

*January–April 2004*



# **EVALUATING THE IMPACT OF ELECTRONIC BUSINESS SYSTEMS LESSONS LEARNED FROM THREE CASES AT THE DEFENSE LOGISTICS AGENCY**

***Jonathan A. Morell, Ph.D.***

This article synthesizes our experience evaluating three electronic business (eBusiness) systems in the Defense Logistics Agency. The focus was on actual impact in real life operational settings. We summarize our experience in terms of lessons learned and make a case that our experience can help others do similar evaluation. Lessons learned are grouped into six categories: metrics and data sources, methodology, program logic, adaptive systems, realistic expectations, and dependencies among the previous five.

**T**his article synthesizes our experience evaluating the impact of three electronic business (eBusiness) systems in the Defense Logistics Agency (DLA). Our intention is to show the tactics that emerged when general principles of evaluation were applied for the context-specific purpose of determining whether, and how, an eBusiness system is affecting its environment. The first section outlines our emphasis on impact assessment and makes a case for evaluating eBusiness systems. The second section presents *lessons learned*

that were abstracted from our experiences and that can be applied to other, similar evaluation exercises. Finally, we illustrate how the lessons learned were combined to produce impact assessments of particular eBusiness programs.

## **IMPACT ASSESSMENT — DIFFICULTIES AND IMPERATIVES**

---

Our evaluation activities assumed that programs that have been deployed should have measurable consequences. In this we

are firmly rooted in the tradition of evaluation for impact assessment. This view, as summarized in a classic evaluation textbook, states that:

The critical issue in impact evaluation, therefore, is whether a program produces desired effects over and above what would have occurred either without the intervention, or in some cases, with an alternate intervention. (Rossi, Freeman, & Lipsey, 1999, p. 239)

Our core challenge in making such an assessment was the need for a methodology that could produce causal information within a context that from our evaluator's perspective, was totally uncontrolled. We had to evaluate a *natural experiment*, i.e., a situation in which "... program variants (or other treatments of interest) are not experimentally controlled but vary in the natural environment and in which causal inference is still desired." (Mark, Henry, & Julnes, 2000, p. 265).

In the present case, not only was the situation uncontrolled, but also entirely

**"Success required knitting together many disparate data sources and analyses."**

post hoc. Evaluation did not begin until after the programs in question were well established. As a result of the timing, it was impossible to influence implementation schedules, to anticipate data needs, or to establish data collection mechanisms. Of necessity, the evaluation design was *quasi-experimental*, an approach defined by Rossi, Freeman, & Lipsey (1999) "An impact assessment

in which 'experimental' and 'control' groups are formed by a procedure other than random assignment" (1999, p. 234). Data limitations, however, made it necessary to formulate tactics that went beyond simple comparisons of non-equivalent control groups. Success required knitting together many disparate data sources and analyses. Much of what will be reported below is the story of the search for those sources and the logic and methodologies used to integrate them.

Because of our emphasis on outcome assessment, we did not dwell on process metrics such as percentage of time a system was running, average time to resolve complaints, or number of users. Rather, we focused on whether, because the system was working, there was measurable impact on dollars, quality, time, or readiness. The objective was to determine whether, for operational eBusiness systems, it would be possible to:

- Obtain relevant data.
- Draw conclusions about what the program accomplished.
- Develop practical recommendations to facilitate further evaluation.

The answer was by no means certain because very few eBusiness programs are implemented in a way that is conducive to impact evaluation. To anticipate the later discussion, limitations on IT systems and inter-organizational agreements conspire to constrain evaluation possibilities. We discovered that despite these problems, it was possible to assess impact for each of these systems. This finding gives

us confidence (but no guarantee) that impact evaluation can also be conducted on other operational eBusiness systems. By presenting this information, we hope to convey a sensibility about how this kind of work can be done, and thus, to spur more such activity by a larger number of people. At the DLA's request, three eBusiness systems were studied: Electronic Document Access (EDA), Central Contractor Registration (CCR), and the Department of Defense (DoD) Emall.

**EDA** ([http://eda.ogden.disa.mil/eda\\_main.htm](http://eda.ogden.disa.mil/eda_main.htm)): The Electronic Document Access Web (EDA Web) combines Internet and World Wide Web technologies with electronic document management to eliminate paper files and facilitate information sharing among DoD communities to provide access to single-source DoD official documents. The information is maintained and available for access to authorized users in Portable Document Format (PDF). Documents included in EDA include contracts and contract modifications, MAAPR (materiel acceptance and accounts payable report), government bills of lading, and DD1716 forms (Contract Data Package).

**CCR** (<http://www.ccr.gov/>): In the past, any vendor who wanted to do business with more than one DoD site was required to submit the same business information to each and every site. This redundancy of paperwork not only created an administrative burden for both the government and the vendor, but also was a major source of administrative error and expense in terms of both time and money. Because DoD is the largest purchaser of goods and services in the world, the cost savings to be incurred

by streamlining these administrative processes are dramatic. CCR was created to be the single repository of vendor data for the entire DoD to avoid this administrative duplication and allow contractors to take responsibility for the accuracy of their own important business information by supplying it directly to the government through a single registration.

