Pergamon

Technovation 19 (1999) 593–604

technovation

www.elsevier.com/locate/technovation

# Science and technology innovation☆

Ronald N. Kostoff *

*Office of Naval Research, Arlington, VA 22217, USA*

Received 22 March 1999; received in revised form 15 April 1999; accepted 2 May 1999

**Abstract**

This paper describes two novel complementary approaches for systematically enhancing the process of innovation and discovery. One approach is workshop-based and the other approach is literature-based. Both approaches have the common feature of exploring knowledge from very disparate technical disciplines and technologies, and transferring insights and understanding from one or more disparate technical areas to another. It is highly recommended that the approaches be combined into a single process. The integrated approach has the potential to be a major breakthrough for the systematic promotion of innovation and discovery. Published by Elsevier Science Ltd.

*Keywords:* Innovation; Discovery; Complementary literatures; Disparate disciplines; Cross-discipline; Database tomography; Computational linguistics; Bibliometrics; Workshops

## 1. Introduction

The process of innovation is of immense social interest and impact, has been studied extensively, and yet remains poorly understood. There is a growing consensus that one important factor in many instances of innovation is the transfer of information and understanding developed in one or more disciplines to other, perhaps very disparate, disciplines. With the explosion in availability of information, scientists and technologists find it increasingly difficult to remain aware of advances within their own discipline(s), much less in other seemingly unrelated ones. As science and technology become more specialized, this cross-discipline transfer of information becomes more difficult. To overcome cross-discipline transmission barriers, systematic methods are required to heighten awareness of experts in one discipline to advances in other disciplines. This paper presents two different, yet complementary, approaches to increase cross-discipline knowledge transfer and provide the framework for enhancing innovation. Each approach represents a major advance in enabling innovation and discovery. The paper concludes that the synergistic combination of the two approaches would provide even greater potential for innovation and discovery compared to conducting the two approaches separately.

## 2. Interdisciplinary workshops for enhancing innovation

The first approach consists of convening workshop(s) of experts from different disciplines focused on specific central themes. The purpose of such a workshop is to achieve multi-discipline synergies and cross-discipline transfers to generate promising research directions for these central themes. The theory behind this approach is described in Appendix A. To test this theory, a workshop on Autonomous Flying Systems was convened in December 1997, and the implementation mechanics and results are described in detail in Appendix A.

The total workshop process consisted of three phases:

1. a two month pre-meeting e-mail phase in which each participant provided descriptions of advanced capabilities and promising research opportunities from his/her discipline to all other participants;
2. a two-day meeting at the Office of Naval Research (ONR) during which the promising opportunities

☆ The views expressed in this article are those of the author and do not represent the views of the Department of the Navy.

* Tel.: +1-703-696-4198; fax: +1-703-696-4274.
*E-mail address:* kostofr@onr.navy.mil (R.N. Kostoff)

0166-4972/99/$ - see front matter. Published by Elsevier Science Ltd.
PII: S0166-4972(99)00084-X

identified beforehand were discussed, crystallized, and enhanced; and
3. a post meeting e-mail phase in which each participant provided additional or embellished opportunities.

A number of important lessons were extracted from the conduct of this workshop, and they can be summarized as follows:
(a) The workshop approach broke new ground toward stimulating innovative thought. It was not easy, simple, or effortless, and required substantial planning and work in order to be effective. One should not throw people from 15 different disciplines together in a room for two days and hope to get new ideas synthesized. There needs to be a common generic thread woven through the different disciplines represented to spark the innovative thought process.

Interdisciplinary workshops, when performed correctly, are the wave of the future in defining new research (and technology) areas and approaches. Because of the intensity and effort involved throughout the process, they are most appropriate for large scale "grand challenges" in full-blown workshop form, but appropriate as well for smaller scale issues.
(b) Representatives from diverse technical disciplines, organizations, and development categories attended the workshop. There was substantial value in having a balance of discipline, category, and organization diversity at the same meeting. The different perspectives presented benefited all participants.
(c) Problem selection is crucial. The problem should be sufficiently general that many diverse disciplines can link to it. Given the choice of equally relevant problems, there is more potential for impact in selecting problem areas for which a large interdisciplinary community is not yet obvious.
(d) It is important to select participants by the most objective processes available. A combination of expert recommendation and strategic topical maps based on computational linguistics, publications, and citations was used for the selection process, and this approach produced highly knowledgeable individuals. Incorporation of the full literature-based approach to innovation in the discipline or participant selection process could further enhance confidence that the most appropriate mix of disciplines and experts has been chosen.
(e) It is extremely important that individuals selected for participation be world-class experts in their particular areas. There are relatively very few individuals producing the seminal works in any field (Kostoff, 1998, 1999), and it is these people who should be central to any truly innovative workshops. However, in addition to these established experts, highly competent individuals new to the field should also be selected. One benefit of transcending selection of known experts is that fresh faces who are new to established communities appear. They can sometimes challenge established paradigms and offer concepts typically not advanced through panels based solely upon well-known, over-used panelists.
(f) The e-mail component of the workshop is crucial. The gestation period between the input of promising ideas and their actual discussion at the workshop allows consideration of many different approaches and syntheses. It also saves substantial time at the workshop by clarifying confusing issues beforehand. However, in the first experience reported here, the stimulation of dialogue in the e-mail phase among most of the participants did not occur. The only participant to raise questions was the author, and this occurred only a few times. Nonetheless, in these instances, the dialogue was extremely valuable in clarifying issues and surfacing points of contention. In future workshops, it is strongly recommended that a few individuals representing different disciplines be asked to assume a role of facilitator, with the task of stimulating dialogue and raising questions during the workshop build-up phase.
(g) All the attendees at the workshop were required to participate; there were no pure observers. This meant that they had to submit accomplishments and opportunities statements by e-mail. They also had to be prepared to lead discussions at the workshop. This participation requirement was valuable in that each attendee obtained a sense of ownership in the workshop and its outcome. His/her contribution tended to be more substantive and creative than is typically the case at standard workshops. Those who contributed more in the e-mail phase tended to contribute more in the workshop phase. In addition, there was a sense of equality among participants when all were required to contribute, as opposed to an audience/performer environment with passive onlookers. On the other hand, the downside of requiring attendees to be active participants was that attendance had to be limited. This may not be totally bad, since audience participation can substantially enrich workshop discussion.
(h) In general, there needs to be some incentive to motivate participation of world-class experts in these workshops. Unless they are able to envision some type of substantive impact resulting from their participation, either on larger science and technology (S&T) issues or in their individual disciplines, they could be reluctant to invest the substantial amount of time required for serious participation. This, however, did not turn out to be a problem for the Autonomous Flying Systems workshop, apparently because of the limited size of the field and the interest of the participants in the type of workshop conducted.

