# Converged Reality: A Data Management Research Agenda for a Service-, Cloud-, and Data-driven Era

Michael Smit
Dalhousie University
Halifax, Canada
Mike.Smit@dal.ca

## Abstract

*We are accustomed to distinguishing activities that occur on or through the Internet as distinct from activities that occur in the physical world: online versus offline, virtual reality versus reality, and so on. As Internet-based services have evolved, this distinction has continued to blur. We now have a* ***converged reality****: the online does not merely augment the offline; rather, the two are increasingly indistinguishable. Mobility, cloud computing, service-driven technology, cognitive computing, and Big Data analytics are some of the distinct but related innovations driving this shift. Because the shift is happening in pieces across multiple areas and sectors, our converged reality is emergent and grassroots, not a carefully planned joint effort. There are therefore areas that have been and will be slow to acknowledge and adapt to this shift; data management is one of these areas. This paper describes how this converged reality grew from previous research into bridging online and offline worlds, and how it will lead to a cognitive reality. It identifies enablers and dampeners, and describes a data management research agenda specifically for converged reality. The proposed research agenda is intended to spark discussion and engage further work in this area.*

## 1. Introduction

Ongoing technical innovation (including product development, industrial R&D, academic research, and others) has converged the online and offline worlds. Original definitions describe *online* as a virtual space, which includes Internet-connected services accessed via some form of connectivity, and *offline* to be the physical world without this connection to services (i.e., the world that predates the Internet). It is clear that the offline world has been substantially influenced by the online world; we go further, and suggest that the current state of technology represents a shift worthy of description and reflection.

This is not the first time the blurred distinction between online and offline has been written about; if we accept the mid-1990s as the point when the Internet began to substantially impact society, there are two decades of articles (and later, blogs and tweets) describing the blurred line between the online and offline world [20]. Nor is it likely to be the last, as technology continues to evolve. It should also be noted that the two are converged but not unified; so long as there remain tasks we would prefer to do in person, and would do more effectively in person, it will be clear that online is still distinctly identifiable from offline. Nonetheless, there has been a recent and a substantial shift with implications for business models, technology development, and data management – among others.

It is natural that the evolution of technology proceeds faster than practices and policies that work toward adoption and maximum utility for that technology. We had databases before we understood privacy, we had the Internet before we understood music piracy, and we had social media before we understood privacy (again). When defining a ground-breaking technology like cognitive computing, or developing a novel interface like a virtual reality headset, or observing an industry-wide evolution to provide everything-as-a-service, data management is often not a priority. However, scholars and businesses should not neglect this important infrastructure that forms the scaffolding of our converged reality.

There are two main contributions in this paper. First, a synthesis of current and predicted technology and their connections and implications, based on academic papers, trade shows, popular media, and trade publications, to better understand the transition to a converged reality. Section 2 describes and defines the converged reality in more detail, Section 3 describes an information flow abstraction for converged reality and places emerging technologies within that abstraction, and Section 4 describes factors that enable converged reality and factors that slow its growth. Second, we present a high-level research agenda for data management in the context of converged reality (built on general principles of data management, Section 5), focused on user-centric data in a business-to-consumer environment (Section 6).

## 2. Converged Reality

The relationship between the online and offline worlds has been of interest since the dawn of the online world [20]. Often the relationship is mediated by human actors; that is, the offline world is impacted because people are impacted. Figure 1 provides a broad overview of different stages of a direct, unmediated relationship between online and offline: initially created separate, the two were soon *connected* directly. Connectedness has gradually transitioned to a state described here as *converged*, and we are currently looking ahead to a *cognitive* relationship. Each stage is described below.

**Connected**. There has been substantial interest in connecting the online and offline world, perhaps best captured by cyber-physical systems (CPS) research (for example, see Kyoung-Dae & Kumar's historical perspective [10]and Rajkumar et al.'s overview [19]). CP Systems seek to improve the physical world by drawing on computing and communications technologies [10]; one of the most active areas of research was how to construct this bridge and on the role of converging technologies; there was less discussion about the converging human experience [10].

Another well-known example of connectedness is mixed reality [13](sometimes referred to as a dual reality, emphasizing that parallel worlds are connected [11]. This includes augmented reality, where elements of the online world are injected into the offline world; for example, using sensor networks to connect virtual worlds (e.g. Second Life) to corresponding physical worlds [11] potentially using smart services [21]. Similarly, augmented virtuality injects elements of the offline world into an online virtual world [5].

