Data Readiness Index: Concept (EN)

Version 1.1 2023-03-08

By Chris Langdon, chris.langdon@cgu.edu

Table of Contents

Introduction: *Are you still storing, or already sharing?* **1**

The problem **1**

The dataspace solution: From Big Data to Better Data **2**

The data iceberg execution danger: Biology **2**

Results: Are you ready? (1) Business impact and (2) actionable **2**

How to do it: Avoid reinventing the wheel, leveraging best practice **3**

Next steps: 2-phase schedule – design & delivery 1.0 ..... **3**

Appendix A: What - Information systems literature foundation **4**

Appendix B: How - Applied behavioral analytics for "world's best car" **5**

References **6**

Introduction: *Are you still storing, or already sharing?*

Think about it: What if new data communication technology were to arrive that would make all the world's data available to anybody? We agree: It would be a game changer. Now, arrival of such technology is no longer hypothetical. It is called dataspace technology. Importantly, it will power a first commercial service, the Catena-X open and collaborative data ecosystem (link). Even more importantly, it will abstract the technology away to make benefits more easily accessible to any business user. And it will come with the first novel applications. Therefore, are you ready for it? And where are your critical gaps?

The problem

With "software eating the world" (Andreessen 2011) data has shifted to the core of business but suffers from a massive productivity gap, turning the 80/20 Pareto principle, a cornerstone of business efficiency (e.g., Neuman 2005), upside down: More than 80% of the time budget of a data analytics project is spent on data wrangling – not with algorithms and results (see Figure 1). Almost always the problem is two-fold: (1) Not the right data, and (2) not enough of it, despite ever bigger data lakes and soaring, painful storage cost.

© 2023 Langdon. All rights reserved.

The dataspace solution: From Big Data to Better Data

Well, problem solved, say "hello" to dataspaces, the evolution from Big Data to Better Data: Dataspaces are part of a Web3 upgrade to enable decentralized, peer-to-peer datasharing with data sovereignty protection (What is a dataspace? Link). No data storing but sharing, and on-demand and just-in-time without losing the power to control one's rights to the data. If everything were connected, any data could be available.