In addition, during the workshop, participants did not appear to have reluctance in sharing new concepts. This is in stark contrast to some workshops the author has attended where novel ideas were held very closely. In the Autonomous Flying Systems workshop, there was a

spirit of comeraderie and cooperation that pervaded the proceedings, and helped overcome the barriers to sharing. This spirit was fostered in the pre-meeting e-mail dialogue phase, and further nurtured during the meeting by having all attendees participate in the proceedings as equal partners.

Finally, interdisciplinary workshops are a powerful potential source of radically innovative ideas if conducted properly. There are three central requirements for success:

1. A problem of significant interest to the sponsoring organization must be selected;
2. An optimal mix of world-class experts appropriate to the problem must be chosen;
3. Conditions must be created which will motivate the participants to share their novel concepts.

The Autonomous Flying Systems workshop addressed these three requirements to a significant degree. A preliminary concept proposal emerged, and a copy of this proposal is available from the author.

## 3. Accessing linked literatures for enhancing innovation

The second approach consists of examining relationships between linked, overlapping literatures, and discovering relationships or promising opportunities not obtainable from reading each literature separately. The general theory behind this approach, applied to two separate literatures, is based upon the following considerations (Swanson, 1986).

Assume that two literatures can be generated, the first literature AB having a central theme “a” and sub-themes “b”, and the second literature family BC having a central theme(s) “b” and sub-themes “c”. From these combinations, linkages can be generated through the “b” themes which connect both literatures (e.g., AB→BC). Those linkages which connect the disjoint components of the two literatures (e.g., the components of AB and BC whose intersection is zero) are candidates for discovery, since the disjoint themes “c” identified in literature BC could not have been obtained from reading literature AB alone.

Some initial applications of the second approach have been published in the medical literature. One interesting discovery was that dietary eicosapentaenoic acid (theme “a” from literature AB) can decrease blood viscosity (theme “b” from both literatures AB and literatures BC) and alleviate symptoms of Raynaud’s disease (theme “c” from literature BC). There was no mention of eicosapentaenoic acid in the Raynaud’s disease literature, but the acid was linked to the disease through the blood viscosity themes in both literatures. Subsequent medical experiments confirmed the validity of this literature-based discovery (Gordon and Lindsay, 1996). (A web site (Swanson and Smalheiser, 1997), overviews the process used to generate this discovery, and contains software that allows the user to experiment with the technique. A 1998 article in The Scientist outlines perceptions of different knowledgeable individuals on Swanson and Smalheiser’s general technique (Finn, 1998)).

The main body of the present paper generalizes and expands upon the literature-based approach, using the Database Tomography (DT) techniques and experience developed since 1991 (Kostoff, 1993, 1994). It outlines the theory of the expanded approach, the implementation details, and overviews the range of applications possible with this technique. However, the present paper differs from previous work in four major ways:

1. The paper’s literature-based component utilizes the phrase frequency and phrase proximity algorithms from DT on any type of full text (free-form or structured) to identify literature themes, associated themes, and the relationships among the themes and associated themes. While the present paper’s focus is the science and technology literature, the DT-based technique is applicable to any type of literature.
2. The paper’s literature-based component generates discovery for more combinations of independent and dependent themes than previous literature-based work (seven combinations in the present paper as compared to two combinations in Swanson and Smalheiser (1997) for the same baseline two literature approach, AB/BC).
3. The workshop-based component (Appendix A) has a more systematic approach than standard workshops for identifying appropriate experts from different disciplines focused on the central workshop theme, and recommends an e-mail-based pre-meeting dialogue.
4. Perhaps the major conceptual breakthrough in this paper for stimulating and enabling discovery and innovation is the recommended coupling of the literature-based approach to the workshop-based approach. Performing these two approaches in tandem:

4.1. eliminates the weaknesses resulting from conducting the two approaches in isolation (e.g., subjective biases of the workshop participants in the absence of systematic literature data, and insufficient participant diversity and scope/breadth of sources in the literature-based approach),

4.2. emphasizes the combined strengths of the two approaches, and

4.3. produces a synergy where the whole (conducting the two approaches in tandem) is greater than the sum of the parts.

## 4. Need for literature/workshop synergy

In the larger technical community, there are a number of separate approaches for enhancing discovery and innovation. While many of these approaches focus on financial or other incentives, there are two main structured approaches aimed at stimulating innovation, workshop-based and literature-based. Most organizations use some variant of a workshop/group dynamics approach. The attendees and participants in these groups tend to be focused subject experts; there is rarely any complementary sophisticated literature analysis performed, and there are rarely experts present from strongly divergent disciplines. While widely used, the workshop techniques tend not to make full use of many of the information technology advances of recent years.

A few performers, individuals or small groups of individuals, pursue the literature-based computer-assisted approach. This literature approach tends to be more sophisticated and technologically advanced than the workshop approach. However, it has not received widespread attention and may fall short of the interpretive and analytical strengths of the workshop approach. As a result, the literature approach is rarely used (e.g., Finn, 1998).

