A Conceptual Framework for Data Governance in IoT-enabled Digital IS Ecosystems

Avirup Dasgupta, Asif Gill and Farookh Hussain
School of Software, University of Technology Sydney, Australia

Keywords: IoT, Data Governance, Framework, Data Management.

Abstract: There is a growing interest in the use of Internet of Things (IoT) in information systems (IS). Data or information governance is a critical component of IoT enabled digital IS ecosystem. There is insufficient guidance available on how to effectively establish data governance for IoT enabled digital IS ecosystem. The introduction of new regulations related to privacy such as General Data Protection Regulation (GDPR) as well as existing regulations such as Health Insurance Portability and Accountability Act (HIPPA) has added complexity to this issue of data governance. This could possibly hinder the effective IoT adoption in healthcare digital IS ecosystem. This paper enhances the 4I framework, which is iteratively developed and updated using the design science research (DSR) method to address this pressing need for organizations to have a robust governance model to provide the coverage across the entire data lifecycle in IoT-enabled digital IS ecosystem. The 4I framework has four major phases: Identify, Insulate, Inspect and Improve. The application of this framework is demonstrated with the help of a Healthcare case study. It is anticipated that the proposed framework can help the practitioners to identify, insulate, inspect and improve governance of data in IoT enabled digital IS ecosystem.

1 INTRODUCTION

IoT is an emerging concept in the Healthcare industry with new applications and devices being manufactured using the Internet of Things (IoT). Sensor fitted wearable devices or implantable devices are increasingly being used to monitor the well-being of a patient. These devices automatically monitor health conditions, notify abnormal situations and propose protective actions such as informing doctors, family and friends (Karahoca et al., 2018, Gill et al., 2016).

The complexity of gathering, storing and processing data, has given rise to many data related problems in particular the governance of data. The General Data Protection Regulation (GDPR) law (Cha et al., 2018) has introduced additional aspects to this issue. Hence, it is vital that we understand the dynamics of data governance and related regulations in the IoT enabled digital IS ecosystem. This includes comprehending data ownership, the process of gathering consent before processing the data as well as understanding data lineage in the IoT enabled IS ecosystem. Thus, data governance issues pertaining to data security, data confidentiality, and data ownership stand as obstacles to the exchange of data among distributed IoT network and applications. Therefore, it is important that organizations address these challenges from a data governance (DG) perspective (Gartner, 2016a).

Data governance is not a new concept and is in use in the financial sector for more than two decades (Kontzer, 2006). However, in an IoT-enabled IS context, it is still at a nascent stage. The evolution of decentralized IoT architectures like Fog, Cloudlets and Edge implies that centralized approaches to governance are not viable (Gartner, 2016a). Several authors have highlighted the unethical use of data (Dastjerdi and Buyya, 2016), reprogramming device to function beyond its intended purpose and lack of network-intrusion-detection mechanism (security) as a major challenge associated with deploying IoT
based applications (Dhillon et al., 2016, Dasgupta and Gill, 2017). This paper focuses on the following research question (RQ): How to establish the data governance in digital IS ecosystem? In order to address the above question, this paper presents the application of the 4I framework in the Healthcare domain.

This paper is organized as follows. Section 2 discusses the data and IoT governance concepts. Section 3 discusses the research method. Section 4 presents the existing frameworks related to data management and governance in the context of IoT. Section 5 summaries the 4I framework. Section 6 demonstrates the applicability of the 4I framework with the help of a Healthcare case study before concluding the paper in section 7.

2 RESEARCH BACKGROUND

This section describes the key concepts of data in the context of IoT enabled digital IS ecosystem in order to provide the research background and context

2.1 Data Management

Data Management is concerned with the use of data to make good business decisions. It focuses on defining data, its storage, structure and data flow. The Data Management Association (DAMA), an association of technical and business data management professionals, defines data management as the development and execution of architectures, policies, procedures and practices to manage entire data lifecycle as well as planning, executing and managing the activities which acquire, control, protect, deliver and enrich data assets (Mosley et al., 2010, Stryk, 2015). Data Management Body of Knowledge (DAMA-DMBOK) identified 10 functions, which constitute Data management (Mosley et al., 2010) as shown in the figure below.

![Data Management Functions](adapted from (Mosley et al., 2010)).

