PATHOLOGY ANNUAL.

series editor SHELDON C. SOMMERS

NINETEEN SIX[illegible]

PATHOLOGY ANNUAL

VOLUME 1

1966

APPLETON-CENTURY-CROFTS
A Division of Meredith Publishing Company
NEW YORK

626-1

Library of Congress Card Number: 66-20355

PRINTED IN THE UNITED STATES OF AMERICA

M-83050

CONTRIBUTORS

Victor M. Areán, M.D.

Professor of Pathology, Chief, Division of Pathologic Anatomy, College of Medicine, University of Florida, Gainesville, Florida

Ronald S. Beckett, M.D.

Senior Pathologist, Hartford Hospital, Hartford, Connecticut

J.M.B. Bloodworth, Jr., M.D.

Professor of Pathology, University of Wisconsin; Chief of Laboratories, Madison Veterans Administration Hospital, Madison, Wisconsin

Jacob Churg, M.D.

Associate Clinical Professor of Pathology, New Jersey College of Medicine and Dentistry; Associate Attending Pathologist, The Mount Sinai Hospital, New York, New York; Pathologist, Barnert Memorial Hospital, Paterson, New Jersey

Charles G. Cochrane, M.D.

Associate Member, Division of Experimental Pathology, Scripps Clinic and Research Foundation; Established Investigator, Helen Hay Whitney Foundation, La Jolla, California

John M. Craig, M.D.

Clinical Professor of Pathology, Harvard Medical School; Pathologist-in-Chief, Boston Lying-in Hospital and Free Hospital for Women, Boston, Massachusetts

Samuel Dachs, M.D.

Director of Laboratories and Research, Montana Deaconess Hospital, Great Falls, Montana

Shirley G. Driscoll, M.D.

Associate, Department of Pathology, Harvard Medical School; Pathologist, Boston Lying-in Hospital, Boston, Massachusetts

Gilbert H. Friedell, M.D.

Lecturer in Pathology, Boston University School of Medicine; Instructor in Pathology, Harvard Medical School; Pathologist, New England Deaconess Hospital, and Research Associate, Cancer Research Institute, New England Deaconess Hospital, Boston, Massachusetts; Consulting Pathologist, Pondville Hospital, Walpole, Massachusetts

Michael S. R. Hutt, M.D., M.C. Path.

Professor and Head, Department of Pathology, Makerere University College Medical School, University of East Africa, Kampala, Uganda; Consultant Pathologist, Mulago Hospital, Kampala, and Ugandan Ministry of Health

Paul E. Lacy, M.D., Ph.D.

Mallinckrodt Professor and Chairman, Department of Pathology, Washington University School of Medicine; Pathologist-in-Chief, Barnes and Allied Hospitals, St. Louis Children's Hospital, St. Louis, Missouri

J.F.A. McManus, M.D.

Executive Director, Federation of American Societies for Experimental Biology, Washington, D.C.; formerly Professor of Pathology, Combined Degree Program in Medical Education, Indiana University, Bloomington, Indiana

Arnold W. Pratt, M.D.

Laboratory of Physiology, National Cancer Institute, National Institutes of Health, Department of Health, Education, and Welfare, Bethesda, Maryland

Leopold Reiner, M.D.

Visiting Associate Professor of Pathology, Albert Einstein College of Medicine; Director, Department of Pathology, The Bronx-Lebanon Hospital Center, Bronx, New York

Russell P. Sherwin, M.D.

Hastings Associate Professor of Pathology, University of Southern California School of Medicine; Member, Attending Staff Association, Los Angeles County Hospital, Los Angeles, California

Herbert B. Taylor, M.D.

Chief, Obstetrics, Gynecology and Breast Pathology Branch, Armed Forces Institute of Pathology, Washington, D.C.

Louis B. Thomas, M.D., F.C.A.P.

Head, Surgical Pathology and Postmortem Service, Department of Pathologic Anatomy, Clinical Center, National Institutes of Health, Bethesda, Maryland

PREFACE

The Pathology Annual Series is being created to remedy an important deficiency in the literature of pathology. It is intended to serve the practicing pathologist, who has the ultimate responsibility for diagnosis. Essays by experienced pathologists on subjects of their long-time special interest and competence will aid the practitioner in his daily job of detailed and accurate analysis of human disease, beyond what basic texts and journal reports provide.

