examine documentation, programming standards, software classification, magnetic tape formats, inter-package communication, and software for the social sciences.

The senior co-author of this working paper, T.W. Carroll, was invited to IIASA for the month of October, 1976, by Dr. A. Butrimenko, Deputy Chairman of the Systems and Decision Sciences Area to consult with IIASA staff on these matters. He returned to IIASA briefly in June and November, 1977, for further discussions. The senior co-author is an Associate Professor at Michigan State University, where he is currently Director of User Services in the Computer Laboratory. He has also served as Head of the Technical Section in the Computer Institute for Social Science Research at M.S.U. The other co-author, M.H. Abkin, is currently an Assistant Professor in the Department of Electrical Engineering and Systems Science of Michigan State University.

Since 1971 both co-authors have been associated with the Project on Adapting the System Simulation Approach to Agricultural Sector Analysis, carried out by the M.S.U. Department of Agricultural Economics and sponsored by the U.S. Agency for International Development (under contract No. AID/csd-2975). In addition to conducting field applications in Korea, a

development analysis training program, and methodological research, project staff were also required under the terms of the contract to:

> Establish a software library, initially at M.S.U., but perhaps to be transferred later to the proposed IDI of A.I.D. The various mathematical models and computational routines referred to here as "software" should be placed in one or more of the standard computer languages, indexed and made available to "borrowers" with appropriate explanations. This will involve generalization of existing components and the provision of a higher-order simulation language for the integration and combination of components into configurations required in various institutional and geographic contexts. Such an arrangement is needed to make the model components available on a national and international basis to potential users in donor, lender, and host country agencies.

This working paper draws on the authors' experience in setting up the Computer Library for Agricultural Systems Simulation (CLASS) [5], under contract AID/csd-2975 and is based on an earlier draft proposal by M.H. Abkin [6]. IIASA's needs for a software library and the library's functions, scope, services, staffing requirements, methodological support, and possible organizational location are discussed.

Because the authors feel that both computer programs and data bases should be stored and maintained as part of the proposed software library, they use the term "program/data library" rather than "software library" in order to indicate explicitly the nature of the library's holdings.

Preparation of this working paper has been supported both by IIASA and by the U.S. Agency for International Development under Contract No. AID/csd-2975.

Laxenburg, Austria
June, 1977

T.W. Carroll
M.H. Abkin

Table of Contents

Preface .....

Introduction 1

Background on IIASA's Research Plan 1

IIASA's Needs for a Program/Data Library 2

Definition and Scope of Holdings 7

Library Standards 12

- Program Selection Standards 12
- Documentation Standards 15
- Programming Standards 15
- Review Process 16
- Data-base Standards .....

Status-Levels of Holdings 16

Incentives for Contributors 16

Use of the Library 17

Basic Services 18

- Acquisition 18
- Cataloging 19
- Storage and Maintenance 19
- Distribution 19

Additional Support Services 20

- Consulting 20
- Training 20
- Upgrading Programs and Documentation 21
- Methodological Development 21

Staffing 22

Organizational Location at IIASA 23

References 25

Establishing a Program/Data Library at IIASA

INTRODUCTION

The basic functions of a program/data library at IIASA would be to acquire, catalog, store, maintain, and distribute computer programs and data bases along with their associated documentations. These computer programs are of generalized models, components, and routines related to IIASA's research program in applied systems analysis. In particular, the library sets standards of admissibility for programs, data bases, and their documentations, catalogs and stores these modules so as to facilitate their retrieval by users seeking a set of modules to be used in a specific problem analysis, and distributes the modules to users in the appropriate machine-readable format.

To further enhance the effectiveness of the library concept, supporting activities also include: identifying and soliciting needed modules, upgrading acquired modules to meet library standards, providing consultation in identifying and implementing appropriate library programs for particular applications, and methodological research to increase the library's usefulness to users. A subsidiary function of the library in conjunction with the identification and solicitation of models is to survey on-going research in applied systems analysis.

The purpose of this paper is to discuss IIASA's needs for a program/data library, the scope of its holdings, potential contributors and users, basic services and supporting activities, and steps required to establish and staff the library.

BACKGROUND ON IIASA'S RESEARCH PLAN

IIASA's *1976 Research Plan* identifies IIASA's special role among international scientific research institutes as being "to perform *integrative studies* of both global and universal issues, concentrating both on the development of *new methodologies* for their study and on the *analysis of alternative strategies* for their resolution." (p.2, authors' italics highlight the key words from the standpoint of the contribution of a program data library to achieving IIASA's research objectives.) To achieve these objectives IIASA has reorganized its research activities along two cross-cutting dimensions.

Integrative studies of major global and universal issues comprise one dimension. Global issues are those issues whose impact and resolution involve groups of nations, while universal issues are those issues which almost all nations face within their own boundaries and can act upon individually. Although IIASA has supported recent efforts at developing global models through sponsorship of our conferences on global models, the current research plan calls for a sector-by-sector approach to building up the knowledge and expertise required for a later attack on the comprehensive problem of global development. The Energy Program is the global sector currently being investigated in depth; food and agriculture has been selected as the next sector for study at the global level. As its cross-cutting universal issue, IIASA has selected "integrated regional development." The Integrated Regional Development Program will draw together portions of existing projects from among the supporting research areas outlined below. Areas being considered for investigation in the Integrated Regional Development Program include: integrated management of regional energy/environmental systems; regional water demand and management; and planning, management, and organization of large-scale regional development programs.

