Tools for Managing and Measuring the Value of Big Data Projects

Abstract

Big Data and analytics focused projects have undetermined scope and changing requirements at their core. There is high risk of loss of business value if the project is managed with an IT centric waterfall approach and classical project management methods. Simply deploying technology on time, to plan, and within budget does not produce business value for Big Data projects. A different approach in managing projects and stakeholders are required to execute and deliver business value for Big Data and analytics focused initiatives.

Introduction

Projects that are designed to drive better decisions in organization can deploy technology on time, to plan, and within budget and completely fail to deliver business value. In the race to extract insights from the massive amounts of data now available, many companies are spending heavily on IT tools and hiring data scientists. Many are struggling to achieve a worthwhile return. Big Data and projects that are treated the same way as IT projects for the most part fail to demonstrate the hoped for business value. This paper will discuss why Big Data and analytics projects must be treated differently to achieve business changing outcomes.

Discovery driven Projects

To obtain value from analytics projects focus must be on solving business problems rather than managing risk of deploying technology. The desire to move to more scientific based management practices using analysis of large and disparate data results in the possibility of change for business processes and the way information is used. This is in contrast to simply optimizing technical processes which is a historical IT strong suit. Organization learning and organizational change as well as decision support and business performance improvement are the outcomes to show value from analytics projects. Standard project management tool and measures are not sufficient to track the delivery or insure the value of analytics efforts. Tools borrowed from the non-profit world of mission and vision driven measures are well suited and can be tied to show applicability to business needs. These include concepts from formal program evaluation and the creation of Logic models which are methods for framing change measures. Logic models frame the hoped for outcomes based on new findings from the analysis. Insights from Information previously hidden can be used to drive change and allow it to be tested in an organizational setting.

Success factors for Big Data Projects

Data-driven decisions and Big Data applications create value by utilizing data sources to discover and put into operations new knowledge that improves business processes or drives innovation. Many analysts have written about the factors that contribute to the success of analytics focused projects.

These include the management of the project of course but also include other factors. Below is a summary of some of these factors.

Technology: Big Data and analytics projects are more than the technology on which they are based. New technology stacks that include Hadoop, columnar databases, visual analytics and technology that parse text based and video based data are all adding to the ability to bring in novel information. It is important to note however that existing technology for pulling information from databases such as SQL and even open source analytics tools like R can enable organizations to gain a foothold in advancing the ability to gain business insights.

People: Like all successful projects it is important to identify the project sponsor to remove obstacles, find the budget, provide organizational support, and serve as project champion. Just as important is the need to identify a project manager and the technical and business teams and have an outline of the roles and responsibilities of each team member. Since it is likely that most team members will have other obligations it is important to determine the team's availability and resource constraints for the project.

The title data scientist brings to mind a group of elite Ph.D. level personal that have highly specialized skills and in very high demand. This is not the case and multidisciplinary teams are best. There are projects that may require this level of expertise but most projects can be staffed with teams of people with quantitative, computational, or business expertise backgrounds. Creativity and attitude to "get the job done" is also critical. Some additional training may be needed to incorporate scripting languages; visual analytics tools and technologies such as Hadoop, however training for these are becoming more widely available.

Scope and Change management: To be successful it is essential that Big Data projects have a determined scope, and be time boxed. A clear scope communicates clearly and specifically what is included and what is not included in the effort. The defined scope is needed to develop a rough budget and to keep tabs on progress. It is also important to establish a set timeline and successful milestones at deliverable commitments set at 2-3 month intervals and should not exceed, 6 months to deliver business value.,

Projects of any sort big or small need to adapt to changes. These changes may be market forces that require refocus on different aspects of the endeavor or late discovered or missed essential tasks that need to be added to make the project successful. Not mater the source of the ability of a project to adapt is key to its success.

Change management is also needed for projects that are prescriptive in nature. These analytics project often involve automated systems that tell customer facing or front-line workers the optimal path or decision to make. Applications that involve changes to work flow affect business process and the people who preform them directly. Therefore change management is needed or the analytics efforts will be at risk for creating unhelpful disruptions or simply a waste of time.