As the gradient in Figure 1 suggests, the dividing line between connected and converged is not distinct; the goal is not to classify technology or research as distinctly one or the other, but rather to identify the overall direction of technology. That is, as cyber-physical systems and mixed reality systems have continued to evolve, they have ushered in the era of converged reality.

**Converged.** In the connected state, a system of related *components* bridges the online and offline worlds. In the converged state, a system of *systems* blurs the distinction between the two.

A key characteristic of a converged reality is that no particular effort is required to navigate between online and offline: one doesn't invoke an app, or click the analyze button. There is no conscious decision to proceed offline or online; one simply performs a task.

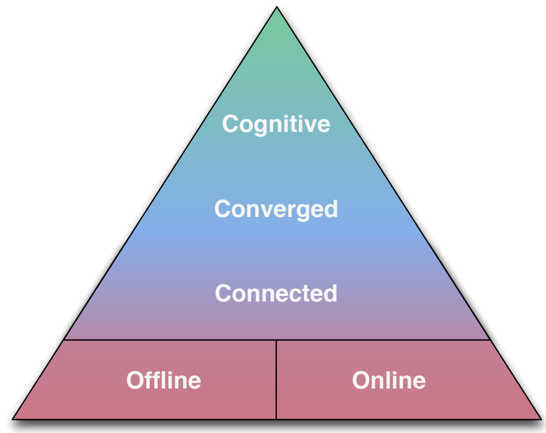


**Figure 1: The stages of integration between online and offline worlds.**

For example, when booking a car using the Uber app, there is no online/offline split: an action happens online, and causes events in the physical world. Events in the physical world (the movement of a car toward your location, for example) similarly propagate to the online world. Substantial behind-the-scenes complexity is masked by an equally complex broker, but to the user the online and offline merge seamlessly. Where once we waited for a modem to dial a number to connect us to online services, and we booked server time in advance, we now do not need to plan in advance or explicitly decide to use an online service. The information needed in a moment is simply present, and the user interaction is seamless.

A second key characteristic is bi-directionality: data flows from the offline world to the online world (through sensors, like smart homes and personal heart rate monitors), and online data is manifested in the offline world (through various forms of user interface, 3D printing, etc.). The offline and online thus have similar information spaces: the online will, for now, be an incomplete representation of the physical world. (Note that although there is bi-directionality, because the physical world predates the online world, it is common to speak of the online world as encroaching upon or subsuming the physical/offline world).

**Cognitive.** In this state, the human element becomes less important in the system of systems of the converged state (though the user *experience* becomes more important), as we augment human intelligence with sophisticated analytics and cognitive computing (computation support that seeks to emulate human thinking processes).

Early converged reality systems enjoy a novelty factor, are experienced primarily by tech-savvy early adopters, and are (relatively) few in number. As their prevalence increases, users will be less willing to engage in the cognitive overhead that comes along with increased immersion in the offline and online worlds simultaneously (they may experience *information overload* [1] or may realize they cannot comprehend enough of the data to be confident they are making the right decision). That is, rather than watch to ensure our Uber driver is not lost, we will want to delegate this task to a smart service powered by analytics backed with cognitive computation. We discuss the challenge of moving past human scale data later in this paper.

While this state is the last state discussed in this paper, it is not terminal; there will be continued evolution.

## 3. Emerging Technology, Flow of Information, and a Converged Reality

The online-offline convergence is driven by existing and emerging technology. While much of this paper describes existing and evolving technology, this section is focused on emerging technology. Gartner Research refers to a "Nexus of Forces" which includes "mobile, social, cloud and information" [6]. Of the "post-Nexus" stage, they write "the concept of blurring the physical and virtual worlds are strong concepts in this stage. Physical assets become digitalized and become equal actors in the business value chain alongside already-digital entities, such as systems and apps." Their focus is on stages of business, and the choice each organization has in terms of embracing a converged reality, but the principle holds true in this research context.

Converged reality relies on emerging technology and vast amounts of data and information. Given our interest in data management, it is important to understand the flow of information. The description of the converged state as a system of systems, and the defining characteristics of seamless, autonomous operation, and the multiple flows of data, is similar to language used when describing adaptive/autonomous systems. We use a model of information flow commonly used to describe adaptive systems, Monitor-Analyze-Plan-Execute with a knowledge base (MAPE-k; Figure 2) [9], as a useful abstraction for the flow of information in converged reality. An adaptive system employs a great deal of information about the environment, and the system it is managing, in order to make decisions to improve the overall system. The Monitor stage uses sensors to measure key attributes. The Analyze and Plan stages assess these attributes,