Medicine has often had a tendency to stray from the patient. There have been eras of preoccupation with phenomena of the pulse, urine, galvanic magnetism and electrolytes. Modern pathology has the capability of refocusing attention on the patient, by exercise of the simple and important medical habit of thoughtful observation.

The publishers have extended themselves to assure prompt publication and pleasing quality. It is hoped that *Pathology Annual 1966* and those of succeeding years will provide useful, timely, and educational reading and reference.

Sheldon C. Sommers

PATHOLOGY ANNUAL

1966

CONTENTS

AN INFORMATION PROCESSING SYSTEM FOR PATHOLOGY DATA*

ARNOLD W. PRATT

LOUIS B. THOMAS

The value of data, in relation to the dynamic process of information exchange, must be measured in terms of quality, completeness, and availability. Data which are difficult to retrieve on a timely basis tend to have limited value irrespective of their excellence. This general statement is particularly applicable in medical science where in-depth teaching or investigative study is dependent on the chance availability of clinical material and the ability to recapture records containing related subject material. Experience has shown that patient data files organized solely by patient registry number are inadequate to serve the informational needs of medical scientists; these numbers serve only the routine archival requirements, i.e., establish order in the physical location of records. Communication among medical scientists is focused on disease or disease manifestations. This communication process can be enhanced if it is possible to identify and retrieve, rapidly and accurately, pertinent data records on the basis of any logical combination of primary data incorporated in the individual records. The primary data comprise any desired set of vital statistics, patient identifying data, and diagnostic data.

The pathologist is continually pressed to make data available. This is understandable. Pathology data arise from the direct examination of tissue and body fluids and are, therefore, indispensible in elucidating, identifying, and investigating

* The authors wish to thank Mr. Martin Epstein of the Computation and Data Processing Branch, Division of Research Services, National Institutes of Health, for technical assistance.

clinical disease and its manifestations. It is self-evident that the effectiveness of the pathologist, acting in his usual consultative role with the clinician, laboratory worker, student, or another pathologist, is importantly dependent on the availability of relevant data records [1] and has need of a responsive information processing system.

An important consideration in organizing an information processing system is the mode of data processing. Any specific system which seeks to be exhaustive in a technical area of science must provide for a minimal number of processing functions. They are:

1. Acquisition of the file documents
2. Identification and conversion of the information content of the documents via coding or indexing into a suitable form for mechanical processing
3. Physical storage of the documents or document images
4. Searches to determine the existence and identity of relevant documents
5. Physical retrieval of documents

It is obvious that some of these functions are closely interrelated. Physical storage and retrieval of documents are dependent on a single system approach, as are the information content coding and the search functions.

In most storage and retrieval applications, the needs are best served by some form of a machine-assisted system. The term machine in this context covers a wide spectrum ranging from the simplest of mechanical devices which require no additional skills for operation to electronic machines, namely, the general purpose digital computer, whose operation may require a number of additional skills. Ideally, the machine on which to base an information processing system is the digital computer. The computer is, in essence, a machine designed to follow a set of instructions rapidly and accurately. It is, however, a revolutionary departure from machines of the past which merely responded to the setting of a switch or the position of a lever; the computer responds to a language! With these computer languages, a user may instruct the computer to form, for itself, long sequences of instructions which can change the in-process mode of operation in intricate ways related to the character of the data and the requirements of the user. The computer languages now available provide a user with an immense capability in both range and diversity of information processing. Certainly, advances in the development of computer programming languages (software) and in the development of computer machinery (hardware) will continue.

Recognizing that the digital computer epitomizes the new, emerging information technology, it follows that the design of a general purpose information processing system for pathology data should incorporate the processing power of the computer. It does not follow, however, that all aspects of a general purpose information system should be exclusively computer-based; all pathologists do not have access to computers. Options in the data processing mode, i.e., punched card equipment or simple hand-sorting systems, must be allowed if the basic system is to provide for communication among pathologists. It is not implied, however, that the processing capabilities of the punched card system and the computer-based system can be totally equivalent.

