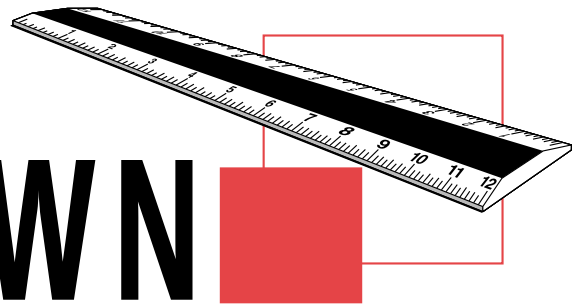
MEASURING THE UNKNOWN



By Garrick Hoadley, General Electric

How much is too much? When should you say when? In the middle of any competitive intelligence project you'll come to a point where you — or perhaps more ominously, someone with their hands on your purse strings — will try to understand exactly how much further it has to go and how much more it is going to cost. In a world of unconnected data points from which you are being asked to find a pattern, how can you truly know the remaining cost of a project when you aren't even sure where the end is? The short answer is you can't, but in this article I'll present a methodology by which you can get a fair assessment.

As a competitive intelligence (CI) analyst who has worked in commercial finance, consumer finance, and capital markets, I know how hard it can be to judge how far into your project you are. It's even harder to state how close you are to the end (if there is one!).

MEASURING THE UNMEASURABLE

What I have sought to create is a tool for measuring the unmeasurable. I feel this is fitting since, when asked what I do by others outside the CI profession, my response has traditionally been "I measure that which others say you cannot." Therefore, to be clear, for those of us who work in tea leaves, this is a tool to help quantify how far a competitive intelligence project has come and how much it will cost to complete.

Up front, let me inject a few caveats. (What would CI be without caveats?) First, the following is a work in progress. I believe it has great potential but interpretations of it, and additions to it, are sure to improve its value. Second, I do not expect the outcome of this analysis to be a decimal point oriented measurement. At best, it will give orders of magnitude, and at worst directional validation.

And third, I would like this analysis technique to be referred to as the Hoadley Suite, as in "Sam, the budget looks good but can you run a Hoadley Suite to support your estimates?" [Why do I request this? Well frankly, I would really like to have something named after me. And since there is no hope of me donating enough money to my alma mater to give even a janitor's closet the sobriquet "Hoadley," anyone who uses this analysis can refer to it freely as a Hoadley (rhymes with Toad - ly) Suite.]

The suite of tools has four distinct analytical stages: data sources, data capture, data mapping, and cost measurement. Each step independently creates value simply by thinking it through.

DATA SOURCES

The first analysis stage is a review of all possible sources of information. Where can you gather data about your topic? As CI professionals, we can all list the various sources of information: SEC filings, equity analyst reports, the press (local, trade, and national), competitors' websites, etc.

For advanced practitioners, there are also primary or human sources. These include contact with reporters in your space, industry conference presentations (particularly those given by officers of your competitors), and even direct conversations with debt and equity analysts. For those in technical fields, you would include patents, journal citations, and symposia.

Our objective here is to group and rate each of these sources on two metrics: completeness of data and timeliness.