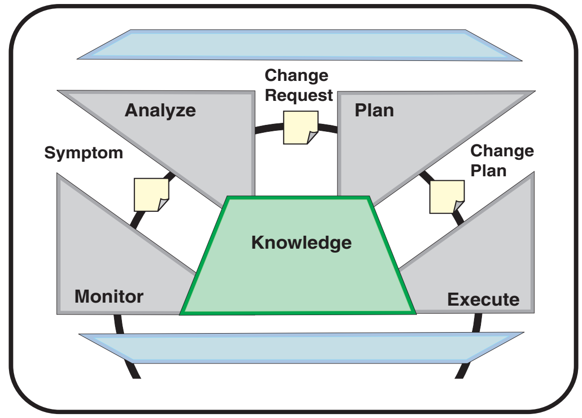


**Figure 2: The Model-Analyze-Plan-Execute feedback loop, with a knowledge base. From** [9].

perform some type of analysis (which may include attempts to model or predict the future proactively), and formulate a plan to respond appropriately. This plan is implemented in the Execute stage via actuators. All the stages make use of a Knowledge Base to store sensor data, analysis, etc. [4].

In this section we describe this information flow abstraction in more detail, describe the current converged reality using this abstraction, and approximately place emerging technologies (as identified by Gartner Research [6]) within the information flow.

**Monitor**: Representing more of the physical world online involves improved sensors that collect vast amounts of data about individuals, from self-posted social media to full-featured location-aware smartphones to quantified self (see Section 4.1). This input data is often unstructured. Improved sensors also require improved understanding and processing to properly understand and aggregate sensor input. One common input mechanism is speech, through natural-language question asking and speech recognition; cognitive computing is well-suited to this role, as are other supporting technologies. Another input mechanism becoming more common is gesture-based control.

**Analyze & Plan**: The increased processing required by rich sensor data is handled by having access to analytics services in the cloud, which provide unprecedented scale. The cloud also provides the infrastructure for the **knowledge base**, storing sensor and analysis data as required. Decision / predictive / prescriptive analytics aim to suggest courses of action.

**Execute**: In this MAPE stage, through the use of actuators, the physical world is modified. In our context, this includes improved display devices, virtual reality headsets, wrist displays, and other novel forms of user interface. It would also include 3D printing of objects; creating physical objects (including food [15])

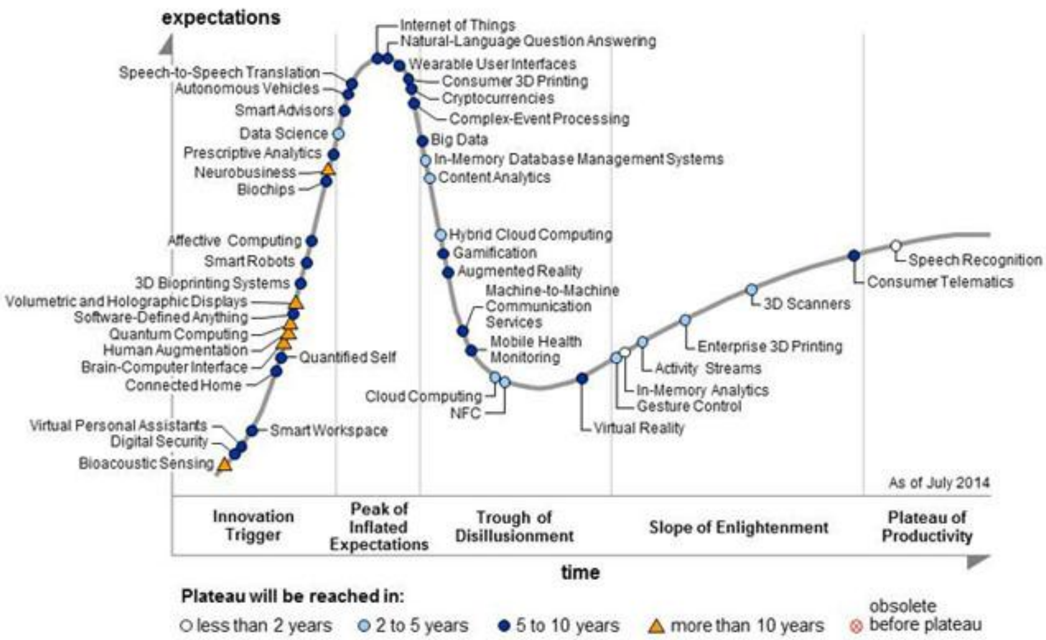


**Figure 3: Gartner Hype Cycle, July 2014. From [6].**

from the online world is a great example of a converged reality. Finally, natural language question answering requires not just speech recognition, but also the structure and synthesis of appropriate spoken responses.