While either the workshop approach or the literature approach can be done independently to help stimulate discovery, they should be done in tandem to maximize the benefit provided by each. There is nothing on record to indicate that this joint approach to innovation has been implemented, or even considered. In a joint workshop–literature effort, the literature approach would be included in the background pre-meeting phase of the workshop approach (as developed in Appendix A). Accordingly, the literature study would provide:

1. background reading for the workshop participants in related yet disparate science and technology areas;
2. strategic maps of the broader science and technology literature as outlined in the DT papers referenced above;
3. promising opportunities for innovation and discovery; and
4. the disparate science and technology disciplines from which the experts for the workshop could be drawn.

## 5. Literature approach

### 5.1. Overview

The theoretical basis of the literature approach mirrors the scientific process in many ways. Information from diverse literatures, with relevant interfaces, is examined. All information is first analyzed and then synthesized to produce discovery and innovation. Myriad pathways through many different literatures can be examined. These pathways are probably necessary, to maximize benefits, for the complex relationships that constitute the real world. Initial work (Swanson, 1986; Gordon and Lindsay, 1996) examined three variable classes or themes (c, b, a) in two literature categories (C and B) using two different approaches (start with “c”, determine “b”, then determine “a”; start with “c” and “a”, then determine “b”).

NOTE: The sequence abc will typically (but not always) represent a time-varying process, such as a procession from research to development to systems. Where this sequence does represent a temporal process, the convention used in the remainder of this paper is that the alphabetical designation of variables follows the arrow of time. Thus, “a” might represent a research variable or theme, “b” might represent a technology variable or theme, and “c” might represent a system variable or theme. The terms “variable” and “theme” are treated as interchangeable; “thematic variable” is used in places to emphasize this congruence.

The principal thematic variables determine a thematic literature. From the previous example, if Raynaud’s disease is the thematic variable specified initially, then the corresponding thematic literature might be all the papers in a given database that contain the phrase Raynaud’s disease. The remaining thematic variables and literatures are determined by applying different algorithms to the initial thematic literature and subsequent derived literatures. Again from the previous example, an algorithm would be applied to the Raynaud’s disease thematic literature to determine the thematic variable blood viscosity, and a derived literature could then be determined as all the papers in a given database that contain the phrase “blood viscosity”.

The first approach in the initial reported work (Swanson, 1986; Gordon and Lindsay, 1996) could be viewed as addressing the question: What variables “a” could influence variable “c” through mechanisms “b”, or, in the example described above, “What treatment factors ”a“ could influence Raynaud’s disease ”c“ through the different mechanisms ”b“. This approach started with thematic variable ”c” (e.g., Raynaud’s disease), and used this variable to develop thematic literature C. Algorithms were applied to this thematic literature database to identify thematic variable “b” values (b1, b2, etc. representing characteristics such as blood viscosity, blood flow, blood platelets, poor circulation, and others) closely linked to thematic variable “c”. Each value or theme of variable “b” (b1, b2, etc.) was used to develop a thematic literature B1, B2, etc. Algorithms were applied to each of the thematic B literatures to identify thematic variable “a” values (a1, a2, etc. representing characteristics such as fish oil, eicosapentaenoic acid, and others) closely linked

to the specific thematic variable "b" of each thematic B literature. Values of the thematic "a" variables in each of the thematic B literatures not found in thematic literature C defined a subset of the thematic B literatures that was disjoint from thematic literature C (e.g., the term "fish oil" was not found in the Raynaud's disease literature). These disjoint thematic "a" variables and their associated thematic B literature subsets became candidates for discovery and innovation.

The other approach reported could be viewed as addressing the question: What are the mechanisms "b" through which variable "a" could impact variable "c". This approach started with variables "c" and "a", and their associated literatures C and A, and identified variables "b" that were linked to both variables "c" and "a". The same types of algorithms as in the first approach were used to identify closely linked variables, and the requirement for disjointness between literatures C and A was used as a basis for discovery.

From the experience of these two approaches, it becomes clear that the independent and dependent variables chosen, and the algorithmic approach selected, depend on the question being asked. Further examination shows that other approaches beyond these two are possible to answer other questions. The present paper examines seven approaches to generate innovation and discovery that are structured to answer seven different questions, and shows how the algorithms and techniques developed in Database Tomography are used in these approaches.

For readers unfamiliar with the mechanics of Database Tomography, the two most recent references are recommended (Kostoff, 1998, 1999).

### 5.2. Specific approaches

The following discussion will be limited to scenarios of three variables "a", "b", "c", and two literatures. In future studies, more complex cases could be candidates for analysis and experimentation.

For the simple two literature/three variable case, seven separate generic cases are possible, where the variables specified can be viewed as "independent" and the variables determined can be viewed as "dependent":

1. specify "a", determine "b" and "c";
2. specify "c", determine "a" and "b";
3. specify "b", determine "a" and "c";
4. specify "a" and "c", determine "b";
5. specify "a" and "b", determine "c";
6. specify "b" and "c", determine "a";
7. specify "a" and "b" and "c", validate linkage existence.

Cases (1), (2), and (3) are the most open-ended and least constrained. In each case, one variable is specified, and the other two are determined using the DT algorithms, the condition of disjointness and, most importantly, expert judgement. Cases (4), (5), and (6) are more constrained, since two variables are specified, and the third is determined using similar processes to the above. Case (7) is fully constrained, and its purpose is to ascertain literature support for validation of a hypothetical relation between specified values of the three variables. Cases (4) and (5) are subsets of case (1); cases (4) and (6) are subsets of case (2); cases (5) and (6) are subsets of case (3); Case (7) is a subset of cases (1) through (6). The solution mechanics for each of these seven cases will now be outlined.