**DoD Emall** (<https://emall.prod.dodonline.net/scripts/EMlogon.asp>): The DoD Emall strives to be the single entry point for purchasers to find and acquire off-the-shelf, finished goods items from the commercial marketplace and government sources.

The evaluation work discussed here took place between 1999 and 2001. The specific findings are frozen in time, while the programs themselves have been evolving. Thus, conclusions concerning the systems that were evaluated may not be useful for current decision making. However, we believe that the lessons learned from that work are applicable to evaluation of other eBusiness systems in government settings.

Few impact evaluations of IT systems take place in government settings. But to calibrate expectations, it is important to realize that few such studies exist for any sector. Most of the research on the impact of IT focuses at its lowest level on the firm, and aggregates up from there. Much of this research deals with what is commonly known as the *productivity*

**“Because DoD is the largest purchaser of goods and services in the world, the cost savings to be incurred by streamlining these administrative processes are dramatic.”**

*paradox*, i.e., the disconnect between our intuitive sense that IT *must* have a beneficial impact, and the failure of researchers to observe that impact (Brynjolfsson, & Hitt, 1998; Chan, 2000; Macdonald, 2002).

**“Despite the difficulties, impact assessment of DoD eBusiness systems is needed to build a fund of knowledge, experience, and wisdom about what works.”**

A second body of research on IT deals with the role that IT plays in particular business processes. For instance, Malone and Crowston (1994) assess how IT affects inter-firm transaction costs, and by so doing, influences decisions about trading partner relationships. A similar fo-

cus is exhibited by Argyres (1999) in his research on how IT affected inter-organizational relationships during the development of the B-2 bomber. Studies like these make a good case that IT can play an important and beneficial role in shaping decisions about how an organization should behave. However, the focus of most existing research and evaluation is on particular processes, and is not cast in the form: “System X was implemented. What impact did it have?”

The reason this kind of evaluation is difficult is because when specific eBusiness systems are evaluated within a larger organizational context, four challenges to good measurement and good methodology are almost always present.

1. The system in question seeks to provide specific and limited improvements within a complex context of multiple interacting business processes and applications.

2. While the system may provide specific assistance to a well-defined group of users, it may also contribute to an overall information infrastructure. In contributing to the infrastructure, the system makes additional, and more diffuse, contributions to the development of other systems and to creative problem solving.
3. At the same time the system is being developed, other systems may also be under development.
4. Plans for impact assessment are not put in place during the programs’ development or initial deployment.

Despite the difficulties, impact assessment of DoD eBusiness systems is needed to build a fund of knowledge, experience, and wisdom about what works. As this understanding spreads within the DoD system development community, new systems will become more effective and more accountable.

In the next section we present lessons learned and examples of their application to the evaluations that were conducted. The subsequent section takes a deeper dive into the EDA evaluation and illustrates the lessons learned in greater detail.

## LESSONS LEARNED

---

Unambiguous instructions for doing post hoc outcome evaluation are impossible because evaluation settings differ with respect to the functionality of the system being evaluated, comparisons that can be drawn, data available, user base, and implementation schedules.

Collectively, these differences are bound to have major consequences for choices about design and analysis. Rather than be prescriptive, the intent of this section is to convey a sense of what issues must be considered, and how choices might be weighed, when deciding on how to conduct post hoc evaluation of eBusiness systems. The discussion is organized by lessons learned in six general categories:

1. Metrics and data sources.
2. Methodology.
3. Program logic.
4. Adaptive systems.
5. Realistic expectations.
6. Interactions among lessons learned.

### **METRICS AND DATA SOURCES<sup>1</sup>**

All relevant metrics are categories and combinations of dollars, quality, time, and readiness. The challenge is to define exemplars of these metrics such that trusted numbers can be found and analyzed. One major problem is that evaluation is usually commissioned by a system's owners. While those owners can provide rich process data (e.g., number of users, up-time, development cost), they usually do not control data relevant to impact. Those data tend to be owned either by a system's users, or a third party data collection function. To illustrate, owners of EDA believed that their system had a positive affect on the ability of the DoD to pay invoices on time. Making that case, however, required getting data from the Defense Finance and

Accounting Service (DFAS), an organization with which the evaluators had neither personal nor contractual relationships.

A derivative problem is that even if data owners are willing to help, their information systems may lack the capacity to yield the fine-grained data needed to evaluate a particular program. Further, no matter how big an organization, any given database is likely to have no more than two to five people who understand the database in sufficient detail to advise as to what information

can, and cannot, be extracted. Moreover, the identities of these people are difficult to ascertain because they tend to be organizationally distant from whatever point of contact an evaluation team may have, and also because job changes often necessitate talking to people about their former, not their present, jobs.

The above problems are exacerbated by the fact that *multiple* sources of data are likely to be needed. Thus efforts to find, get, and access information are multiplied. To illustrate, consider the complexity of information used in our evaluation of CCR.

- Relevant information came from the Department of the Treasury, data archives at three different DLA organizations, and the personal knowledge of many different people.
- Electronic Funds Transfer (EFT) volume and contract transaction volume

**“Further, no matter how big an organization, any given database is likely to have no more than two to five people who understand the database in sufficient detail to advise as to what information can, and cannot, be extracted.”**

were needed to construct ratios of actual savings to real savings. To do this, two different sources of contract volume were helpful in improving estimation accuracy.