Clear business objectives: One of the first steps in determining if a project is going to succeed is to set specific success criteria. These criteria must be measurable just like a specific key performance metric. The objective or success criteria must be documented as precisely as possible to insure its correct communication. Each business objective needs to have measurable criteria that will determine if that

objective has been met successfully. These criteria just like the need for clear project scope work best when they are shared and approved among the project key stakeholders.

Part of the purpose of Big Data projects is to shift through mountains of data to find meaningful relationships between variables and data elements that can be adapted for business use. To know the value of the relationships, however, requires a well-articulated business question that is understood by the stakeholders and individual contributors.

Identifying individuals who have a higher than average propensity to submit fraudulent insurance claims or identifying telecom customers who have a higher than average likelihood to churn are both measureable goals. Shifting aimlessly thru piles of data may find a nugget of knowledge that may produce some surprising and useful results. The same discovery can be made, however, when analysis is focused to meet a specific goal and produces more consistent positive outcomes.

Project management: Even well managed project sometimes fail. It is reported that Big Data projects fail 30% more often than other IT projects. The biggest reasons these projects failed was inaccurate scope followed by unconstrained expansion of requirements or lack of a good set of project objectives. Without having firm success criteria, projects continue to expand without realigned timelines, fail to demonstrate positive return on investment or in the worst case – fail to meet objectives and provide business value.

A fundamental aspect of project management are that all project parties are either bound by, or incented by one or more of the triple constraints (budget, scope, and schedule) to attain or measure success. This is also true of Big Data Projects.

In an endeavor that uses project management concepts the project team measures progress and status against budget (cost or funding), schedule (or timing), and scope (requirements or quality). Project parties use one or more of these three elements to measure success from their particular vantage points. This is not different with Big Data projects and the same constraints apply.

Continued cooperation between departments and issue management are also essential to the success of Big Data Projects. The insights gained from an analytics focused Big Data project are broader than helping one stakeholder. The insights gained in analytics based projects are often something that improves performance for a lot of different parties including IT teams, data scientists and analysts, and people across the organization from line of business managers to executives. Open lines of communication and reminders of the goals of the project are needed to keep projects on track and to deliver business value.

Big Data projects are more like R&D than production applications and therefore have a higher risk of not aligning with original expectations or failing. Collecting and communicating lessons learned from each project or project iteration will increase the change for success in future projects.

Iterative discovery for better business outcomes and adaptive learning

Because of the different way people create and use information, iteration and discovery are essential components of Big Data projects. The iterative nature of analysis and discovery also means that projects cannot be mapped out in a neat fashion. At the beginning of the project specific goals and scope is

required but no one knows the exact decisions the tools will be asked to support and the questions they will be expected to help answer as the project progresses.

There are many aspects of Big Data or analytics projects that benefit from formal project management methods, but not all. Conventional large IT projects include defined outcomes, required tasks, and detailed project plans for carrying them out. The iterative nature of Big Data projects don't fit neatly into being managed this way. Big Data projects may be small and shorter initiative efforts designed to frame questions to which the data might provide answers and develop hypotheses. This can be followed by iterative experiments and data gathering to gain knowledge and understanding and apply the learnings to organizational processes.

Over all the critical success factors for Big Data and Analytics projects is similar to those for general Business Intelligence Implementations and represents 'system thinking' rather than waterfall development. These include the following:

1. Business alignment/partnership
2. Business executive sponsorship (by definition, adequate funding)
3. Adequate resources, project management and governance
4. Architectures for Enterprise Information Management
5. Rapid, incremental and iterative development
6. Metadata foundations (especially business metadata)
7. Information quality
8. Adequate development and technical skills
9. Appropriate tools and processes
10. End-user training and change management.