#### 5.2.1. Opportunity driven

This first case addresses the question, "What are the potential variable 'c' impacts which could result from variable 'a,' and what are the variable 'b' mechanisms through which these impacts occur?" One specific variant of this question is of particular interest and importance to the science and technology community, "What are the potential impacts on research, development, systems, and operations that could result from research on a given topic?"

If the generic question of this first case is applied to the above example for the case where variable "a" is "fish oil" only, it could be phrased as, "What are the potential impacts or benefits (positive or negative) resulting from fish oil that would not be obvious from examining the fish oil literature alone?" This is an open-ended question, and places no restrictions on the mechanisms "b" or the types of impact "c". The first case is represented schematically as:

$$a \longrightarrow b \longrightarrow c$$

Here, "a" is the independent variable, and "b" and "c" are the dependent variables that result from the solution process. The operational sequence is to start with the variable "a" and generate a literature A. Again following the above example and using the abbreviations FO (fish oil), BV (blood viscosity), and RD (Raynaud's disease), this means that the process would start by identifying the FO literature (call this A1). Many approaches could be used to define this literature; the approach recommended here is the one used in recent DT studies (Kostoff, 1997, 1998) for defining literatures. As an example of one literature definition approach, the iterative Simulated Nucleation method (Kostoff, 1997) would be used to identify all the papers in the Science Citation Index (SCI) which contained FO (and other related terms in the query) in the title, keywords, and abstract fields. This collection of papers would constitute the FO literature.

NOTE: Use of the SCI is one example only. Because DT uses full text databases, there is no limitation in any database selected to titles or key words or index words,

and many different types of databases or free text can be used for the analysis.

The next step in the process is to identify the variables “b” (b1, b2, . . .) linked closely to variable “a1”, and then identify the literatures B associated with variable “b” (B1, B2, . . ., the BV literatures). For this step, the proximity analysis method used in the recent DT studies (Kostoff, 1998, 1999) would be employed. For a journal based database, this method conceptually identifies phrases in paper titles or abstracts or main texts physically located near the term of interest. As an example, if the term of interest in a given database is Raynaud’s disease, then the proximity analysis method would provide a list of all phrases in close physical proximity to the term Raynaud’s disease for all occurrences of this term in the text. The proximity analysis approach of DT is based on the experimental findings that phrases within a semantic boundary (same sentence, paragraph, etc.) which are located physically close to the term of interest are contextually and conceptually close to the term of interest. Continuing the above example, this step uses the proximity analysis of DT to identify phrases in the FO literature physically close to the term FO, such as “b1”, “b2”, etc.

For each of these identified phrases “b1”, “b2”, etc., a literature (B1, B2, . . .) is established by querying the SCI. The next step is, for each of these B literatures, to identify the linked variables “c” (c1, c2, . . .) The process used to identify the variables “b1”, “b2”, etc. linked to variable “a1” is repeated to obtain the variables “c1”, “c2”, etc. linked to each value of variable “b”. The subsets of the B literatures which are disjoint from literature A1 (e.g., the B literatures which don’t contain the term FO) must then be identified, and the variables “c” (and their associated linking mechanisms “b” to variable “a1”) within these disjoint B literature subsets then become the candidates for discovery and innovation.

It is obvious that the process can easily mushroom out of control unless stringent limiting constraints are placed on the number of B literatures and “c” variables selected. For example, suppose that three “b” variables “b1”, “b2”, “b3” (and their associated three B literatures (B1, B2, B3) are identified as closely linked to FO. Suppose also that each of these three “b” variables is closely linked to five “c” variables. Then four literature searches are required (A1, B1, B2, B3), and fifteen abc linked pathways must be examined for disjointness and discovery, according to the following:

a1→b1→c11

a1→b1→c12

a1→b1→c13

a1→b1→c14

a1→b1→c15

a1→b2→c21

a1→b2→c22

a1→b2→c23

a1→b2→c24

a1→b2→c25

a1→b3→c31

a1→b3→c32

a1→b3→c33

a1→b3→c34

a1→b3→c35

In reality, there will be hundreds, if not thousands, of candidate “b” and “c” variables. However, there are different ways by which the “b” and “c” variables can be sharply limited in number. First, the analysts performing the study would eliminate all non-technical content phrases which passed through the trivial word filter in the DT algorithm. Second, the numerical indices for each phrase generated by the DT proximity algorithm would be used as one figure of merit for pre-selection of key phrases. Third, those “c” variables which reappear in different abc pathways would have a higher priority for selection. Fourth, analyst judgement would be applied to weight the potential value of the different abc pathways in computing figures of merit.

The literature searches and proximity analyses are fairly straightforward, and have been refined in the DT process. The main intellectual efforts must be focused on prioritizing and reducing the number of linked variables or literatures to be examined, and interpreting the relationships among the final disjoint literatures to generate potential discovery relationships.

### 5.3. Requirements driven

This second case addresses the question, “What are the variables ‘a’ that could impact variable ‘c,’ and what are the variable ‘b’ mechanisms by which these impacts are produced?” Applied to the above example for the case where “c” is Raynaud’s disease only, it could be phrased as “What are the factors and their associated mechanisms that could impact the course of Raynaud’s disease that would not be obvious from examining the Raynaud’s disease literature alone?” This second case is represented schematically as:

a←b←c

Here, “c” is the independent variable, and “b” and “a” become the dependent variables. The operational sequence is to start with variable “c”, and generate a literature C. Again following the above example, this

means that the process would start by identifying the RD literature (call this C1). The same literature definition process as in the first case would be used. The next step would be to identify the linked variables "b" (b1, b2, etc.) to variable "c1", and then their associated literatures B (B1, B2, the BV literatures). For this step, the proximity analysis method used in the recent DT studies would be employed again as in the first case. Continuing the above example, this step uses the proximity analysis of DT to identify phrases in the RD literature physically close to the term RD, such as "b1", "b2", etc. For each of these identified phrases b1, b2, etc. a literature (B1, B2, etc.) is established by querying the SCI. The next step is, for each of these B literatures, to identify the variables "a" (a1, a2, etc.) linked to variable "b". The process used to identify the variables "b1", "b2", etc. linked to variable "c1" is repeated to obtain the variables "a1", "a2", etc. linked to each value of variable "b". The subsets of the B literatures that are disjoint from literature C1 (e.g., the B literatures which don't contain the term RD) must then be identified, and the variables "a" within these disjoint B literature subsets (and their associated linking mechanisms "b" to variable "c1") then become candidates for discovery and innovation. The same stringent limits on variables and literatures used in the first case are applicable here.