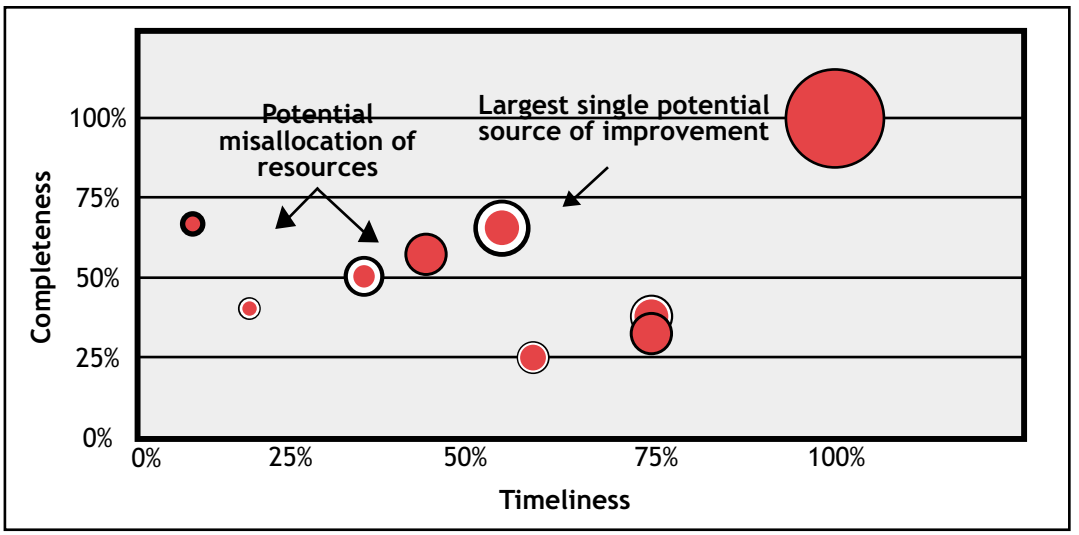


Figure 1: Data capture rate

Now that we have the two metrics for each source, we are going to multiply them together. This represents the data value for a source. As a quick take-away, the results of this calculation will allow us to rapidly sort those sources that we should spend most of our time with versus those sources that are merely supportive.

Does a particular source have a lot of good, accurate data about our target? How fast is this source as a channel for that data? For example, an industry publication is often very timely, but due to the quality of the journalism, the value of the data is often suspect or incomplete. On the other hand, a 10-K is very accurate but not very timely as it is, by default, a backward looking document.

CEO Head Test

The absolute measurement for the quality of these sources is the *CEO's head* test. The CEO of your CI project target — or whoever is the most relevant decision-maker — is the gold standard for an information source's timeliness and completeness. In essence, only she knows for sure what the company is going to do – and she knows it instantly. She scores a 100 for each of the two metrics.

When we think about our previous examples, the industry publication might rate 25 for completeness (be able to provide you with 25% of the complete picture, depending on how good it is and how important the company you are analyzing is to your industry). It may scores 50 for timeliness (contain information on 50% of what the company's CEO knows). The 10-K, however, scores 75 on completeness, but on timeliness is only a 25.

Data Value

These two figures are the first in a series of estimates you are going to make. Take your time and use your own complete data set as a self referencing tool. Check your estimates for one source against your estimates for each of the other sources. It is more important that the sources are all comparatively right with each other and, as a group, drift in one direction or another, rather than one source being absolutely correct on both axes but the others be completely off.

DATA CAPTURE

Now make an assessment as to how well you are doing with each of these sources. For example, when reviewing a new topic, I generally start with one of the online public press data sources. I run several key word searches and can expect to read through several hundred, if not thousand, articles. This is a painstaking process, but it ensures that I have a good grounding in my topic. It also allows me to understand not just the broad outlines of my quarry but, and perhaps more importantly, what I still need to find out. As Professor J. Edward Russo of the Johnson School would describe it, this is your metaknowledge: the knowledge of that which you do not know.

After a day of this I would like to say that I have captured 100% of the relevant information from this source. The reality is I have not. Yes, I have done a better article search than could possibly have been done 10 years ago, but it is not perfect.

First and foremost, my choice of key words may have excluded certain important variations but also, no online press source is complete nor timely. Due to competition in the industry, no one provider has a monopoly on the publishers and many publishers delay the uploading of their information into the services by up to a month. Even after a whole day's effort and follow up sessions instigated by the discovery of other key phrases, I would probably never have aggregated from this information source more than 75% of the total data available on my analytical target.

This amount will fluctuate. If the topic is incredibly rich (such as interest rate changes), I would expect a lower aggregation rate simply because it is unlikely that I would be able to capture all of the data and permutations needed in the key word searches. If the topic was obscure (such as AAA bond insurers), I would probably expect to capture almost the entire set of data.

