A case for Data Democratization

Pranjal Awasthi
Mays Business School, Texas A&M University, pranjal3103@tamu.edu

Jordana J. George
Mays Business School at Texas A&M University, jgeorge@mays.tamu.edu

Follow this and additional works at: https://aisel.aisnet.org/amcis2020

https://aisel.aisnet.org/amcis2020/data_science_analytics_for_decision_support/data_science_analytics_for_decision_support/23

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2020 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
A Case for Data Democratization

Completed Research

Pranjal Awasthi
Texas A&M University
pranj3103@tamu.edu

Dr. Jordana George
Texas A&M University
jgeorge@mays.tamu.edu

Abstract

The 21st century is quickly becoming the century of data. Growth in data volumes, data types, analysis tools, data mining, computing power, and machine learning algorithms have given organizations unprecedented fodder for data-driven decisions. The democratization of data, which we define as intra-organizational open data, provides an opportunity to transform employees from data users into citizen data scientists who provide valuable insights. This conceptual exploratory paper investigates the concept of data democratization, including processes, definitions, implications, and examples.

Keywords

Data democratization, Data democracy, Open data, Data-driven decision making, Big data, Analytics, Data science.

Introduction

The internet has played a crucial role in ushering in a data-driven age, where firms, governments, non-profits are drenched in data (Desjardins 2019; Srujana et al. 2016). This data has the potential to significantly improve organizational decision making and adaptation to changing environments (The Economist 2017). Opening data to the general population has a robust history with open data stalwarts like Tim Berners Lee voicing their support (Berners-Lee 2010), and open government initiatives increasing on every continent (Data.gov 2020; Data.gov.in 2020; Data.gov.uk 2020; Dataportal.opendataforafrica.org 2020). However, while open government data (OGD) is gaining acceptance, we have not seen the same enthusiasm within organizations except for a notable few (Court 2015; Moore 2017). The concept of intra-organizational open data is the democratization of data.

The 21st century is an opportune time for Data Democratization. Growing volumes of data, cost reductions for information systems, and data skill growth in the pool of potential users all contribute to a favorable environment for liberating data within an organization (Sawicki & Craig 1996). In this paper we unpack the concept of data democratization and explore how it provides competitive advantage. As data is central to Information Systems (IS), the democratization of data is a relevant topic for the field. This explorative study expands on research in open data, strategic use of IS, and data management. Our primary research questions include:

• What is Data Democratization?

• How can Data Democratization be leveraged for competitive advantage?

• What are the chief concerns in the present Data Democratization paradigm?

Data Democratization – A concept

Data Democratization is the act of opening organizational data to as many employees as possible, given reasonable limitations on legal confidentiality and security (Cornelissen 2018; Lee 2013; Marr 2017;
To operationalize data democratization, data is made available not only to analysts and executives, but also to non-technical or non-specialist employees to observe, analyze, use in data-driven decision making, or make other use of it. Data Democratization allows data to transition from the hands of a select few employees into the hands of the masses within a firm. Limiting data limits data usefulness. Limiting data to a handful of people creates frustrations such as difficulty in communicating data-driven insights to long delays for receiving reports from overworked analysts. Yet, simply placing data sets in an open shared drive will not promote data democratization, either, when data seekers have few technical data skills (Cornelissen, 2018). Data Democratization encompasses a gamut of steps a firm can take to ensure effective data use and allow employees to make data-driven decisions (Court 2015). Data must be cleaned, categorized, put into context, and placed in an easy to access and use repository. Users must have a minimal level of skills to import, export, document, and analyze data. And last, organizations need to embrace data-driven decision making into their culture (Ruijer et al. 2017; Safarov et al. 2017).

Data Democratization derives its core principles from related movements like Data Philanthropy and Open Data, in that it aims to increase data access and reduce gatekeepers for data stores (Brodsky & Oakes 2017; George et al. 2019; Schlagwein et al. 2017). It also includes shared responsibility for the data. Figure 1 illustrates these relationships and Table 1 summarizes the definitions of data democratization, data philanthropy, and open data.

