ICEB 2018 Proceedings

International Conference on Electronic Business (ICEB)

Winter 12-6-2018

# Big data values deliverance: OSS model

John R. Hamilton

Sing What Tee

Hamilton, J.R. & Tee, S. (2018). Big data values deliverance: OSS mode. In *Proceedings of The 18th International Conference on Electronic Business* (pp. 154-162). ICEB, Guilin, China, December 2-6.



# Big data values deliverance: OSS model

*(Full Paper)*

John R. Hamilton*, James Cook University, Australia, John.Hamilton@jcu.edu.au
SingWhat Tee, James Cook University, Australia, SingWhat.Tee@jcu.edu.au

**ABSTRACT**

Open source software (OSS) repositories, like GitHub, conjointly build numerous big data projects. GitHub developers and/or its responders extend/enhance a project's software capabilities. Over time, GitHub's repositories are mined for new knowledge and capabilities. This study's values-deliverance staging system data mines, isolates, collates and incorporates relevant GitHub text into values deliverance model constructs. This suggests differential construct effects influence a project's activities levels. The study suggests OSS big data platforms can be software data mined to isolate and assess the values embedded. This also elucidates pathways where behavioral values deliverance improvements to GitHub can likely be most beneficial.

*Keywords*: GitHub, value; open source; big data; values deliverance; text data mine; MapReduce

*Corresponding author

## INTRODUCTION AND MOTIVATION

Big data generation exceeds 2.5 Exabytes per day (IBM 2015). Big data accumulates from information generation sources including: sensors, climate, health-fitness activities, transport, processing, logistics, financial actions, social media, knowledge accumulation and research. In the last 2 years 90 per cent of all data has been accumulated.

Big data views by Gartner (2012) and Schroeck *et al.* (2012) offer the definition: big data is 'uncertainties-embedded (veracity), high-volume, high-velocity, high-variety information assets that collectively demand cost-effective, innovative forms of uniform processing for enhanced insight and decision making.

Today the uniform processing of big data analysis still remains a misnomer. Big data is often mined for specific new-knowledge purposes such as: productivity and profitability (McAfee and Brynjolfsson 2012) competitiveness (Davenport 2006), advertising (Trattner and Kappe 2013), health (Bodanar and Salathe 2013; Chen *et al.* 2016), retailing (Bradlow *et al.* 2016), or investigations (Gerber 2014).

Big data mining techniques are also diverse and purpose-driven - with libraries like Mahout and SparkMlib utilizing targeted, machine-learning, efficiency-related applications. The combination of big data and traditional machine learning is also generating challenges within the social big data spectrum - with data-patterning trialed across fields including behaviors, visualizing and tracking (Bello-Orgaz *et al.* 2016).

### GitHub

GitHub is a big data storage site that accepts or rejects global developer contributions into its embedded private or public (free) projects repositories (Badashian et al 2015). It houses over 20M users and 57M repositories. It is the largest open source software (OSS) of communities that combine their talents to develop software (Cosentino *et al.* 2017). It promotes rapid software development projects. GitHub incorporate collaborative development features including: bug-tracking, feature-requests, task-management and wikis (Williams, 2012; Marlow *et al.* 2013). This on-line platform hosts, and permits the building of collaborative OSS development projects.

GitHub mixes traditional software capabilities (like version control, free hosting and project storage) with an extended social reach (Squire 2014). Here, crowd-sourced coding contributors also participate indirectly through microblogs, and other websites (like Hacker.com). These crowd-sourced coding contributors comment on projects, analyze the merits of a project or possibly refine a specific project component coding modification and so add value towards an ongoing GitHub software development project (Taay *et al.* 2014).

GitHub's big data is stored in projects within its data storage repository. GitHub allows the use of restrictive API's to investigate information within these repositories. Project developers can use these API's to follow the relative activities contributions around their project. To support such project developer knowledge acquisitions this study investigates big data analytics and the deliverance of values.
Hence, this study investigates the first question:
RQ1: Can the OSS big data platform GitHub be data mined to assess the values embedded within its participant's actions.

## ANALYSING BIG DATA: GITHUB

### Big Data

GitHub's big data varies across time depending on project: (1) type, (2) scale, (3) duration, (4) team strength, (5) language, and (6) complexity (Cosentino *et al.* 2017). It is a multi-repository site where motivated contributors or responders can: (1) join a chosen harmonious community, (2) engage with others, (3) test the fit of their intentions, and (4) possibly be further motivated when other responder additions also provide unique personal value adds.

### Analyzing Data Streams

GitHub offer API's (https://developer.github.com/v3/) that allow its contributors to mine and retrieve the contents of files within a big data repository as Base64 encoded content. This allows its contributors to mine, investigate, and then develop effective, enhanced systems. These, in-turn, drive: (1) advanced project operational efficiencies, (2) project responsiveness, (3) expansion of project communities' interactive connectivities, (4) teams of global social-communities (developers, contributors/responders, assessors, and marketers), and (5) lower project completion times.