Figure 3 shows the Gartner Research hype cycle from July 2014 [6]; this chart shows emerging technologies and their maturation, through a period of inflated expectations and into productivity. All but a few of these items reflect a convergence of online and offline, or a technology that enables such a convergence. Table 1 presents these categorized into categories loosely based on the MAPE-based information flow we just described: advanced sensors, supporting technologies that enable sophisticated responses, and advanced actuators that bring online data to the offline world. There are also some examples of complete systems, "converged reality features", that use the supporting technologies to provide value directly to consumers. Although we refer to reality as being converged, it is clear that the overlap can increase further. Items currently ascending Gartner's 2014 hype cycle include Brain-Computer Interfaces, 3D Bioprinting, and Human Augmentation, each of which will contribute to further merging the online and offline realities.

## 4. Enablers and Dampeners of Converged Reality

This section describes the most relevant factors to the convergence of the online and offline worlds, both those that encourage convergence and those that inhibit. There is no clear line between connected reality and converged reality, so these factors are divided into enablers (which generally move us toward the top of the triangle in Figure 1) and dampeners (which generally slow the move toward the top of the triangle). Some of these dampeners are reasonable and important, and thus the less pejorative term "dampener" is used in lieu of inhibitor or barrier.

### 4.1. Enablers

Some enablers have existed for long enough they now seem prosaic: the prevalence of online maps; the ubiquity of wireless communication technology (3G, LTE, 802.11, etc.); mobile devices that are always on and always connected; virtual reality (though there has been significant improvement in this area); services-oriented computing; mobile app marketplaces; etc. Other enablers are generally understood: the Internet of Things is important to appropriate sense the physical world and create quantified data the online tools can work with; novel input/output devices ranging from gestures to brain waves to 3D printing further cement the online/offline link; etc. However, a few enablers merit additional explanation, as follows:

**Social media.** Individual technologies contribute to a converged reality in different ways through direct data collection and sharing, but the broader impacts are more important. The first is the ability for everyone to share and document their lives digitally, in a far more accessible and immediate way than previously

**Table 1: Items from the July 2014 Gartner Hype Cycle, categorized**

| Sensors | Supporting Technology | Actuators | Whole System (Converged Reality Example) |
|---|---|---|---|
| Bioacoustic sensing<br>Human augmentation<br>Brain-computer interface<br>Mobile health monitoring<br>Quantified self<br>Gesture control<br>3D scanners<br>Consumer telematics<br>Speech recognition<br>Activity streams<br>NFC | Software-defined anything<br>Affective computing<br>Neurobusiness<br>Prescriptive analytics<br>Data science<br>Speech-to-speech translation<br>Internet of Things<br>Complex-event processing<br>Big Data<br>In-memory DBMS<br>In-memory analytics<br>Content analytics<br>Prescriptive analytics<br>Hybrid cloud computing<br>Cloud computing | Volumetric / holographic displays<br>3D bioprinting<br>Wearable user interfaces<br>Consumer/Enterprise 3D printing<br>Virtual reality | Connected home<br>Smart workspace<br>Virtual personal assistant<br>Smart robots<br>Smartadvisors<br>Augmented reality<br>Natural-language question answering<br>Autonomous vehicles |

possible. The second is related: the general *willingness* to engaged in social sharing. The motivation to share data is social, not economic, which also means that the cost of gathering this data is relatively low. This willingness did not occur immediately, though it did occur rapidly; our social norms around sharing personal information have changed. If we offered a Twitter account to Warren and Brandeis in 1890, after they famously defended privacy in response to the advances in technology that made printing photos in newspapers feasible [3], they might not be as enthusiastic as Twitter's 300 million active users [23]. Mark Zuckerberg, founder of Facebook, was widely quoted on this willingness to share:

> "People have really gotten comfortable not only sharing more information and different kinds, but more openly and with more people… That social norm is just something that has evolved over time." [7]

Of course, it should be noted that while most recognize the shift in social norms, researchers suggest we care about privacy differently than before and that Facebook does not adequately capture the changed social norm (see for example [2]).

**Quantified self.** There is a growth in devices that sense the human body, converting our activity, autonomic systems, and even our thoughts into quantified data. These devices have moved from a niche market (e.g., running watches) to common consumer items (e.g. Fitbit, Apple Watch). Analytics are used to provide features like alarm clocks that adjust the wake-up time based on the detected sleep cycle, or the ability to share one's running route with friends.

