Available online at www.sciencedirect.com

SCIENCE DIRECT®

Technological Forecasting & Social Change 72 (2005) 1122–1136

Technological Forecasting and Social Change

Evaluation of Laboratory Directed Research and Development investment areas at Sandia

Kevin W. Boyack*, Nabeel Rahal

Sandia National Laboratories,[1] *P.O. Box 5800, Albuquerque, NM 87185, United States*

Received 13 May 2004; received in revised form 15 September 2004; accepted 16 September 2004

Abstract

Sandia National Laboratories conducts a variety of research projects each year under its Laboratory-Directed Research and Development (LDRD) program. Recently, information visualization techniques have been used with corporate data to map several LDRD investment areas for the purpose of understanding strategic overlaps and identifying potential opportunities for future development outside of our current technologies. Tools, techniques, and specific analyses are presented here. We find that these tools and techniques hold great promise for aiding the future direction of the science and technology enterprise.

Keywords: Laboratory-Directed Research and Development; Investment; Sandia

1. Introduction

The Laboratory-Directed Research and Development (LDRD) program at Sandia National Laboratories conducts world-class research on a variety of subjects that are relevant to Sandia's missions and potentially useful to other national needs. Much of the technology that has been developed

* Corresponding author. Tel.: +1 505 844 7556; fax: +1 505 844 5670.
E-mail addresses: kboyack@sandia.gov (K.W. Boyack), nrahal@sandia.gov (N. Rahal).

[1] Sandia is a multiprogram laboratory operated by Sandia, a Lockheed Martin Company, for the United States Department of Energy under contract DE-AC04-94AL85000.

doi:10.1016/j.techfore.2004.09.007

at Sandia has its roots in the LDRD program. Research investment decisions made 10 and 15 years ago are having a direct impact on national security programs today.

Sandia's LDRD program is divided into roughly a dozen different investment areas (IAs) including five that we focus on in this paper: *Computational and Information Sciences (CIS)*, *Engineering Sciences (ES)*, *Electronics and Photonics (EP)*, *Materials Science and Technology (MST)*, and *Pulsed Power Sciences (PP)*. The LDRD process occurs annually at Sandia, starting early each spring. First, staff members submit short ideas answering written calls (i.e., requests for proposals). Then, internal teams of experts in the technologies comprising each IA review the short ideas and select some fraction of them for full proposals. The expert teams then review the proposals and select those to receive funding. LDRD projects have a maximum duration of 3 years. Continuation proposals and an annual review are required for each existing project that has not completed its term. In an effort to improve the inputs to the request for proposal and decision-making process, and given the availability of relevant data, we have embarked on a program to map our LDRD IAs. We have applied advanced information visualization tools to understand historical development, validate strategic and tactical directions, and identify opportunities for future development for each of the five IAs mentioned above.

This paper describes the project plan, detailed processes, data sources, tool sets, and sample analyses and validation activities associated with the mapping of Sandia's LDRD IAs.

2. Project plan

The original plan associated with this assessment activity consisted of several steps, which included ways both to benchmark our methods and to deliver practical results.

The first step was to create Sandia-specific visualizations of the IAs. The purpose of these visualizations was to identify past and present technological competencies, and overlaps of competencies, within the IAs. Although IA experts are expected to be very versed in the competencies of their own IAs, they may not have such detailed knowledge about the other IAs. Mapping enables experts to extend their expertise outside of their own IAs, thus enabling them to better leverage investments in other areas. Benchmarking was accomplished by comparing the visualizations with the mental models of IA leads—experts who have, in the past, used traditional processes to understand their areas and make funding decisions. Time was built into the plan to iterate the visualizations if large differences were found between them and the leaders' mental models of their areas. Meetings with the

Table 1
Data used to produce Sandia-specific and DOE LDRD maps and mapping between fields from different data sources

Calls (RFP)	New proposals	Continuation proposals	Project reports	Publications	U.S. DOE LDRD data
ID number	ID number	ID number	ID number	ID number	ID number
IA	IA	IA	IA	IA	
					Laboratory
	Title	Title	Title	Title	Title
Year	Year	Year	Year	Year	Year
	PI[a]	PI	PI	Author	
				Source	
Text	Text	Text	Text	Abstract	Text

[a] PI=principal investigator.

IA leads were designed to not only benchmark the visualizations, but to educate the leaders and add detail to their mental models.

After completion of the benchmarking activity with Sandia-specific visualizations, a second set of visualizations was created to include data on all U.S. Department of Energy (DOE)-funded R&D activities related to the IAs. The purpose of this set of visualizations, hereafter referred to as "DOE LDRD," was to place Sandia's IA activities within a broader context, thus allowing for the identification of new opportunities by semantic association with activities outside Sandia. These visualizations were also presented to IA leaders. Copies of the data, visuals, and navigation tools were also provided to IA leaders to allow them to explore the data independently.

3. Process, data, and tools

Two different types of visualizations, each designed to provide different types of information, were created for this activity. The first can be described as a landscape map, which is particularly suited to looking for patterns and trends in large data sets. The second type is a link analysis map, which is valuable for identifying specific topic-based relationships within large data sets.

The landscape maps were created using a process consistent with commonly accepted methods of mapping knowledge domains [1] (see Fig. 1):

- Appropriate textual records were identified and combined in a database.
- Latent semantic analysis (LSA) [2,3] was used on the titles and descriptive text for each record to generate a document–document similarity matrix.
- A graph layout program, VxOrd [4], was used to calculate the document graph.
- The resulting graph or map was explored using VxInsight [5], a visualization tool that enables interactive navigation and query of an abstract information space.

Link analysis maps were generated using ClearResearch, a product developed by ClearForest[2] that extracts entities (e.g., person, company, technology, product, university, etc.) and relationships from unstructured textual sources. Using rules to define categories, ClearResearch produces link analyses at multiple levels of detail. The steps involved in producing these maps are as follows:

- The same textual records and database described for the landscape maps above were used here.
- A rule-based unstructured text tagging module was used on titles and abstracts to extract and categorize technology terms and organization terms (e.g., CIS, MST, PP, ES, and EP).
- Technology and organization terms were linked together on a document basis and visualized in a network or link analysis map.

Both types of visualizations, the landscapes and the link analyses, were used for both the Sandia-specific and DOE LDRD analyses, as detailed below.