Those researching GitHub big data continually add new understanding into this global open source software (OSS) community. Approaches vary from: (1) mathematical data mining techniques (Cosentino *et al.* 2017), (2) investigating readme files (Weber and Lou 2014), (3) tracking lines-of-code, forks, pulls, watchers, stars, reviews and relating these project additions/subtractions (Aggarwal *et al.* 2014; Borges *et al.* 2015; 2016; Xavier *et al.* 2014), (4) modelling the factors effects delivering project popularity and/or project activities levels (Aggarwal *et al.* 2014; Alshomali *et al.* 2017; Cosentino *et al.* 2017), (5) specific software/language studies, (6) social network studies (Jiang *et al.* 2017; Ma *et al.* 2016; McDonald *et al.* 2014; Sheoran *et al.* 2014; Singer 2014), (7) global software and economic trends (Grewal et al 2017), and (8) behavioral studies (Phipps et al 2013).

GitHub is also a social network, and it services its OSS communities (Lima *et al.* 2014). Here, its project hosting programs promote their collaborative software, and the activities/interests of one developer are promoted through to other developers (Thung *et al.* 2013). Thus, GitHub is a social interactive platform that interconnects all its stakeholders. Hence, although many different approaches are used to mine its big data, GitHub likely acts as a services business, and so likely provides a source of values deliverance similar to what is found in service business.'

### Analyzing for Value

The Oxford Dictionary (https://en.oxforddictionaries.com) considers value as a verb as something holding economic value, being deemed important/beneficial; or generating a high opinion. This is a values acquisition process. In contrast value as a noun is something (in one's judgment) that is held of importance, of worth, or of usefulness. This is a post-event measure.
The Collins dictionary (https://www.collinsdictionary.com) sees value as a verb as 'if you value something (or someone) you think it is holding personal importance, and you appreciate this.' This is a values acquiring process. Value as a noun is seen as 'the value of something such as a quality, an attitude, or a method is its importance or usefulness' to you. This is an established or after-the-event measure.

Merriam-Webster (https://www.merriam-webster.com) defines value as a measurement outcome (noun): 'a monetary worth of something; a fair return in goods, services, or money; relative worth, utility, or importance; something intrinsically valuable/desirable, sought material values; a numerical quantity assigned or determined; a musical note; relative color or picture part lightness or darkness.' Merriam-Webster gives scant consideration to value as verb, but does indicate that value is a complex mix that is determined over time.

On closer examination such definitions suggest value as something currently being built (verb) around an item, and as something already gauged as present in an item (noun), and as a multi-dimensional consumer perception of an item.
Like the above definitions, Boksberger and Melsen (2011) recognize value (value as a noun) as a preferential, after-event judgment involving a single transaction or end-state. They also recognize that the term values (value as verb) represents a series of criteria, norms or ideals that are framing towards a deduced or derived value position.

Value is a perception framed through Equity Theory, where consumer's outcomes-versus-inputs are gauged against the service provider's outputs-versus-inputs (Oliver and DeSarbo 1988). The consumer determines a value by evaluating what is fair, correct, and/or deserved against some perceived net-worth (Bolton and Lemon 1999). This net-worth is the values being acquired by the consumer - possibly as: (1) time being consumed; and/or (2) energies/actions being applied; and/or (3) rewards being acquired; and/or, (4) perceived sacrifices being incurred, and/or (5) feelings being incurred; and/or (6) affordability of costings incurred (Yang and Petersen 2004).

Parise *et al.* (2012) enlist a four quadrant model (performance management, data exploration, social analytics, and decision science) to: (1) integrate big data strategies, (2) build a data capability, and (3) be proactive in creating big data policies. They recognize the real value of big data (transactional and non-transactional) and related strategies, offers a pathway towards competitive advantage. They recognize the cost of data capture/acquisition is continually decreasing, and that big data value extends business capabilities to: (1) answer known questions, (2) experiment, (3) discover trends and opportunities, and (4) help managers in their decision making in ways never previously imagined.

LaValle *et al.* (2011) see value from big data as a value creation that occurs early in the business's development of analytics sophistication. Gandomi and Haider (2015) recognize the massive volumes of unstructured text, audio and video heterogeneous data (big data) can be efficiently text-mined with structured computational algorithms, delivering to the business a unique set of value characteristics.

Sivarajah *et al.* (2017) conclude big data holds the potential to release knowledge and value insights without loss from unstructured data. This values extraction then offers advanced business opportunities including: (1) strategic decision making, (2) improved strategic operational efficiency, (3) new revenue streams and (4) competitive business advantages. However such business values competitive opportunities remain complex.

Bradlow *et al.* (2016) adopt a knowledge investigation process to extract new predictive value components from retail-related big data sources and against five dimensions: (1) the customer, (2) the product, (3) the time, (4) the geo-spatial location and (5) the channel.

Erevelles *et al.* (2016) consider analyzing consumer-related big data as a pathway towards transforming tomorrow's marketing approaches. They propose tapping big data through: (1) physical resourcing - utilizing evidenced consumer activities; (2) human resourcing - extracting embedded consumer commentary perspectives; and (3) organizational capital resourcing - engaging embedded consumer insights. This process is contextualized against each firm's unique big data resourcing requirements.
Thus value can be something already present (noun), or something currently being built (verb) through multi-dimensional consumer perceived compilations around an item. In addition, a value perspective can change (or develop) over time as new valuing perspectives are being acquired (or are developing further), and this is time-line development represents an ongoing 'values deliverance' approach.

