White Paper

The New Science of Data Geography, Datography and Data Cartography

By Robert Grant Beauchamp

Contents

Introduction

Most of us don't think about geography very often. Probably not since memorizing state capitals in middle school, or as a go-to-class for liberal arts majors to meet their science requirement. However, the science of geography can provide useful instruction on how to approach an organization's data, data environment and data-related capabilities.

Geography is a big topic. It's multi-disciplinary and wide-ranging with multiple sub-disciplines and areas of study. This makes it difficult to provide a concise definition of geography. These excerpts from National Geographic, however, work for our purposes: "'Geography is the study of places and the relationships between people and their environments' and, 'Geography seeks to understand where things are found, why they are there, and how they develop and change over time' as well as 'the why of the where.'"

Most definitions of geography emphasize concepts such as location, classification, purpose, relationships, inter-action, dependencies, navigation and historical context. But overall, ***geography provides a comprehensive picture of the world that is recognizable, useable and shareable.***

Which makes it a useful starting point in the development of a new data-oriented science. We can call this new science ***data geography***, ***datology***, or even ***datography***. It is still early days, and terminology, definitions and labels tend to take on a life of their own. As when data artisans morphed into data scientists. A practitioner might title themselves a data geographer, datoligist or datographer. Datography is an excellent term, first coined by Wayne Wrigley, a South African technologist and the world's first documented datographer at datography.net.

For our purposes, however, we are going to use the term ***data cartography***. We might think of data geography as the field, and datography as the discipline. Cartography, however, implies the ***act*** of recording, mapping, classifying, documenting and presenting. This makes data cartography easier to explain to business leaders. It is also the activity where businesses will see the most immediate value as ***data cartographers*** help them see their businesses in ways they never did before.

Purpose

The purpose of this document is to:

- Provide an introduction into the concept of data cartography.
- To provide 'the what' in order to set up 'the how' in future white papers.
- Provide a case to start utilizing data cartography at your organization.

A Note about GIS

The concept of data cartography should not be confused with (GIS) Geographical Information Systems (or Science). GIS is a category of data science that concerns systems and disciplines that utilize geographic location data for analysis. This might include, for example, the locations of a competitor's retail outlets in a city, the distribution of 911 calls or the population of wildlife in different environments.

The subject of this white paper is how an organization can better understand its own environments of data, databases, systems and networks.

Why Data Cartography is Important

It should be no surprise to anyone that data is the key to the future. Everything from IOT to machine learning to health management not only depends on data but also generates seemingly limitless amounts of data, which in turn can be leveraged for new purposes. That message is loud and clear.

Unfortunately, most people in most organizations don't have a comprehensive picture of their own data world that is recognizable, useable and shareable.

Obviously, every organization knows that it has data. For most people in most organizations, however, the location, composition, quality and potential of that data remains vague, miss-understood or unknown. They do not have the tools to navigate this world to avoid everyday data disasters and make the best use of their data to meet technological challenges.

It's obvious as well that most organizations already have reams of documentation that relate to an organization's data. This includes configuration management, architecture, asset tag lists, and data dictionaries to name just a few. These documents, however, don't reveal much about the business of the organization as a ***business***.

In addition, data-related skills, knowledge and tools are racing into the future. However, any potential they hold for these companies is blunted by the collective lack of knowledge of the organization's own data, data environments and data-related capabilities.

This clearly indicates that there is a very large hole in the ability of these organizations to deliver on their future.

Before they can ***use*** their data to realize their potential, they must ***know*** their data. That comprehensive picture of their ***data*** world that is recognizable, useable and shareable has become essential.

Why Data Cartography is Important, *NOW*

Several significant technology forces are currently in play, causing corporate leadership to understand that their data and how they manage it is strategic. ***It is necessary for their success if not for their very existence as a business***.

You've been reading about these trends and you may already be neck deep in some of them. But what is significant is that they are all happening at once, they often overlap, and they are all happening *now*.