TABLE 1:

Source Secondary	Completeness	Timeliness	Data Value	Current Capture	Capture Value	Missed Value
Source 1	35%	50%	18	25%	4	13
Source 2	15%	40%	6	40%	2	4
Source 3	45%	60%	27	90%	24	3
Source 4	75%	40%	30	70%	21	9
Source 5	75%	35%	26	90%	24	3
Source Primary						
Source 1	60%	25%	15	50%	8	8
Source 2	55%	65%	36	25%	9	27
Source 3	10%	65%	7	75%	5	2
			164		97	

DATA MAPPING

We can now overlay the percentage of capture for each source onto the Data Value for those sources. This gives us an estimate of how much of the potential data regarding our target we are actually capturing. The example in Table 1 reveals two important facts: our data capture rate is only at 60%, and we need to focus on one primary source first. Table 1's information can be graphed. (See Figure 1.)

Visually, we can now see the specific data source that is being underutilized. In addition, two other learnings from the diagram can be derived. First, the diagram should help us judge if we are devoting our efforts on the right place. We may have one source perfectly covered. But if its data value (or circle size) is small, and we know that capturing 100% of that source takes a lot of time, we may need to rethink where we are concentrating our efforts.

Second, we may need to decide between the source's timeliness and completeness. In certain instances, we may need to concentrate on those sources that can provide us with data, even if it is not complete, simply to meet internal or external demands. At other times, we may be willing to sacrifice time for completeness. This could be particularly true when results would inform the decision to invest in large capital outlays or other long-term decisions.

COST MEASUREMENT

We now look at the fourth stage, the cost of our project to date. At this point we need to be ruthlessly honest about the amount of time and resources we are using to complete the project. A simple list of inputs would look like this:

- your time
- colleagues' time
- support personnel
- outside resources: reference sources, online research tools, conferences
- consultants

Your goal is to honestly tally the cost of this project to date.

I am now going to make two thunderous assumptions.

1. The law of diminishing returns dictates that costs go up for each incremental gain in data capture.
2. A CI practitioner's analysis time runs concurrent with their data capture time.

Cost Curve

Assumption 1 allows us to presume an upward sloping cost curve as we move from 0% of data capture to 100%. When you first start your research, you may simply pick up your industry newsletter. This is simple low cost work.

Eventually you find yourself searching for a specific piece of data to complete or refute a specific argument that has arisen from your work. You may find yourself searching for hours through esoteric websites. This occurs towards the end of your information search process, and you are now expending far greater resources per fact.

The intrinsic value of the data will revolve around the mean as you move from 0% capture to 100%. Some items will be incredibly valuable and others will be useless. Your problem is that you cannot tell in which end of the value spectrum your next fact will reside. Therefore, this assumption subsumes the concept that the value of each data point is equivalent as we move up the capture scale.

Although I have used a linear relationship between data and cost, which I believe is accurate enough for simple analysis, I suspect this is actually an upward trending curve. Obviously any increase in the curvature will make the final data points that much more expensive the first.

Concurrent Analysis Time

Assumption 2 implies that the time spent by a CI practitioner to analyze, collate, and develop forward looking assessments is materially concurrent with the time spent on the data capture. This assumption is the harder to accept at face value but I believe it can be proven fairly simply.

