Do we have a Data Culture?

Wolfgang Kremser
Salzburg Research Forschungsgesellschaft m.b.H.
Salzburg, Austria
wolfgang.kremser@salzburgresearch.at

Richard Brunauer
Salzburg Research Forschungsgesellschaft m.b.H.
Salzburg, Austria

Abstract—Nowadays, adopting a "data culture" or operating "data-driven" are desired goals for a number of managers. However, what does it mean when an organization claims to have data culture? A clear definition is not available. This paper aims to sharpen the understanding of data culture in organizations by discussing recent usages of the term. It shows that data culture is a kind of organizational culture. A special form of data culture is a data-driven culture. We conclude that a data-driven culture is defined by following a specific set of values, behaviors and norms that enable effective data analytics. Besides these values, behaviors and norms, this paper presents the job roles necessary for a data-driven culture. We include the crucial role of the data steward that elevates a data culture to a data-driven culture by administering data governance. Finally, we propose a definition of data-driven culture that focuses on the commitment to data-based decision making and an ever-improving data analytics process. This paper helps teams and organizations of any size that strive towards advancing their – not necessarily big – data analytics capabilities by drawing their attention to the often neglected, non-technical requirements: data governance and a suitable organizational culture.

Keywords—data culture; big data; data-driven culture; data governance; big data analytics

I. INTRODUCTION

The term "data culture" repeatedly appears in (online) publications that are thematically close to business analytics and big data. While scholars repeatedly emphasize its significance for a successful, long-term data-centric venture [1][2], they do not provide a clear definition. We believe data culture to be insufficiently defined relative to its already recognized importance. What distinguishes organizations that follow a data culture is vague and "there are very few attempts made to really define what data culture means" [3, para. 2].

This paper aims to provide a definition of data culture by identifying its set of shared values as well as the corresponding job roles and their responsibilities towards a healthy data culture. It concludes by offering a definition centered on data quality and the associated job role of the data steward. This definition makes it clearer in which ways a healthy data culture benefits data analytics processes. Managers will understand the commitment and talent required to establish a fruitful data culture. The job role descriptions will help data analytics professionals to focus their efforts and communicate their needs.

A. How "Data Culture" is used

We start our exploration of the meaning of "data culture" by looking at examples of how it is used in web and literature. Three examples have been selected because they emphasize different aspects and will serve as a guide towards a comprehensive definition:

For Microsoft, a good data culture is one in which "every team and every individual is empowered to do great things because of the data at their fingertips" [4, para. 6]. This description puts strong emphasis on the democratization of data. Empowering every individual within an organization to extract value from data requires a robust set of tools and methods.

A whitepaper by the IT service provider Cognizant describes data culture as a "workplace environment that employs a consistent, repeatable approach to tactical and strategic decision-making through emphatic and empirical data proof" [5, p. 2]. Again we notice an emphasis on robust tools and methods to make data analytics consistent and repeatable. Furthermore, it links data culture to a form of decision making based on data.

Marr says that "in a data culture, data is recognized as a key business asset, and it is used, wherever possible, at every level of the business to make improvements" [6, p. 172]. This description corroborates the democratization and decision making aspects of data culture. It also introduces the aspect of constant improvement associated with data culture.

All three attempts to characterize data culture have in common that data is considered as an asset of an organization and that organizations either "have" or "do not have" a data culture. Nevertheless, they do not outline what a "culture" exactly is: Is a (data) culture a set of common values, a common agreement, a set of recommendations, a way of structuring an organization or a set of business processes?

In their recent systematic literature review about big data capabilities, Mikalef et al. [7] used the term "data-driven culture" instead of "data culture". They classified data-driven culture as an intangible resource within the managerial framework of resource-based theory. A resource is defined as "stocks of available factors that the organization owns or controls" [8, p. 35]. Other examples for intangible resources are organizational culture and organizational learning [9]. Mikalef et al. [7] extracted three aspects that contribute towards data-driven culture: (1) prioritization of business analytics investments, (2) support from top management, and a (3) fact-based and learning culture.

B. Data Culture vs. Data-Driven Culture

Before defining data culture, we first want to place the term ontologically. Starting off the considerations of Mikalef et al., data culture is an intangible resource of organizations [7]. This

puts it on the same level as organizational culture [9]. Indeed, scholars treat data culture as a special form of organizational culture [10]. While there exists no generally accepted definition of organizational culture [11], we use the following working definition: An organizational culture is the collection of common values, behaviors and norms which are postulated by top management and are followed by most individuals within an organization. Data culture therefore refers to the values, behaviors and norms shared by most individuals within an organization with regards to data-related issues.