- CCR implementation timelines were needed to assess the likely course of events, had CCR not been available. Transactions costs from the Treasury study were combined with transaction volume data to assess overall impact.
- Qualitative knowledge about CCR's role in process improvement led to a logic model, which dictated the analysis strategy.

**"Methodology is the logical structure in which data collection and analysis are carried out."**

#### **METHODOLOGY<sup>2</sup>**

Methodology is the logical structure in which data collection and analysis are carried out. Without a clear sense of that logic, there is no way to know what to do with metrics. For instance, an evaluation of EDA might re-

quire using the metric *time from a contract being finalized to its arrival at the Defense Contract Management Agency (DCMA)*. But how should this metric be used to draw inference about EDA? Is it necessary to track the metric weekly, monthly, or annually? Is it necessary to compare data at different locations within DCMA? Is there a need to differentiate between kinds of contracts? Is it necessary to obtain historical baseline data, or will current information suffice? Would it be beneficial to compare contract transmittal time to DCMA with

transmittal time to other agencies? Answers to these kinds of questions make a practical and significant difference for the kind of evaluation that can be done.

While the above example deals with a fine-grained metric, the problem scales. For instance, another metric might be development costs for IT systems, to be measured as part of an assessment of the accomplishments of the Clinger-Cohen Act. There is no doubt that the federal government has many metrics relating to the cost of IT systems. But would it be possible to compare these costs over a twenty-year period? Have the components of the metric changed over the years, and if so, have they changed in a way that invalidates historical comparison? Or, perhaps different federal agencies implemented the act in different years. Is the time difference in implementation, compared to the time scale of the metrics, conducive to comparison across agencies? Would the data allow sub-department level comparison? Depending on the answers to these questions, it may or may not be possible to implement different evaluation methodologies.

#### **PROGRAM LOGIC**

Choosing metrics and methodologies is greatly aided by developing a *program logic model* in order to answer the question: If the system works as planned, what will be different? This may not be an easy question to answer. A program's impact can be broader than indicated by meeting requirements for well-defined user groups. Proximate impact may induce secondary change. Time frames for impact may vary — some changes may occur immediately

upon system implementation, while other changes may develop over years.

Outcomes may interact with each other. By representing these phenomena in pictorial or tabular form, logic models force evaluators to identify what to measure, what measurements to compare, and when data analysis should take place. Developing these models has the added advantage of forcing collaboration between evaluators and stakeholders, and in achieving consensus among stakeholders as to what outcomes should be measured. (The field of Evaluation has a long history and extensive literature on developing logic models to drive evaluation. For an introduction, see Renger and Titcomb [2002].)

### **ADAPTIVE SYSTEMS**

The uses of eBusiness systems are not static. Of course all such systems have *core uses* that are enshrined in requirements and justification documents. These uses represent the main reasons a system was built, and their evaluation must carry through time. Focusing *only* on these uses, however, is almost certain to miss many important impacts. (Whether these are desirable or undesirable is an empirical question.) Any new eBusiness system represents a bundle of functionality that constitutes a tool people can use to solve problems.

As users become comfortable with their new tools, they will recognize new uses for the tools. These uses cannot be anticipated because experience with a tool is often a prerequisite for appreciating its value. Another reason is that personnel change over time and bring new skills and new perspectives to their jobs.

Additionally, the environment in which systems operate is not stable. It is entirely possible that by the time a system is fully deployed, new reasons to use it will appear. (The opposite may also be true. The original need for a system may disappear. This too, must be included in evaluation.)

A good example of newfound use is the case of CCR. CCR was originally conceived as a method of decreasing labor for data input by government personnel, decreasing the number

of times contractors had to provide the same data, and increasing data accuracy. All these were worthy goals, which may have justified CCR. However, as CCR developed, its true power came to be realized. For the first time, the government had a single, unambiguous identifier for all government contractors, a number that remained constant and reliable across contracts and across contracting activities. This ability turned out to have major benefits. For instance, it was instrumental in facilitating the government's move to electronic payment of invoices.

### **REALISTIC EXPECTATIONS**

One of the most frequent questions evaluators asks is some variant of: "What are your expectations for what this system will accomplish?" The usual answers are almost always wildly optimistic. Perhaps a system's owners can't get out of selling mode, or perhaps they have come to believe their own rhetoric — but for whatever reason, claims about a