[2] ClearForest (see http://www.clearforest.com/).

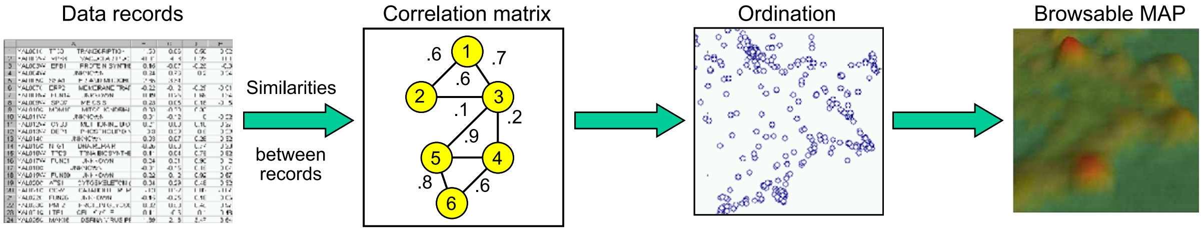

Fig. 1. Process of putting data into a VxInsight map.

3.1. Data collection

Two different sets of data were compiled from multiple sources for our analyses—one for the Sandia-specific visualizations and one for the DOE LDRD visualizations.

The data for the Sandia-specific visualizations consisted of 1209 records from the five IAs and included the following: LDRD call text (i.e., request for proposals), descriptive text for both new proposals and continuation proposals, report abstracts for funded projects, and abstracts from peer-reviewed publications resulting from funded projects from FY2001[3] through FY2004 (see Table 1). Given that we are in the midst of FY2004, no project reports are available for the current year. These data are proprietary to Sandia, and are not generally available externally.

To create the DOE LDRD visualizations, an additional ~4300 LDRD records from the entire U.S. DOE complex (e.g., including Los Alamos National Laboratories and Lawrence Livermore National Laboratories) for FY2001 and FY2002 were added. FY2003 data were not yet available. Of these, 180 duplicated existing Sandia-specific records and another 200 had no titles or descriptive text, and were thus removed from the data set. A total of 990 of the new records had both titles and descriptive text, while the balance only had titles. With the Sandia-specific and additional DOE data, this set consisted of 5112 records.

3.2. Similarity calculation

LSA is a technique based on the vector space model that has found recent application in information retrieval. Its relative strengths are that it can represent aspects of the meanings of words, and effectively deals with synonymy and polysemy. Traditional LSA uses the singular value decomposition (SVD) technique to deconstruct a term–document matrix $\{\mathbf{X}\}$ into the product of three other matrices, with $\{\mathbf{X}\}=\{\mathbf{W}\}\{\mathbf{S}\}\{\mathbf{P}\}'$, where $\{\mathbf{S}\}$ is the matrix containing the singular values. This matrix is then truncated to the highest ~300 singular values. To calculate the document–document similarity matrix, matrices $\{\mathbf{W}\}$ and $\{\mathbf{S}\}$ are multiplied. The resulting vectors are then normalized to unit length, and the inner products are calculated. These inner products are the document–document similarity values.

Our LSA methodology differs slightly from that above in that we use semidiscrete decomposition [6,7] rather than SVD to do the term–matrix deconstruction. Although this typically reduces the precision by a small amount (~2%), it is much less memory-intensive due to its discrete nature, and runs easily on a PC. We also used an optimized stopword list prior to construction of the initial term–

[3] FY=fiscal year, which goes from October 1 to September 30.

document matrix. This stopword list was designed to allow the LSA to focus on technical content, and thus removed common words, many verbs, adjectives, adverbs, and words that were tied closely to only one type of document (e.g., project, proposal, report, etc.).

LSA generates a full $n \times n$ similarity matrix. Our experience with many data types and sets indicates that the use of the full similarity matrix is not necessary. Rather, use of the top few similarities per record is sufficient to generate a meaningful map. Thus, we used only the top 15 similarities per record to generate the landscape maps.

3.3. Ordination

Ordination using the similarity files generated from LSA was done using VxOrd, a force-directed graph layout algorithm that preserves both global and local structures for a range of graph sizes ($1k-1M$ nodes). VxOrd has been used for many different types of maps with good success [5,8,9,10]. This step is referred to as ordination rather than clustering because VxOrd generates x,y coordinates for each record (calls, proposals, reports, etc.) but does not assign cluster numbers. The ordination places similar documents close to each other on the graph.

3.4. Visualization using VxInsight

After calculating coordinates, the data set is loaded into VxInsight for exploration and analysis. VxInsight is a tool that allows visualization and navigation of an abstract information space, such as a large document set. It uses a landscape metaphor and portrays the structure of the space as peaks and ridges of documents. The size of a peak and its relative position in the layout provide valuable clues to the role of that group of documents in the overall structure. Labels on dominant peaks are based on the two most common words in the titles (or other fields) that comprise that peak, thus revealing the content of the various peaks. Users can navigate the map terrain by zooming in and out, querying metadata fields (e.g., titles, abstracts, etc.), or restricting the data displayed to a certain time span and sliding through sequences of years with a slider. Relationships among the individual data records may be displayed as arrows between documents and understood at many levels of detail. Details about any data record are also available upon demand.

Effective use of the labels, zooming, and query, and details on demand capabilities within VxInsight allow the analyst to both pose and find answers to questions of a strategic nature. However, the tool is tuned to interactive exploration rather than static presentation. Thus, in most cases, the analyst will make screen captures and note findings to be presented to others using more traditional static forms. Software tools that are very good at both exploration and presentation have yet to be developed.

4. Analysis and discussion

4.1. Landscape mapping of IAs

As mentioned above, one of the purposes of generating a Sandia-specific map of IAs was to benchmark the map against the mental models of IA leads. One such exercise is described here for the *Computational and Information Sciences (CIS)* IA. During our meeting with the CIS area leader, we first gathered information about his mental model of: (1) the CIS area, and (2) perceived overlaps between CIS and the other four areas, and then we presented our maps to the individual.