![Figure 1. Data Democratization Pillars](image)

**Table 1. Variations of Open Data**

<table>
<thead>
<tr>
<th>Data Action</th>
<th>Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Data Philanthropy</td>
<td>Sharing of private data assets to serve public good</td>
</tr>
<tr>
<td>2. Data Democratization</td>
<td>Data made available to the non-technical or non-specialist employee for them to make use of it</td>
</tr>
<tr>
<td>3. Open Data</td>
<td>Proprietary data made available to the public via open source cloud frameworks, open APIs</td>
</tr>
</tbody>
</table>
Effecting better use of talent resource

The traditional literature around Data Democratization has mostly been about breaking in-firm silos and transforming employees into citizen data scientists by providing them unfettered data access and the appropriate training. From an organizational point of view, it makes sense to not relegate the ever-increasing data science operations to a small group of specialists. This concept of job enrichment through vertical loading, self-efficacy enhancement and increasing intrinsic motivation (Conger and Kanungo 1988) has long held currency amongst management scholars (Menon 2001). The concept of employee empowerment has shown to be contingent on resource allocation and the ability to control and leverage a proprietary resource of a firm (Astley and Sachdeva 1984). In a data-centric workforce, data has emerged as a resource of prime importance, the judicious use of which can be a source of sustained competitive advantage. (George et al. 2019). In such a workforce paradigm, Data democratization helps increase empowerment at all levels of organization and instill a sense of shared responsibility and onus. (Matisoff 2018). Bernard Marr, in his book “Big Data in Practice,” highlights the same principle when talking about Data Democratization. The goal of data democratization, Marr notes, is to get employees easier and faster access to data to get the business insights they need without external help (Marr 2017).

One benefit of spreading out data work is relieving data scientists of low-level tasks. Data Scientists swamped with basic data analysis tasks are not able to focus on tasks which require deeper technical expertise. Further, if all the diverse data requests from disparate departments and business stakeholders are scheduled to be cleared and answered by just one team of specialists, it increases the turnaround time for complaint addressal and makes scaling un-sustainable in the longer run (Cornelissen 2018). Another corollary of this is the lack of domain expertise data scientists might have. They might not be able to see trends and patterns in the domain data. In such a scenario, it becomes imperative for data scientists to empower domain experts with data tools and allow them to put data to use as best as the domain demands. This distribution of information and tools to make sense out of that information can increase the entrepreneurial spirit and add agility to the enterprises work structure. Honing employees’ data skills and endowing them with data analytics tools can help a firm reap better insights from their data, make inter-colleague communication clearer and help make individual employee a better decision maker in his/her department (Schatsky et al. 2018; Tapadinhas and Idoine 2016).

A study by IBM found that executives worldwide spend about 70% of their time looking for data and only 30% of their time analyzing it (Srujana et al., 2016). This is understandable for two reasons. First, data often comes in vast proportions in a variety of forms from a range of sources. Second, there is a shortage of data scientists in the market (Schatsky et al. 2018). McKinsey & Company predicts that in the United States alone, there could be a shortage of 140,000 to 190,000 people with requisite analytical skills as well as 1.5 million managers with the required expertise to make data-driven and meaningful decisions (Manyika et al., 2011). In such a scenario, it makes sense for an organization to start developing in-house data skills and empower its human capital to become minimally data proficient (if not a complete data scientist) in their day to day tasks.

From this field of thought emerges another proximal concept of citizen data scientist. Citizen data scientists are people who are not professional analysts but learn data skills up to moderate level of expertise (Idoine 2018; Tapadinhas and Idoine 2016).

A citizen data scientist can bridge the gap between mainstream self-service analytics by business users and advanced statistical techniques of a data scientist (Idoine 2018; Moore 2017; Tapadinhas and Idoine 2016). The main characteristic of a citizen data scientist is the fact that his/her primary domain of work lies beyond analytics and statistics. A citizen data scientist is generally a domain expert who also happens to be proficient user and manipulator of data thereby complementing the traditional data scientist (Idoine 2018; Moore 2017). Today, sophisticated Business Intelligence tools and analytics frameworks are allowing employees to gain access to data analytics capabilities and are allowing firms a great pedestal engender a data democratization paradigm whole heartedly. Figure 2 illustrates the key traits of a citizen data scientist as described by Gartner (Idoine 2018).
A Case for Data Democratisation

The approach of Data Democratization and turning employees into citizen data scientists has helped firms like AirBnB (Pratt 2017), Viacom (Berman 2017) and Royal Bank of Scotland (Matisoff 2018) secure better insights from their data and empower their employees to make data-driven decisions.