2.2 Data Governance

As corporations recognize the importance of data and the challenges they face in integrating the data from various disparate source systems, an increasing number of companies have started exploring data governance. Data governance enables corporate-wide accountabilities and decision rights for data quality management (Weber et al., 2009) and is essential for the existence of an organization (Stryk, 2015). It is defined as an organizational approach to data and information management (Janne J. Korhonen et al., 2013) that formalizes a set of policies and procedures to include the full life cycle of data, from acquisition to use and to disposal. Gartner defines it as the procedure of setting decision rights and answerability for an asset, establishing policies aligned to business objectives, investing in assets that aid business objectives, establishing measures to ensure compliance to corporate policies, and ensuring adequate corporate risk management (Gartner, 2016b).

While governance refers to the decisions that are taken to ensure effective use and management of resources, management is focused on executing the decisions. Thus, management is influenced by governance (Ibrahim Alhassan, 2016). Data governance defines standards and procedures to ensure the proactive and effective handling and guidance of data management practices such as data replication, data archival, security, data backup, metadata management (MDM), data traceability and lineage, business glossary mapping, governance council, release and change management, master data and business (Infotech, 2016). Effective data governance results in profitable data use in an organization (Panian, 2010). With appropriate data governance, businesses can make insightful decisions by putting context to the data and transforming the information into knowledge and intelligence. This includes ensuring data has the necessary quality, availability, integrity and security throughout its lifecycle (Al-Ruitha et al., 2018).

2.3 IT Governance

IT governance is different from data governance and is concerned with the overseeing of IT resources such as computer networks, servers and applications through risk monitoring and control (Peterson, 2004) in alignment with the aims and strategies (Tallon et al., 2013) of an organization. Traditionally, financial assets and services (Gill et al., 2015) were administered using Governance, however, in last few
decades it has been extended to data and IT assets (Robert C. Rickards, 2012). There are several IT governance frameworks such as ISO 27001, Information Technology Infrastructure Library (ITIL) and Control Objectives for Information and Related Technology (COBIT) (Gehrmann, 2012). While ITIL was established to provide best practices for the IT services to its customers, COBIT framework supports governance of IT assets with a distinctive focus on ensuring IT procedures and activities align with the strategic goals of an enterprise (Egelstaff and Wells, 2013); (Juiz and Toomey, 2015).

2.4 IoT Governance

IoT governance is an extension to IT governance, where IoT governance is specifically focused on the lifecycle of IoT devices, data managed by the IoT solutions, and IoT applications in an organization's IT landscape (Gantait et al., 2018). IoT governance is can be considered a part of the existing IT governance landscape. It comprises of organizations such as Internet Engineering Task Force (IETF), Regional Internet Registry (RIRs), Information Security Operations Centre (ISOC), IEEE, The Internet Corporation for Assigned Names and Numbers (ICANN), Internet Governance Forum (IGF), and W3C and should leverage or tailor IT governance frameworks available to govern IoT (Virgilio A.F. Almeida 2015).

There is a need to have a clear distinction between the IT governance and data governance. While data governance deals with the data assets to improve business outcomes for business stakeholders, where IT governance is primarily focused on the IT assets. Further, these two concepts can be linked to strategy and enterprise architecture (Korhonen et al., 2016) in modern adaptive enterprises. These are two different but related concepts and thus there are some overlapping areas between IT and Data Governance as shown in table below. Thus, the scope of this paper is limited to data governance in IoT (a kind of IT) enabled IS.

### Table 1: Difference between IT and Data Governance (adapted from (Dimick, 2013)).

<table>
<thead>
<tr>
<th>IT GOVERNANCE</th>
<th>DATA GOVERNANCE</th>
</tr>
</thead>
<tbody>
<tr>
<td>IT driven led by (Chief Information Officer)</td>
<td>Business Driven</td>
</tr>
<tr>
<td>Oversee implementation of IT policy process and extract business benefits</td>
<td>Operational Focus</td>
</tr>
<tr>
<td>Policy and process ensuring effective evaluation, selection, prioritization, funding of competing IT assets and investment</td>
<td>Policy, process and practice that address accuracy, validity, completeness, timeliness, data integrity</td>
</tr>
</tbody>
</table>

3 RESEARCH METHOD

This research aims to address the data governance challenges in IoT and proposes the development and evaluation of the 4I framework using the Design Science Research (DSR) (Prat et al., 2014). DSR is problem focused (Kuechler and Vaishnavi, 2008) and seeks to design and evaluate an innovative product, or artefact, that provides a potential solution to a real-life problem within an organization as shown in figure 2 below. In DSR, the artefacts usually can be technical elements such as concepts, models, methods, frameworks or instantiations (March and Smith, 1995) as well as social elements such as humans, roles, work processes, teams, groups of organizations (Drechsler, 2015).