### 5.3.1. *Mechanism driven*

The third case addresses the question, "For a given mechanism 'b', what are the variables 'a' that could impact the variables 'c'?" Applied to the above example for the case where "b" is blood viscosity, it could be phrased as, "What combinations of variables that could effect a change in the blood viscosity mechanism and could be impacted by a change in the blood viscosity mechanism are candidates for discovery that were not obvious from examining only the blood viscosity literature?" The third case is represented schematically as:

$a \leftarrow b \rightarrow c$

Here, "b" is the independent variable, and "a" and "c" are dependent variables. The operational sequence starts with variable "b", and generates a literature B. Again following the above example, this means that the process would start by identifying and generating the BV literature (call this B1). The same literature definition and generation process as in the first case would be used. The next step would be to identify the variables "a" (a1, a2, etc.) and "c" (c1, c2, etc.) linked to variable "b1", and then their associated literatures A (A1, A2, the FO literatures) and C (C1, C2, the RD literatures). For this step, the proximity analysis method used in the first two cases would be employed for the BV literature (B1). Continuing the above example, this step uses the proximity analysis of DT to identify phrases in the BV literature physically close to the term BV, such as "a1", "a2", etc. (FO literature) and "c1", "c2", etc. (RD literature). However, an arbitrary step is required at this point, since the proximity analysis only provides the aggregate of the linked variables "a" and "c". The analyst is required to divide the aggregate linked variables obtained from the proximity analysis into two groups, "a" variables and "c" variables. In the above example, the proximity analysis would generate the linked variables such as fish oil and Raynaud's disease. The analyst would be required to specify two categorizations for these variables, such as "dietary factors" for the "a" variables and "diseases" for the "c" variables. This step will depend heavily on the analyst's expertise in the technical area and ability to create taxonomies.

The next step is to identify/generate the A and C literatures using the approach described above. The final step is to identify the subsets of the A literatures and C literatures that are disjoint. Each group of articles from the A literature and the C literature that contains a "b1" variable is considered to be a linked group. The subsets of these literatures that are linked through the common "b1" variable and that are disjoint (i.e., the C literature does not contain the "a" variable and the A literature does not contain the "c" variable) must then be identified. The variables "a" and "c" within these disjoint A and C literature subsets linked through the "b1" variable then become the candidates for discovery and innovation. The same stringent limits on variables and literatures used in the first approach are applicable here.

### 5.3.2. *Opportunity-requirements driven*

This fourth case addresses the question, "What are the mechanisms 'b' through which variable 'a' could impact variable 'c'?" Applied to the above example for the case where "c" is Raynaud's disease only, and "a" is fish oil only, it could be phrased as, "What are the mechanisms through which fish oil could impact Raynaud's disease that would not be obvious from examining only the Raynaud's disease literature or the fish oil literature?" The fourth case is represented schematically as:

$a \rightarrow b \leftarrow c$

Here, variables "a" and "c" are independent, and variable "b" is the dependent variable. The operational sequence is to start with the variable "c", and generate a literature C, and with variable "a", and generate a literature A. Again following the above example, this means that the process would start by generating the RD literature (call this C1) and the FO literature (call this A1). The same literature definition and generation process as in the first case would be used. The next step would be to identify the linked variables "b", and then their associated literatures B for both the A1 literature and the C1 literature. For this step, the proximity analysis method used in the first two approaches would be employed, for the FO literature (A1) and the RD literature (C1). Continuing

the above example, this step uses the proximity analysis of DT to identify phrases in the RD literature physically close to the term RD, such as “b1”, “b2”, etc. and to identify phrases in the FO literature physically close to the term FO, such as b51, b52, etc. The next step is to identify the subsets of the A1 literature and C1 literature which are linked. Each group of articles from the A1 literature and the C1 literature that contains a “b” variable is considered to be a linked group. The subsets of these literatures linked through the common “b” variables that are disjoint (i.e., the C1 sub-literature that does not contain the “a1” variable and the A1 sub-literature that does not contain the ‘c1’ variable) must then be identified, and the variables “b” within these disjoint A1 and C1 literature subsets then become the candidates for discovery and innovation. The same stringent limits on variables and literatures used in the first case are applicable here.

*5.3.3. Opportunity-mechanism driven*

The fifth case addresses the question, “What are the variables ‘c’ which could be impacted by variable ‘a’ through mechanism ‘b’?” Applied to the above example for the case where “b” is blood viscosity only, and “a” is fish oil only, it could be phrased as, “What abnormalities could be influenced from the impact of fish oil on blood viscosity that would not be obvious from examining only the abnormality’s literature or the fish oil literature?” The fifth case is represented schematically as:

a⟶b⟶c

Here, “a” and “b” are the independent variables, and “c” is the dependent variable. The operational sequence is to start with the variable “a”, and generate a literature A, and with variable “b”, generate a literature B. Again following the above example, this means that the process would start by generating the FO literature (A1) and the BV literature (B1). The same literature definition and generation process as in the first case would be used. The next step would be to identify the linked variables “c”, and then their associated literatures C (the collection of RD literatures) for the B1 literature. For this step, the proximity analysis method used in the previous cases would be employed for the B1 literature only. Continuing as before, this step uses the proximity analysis of DT to identify phrases in the BV literature physically close to the term BV, such as “c1”, “c2”, etc. The resulting C literatures are automatically linked to the A1 literature through the linking variable “b1”. The “c” variables which are disjoint to the A1 literature (i.e., the C sub-literature that does not contain the “a1” variable and the A1 literature that does not contain the “c” variables) must be identified, and become the candidates for discovery and innovation. The same stringent limits on variables and literatures used in the first case are applicable here.