Four supporting research areas comprise the second dimension of IIASA's research program: Resources/Environment, Human Settlements/Services, Management/Technology, and System and Decision Sciences. The first three research areas will build up IIASA's competence in the substantive areas needed to support integrative research at the global and regional levels; the earth's natural endowment, its human endowment, and its man-made social, economic, and technical mechanisms. The fourth research area, System and Decision Sciences, is concerned with mathematical and computational methods of treating systems and decision problems.

IIASA'S NEEDS FOR A PROGRAM/DATA LIBRARY

The section examines IIASA's needs for a program/data library from the standpoint of outputs from and inputs to IIASA's scientific research effort.

The output from IIASA's scientific research activities takes three main forms:

1. Scientific publications
2. Increased scientific expertise (IIASA "alumni")
3. Computer-based models and data bases.

Greatest progress has probably been made in establishing the formal procedures for editing, printing, and distributing the scientific publications. The different publication series now

include the Research Report (RR), Collaborative Paper (CP), Research Memorandum (RM), the Professional Paper (PP), and the Working Paper (WP). Total publication for 1973 was 36 publications; 1974, 59 publications; and 1975, 124 publications. The Publications Department is set up as one of four scientific services departments. This output activity seems well in hand and has already contributed to establishing IIASA as an important international research center. Although there are currently only 15 - 20 annual subscribers to IIASA publication series at a cost of $250 per year, 1500 individuals or organizations on IIASA's mailing list receive indexes or selected publications on a regular basis.

In addition, the special research project on the International state-of-the-art in applied systems analysis is preparing two products. The first in the State-of-the-Art Series of individual volumes which survey specific aspects of applied systems analysis. The second product will be a Handbook of Applied Systems Analysis in three volumes, which will provide a comprehensive survey of the field.

The second form of IIASA output is the increased expertise and insights gained by scientists while working in the unique international atmosphere of exchange provided by IIASA. In an informal way the enthusiasm carried forward by IIASA's alumini helps establish the reputation of IIASA. In many cases alumini still continue to contribute directly to IIASA's research objectives through formalized collaborative research programs.

The final area of scientific output includes computer programs (models, components, routines) and data bases developed by scientists working at IIASA or participating in collaborative research programs. The development of "new methodologies" often takes the form of programs and supporting data bases, which represent an output of scientific "capital stock,"* which is not currently being adequately "packaged", indexed and made available as a product of IIASA. A program/data library could perform these activities as one of its functions. The primary objective here is to serve the scientific community outside of IIASA and to help establish IIASA as a source of quality programs and data bases for applied systems analysis.

---

*Referring to the modules as "capital stock" highlights certain characteristics of software modules which are similar to those of any piece of capital. Both capital stock and software modules, (1) require an investment of resources to develop, (2) once designed, can be duplicated for use in multiple applications, (3) usually must follow industrial standards to facilitate utilization and maintenance, (4) once developed, save the user time and labor in both development and utilization, (5) have a useful life of 5 to 10 years before being made obsolete by technological advances.

Equally important, if not more important, is the need for a program/data library to support IIASA's own scientific personnel in achieving IIASA's research objectives as outlined above. As an international scientific research institution, IIASA has many strengths. However, it operates under the handicap that most of its scientific research appointments are made for only one or two years. Thus, in terms of staff, IIASA has a "short memory." The publications series, the document library, collaborative research projects, and consulting visits by former alumni all contribute to overcoming this handicap. However, in the special area of computer programs and data bases there currently is no formal mechanism for preserving the IIASA computer-based work. Much of the information about models and data bases now resides in the research groups, not in the Computer Services Division. When researchers leave, knowledge about models and data bases tend to get lost. It is true that some of these models have been documented in IIASA's regular publication series, often with a computer listing as an appendix. However, useful as such documentation is for describing the theory behind the model or the structure of a data base, this is not an efficient way to index and distribute computer-based models and data bases. The basic element being stored in a library is the machine-readable source code and/or object code along with the necessary documentation which itself may be in machine-readable form. Thus, a library would serve as an efficient memory source of software and data bases developed at IIASA for the various research areas.

Sometimes it is argued that researchers are reluctant to use a computer program developed by someone else because they do not trust it or do not understand it. This argument should be considered as an argument in favor of establishing a software library, not an argument against it. One aspect of the *modus vivendi* of scientific research is to build incrementally on the work of others as reported in the scientific literature. The same approach applies to the concept of a program/data library as a storehouse of one type of scientific knowledge. Researchers often avoid using computer programs developed by other researchers because the programs may be poorly documented, poorly coded (making them difficult to adapt), or written in an incompatible language. One function of a library is to develop and foster the use of programming and documentation standards to facilitate the use of its contents by other researchers. At least one IIASA research memorandum has cited the need for "a library of generalized and tested modules" as "the first important ingredient in developing a new generation of models which focus on specific resource or environmental problems" [7]. The library will thus contribute directly to developing new methodologies and making them available for researchers both within and outside IIASA.

The purpose of the library is not only to contribute to the effective use of new methodologies and conceptual models within research areas, but also across research areas.

With respect to new methodologies, IIASA scientists, particularly in the Systems and Decisions Sciences Area, are developing generalized methodologies for applied systems analysis which are useful across a number of substantive research areas. In building, modifying, testing, and validating substantive models, a number of necessary activities could be simplified with the aid of computerized technical routines. Some examples are senstitivity testing and/or Monte Carlo analysis of model parameters, gradient search techniques for parameter estimation, spectral analysis of simulation generated data, tracking model output against historical time series data, regression analysis on time series data to estimate model parameters, database/model linking to initialize variables at levels of model aggregation, and providing conversational interfaces between models/databases and policy planning personnel who are not likely to be sophisticated computer users.