Related Work

Being a relatively new phenomenon, existing research on data democratization is most commonly associated with the extant data culture in firms (Buehler and Henke 2018), theories proposing a more holistic CRM approach (Su et al., 2010), practices about Democratizing Big Data (Cornelissen 2018), and case studies of firms and non-profits who have taken data democratization initiatives (George 2019). Further, some studies conflate the concepts of Open Data and Data Democratization to propose a hybrid data democratization which entails firms giving data access and sharing data skills with the public at large and not limit it within company boundaries (Inta 2019).

In light of the amounts of data being collected, processed and stored, many company cultures are transforming into data cultures (Clark 2018). McKinsey research has shown a growing dichotomy between businesses who have been leaders in adopting better to this data revolution with better data analytics proliferation and laggards who are yet to start making effective use of their data sources (Buehler and Henke 2018). Rob Casper, the chief data officer of JP Morgan Chase has talked about how data is a great determinant for solving business problems (Casper 2019). An effective Data culture goes beyond mere volume and talks about getting better solutions from a firm’s data stores. Red flags are raised when the executive teams lacks an understanding of predictive analytics and when analytics capabilities are isolated from business, effecting an inefficient corporate structure (Court 2015). Such reports and research, support democratization of data in favor of seamless and cross-departmental flow of data.

Data Democratization finds support in another field of research i.e. Building effective Customer Relationship Management (CRM) strategies. CRM can be best defined as the set of business process and policy to acquire, retain and service the customer (Scott 2001). Fingertip access to data about the customers to a firm’s employees (not just data scientists and data analysts) plays a crucial role in enhancing customer relation at each point of contact (Crosby and Johnson 2001). Data democratization gains support from the rise of customer strategies like 1-to-1 marketing, mass customizations, service-based competition etc. (Crosby and Johnson 2001). In this day and age CRM strategies have been taking a wider view towards customers and the organization itself and expanding its ambit to integrate sales, marketing and customer care services to create a unique value proposition (Chalmeta 2006). This extrapolates to customer data begin accessed by various departments and the employees working holistically to provide the customer a better experience. This customer centric approach is entirely compatible and finds its place in the core tenets of data democratization.

Examples of firms which have benefitted from Data democratization also form an important part of the sources we have chosen to refer to for this study. Firms like AirBnB with their Data University and Viacom with their analytics platform are trying to increase data accessibility and ensure that solutions which can be delivered without the expertise of a data scientist are delivered so by regular employees. Ted Colbert, Chief
Information Officer at Boeing, talks about how there should be firm-wide platforms where people can easily access data. Colbert remarks that when people begin to visualize data, it is a game changer. Data changes their perceptions and makes them inherently more data-driven (Buehler and Henke 2018).

**Methodology**

For this study we look at the extant data culture in corporate management, how firms are using data in unique ways to enhance their competitive advantage, and where data democratization fits in. To understand data democratization and its importance, we have also examined academic theories around resources and how data qualifies as a resource capable of boosting strategic positions, thereby giving firms ample rationale to adopt the democratization of data. We also refer to classic management studies around employee empowerment, and maximizing human resource potential to give credence to arguments that promote data democratization.

In addition to management studies, we consider the examples of pioneering firms which have taken substantial steps to engage with their employees and work towards democratizing data, ensuring more data access, greater data skills and added data responsibility to allow them to take better data-driven decisions. The choice of firms can be considered a sample of convenience. Given the nascency of Data Democratization as a field, there is a dearth of public data or company statements available at the time of conducting this research.

**Theoretical Foundations**

We derive the importance of data democratization by understanding the importance of data as a resource. 21st century technologies have led to an exponential increase in data generations. Any proposition or theory concerning data use first affects a study of data as its importance. This paper views uses the Resource Based View (RBV) and Resource Dependent Theory (RDT) as our primary theoretical foundations. These help us understand the nature of data and its role as a catalyst of competitive advantage.