A graphic describing the CIS IA lead's mental model of the overlap between CIS and the other four areas is shown in Fig. 2. First, the lead perceived that there was a significant fraction of the CIS space that was unique to the IA, and thus had no overlaps. The two largest perceived overlaps were between CIS and the *Engineering Sciences (ES)* and *Materials Science and Technology (MST)* areas, each accounting for a significant fraction of the space. The perceived overlaps with the *Electronics and Photonics (EP)* and *Pulsed Power* (PP) areas were much smaller, but were thought to be increasing with time. Potential three-way overlaps were not considered.

The Sandia-specific IA map, generated using the process described above, is shown in Figs. 3 and 4. Fig. 3 shows the VxInsight view, which does not translate well to paper or grayscale, but which is very useful for interactive exploration. Fig. 4 shows the same data in a scatterplot view, where different symbols are used for the different IAs. The calculated positions of the documents have been dithered slightly for this view to allow more data points to be viewed. The CIS-related documents are shown by two different symbols: those directly funded by CIS (filled squares), and those funded by other IAs but using the term "computation" (filled circles).

Fig. 4 shows that the main area for CIS is the large cluster of filled squares at the middle left. This comprises the portion of the CIS area that is unique and not overlapping with other IAs. ES (open squares) tends to form a bridge between the main CIS cluster and the large MST-based cluster at the lower middle. ES also dominates the smaller clusters at the far and upper left of the graph. MST (open diamonds) is divided into three main components: one at the lower middle, one just to the right of the middle, and one near the lower right of the graph. EP (open triangles) likewise has two main components—one near the upper right, and one at the middle right—while PP (open circles) is focused at the far middle right. MST seems to have the most central position of the five IAs.

Specific overlaps between CIS and the other IAs are labeled on the figure for convenience. Areas occupied solely by CIS are a significant fraction of the CIS total and correspond well to the mental model's view of unique space. Significant overlaps between CIS and ES at the bottom of the main CIS cluster, and in the smaller clusters to the far and upper left mainly deal with algorithms and transport phenomena. Several smaller regions of overlap between CIS and MST, mostly in the lower part of the graph, are all related to microsystems and related technologies. These two areas, ES and MST, show the

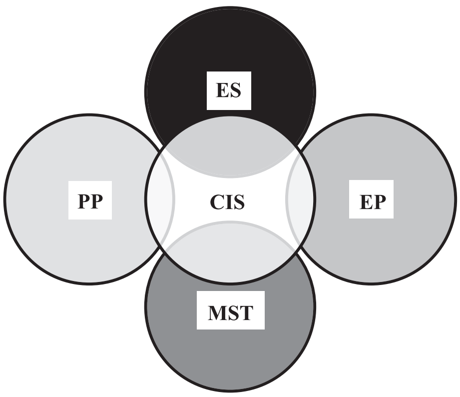

Fig. 2. CIS IA leader's mental model of CIS overlaps.

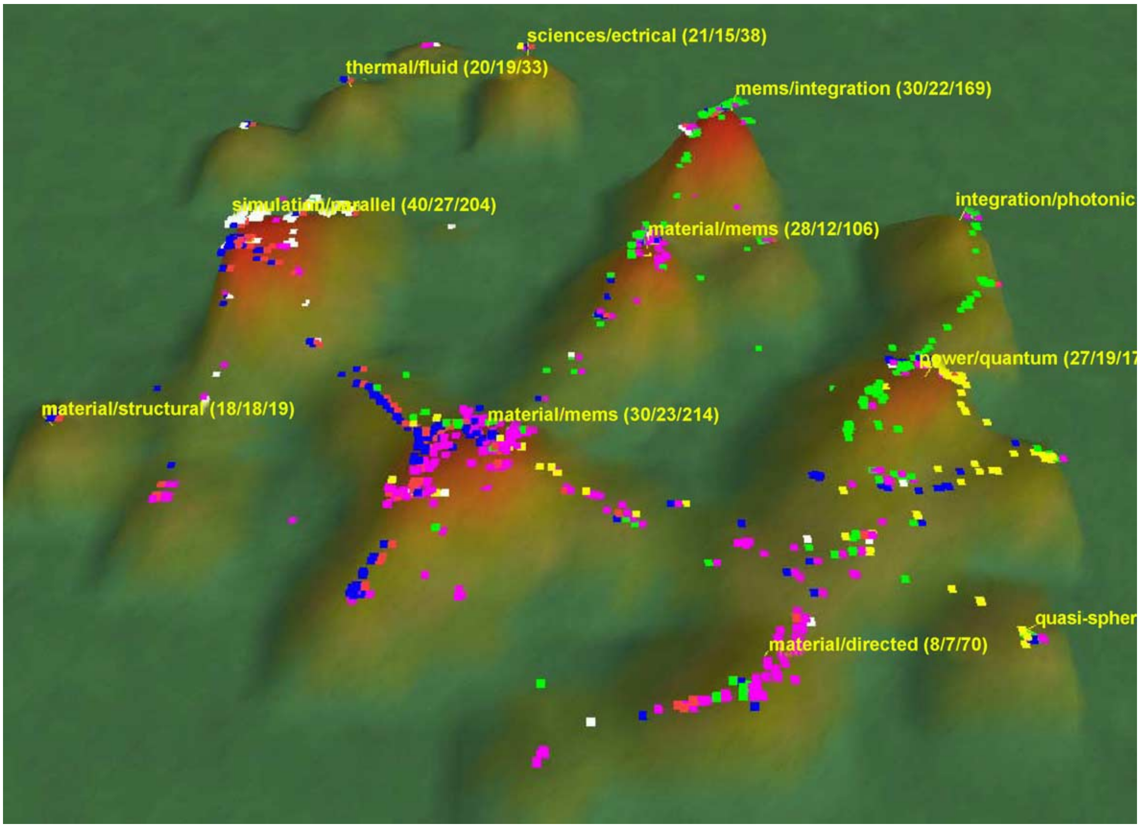

Fig. 3. VxInsight map of five Sandia LDRD IAs. Each area (CIS, ES, EP, MST, and PP) is indicated by a different colored dot on the landscape. A sixth dot color represents CIS indirect investments. Overlaps between areas can be seen where dots of different colors are shown together. Labels indicate the most dominant "title" words and their frequencies for each peak on the landscape. The VxInsight views are meant more for active navigation of data than for presentation of results.

greatest overlaps with CIS, which correlates well with the mental model of Fig. 2. EP shows two small areas of overlap with CIS, while PP shows only one overlap at the far right. This also matches the mental model quite well.