**Rich individual spatiotemporal data.** This data is itself enabled by social media and cheap, accessible geolocation technology (including GPS and wireless-based location services). It is now commonplace to share, or at least log, one's location, a timestamp, and an activity or thought. This is incentivized socially, but also by providing tools and analytics that enable services. For example, our devices can track our daily commutes to estimate our morning drive time in advance based on traffic and weather. The promise of quantified self is to track our location and heartrate when running or cycling to provide personalized fitness advice. Analysis of this data is still challenging [25], but it is an active field of research and interest is growing. Nor is the data necessarily personal; industrial and consumer telematics collect detailed data about vehicles and their locations; AIS gives us readily available data about ships; smart luggage tracks its own location [22].

**Apps that integrate online and offline.** Uber is one of the best-known examples of an app that better connects users to the physical world; any other delivery app or taxi app achieves the same goal. This class of applications is a growing force in the international startup community.

**Cognitive computing.** Perhaps the best-known example is IBM's Watson competing on Jeopardy, and the various services and tools based on this technology that IBM has released. Generally, cognitive computing is "the goal of creating machines that do much more than calculate and organize and find patterns in data—they sense, learn, reason and interact naturally with people in powerful new ways" [8]. It is a key component in the converged reality described here, because it does not rely on consistently and predictably structured data. Relational databases have been the norm for 40 years, and with them comes carefully modeled, rigidly structured data. NoSQL and NewSQL have changed the landscape of data storage, but while some of the technology has matured to reliability, data management practices for NoSQL have not. Cognitive computing aims to establish more natural, seamless integration with humans. It should be clear that while it is the confluence of all of these enablers that created a converged reality, cognitive computing plays (or will play) a key role in ensuring interactions are seamless. While many of these technologies have existed for several years or longer, cognitive computing is a relatively recent technology that is still maturing.

### 4.2. Dampeners

**Privacy concerns.** This is the *de facto* issue raised any time more data is being collected, stored, shared, analyzed, and converged. As discussed in the research agenda, there are many open issues with regard to protecting privacy in a Big Data context.

**Lack of openness.** While some companies have adopted transparent policies, in general the mass collection of data is being conducted by businesses. These businesses aren't necessarily intending to deceive, but we know that online Terms of Agreement are read only infrequently. Even if users are aware of the data that is being collected, they may not be aware of the capabilities of the collector, or of the pooling of their information with other collectors.

**Interoperation.** The piecemeal evolution of this converged reality means that technologies have diverged. There is no widely accepted standard for the interoperation of these novel devices (including sensors and actuators). For example, there are several competing smart home technologies, with mixed compatibility. This is an issue that has impacted every major technical revolution; for example, the world wide web turned 25 years old in 2014, and marked 25 years of web browser compatibility issues.

**Bandwidth.** The rich data of a converged reality requires substantial bandwidth. This is particularly true when contemplating virtual reality, increased video use, and a substantial number of sensors. Adding bandwidth is expensive for all concerned, and a large number of people do not have access to high-speed Internet at home. As some of this data is more valuable to other stakeholders than to the data creator, alternative billing/business models may be required to provide this bandwidth while not increasing the digital divide.

**Big Data.** Often cited as an enabler, we suggest it is a dampener; Big Data is a problem. There are multiple contributors to the problem of Big Data. Social media turned everyone into a content creator, and the Internet of Things turns every*thing* into a content creator; worse, the value of this content is difficult to assess at the time of creation. The advent of cloud storage has made data accumulation less visible; cloud-enabled offices are not cluttered with paper files or USB drives, but the data stored in the cloud is often just as disorganized as a stack of paper files. Worse, the (effectively) infinite capacity of cloud storage avoids a decision point: individuals and companies never have to decide to spent money on adding additional storage; the storage is just there and available. The cost of storage continues to drop exponentially, from around $1,000 per GB in 1995 to $.03 per GB in 2015. The final key cause of Big Data is hope: the optimism that our data has great value, if only we can unlock it. Collectively, the result is the collection and storage of increasing amounts of data with the ambition of making use of it. But left unused, this data is a liability, firmly in the expense column. The use of appropriate tools – records management, data analytics, cognitive computing – can mitigate the impact, but given the volume, variety (lack of, or varied, structure), and velocity (rate of new data being created) of Big Data, existing techniques fall short.

## 5. Data Management

We are a data-rich society; perhaps even data-driven [17]. In 2012, analysts estimated 90% of the world's data had come into existence within the previous 2 years [24]. Data management is one of the tools to coping with this data, and ensuring that it is an asset, not a liability.