**Resource Based View (RBV)**

A resource is defined as an asset or capability which is helpful in gauging market threats and identifying opportunities (Zuiderwijk et.al, 2015). The RBV encompasses creating competitive advantage using unique resources and capabilities (Wade and Hulland 2004). These attributes have been encapsulated in the VRIN framework (Barney 1991). Data is a strategic resource of the firm because, as illustrated in Figure 3, a firm’s proprietary data possesses VRIN (Valuable, Rare Inimitable, Non-Substitutable) capabilities (George et al. 2019; Picciano 2014).

![Figure 3. Data as a VRIN resource (based on George et al. 2019)](image-url)
Resource Dependent View (RDV)

Related to RBV, the Resource Dependent View (RDV) looks at the internal firm resources to single out resources which can provide sustained strategic advantage to a firm (Hillman et.al. 2009; Pfeffer, et.al, 2003). Unlike RBV, RDV does not assume that resources just “exist.” Firms lacking VRIN resources usually find other ways of procuring such resources from their external environment (Zuiderwijk et.al. 2015). RDV manifests itself in multiple ways such as Merger and Acquisitions, Joint Ventures and other interorganizational relations which help reduce uncertainty (Hillman, et.al. 2009).

Data Democratization Examples: Ameliorating the Organizational Digital Divide

AirBnB’s Data University

AirBnB’s implicit belief in empowering every employee to make data informed decisions and embracing Data Democratization as a whole has led the firm to create a platform which they call Data University (Feng et al. 2017; Mannes 2017). Data University imbibes the core tenet of making data, data tools and data skills available to every employee of a firm. AirBnB has a fundamental belief that every employee ought to be enabled to be able to data informed decision. Data University combines Data Access, Data Education and Data Tools to make advanced data analytics capabilities accessible to anyone (Feng et al., 2017). AirBnB’s Data University helped unfurl the real benefits of data democratization in that it made the employees more data conscious (Airbnb 2020). With AirBnB’s automated end to end machine learning tools and data pipelining frameworks, non-specialists employees were able to realise the potential of the vast data chunks and use AirBnB’s data stores to solve their individual problems. Further Data Scientists saw a 50% decrease in ad hoc requests when the teams they support went through Data University and became self sufficient in answering questions using dashboards and basic SQL queries (Feng et al. 2017). The solutions which were now being provided were more inline with the firm’s overall goals and much more global.

Viacom’s Science Central

Viacom’s executives had been finding it frustrating to ask data teams for information to help make decisions about specific shows. This, in the words of Chris McGrath Senior VP of Data Strategy, was a “never ending” process (Tribbey 2016). To expedite the process of decision making and data access, Viacom came with their Audience Science initiative which eventually manifested itself into the Science Central. This was particularly important because data scientists, although understand the nuances of data in and out, but might not have the credentials to make a show which people like and watch. In the presence of such obvious dichotomies in terms of expertise, it made more sense to reduce the gatekeepers and democratize data in order to get it in the hands of the experts in a way so as to get the best of both set of skills (Viacom 2020). Science Central was created in order to give employees from around 20+ Viacom’s cable networks to access data like television audience insights, show stats etc. (Berman 2017). This data helps the employees like show creators, directors, producers understand the audience and content better. Including Data Democratization in its Data Strategy has helped Viacom’s network operators plan better marketing campaigns, gauge audience response better to show pilots and above all understand .

Royal Bank of Scotland

Royal Bank of Scotland (RBS) is a prominent adopter of Data Democratization (Groenfeldt 2018). RBS used data to integrate customer experience across all the touchpoints. RBS remapped their organization structure to introduce digital marketing leaders from across departments. This helped experts interact with each other and use the same firm data to help optimize overall customer experience from different perspectives (Matisoff 2018). With this data-driven and integrated approach, RBS was able to understand customer behavior much better. They, for example, understood that 30% of all their customers were using mobile phones to apply for loans, albeit with low conversion rates. This caused RBS to concentrate their efforts on rolling out a better, more intuitive and streamlines mobile experience. This almost immediately increased their customer conversion rate substantially by over 20% (Wareham 2015).
Walmart’s Data Café

Walmart’s multiple departments, each working towards delivering a solution for one of the firm’s many problems, often lacked data which might provide a better solution. Walmart’s Data Café looks to solve this very problem of paucity amidst plenty (Marr 2017). This 2.5 petabytes of data in Walmart’s private cloud, would help their individual departments to combine their respective expertise with other (even seemingly tangential) data streams outside their department to actually create a unique proposition to their problems at hand (Reilly 2017). We suggest, with corroborated data, the individual department’s contributions are more in line with the firm’s overall vision and goals. In addition to this, a non-circuitous access to data reduces time and/or addressing problems which demand immediate attention. This theory begins to reap benefits in real time too. In one example, Walmart’s grocery team was at a loss as to why sales had suddenly plummeted in a particular product type. Running numbers and investigating the data from the Café, made them realize that along the supply chain, pricing miscalculations had been made at a particular stage which led to the products being listed at a higher price at the store that what had been intended (Marr 2017).