Hence, this study pursues a multi-dimensional, ongoing values deliverance approach. Firstly, it aims to extract values indicator items from the OSS big data contributions provided by the GitHub repository contributors. Secondly, it aims to group these values indicator items against consumer literature-supported (and ongoing) perceptive behavioral values dimensions. Finally, this study aims to add further big data contributor insights into the retail business domain.

Hence, this study investigates the second question:
RQ2: Can the relative strength of behavioral value dimensions support pathways to improve GitHub as an OSS services provider.

## VALUES DELIVERANCE

Bradley *et al.* (2012) see values deliverance as a dynamic construct encapsulating valuing across several dimensions and as changing over time. Sanchez *et al.* (2006) notes values deliverance as a dynamic acquisition process occurring across time and in stages. Values deliverance also involves the consumer in perceptive and behavioral considerations. Here a consumer extracts item valuing whilst considering what they are requiring of an item against their pre-conceived expectations of this item. This links into Aijen's Theory of Planned Behavior, and into the Theory of Reasoned Action. This implies that where values deliverance is mapped behaviorally it can result from a consumer's planned engagements and expectations, then into the consumer's actioning and trialing, and then through to the consumer's deductively reasoned concluding decisions.

Initially, values deliverance arises through the consumer's prior experiences around a desired item (ref). This pre-event knowledge can be sufficient to motivate the consumer into an item valuing pathway that assesses the items value dimension possibilities. Here the motivated consumer moves to a consumptive processing. This at-event item processing sees the consumer considering and trialing the item's potential values combinations against: (1) delivering performance, (2) delivering qualities, (3) delivering servicing, (4) delivering suitable economic-worth, and (5) delivering internal emotive-satisifier capabilities (Hamilton and Tee 2013). Eastman and Eastman (2015) also support behavioral motives as pre-cursors to delivering consumptive actions. Thus, in the value as verb phase of values deliverance, a Motivation Theory phase drives the consumer into a Consumption Theory phase.

The consumptive valuing perceived by the consumer also acts as a behavioral regulator (Sirdeshmukh, Singh, and Sabol 2002). At this point, the values deliverance process extends further - into the value as a noun phase. Here, the consumer's consumptive item valuing translates into their adopted reflective, after-use, self-gratification perspectives (Phipps et al 2013). This part of the values deliverance process is similar to a User Gratification theory approach, and it incorporates Goal Theory and Action Identity Theory.

This reflective value as a noun approach also allows for a re-motivation feedback loop that can then deliver further ongoing (loyal) consumer engagements around the item (Yang and Petersen 2004).

Rintamaki and Kirves (2017) also adopt reflective item considerations when developing value model feedback loops between a business' proposed values and its consumers' perceived values. They contextualize (country, channel, product-category, and competitive situation) their feedback loops against economic value, functional value, emotional value, and symbolic post-considerations of value acquired (gauged as reflective satisfaction and word-of-mouth comments).

Thus, values deliverance is a dynamic occurrence that builds behaviorally within the consumer across a period of time. It builds from the consumer's pre-event expectation and engagement motivational perspectives (Hamilton and Tee, 2013). Where the consumer accepts these motivations as sufficient the consumer then engages in at-event consumptive values experiencing actions. These then lead to the consumer generating post-event gratification reflections regarding their valuing experiences. These post event measures are typically gauged as satisfaction, and/or trust, and/or loyalty (Hamilton and Tee, 2013). Finally, values deliverance is also a dynamic process, and it embeds feedback loops between the post-event gratification decisions, at-event consumptive actions, and pre-event motivations (Hamilton *et al.* 2013). This values deliverance process is displayed as Figure 1.