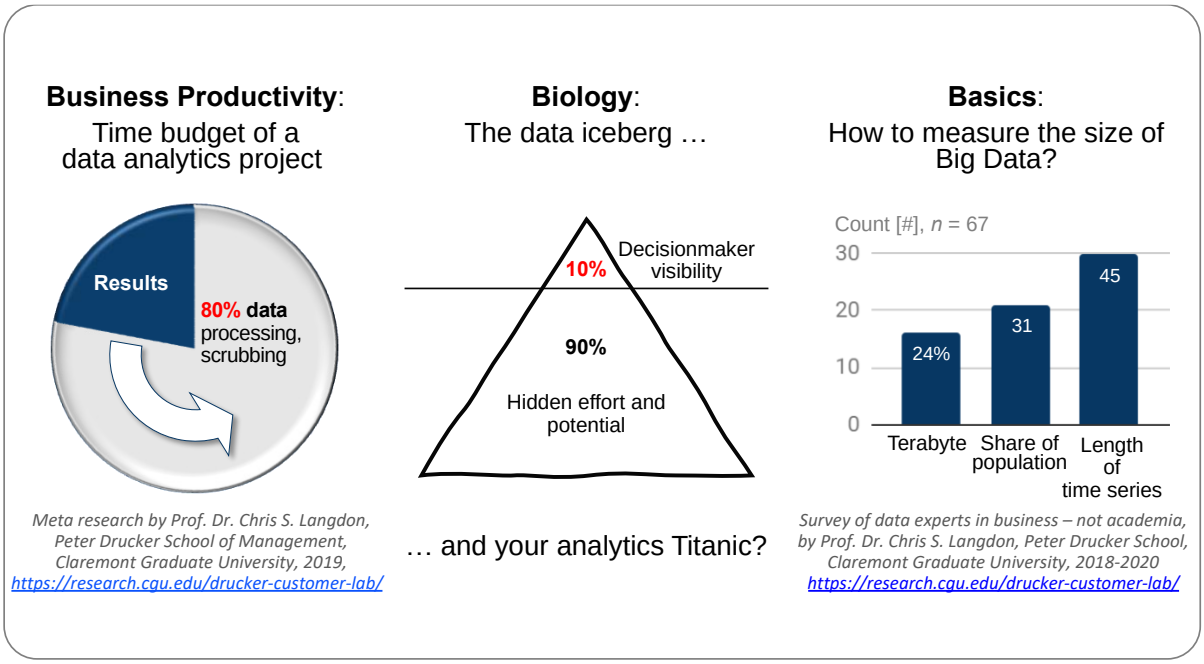

Figure 1: Data problems

The data iceberg execution danger: Biology

One issue remains, the data iceberg problem: With data, we can't see much, we may see 10%, the tip of an iceberg, not the 90% lurking beneath the surface (see Figure 1). It's simply biological. We humans are highly visual creatures and can see danger from afar, but data is like an iceberg, our brain struggles to "see" much in reams and reams of numbers (Kaas 2013, 1992). This makes for funny jokes but is terrible in business. For example, we dream of data monetarization riches but don't even know how to measure data quantities properly. We know to buy eggs by the dozen, a pound of butter, and a liter of milk. But how do we buy data? By byte or length of a time series or share of population? Figure 1 illustrates this data sizing dilemma. Therefore, we need aids, visualizations, actionable guides, like, well, a Data Readiness index or DRi.

Results: Are you ready? (1) Business impact and (2) actionable

Readiness is the state of being ready (Oxford Dictionary). But what is readiness for data? And how to operationalize, measure and act on it?

- **Business impact**: Following the advice of management legends Peter Drucker and Steven Covey we start from results (Covey 1989, Drucker 1963). Therefore, our

© 2023 Langdon. All rights reserved.

definition captures data readiness from a business impact or outcome perspective … not from a technology standpoint (see Schlueter Langdon 2006, for example).

- **Actionable**: Furthermore, our index will be designed for actionable results. Here we borrow from a proven recipe, the protocol of a doctor visit: As we don't swallow medication without prior diagnosis, let's not jump to conclusions and carefully capture your data readiness level or symptoms first, as the starting point for a proper diagnosis in step two.
  - **Symptoms** (Zustandsindikator): Establish data readiness through data analysis: **Where do we stand as a participant?**
  - **Diagnosis** (Prognoseinstrument): First analysis of variance of index values. What is causing low scores? **Why are we in the middle scoring range of results?**

How to do it: Avoid reinventing the wheel, leveraging best practice

The index will leverage (a) science and relevant data studies (Roehl & Scheufen 2022, Demary et al. 2019), as well as (b) best practice and apparatus of similar implementations.

- Science foundation: Our approach will be grounded in pertinent **scientific, peer-reviewed literature** with vetted publications (not just white papers written by vendors). Because of a focus on business impact as opposed to technology, we will borrow primarily from the information systems, human-machine-interface, and cognitive sciences literature (see Appendix A).
- Best practice implementation: Borrow from proven success stories, such as related **index measurement efforts** (e.g., Ifo Geschäftsklimaindex) and **behavior analytics in automotive**, for example, as developed by Chris Langdon and operated by Pacific Coast Research Inc. for global automakers (see Appendix B).

Next steps: 2-phase schedule – design & delivery 1.0

Based on previous data analytics projects (Appendix B, for example) a 2-phase implementation schedule is proposed: (1) development, and (2) delivery 1.0. Both phases will borrow from proven methods. Figure 2 provides an overview.

- **Development** … will borrow from design thinking and rapid prototyping utilized in teaching (e.g., "prototyping propeller" in MGT 317: Creating Smart Products, link) and industry projects. Specifically, iterative development from existing studies to measurement concepts and instruments. One key challenge is establishing the truth with minimal subjective bias, minimizing data capture with surveys of focal firms (see Kahneman & Krueger 2006).
- Delivery 1.0 … will benefit from best practice utilized in teaching (e.g., 5-phases from MGT 378 Social Media Analytics, link) and industry projects (see Appendix B, for example). Specifically, we envision the step of data strategy (variables, data types, tabulation), systems (instruments, integration, calibration), data collection (scheduling, setup, logging), and results (baseline, analysis of variance, extrapolation).

© 2023 Langdon. All rights reserved.