In addition to the technical routines, there exists a class of generalized, functional components which are useful in a number of different application areas. For example, demographic routines have been developed [8, 9], which model not only populations of people, trees and cattle but also other populations, e.g. tractors, houses, bacteria, and insects. This is possible because the demographic process of aging, attrition, and regeneration is a fundamental, widely occuring process in nature. The basic dynamics are independent of the areas of application. The library would provide considerable efficiences by making those basic routines and functional components available in a well-tested, well-documented, computationally efficient format. Thus, a primary objective of the software/data library is to reduce duplication of effort. For example, the *1976 Research Plan* (p.63) Subtask 3 under Biomedical Systems states:

> During the first stages of the work, special attention shall be paid to modelling the sex-age structure of the population in order to find out its effect on HCS resource consumption and vice versa.

The implication is that the Biomedical Systems staff will develop a population model from scratch. Surely, either at IIASA or somewhere else there exists a population model which could be adapted to meet this need. The function of the program/data library would be to acquire and select suitable models for such tasks.

In order to support the primary research objective of preparing "integrative studies of both global and universal

issues" and "analyses of alternative strategies for their resolution," IIASA also needs to store in a centralized library the substantive computer models and supporting data bases, developed in the supporting research areas. The basic strategy underlying both global and regional integrative studies is that it will be possible to link together pairs and clusters of substantive models each of which will have been well-developed and tested by a research task group. IIASA does not have the resources to develop these component models simultaneously. The global research program has outlined a sector-by-sector approach to develop the basic component sub-models of a global model with the linking to follow later. The Integrated Regional Development Program already has some tested models and is ready to begin linking some of these models to explore alternative strategies for specific applications.

Thus, the outputs from IIASA's research areas will become the inputs to the global and regional research programs. The program/data library will have the crucial role of storing the models developed by the research tasks and encouraging programmers to follow a set of programming and documentation standards which will facilitate utilization of the models either singularly or linked together.

Another important function of the program/data library is to keep abreast of scientific research going on outside of IIASA in order to identify models and data bases which could be utilized by researchers at IIASA. The IIASA research reports prepared by Jean Pierre Charpentier are one example of a very useful survey of existing models in the energy field [10, 11]. Similar surveys need to be done in other fields. Although it would not be possible nor desirable for the library to acquire and implement very many of the models identified in such surveys to serve a useful "clearing house" function in their own right and also help IIASA researchers identify selected models which might be brought to IIASA, implemented on the computer and cataloged into the library.

The emphasis in this discussion has been placed on establishing a "working library" rather than an "archival library" -- the same approach being followed by IIASA's document library. However, it is possible that over time IIASA would become known as an international repositary of programs and data bases. Without being solicited by IIASA, researchers outside of IIASA might on their own volition begin submitting models, components, routines, and data bases to be considered for acceptance in IIASA's library. Also, as the computer network becomes a reality, the clearing-house function of the library will become more important, if only to the point of indexing what is accessible through the network.

In summary, the purposes for establishing a program/data library at IIASA are to:

1. Preserve and distribute important products of IIASA's research to users outside IIASA.

2. Preserve products of IIASA's research for internal use in order to:

   a) counteract IIASA's "short memory" problem resulting from high staff turnover;

   b) avoid duplication of effort;

   c) make methodological advances in ASA techniques available to all substantive research areas;

   d) support integrative research objectives by providing a library of standardized, modularized models and submodels.

3. Locate and acquire modules developed outside IIASA which are useful to IIASA's research program.

4. Support an activity which may evolve into an internationally known repository of programs and data bases supporting applied systems analysis of global and universal problems.

5. Support an activity which can make a significant contribution to the successful utilization of IIASA's computer network.

DEFINITION AND SCOPE OF HOLDINGS

The various models, submodels, components, subcomponents, technical routines, and data bases stored in the library will be referred to as "modules". The term "module" is analogous to the term "document" in the traditional manuscript library. The card catalog in a manuscript library points to a document sitting on the shelf. In most cases (except multiple-volume work), there is one physical document.

In the case of a software/data library the term "module" is used to highlight the differences between what is stored in a software/data library and what is stored in a manuscript library. The term module is used to suggest two important characteristics of the holding beging stored in a software/ library: (1) the module (particularly in the case of a computer program) models a process by transforming inputs into outputs by means of a well-defined algorithmn and is potentially linkable to other modules, and (2) is itself comprised of subcomponents. We have already discussed the importance of the first set of characteristics in terms of IIASA's research

program, and will discuss it again in terms of the need for standards.

The second characteristic -- that a software module is comprised of subcomponents -- complicates the consideration as to what gets "stored on the shelf." For example, a computer program often is made up of the following components:

1. Abstract symbolic model of a process, which is usually described by a verbal and/or mathematical description in a standard scientific publication.

2. Algorithmic representation of the process written in a higher level language (source code).

3. An associated compiler which translates the higher level language to object code for a particular machine.

4. Object code - the machine level instructions including any required auxiliary library routines.

5. User documentation which explains how to prepare inputs and outputs at the substantive process level.

6. Internal program documentation which explains the structure of the computer program to a programmer in order to aid in implementation, linkage with other modules, and modification.

7. A test deck with which to test out the implemenation of the program on a particular computer.

A data base also is made up of subcomponents which might include:

1. A conceptualization of the observational process including the relation between operational measures and theoretical concepts, e.g., income.

2. A BCD file of the basic data matrix (possibly in several forms: raw, cleaned, transformed, and inverted).

3. Binary file of above.

4. Structured data file (hierarchical, network, or relational) which include additional pointers to facilitate certain types of processing, e.g. aggregation.