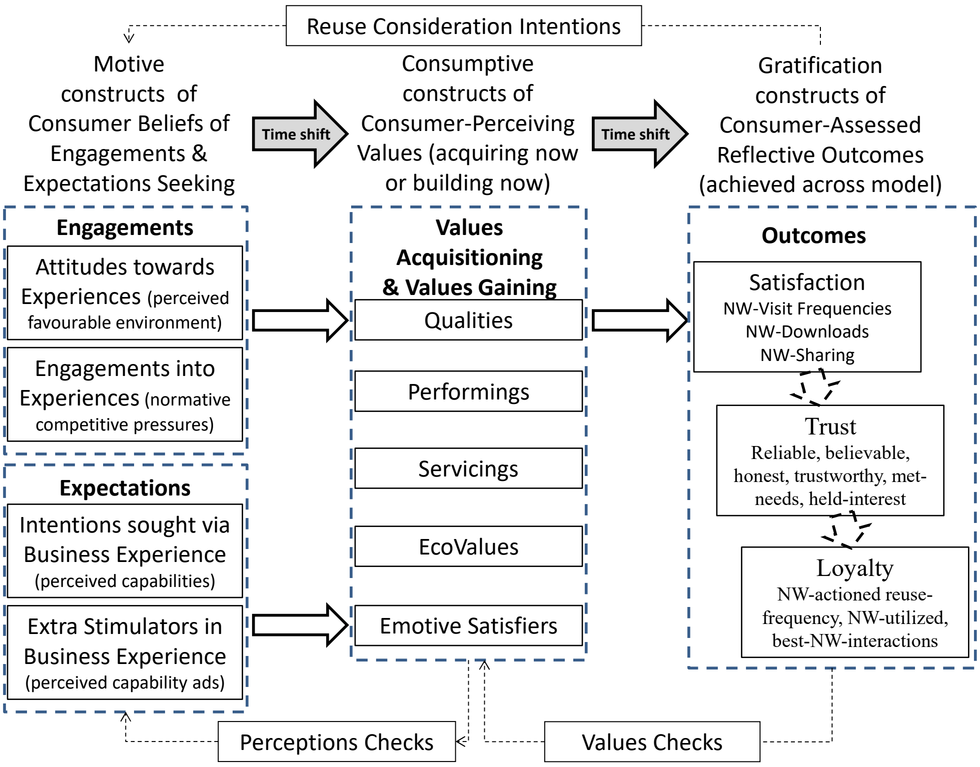

Figure 1: Values deliverance framework (adapted from Hamilton *et al.* 2013; Hamilton and Tee 2013)

Values deliverance is by nature multi-dimensional. Values deliverance is an acquisition process involving perceptions, and inducing actioning and trialing behaviors, and reflective summations. These can also be tangible, and/or intangible, and/or emotive, and they can be positive, or neutral, or negative.

**Behavioral Values in Service Business'**

The literature offers many variants when modelling values deliverance, but a sequential time-line and multi-dimensional approach offers a solid behavioral framework. Figure 1's values deliverance approach enlists the three stage, theoretically-supported construct blocks (West and Turner 2010).

Considering service business' Woodruff (1997) believes values are hierarchically developed, and they can be progressed from a consumer's desired values perspective (as pre-event expectations) through to the consumer's post-event perceived values deliverance outcomes. Mattson (1991) adds that these value dimensions can exist as mixes of emotional, practical and logical values.

Sheth *et al.*'s 1991 Theory of Consumptive Value delivers five value dimensions: (1) functional (performance), (2) social (servicing), (3) emotional (satisfiers), (4) epistemic (economic-worth of some novel item), and (5) conditional (quality). Sweeney and Soutar (2001) refined this behavioral work using: (1) price/value-for-money, (2) quality/performance, (3) social value and (4) emotive value, as their values measures. Roig *et al.* (2009) followed a similar path, measuring behavioral value across services through: (1) social, (2) emotive, and (3) functional (performance/reputation, qualities delivered, servicing, and economic) component measures. These value components also modelled into the deliverance of satisfaction and then loyalty.

Boksberger and Melsen (2011) offer values as multi-dimensional, and encompassing social behaviors including: (1) attitude, (2) ideology, (3) beliefs, and (4) justifications - but they recognize values must be set against rules, criteria, norms or ideals and preferential judgments.

Parente *et al.* (2015) support the behavioral perception of value - using a service's: (1) quality (quality experiences), (2) reputation (performing), (3) convenience (servicing), and (4) monies committed (economic-worth) measures. Hamilton *et al.* (2013) build behavioral values from pre-event motivational factors through to five at-event consumptive values, and then through to reflective post-event gratifications. Here, value acquisition is a real-time trialing and 'occurring-now' suite of the consumer's actioning experiences – gauged against the various dimensions of the business' service offering. Thus the value being derived is actually the net combination of the Figure 1 suite of 'at-event' consumer actioning (experiential) investigations.

Over time, these, and other researchers, have literature-developed, and empirically-tested, various consumptive consumer behaviors regarding the business' specific service offerings. Many have modelled the input and output constructs that contextualize the consumer's perceived value acquisitions. The input and output constructs of these research works are summarized in this study into Figure 1's Value Deliverance framework.

West and Turner (2010) support that real-time values acquisitions by the consumer are consumptive behaviors which in-turn deliver reflective gratification outcomes. Hamilton and Tee (2013) add that the values acquisitioning process commences when pre-event motives are sufficient to then stimulate a consumer to action their consumptive trialing processes - checking (1) performance, (2) quality, (3) economics, (4) servicing, and their (5) emotive satisfier feelings.

Thus, the Figure 1 progressive timeline-shift likely occurs from the consumer's pre-event item motives, through to their at-event consumptive item trialing behaviors, and then through to their resultant post-event item gratification conclusions (Sanchez *et al.* 2006; Bradley *et al.* 2012), and this values deliverance process has embedded feedback loops that update (and reengage) the at-event, or the pre-event constructs (Hamilton and Tee 2013).

Researchers support behavioral motives as delivering consumption (Eastman and Eastman 2015), such reflective behavioral and conjoint value approaches offer pathway inclusions into a consumer's 'post-event' reconsideration intention feedback loop, and/or into a values checking feedback loop. However, they do not capture the values acquiring processes that comes through a service experience, such as occurring in continual big data contributions processes or in a service experience process. Thus, as shown in Figure 1, the deliverance of value remains an ongoing, and cyclical process arising through continuing accumulations across the suite of values deliverance measures.