NCI Information Processing System for Pathology Data

The National Cancer Institute information processing system for pathology data is a computer-based system* and includes a number of computer programs to provide for:

1. The creation and maintenance of a machine-stored data file of final diagnostic statements organized as patient-records;
2. The search and identification of a subset of patient-records based on any logical combination (and/or/and not) of subject matter data within a record;
3. The retrieval and printout of the complete data record(s);
4. The printout of summary tabulations pertinent to the subset of records identified by a search.

Three separate data files are maintained in the system, an autopsy data file, a surgical pathology data file, and a cytopathology data file. The format, informational content, and computer programs for processing the three data files are very similar, so that in the following discussion it will suffice to discuss the autopsy record processing almost exclusively. It is worth recording the size of the respective files in approximate numbers. The autopsy file contains 2,300 patient-records and 78,200 diagnostic data entries, the surgical pathology records number 25,000 with 62,500 diagnostic data entries, and there are 16,000 cytopathology reports with 20,000 diagnostic data entries. Coding of all diagnostic entries for the records in these data files was done using the Field Trial Edition [2] of "Systematized Nomenclature of Pathology" (SNOP).†

Input Data: Patient-Record Summary

A requirement for the organization of an information processing system is the identification of the documents which will compose the information file. It is essen-

* The current system utilizes Honeywell H-800 and H-200 equipment. The programs are written in Argus-enriched Fortran II programming language. For experimental investigation during the evolution of the system, less complete systems were organized using a small IBM 1620 digital computer and punched card equipment. Beginning in February, 1966, the NCI information system will be reprogrammed for the IBM 360 series, model 65; these programs will be made available to interested users.

† "Systematized Nomenclature of Pathology" was developed under the auspices of The College of American Pathologists, 230 North Michigan Avenue, Chicago, Illinois 60601, by the Committee on Nomenclature and Classification of Disease, Arthur H. Wells, M. D., Chairman. The Field Trial Edition of the "Systematized Nomenclature of Pathology" was distributed in January, 1963, to approximately 400 pathologists in the United States and in other countries so that the nomenclature could be tested for completeness and accuracy. More than 3,000 additions, corrections, and changes in the organization of the terms were suggested by the pathologists participating in the field trial. Most of these suggestions were used in the preparation and editing of the First Edition of SNOP which was published in January, 1965. The NCI system was established in 1963 based on the organization of the Field Trial Edition. The system is being revised and expanded using the new, First Edition of SNOP which has a larger vocabulary of terms. The basic organization of the information system as described here remains the same. This new nomenclature will be discussed in a following section, but it can be noted here that SNOP is, basically, a rather complete vocabulary of pathology terms and concepts organized in a manner which minimizes ambiguity and redundancy of the coded diagnostic data. It was created specifically to help pathologists organize and utilize their data files.

tial that the primary data in the file satisfy the purposes for which the system is intended: to yield, by an exhaustive and efficient file search, a subset of file documents (patient-records) which are precisely relevant to the search request. The pathologist's needs focus on identifying study material on the basis of primary data contained in the patient's record.

The standard autopsy protocol is not the ideal record; it is too detailed and usually restricted to terms descriptive of pathologic anatomy. The pathologist's concept of disease is considerably broader than pathologic anatomy. The more suitable document for the information file appears to be the face sheet of the autopsy protocol (Fig. 1). Customarily, the diagnostic information in the face sheet is presented as a number of final diagnostic statements incorporating information derived both from a consideration of the important clinical and laboratory data and from the gross and microscopic examination of tissue. These statements taken together represent the pathologist's terse but adequate summary of the disease and the disease manifestations found in an individual case. As used by the pathologists, these statements tend to be stylized and well ordered; they can be, and usually are, written to present a minimum of syntactical complexity.