Program Logical Models to guide and measure stakeholder learning

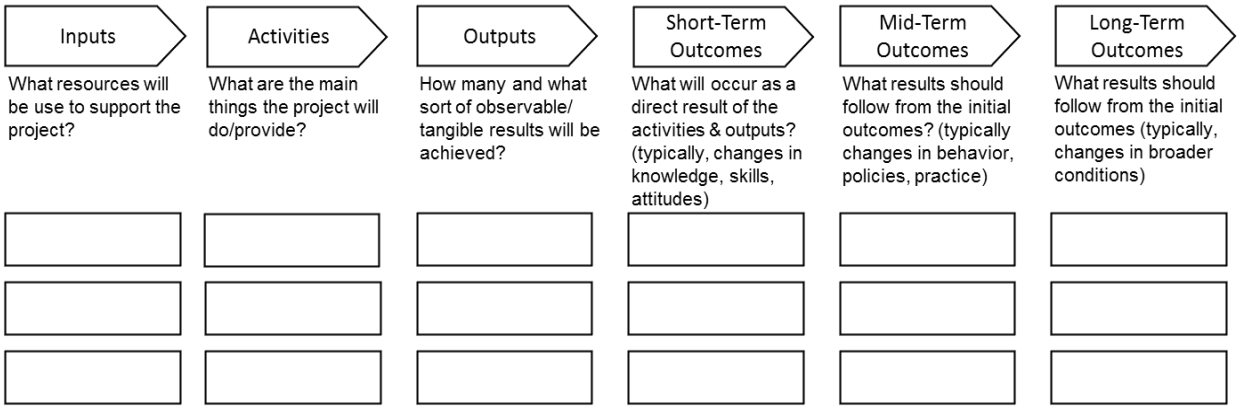
Program evaluation is a systematic method for collecting, analyzing, and using information to answer questions about projects, policies and broader programs. It enables evaluation of effectiveness and efficiency of interventions which can be applied to organizational change. One of the key tools of program evaluation and planning is the logic model.

The logic model is often used in government or not-for-profit organizations, where the mission and vision are broader than achieving a directly measured financial benefit. The process for creating Logic Model has been used successfully for at least twenty years to clearly communicate program aspirations and assumptions and document outcomes. They are increasing becoming a requirement for funding for socially focused programs and have proven a very valuable communication tool. A Logic Model presents a plausible and sensible model of how a project or program will work under certain conditions to solve identified problems. The Logic Model is the basis for a creating a convincing story of the program's expected performance to a broad audience. The elements of the Logic Model are resources, activities, outputs, customers reached, short, intermediate and longer term outcomes, and the relevant external influences. These same concepts can be used for broader business focused projects as well especially if longer term outcomes are the goal.

The benefits of using a Logic Model in setting the framework for Big Data projects include:

- Building a common understanding of the program and expectations for resources, customers reached and results, are good for sharing ideas, identifying assumptions, team building, and communication;
- Helpful for identifying opportunities for iterative improvement, identifying projects and points of analysis that are critical to goal attainment, redundant, or have inconsistent or implausible linkages among program elements;
- Communicating the place of an analytics project or program in the organization, particularly if there are shared logic models at various management levels;
- Points to a balanced set of key performance measurements and evaluation issues,
- Highlights areas for improvement for data collection and future iterations.

Figure 1 – Elements of a Logic Model



Inputs/Resources include human and financial resources as well as data and information inputs required to support the program.

Activities include all those action steps necessary to produce program and project outputs.

Outputs are the products, goods and services provided to the project or programs direct customers. For example, conducting research is an activity and the reports generated for other researchers and technology developers could be thought of as outputs of the activity.

Outcomes are characterized as changes or benefits resulting from activities and outputs. Programs and projects typically have multiple, sequential business or social outcomes. First, there are short term outcomes, those changes or benefits that are most closely associated with or "caused" by the program's outputs. Second, there are intermediate outcomes, those changes that result from an application of the short term outcomes. Long term outcomes or program or business process impacts, follow from the benefits accrued though the intermediate outcomes. For example, results from a laboratory prototype for an energy saving technology may be a short-term outcome; the commercial scale prototype an intermediate outcome, and a cleaner environment once the technology is in use one of the desired longer term benefits or outcomes.

Conclusion

The way Big Data projects are organized and managed is as important as the technology and data sets involved. Big Data projects however cannot be fully successful if constrained by traditional methods of implementing IT systems. Big Data projects do not have full answers when they start, but focus on finding them through evidence based discovery. This discovery is best accomplished when viewed with overall organizational goals and aspired outcomes in mind that an evaluation logical model can provide.

References

Donald Marchand and Joe Peppard. *Why IT Fumbles Analytics* - Harvard Business Review Jan-Feb 2013.

CONTACT INFORMATION

Name: Rob Phelps
Enterprise: HP Enterprise Services – Analytics and Data Management

City, State ZIP: Winston-Salem NC 27106
Work Phone: (336) 602-3177
E-mail: rphelps@hp.com