*5.3.4. Requirements-mechanism driven*

The sixth case addresses the question, “What are the variables ‘a’ that could impact variable ‘c’ through mechanism ‘b’?” Applied to the above example for the case where “b” is blood viscosity only, and “a” is fish oil only, it could be phrased as, “What factors could impact Raynaud’s disease by impacting blood viscosity that would not be obvious from examining only the factors’ literature or the Raynaud’s disease literature?” The sixth approach is represented schematically as:

a⟵b⟵c

Here, ”b“ and ”c” are the independent variables, and “a” is the dependent variable. The operational sequence is to start with the variable “c”, and generate a literature C, and with variable “b”, and generate a literature B. Again, this means that the process would start by identifying and generating the RD literature (C1) and the BV literature (B1). The same literature definition and generation process as in the first case would be used. The next step would be to identify the linked row of variables “a” (a1, a2, etc.), and then their associated literatures A (the FO literatures) for the B1 literature. For this step, the proximity analysis method used in the previous cases would be employed, for the B1 literature only. Continuing as before, this step uses the proximity analysis of DT to identify phrases in the BV literature physically close to the term BV, such as “a1”, “a2”, etc. The resulting A literatures are automatically linked to the C1 literature through the linking variable “b1”. The “a” variables which are disjoint to the C1 literature (i.e., the A sub-literature does not contain the “c1” variable and the C1 literature does not contain the “a” variables) must be identified, and become the candidates for discovery and innovation. The same stringent limits on variables and literatures used in the first case are applicable here.

*5.3.5. Opportunity-mechanism-requirements validation*

The seventh case addresses the question, “Does the literature support the possibility that variable ‘a’ could impact variable ‘c’ through mechanism ‘b’?” Applied to the above example for the case where “a” is fish oil only, “b” is blood viscosity only, and “c” is Raynaud’s disease only, it could be phrased as, “Does the literature support the possibility that fish oil could impact Raynaud’s Disease by altering blood viscosity in a way that would not be obvious from examining only the fish oil literature or the Raynaud’s disease literature?” The seventh approach is represented schematically as:

a⟷b⟷c

Here, “a” and “b” and “c” are independent variables. The operational sequence could start with either “a” or “b” or “c.” For the present discussion, the operational sequence starts with the variable “b”, and generates literature B. Again following the above example, this

means that the process would start by identifying and generating the BV literature (B1). The same literature generation process as in the first approach would be used. The next step would be to extract the B1 sub-literatures which contain the variables “a1” (literature A1) and “c1” (literature C1).

The final step is to validate the existence of disjoint A1 and C1 sub-literatures (i.e., A1 sub-literature that does not contain the “c1” variable and a C1 literature that does not contain the “a1” variable). The “a1”-“b1”-“c1” sequence then becomes a candidate for discovery and innovation. The same stringent limits on variables and literatures used in the first approach are applicable here.

### 5.4. Conclusions

The advent of large databases, and the parallel advances in computer hardware and software, provide the opportunity to augment and amplify traditional approaches of human creativity in generating discovery and innovation. This document has shown that multi-discipline structured workshops can enhance the S and T innovation process, and has shown that multi-discipline literature-based analyses can enhance the S and T discovery process. The document has shown conceptually that the combination of computer-enhanced literature-based analyses and multi-discipline structured workshops has the synergistic potential to dramatically improve the discovery and innovation process relative to the already strong capabilities available from each process separately. This literature–workshop synergy represents a potential major breakthrough for systematically identifying the most promising disciplines to be used in the workshop and, from the strategic maps, systematically identifying specific experts from these different disciplines as well.

## Appendix A. Crossing the bridge: interdisciplinary workshops for innovation

### A.1. Abstract

This appendix describes the planning, conduct, and initial results of an interdisciplinary workshop on the theme “Autonomous Flying Systems”. The purpose of the workshop was to assemble representatives and achieve multi-discipline synergies and cross-discipline transfers to generate promising research directions for “Autonomous Flying Systems”.

The total workshop process consisted of three phases:

1. a two month pre-meeting e-mail phase in which each participant provided descriptions of advanced capabilities and promising research opportunities from his/her discipline to all other participants;
2. a two-day meeting at the Office of Naval Research (ONR) during which the promising opportunities identified beforehand were discussed, crystallized, and enhanced; and
3. a post meeting e-mail phase in which each participant provided additional or embellished opportunities.

These three phases are subsequently labeled “buildup”, “meeting”, and “cleanup”. This appendix summarizes some lessons learned from the workshop experience.

### A.2. Background

On 28 April 1997, at the 134th Annual Meeting of the National Academy of Sciences (NAS), Dr Bruce Alberts (NAS President) made a presentation entitled “Measuring what counts in science”. One point in his speech focused on a quote from Jared Diamond, a respected biologist: “Great scientific advances come especially from applying discoveries in one field to another entirely”. This interdisciplinary theme pervaded Dr Alberts’ presentation, and stimulated the establishment of workshops at ONR with the goal of enhancing interdisciplinary communications. Specifically, the aim of these workshops was to add discipline, category, and organizational diversity to ONR planning and other related processes.

The present appendix describes a novel workshop process developed under the above guidelines targeted at promoting innovation while also incorporating the organization, category, and discipline diversity components. This novel process was designed starting with a clean slate and was intended for application to very significant technical challenges.