The Data Management Association, in their Data Management Body of Knowledge (DAMA-DMBOK) [14] defines data management as "an overarching term that describes the processes used to plan, specify, enable, create, acquire, maintain, use, archive, retrieve, control, and purge data" [14]. This task is embraced by data management professionals, including dedicated information managers and IT staff. It often involves managing data throughout its lifecycle (usually collection, primary use, secondary use, archiving or deletion).

While the next edition, DMBOK2, is expected to incorporate best practices for new and emerging data technology, DMBOK2 is still limited to including best practices that exist. Many existing best practices don't scale well when the volume, variety, and velocity exceed human capacity. Even policy-based solutions, which theoretically scale well, don't adequately address unstructured, semi-structured, or variably structured data once implemented. DMBOK organizes data

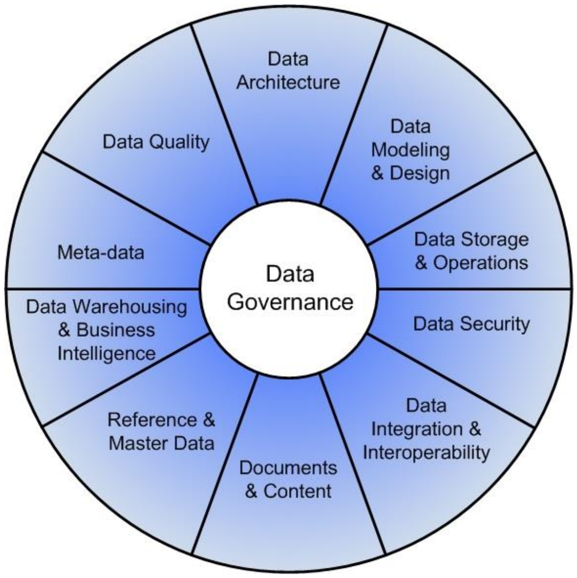


**Figure 4: The important areas of data management, from the Data Management Association, in their Data Management Body of Knowledge. From** [14].

management into 11 main areas, including Data Governance, a super-area that guides the other 10 areas. These main categories are shown in Figure 4, and described as follows (quoted from [14]):

- Data Governance – planning, oversight, and control over management of data and the use of data and data-related resources. Note that governance covers 'processes', not 'things'.
- Data Architecture – the overall structure of data and data-related resources as an integral part of the enterprise architecture
- Data Modeling & Design – analysis, design, building, testing, and maintenance
- Data Storage & Operations – structured physical data assets storage deployment and management
- Data Security – ensuring privacy, confidentiality and appropriate access
- Data Integration & Interoperability – acquisition, extraction, transformation, movement, delivery, replication, federation, virtualization and operational support.
- Documents & Content – storing, protecting, indexing, and enabling access to data found in unstructured sources (electronic files and physical records), and making this data available for integration and interoperability with structured (database) data.
- Reference & Master Data – Managing shared data to reduce redundancy and ensure better data quality through standardized definition and use of data values.
- Data Warehousing & Business Intelligence – managing analytical data processing and enabling access to decision support data for reporting and analysis.
- Metadata – collecting, categorizing, maintaining, integrating, controlling, managing, and delivering metadata.
- Data Quality – defining, monitoring, maintaining data integrity, and improving data quality.

## 6. Research Agenda for Data Management

Based on the description of converged reality, and the identified enablers and dampeners, we suggest a starting point for a high-level, user-centric research agenda. The goal of this agenda is to engage related research communities in a common effort to advance the state of data management in this emerging paradigm. There is not a robust body of data management research (as DMBOK defines data management) to build on in the connected reality space (for example, cyber-physical systems). The vitality of data to a converged reality, however, suggests the following initial set of research agenda items:

**Cognitive assistants**: The cognitive overhead of engaging in the online and offline world simultaneously may be tractable for humans if the number of systems is small and their attention is not divided. However, with more applications reaching the converged reality level, our ability and willingness to navigate these streams of information will decline. Cognitive assistants employ computation advances and cognitive computing to intelligently provide decision support to users. The human remains a part of the system, but with a reduced cognitive load. What is not clear is how cognitive assistants will be created, and by whom? To what extent, and when, are humans willing to delegate cognitive thought? How will humans interact with these assistants in a way that provides a positive user experience but also scales?

**Privacy.** With more personal data, stored in more places, the research challenge isn't even how to protect privacy: it is how to identify privacy information. Best practices for data privacy tend to rely on structured data. Given the past three-four years of Big Data and privacy research, how might that be applied to this context? What tools and methods are required to identify private data in unstructured, high-volume text? How do we securely store and manage data throughout its lifecycle?