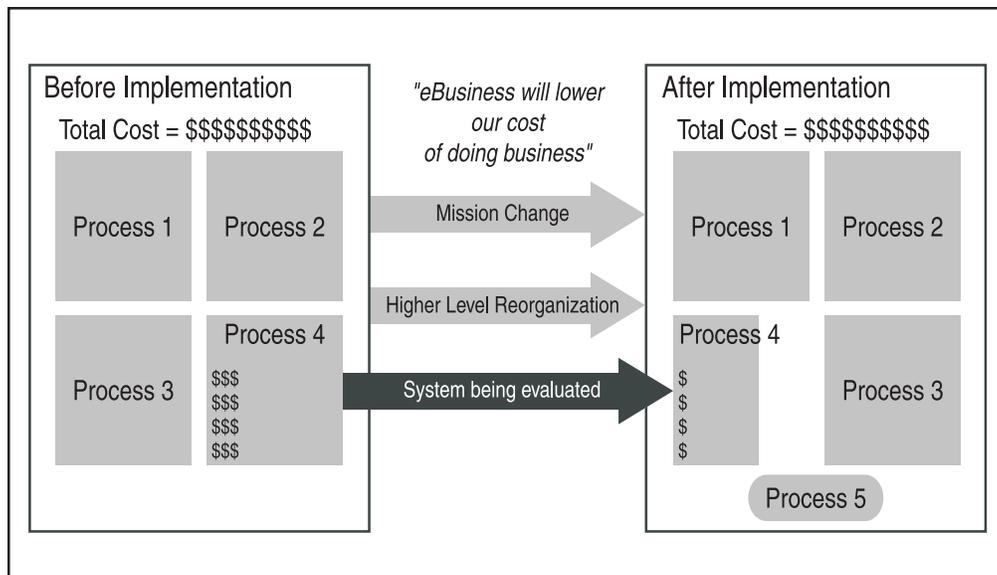
**"Any new eBusiness system represents a bundle of functionality that constitutes a tool people can use to solve problems."**

system’s accomplishments are often far beyond any reasonable boundaries of real world impact. Woe to the evaluator who takes these statements at face value and proceeds to do an excellent job of measuring the program relative to those projected outcomes. And woe to the program’s owners, who will receive only bad news about the value of their efforts. The disappointment has real and important consequences.

First, program managers do need to justify their programs. Evaluation relative to impossible goals will not provide that justification. Second, program managers need evaluation data to help them build on accomplishments. Without knowledge of what actually happened, needed guidance is missing. Third, evaluation almost always requires the cooperation of those being evaluated. Over time, assessment that brings only bad news will poison the climate for doing evaluation.

While almost everyone has an intuitive understanding of these dynamics, we have found that the generic logic model shown in Figure 1 is extremely useful in driving the point home and in facilitating the kinds of conversation needed to identify measurable achievements. Figure 1 depicts a program made up of four processes. An eBusiness system is implemented for the purpose of lowering the program’s overall costs. Upon close inspection though, it’s obvious that the new business system will affect only Process Four.

While the new eBusiness system may improve Process Four, it may not change the total cost of doing business because mission change, or high level reorganization, may affect the scale of the program’s activities. Also, changes in Process Four may facilitate other internal changes within the program. Using a picture like Figure 1 helps get stakeholders to address crucial questions about scope.



**Figure 1. Realistic Expectations**

What specific process will be affected? If those processes were improved, how much total change in the organization could be expected? If new functionality became available, what new processes might appear? What external forces are operating that might affect the impact of the system being evaluated?

Of course evaluators must not cook the books. There is a duty and an obligation to provide accurate information, even when that information will work to the detriment of some stakeholders. Programs are justified to funders based on specific claims, and it is important to hold managers to their claims. The solution is to employ a variety of tactics. First, the messy world of program justification and development is a web of political, budgetary, and bureaucratic forces that requires successful managers to make different claims, in different ways, to a variety of groups. While some of those claims will be core justifications that must be evaluated, others will not.

Second, eBusiness systems will have intermediate and localized impacts that are desirable, and that provide useful feedback for program improvement. These must be measured. (Of course not all the local or intermediate outcomes may be desirable, and these too must be assessed. Not only is doing so necessary for a fair evaluation, but the information can also be extremely useful for designing mid-course corrections.)

### **INTERACTIONS AMONG LESSONS LEARNED**

For the sake of exposition, the lessons learned were presented as if each were distinct and independent. In reality, they are inextricably linked. The process of evaluation should be seen as a continual

scanning for these relationships as the life cycle of an evaluation unfolds. A good example of this process involves the interaction between *data sources* and *methodology*.

One of our early plans for an evaluation design was a time series analysis of a particular transaction at a particular agency. The idea was to compare trends before and after implementation. The plan seemed especially appealing be-

cause we knew that the system had been introduced at different times in different parts of the organization. Our team was attracted to the possibility of making comparisons both over time and across organizational subunits. We formed this plan because trusted informants assured us that the data we needed had been collected over a long period. This information proved correct, but other facts emerged as we investigated the possibility of getting that data.

First, the older information was contained in a system that had been phased out and, while theoretically available, was not obtainable in practical terms. Second, the data were not collected at frequent enough intervals over the several years we needed to provide enough data points. Third, the way in which a critical data field was defined had changed over time, thus making historical comparisons problematic. Finally, the agency itself had changed organizational structure over the years. As a result, it was not possible to compare change over time either within, or across, the various subunits. In light of these

**“The process of evaluation should be seen as a continual scanning for these relationships as the life cycle of an evaluation unfolds.”**

discoveries, it was necessary to abandon the time series methodology in favor of more localized comparisons.

In terms of the practice of evaluation, it is important to note that our initial plan was based on information from well-meaning people with good knowledge of the eBusiness system involved, the agency in which it was used, and the data that were generated. However, it was only after we had a chance to talk to many mid-level and lower-level personnel that we were able to get the specifics needed to make an informed judgment about whether a time series methodology was practical.

