

This work was written as part of one of the author's official duties as an Employee of the United States Government and is therefore a work of the United States Government. In accordance with 17 U.S.C. 105, no copyright protection is available for such works under U.S. Law.

Public Domain Mark 1.0

<https://creativecommons.org/publicdomain/mark/1.0/>

Access to this work was provided by the University of Maryland, Baltimore County (UMBC) ScholarWorks@UMBC digital repository on the Maryland Shared Open Access (MD-SOAR) platform.

Please provide feedback

Please support the ScholarWorks@UMBC repository by emailing scholarworks-group@umbc.edu and telling us what having access to this work means to you and why it's important to you. Thank you.

Framework for Scientific Communication in Earth and Space Science

G. Howard Burrows
Raytheon STX
7701 Greenbelt Road #400
Greenbelt, MD 20770 USA

Susan Hoban
University of Maryland
Baltimore County
Baltimore, MD 21201

Robert Harberts
Raytheon STX
7701 Greenbelt Road #400
Greenbelt, MD 20770 USA

Nand Lal
NASA Goddard Space Flight
Center
Greenbelt, MD 20771 USA

Abstract--We propose a framework for scientific communication that focuses on the value of the answers provided by science. The implementation would combine the advantages of a flexible index/thesaurus, a comprehensive taxonomy/ontology, and an annotation/discussion environment. Using our experience with the NASA Digital Library Technology program we will examine the social and economic issues involved in the acceptance of this new information technology by a community of scientists. We then speculate on opportunities provided existing publishers within this new model of scientific communication.

I. INTRODUCTION

Our intention is to design a framework for scientific communication that would make it easier to answer questions. The first step will be to develop a knowledge system that helps organize and capture the relationships between sets of questions. This would lead to a richer base of questions to provide the impetus for scientists to propose and evaluate possible answers. Validated answers provide value that can be used to fund future research.

A. What motivates the more formal framework that we propose?

Recent developments in information technology provide significant opportunities to increase the efficiency and productivity of scientific communications. The motivation for our proposal comes from our experience with an institutional Framework for Information Technology infusion into Earth and Space Science (FITNESS). Based on this experience we feel that effective communication involves three components: a knowledge base, active promotion and brokering, and an incubator in which the implications of new concepts can be tested.

Currently, the basic unit of communication of scientific results is the peer-reviewed journal article. The questions raised and the context for the answers provided are developed within the article itself, using scenarios and language specific to the immediate discipline of the study. Periodically, after a number of papers have been published, authors consolidate

the results and prepare a review article in which they suggest the areas of consensus and the areas still open to conjecture.

This process is results-based, with the focus starting from the detailed results of a specific study expanding to more general results. The process is slow and provides answers in a form that often makes it difficult for parallel comparison of alternative solutions. There is little room to report the many interesting approaches tried that have not yet produced positive results. The authors, subject to limited, usually confidential commentary by reviewers and publishers, choose which data to report and what issues to raise in their comments and analysis. Long periods of development occur without consolidation into a unified summary of current progress.

B. How will the new framework influence science?

We envision a framework for communication that starts with general questions and hones in on specific detailed studies through the use of a formal knowledge system coupled to a program of dialog and promotion in which the implications of new ideas can be studied. The knowledge system focuses from the onset on questions and their possible answers and consolidates additional information immediately as it becomes available.

Under this framework new incentives favor early reporting of interesting questions and early suggestion of the most plausible answers to be tested. There will be no advantage to selective reporting in order to withhold data for further analysis prior to publishing. Publishers will compete to develop structures that present the best status report for each family of questions providing different presentations for each community of users. It will be immediately clear which answers are generally accepted within a field.

II. FOUNDATIONS

The structure for a knowledge system with the desired properties is suggested by three well-known tools, an index/thesaurus, a taxonomy/classification system, and an enriched discussion environment for resolving conflict. The rich index allows people to jump quickly to the neighborhood of the information that they need. The underlying

classification system helps expand both the questions and the possible answers and establishes neighborhoods of related material along multiple dimensions. Finally, an expansive discussion environment provides consolidation and filters that assure that the most sensible, well-stated, and agreed upon conclusions are presented to the users first when they seek answers to their questions. Lessons for the new structure derive from the following three existing systems which already enjoy widespread, successful application.