The focus of the first novel workshop founded on this plan was “Autonomous Flying Systems”, an area of perceived long-term interest to not only the Navy and DOD, but also to NASA and other governmental and industrial organizations. The present appendix further describes the process that was used to identify the technical theme of the workshop, select the participants, and conduct all three phases of the total workshop.

### A.3. Workshop theme identification

It was decided that the initial workshop theme should focus on problems related to the main science and technology emphasis area of the author’s home organization, which was Strike Technology. The types of problems that could have been addressed ranged from fundamental science problems to advanced development and engineering problems. However, it was felt that the first workshop should be operated under conditions that would establish the most supportive environment for inno-

vation. There, the problem selected should be focused and understandable, and it should have a generic technical base amenable to soliciting people from many different disciplines. The less classified the problem is, of course, the more potential exists to draw individuals from a variety of disciplines. A number of technical problems were considered as candidates for the workshop theme, and the topic finally selected was autonomous control of unmanned air vehicles, including takeoff and landing from limited areas on smaller Navy ships. It was apparent that the underlying science and technology permeated many different disciplines, including aerodynamics, controls, structures, communications, guidance, navigation, propulsion, sensing, and systems integration. Also, the naval applications for some aspects of this problem were sufficiently unique that probably not a great deal of work had been done in this area. Subsequent literature analyses validated this assumption.

Present naval air systems are either manned (most aircraft) or teleoperated, semi-autonomous (weapons and some aircraft). The weapons are a mix ranging from "dumb" bombs and shells to "smart" missiles. The future trend is toward "smart" autonomous or semiautonomous aircraft and weapons. Since a major role of ONR is to proactively address the technology that will influence future naval forces, it seemed natural to examine science and technology roadblocks on the path to unmanned autonomous "smart" flight systems. Consequently, the focus of the initial workshop was defined as identification of the fundamental operational principles of autonomous flying systems over a fairly wide range of flight environments.

In particular, the workshop was aimed at examining what has been learned about autonomous or semiautonomous operation from the animal (mainly flying) kingdom and from other unmanned autonomous/semiautonomous teleoperated systems such as autonomous underwater vehicles and locomoted robots. Animals are now being studied as integrated systems by scientists on the forefront of biological research. The issues of aerodynamics, flight mechanics, dynamic reconfiguration, materials, control, neurosciences, and locomotion are not being studied as separate disciplines by these scientists, but rather are being studied in parallel in the same animal system and in their relation to the function and mission of the animal system. While this integrative biological research is in its infancy, and results are only starting to emerge, the time seemed appropriate for assembling these diverse groups and exploiting their synergy. Not only could there be benefit to the Navy from such cross-discipline interaction, but benefit could be possible for each of the contributing disciplines as well.

A major thrust of the workshop was projected to be identification of the autonomous operational principles for each unique system and the relation of these principles to mission and function, then extraction of the generic operational principles that underlay all the systems, both biological and man-made. It was hoped that the cross fertilization of disciplines would be able to further elucidate and clarify the more important generic concepts, and then provide insight that could be utilized to enhance the autonomous operation of naval flying systems.

### *A.4. Participant selection*

Once the theme of the workshop was established, a sub-theme taxonomy was developed to focus the agenda and to identify workshop participants. A dual approach was followed to generate the taxonomy.

Discussions were held with agency experts on the generic theme concerning the taxonomy structure. In parallel, the Science Citation Index (SCI) was queried for papers related to the generic theme. Both bibliometric and computational linguistics analyses of these papers were performed to provide strategic maps of the topical area, identifying key performers, journals, institutions, and their relations to the technical themes and subthemes of the workshop. A taxonomy was constructed based on these strategic maps. (For a description of how the bibliometric and computational analyses are combined to generate strategic maps, see Kostoff, 1997, 1998).

NOTE: If a combined literature-based and workshop-based approach is taken for stimulating innovation, as recommended in this paper, the literature-based analysis would be done at this stage.

Both of these taxonomy sources, in-house experts and the SCI, then provided initial candidates for participation in the workshop. These candidates were contacted, and asked to suggest additional candidates. This procedure continued until a large pool of potential candidates was established. Three main selection criteria for workshop participants were established:

1. multiple recommendations,
2. significant publications in the field, and
3. literature citations.

These three criteria were tempered with judgement to insure that bright young individuals, who had not yet established a track record, were not excluded from the pool, and that the panel as a whole had the correct level of discipline, category, and organization balance. In addition, a guideline was established that all workshop attendees would be active participants, so the number of attendees was limited to facilitate discussion and interactions.

All these constraints, guidelines, and selection criteria were used to arrive at the final panel size and structure. The result was a panel of slightly more than twenty

people representing a mix of disciplines that included biologists (experts in bird, bat, frog, fish, or insect studies), robotics, artificial intelligence, controls, autonomous aircraft, fluid dynamics, sensors, neuroscience, cognitive science, autonomous underwater vehicles, aerodynamics, propulsion, and avionics.

### *A.5. Overview of workshop process steps*

#### *A.5.1. Workshop buildup*

The buildup period for the Workshop in question started about two months before the meeting. Specific guidance for the conduct of the workshop was sent to the participants by e-mail, including a statement of the naval technical problems to be addressed. The technical component of the buildup phase was then conducted by e-mail.

The main purpose of this buildup phase technical component was to have each participant generate new ideas from his/her discipline for all other participants to consider. The other participants could then dialogue by e-mail to clarify/modify/embellish these ideas. At a minimum, even if no dialogue resulted, there would be a gestation period of about two months for each participant to absorb these concepts from other disciplines. Specifically, each participant was requested to:

- submit a half dozen leading edge capabilities or accomplishments in his/her discipline(s) that could potentially impact the naval technical problems; and
- identify several leading edge capabilities or accomplishments projected in his/her discipline(s) over the next decade that could potentially influence the naval technical problems; and
- submit a few leading edge capabilities or accomplishments in his/her discipline(s) whose impact on the naval technical problems was not obvious to him/her, but might be obvious to someone else.