Table 2. Examples of Data Democratization

<table>
<thead>
<tr>
<th>Firm</th>
<th>Example</th>
<th>Benefits Accrued</th>
</tr>
</thead>
<tbody>
<tr>
<td>AirBnB</td>
<td>Data University</td>
<td>50% decrease in ad hoc data requests.</td>
</tr>
<tr>
<td>Viacom</td>
<td>Audience Science</td>
<td>More effective marketing campaigns</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Audience response better recorded</td>
</tr>
<tr>
<td>Royal Bank of Scotland</td>
<td>Advanced Analytics Platform</td>
<td>More effective customer management with 30% increase in loan applications</td>
</tr>
<tr>
<td>Walmart</td>
<td>Data Cafe</td>
<td>Better sales and quicker analysis time</td>
</tr>
</tbody>
</table>

Concerns with Data Democratization – Privacy and Access Controls

Data Democratization as a practice comes with a set of quality, cybersecurity and data privacy concerns. When the data floodgates open, quality is paramount as new data workers may have little experience with judging data quality. Basing decisions on poor data will result in poor decisions (Janssen et al. 2017). In terms of access, each firm has its own level of access control and data dissemination. There are still questions about how much access should be given to employees, how and when strategic advantage is compromised if proprietary data gets leaked, and concerns about data ethics (Marr 2017). One of the primary concerns with data democratization comes in the form of potential misuse of data (Barrett 2018). Without appropriate training, employees poorly versed in data analysis can easily read the data incorrectly and make erroneous assumptions, especially if the data is taken out of context (Moreau 2010). The problem is further exacerbated in industries like healthcare wherein incorrect decisions can have disastrous consequences. Another concern which is cited by critiques is that of redundancy (Brooke 2018). With different teams working on same data in a de-centralized manner, there is a potential for work redundancy and similar insights being worked out by different teams, essentially wasting effective man hours.

Future Implications and Conclusions

Data Democratization is a new and promising paradigm for growing an organizational data culture. In this study we examine the conceptual origins of data democratization along with examples of firms who have adopted Data Democratization and receive benefits through decentralizing data stores and creating citizen data scientists. The theoretical foundations of the resource based and resource dependent views, existing open data paradigms and open data forms, and current organizational shared data activities offer implications for both researchers and industry. This can have long lasting effects on the multiple dimensions of corporate governance ranging from changing human resource roles to firms investing in post-employment education and training. The concept of citizen data scientist has been referred to by
Gartner as one of the biggest disruptors in the ways we see data and data analysis. With more than 40% of conventional data science jobs getting automated, broader usage of data analytics is likely (Gartner. 2017). Further, since data is a resource, data democratization also has implications on how firms use this strategic resource as a source of competitive advantage. With promises of enhanced customer management, better insights from data streams and increased individual responsibility of employees effecting better talent use, data democratization is an opportunity for digital leaders to become more agile and responsive to customers. Its repercussions spread out to organizational structure, HR management, technological ability and strategic advantage.

This study is a first look into a promising new organizational movement. Given how recent the concept of intra-organizational open data is, we were limited by the availability of data on the topic. This limitation guides future work on the democratization of data, where primary empirical data from subject organizations may provide deeper insights. The implications for organizations is clear. Firms may benefit strategically from an enterprise level shift towards data democratization. When implementing data democracy, the shift may be eased when the foundational cornerstones of data literacy, data access and a high level framework for abstractions have been laid. This approach allows firms substantial leeway and individuality to deal with the concerns highlighted above in a manner most resonant with the firm’s data culture. With the three pillars of data democratization secured, small individual wins and micro decisions based on data will accrue larger merits to the firm on the whole.

REFERENCES