5. For structured files, a data management system.

6. Directory or codebook.

Figure 1 indicates the range of software support for applied systems analysis in a typical university computer center. The examples below the headings are related to IIASA's computing activities. At the core, of course, lie the particular hardware facilities. The highest level of support in a typical computer facility is usually given to the operating system and then the compilers; this support may be provided by the computer center staff or the manufacturer or both. The next level of support usually is given to mathematical, statistical, and utility libraries (e.g. IMSL, SPSS, etc.) Support for these libraries usually comes from some other institution although the local computer center staff is responsible for installing updates and making documentation available. In similar fashion, computer centers often make available one or more of the more widely used simulation languages (e.g., GPSS, DYNAMO, GASP, etc.) It is at this point that responsibility for the support of the more applied programs usually passes to the individual academic departments or research centers which use the programs. At IIASA a similar situation exists at present: the various research areas have been responsible for developing and maintaining their own programs and data bases. For reasons discussed in the preceding section, a centralized IIASA program data library should take over the storage functions now being informally done by the research groups.

Hardware	Operating Systems	Assemblers Compilers	Math. Stat. Util.	Simulation Languages	ASA Components, Routines	Models Databases Applications
PDP 11/45	UNIX	APL	APEX	GPSS	WISSIM [16,17]	Energy
CDC CYBER 74	SCOPE	BASIC	MPX	DYNAMO	PAL [18]	Food
IBM 370/155	OS	FORTRAN	IMSL [14]	GASP	DEL [13]	Water
UNIVAC 1106	EXEC	PL/I	NAG	SIMULA	TAB [12]	Regional Planning
etc.	etc.	COBOL	SPSS	etc.	SYSOPT [19]	Environment
		etc.	SSP		etc.	Population
			OURLIB			Health
			BOYD			Settlements
			etc.			Managements
						etc.

CORE SUPPORT ——————————————————————————— FRINGE SUPPORT

Figure 1. Core and Fringe Areas of Software Support for Applied Systems Analysis in a Typical University Computer Center.

The types of modules to be stored in the library are *generalized subject-matter models, function-oriented components, technical routines, and scientific data bases.*

Generalized subject-matter refer to models constructed to handle sets of problems in a subject-matter area, such as agricultural sector analysis, energy consumption analysis, environmental systems analysis, etc. The model should be general enough to be useful in more than one specific application. Larger-scale models are often composed of sub-models, components, and sub-components, and usually cut across the traditional academic disciplines.

Function-oriented components focus on fundamental processes which occur in a number of different fields. For example, generalized input/output components could be used to handle production processes associated with the production of fish, livestock, crops, industrial goods, and public and private services. A generalized demographic component would model populations of people, cattle, trees, and even capital goods such as housing, tractors, and factories [8, 9]. Prey-predator processes could be modelled by a standardized function-oriented component. Accounting components could handle industry, sector, and natural accounting functions.

Technical routines can be considered as utility routines for performing specific operations often encountered in applied systems analysis. As already outlined in a previous section, examples of technical routines would include table look-up functions [12], delay routines [13], matrix operators [14], linear programming subroutines [15], simulation executive language [16, 17], conversational user/model interface language optimization routines for parameter estimation [9], graphics, etc.

Finally, the importance of access to machine-readable scientific data bases to support applied systems analysis in being recognized. Hand copying information from statistical yearbooks for punching onto data cards is a slow and inefficient way to access data. As a result, the processing of data can take a high percentage of the resources devoted to a systems modelling project. Data is used in applied systems analysis in three ways:

a) initializing model variables;

b) estimating model parameters; and

c) validating model structure by historical tracking.

The data usually is not aggregated to the level desired by the analyst to run his model. For example, most global models require aggregation of national-level data to the regional levels considered in the model. IIASA is exploring the

development of data bases along with data management systems to facilitate access to data and its utilization. For example, the Energy Program at IIASA under the WELMM Approach is developing data bases for natural resources, data bases for resource content of basic materials, goods, and processes, and data bases for energy chains. The data bases being developed are more complex than the simple rectangular data matrices (variables x observations) usually encountered in scientific research; structure is being built into the data to facilitate processing. Thus, inasmuch as there is a strong interdependence between data and models in scientific applications, structured data bases have an important place in a computer-based library.

LIBRARY STANDARDS

If the program/data library is to operate effectively and efficiently in carrying out its obligations of maintenance and distribution, and if the programs and data bases are to be applicable to the successful building of models in specific problem areas, then it is essential that a set of standards be defined as maintained. Although these standards should not be viewed as rigid, nevertheless they establish criteria against which to judge the admissibility and continued inclusion of library programs and data bases. The programs include models and components designed specifically to simulate processes and/or address problems related to the library's areas of application and utility routines which perform generally useful mathematical or logical operations. Each program is operationally realized in the library as both source code, object code, and written documentations. Standards are set separately for the selection of programs, for the documentations, and for the source code. Standards for data bases are not discussed in detail here but would follow along similar lines.

Program Selection Standards

Programs are evaluated for admission to, and continued inclusion in the library in three broad respects: *general applicability*, theoretical and *empirical validity*, and *compatibility* with other library software. Since quantitative or objective criteria cannot be set for some of these concepts, subjective evaluations often have to be made.

General Applicability. Library components are expected to be of general interest and applicability in IIASA's research areas. That is, specialized models designed to simulate processes and/or address problems unique to a particular country or region within a country will *not* be included in the library. Models built for specific countries or regions will be eligible for admission to the library after reworking to generalize their structure and programming and otherwise made to meet the library's standards.

Similarly, utility routines will be included if they are judged to be of general usefulness. These may include matrix operations, simplex algorithms, gradient search procedures, distributed and discrete delays, table look up functions, plot routines, etc.