**APPLYING LESSONS LEARNED:  
THE EXAMPLE OF ELECTRONIC  
DOCUMENT ACCESS**

---

Electronic Document Access affects so many processes that a wide variety of input was needed to make decisions about

logic models, metrics, and methods. The organizations whose input influenced the EDA evaluation included the: Defense Contract Audit Agency, Defense Contract Management Agency, Defense Finance & Accounting Service, Defense Information Systems Agency, Defense Logistics Agency, Directorate for Information Operations and Reports, Fitting Out and Supply Support Assistance Center (FOSSAC), Navy/Air Force Interface, Office of the Secretary of Defense — CIO Office, and several Army Commands.

At the time this work was carried out, the most extensive use of EDA was for the management of contracts and contract modifications. Thus “contracts” became our primary focus. Potential metrics were cast within a Balanced Scorecard framework because at the time of this project, Balanced Scorecard was being heavily used in the DLA.<sup>3</sup> We felt that even though our work was unrelated to that Balanced Scorecard activity, using Balanced Scorecard categories would help our

**Table 1. Reasons for EDA Impact**

Business Process	Balanced Scorecard Category	Reason why EDA may be Helpful
DFAS invoice processing	Financial	DFAS requires complete paperwork before it can process an invoice. EDA: 1 - reduces time from document creation to its arrival at DFAS, and 2 - assures a single complete set of contracts and associated modification. The result is decreased time for invoice processing, fewer aged invoices, and better compliance with the Prompt Payment Act.
Contract/mod creation, distribution	Financial  Internal process	EDA has the potential to decrease labor effort for contract management, and as such, has financial implications.  Consistent with any organization's ability to adapt to circumstance, decreased labor effort for any given task will result in a reordering of work priorities, or the development of new processes.

stakeholders form useful linkages among parallel, but conceptually related, activities.

Using a logic model perspective, we articulated why EDA should affect the metrics identified. The mechanisms of action are presented in Table 1. (Table 1 also illustrates the notion that while logic models are usually represented in graphical fashion, tabular descriptions can also be useful.)

**EDA IMPACT: CONTRACT PROCESSING LABOR AND INTEREST SAVED ON OVERAGED INVOICES**

Data used in this analysis, and their sources, appear in Table 2. This analysis again illustrates the need for multiple sources of data, some of which reside in data archives, and some of which were

developed for a specific, empirical investigation of a program. In the present case, the data came from FOSSAC's detailed and careful assessment of how EDA affected their contract processing efforts. For their contracts, we had good information on labor hours and interest payments due to over aged invoices for the 2000 and 2001 fiscal years, i.e., the time immediately before and immediately after the adoption of EDA.

The limitation on the FOSSAC assessment was that it covered only a small number of contracts. To scale up the findings, it was necessary to determine the historical number of similar *paperless* transactions for the whole Department. The extra effort to determine the percent paperless

**Table 2. Data Used in Assessment of EDA Impact on Contract Processing Labor**

Data	Use	Source
Historical data on DFAS workload	Contextual understanding of how DFAS worked, the pressures operating on the Service, and why better access might be important.	DFAS
Per-contract impact of EDA, time before and after EDA implementation at the Fitting Out and Supply Assistance Center (FOSSAC)	Hard data on change due to EDA. Used as basis for scaling up estimate to the DoD.	FOSSAC*
Contract volume per year for DLA, Air Force, Army, Navy	Used to scale up local impact to DoD.	1 - OSD CIO Office 2 - DD350 database
% paperless transactions	EDA only contributes to change for processing of paperless transactions. "% paperless" is needed to avoid applying analysis to the total contract volume.	1 - OSD CIO Office 2 - DD350 database
* Data courtesy of Bonnie Brown-Murphy, Management Program Analyst, Fitting Out Supply Support Assistance Center, Special Projects. Original source: Electronic Commerce Solutions Corporate Information Management Board, Paperless Working Group, Oct 9, 2001.		

**Table 3. Hours Made Available Due to EDA**

Year	98	99	00	01	02	Total
Hours/year	3.8K	42.4K	49.8K	51.7K	51.7K	250.9K

was critical because although it is relatively easy to find the total number of contracts, EDA provides labor savings *only* for that percentage of the transactions that were paperless. As a result, two data sources had to be used: the first on contract transaction volume, and the second, on paperless transactions. To make this determination, two data sources were combined.

The first was information on total contract volume. The second was percent paperless data that began with FY98 and

ended with the third quarter of FY01. Using all this information, it was possible to calculate both the number of labor hours that were no longer required for contract processing due to EDA and the savings in interest payments due to EDA. This information is summarized in Tables 3 and 4. Data were projected several years into the future. We ended the analysis at FY03 because while projections into the future are legitimate, the further the projection, the greater the inaccuracy. Also, we had reason to believe that another program —