There are other characteristics of the final diagnostic statements which are attractive to the documentalist responsible for organizing an information system. It is relatively easy to identify key words or combinations of key words which are used routinely as information content indicators in conjunction with surrounding words—for example, "metastatic to," "treated with," "due to." The information content terms which appear singly or in combination in the diagnostic statements, and which compose the vocabulary of pathology, are limited in number and have almost uniformly a highly specific denotation. It is possible, therefore, to construct an organized lexicon of the terms and concepts of pathology which can serve to identify and validate all data admissible to the information file. Given a lexicon of specific terms organized in some meaningful way, it is possible to assign codes which provide ease and economy of processing whether done by clerks or machines. Further, it is possible to impose a rudimentary syntactical form which insures completeness in coding a diagnostic datum. These considerations were fundamental to the development of the "Systematized Nomenclature of Pathology."

Systematized Nomenclature of Pathology*

Organization and classification of diagnostic data in the patient-records are conveniently done using the "Systematized Nomenclature of Pathology" (SNOP). This new nomenclature represents a special purpose language for pathologists and is composed of a relatively word-rich vocabulary and a primitive grammar.

The vocabulary consists of approximately 14,400 English language entries of specific pathology terms and concepts, each of which carries an assigned code of four alphameric characters. A term (or concept) is listed only once in the vocabulary and in addition, is assigned to one of four information fields:

* The codes for the pathology terms used in this section are taken from the First Edition of SNOP.

Standard Form 503
Revised August 1954
Bureau of the Budget
Circular A-32 (Rev.)

* U.S. GOVERNMENT PRINTING OFFICE 1961 O—582527

CLINICAL RECORD | **AUTOPSY PROTOCOL**

DATE AND HOUR DIED		DATE AND HOUR AUTOPSY PERFORMED		CHECK ONE		
10/13/58 3:10 P.M.		10/13/58 4:30 P.M.		FULL AUTOPSY	HEAD ONLY	TRUNK ONLY
PROSECTOR	ASSISTANT					
Rowan C. Boylan, M.D.				X		

CLINICAL DIAGNOSES *(Including operations)*

1. Choriocarcinoma with metastases in lungs, liver, adrenal glands, kidneys, and brain.
2. Recent basilar subarachnoid hemorrhage.

PATHOLOGICAL DIAGNOSES

1. Choriocarcinoma of uterus treated with nitrogen mustard, methotrexate, 2 - desoxy-glucose.
2. Metastatic choriocarcinoma in lungs, liver, brain, right adrenal gland, left kidney.
3. Hemorrhage in subarachnoid space and cerebellum (due to metastatic choriocarcinoma).
4. Organizing hemorrhagic infarct in region of pars intermedia.
5. Subpleural fibrocalcific granuloma in middle lobe of right lung probably due to Histoplasma capsulatum.
6. Bilateral obliterative pleuritis (due to metastatic choriocarcinoma).
7. Surgical absence of uterus, fallopian tubes, ovaries (March 1958).

Rowan C. Boylan, M.D.
Pathologist

Reviewed:

APPROVED—SIGNATURE

E. B. Price, M.D., Pathologic Anatomy Department

MILITARY ORGANIZATION (*When required*)	AGE	SEX	RACE	IDENTIFICATION NO.	AUTOPSY NO.
	35	F	W		A58-234

PATIENT'S IDENTIFICATION (*For typed or written entries give: Name—last, first, middle; grade; date; hospital or medical facility*)

REGISTER NO.	WARD NO.
02-04-77	

AUTOPSY PROTOCOL
Standard Form 503
503-104

Clinical Center, National Institutes of Health
NCI

Fig. 1. An example of a typical face sheet from an autopsy protocol. The patient's name has been deleted.

1. TOPOGRAPHY—the site of the body affected;
2. MORPHOLOGY—the structural changes produced compared to the normal;
3. ETIOLOGY—the causative agents or factors, such as microorganisms, drugs, chemicals, physical agents;
4. FUNCTION—the physiologic manifestations associated with diseases, including clinical signs and symptoms, and a limited number of specific infectious diseases and complex disease entities.