Anyone of the participants was free to comment on potential relations among any of the capabilities, accomplishments, or combinations of capabilities and accomplishments, and any of the naval technical problems, or combinations of problems. All of the comments received were then sent to all the participants. This exercise helped stimulate the thinking of the participants, and provided a documented record of the process. One of the functions of the participants from the author's organization was to facilitate and stimulate dialog by raising questions and issues on the submitted information.

If any of the participants saw a capability or accomplishment from another participant which could impact a problem in his/her discipline, but not impact a naval technical problem, then the two participants were free to dialog together without informing all the participants. However, these two participants engaged in independent dialog were requested to keep a record of their exchange that might be included with the final workshop report as a potential innovation. This would cover the real possibility of innovation occurring in topics other than the one targeted.

#### *A.5.2. Workshop meeting*

As a result of the ideas presented during the buildup phase, it appeared that the seeds existed for a new science and technology (S and T) program on Autonomous Flying Systems. Therefore, an agenda was sent to the participants with further guidance to address promising S and T opportunities at the workshop, that would serve as the foundation of such a program. Specifically, the participants were asked to address the following issues at the workshop:

- What are the present leading-edge capabilities in your discipline?
- What are the desired future capabilities in your discipline?
- What are the leading research opportunities in your discipline and what additional capabilities could they provide if successful?
- What is the level of risk of these opportunities successfully achieving their targets?
- How would these potentially enhanced capabilities contribute to, or translate into, improved understanding and/or operation of autonomous flying systems?

The meeting occurred on 10-11 December 1997 at ONR. Since some of the leading edge capabilities and potential accomplishments appeared to have applicability to naval technical problems (identified during the e-mail buildup period), the proponent for the capability or accomplishment item took the lead in fleshing out his/her ideas and leading the discussion at the meeting. As a result, the workshop meeting tended to evolve into full panel discussions on each of these potential capabilities.

There were two rounds of discussion at the workshop. The first round consisted of presentations and discussions by each proponent. The second round of the workshop consisted of each participant identifying his/her leading promising research opportunities.

#### *A.5.3. Workshop cleanup*

The participants were requested to provide any additional narrative information that added to or modified their ideas as a result of the workshop experience. The outcome of the workshop included both the tangible and intangible.

Three immediate tangible outcomes were projected:

1. A concept proposal for an S and T program focused on Autonomous Flying Systems would be generated;

2. Technical papers may be submitted to leading science journals based on innovations identified; and
3. One or more papers on the complete workshop experience might be submitted to leading science journals.

In addition to developing specific topics, it was anticipated that new, un-exploited ideas in interdisciplinary research and development might surface during contact between panelists. These novel subjects might form the basis of additional workshops. In addition, extensive lessons were learned as a result of the workshop process. These lessons were summarized in Section 2.

## References

Finn, R., 1998. Program Uncovers Hidden Connections in the Literature. The Scientist May 11, 1998.

Gordon, M.D., Lindsay, R.K., 1996. Toward discovery support systems: a replication, re-examination, and extension of Swanson's work on literature-based discovery of a connection between Raynaud's disease and fish oil. JASIS 47, 2.

Kostoff, R.N., 1993. Database tomography for technical intelligence. Competitive Intelligence Review 4, 1.

Kostoff, R.N., 1994. Database tomography: origins and applications. Competitive Intelligence Review, Special Issue on Technology 5, 1.

Kostoff, R.N., 1997. Database tomography for information retrieval. Journal of Information Science 23, 4.

Kostoff, R.N., 1998. Database tomography for technical intelligence: a roadmap of the near-earth space science and technology literature. Information Processing and Management 34, 1.

Kostoff, R.N., 1999. Hypersonic and supersonic flow roadmaps using bibliometrics and database tomography. Journal of the American Society for Information Science, 15 April.

Swanson, D.R., 1986. Fish oil, Raynaud's syndrome, and undiscovered public knowledge. Perspectives in Biology and Medicine 30, 1.

Swanson, D.R., Smalheiser, N.R., 1997. An interactive system for finding complementary literatures: a stimulus to scientific discovery. Artificial Intelligence 91, 183.

**Ronald N. Kostoff** received a Ph.D. in Aerospace and Mechanical Sciences from Princeton University in 1967. At Bell Labs, he performed technical studies in support of the NASA Office of Manned Space Flight, and economic and financial studies in support of AT and T Headquarters. He invented many concepts, including the Orbiting Molecular Shield (aka Wake Shield). This concept pioneered the capability of high vacuum in low orbit, presently exploited by all manned space vehicles. His initial aerobraking research reported in 1970-71 pioneered the Aeroassisted Orbit Transfer subfield of Orbital Transfer Vehicles. His economic and financial studies, which supported AT and T's operations, resulted in potential savings to the Bell System of over one billion dollars. At the US Department of Energy (DOE), he managed the Nuclear Applied Technology Development Division, the Fusion Systems Studies Program, and the Advanced Technology Program. He published numerous technical papers in the fields of pulsed fusion operation, impact fusion options, and fissile fuel production using advanced breeders. At the Office of Naval Research, he was Director of Technical Assessment for many years. He invented and patented the Database Tomography process, a textual data mining approach that extracts relational information from large databases. His interests continue to revolve around improved methods to assess the impact of science and technology, incorporating maximal use of the massive amounts of data available. He has published many papers on technical, evaluation, and data mining topics, and has edited three journal special issues since 1994 (Evaluation Review [Feb. 94], Scientometrics [July 96], Journal of Technology Transfer [Fall 97]).