A. Roget's Thesaurus

One hundred and fifty years ago Peter Mark Roget, then secretary of the Royal Academy of Science in London, retired to pursue an interest that he had had for forty years. Roget had been a public health physician, admitted to the Royal Society for a paper in which he described the slide rule as a mechanical device for calculation. Four years after retiring, he published his first Thesaurus of English Words and Phrases [2], a book he continued to expand into his late eighties.

Roget's interest with the Thesaurus was to build a new "Philosophical Language" with which to break down the "barrier... between man and man." His initial attempt was based on a broad taxonomy of human thought. The analysis and classification of our ideas was used to disambiguate words and help clarify expression. He originally thought that the taxonomy alone would be sufficient to help people navigate to the exact expression that would fit their need.

The lesson from this unified system is the provocative nature of a well-structured classification system. It makes transparent the relation between similar concepts. Its value is in consolidating ideas and establishing neighborhoods for clarifying and making precise the application of those ideas. In addition to the classification system which aids navigation between ideas, Roget added an index that facilitates rapid entry into the appropriate neighborhood for a family of ideas.

B. Ranganathan's Classification

Shiyala Ranganathan was a mathematician in India who early in his career was invited to become the University Librarian in Madras, India. Unhappy with the state of library science in India, he traveled widely and studied literature on classification ranging from the ancient Vedas to such modern philosophers as Kant and Hegel [3]. He incorporated the results of these studies and extended their work with twenty years of his own research. By 1948 he had developed a theory of classification so powerful and innovative that he was invited by the Rockefeller Foundation [4] to attempt to extend it into a formal language of communication. His Colon Classification system is still used widely in India and throughout many parts of the world.

In the Rockefeller Foundation study Ranganathan broadened his classification scheme to include the classification and cataloging of complex sentences. The power and flexibility of his system was based on the notions of "facet" and "phase" of topics. In facet analysis he established rules for unambiguously isolating and assigning an order to topics based on facets of time, space, energy, matter, and personality. Complex subjects composed of multiple topics could be classified by capturing the phase relationship as a type of loose assemblage of the component topics.

"Pre-natal" classification of materials was suggested to help authors establish the context of a work before they wrote it.

The lesson that we take from this work is the formal enriched taxonomy of ideas and the concept that communication and knowledge growth can be facilitated by this taxonomy.

C. Adler's Syntopicon

Mortimer Adler is a philosopher who became involved in the editing of the Encyclopedia Britannica. His gift was in capturing issues and establishing the relationship between them in annotated indexes. In 1948 he took on the job of selecting and promoting the Great Books of the Western World. He included such works as Shakespeare, Aristotle, the Bible, and works from more than fifty of the greatest authors.

He then raised nearly a million dollars and spent seven years analyzing the issues raised in the books he had selected. A team of over thirty scholars read through the books and extracted thousands of ideas that the books contained on index cards. By sorting these cards they established that 102 Great Ideas appeared in a preponderance of the books. A global index of all the ideas and a summary and outline of coverage of the 102 great ideas was published as the Syntopicon [5].

Immediately following the publication of the Syntopicon, Adler began a massive five year effort to expand and develop his social "dialectic" method of analysis in which all the issues surrounding a topic are raised and all sides of a controversy are revealed for comparison. The context for this study was the analysis of one of the Great Ideas, the idea of freedom [6]. His process is described in detail in the first volume of this work.

Adler developed such a strong faith in the organizational power of well-structured indexes, he determined to build the index first and use it to change the layout and apportion the content of a new edition of Encyclopedia Britannica. The categorical subject catalog called the "Propaedia" was first

used to select and allocate coverage to the topics in the Fifteenth Edition of Encyclopedia Britannica.

The lessons we derive from Adler's work add the social role to the corporate enterprise of structuring and elaborating the system of understanding. This social process is applied both in the selection of topics and in the dialectic expansion of issues that arise as we develop the implications and relationships between topics.

III. PROPOSED KNOWLEDGE SYSTEM

The system that we propose for scientific communication is modeled after the work of these three giants. We begin with the flexible "thesaurus" or treasury of related ideas and questions, after the model of Roget. Instead of cataloging and indexing only words and phrases, we would also focus on propositions and questions. The process is to capture the diversity of interest through global "brainstorming." An area for research is to create mechanisms that encourage broad participation of people from all walks of life and all geographic locations.