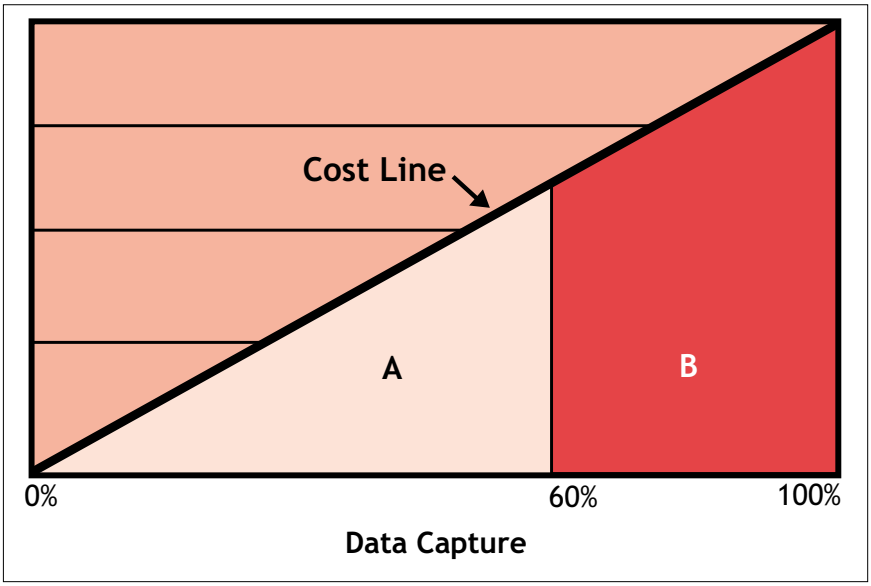


Figure 2: Projected data capture cost

The simplest logical result of Assumption 2 is to realize that the CI process is not broken into two pieces. We do not collect all the information we can for several weeks and then analyze the data. In a Bayesian process, our research affects our hypotheses, which directs us to more research and so forth in a continuing process. If you reflect on the time you have spent on CI, I am sure you will remember how a piece of data changed your position and drove you to look for more information. A gap in logic is discovered or an argument is proven weak, and you, the CI professional, then search for data that can support or refute the revised hypothesis.

What we have proven here — and this is important for the validity of our final step in this process — is that the analysis expense (effectively the time spent by the CI professional reviewing existing data) is always captured when we measure the hours spent at a certain time in the process and are inherent in the upward sloping cost line.

For our example, let us assume that we have spent $100,000 to date on our competitive intelligence project and have reached the 60% mark for data capture (X). Assuming a slope of 1 for the diminishing returns line, we get Area equal to $100,000. (See Figure 2.) The question remains: how much more expensive will it be to get to 100% data capture? The simple calculation is:

$A / (X^{\Delta}2) - A = B$

So in our case:

$100,000 / 0.62 - $100,000 = $177,778

For us to truly approach 100% capture we would need to spend an additional $178K. Interestingly, that may or may not be a hard pill to swallow. I have worked for companies that would baulk at the original $100K and I have worked at companies that would think nothing of allowing me an additional $178K to complete my work. What you, as a CI professional, must now do is assess the value and need for these outlays.

MOVING FORWARD

Those are the tools as they stand today. I believe this simple analysis of cost expectation can be applied in a variety of situations, such as large scale project assessment, and scientific research cost containment. However, I created this process while working as a CI professional and see it as a definite value in a field that traditionally has had a hard time quantifying its cost/benefit ratio.

I would welcome a dialogue with other CI practitioners regarding this analysis technique. I am a strong proponent of the need for accountability in our field. I created the basic concept above to clarify the cost of my research, but this is just one piece of the CI profit and loss statement. Our greatest challenge is to demonstrate that our value is greater than the costs we have delineated above and weigh that value against the continuing costs for our work.

In a future article I will outline my view on the organizational requirements a company must meet before it can begin to assess the value of CI. In the meantime, please feel free to forward comments, suggestions for improvement of the process, and, perhaps more fundamentally, your views on the actual ability to apply this analytic technique.

[Editor's note: Mr. Hoadley is the inventor of a competitive intelligence database and user interface structure developed to collect, collate, and disseminate information throughout an entire organization. A patent is pending.]

Garrick L. Hoadley has an undergraduate degree in economics and political science from Yale University and an MBA from the Johnson Graduate School at Cornell. He has worked as the head of finance and marketing for a high tech manufacturer and for BTS, a Swedish consulting company focused on strategic change implementation, where his primary clients were Texas Instruments and Nortel. Since 1999, Garrick has been with General Electric in various positions including competitive intelligence roles in insurance, capital markets, and consumer finance. He currently works in mergers and acquisitions for General Electric Consumer Finance – Americas. Garrick can be reached at garrick-hoadley@excite.com.