Value resulting from an available service remains the accumulative processing of consumer's behavioral consumptive actioning 'now.' It occurs at the time it is being consumed or experienced. Thus, accurate value measures are actually behaviorally consumptive measures and vice-versa. The key Figure 1 consumer-assessed, post-event, reflective outcomes in value deliverance fit under the satisfaction, loyalty, and trust construct domains.

## SERVICE BUSINESS' APPROACH

Services business' deliver valued actions, deeds, efforts, costs or qualities that target satisfying a consumer's: (1) needs, (2) wants, (3) demands, (4) desires, and (5) fun. Such services can be highly-skilled, and can involve intangible and tangible expert solutions. Services business' also spread across sectors including: accounting, banking, benchmarking, consulting, consumer management, education, insurance, operations, OSS development, professional services, social networking, retailing, tourism, and transport.

### Locating GitHub Behavioral Values

Big data OSS developer sites like GitHub behave as a development and storage hub for all aspects of a software project's development. Thus, GitHub acts somewhat like a service business.

Hence, a service business research approach may offer useful insights into the values building in repositories, and this approach may indicate how big data can be analyzed from a values deliverance perspective. In GitHub the consumer's role in values deliverance is replaced by the contributions of those participating. These OSS developers, and their associated responders add value into the GitHub repositories through: (1) their comments, (2) their 'readme' contributions, (3) their internal/external assessments and (4) their code development additions or deletions contributions.

Today, those working in the big data behavioral values field are now considering service business' values-related issues including: (1) approaches to extract data, (2) purposeful data extraction techniques, (3) values measurement, and (4) new business engagement opportunities (Bradlow *et al.* 2017; Cosentino *et al.* 2017; Erevelles *et al.* 2016; Grewal *et al.* 2017; Hamilton *et al.* 2017). However, the extraction of the participant-contributed value items from within a big data development and storage hub like GitHub still remains unfocussed, and big data values deliverance remains an ongoing domain of research investigation.

**Applying Behavioral Values to Big Data Streams**

The above values delivery systems are actually survey based and point-in-time text response models. With big data providing the input constructs to Figure 1's Value Deliverance Model. GitHub projects continually acquire text information from multiple sources - stars and watchers are attitudinal; pulls and forks are engagements; contributors and commits are intentions; and project releases are stimulators. These all link to the 'text' values capture and then culminate in project activities levels. The linking of big data text storage to values deliverance offers an opportunity to recompile the values system for: (1) a business, and/or (2) a service, and/or (3) a developer/R&D environment. Many big data mining approaches exist (Bello-Orgaz *et al.* 2016; Cosentino et al 2017). Several text mining approaches also exist. Dean and Ghemawat (2004; 2008), Zaharia *et al.* (2008), and Shim (2012) mine large scale data-intensive applications in clusters using MapReduce algorithms.

MapReduce algorithms often engage with large (terabyte) data sets and can generate intermediate value pairings as a reduced function containing all the intermediate value pairings against a selected intermediate key. Google uses query languages built into various parallel databases systems to execute over 100,000 MapReduce tasks per day. MapReduce uses Hadoop OS algorithms to delivers on small pieces of data including: (1) large-scale graphics, (2) text, (3) machine learning, and (4) statistical machine translation (Pavlo *et al.* 2009; Fernandez *et al.* 2014) - with multi-data assessments processing in-parallel at the working node.
In text mining GitHub repositories, MapReduce parallelizes and executes across selected repositories and it runtime solves: (1) input-data-partitioning, (2) repository-execution-scheduling, (3) repository-handling-failures, and (4) inter-repository-management-communications (Dean and Ghemawat, 2010).

MapReduce uses 'map' functions to deliver each chosen word aligned to a paired count-of-occurrences. It splits text data into small, independent, intermediate, key word/value pairs (or 'subset chunks'). The 'reduction' functions then shuffle and combine these key words/values at the same working node. Word counts and a common concept/output are generated. Next, an execution wrapper reduces these words/concepts into a sorted output list of words/values counts (Fernandez *et al.* 2014). Each output count can then be collated (or word-parceled) into its specific value construct, and analyzed using the Figure 1 values deliverance system. This staging system is shown in Figure 2 – with the values deliverance included during stage 3.