Theoretical and Empirical Validity. Generalized models cannot be validated to the extent of faithfully representing a particular real-world system. Such a validation can only be made once the models have been adapted in the context of a specific application with its particular data. Rather, they are idealized representations of reality whose validity can only be determined with respect to internal logical consistency and conformity with accepted social and economic theory. Nevertheless, successful utilization of generalized models and components in specific application increases their credibility and perceived utility.

Compatability. Since components and routines will be combined in a variety of configurations, as required for specific applications, library modules are expected to be compatible with one another. Compatibility is insured and facilitated by the use of a standardized programming language, such as FORTRAN, a structured programming style, clearly specified and compatible (as to definitions and formats) inputs and outputs, and documentation illustrating how software combinations may be made.* Documentation and programming standards are discussed in more detail below.

Documentation Standards

The program/data library should maintain three types of documents for each program: announcement, abstract, and documentation. Format and content standards for each of these will be discussed separately.

---

*It is recognized that there are two approaches for linking modules to form larger programs. The first approach involves linking, prior to compilation, subroutines written in a standardized language such as FORTRAN. The modules communicate with each other via parameter lists, COMMON blocks, and auxiliary files. The second approach involves linking object-code programs which communicate through auxiliary files. In this latter case the use of a single standardized programming language and style is less important. The UNIX operating system on IIASA's PDP 11/45 computer facilitates this latter approach. Also, communication between programs located at different nodes in a computer network will also follow the second approach.

Announcement. Announcements are published and distributed--for example, in a newsletter or index--to actual and potential library users as a means of publicizing what the library has to offer. Thus, in most cases it will be on the basis of announcements that a user will decide, at least initially, whether certain software may be useful for his particular application, and, hence, whether to seek further information by requesting full documentation and/or consultation with the library staff.

An announcement, therefore, while it must be brief (a short paragraph), must include enough information so the above decision may be made. As a minimum, an announcement must give a verbal statement of what the software does, i.e. the problem or process simulated by the model or the utility function or algorithm performed by the routine, including any special features of the software which may distinguish it from other possible models of the same process or similar processes. In addition, examples may be given of possible uses of the software, e.g. larger models into which it may be embedded.

Abstract. Abstracts are intended to provide technical experts, analysts, and advisers with brief technical description of library software for use in preliminary evaluation and selection. In addition to a more technical presentation, where appropriate, of the information contained in the announcement, an abstract states the inputs required by the software, the outputs it supplies, and the core storage and sample run execution time for the computer on which it was tested.

Documentation. There are actually two audiences for which documentations are intended: consultants and analysts, and programmers. One effective way of meeting the needs of these distinct audiences is to compartmentalize the documentation in three clearly identifiable sections. The first section includes as much explanatory material which would be required in order to understand what the software does and where it may properly be used; the second goes deeper into the theoretical and mathematical detail required by the professional analysis and the third provides the information a computer programmer will need in order to successfully implement the program at the local computer installation. The documentation will be essential in (1) selecting appropriate software elements for a particular application, (2) validating models composed of a number of components, (3) analyzing the results of simulation experiments, (4) implementing software on a local computer in the context of a larger program comprising a model, and (5) making modifications to software as necessary to suit a specific need.

A discussion of the contents and formats of each of the three sections of the documentation and a sample set of documents for a demography component, are included in the *Software Standards Manual* prepared for the Computer Library for Agricultural Systems Simulation (CLASS) [20].

Programming Standards

The objectives of programming standards are to:

(1) maintain a consistent structured programming style

(2) maintain a compatibility among computer programs

(3) ensure and facilitate adequate error checking

(4) facilitate further developments

(5) enhance readability

(6) ensure as much machine independence as possible.

These objectives are set with the belief that their attainment will resolve a large number of problems which might otherwise be encountered by users of programs contained in the library. FORTRAN has been chosen as the standard language for CLASS and in general CLASS programming standards are compatible with ANSI FORTRAN. Detailed programming standards also appear in the CLASS Software Standards Manual [20].

Review Process

In order to assure completeness, accuracy and clarity, at least two reviewers check each program's set of documents. A computer-oriented reviewer evaluates primarily the documentation and the computer program itself in terms of:

(a) the accuracy and efficiency of the program as a computer model of the mathematical model described in the documentation and

(b) the sufficiency of the documentation for implementing the computer program.

The other reviewer, with expertise in the subject areas (disciplines) covering the problem or process addressed by the computer program, considers:

(a) the theoretical and empirical validity of the mathematical model described in the documentation and the suitability of the component as a whole, in terms of its intended applications;

(b) the sufficiency of the documentation for understanding the program and its intended applications (e.g. the completeness of the discussion of implicit and explicit assumptions of the model) and as a basis for deciding the suitability of the module for particular applications.

Data-base Standards

The utilization of structured data bases to support applied systems analysis is an emerging field. Consequently, a set of standards for data bases must be prepared which parallels the standards for programs with respect to selection, documentation, and structure of data files.

STATUS-LEVEL OF HOLDINGS

We have emphasized that the program/data library at IIASA should be a *working* library rather than an archival library. It should serve the needs of the scientists working in IIASA's research areas and integrative programs. Nevertheless, certain programs or data bases which have been developed at IIASA or initially developed elsewhere but brought to IIASA to be upgraded to meet the library standards will be identified as being of "export" quality. With this in mind, it may be helpful to categorize the library's holdings into three status levels: "state-of-the-art," "working-level," and "preliminary."

The "state-of-the-art" category would include the best products of IIASA's scientific research which would be of interest to the larger scientific community. Distribution of state-of-the-art programs also required a commitment by IIASA to maintain programs.

