Every so often a technology movement comes along that promises to change everything. Many say big data—which can encompass anything from a few hundred million names in a census survey to billions of data points in a scientific research project—is one of those game changers. In its report "Demystifying Big Data: A Practical Guide to Transforming the Business of Government," the industry group TechAmerica Foundation declared: “The impact of big data has the potential to be as profound as the development of the Internet itself.”

That is big.

As part of the Obama administration’s 2012 Big Data Research and Development Initiative, six federal agencies are investing more than $200 million in technologies that can leverage massive amounts of information to identify trends and improve government services.

The U.S. Geological Survey for instance is analyzing data on Great Lakes deepwater fisheries and invertebrates dating back to 1927 to look for regime shifts in ecological systems. The Energy Department is providing $25 million to establish the Scalable Data Management, Analysis and Visualization Institute, where big data projects will accelerate the pace of scientific discovery in the department’s research facilities. The National Oceanic and Atmospheric Administration is devoting a portion of its annual $1 billion IT budget to improve weather and climate forecasting through big data.

Why do we need to corral big data? For one thing, the signal-to-noise ratio is overwhelming. IBM, for example, has observed that Twitter generates 12 terabytes of tweets daily—the equivalent of 5.3 billion single-spaced typewritten pages. And that’s just tweets. In a world that produces data in such volumes, some mechanism is needed to make sense of it all.

The big data push in government has spawned computing techniques to process vast amounts of information at speeds that make interpretation practical. They can capture meaningful intelligence, even from unstructured information, and apply the analytics that drive agency decision-making.

Still, advances in big data are in their infancy. A recent survey of security and IT professionals conducted by LogLogic and Echelon One showed 49 percent of respondents were somewhat or very concerned about managing big data, with 38 percent acknowledging they do not have a clear understanding of what big data is.

Here’s how to avoid five common mistakes in managing big data.
**START SMALL**

While the overarching goal should always be top of mind, organizations shouldn't aim to put in place the machinery of massive analytics in one fell swoop. A deployment that is too fast and too big could set agencies up for failure.

The move to big data “will be iterative and cyclical, versus revolutionary,” the TechAmerica report states. “Successful big data initiatives commonly start with a specific and narrowly defined business or mission requirement versus a plan to deploy a new and universal technical platform to support perceived future requirements.”

Such measured beginnings can then lead to further expansions. “After completing their initial deployments, government leaders typically expand to adjacent use cases, building out a more robust and unified set of core technical capabilities,” the report says.

Big data projects require an element of restraint. That doesn't mean putting on the brakes. Rather it means building capabilities incrementally. At first, analytics can be incorporated into just one element of the operation; it may drive a single project or fulfill a single function across a broad range of departments. Success comes in measured steps, rather than in one giant leap.

**THE BIG PICTURE**

In the push to deploy big data tools for a particular mission, agencies risk losing sight of the broader IT strategy. Creating technology silos or independent channels is a sure route to inefficiency and suboptimal use of resources.

The temptation to wall off fledgling big data initiatives is understandable. As new systems evolve, they may need special care and feeding and the learning curve can be prolonged. At the same time, agencies are reluctant to integrate a powerful new capability into existing systems, until they are sure how it will behave.

But these tools are undoubtedly going to be a big part of an agency's overall data infrastructure. It makes sense to think in terms of integration now, rather than cobbling together the pieces down the road.
THE BUSINESS CASE

Big data, whether it tracks citizen demographics or weather, is complex, resource intensive, and expensive to collect and analyze. IT and program managers must sit together to develop a business rationale for ramping up new technology and processes to cull big data.

Well before implementation begins, players must understand the agency’s goal and the tools that will be needed to achieve it. Big data can be put in play to improve processes, expand business intelligence, or identify patterns or trends that might otherwise go unheeded. Whatever the case, designers will have to determine the types of analytics needed, the form of reporting and other fundamentals to ensure the project yields not just data but actionable information.

Again, mere volume is not a measure of success. By collecting and analyzing data, project managers can generate detailed and revealing information. But the real result comes only when that data is used to generate specific and significant improvements to the agency’s mission.

QUALITY VS. QUANTITY

It is easy to place a premium on big data merely because it is just that—big. The scale of knowable information to which government has never before had access is staggering. Traffic information is compiled to promote environmental policies, fraud detection data informs health care benefit policies and cybersecurity data helps in building defenses for agency IT programs.

It is exactly this wide scope of possibilities that threatens the viability of big data. People often speak of these projects in terms of volume and velocity—how much information gets processed and how fast. But they sometimes overlook a third imperative, which must be accounted for in any big data effort: Variety. To get the full range of analytics, planners must reach across a spectrum of data points that might not traditionally fit together. This can include images, social media, maps, documents and form data, to name a few.

It takes broad analytical thinking to put together a full picture of the metrics that are consequential to business decisions. For these projects to be worthwhile, program designers must look beyond the sheer volume data to ensure they encompass relevant information all its forms.
THE HUMAN FACTOR

It’s a big mistake to assume that built-in analytics of big data are entirely objective and rational.

Despite breakthrough software and automated processes, human intelligence is required to interpret the outcomes of big data analysis. That intelligence is fallible and its shortcomings can be underestimated.

Data can show unexpected spikes or dips, but haphazard interpretation can give too much, or too little, weight to these variations from the norm. Often we see what we want to see. Just because the numbers were objectively generated, that doesn’t mean they will be read objectively.

As with any automated process, it’s only as efficient as the people who design the systems and interpret their output. In the design phase, managers should be careful not to include too many data sources, and be sure to weed out those that are irrelevant. They also should guard against organizing the outputs in a way that inadvertently bolsters preexisting ideas.

The promise of big data far outweighs the perils. Government has near-infinite volumes of data to work with, but the key will be an upfront commitment to making sure new analytics software and processes can deliver on that promise.