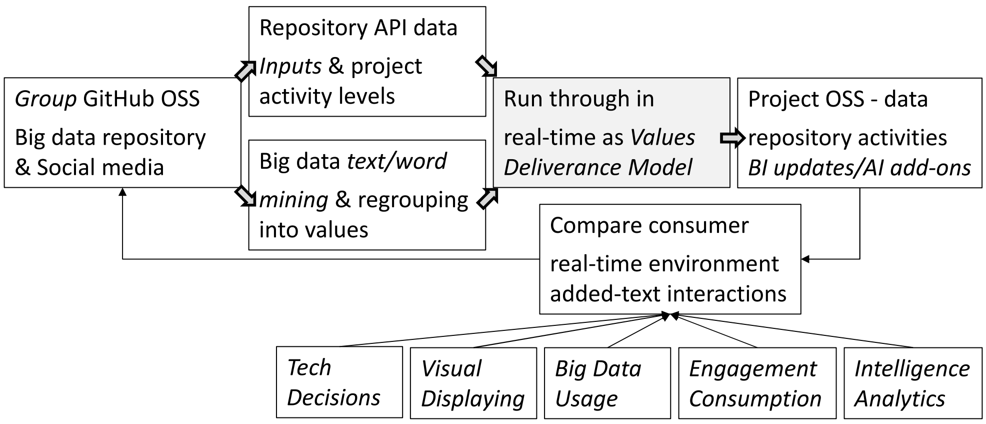


Figure 2: Values deliverance: Big data staging system

The specifically-targeted consumer-business engagement environment, and its chosen service offering, is first defined. In this case, it is GitHub JavaScript, and the strength of the values deliverance system mined from within its repository projects. Next, the values deliverance system construct items are added (and can be extended by co-tracing and adding word synonyms). Then, a reduction algorithms approach is applied at the working node. Measurement of these reduction outcomes then allows application through real-time mechanistic path models.

Considering Figure's 1 and 2, the emergent individual construct items within the values delivery system can be selectively gauged against the above gratification groupings, and/or against activities outcomes measures such as GitHub project activities levels

(Hamilton *et al.*, 2017). Interpretable, standardized path strengths, and standardized total effects then offer insights into business-applicable outcomes. This is shown in Figure 3. Here the gratification post-event measures are encapsulated through the project's activities levels. Other measurement items involved can also MapReduce into satisfaction, trust, loyalty, and feedback measures in a similar way to the motive and consumptive MapReduce measures discussed across Figures 1 and 2.

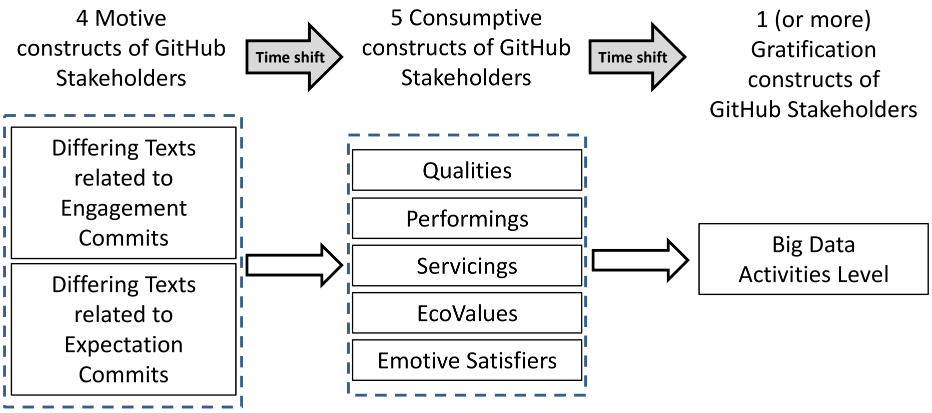

Figure 3: Values deliverance model

This real-time analysis can also incorporate artificial intelligence and machine learning responsive solutions. For example, Grewal *et al.* (2017) link retail consumer engagement decision making to a raft of mineable considerations: (1) technologies-deployed, (2) aesthetic visual displays, (3) big data usage, (4) activities analytics levels, and (5) consumptive engagements. These can be intelligently-designed to behaviorally-monitor (and respond to) external interactors, and they can be incorporated via feedback loops into Figure 2's value deliverance staging system. This adds further intelligences towards determination and provision of near-real-time, business-relevant, consumer-targeted, value deliverance system outputs.

## CONCLUSION

Our approach suggests the OSS big data platform GitHub can be data mined to assess the values embedded within its participant's actions (RQ1) , and that the relative strength of behavioral value dimensions likely does support pathways to improve GitHub as an OSS services provider (RQ2).

Hence, this study now follows Figures 1, 2 and 3 to API extract information within chosen GitHub repositories, and to assessment the relative activities contributions into a GitHub project. Like Google, we now deploy the values deliverance framework measures using MapReduce to text mine the projects' information (previously extracted through GitHub's restrictive API's). These text components are generated by GitHub's OSS developers and their associated follower ecosystems. This approach allows the extraction and collation of the text components (text comments, their text 'readme' contributions, their internal/external text assessments, and their text code development additions or deletions contributions).

Data mined text items are collated, regrouped, and deployed into the values deliverance model (Figure 3) as motive, consumption and gratification constructs. The values-deliverance model allows GitHub and its project creators (and their core team of lead developers) to better understand the significant path relationships that exist within GitHub's ecosystem, and within its projects and their communities of stakeholders.

The values deliverance model exposes the significant total effects pathways. These can be interpreted by GitHub and its project creators (and their core team of lead developers). Decisions can then be applied towards influencing GitHub project activities. This can expedite GitHub project release solutions delivering enhanced qualities, efficiencies, cost savings, and time reductions.