Implicit in the use of the four information fields is the assumption that disease can be described adequately by these four informational elements and that it is necessary to use the fields in combination to confer full descriptive specificity on any diagnostic finding. Using the fields in combination implies further, particularly with coded information, that the order of the reading of the fields is rigidly fixed, i.e., Topography, Morphology, Etiology, and Function in that order. All four fields must be considered in the coding and description of a single diagnostic datum. Restricting the use of any term to a single field has the advantages of minimizing the ambiguous use of a term, of identifying immediately the general information content of the term—e.g., a topographic site, a morphologic change—and of specifying the field in which the term must be coded by the human coders or the computer.

It is apparent then, that SNOP is not just another code. It is a special purpose language which, by its vocabulary, identifies all admissible primary data terms. Its organization specifies the syntactical form of any diagnostic message, first, by restricting a term to a single information field, and second, by fixing the order in which the four information fields must be read. The utility of SNOP is further enhanced in that the terms assigned to each information field have been organized into a hierarchically structured classification system; the position of any term in the classification system has been determined relevant to the meaning of the term in relation to all other terms assigned to that field. For example, the term "pulmonary alveoli" is listed within the generic subset "lung," which in turn is listed within the generic set "respiratory tract." All of these terms are assigned to the Topography field. It is obvious that the organization of the Topography field is based quite naturally on anatomic relationships.

The other information fields—Morphology, Etiology, and Function—are organized in analogous fashion using scientific principles or natural relationships, appropriate to each field, on which to base the arborized structure of terms. Thus, the organization of SNOP makes it possible to form almost any diagnostic statement and makes the coding of the complete diagnostic statement a straightforward matter. The four-character, alphameric codes assigned to each of the English language entries are simply synonyms; the structure of the code is dictated by the organization of terms within an information field and the combinational organization of the information fields needed to convey completeness of a diagnostic datum.

The organization of the first edition of SNOP consists of five numerical indexes: 2-Digit Topography Numerical Index, 4-Digit Topography Numerical Index, 4-Digit Morphology Numerical Index, 4-Digit Etiology Numerical Index, and 4-Digit Function Numerical Index. There are, in addition, four alphabetic indexes: 4-Digit Topography Alphabetic Index, Chemicals and Drugs Alphabetic Index, Enzyme Disorders Alphabetic Index, and a General Alphabetic Index.

The 4-Digit Topography Indexes provide for extensive anatomic detail; for example, colon is coded **T-6700** and descending colon is coded **T-6760**. They also include codes for secretions and cytologic material as purely anatomic terms to facilitate the processing of cytologic diagnoses. The Morphology Indexes include terms for structural alterations ranging from gross observations to intracellular, ultrastructural changes. The terms in Morphology have been organized in recognized classes of structural alterations as illustrated in section 4, Inflammation and Fibrosis, and sections 8 and 9, Neoplasms. For example **M-4000** is "inflammation, not otherwise specified."

To provide for certain adjectives commonly used in connection with inflammation and fibrosis, the second digit of the 4-digit code number has been used to designate acute, subacute, chronic, or granulomatous types of inflammation. Thus acute inflammation is **M-4100**. Similarly the third digit denotes specific adjectives, such as focal, diffuse, serous. Acute focal inflammation is, therefore, **M-4110**.

For neoplasms the fourth digit of the code number has specific meanings designating benign, primary, metastatic, and so forth. Code numbers also were assigned for the absence of morphologic alteration ("no pathologic diagnosis") and for the absence of residual tumor, since it might be desirable to code these negative diagnoses.

The 4-Digit Etiology Numerical Index includes pathogenic organisms, physical agents of injury, and a nonexhaustive list of therapeutic drugs and chemicals. The 4-Digit Function Numerical Index completes the pathology concept of disease. It allows recording of postmortem and antemortem laboratory findings, signs and symptoms, and other pertinent clinical observations together with the morphologic alterations and known etiologic agents.

The taxonomic structure of SNOP can be illustrated by the following examples. Inflammation of the colon is coded as:

6700 4000 0000 0000 colon, inflammation, NOS. The general morphologic change becomes site specific by specification of the topographic site.