One area of particular interest on the map is that found in cluster at the lower middle of the graph. Although this region is dominated by MST and CIS, careful inspection shows that ES and EP also have a presence here. Thus, four of the IAs overlap, suggesting that all four IAs could share and benefit from joint calls and proposal review in the central subject of this region of the map—that of microsystems and related materials.

Using the VxInsight time-sliding capability, we investigated trends in the IA overlaps, some examples of which are mentioned here. The extent of overlap between CIS and ES has remained roughly constant over the period from 2001 to 2004, with the areas of focus shifting towards optimization of algorithms. The conceptual overlap between EP and MST has increased significantly in the past 2 years, especially in the area of integration for product application. Shifts in focus in the individual IAs can be seen as well. For instance, a portion of the EP portfolio dealing with microelectromechanical systems (MEMS) technology has shifted from component integration to applications.

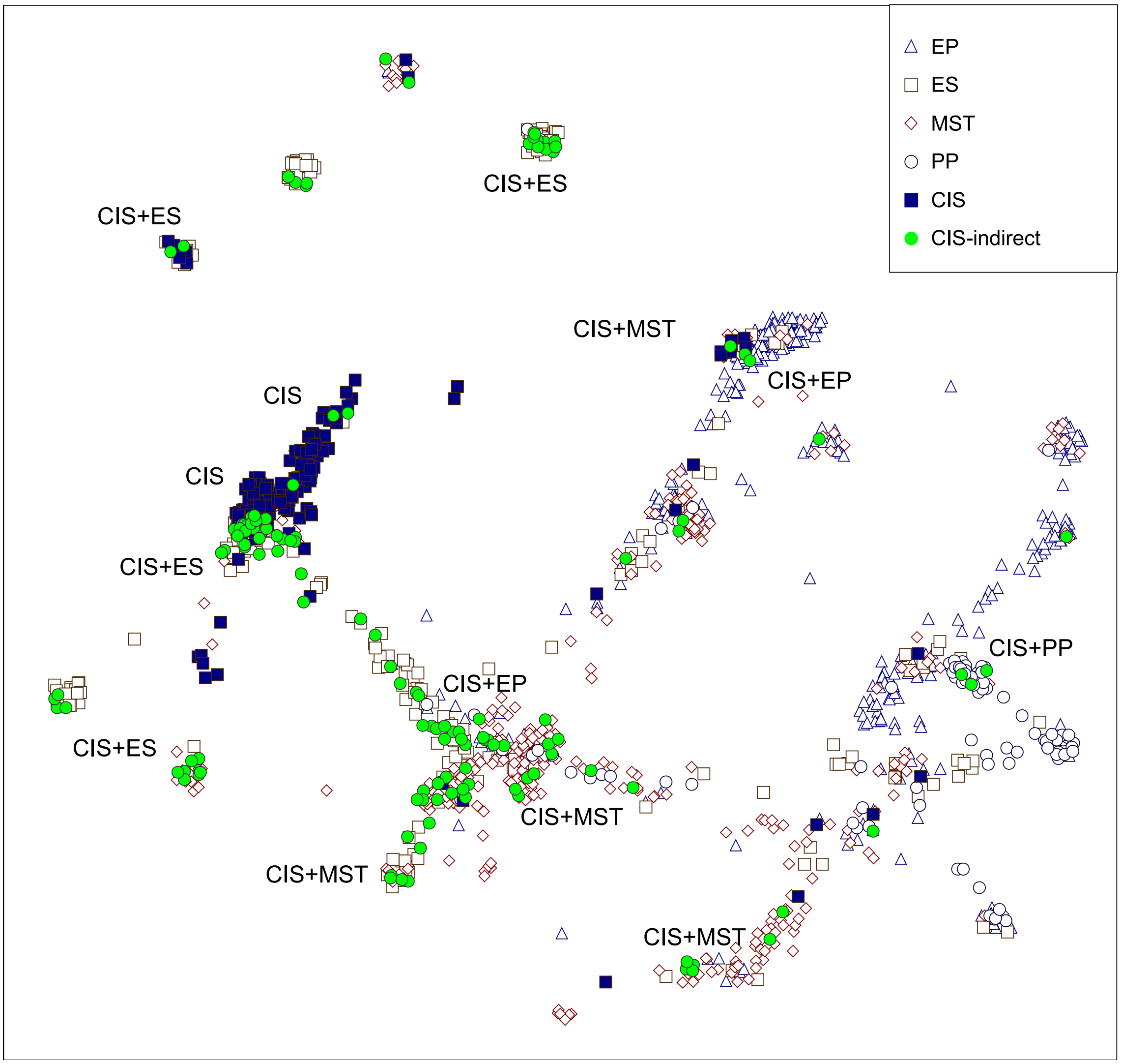

Fig. 4. Scatterplot of the five Sandia LDRD IAs using the same map coordinates as shown in the VxInsight map of Fig. 3. Overlaps between the CIS IA and other IAs are specifically labeled.

Another significant outcome of the meeting with the IA leader was his desire to put the VxInsight tool and data sets on his computer so that he could explore the data independently and draw his own conclusions related to both assessment and potential future directions.

4.2. Link analysis of IAs

The analyses of the visualizations in Section 4.1 tend to strongly convey the patterns and trends occurring within the IAs. However, specific information indicating the relationships between technology and IA and the explicit nature of the relationships between the technologies is still hidden. In order to extract the hidden relationships within the landscape visualization, many hours of exploring, including reading abstracts, would be required. An alternative approach to tedious review is the development of a link analysis map coupled with an unstructured text-tagging rulebook.

The link analysis map was crucial in portraying to the IA leads the direct and indirect relationships that occurred between technologies within their IAs, as well as relationships that occurred between all five IAs. This analysis added value in that the IA leads obtained information that assisted them in the evaluation and redirection of their R&D activities.

The first level of analysis consisted of identifying relationships between technologies and multiple IAs. The relationships exposed by this analysis were intended to reveal potential overlapping or complementary technology spaces that can be jointly leveraged in future LDRD calls. Fig. 5 is an example of the link analysis visualizations that were created and shared with the IA leads. Common technologies that indirectly link two (or more) IAs appear between the IAs, showing direct links between a technology and the associated IAs. Thicker lines indicate stronger relationships. Technologies that are unique to an IA are depicted by the collection of links that extend out from each IA label. These are not shown in the figure to focus attention on the overlaps. The actual visualizations reviewed by the leads were often more detailed, using lower-linking thresholds.