The "working-level" category includes holdings primarily utilized in IIASA's on-going research. They may include programs and data bases which are available elsewhere or are still considered to be in a developmental stage, and thus, not ready for the export-level category.

Finally, the "preliminary" category includes programs inventoried on an as-received basis. Documentation and programming may fall short of the standards required to make the items qualify for the working-level holdings.

INCENTIVES FOR CONTRIBUTORS

Why should a researcher take the extra time and effort required to clean up and document a program or data base for submission to the library? Past experience indicates that models are primarily developed as a one tool among many needed to complete a research report. Usually, there is little time or incentive to do the added work required to meet criteria of a well-documented program suitable for submission to a library for use by other researchers.

Certainly, once a formal program/data library exists, IIASA can administratively require that as a condition for

using computer service and programming support, researchers must provide at minimum an abstract describing the module(s) being developed. Also, when researchers leave IIASA, as one step in "clearing post" they should leave computer-based materials at the program/data library. This would insure that IIASA has at least obtained copies of "preliminary-level" software.

However, in order to encourage development of "working-level" and "state-of-the-art" programs and data bases, there must be professional incentives. As the reputation of IIASA's program/data library grows, the incentives will come from the professional recognition which the researcher receives from being a contributor to a library which maintains standards for admissability. Already, several other software libraries have reputations which provide these incentives. For example, having a program accepted into the Quantum Chemistry Program Exchange at Indiana University is equivalent in terms of professional recognition to having a scientific article published in one of the leading quantum chemistry journals.

USE OF THE LIBRARY

It is expected that the main use of the library will be in the application of the methodologies, which have been developed at IIASA and made operational in the form of computer programs, to analyze alternative strategies for the resolution of specific problems.

In keeping with the building block approach of applied systems analysis, the generalized function-oriented components, subject-matter models, and technical routines can be combined to form larger models for use in the investigation of a wide spectrum of specific problems. For example, to look at the problem of low productivity of agricultural resources in a country or region we might build an agricultural sector model utilizing a population model, a technological change component, an agricultural resource allocation model (using a recursive linear programming algorithm), an accounting component, linked together with an overall executive program. Selected data bases may be called to estimate model parameters and initialize model state variables. The model may be tested against historical data also contained in the data base. Finally, selected runs would be made to test alternative policies and strategies. In this way problem-oriented models can be built to examine a wide range of problems in the agricultural sector of any other sector.

Problem areas of concern to overall agricultural sector development, for example, fall in three main areas: economic--including resource allocation, resource productivity, inflation, income distribution, etc.; social--including health care, population, education, housing, law enforcement, etc.;

physical infrastructure--including power systems, communication systems, public works, etc. These are related, and overlapping fields; and analysis of a particular problem may require components and models from all of them.

Of course, the library is not likely to contain all the modules needed to analyze particular problems. New modules may have to be built, and existing modules may have to be modified to be of use. As new and improved modules become available, they will join the library, perhaps replacing modules already there. Furthermore, if certain modules fall into disuse for whatever reasons, they may be purged from the active library. Such operation is in keeping with the concept of a working library.

BASIC SERVICES

The program/data library will perform those basic services expected of any library: acquisition, cataloging, storage, maintenance, and distribution (comparable to circulation for a document library).

Acquisition

Models, components, routines, and data bases will be acquired by the library either through contribution, solicitation, or internal development.

In the beginning it is expected that the primary source of programs and data bases, which are contributed to the library, will be the products of IIASA's on-going internal and collaborative research projects. At a later date, after the library becomes more widely known as a repository of programs and data bases, they may be contributed by outside research projects.

After a periodic review of its stock or in response to a specific request, IIASA may actively solicit certain kinds of programs and data bases following a survey of the literature, outside research projects, and public and private research institutions. If suitable candidates are found, one or more is selected for further review. If the search is unsuccessful, the library may request that the development of the needed program or database be undertaken as an internal or collaborative IIASA research task.

Following initial receipt of programs and data bases, the library staff will review them for general applicability, validity, and compatibility, their potential contribution toward achieving IIASA's research objectives, and their conformity to the library's programming and documentation standards. The program will be tested using the supplied

test deck after the preliminary internal review, the submitted program or data base may be conditionally accepted into the preliminary status category. Then the item will be sent out for formal review as outlined earlier. During the internal and external review process it may be determined that the item does not meet the library's programming and documentation standards, but is otherwise acceptable. In this case, it must be decided whether to invest the additional resources required to bring the program up to standards given the library's limited resources for this activity (see later discussion).

Cataloging

Cataloging will take the form of adding the module to the library's published index of holdings. It is likely that this index will be stored on the computer for quick updating and retrieval through a standard computer-based information retrieval system. IIASA's Library already has an information retrieval system in operation, and the Documents and Publications Service will soon have one in operation. Since the text of the announcement for each item should appear as part of the index, the information retrieval system being developed for IIASA's publications series may be most suitable. DINFLO [2] might also be considered for this application. In any case, both services should be consulted on the best procedure for developing a computer-based index for the software/data library.

Storage and Maintenance

Library software and data bases exist in three forms: an abstract conceptualization (e.g. the mathematical model or data structure), a computer code (both source and object) or data file, and a set of documents. Physically, the items exist and are stored in the latter two forms. The media for storage are magnetic tapes and listings for programs and data bases and duplicated typescript, or in some cases machine storage, for documents.

The library must actively maintain the quality of the software and data base holdings by correcting and updating items as errors are discovered and enhancements made. Infrequently used and non state-of-the-art modules are purged in order to keep the library up to data as a working collection to serve the needs of users.