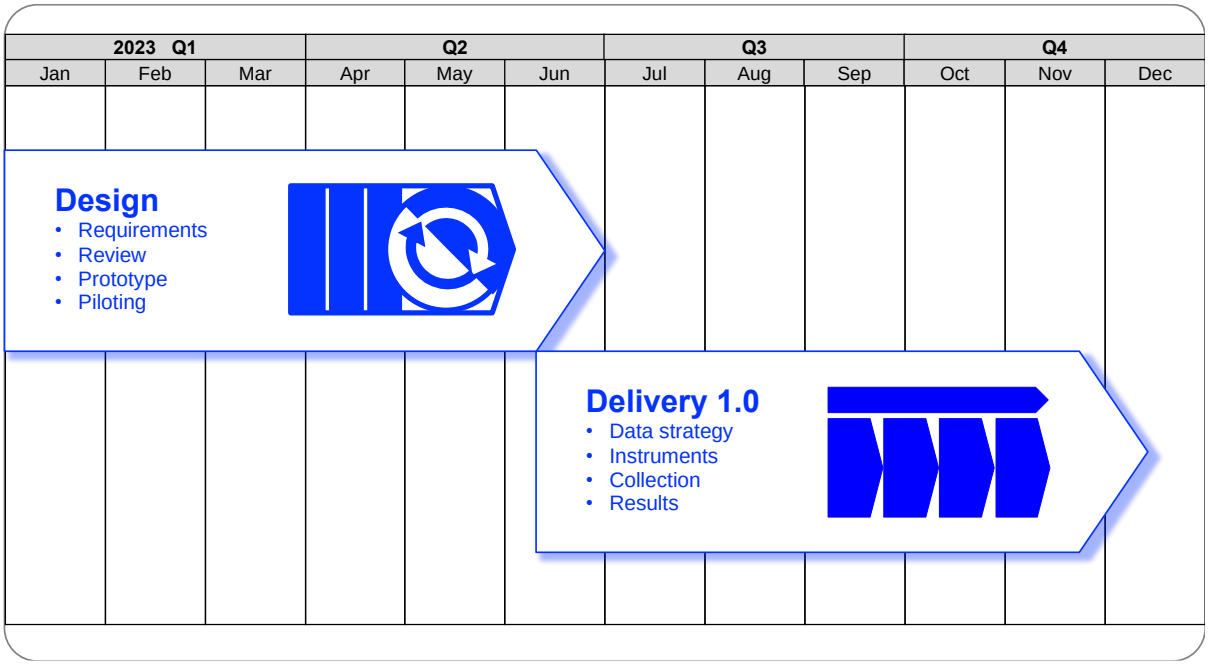

Figure 2: Schedule - Piggybacking on research and industry best practice

Appendix A: What - Information systems literature foundation

Following a long tradition of information systems research grounded in economics, such as transaction cost theory, a clear picture has emerged of the chain of activities required for data-driven business value creation (also platforms, see Otto & Jarke 2019). Figure 3 depicts a conceptual framework derived from elaborate systematic literate reviews (SLRs) and in-depth case studies of digital pioneers with large-scale data processing systems including Facebook and Netflix (adapted from Schlueter Langdon & Sikora 2020, Figures 1 and 2).