**Table 4. \$M Savings, Interest on Overaged Invoices Attributable to EDA**

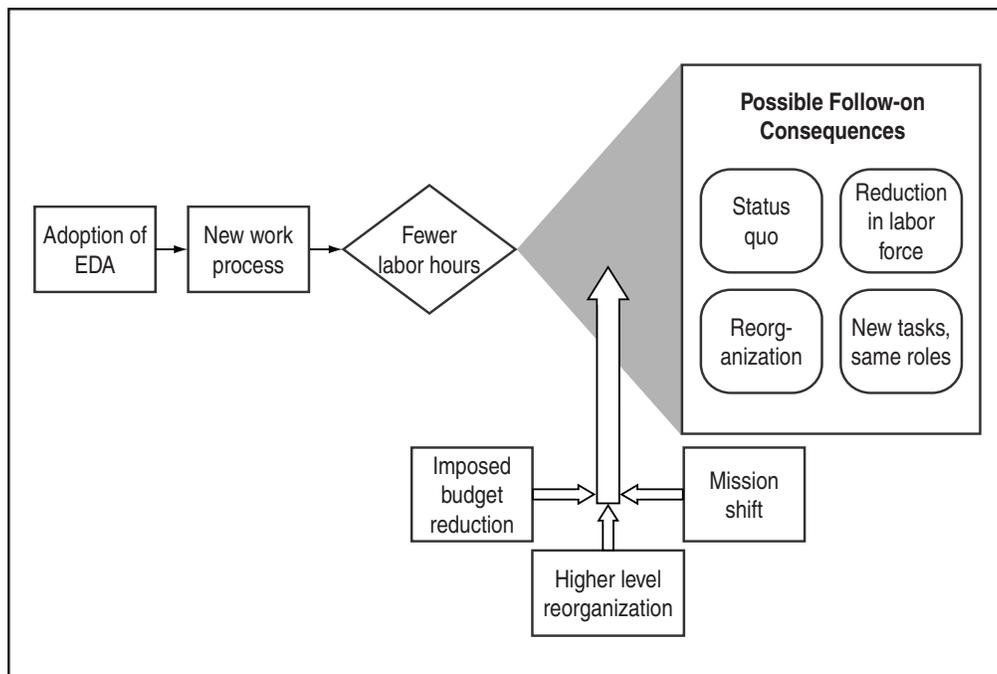
Agency	FY97	98	99	00	01	02	03	Total
Army		0.35	0.79	0.95	0.95	0.95	0.95	4.94
Navy			0.45	0.99	1.30	1.30	1.30	5.34
Air Force				0.44	0.52	0.52	0.52	2.00
DLA			7.14	7.43	7.44	7.44	7.44	36.89
Total		0.35	8.38	9.81	10.21	10.21	10.21	49.17
Cumulative		0.35	8.73	18.54	28.75	38.96	49.17	
ROI								
NPV of Savings								43.31
Investment	2.6	1.0	1.0	1.0	1.0	1.0	0.5	
NPV of Investments								7.33
ROI								5.91
Notes:								
1. Conservative estimate. Does not include impact on discounts earned, bills of lading, vouchers, MAAPR, DD1716, or \$ value of new activities.								
2. 02/03 projections based on 00/01 data.								
3. Unadjusted \$.								
4. Return on Investment (ROI).								
5. Net Present Value (NPV)								

Wide Area Work Flow (WAWF) — would come on-line in about three years, at which point the unique impact of EDA would be blurred by the combined consequences of both programs.

The approach taken here highlights possible interactions between decisions made about metrics and decisions made about methodologies. Our initial inclination was to find one or two metrics that indicated the impact of EDA and that could be collected on an organizationwide basis. Had we been able to do this, some relatively simple comparisons or time series analyses would have sufficed to provide the information we were after. Once we learned that no such metrics were possible, we began to cast about for alternate metrics and, as we did so, for methodologies that could exploit those metrics. This process led to the tactics we actually used, i.e.,

we took a micro-level view of good impact data and brought in multiple data sources to scale up the findings to a broader level.

The data in Tables 3 and 4 illustrate some of the limits that must be accepted when doing post-hoc evaluation of this type. While we could estimate the number of hours that no longer had to be devoted to contract processing, we were *not* able to determine how organizations adapted to that change. Unanswered questions included: Did they decrease their labor force? Did they reorganize? Did they deploy the workforce to other, truly value added activities? Any of these (in multiple combinations) were possible and were likely to vary from setting to setting. Because no mechanisms were in place to get the needed data, a comprehensive evaluation would have required



**Figure 2. EDA: Direct Impact and Second-Order Consequences**

an impractical plan that far exceeded available resources. First, it would have been necessary to identify the locations where these changes had been taking place. Second, different methodologies would have been required for each set of outcomes.

The problem of data access is a practical limitation, but another limit touches on the fundamental question of what impacts should be expected from any given program. To understand the issue, logic modules can be of assistance. Figure 2 illustrates that while *labor hour savings*

can reasonably be expected to result directly from EDA, the follow-on consequences of labor hour savings are affected by powerful forces that EDA cannot influence. The immediate impact of EDA is that as people start to use it, they spend less time in the paper processing aspect of contract management. But what happens once the time is saved? There could be a change in the size of the workforce, or in the nature of the organization, or in the nature of work. However, none of these changes are direct and immediate impacts of EDA.