Under this lens, data culture becomes a very broad term. All modern organizations collect some form of data (e.g. billing addresses, transaction data, supplier data, etc.). Hence all modern organizations have a data culture. The term is too general when trying to talk about specific data-related capabilities of an organization. What is more often meant when talking about data culture is data culture of a certain quality. We call a data culture of this quality data-driven. This differentiation fits the usage of the term "data-driven culture" in [8]. The following sections will isolate the distinct values, behaviors and norms that make a data culture a data-driven culture.

II.   TOWARDS DEFINING DATA-DRIVEN CULTURE

As established in the last section, there are certain qualities that distinguish a data-driven organizational culture. To understand these qualities we examine data-driven culture's relation to other organizational resources.

A. Organizational Values

As already stated, a data-driven culture is a form of organizational culture. As such it represents in our understanding a set of values etc. We stipulate three values which have to be shared by most individuals following a data-driven culture. These values can be summarized by an organization-wide preference for data-based decisions:

Value 1 – data-driven decision making is valuable: Decision making on the basis of insights extracted from data is actively encouraged throughout the whole organization. This behavior is favored compared to the "classic" decision-making process in which the hierarchically highest person decides alone and based on experience, costs and intuition [10][12].

Value 2 – seamless data access and data provisioning between business units is valuable: The previous value goes hand in hand with the democratization of data throughout the whole organizational structure. Sharing data between business units enhances isolated data and provides decision-makers on all levels with a basis for their decisions [13].

Value *3 – discussing and distributing knowledge about data processing is valuable*: As a consequence to the distribution of decision-power, a data-driven culture embraces organizational learning which is the process of creating, retaining, and transferring knowledge within an organization [14]. The tools available for manipulating data are numerous, with new solutions for sub-processes being released constantly [15]. This requires professionals to regularly update their knowledge and methods. They consistently share their skills and practices with others to improve data-related processes.

B. The Relation to (Big) Data

One resource related to data-driven culture (DDC) is data itself. As a tangible resource, data cannot create any business value on their own. They lie dormant until they are actively leveraged [9]. Barring directly selling data, the method to generate business value, i.e. insight, from them is called *data analytics* [16].

Data culture has emerged as a topic of discussion in the context of big data and big data analytics. At least since the early 2010's, the term big data has found its way into literature. It is often characterized by the three V's – volume, variety and velocity [17]. De Mauro et al. [18] include big data's requirements on methodology: "Big Data is the Information asset characterized by such a High Volume, Velocity and Variety to require specific Technology and Analytical Methods for its transformation into Value" [18, p. 131]. These definitions all include the size of the datasets that are processed.

Others define big data without referencing the size of the datasets. The MIKE 2.0 project characterizes big data by the degree of complexity that datasets possess and the amount of value that is gained by applying innovative techniques in analyzing them [19]. Gantz and Reinsel [20] state that big data focuses on three main characteristics: the data itself, the analytics of the data, and presentation of the analytics' results that allow the creation of business value in terms of new products or services.

Whichever definition one chooses, there is no reason why the values listed earlier could only be present in organizations that process big data (BD). Data-driven decisions can be made on the basis of simple and small, yet representative datasets. These datasets can just as well be made accessible, and organizational learning is not bound to any specific kind of dataset. This suggests that a data-driven culture can be established regardless of the size or complexity of the datasets involved. Hence, processing big data cannot be a necessary requirement for data-drive culture (formally: DDC $\not\Rightarrow$ BD).

There is also no reason to believe that all organizations that perform (big) data analytics follow a data-driven culture. It is possible to generate insights from data without ever using these insights as a decision basis, or while keeping data isolated within business units. Simply having data analytics capabilities (DAC) is not enough to claim data-driven culture (DAC $\not\Rightarrow$ DDC).

Instead, it seems that data-driven culture is more closely related to *data quality* (DQ) because "the most significant benefit of quality data is the willingness of managers to trust the data and to act on the information" [21, p. 5]. Therefore we include data quality as a necessary (and reasonable) prerequisite for data-driven culture (DDC $\Rightarrow$ DQ). Data is of high quality if they meet the expectations and needs of the job role that is working with them [22]. The Strong-Wang framework [23] describes 15 dimensions of data quality which later have been summarized into six core dimensions [22]: completeness, uniqueness, timeliness, validity, accuracy, and consistency. Furthermore, in an ongoing effort to standardize data quality, the ISO 8000 [21] defines its own five key characteristics of quality data. One of them, *provenance*, is knowing the source of the data to establish trust.