**Data literacy.** Data literacy is the ability to *comprehend*, *create*, and *communicate* data, and is the first level of the tri-level literacy, fluency, mastery scale. Data-literate individuals have the knowledge, understanding, and skills to connect people to data, and to use and comprehend data themselves. A converged reality, with access to data visualizations in previously unknown volumes, is only useful if consumers are capable of using and comprehending the data. Core elements of data literacy would include basic analysis (how to transform raw data into usable information and/or knowledge), data visualization, metadata, and evidence-based decision-making (the effective and ethical use of data to inform policy-making, decisions, or even personal opinions). Elements of data literacy are taught, explicitly or implicitly, across all disciplines and at all levels of post-secondary institutions. Faculty have substantial expertise in these areas and many students will graduate with some level of data literacy. However, while the necessity for data literacy spans disciplines, best practices for teaching it do not. There are pockets of excellence in providing the knowledge, understanding, and skills each academic program has identified as important (such as data analysis in business or data collection in sociology), but there is no systematic approach to understanding how best to teach data literacy across programs, and no common standard for certifying data literacy. Finally, and perhaps most importantly, data literacy is not taught as a transferable skill; students learn how to work with data in their specialty, often in a research context, but are not cognizant of the broad applicability of such skills. These best practices from specific disciplines and for particular kinds of data literacy have not been converged into a transdisciplinary pedagogy. This research area investigates how we can best equip graduates with the knowledge, understanding, and skills required for the data-rich knowledge economy.

**Records management.** This is the practice of systematically controlling the creation, distribution, use, maintenance, and disposition of recorded information stored for business purposes [16].

One of the most valuable benefits of effective records management is the capacity (often the requirement) to decide which records no longer have value to your organization, based on legal and regulatory requirements, levels of use, or other factors. This process is difficult in unstructured, semi-structured, and variably-structured data environments. For example: what constitutes a "record"? A cognitive system will need to be self-records managing: identifying records that do not add value, or are outdated, and automatically pruning them; this will require tools and methodologies.

**Personal information stewardship.** The approach to personal data management has always been "entrust it to your provider". People are comfortable with this is some scenarios (e.g., health [12]). But this approach makes personal control difficult, and is not archival: customer data is routinely purged after a period of inactivity. What is the right approach to personal information stewardship? Are the companies using the data the best stewards? What access policies should be in place (see, for example, the approach used by Ontario, Canada with their Green Button initiative to provide individuals access to their own minute-by-minute utility records [18]).

**Internet of Me.** The Internet of Things has potential, but generally things that belong to others are of little interest to users. Their focus is on the things that can improve their own perceived quality of life. How do we get from the Internet of Misconfigured Incompatible Devices to the Internet of Things, and from there to the Internet of Me (or, less narcissistically, Internet of Us)? The theme of personalization has come and gone as a hot topic in the software research area, but this research agenda suggests resurgence is necessary.

**Interoperability**. Competing standards from competing manufacturers are common in nascent technologies, and wireless communication in smart homes is a good example of a crowded space. The format of data captured and stored by modern sensors (which range from smart home devices to telematics to smart dust) varies depending on the manufacturer. For a truly converged reality, these multiple data sources must be combined. When in the data lifecycle, and how, do we ensure data is interoperable?

**Data quality.** One of the challenges of cognitive computing is explaining how the system arrived at the answer it found. There are issues of trust at stake: one design goal is explainability. An important premise of cognitive computing is it can tolerate data of mixed quality, but how do we ensure, measure, and assert the quality and reliability of the information produced by a modern converged reality system? This is closely tied to issues of trust in increased automation.

**Data governance.** How do we manage, control, protect, test, and regulate data management systems that are handling data that is beyond human scale? When the right answer isn't known, it is difficult to assert that a data management system is operating correctly. What information is required to ensure these processes are sufficiently governed?

**Data modeling**. There is a great deal of collective wisdom captured in best practices for modeling data for relational databases, and increasingly for NoSQL databases as well. What data modeling techniques are relevant when there are more data items than ever before, but each individual data item is of less value? What new techniques are required? Which data models are required for cognitive assistants and other core technologies developed to support converted reality?

## 9. Conclusion

This paper provides a high-level view of the phenomenon we label converged reality, and of a data management research agenda. Several future research directions are suggested at a high level. The next step is to explore these directions in more detail with appropriate research communities from both academia and industry. Workshops and panels at carefully selected venues will allow us to actualize this agenda, and further explore and address the growing overlap among the enablers and dampeners.