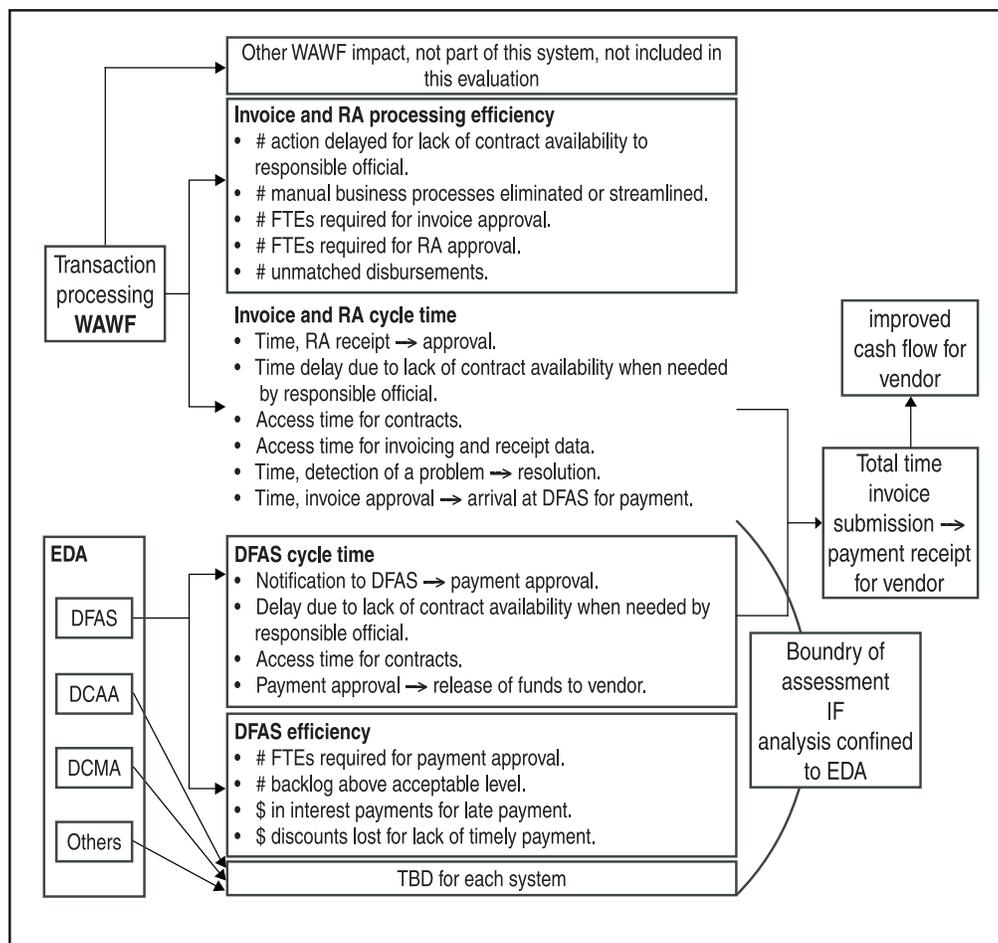


Figure 3. EDA - WAWF Interaction

In addition to the operation of outside forces, the impact of any single eBusiness system is constrained by interactions among multiple eBusiness systems. In any large organization, many different process improvements and eBusiness implementations are likely to be under way. Any single system is part of a larger developing infrastructure. Change in multiple parts of the infrastructure is needed to have truly profound impact. (Multiple systems are also the root of many methodological difficulties because evaluation requires teasing out the impact of one system from the combined impact of several.) The need to limit expectations for any single eBusiness system is illustrated by the relationship between EDA and WAWF.

One of our early logic models (Figure 3) took a very broad view of EDA. In doing so, it included the expected advent of WAWF, and it also took a longer-range view of likely outcomes. As Figure 3 shows, EDA alone can be expected to improve internal processing efficiency at DFAS. DFAS processing time, however, is only a part of the total cycle time from when a vendor submits an invoice, to the time payment is received. For the entire cycle time to be improved, WAWF would be needed to shorten many other cycle times that are part of the entire process.

## **CONCLUSION**

---

It is difficult to evaluate the impact of eBusiness systems in real life operation

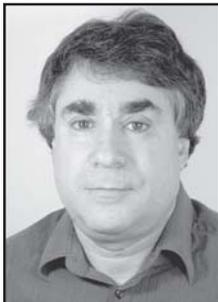
because the data needed do not cleanly follow the contours of a system's application. This is true both organizationally and temporally. From an organizational point of view, existing data often cannot differentiate those parts of an organization that are using a system from those that are not. From a temporal point of view, data may not be available over time periods that will allow before and after comparisons to match a system's implementation schedule.

Many variations on these themes exist, and many problems derive from these difficulties. For instance, useful data may be trapped in archaic systems. The definition of data elements may change over time. Because clean data cannot be found, multiple data sources are needed to triangulate on a conclusion, and the greater the number of data sources, the greater the likelihood of having to negotiate with recalcitrant data owners. Despite these problems, successful impact evaluation can be carried out, and guidelines — lessons learned — can be abstracted from past efforts that are applicable to future efforts. (To aid in this application, Table 5 summarizes critical issues.) We hope we have convinced the reader of this conclusion, and that by so doing, spurred further efforts at eBusiness system impact assessment.