## REFERENCES

[1] Aggarwal, K., Hindle, A., & Stroulia, E. (2014). Co-evolution of project documentation and popularity within GitHub. In Proceedings ACM 11th Working Conference on Mining Software Repositories (pp. 360-363), Hyderabad, India May 31 – June 1.
[2] Ajzen, I. (1991). The theory of planned behaviour, Organizational Behavior and Human Decision Processes, 50(2), 179–211.
[3] Alshomali, M.A., Hamilton, J.R., Holdsworth, J., & Tee, S. (2017). GitHub: factors influencing project activity. In Proceedings 17th International Conference on Electronic Business (pp. 1-6), Dubai, UAE, Dec 4-8.
[4] Anderson, C. (2008). Marketing, scientific process, and scientific method. Journal of Marketing 47(4), 18-31.
[5] Anderson, P.F. (1983). What are dynamic capabilities and are they a useful construct in strategic management? International

Journal of Strategic Management Reviews, 11(1), 29-49.
[6] Badashian, A.S., Shah, V., & Stroulia, E. (2015). GitHub's big data adaptor: an eclipse plugin. In Proceedings 25th Annual International Conference on Computer Science and Software Engineering (pp. 265-268), Markham, Canada, Nov 2-4.
[7] Barnes, J. (1984). The Complete Works of Aristotle. The Revised Oxford Translation (2 vols), Bollingen Series, 71(2), Princeton, NJ: Princeton University Press.
[8] Bello-Orgaz, G., Jung, J.J., & Camacho, D. (2016). Social big data: Recent achievements and new challenges. Information Fusion, 28, 45-59.
[9] Bissyande, T.F., Thung, F., Lo, D., Jiang, L., & Reveillere, L. (2013). Popularity, interoperability, and impact of programming languages in 100,000 open source projects," In Proceedings 2013 IEEE 37th Annual Computer Software and Applications Conference (pp. 303-312), Kyoto, Japan, July 22-26.
[10] Bodnar, T., and Salathe, M. 2013. Validating models for disease detection using twitter. In Proceedings. ACM 22nd International Conference on World Wide Web (pp. 699-702), Rio de Janeiro, Brazil, May 13-17.
[11] Bolton, R.N, & Lemon, K.N. (1999). A dynamic model of customers' usage of services: Usage as an antecedent and consequence of satisfaction. Journal of Marketing Research, 36(2), 171-186.
[12] Borges, H., Hora, A., and Valente, M.T. (2016). Understanding the factors that impact popularity if GitHub repositories. In Proceedings 32nd International Conference on Software Maintenance and Evolution (pp. 334-344), Raleigh, North Carolina, Oct 2-10.
[13] Bradlow, E.T., Gangwar, M., Kopalle, P., & Voleti, S. (2016). The Role of Big Data and Predictive Analytics in Retailing. Journal of Retailing, 93(1), 79-95.
[14] Capra, E., Francalanci, C., Merlo, F., & Rossi-Lamastra, C. (2011). Firms' involvement in Open Source projects: A trade-off between software structural quality and popularity. Journal of Systems and Software, 84(1), 144-161.
[15] Cassidy, L.J. (2016). Website Benchmarking: an abridged WAM study. PhD Thesis, James Cook University, Australia pp. 1-239.
[16] Cattell, R.B. (1943). The description of personality: Basic traits resolved into clusters. Journal of Abnormal and Social Psychology 38(4), 476-506.
[17] Chen, Y., Argentinis, J.E., & Weber, G. (2016). IBM Watson: how cognitive computing can be applied to big data challenges in life sciences research. Clinical Therapeutics, 38(4), 688-701.
[18] Chiu, C.M., & Huang, H.Y. (2015). Examining the antecedents of user gratification and its effects on individuals' social network services usage: the moderating role of habit. European Journal of Information Systems, 24(4), 411-430.
[19] Cosentino, V., Canovas Izquierdo, J.L., & Cabot, J. (2017). A systematic mapping study of software development with GitHub (pp. 7173-7192). IEEE Access, Mar 27.
[20] Cunningham, E. (2008). A Practical Guild to Structural Equation Modeling using AMOS. Melbourne, Australia: Streams Statsline.
[21] Dabbish, L., Stuart, C., Tsay, J., & Herbsleb, J. (2012). Social coding in GitHub: transparency and collaboration in an open software repository. In Proceedings ACM 2012 Conference on Computer Supported Cooperative Work (pp. 360-363), Seattle, Washington, Feb 11-15.
[22] Davenport, T.H. (2006). Competing on analytics. Harvard Business Review 84(1), 98-109.
[23] Dean, J., & Ghemawat, S. (2004). MapReduce: Simplified Data Processing on Large Clusters. In Proceedings. 6th USENIX Symposium on Operating Systems Design and Implementation, OSDI 2004 (pp. 137-149), San Francisco, CA, 6-8 Dec.
[24] Dean, J., & Ghemawat, S. (2008). MapReduce: simplified data processing on large clusters. Communications of the ACM. 51(1), 107-113.
[25] Dean, J., & Ghemawat, S. (2010). MapReduce: a flexible data processing tool. Communications of the ACM, 53(1), 72-77.
[26] Du, R.Y., & Wagner A.K. (2012). Quantitative trendspotting. Journal of Marketing Research, 49(4), 514-536.
[27] Eastman, J.K., & Eastman, K.L. (2015). Conceptualizing a model of status consumption theory: an exploration of the antecedents and consequences of the motivation to consume for status. Marketing Management Journal, 25(1), 1-15.
[28] Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big data consumer analytics and the transformation of marketing. Journal of Business Research 69, 897-904.
[29] Fernandez, A., del Río, S., Lopez, V., Bawakid, A., del Jesus, M.J., Benítez, J.M., & Herrera, F. (2014). Big Data with Cloud Computing: an insight on the computing environment, MapReduce, and programming frameworks. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 4(5), 380-409.
[30] Fishbein, M., & Ajzen, I. (1975). Belief, attitude, intention, and behavior: An introduction to theory and research. Reading, MA: Addison-Wesley.
[31] Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods and analytics. International Journal of Information Management, 35(2), 137-144.
[32] Gangwar, M., Kumar, N., & Rao, R.C. (2013). Consumer stockpiling and competitive promotional strategies. Marketing Science, 33(1), 94-113.
[33] Gartner, (2012). IT glossary - Big Data. Accessed 10th July 2017 at: http://www.gartner.com/it-glossary/big-data/
[34] Gerber M.S. (2014). Predicting Crime Using Twitter and Kernel Density Estimation. Decision Support Systems, 61(May),