Distribution

Distribution will take several forms depending on the level of interest on the part of the user. Initial communication with the user will take place through the distribution of published indexes of programs and data bases similar in format

to the list of publications prepared and distributed by IIASA's Documents and Publications Service. This index will include sufficient description of the item for the user to tell if he would like to examine the complete set of documentation. (As indicated above, the index itself will be stored in an information retrieval system to facilitate library searching.) If the user then indicates he wants a program or data base, it will be copied onto tape or punch cards in the format requested by the user. To facilitate access by users at IIASA an auxiliary library system should be implemented in order to permit direct access to the library from a terminal.

ADDITIONAL SUPPORT SERVICES

To enhance the usefulness of the program/data library concept, other supporting services would include consulting, training, upgrading (by further generalizing and standardizing selected programs or data bases), and methodological development.

Consulting

Most libraries offer a reference service to aid users in locating materials. At the very least the program/data-base library should offer this service. But, because of the complexity of interlinking separate program components and utilizing data bases, it will be necessary to provide consulting on the best procedures for utilizing the holdings in actual applications. The consultants also serve as a point of contact for users who think that they have found errors in the programs. The consultants help to sort out users errors from program errors and report the latter back to those responsible for maintaining the library.

Training

Consulting with users on a one-to-one basis is an expensive way to communicate basic information on how to use effectively the contents of the library. Short courses and/or programmed audio-visual presentations are a more cost-effective way of communicating the basics. This is particularly true if IIASA wishes to encourage outside institutions and agencies to utilize the products of its research program.

Upgrading Programs and Documentation

As discussed earlier, sometimes programs and data bases submitted to the library may fall short of the standards established by the library, or else could be made more useful by further generalizing the structure of the program or data base. In such cases, a decision must be made as to whether to invest staff resources in upgrading the program or data base. Upgrading can be a costly activity and it is not likely that many items could be upgraded in this fashion. It is far more effective if the library staff work with programmers during the development of models and data bases in order to see that source code and documentation conform to accepted standards from the beginning -- an ideal which may not always work out in practice. Therefore, some effort at upgrading certain items will be necessary within the budget provided for that activity.

Methodological Development

Another important activity supporting the effective utilization of program data library is further methodological development of the basic software tools to be stored in the library. The concept of linking routines, components, models and data bases for application to analyzing particular problems in the social sciences shows promise, has seen limited applications, but remains to be fully demonstrated. It is consistent with IIASA's objectives both to extend the applications and to develop further the methodologies required to make the concept operational. This supporting methodological development should be carried out by Systems and Decision Sciences Group rather than by the staff of the program/data library.

Without further elaboration the following areas for further research and development may be identified:

1. Programming and documentation standards including use of structured programming techniques;

2. Design strategies for modularity and linkability of components including communication of variables;

3. Model/data-base interfaces including problems of updating projections and flexible aggregation of data;

4. Selection of languages for simulation;

5. Executive routines for structuring large simulation models including sequencing and overlaying;

6. User interface languages;

7. Specialized technical routines for applied systems analysis, e.g. gradient search methods, sensitivity testing and output routines.

8. Techniques for self-documentation of programs, e.g. automated variable directories.

STAFFING

To begin to carry out the various functions discussed above, the staff of the library should include at the minimum: a manager, a programmer, and a part-time secretary.

The library manager is responsible for coordinating the library's activities and maintaining quality control over the programs and data bases in the library and their associated documentation. He oversees and participates in the work of identifying needed programs and data bases, and acquiring, testing, storing, and upgrading (if necessary) them. He consults with users on their needs and how the library can serve them. He works with the professional scientists who are serving as consultants to the library.

At least one programmer is needed to develop and operate the auxiliary library system of programs and data bases on the computer and the information retrieval system for the index. He will test programs as they come in and review the documentation. He will consult with other programmers working in the various research groups to encourage conformity to library standards. He will prepare selected programs and data bases for export.

A secretary handles the clerical operations associated with the library, such as typing, filing and copying documentations, answering requests from outside of IIASA, perhaps even to the point of submitting standardized jobs to prepare tapes and cards.

Part-time editorial services would be useful in checking documentations, which have been prepared at IIASA, for their conformity to standards and their clarity. These services could probably be provided by the existing Documents and Publications Service.

Since the program/data library is designed to serve the needs of the research scientists at IIASA and elsewhere, there must be provision for input and feedback on the direction and priorities for the library staff work. The overall work of the library should be under the general direction of one of the senior scientists, perhaps either the Chairman or Deputy Chairman of the Systems and Decision Sciences Area. In addition a liaison scientist in each of the research areas and programs should be identified to coordinate the work of

the library with the work in the particular research area. Although this coordination could be done by setting up a formal advisory committee, it may be sufficient to allow for contact with liaison scientists as needed and allow the director and the manager to synthesize feedback from users regarding:

1. The quality and coverage of the models;

2. Needed revisions in existing programs and data bases;

3. Needs for new programs and data bases;

4. Quality and scope of library services.

It still might be useful to have at least an annual review of the library's operation by the liaison scientists.

Input regarding the more technical aspects of the library's operation, related to the methodological areas outlined above, can come directly from research scientists in the Systems and Decision Sciences Area.

The program/data library staff should also draw on IIASA's Documents and Publications Service for editing and publication of documentation for the programs and data bases. In addition, the program/data library staff should closely coordinate with IIASA's Library staff in utilizing techniques for cataloging, automated bibliographic searches, and information retrieval.

ORGANIZATIONAL LOCATION AT IIASA

There seem to be two alternatives regarding the location of the program/data library within IIASA's organizational structure. The library could be located as a sub-unit within Computer Services or as a new and separate scientific service.