**“It is difficult to evaluate the impact of eBusiness systems in real life operation because the data needed do not cleanly follow the contours of a system's application.”**

**Table 5. Critical Questions Within Lessons Learned**

Critical Questions Within Lesson Learned
<p><b>Metrics and data sources</b></p> <ul style="list-style-type: none"><li>What data are needed?</li><li>Who owns the data?</li><li>If data are not owned by group that commissioned the evaluation, can the necessary data be obtained?</li><li>Who/where are the few people who truly understand how needed data bases are constructed?</li><li>Can the data be extracted for the time period, and at the level of granularity, needed for the evaluation?</li><li>Are the data reliable?</li></ul>
<p><b>Methodology</b></p> <ul style="list-style-type: none"><li>What comparisons can be made to determine the program's impact?</li><li>What are the specific targets (e.g. users, business processes) of each comparison?</li><li>What are the threats to validity for each comparison?</li></ul>
<p><b>Program Logic</b></p> <ul style="list-style-type: none"><li>Who are the groups that must agree on what the system should be able to do?</li><li>What groups and business processes should be affected?</li><li>What are the proximate and secondary impacts?</li><li>What elements of a system must be in place before any particular impact can be manifest?</li><li>What are the key dependencies in the system and among impacts?</li><li>What are the time frames for particular impacts to appear?</li></ul>
<p><b>Adaptive Systems</b></p> <ul style="list-style-type: none"><li>As a system becomes known, how does its availability affect decisions about what problems should be solved or opportunities pursued?</li><li>How is the business environment affecting beliefs about how a system should be used?</li><li>What new systems are being implemented that draw on the functionality of the system being evaluated?</li><li>As new uses of a system develop, which ones are important enough to be assessed?</li><li>Can evaluation tease out the contribution of one system from another?</li></ul>
<p><b>Realistic Expectations</b></p> <ul style="list-style-type: none"><li>What are the critical claims for a system's value that must be measured?</li><li>What claims on their face are unlikely to occur?</li><li>What reasonable impacts were not originally envisioned for the program?</li></ul>
<p><b>Interactions Among Lessons Learned</b></p> <ul style="list-style-type: none"><li>Does development of the evaluation methodology follow a "waterfall" of a "spiral" model?</li><li>Is there a process in place to detect how developments within one lesson category may affect the others?</li><li>Does the evaluation team have the expertise needed in qualitative and quantitative methods to integrate an evaluation approach across all lesson categories?</li></ul>



**Dr. Jonathan "Jonny" A. Morell** is an organizational psychologist with expertise in evaluating the products, services, and activities that constitute electronic business. He is Editor-in-Chief of the *Evaluation and Program Planning* journal, and is on the editorial board of the *International Journal of Electronic Business*. He has a doctorate from Northwestern University and is a recipient of the American Evaluation Association's Distinguished Service Award. Presently, Morell is a senior policy analyst at the Altarum Institute's Enterprise Solutions Division.

(E-mail address: Jonny.Morell@altarum.org)

**REFERENCES**

---

- Argyres, S. N. (1999). The impact of information technology on coordination: Evidence from the B-2 "Stealth" Bomber. *Organization Science*, 10(2), 162–180.
- Brynjolfsson, E., & Hitt, L. M. (1998). Beyond the productivity paradox. *Communications of the ACM*, 41(8), 49–55.
- Chan, Y. E. (2000). IT value: The great divide between qualitative and quantitative and individual and organizational measures. *Journal of Management Information Systems*, 16(4), 225–261.
- Kaplan, R., & Norton, D. (1996). Using the balanced scorecard as a strategic management system. *Harvard Business Review*, 74(1), 75–85.
- Macdonald S. (2002). The IT productivity paradox: Technological determinism masked my managerial method? *International Journal of Information Technology and Management*, 1(1), 1–29.
- Malone, W. T., & Crowston, K. (1994). The interdisciplinary study of coordination. *ACM Computing Surveys*, 26(1), 87–119.
- Mark, M. M., Henry, G. T., & Julnes, G. (2000). *Evaluation: An integrated framework for understanding, guiding, and improving policies and programs*. San Francisco: Jossey-Bass.
- Martinsons, M., Davison, R., & Tse, D. (1999). The balanced scorecard: A foundation for the strategic management of information systems. *Decision Support Systems*, 25, 71–88.
- Renger, R., & Titcomb, A. (2002). A three-step approach to teaching logic models. *The American Journal of Evaluation*, 23(4), 493–503.
- Rossi, P. H., Freeman, H. E., & Lipsey, M. W. (1999). *Evaluation: A systematic approach* (6th ed.). Thousand Oaks, CA: Sage Publications.

**ENDNOTES**

---

1. This article cannot serve as a complete treatment of measurement issues in evaluation. For a good introduction to this topic, see Rossi, Freeman, & Lipsey (1999).
2. As with the topic of measurement, this article cannot serve as a complete treatment of all-important issues in evaluation. For a good introduction, see Rossi, Freeman, & Lipsey (1999).
3. Balanced Scorecard in an organizational planning and assessment approach that casts leading and trailing indicators into four general categories: financial, customer, internal business process, and growth. It has been adapted for other contexts, but the principle of using measures from multiple domains is consistent. Diversity of measures is the Balanced Scorecard's greatest strength. When a single overriding metric is imposed on a system, the system will maximize that metric. Other crucial aspects of organizational functioning will be ignored, thus threatening the organization's long-term viability. The power of the Balanced Scorecard is that it helps organizations pursue the joint optimization of metrics that relate to different critical domains. For a general discussion of the Balanced Scorecard, see Kaplan and Norton, (1996). For a discussion of applying Balanced Scorecard to information systems, see Martinsons, Davison, and Tse (1999).