These trends include:

- Digitization
- Data as Asset
- Data-Driven Decision Making
- Consumerization
- Internet of Things
- Cloud Computing
- Advanced Analytics and Business Intelligence
- Artificial Intelligence, Algorithms and Machine Learning
- Robotics
- Data Warehouse Modernization
- Increased Regulatory Scrutiny and Reporting Requirements
- Data Security and Privacy Protection
- Preparation for, and response to, legal actions related to data
- IT Consolidation resulting from Mergers and Acquisitions

The industry leaders in these areas have known this for a while, but for those organizations just entering the fray, they are quickly coming to the realization that ***if you can't use your data you can't do these things.***

And it can be asserted that ***you can't use your data in the way you want to, and will need to, if you lack an understanding of your organization's data geography.***

The Atlas

What is cartography without an Atlas? What is geography without a detailed description of the environment?

The documents described in this white paper focus on the most crucial component of a company's data-related capabilities: ***Identification of Data. The ability to find it, identify it, understand it, classify it, map it, and share this information with others.***

Every organization is unique, of course, and the exact mixture of documents will be unique as well.

The Data Atlas

The goal of 'The Maps' that comprise the Data Atlas is to provide immediate value with a minimal amount of time and resources. Below are just four examples of Data Atlas maps.

Often stakeholders have an 'a-ha' moment; a moment of recognition. They see their business reflected in the maps as they have never seen it represented before.

Data Cartographers know they have been successful when a stakeholder takes the maps and immediately starts showing them around at meetings and displaying them on walls.

Landscape Map
An illustrated view of the systems, sub-systems, ancillary systems and network communications of the enterprise and how they interact with each other. The purpose of this document is to ensure completeness. The stakeholders see that the systems that impact their accountability are accounted for. This also provides a powerful moment of recognition as they see their own business accountabilities represented graphically.

Data Source Map
This can be standalone or an overlay of the Landscape Map. The purpose of this document is to show the major databases that support the major systems along with their proper names and supporting platforms e.g. Oracle, DB2 etc. The client will usually be surprised at the number of data sources and the variety of platforms. Note that this

	document shows the sources of data for the major business functions. It is not meant to be comprehensive. That comes later. **Data Flow Map** This document can also be stand alone or an overlay. The purpose of this document is to show the electronic transactions that move data through the environment and their supporting technologies such as the network, EDI, ETL, connections, interfaces, manual uploads etc. Depending on the type of business this can become quite involved, involving internal transactions as well as transactions with external business partners and stakeholders. **Business Data Process Flow Map** The point of this map is not to conduct a full business process decomposition. Those types of projects can be lengthy and involve a great number of stakeholder time. Again, the idea is to provide immediate value to the stakeholders. The map just needs to show how the major business processes flow through the computing environment through the utilization of data. When all the component maps are assembled stakeholders will have a visual representation of their world that is comprehensive, understandable and shareable.

The Encyclopedia

The Data Encyclopedia includes detailed and comprehensive inventories, descriptions and classifications of the client's data and data environment.

The key messages regarding these documents are***: know what you own*** and ***have a common vocabulary to describe what you own.***

The extent an organization will wish to take this will be highly dependent on its needs, available resources and level of interest. The inventories and their accompanying descriptions may include:

- Servers hosting databases
- Reports
- ETL, ELT and EDI transactions
- Connections and interfaces
- Jobs involved in overnight processing
- Databases hosted in the cloud as opposed to on-site.
- Etc.

Possibly the most useful aspect of the Data Encyclopedia is the data classifications. The key message here is that early in the process it is less important to strive for quality data than it is to ***know the quality of your existing data*****.**

Each component of the data environment is rated according to:

- Availability
- Accessibility
- Reliability
- Usability
- Security

This simple and flexible system can be used to separately rate the quality of a database, a table within that database or an element within the table. It can be used to rate files, automated reports and ETL transactions.