115–125.
[35] Grewal, D., Roggeveen, A.L., & Nordfalt, J. (2017). The future of retailing. Journal of Retailing, 93(1), 1-6.
[36] Hair, J.F., Black, W.C., Babin, B.J., & Anderson, R.E. (2013). Multivariate Data Analysis. Harlow, UK: Pearson, Education.
[37] Hamilton, J.R., Prideaux, M.C., & Tee, S. (2013). Linking car racing expectations and value with satisfaction, trust and loyalty outcomes. Proc. 12th International Decision Sciences Institute and 6th Asia-Pacific Decision Sciences Institute Conference (pp. 875-888), Nusa Dua, Bali, Indonesia.
[38] Hamilton, J.R., & Tee, S. (2013). Social network site value deliverance feedback loops. In Proceedings 13th International Conference of Electronic Business (pp. 108-110), National Chengchi University, Singapore, Dec 1-4.
[39] Hamilton, J.R., Tee, S., & Franklin, T. (2016). Gauging loyalty: A local swim school fitness training study. In Proceedings. 16th International Conference on Electronic Business and 1st Global Conference on Internet and Information Systems (pp. 1-7), Xiamen, China, Dec 3-7.
[40] Hamilton, J.R., Holdsworth, J., Tee, S., & Alshomali, M.A., (2017). Analysing big data projects using GitHub and JaveScript. In Proceedings 17th International Conference on Electronic Business (pp. 1-5), Dubai, UAE, Dec 4-8.
[41] Hata, H., Todo, T., Onoue, S., & Matsumoto, K. (2015). Characteristics of sustainable OSS projects: A theoretical and empirical study. In Proceedings 2015 IEEE/ACM 8th International Workshop on Cooperative and Human Aspects of Software Engineering (pp. 15-21), May 18, Florence/Firenze, Italy.
[42] Henderson, S. (2009). How do people manage their documents?: An empirical investigation into personal document management practices among knowledge workers. PhD Thesis (pp. 1-285). Retrieved from ResearchSpace@Auckland.ac.nz (20 June 2017).
[43] IBM. (2012). "IBM What is Big Data: Bring Big Data to the Enterprise. IBM. Retrieved from http://www-01.ibm.com/software/data/bigdata/ (20 May 2017).
[44] IBM. (2015). "Data IB, Hub A. The Four V's of Big Data," IBM. Retrieved from http://www. ibmbigdatahub.com/infographic/four-vs-bigdata (6 June 2017).
[45] Jiang, J., Lo, D., He, J., Xia, X., Kochhar, P.S., & Zhang, L. (2017). Why and how developers fork what from whom in GitHub. Empirical Software Engineering, 22(1), 547-578.
[46] Kalliamvakou, E., Gousios, G., Blincoe, K., Singer, L., German, D.M., & Damian, D. (2017). An in-depth study of the promises and perils of mining GitHub. Empirical Software Engineering, 21(5), 2035-2071.
[47] LaValle, S., Lesser, E., Shockley, R., Hopkins, M.S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. MIT Sloan Management Review, 52(2), 21-31.
[48] Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: traps in big data analysis. Science 343(6176), 1203-1205.
[49] Lee, M.J., Ferwerda, B., Choi, J., Hahn, J., Moon, J.Y., & Kim, J. (2013). GitHub developers use rockstars to overcome overflow of news. In Proceedings 2013 ACM Proceedings CHI'13 Extended Abstracts on Human Factors in Computing Systems (pp, 133-138), Paris, France, Apr 27 - May 2.
[50] Lima, A., Rossi, L., & Musolesi, M. (2014). Coding Together at Scale: GitHub as a Collaborative Social Network In Proceedings 8th International AAAI Conference on Weblogs and Social Media (pp. 295-304), Ann Arbor, Michigan, July1-4.
(*Full reference list is available upon request from the corresponding author.)