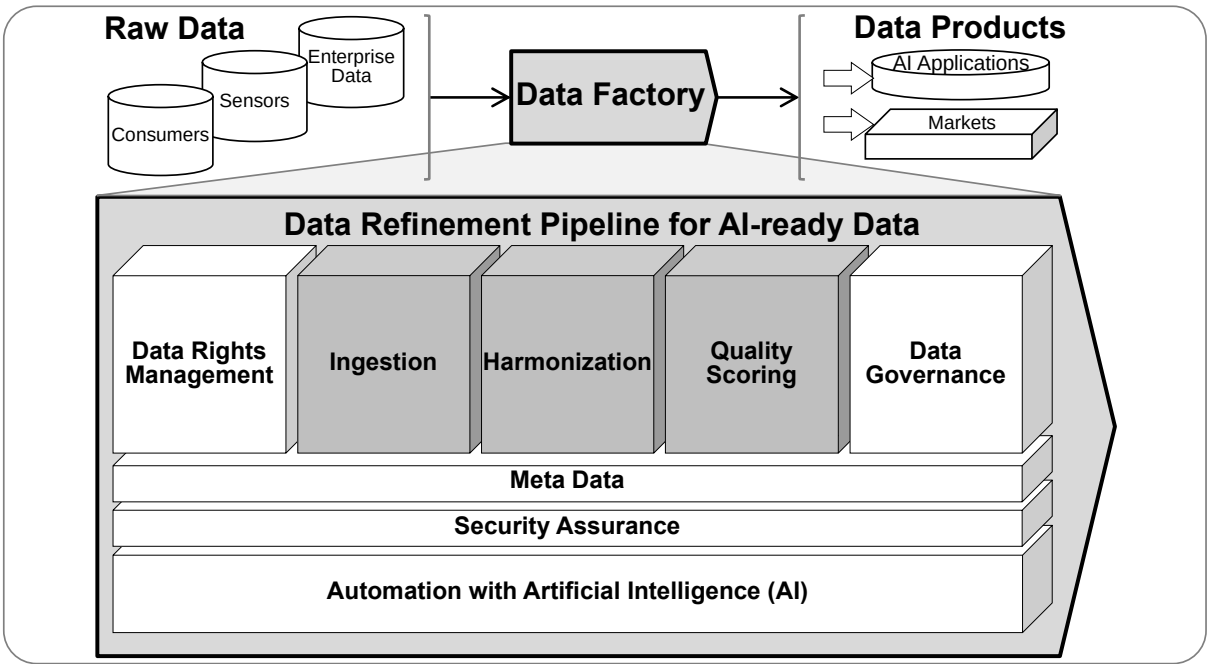

Figure 3: Data factory for data products (adapted from Schlueter Langdon & Sikora 2020)

© 2023 Langdon. All rights reserved.

It illustrates the proposition of (a) treating **data as a product** that requires processing to be ready for use in applications (Crosby & Langdon 2019), and (b) the **industrialization** of this process.

- In terms of **data product**, think about water. At an executive level "data" tends to be thought of as a product that can be grabbed off the shelf … well, like bottled water, for example. However, for a data scientist it is more like a puddle of water, raw water. Nobody would drink from a puddle. The water would have to be "treated," filtered, analyzed for harmful ingredients, labeled, and bottled to an exact quantity. This is time-consuming and not cheap, just look at the difference in price: the price of one liter of branded, bottled water at retail, buys one thousand liters of raw water in Berlin, for example (BWB, link).
- The gist of our analogy is this: Data needs refinement and in order to economize on the process it takes – no surprise – **industrialization or a "factory"**. We use factories for everything else. Gottlieb Daimler invented the motor car in 1886 but Henry Ford invented the modern factory and with it automotive as an industry. His factory turned auto-making from a hand-made affair to mass production through automation (Womak et el. 1990).

In Figure 3 raw data rights must be verified before any data can be ingested or harvested (rights, licensing, user consent). Then data ought to be properly labeled or tagged for it to be made discoverable through a catalog of categories and search engines (harmonization, classification, etc.). Furthermore, it needs to be scored to provide some indication of quality, because without it any subsequent analytics is pointless – "garbage in, garbage out" (GIGO, quality scoring; see Otto 2015, Otto et al. 2012). Finally, governance mechanisms are required to ensure that data can be exchanged while data sovereignty is maintained for each data provider (Otto 2011).

Appendix B: How - Applied behavioral analytics for "world's best car"

At the turn of this millennium product development in automotive was still reliant on demographics and psychographics (who people are), and gathered in surveys (what people say). Then science (Nobel Prize winning work by Kahneman, for example, link) and digital technology evolved to enable a **shift to behavioral data** (what people do) captured with in-car data- and video-loggers (observing actual user-product interaction). We have been one of the pioneers helping with this shift toward behavior variables. With one automaker, our program was called, FUMP - feature usage measurement project. It delivered decision support for better targeted cabin investments to create "the world's best car" (the 2012 Mercedes-Benz S-Class or W222, Top Gear 01/14) and boost JD Power customer satisfaction scores leading to record-breaking sales.

- Grounded in **scientific research**: FUMP was developed by Chris Langdon, who is a Research Associate Professor at the Peter Drucker School, Claremont Graduate University in Southern California, home of Peter Drucker, founder of modern management. Chris also runs **Drucker Customer Lab** (link), co-founded with his colleague Bernie Jaworski who is recipient of the three prominent Journal of Marketing awards – the Alpha Kappa Psi award which he received twice, the Maynard award, and the Sheth award. He also received the Converse award and Vijay Mahajan lifetime achievement award for contributions to marketing.