A key function of the classifications is to guide stakeholders toward more useful data and away from less useful data. It also gives the organization an exceptional baseline from which to prioritize data quality initiatives.

Decision Making and Regulatory Tools

		Understanding your organization's data geography and employing data cartography makes a myriad of everyday things possible that many organizations have difficulty with. These include: • Data asset valuation • Data capabilities assessments • Project planning, risk management and level-of-effort estimating • Transaction development • Data mapping • Data migrations Examples of some of these are described below. • Documentation of the organization's 'underground data economy.' • Traceability reports • Data governance
The Underground Data Economy.		The 'Underground Data Economy' documents are highly valuable, and many stakeholders will find them intriguing and irresistible to review. The underground data economy is comprised of the massive number of Excel Spreadsheets, MS Access Databases and database queries that are regularly used in the operational functioning of the business. These spreadsheets and ad hoc databases and queries are used without any accountability or oversight. Data locked in spreadsheets and ad hoc databases is extremely hard to account for and monitor. These documents can include inventories of spreadsheets and databases. Not only will the vast number surprise the stakeholders, but it will also help expose the rampant reuse and recycling of data and the resulting persistence of errors.
Data Traceability Reports		A Data Traceability Report involves taking a document or report and working backwards to trace the origin and transformation of each piece of data in the report.

		The Data Traceability Report can authoritatively describe what each field means, how it was calculated, describe the path the data took through the enterprise computing environment and the transformations it underwent. The Data Traceability Report is useful not only for audits, and legal actions, but also for data quality checks. Data traceability is only made possible with the extensive knowledge gained and effective tools available that comes with data cartography capability.
Data Governance		Data cartography plays a crucial part in effective data management, including: • Data Governance • Master Data Management • Meta Data Management • Data Quality Assurance • Data Lifecycle Management • Data Security It is a perfectly legitimate question to ask how an organization can manage something it doesn't completely understand.

Overcoming Internal Questions, Issues and Concerns

When data cartography is brought into an organization questions and concerns will inevitably be raised as they are with any corporate initiative. Fortunately, there are excellent responses to these questions that may help stakeholders overcome their concerns. These questions and concerns often include:

Our previous data quality efforts have gotten bogged down in the weeds. What will be different this time?
Contentious and detailed oriented people tend to create lot of detailed and thorough information. However, Data Cartography is not about neatness. ***Data Cartography is about knowing. It is about speed. It is about movement.*** The ability to quickly know as much as possible now, even if the news is bad, is more important than being presented with perfectly coifed, but limited, data sets at some unspecified time in the future.

We just did a data classification assessment for our security initiative. How is this different?
That is an excellent start. And that information will be useful. However, most efforts of these types classify risk relative to a potential breach. It does not necessarily help understand the meaning of the data or enlighten us to how it can be used for purposes such as analytics.

We have so much data, I can't imagine getting our heads around all of it.
One key to understanding this effort is that even though you may have petabytes of data, you have a finite number of data sets, databases, tables, rows, columns, fields and data elements. 'Eating the elephant' becomes a matter of tackling these one at a time until you are through.

We have so many projects underway and we can't afford to put any of them on hold.
One of the many advantages of developing Atlases and Encyclopedias is that you don't have to put anything on hold. Their creation is designed to 'run in the background' delivering value as documents are created.

My staff is already overwhelmed. How will they have time for this?
There will, as always, be a certain amount of time required of your subject matter experts. But as we already know they are being pulled in many directions and their time is very valuable,

there are strategies we can use to lessen their time commitment. However, in our experience once the SMEs start to see the value of the data sourcing deliverables for their own work lives, they become eager to participate.

This seems like a lot to bite off. Plus, we really don't have a budget for this right now and we are months away from our yearly budgeting exercise.

The good news about implementing a data cartography capability is that, Atlases and Encyclopedias can be created at a pace with which your organization is comfortable. Projects can:

- Start small with a minimal team to keep initial phases under budget signing thresholds.
- Be broken out into easily digestible stages.
- Go as wide and deep as you are comfortable.
- Show value quickly, building enthusiasm and a demand for more, making future budget request easier.