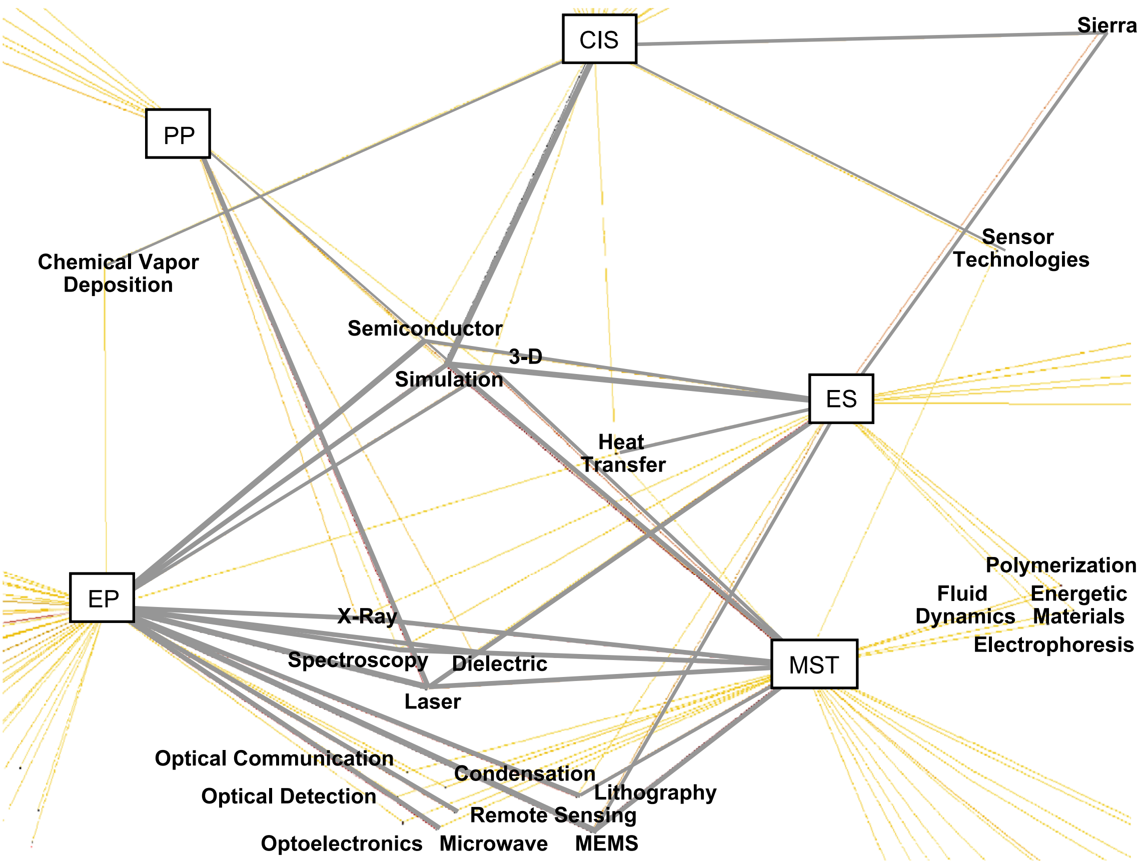

Fig. 5. ClearForest link analysis map of specific technology linkages between the five IAs. Thicker lines indicate stronger relationships.

The first level of analysis identified a macroscale understanding of the overlaps as well as the unique competencies and capabilities that each IA possessed. This understanding was then used as a validation model for the IA leads. Fig. 5 indicates that each IA has a robust set of unique technologies indicated by the unlabeled lines extending out from the IA markers. This unique set of technologies represents the development of a strong and innovative R&D portfolio. The figure also validates the proper roles assigned to each IA. For example, MST idealistically should support EP, PP, and ES, with very little support to CIS. The rationale behind this is that MST provides the expertise in materials for the development of devices in EP and PP; however, MST needs the simulation expertise that resides in ES to develop materials, and ES needs the hardware and software expertise in CIS to develop and apply simulations. Fig. 5 is a visualization of the current relationships, which seems to be consistent with the ideal state mentioned above.

The second level of analysis consisted of the identification of specific relationships between IAs. Fig. 5 depicts a very strong relationship between EP and MST. The thickness of the links between EP and MST indicates a strong potential collaboration based upon MEMS and lithography. In addition, optical detection, communication, optoelectronics, and remote sensing should also be taken into consideration as potential areas of collaboration. As a result of the findings above, it was advised that EP and MST work together to identify a collaborative approach for a portion of their future LDRD calls, and to establish a funding pool for joint EP/MST proposals.

The third level of analysis consisted of a technology-to-technology relationship assessment within a single IA. The assessment was used to assist the IA leads in portfolio management activities. The visualization contained very detailed (and thus proprietary) information, and is not shown here. The result of the visualization pointed to specific technological efforts within an IA that could be combined to create a larger effort that could in turn attract future funding outside of the LDRD program. In addition, the IA lead was able to identify, compare, and leverage objective technological strengths to attract new external customers.

4.3. Landscape mapping of DOE LDRD

A map of the DOE LDRD data set was created using the same technique described previously and is shown in Fig. 6. The purpose of this map was primarily to identify additional opportunities by comparison of Sandia IA data with work of national interest that is being funded at other DOE laboratories. The roughly 3800 records added to the Sandia IA data add significant context and content that provide fodder for new ideas. It is worth noting that the visualizations themselves do not generate new ideas. Rather, it is the analyst or IA lead interacting with the visualizations that formulates questions and new ideas based on the information and patterns seen there.

Fig. 6 shows the overall scatterplot comprising investment in LDRD by all of the U.S. DOE's laboratories. In the context of this type of map, we define an opportunity as a space where other laboratories are performing work, but Sandia is not. In reality, the opportunity space is somewhat broader than this definition given that the map can cause the analyst to form questions or ideas outside the technology clusters within the map. We note that a more global map, one based on current global literature rather than just the work of one U.S. government agency, would show a much broader opportunity space. However, such a map would take much more data and time to construct.