## 10. References

[1] S. P. Anderson and A. De Palma. Competition for attention in the information (overload) age. *The RAND Journal of Economics 43(1),* pp. 1-25. 2012.

[2] d. boyd and E. Hargittai. Facebook privacy settings: Who cares? *First Monday 15(8),* 2010. Available: http://firstmonday.org/ojs/index.php/fm/article/view/3086/2589.

[3] L. Brandeis and S. Warren. The right to privacy. *Harvard Law Review IV(5),* pp. 193-220. 1890.

[4] Y. Brun, R. Desmarais, K. Geihs, M. Litoiu, A. Lopes, M. Shaw and M. Smit. "A design space for self-adaptive systems," in *Software Engineering for Self-Adaptive Systems II*, R. de Lemos, H. Giese, H. A. Müller and M. Shaw, Eds. 2013.

[5] R. Freeman, A. Steed and B. Zhou. Rapid scene modelling, registration and specification for mixed reality systems. In Proceedings of the ACM Symposium on Virtual Reality Software and Technology. 2005, DOI: 10.1145/1101616.1101647.

[6] (August 11, 2014). *Gartner's 2014 Hype Cycle for Emerging Technologies Maps the Journey to Digital Business*. Available: http://www.gartner.com/newsroom/id/2819918.

[7] B. Johnson. Privacy no longer a social norm, says facebook founder. *The Guardian ([online]),* 2010. Available: http://www.theguardian.com/technology/2010/jan/11/facebook-privacy.

[8] J. E. Kelly and S. Hamm. *Smart Machines : IBM's Watson and the Era of Cognitive Computing* 2013.

[9] J. O. Kephart and D. M. Chess. The vision of autonomic computing. *Computer 36(1),* pp. 41-50. 2003.

[10] Kyoung-Dae Kim and P. R. Kumar. Cyber-physical systems: A perspective at the centennial. *Proc IEEE 100(Special Centennial Issue),* pp. 1287-1308. 2012. DOI: 10.1109/JPROC.2012.2189792.

[11] J. Lifton and J. A. Paradiso. Dual reality: Merging the real and virtual. In First International Conference on Facets of Virtual Environments. 2010.

[12] Michael Smit, M. McAllister and J. Slonim. Privacy of electronic health records: Public opinion and practicalities. In Networking and Electronic Commerce Research Conference, Lake Garda, Italy. 2005.

[13] P. Milgram and F. Kishino. A taxonomy of mixed reality visual displays. *IEICE Trans. Inf. Syst. 77(12),* pp. 1321-1329. 1994.

[14] M. Mosley, M. H. Brackett, S. Earley and D. Henderson. *The DAMA Guide to the Data Management Body of Knowledge: (DAMA-DMBOK Guide)* 2010.

[15] (May 25, 2015). *Introducing Foodini, a 3D food printer*. Available: https://www.naturalmachines.com/.

[16] I. A. Penn, G. B. Pennix and J. Coulson. *Records Management Handbook*. 1994.

[17] A. Pentland. The data-driven society. *Scientific American 309(4),* pp. 78-83. 2013.

[18] (April 21, 2015). *Ontario's Green Button*. Available: http://greenbuttondata.ca/.

[19] R. R. Rajkumar, I. Lee, L. Sha and J. Stankovic. Cyber-physical systems: The next computing revolution. In Proceedings of the 47th Design Automation Conference. 2010.

[20] J. Rutter and G. Smith. Presenting the offline self in an everyday, online environment. In Identities in Action Conference, Gregynog. 1999.

[21] E. Stroulia, "Smart services across the real and virtual worlds," in *The Smart Internet*, M. Chignell, J. Cordy, J. Ng and Y. Yesha, Eds. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 178-196.

[22] (May 12, 2015). *Trakdot Luggage | Trakdot Luggage Tracker*. Available: https://www.trakdot.com/en/trakdot-luggage.

[23] (March 31, 2015). *Company | About*. Available: https://about.twitter.com/company.

[24] D. Vesset, C. W. Olofson, D. Schubmehl, B. McDonough, A. Woodward, C. Stires, M. Fleming, A. Nadkarni, A. Zaidi and M. Dialani. *IDC FutureScape: Worldwide Big Data and Analytics 2015 Predictions* 2014.

[25] C. Zhong, T. Wang, W. Zeng and S. M. Arisona. "Spatiotemporal visualisation: A survey and outlook," in *Digital Urban Modeling and Simulation*, S. M. Arisona, G. Aschwanden, J. Halatsch and P. Wonka, Eds. 2012.