We can't wait months to see results from another 'methodology.'

A significant aspect of Data Cartography is its modularity and scalability. Features include:

- Documents are such that they can be completed at a pace, team size and price the organization is comfortable with.
- Each document stands on its own and demonstrates immediate value upon completion.
- Documents are easily integrated into other initiatives in which the client may be engaged.
- Starting does not require any special tools. You can go a long with just by using Word, Excel, Access and Visio.

We don't have time for a lengthy awareness, education and adoption process.

While some education is required, the good news is that the things that are being documented, such as databases, networks, ETL, data warehouses, analytics, etc. the organizations is already very familiar with. And, of course, the stakeholders are already familiar with the difficulties they experience when trying to leverage their data.

All things considered, with data cartography, it's never been easier to get started and do the right thing.

The Case for Starting NOW!

The Case for Data Cartography	**Bring your data and your data environment out from the shadows.** How can you use what you don't know you have? How can you turn your data into a corporate asset usable by everyone, if it is hidden behind layers of technology, jargon and organizational turf wars? **Answer the tough questions** When it comes to audits, data breaches and litigation, demonstrating a mastery of your data and data environment goes a long way to protecting your reputation when you are managing a crisis. **Resilience** No matter how technology changes, indeed, how the world changes, it will require you to be able to know your data more than ever before. **Don't stumble on your way to the future** Whatever forward-thinking strategic initiatives you will engage in, whether digitization, advanced analytics or artificial intelligence, you don't want to be held back because you don't know your data or your data environment. Get ahead of the game and reap the benefits of your strategic initiatives sooner. Remember, your data is literally the digits in digitization. ***Data Cartography builds knowledge and mastery over, what in the future, will probably be the single most important asset your company owns, your data.***

Conclusion

More to come	The purpose of this white paper is to provide an ***introduction to the concepts*** of data geography, datology and data cartography. In addition, the paper describes why it is important to get started now. This is the beginning of a long adventure. There are more white papers to come, along with videos, webinars, conferences and certifications that will describe the 'how' and not just the 'what' and the 'why.' There will be many contributors, as well, bringing their own experiences, ideas, methods and techniques to the table. In addition, as more automated tools become available, they will need to be mastered and integrated into the work. So, in the end, this white paper is not just an ***introduction***, but an ***invitation*** to join in on the adventure.

About Robert Grant Beauchamp

Robert Grant Beauchamp is a consultant, architect, and former CIO with a proven record of helping organizations understand and improve their data quality, data environments and data-based capabilities. As a systems integrator, Robert has successfully introduced and implemented data-related technologies such as BI, EDI, ETL, data warehousing and three-tier architectures.

If you would like to learn more about data quality or would like help improving data quality within your organization, connect with him on LinkedIn.

Robert has filled the roles of computer journalist, tech writer, business analyst, marketing communications manager, business architect, project manager, program integrator, program manager, account manager, data security consultant, solutions architect and trusted advisor, including:

- Thirty years of business and information technology experience including over five years as Chief Information Officer of a rapidly growing health plan.
- Proven track record of successfully strategizing, developing and implementing enterprise-level business and technology initiatives in the health care and financial services industries.
- An experienced and practiced consultant with the ability to work with C-level executives to develop strategy, assess capabilities, manage risk, and offer solutions that can be successfully implemented in an organization's unique environment.
- A proven communicator well versed in public speaking, meeting facilitation, webinars, journalism and video.
- A proven history of developing and implementing successful service offerings for a major IT consulting firms including Y2K, HIPAA Security and Privacy and HIPAA Electronic Transactions.

Currently Mr. Beauchamp is championing a capability-based approach to data quality. He is a leader in the adoption, education, and implementation of Data Sourcing as a corporate capability. He is in the process of writing a book on Data Sourcing.