© 2023 Langdon. All rights reserved.

- Delivered on-time, on-budget with **operational excellence**: All projects have been executed and delivered by Chris Langdon's Pacific Coast Research Inc. Projects were delivered end-to-end from ideation, measurement and sensor development, systems integration and calibration, data collection and refinement, data analytics and results, to visualization in interactive dashboards (**"Calculator" Powered by ML: Auto Interior & UX**, link). Data collection was standardized and conducted (a) as triangulation of micro-controller (ECU) data, in-car video, and customer surveys, (b) across vehicle lines (small, medium, large), generations, and model years; and (c) globally in Germany, the USA (East and West Coast), and China (Beijing, Shanghai, and Guangzhou).

References

Andreessen, M. 2011. Why Software Is Eating the World. The Wall Street Journal (2011-08-20), link

Clark, K. B., and T. Fujimoto. 1991. Product Development Performance: Strategy, Organization, and Management in the World Auto Industry. Harvard Business School Press: Boston, MA

Covey, S. R. 1989. The 7 Habits of Highly Effective People – Powerful Lessons in Personal Change. Simon & Schuster, New York, NY

Crosby, L., and C. Schlueter Langdon. 2019. Data as a Product to be Managed. Marketing News, American Marketing Association (October 10th), link

Demary et al. 2019. Readiness Data Economy - Bereitschaft der deutschen Unternehmen für die Teilhabe an der Datenwirtschaft. Eine Veröffentlichung im Rahmen des BMWi-Verbundprojektes: Demand – Data Economics and Management of Data Driven Business. Institut der deutschen Wirtschaft Köln e.V., Koeln

Drucker, P. F. 1963. Managing for Business Effectiveness. Harvard Business Review (May): 53–60

Kaas, J. H. 2013. The evolution of brains from early mammals to humans. Wiley Interdisciplinary Reviews: Cognitive Science Wiley 4(1): 33–45, link

Kaas, J. H. 1992. Do humans see what monkeys see? Trends in Neurosciences 15(1): 1–3, link

Kahneman, D., and A. B. Krueger, 2006, Developments in the Measurement of Subjective Well-Being, Journal of Economic Perspectives 20(1): 3-24

Newman, M.E. 2005. Power laws, Pareto Distributions, and Zipf's law. Contemporary Physics 46(5): 323–351

Otto, B., and M. Jarke. 2019. Designing a multi-sided data platform: Findings from the International Data Spaces case. Electronic Markets 29(4): 561-580

Otto, B. 2015. Quality and Value of the Data Resource in Large Enterprises. Inf. Syst. Manag. 32(3): 234-251

Otto, B. 2011. Organizing data governance: Findings from the telecommunications industry and consequences for large service providers. Communications of the AIS 29(1): 45-66

© 2023 Langdon. All rights reserved.

Otto, B., K. M. Hüner, H. Österle. 2012. Toward a functional reference model for master data quality management. Inf. Syst. E Bus. Manag. 10(3): 395-425

Roehl, K-H., and M. Scheufen. 2022. Hemmnisse beim Data Sharing: Empirie und Handlungsempfehlungen. Arbeitspapier, Institut der deutschen Wirtschaft Köln e.V., Koeln

Schlueter Langdon, C. 2016. What your customers are doing in their vehicles, why you want to know, and how to profit from it – 'Auto Google' and 3 steps to success. 15th Automobile Interior Conference (October 18-19), SV Veranstaltungen, Audi Forum Ingolstadt, Germany

Schlueter Langdon, C. 2006. Designing Information Systems Capabilities to Create Business Value: A Theoretical Conceptualization of the Role of Flexibility and Integration. Journal of Database Management 17(3) (July-September): 1-18

Schlueter Langdon, C., and R. Sikora. 2020. Creating a Data Factory for Data Products. In: Lang, K. R., J. J. Xu et al. (eds). Smart Business: Technology and Data Enabled Innovative Business Models and Practices. Springer Nature, Switzerland: 43-55, link

Womack, J, D. Jones, and D. Roos. 1990. The Machine That Changed the World: The Story of Lean Production. Free Press, Simon & Schuster: New York, NY

© 2023 Langdon. All rights reserved.