The main argument for putting within the Computer Services is that the library's services are so computer dependent that the library naturally belongs with Computer Services. If this is done, the library should be a recognizable sub-unit within Computer Services with "operations" and "applications programming" perhaps being the other sub-units.

The main argument in favor of establishing the library as a new service is that the function of the library is so important and different enough from the function of Computer Services that it should be an independent unit. It could also be argued that the present Computer Services has a large enough task in providing basic computer service on four different machines and making preparations for providing network service.

Regardless of whether the library is a sub-unit within Computer Services or a new service, a line dividing responsibilities for the maintenance and service of software between the operations area of computer services and the program/data library should be made. (See Figure 1 cited earlier).

It should also be obvious that at the beginning of its operation, the library is not likely to have either the budget or staff to carry out the full range of basic and additional support activities discussed in this paper. These activities must grow as needs and experience warrant.

Nevertheless, a start must be made. An considerable contribution can be made just in gathering and cataloging what is currently available. Perhaps for the next several years it would be best to locate the program/data library as a clearly identified and budgeted pilot project within the System and Decision Science Area. This would leave open the question as to where in IIASA's organization structure a fully functioning program/data library might ultimately be located. At present, the methodological questions probably outweigh the service function required for a library with a limited inventory.

REFERENCES

[1] Introduction to IIASA Computer Services, September 1976.

[2] Schweeger, Bernhard. "The Small Size Multipurpose Information Retrieval System DINFLO." IIASA Software Library Series LS-1, January, 1976.

[3] Fischer, G. "Global Modelling, Part 1: WDYN, A FORTRAN Implementation of J.W. Forster's 'World Dynamics'." IIASA Software Library Series LS-2, August, 1976.

[4] Fischer, G. "Global Modelling, Part 2: World 3, A FORTRAN Implementation of D. Meadow's 'Dynamics of Growth in A Finite World'." IIASA Software Library Series, LS-3, August, 1976.

[5] Abkin, Michael H., and Carroll, Tom W. "Computer Library for Agricultural Systems Simulation: A Progress Report." CLASS Document No. 1, Computer Library for Agricultural Systems Simulation, Dept. of Agricultural Economics, Michigan State University, East Lansing, July, 1976.

[6] Abkin, Michael H. "Draft Proposal for an International Software Library for Applied Systems Analysis." Typescript, October, 1975.

[7] Holling, C.S. "Development and Use of Ecological Modules in Resource Development Simulation." IIASA Research Memorandum No. RM-73-3, October, 1973.

[8] Abkin, Michael H., and Wolf, C. "DEMOGC: Demography with Distributed Age Cohorts." Program No. DEM-1, Document No. 6, Computer Library for Agricultural Systems Simulation, Dept. of Agricultural Economics, Michigan State University, East Lansing, July, 1976.

[9] Abkin, Michael H., and Wolf, C. "DEMOGC: Demography with Discrete Age Cohorts." Program No. DEM-2, Document No. 7, Computer Library for Agricultural Systems Simulation, Dept. of Agricultural Economics, Michigan State University, East Lansing, July, 1976.

[10] Charpentier, Jean-Pierre. "A Review of Energy Models: No. 1 -- May 1974." IIASA Research Report No. RR-4-10, July, 1974.

[11] Charpentier Jean-Pierre. "A Review of Energy Models: No. 2 -- July 1975." IIASA Research Memorandum No. RR-75-35, July, 1975.

[12] Abkin, Michael H., C. Wolf, and Tom W. Carroll. "Table Functions: TABEL, TABEX, TABLIC, TABUX." Program No. UTR-2 and CLASS Document No. 9, Computer Library for Agricultural Systems Simulation, Dept. of Agricultural Economics, Michigan State University, East Lansing, July, 1976.

[13] Abkin, Michael H., and Wolf, C. "Distributed Delay Routines: DEL, DELS, DELF, DELLF, DELVG, DELLVG." Program No. UTR-1 and CLASS Document No. 8, Computer Library for Agricultural Systems Simulation, Dept. of Agricultural Economics, Michigan State University, East Lansing, July, 1976.

[14] International Mathematics and Statistics Library. Third edition, 7500 Bellaine Blvd., Houston, Texas, August, 1976.

[15] "DULPDX/DUPLX: Linear Programming Subroutines." The University of Wisconsin Computing Center, Madison, May, 1971.

[16] Kishline, P.D., and Buehring, J.S. "WISSIM...A Simulation Command Language Part 1: Users guide to WISSIM Commands." Energy Systems and Policy Research Report No. 9, Institute for Environmental Studies, University of Wisconsin, Madison, December, 1974.

[17] Buehring, J.S. and Kishline, P.D. "WISSIM...A Simulation Command Language Part II: Users Guide for Interfacing WISSIM and Simulation Models." Energy Systems and Policy Research Report No. 13, Institute for Environmental Studies, University of Wisconsin, Madison, December, 1974.

[18] Winer, C.S. and Wolf, C. "PAL: Policy Analysis Language." Program No. SML-1, Computer Library for Agricultural Systems Simulation, Dept. of Agricultural Economics, Michigan State University, East Lansing, July, 1976.

[19] Buchner, Marcus. "SYSOPT: Interactive Optimization Component." Program No. SPT-3, Computer Library for Agricultural Systems Simulation, Dept. of Agricultural Economics, Michigan State University, East Lansing, July, 1976.

[20] "Software Standards Manual." CLASS Document No. 2, Computer Library for Agricultural Systems Simulation, Dept. of Agricultural Economics, Michigan State University, East Lansing, July, 1976.