Fig. 6 shows that significant areas of the graph, especially at the top and right, are not covered at all by any of the Sandia IAs. Although we could consider these clusters to be opportunities, they are not the

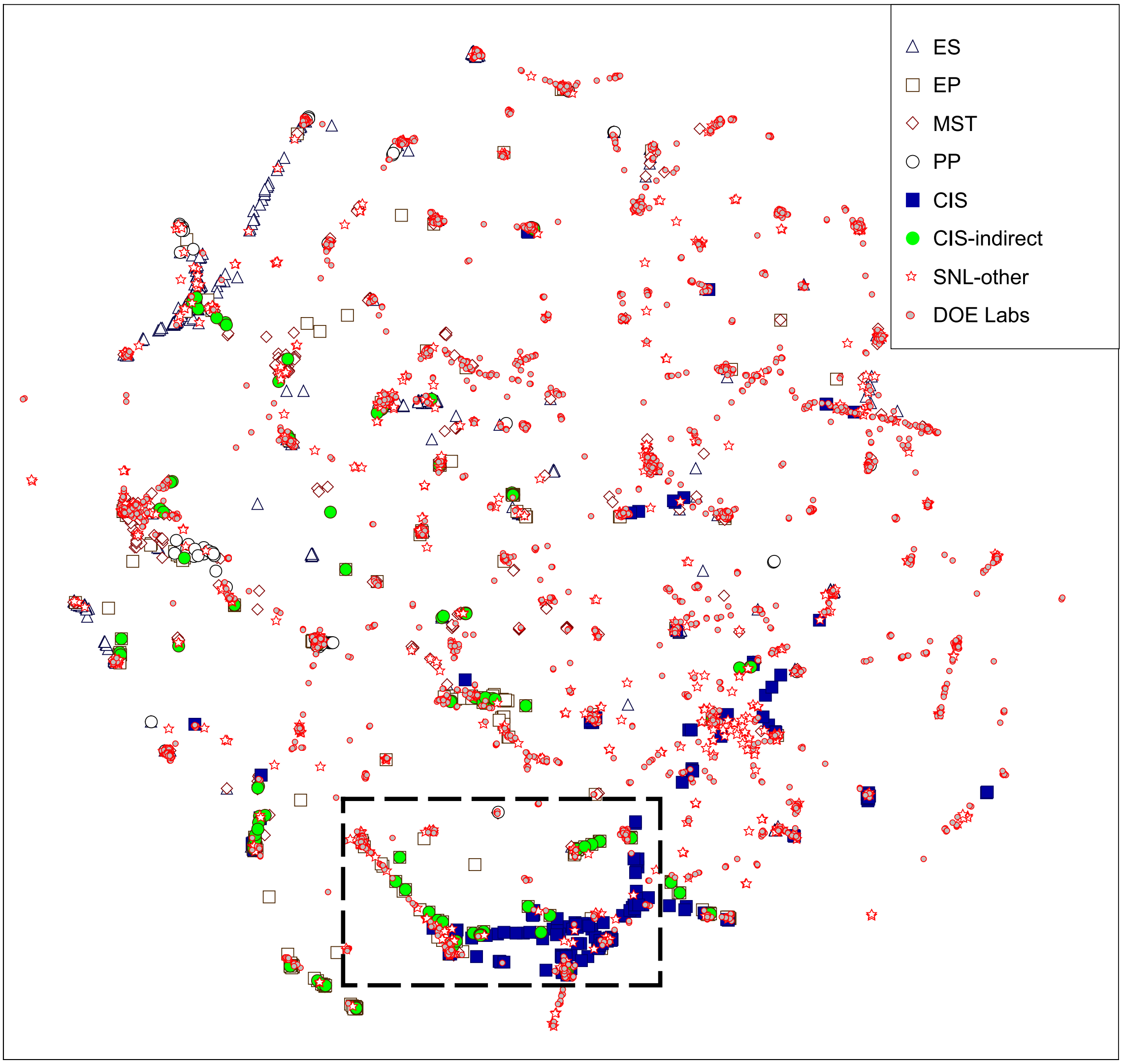

Fig. 6. Scatterplot of the entire DOE LDRD space. The five Sandia IAs are shown using the same legend shapes as used in Fig. 4. Stars indicate Sandia LDRD projects in IAs other than the five specifically called out here. Small filled circles indicate LDRD investments made by all other DOE laboratories. The area inside the dashed box is explored further in Fig. 7.

areas of interest to Sandia since the map indicates that they are well outside our core competency areas. We are more interested in new opportunities in areas very related to our own competencies given that the barriers to entry would be much lower given our expertise. Thus, we have looked at the map to specifically identify clusters of non-Sandia work that are very close to our competencies. These are given first consideration as potential opportunities for expansion.

Fig. 7 shows a detailed VxInsight view of the lower middle region of the DOE LDRD map. This region is dominated by Sandia's CIS IA, as shown by the large filled squares and circles within the dashed region of Fig. 6. All of the non-Sandia records have been marked as black dots in Fig. 7. Examination shows several small clusters of data in areas that are very related to our computational