The more specific information, *acute* inflammation of the *descending* colon is coded as:

6760 4100 0000 0000 descending colon, acute inflammation.

A specific etiologic agent such as *Salmonella typhosa* can be included by:

6760 4100 1361 0000 descending colon, acute inflammation due to *Salmonella typhosa*.

An associated clinical manifestation, for example diarrhea, can be added to the morphologic and etiologic data by:

6760 4100 1361 7225 descending colon, acute inflammation due to *Salmonella typhosa* with associated diarrhea.

Organizing a nomenclature into hierarchical sets of related terms and conditions allows retrieval of a related class of information. If a data file has not been organized so that all the specific terms can be found by retrieving the generic set, one must retrieve on each of the specific terms separately. There are many terms which connote pain. Some of these are pleurodynia, myalgia, lumbago, colic, tenesmus, ache, migraine, neuralgia, anginal pain, phantom limb pain, and, of course, the term "pain" itself. In SNOP these terms are organized in a hierarchical

group in the Function field and are assigned within the set of numbers **F-753***. **F–7530** is the number assigned to the word, pain NOS (not otherwise specified). Pleurodynia, mastodynia, myalgia, and lumbago are also given the number **F-7530** because each of these terms connotes pain of a particular body site. These site-

```
PATHOLOGY RECORD SEARCH

 20477  A58- 234

101358  035 F   W C
F CLINICAL CENTER, NIH


8300 8806 0000 0000 0  UTERUS, CHORIOCARCINOMA
8300 1510 0000 0000 0  ,TOTAL HYSTERECTOMY, MARCH 1958
8730 1510 0000 0000 0  ,BILATERAL SALPINGO-OOPHORECTOMY, 1958
YY00 0000 8302 0000 0  ,NITROGEN MUSTARD RX OF CHORIOCARCINOMA
YY00 0000 8315 0000 0  ,METHOTREXATE RX OF CHORIOCARCINOMA
YY00 0000 7000 0000 0  ,2-DESOXY-GLUCOSE RX OF CHORIOCARCINOMA
2800 8809 0000 0000 0  LUNGS, MET CHORIOCARCINOMA
5600 8809 0000 0000 0  LIVER, MET CHORIOCARCINOMA
X200 8809 0000 0000 0  BRAIN, MET CHORIOCARCINOMA
9300 8809 0000 0000 0  RT ADRENAL GLAND, MET CHORIOCARCINOMA
7100 8809 0000 0000 0  LT KIDNEY, MET CHORIOCARCINOMA
X390 3850 0000 0000 0  ,INTRACEREBELLAR HEMORRHAGE, MET CHORIOCARCINOMA
X180 3850 0000 0000 0  ,SUBARACHNOID HEMORRHAGE, MET CHORIOCARCINOMA
9103 5482 0000 0000 0  PARS INTERMEDIA OF PITUITARY GLAND, HEMORRHAGIC INFARCT
2830 4473 0000 0000 0  MIDDLE LOBE OF RIGHT LUNG-SUBPLEURAL FIBROCALCIFIC GRANULOMA
2830 0000 4315 0000 5  MIDDLE LOBE OF RIGHT LUNG, H CAPSULATUM PROBABLE
2901 4833 0000 0000 2  ,BILATERAL OBLITERATIVE PLEURITIS DUE TO TUMOR
```

Fig. 2. An output display of a coded autopsy record.

specific terms mean pain of the pleura, pain of the breast, muscle pain, and pain of the hip region, respectively. The number **F-7531** is assigned to hyperesthesia, **F-7532** to phantom limb pain, **F-7533** to anginal pain, **F-7534** to neuralgia, **F-7535** to causalgia, **F-7536** to ache, **F-7537** to migraine, **F-7538** to tenesmus, and **F-7539** to colic. Headache, earache, and toothache are other examples of site-specific terms and are assigned to the same number, **F-7536**, as ache. Thus, retrieval of the **F-753*** set recovers the information class "pain," which includes the above specific terms.

Within this organized format of coded information, it is possible to write an immense number of detailed and specific diagnostic messages, and for retrieval purposes it is possible to recognize each diagnostic message as a specific item within