Ranganathan's deep analysis of classification can be applied to assure wide digression and elaboration of each idea. New technology allows us to relax the traditional librarians' requirement for ordering manuscripts in one linear array to fit the shelves in a library. Similar to using a card catalog for one dimension, readers can look for related questions and statements along the dimensions of any of the facets and phases in the catalog. New technologies, such as hypertext, could be used to facilitate this process.

Collaboration and group dialog will be modeled after the work of Adler's dialectic as established in the Great Books and extended in his group analysis of the concept of freedom. This will be used to establish the validity of the answers and to help with the market analysis that can be used to exploit the commercial value of the ideas. One can envision the application of new technologies to capture the facility of the index cards used in the creation of Adler's Syntopicon.

IV. SUMMARY AND IMPLICATIONS

We propose a framework for scientific communication based on a consolidated catalog of questions with an associated index of evaluated possible answers. Since validated answers have market value, we anticipate that this framework establishes an environment in which the value of answers provided by past science will produce revenue to support future science research. Since many of the physical processes involved in printing and shipping will be removed, there may be cost savings. The combination of these cost savings and the additional funds from the more efficient knowledge system could even generate funds that can be

applied for the additional pure or basic research necessary to maintain a constant flow of new answers for the future.

The framework that we propose provides a flexible social environment for rapid communication, both within and between populations of scientists and between the scientists and the people who can benefit from the answers. In this environment, publishers raise money from the inherent value of the current knowledge base. There will be additional market value when publishers re-apply the profits to buy more applied science for the future. The knowledge base not only reveals fundamental new questions, it also reveals possible answers and serves to advertise potential widespread implications of the answers affecting many users.

The economic model used with Encyclopedia Britannica where readers pay a one time entry fee followed by smaller payments for yearly updates might prove successful in the context of a scientific knowledge system. Since the framework provides the index and content for all questions, users would be expected to want frequent access as provided by a service in which users pay monthly fees. Alternatively, the science industry might follow the food industry with numerous vendors specializing in different forms of information. As in the food industry, the retail providers would be expected to channel funds to the farmers, ie publishers would be expected to fund research. Specialized answers, requiring new studies (cure my disease, save my crop, fix my TV) could be handled separately to allow efficient application and assembly of resources to validate the most valuable and promising among possible answers.

The increased role of social processes in the new framework might serve to bring many new types of people into the field of science. The current stereotypic scientist is intelligent, works alone with single-minded focus on one problem at a time. By enhancing the tools for collaborative research, this proposed framework opens science to other types of people. By focusing on the relationship between sets of questions and their possible answers, creative innovative people with ideas, but few skills to test those ideas can participate directly. Some technicians who know how to test the ideas, but who have little patience to speculate on new plausible answers, can find support for work on validating proposed answers. Clearer public understanding of the nature of validation and the implications of possible answers might encourage more people to contribute resources or even participate in the work of discovering and validating answers.

The role of the publisher in this new scenario involves first providing the software and skilled knowledge engineers for classifying, indexing, and brainstorming. This provides the structure of science. Publishers compete to develop many approaches to build the hierarchies that aid navigation and validation of possible answers. Publishers then provide the structure for communication between people of different

ages, backgrounds, and socio-economic situations. The publishers not only build to this knowledge base, they also provide the tools for the rich dialectic evaluation of divergent types of answers. In this way they actively promote answers with great implications and potential value. They will advertise and market viable, resilient concepts, rather than packages of information in the form of documents, journal articles, and monographs. From the results of the dialectic, areas of broad interest can be recognized and appropriate forms of the answers provided to meet the needs of different communities.

- [1] D.L. Embler, *Peter Mark Roget: the word and the man*. New York: Crowell, 1970.
- [2] P.M. Roget, *Thesaurus of English Words and Phrases*. New York: Grosset & Dunlap, 1937.
- [3] S.R. Ranganathan, *Prolegomena to Library Classification*. New York: Asia Publishing House, 1967.
- [4] S.R. Ranganathan, *Classification and Communication*. Bangalore: Sarada Ranganathan Endowment for Library Science, 1952.
- [5] M.J. Adler, *The Syntopicon: an index to the great ideas*. Chicago: Encyclopaedia Britannica, Inc., 1952.
- [6] M.J. Adler, *The Idea of Freedom*. Garden City, New York: Doubleday & Company, 1958.