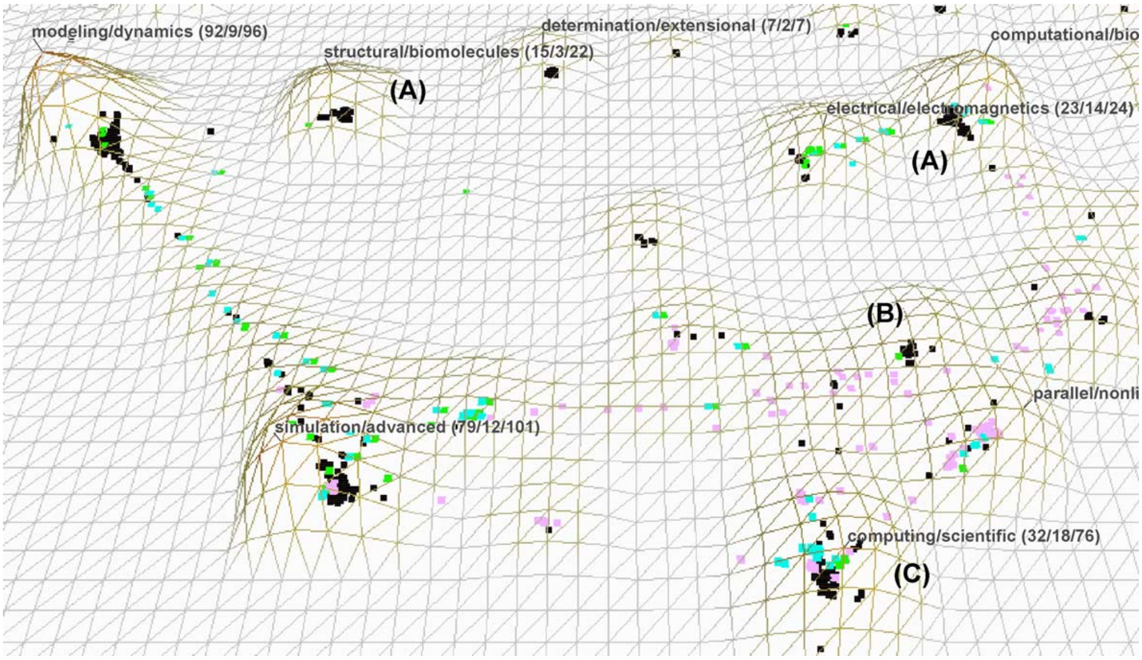

Fig. 7. VxInsight view of the dashed inset from Fig. 6 of the area comprising most of the CIS IA. Light-colored dots show the Sandia CIS projects, while black dots indicate projects from the other DOE laboratories. Dark dots surrounded by few light dots thus show areas semantically within the Sandia CIS space, but that currently receive little or no Sandia investment.

competencies, and that are potential areas of future opportunity for the CIS IA, given its current portfolio and competency base. These areas include computational biology (both structural and genomics), nonlinear algorithms, and climate modeling, labeled A, B, and C, respectively in the figure. Some of these were anticipated by the CIS investment team in that the FY2005 calls (issued in March 2004) reflected an increased interest in informatics, of which computational biology is one type.

One can carry the analysis even further by looking at the distribution of projects in the "potential" spaces by laboratory. If the "potential" space in a given cluster is dominated by a single laboratory with many projects, then the barrier to entry (in terms of future competition for funding) would be high given the unique expertise of that laboratory. Conversely, if the space is spread among many laboratories, each with just a few projects, the barrier to entry, and ability to compete in the near future, is low. Using this metric, we find that the barrier to entry is reasonably low for the computational biology and climate modeling areas, and somewhat higher for nonlinear algorithms. Of course, in a final analysis, barriers to entry would be weighted against specific competencies and the people with those competencies in making decisions about which future opportunities to fund.

4.4. Link analysis of DOE LDRD

The Sandia-specific link analysis assisted in the understanding of the technologies within, and the relationships among, the technologies from different IAs. The next step was to take the localized knowledge extracted from the IA analysis and compare the strengths and weakness with the rest of the DOE complex. The first analysis in this section consisted of only using LDRD projects, in addition to rolling up all of the IAs to an overall Sandia category. The second analysis consisted of analyzing each IA in the context of the entire DOE complex. The data used for this analysis consisted of LDRD calls, proposals, and projects for the IAs, and LDRD projects for the DOE complex.

The link analysis visualization for the entire DOE complex is represented in Fig. 8. Although there are several laboratories in the original analysis, only the strongest links between technologies and laboratories were extracted and visualized. Fig. 8 identifies the relationships between laboratories and technology, and thus laboratories with common technology competencies. For example, laboratory B has an area of common technical focus with laboratory A through lithography, laboratory C through fuel cells and biological systems, and laboratory D through biological systems and semiconductors. The identification of these common points directs us to "technology categories" that can be further analyzed to identify the portfolio of technology that characterizes the capabilities of each laboratory. For example, when clicking on the fuel cells node in Fig. 8 when using the ClearForest link analysis tool interactively, a large number of additional relationships appear. The relationships consist of additional laboratories and technologies that have weaker links than in the original visualization. Drilling down into a technology is a powerful analysis technique, and provides greater detail for the laboratory and IAs. The value of this analysis lies in its ability to identify the technological capabilities of each laboratory, in addition to determining whether duplication or collaborative opportunities exist.

The second analysis consisted of linking each individual IA to other laboratories in the DOE complex through common technologies. The analysis was conducted by selecting each IA in turn and exposing all laboratory and technology relationships associated with it. The result was a visualization that placed the IA in the middle of the link map with a minimum of 50 nodes identifying direct and indirect

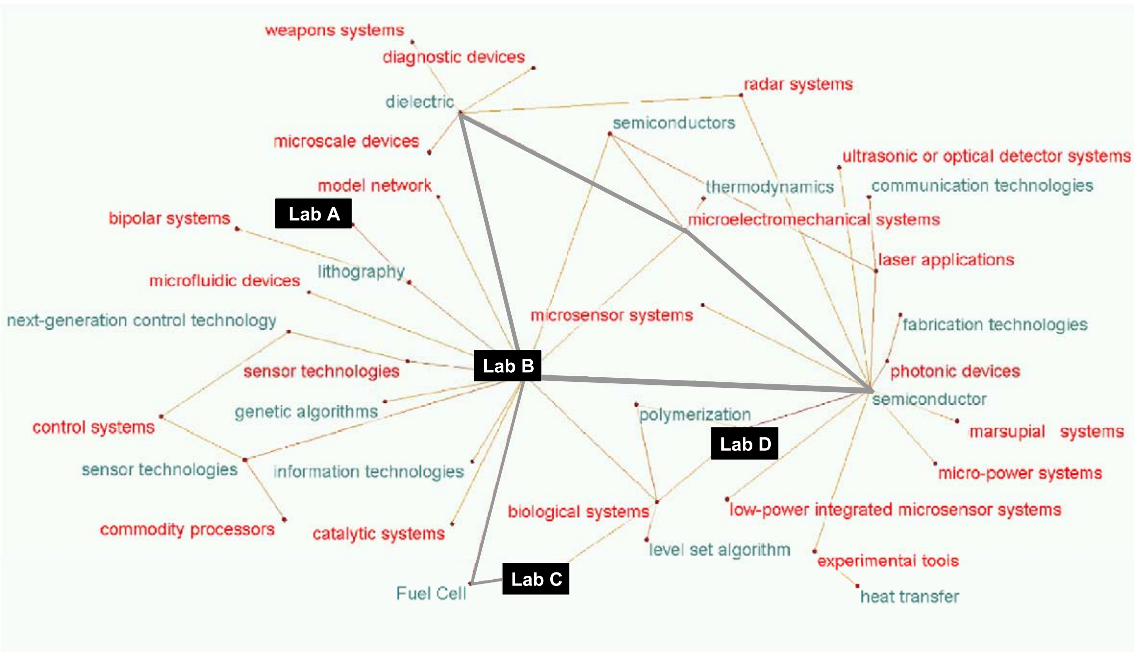

Fig. 8. ClearForest link analysis map of specific technology linkages between different laboratories within the U.S. DOE complex. Thicker lines indicate stronger relationships.

relationships. The direct relationships were explored to identify duplication or complimentary efforts. The indirect relationships were explored to identify complimentary technology outside of Sandia, and thus to assist in the identification of new but related applications outside of Sandia's original intended use, or to suggest potential collaborative opportunities between laboratories.

5. Future directions

This is the first year that we have applied such analyses to our LDRD process. Coming late in the annual process, the results have been more modest than they could have been. We plan to start a similar process for the FY06 LDRD process, and carry it out much sooner in the annual process.

We have learned that it is important not to saturate the IA leads with the information from these analyses, but rather to present some information, and then allow them to further explore the information on their own. It is only as those with funding authority internalizing the results of such analyses, integrating them into their mental models, and foreseeing how overlaps, collaborations, and new opportunities can benefit the return on investment to their IAs, that they will put the results into practice. We have also learned that one tool does not fit all situations, but that different approaches offer different perspectives and levels of detail that can each be of benefit to the analyst or manager.

The current approach of applying information visualization tools to the analysis of the LDRD portfolio enables a comprehensive assessment of the technological development trends occurring within our IAs. Insights gained from the novel application of visualization tools, coupled with the tacit knowledge that comes from years of personal experience as experts in a technical field, aid the IA leaders in forecasting the direction of technology development. Although this is not equivalent to more traditional and long-term forecasting methods such as Delphi studies or scenarios, it is nonetheless an effective means of guiding the science and technology enterprise in the shorter term. In particular, it can help to redirect or consolidate efforts to create a more focused and effective technology development program.

In the near future, we plan to expand our scope to include not only the LDRD information from DOE laboratories, but also much data from industry and academia. This will allow us to broaden the technology intelligence that forms the context of our maps, and thus broaden the 'opportunity space' that can be mined by our IAs. Better context also enables better avoidance of duplicative efforts and better knowledge of research risks. In a parallel effort, we plan to investigate different models of impact and join the best of those to our visualizations to answer questions related to return on investment [11]. As we expand our efforts and grow our maps, it is our hope that a global-mapping context will allow us to identify and forecast technology paradigm shifts, which in turn will allow us to take a stronger role in accelerating the development of cutting-edge technology. This is a research question and possible future that is worthy of exploration.

Acknowledgements

The authors gratefully acknowledge the support of the LDRD Program, Sandia National Laboratories, U.S. DOE under contract no. DE-AC04-94AL85000. Insightful comments and suggestions by anonymous reviewers are also acknowledged.

References

[1] K. Börner, C. Chen, K.W. Boyack, Visualizing knowledge domains, Annu. Rev. Inf. Sci. Technol. 37 (2003) 179–255.
[2] S. Deerwester, S.T. Dumais, T.K. Landauer, G.W. Furnas, R.A. Harshman, Indexing by latent semantic analysis, J. Am. Soc. Inf. Sci. 41 (6) (1990) 391–407.
[3] T.K. Landauer, D. Laham, M. Derr, From paragraph to graph, Proc. Natl. Acad. Sci. 101 (Suppl. 1) (2004) 5214–5219.
[4] G.S. Davidson, B.N. Wylie, K.W. Boyack, Cluster stability and the use of noise in interpretation of clustering, Proc.-IEEE Inf. Vis. 2001 (2001) 23–30.
[5] K.W. Boyack, B.N. Wylie, G.S. Davidson, Domain visualization using VxInsight for science and technology management, J. Am. Soc. Inf. Sci. Technol. 53 (9) (2002) 764–774.
[6] J. Dowling, Information retrieval using latent semantic indexing and a semi-discrete matrix decomposition, Unpublished Bachelor's thesis, Monash University, 2002.
[7] T.G. Kolda, D.P. O'Leary, A semidiscrete matrix decomposition for latent semantic indexing in information retrieval, ACM Trans. Inf. Sci. 16 (4) (1998) 322–346.
[8] K.W. Boyack, K. Börner, Indicator-assisted evaluation and funding of research: visualizing the influence of grants on the number and citation counts of research papers, J. Am. Soc. Inf. Sci. Technol. 54 (5) (2003) 447–461.
[9] K.W. Boyack, K. Mane, K. Börner, Mapping Medline papers, genes, and proteins related to melanoma research, Proc.-IEEE Inf. Vis. 2004 (2004) 965–971.
[10] S.K. Kim, J. Lund, M. Kiraly, K. Duke, M. Jiang, J.M. Stuart, et al., A gene expression map for *Caenorhabditis elegans*, Science 293 (2001) 2087–2092.
[11] K.W. Boyack, Mapping knowledge domains: characterizing PNAS, Proc. Natl. Acad. Sci. 101 (Suppl. 1) (2004) 5192–5199.

Kevin W. Boyack is a Principal Member of Technical Staff (PMTS) in the Computation, Computers, Information, and Mathematics Center at Sandia National Laboratories. His PhD in chemical engineering is from Brigham Young University. Dr. Boyack's current interests and work are related to information visualization, knowledge domains, semantics, metrics and indicators, cognition, and the integration and analysis of multiple data types.

Nabeel Rahal is a business intelligence researcher and analyst in the Business Development and Corporate Partnerships Center at Sandia National Laboratories. Nabeel holds a Master's degree in Management of Technology (MOT) from the Anderson Schools of Management at the University of New Mexico. His current areas of interest include the integration of information visualization technology with business intelligence.