We applied the guidelines of DSR (Hevner and Chatterjee, 2010) to conduct this research. This paper focuses on the evaluation aspect of the proposed 4I framework.

4 RELATED IOT DATA MANAGEMENT AND GOVERNANCE FRAMEWORKS

The traditional data governance practices comprising of people, process and technology (Merkus, 2015) are going through a fundamental shift or transformation phase. This can be attributed to the changes in regulations (Wachter, 2018) as well as advancement in technologies such as Big Data, Blockchain, Cloud, IoT and Mobile (Copie et al., 2013). Thus, there is a need to tailor the data governance practices in an IoT context (Al-Ruithe et al., 2016), (Porambage et al., 2016), (IERC, 2015), (Banerjee and Sheth, 2017)), (IOTAlliance, 2017) in order to address IoT specific issues as indicated in figure 3 above. Few studies on
governance of data have been conducted in the domain of healthcare such as HeathFog in (Verma and Sood, 2018) and attribute based smart health in (Fuentes et al., 2018). In (Sajid and Abbas, 2016), authors discussed encryption based data privacy in cloud based healthcare systems. In another study (Banerjee and Sheth, 2017), the authors put forward an evaluation model to contextually evaluate the data quality based on two use cases.

To the best of our knowledge, currently there is no concrete IoT framework, which is available and can provide a blueprint to establish a data governance environment, particularly from a regulatory perspective. However, it is evident from the recent commercial works and analyst reports such as Gartner (Gartner, 2016a) and Frost and Sullivan (Sullivan, 2018), that there is an urgent need for more research and development in this area of governance.

5 THE 4I FRAMEWORK

This research developed a 4I framework (Dasgupta et al., 2019) (Identify, Insulate, Inspect and Improve) for managing and governing the data assets in the IoT-enabled Digital IS ecosystem. The 4I framework (Version 2) depicted in figure 4 is an updated version of the framework introduced in (Dasgupta et al., 2019).

It is based on the extensive review of existing data governance literature from academia and industry such as DAMA-DMBOK2 Framework (Mosley et al., 2010), The Data Governance Institute Data Governance Framework (Proença and Borbinha, 2016), The IBM Data Governance Council Maturity Model (A. Wröbel, 2017), The Gartner Enterprise Information Management Framework, EDM Council

![Figure 4: The 4I framework.](image-url)
Data management Capability Maturity Model (Council, 2015), Generic Framework (Al-Ruithe et al., 2016), and DGMM Framework (Merkus, 2015).

It is intended for use by data governance personnel as a guide to ensure appropriate data collection (“what to use”), processing (“how to use”), and retention (“until when to use”) mechanisms as well as significance (“why to use”) of data. It is composed of four stages or phases and explained in detail with the Fitbit case study in Section 6.

6 APPLICATION OF THE 4I FRAMEWORK: HEALTHCARE CASE STUDY

Wearables are the main fitness trend for 2019(Thompson, 2018) according to American College of Sports Medicine (ACSM). Wearables rely on the collection of the consumer’s private and personal data. Personal Identifiable Information (PII) can include First name, Last Name, Data of Birth, Address, Phone number, Financial and health of an IoT Fitbit user. This also constitutes sensitive personal information (SPI).

2) Identifying potential threats or risks around data management. For example, exploitation of security vulnerabilities to obtain user data is a common phenomenon. This is particularly important in Fitbit app context where apps running on Android are impacted by vulnerabilities from time to time(Linares-Vásquez et al., 2017), As a part of “identify” stage, Security advisories on vulnerability published periodically by Android can be documented and included in the patch management policy.

3) Classifying the sensitivity of data collected and determining the impacts of

- Sharing of health data publicly
- Sharing of data (health, PII, PCI and location) to a 3rd party such as medical providers(hospital, doctors), healthcare service provider, cloud or fog hosting service provider, network carrier
- Tracking of movement of individuals (including elderly patients) using motion sensors, camera, GPS tracker.
- Retention of data after customer stops using the device and its services
- Inferring customer’s traits based on data.

Figure 6: Healthcare data related Laws.

Figure 5: Wearable smart IoT-enabled ecosystem.

The “Fitbit” data is transmitted using Bluetooth technology to the consumer’s mobile or desktop application before it is transferred to the cloud. As shown in figure 5, the data exchange involves stakeholders such as App manufacturer, Cloud Providers, Health Service Provider and several systems to provide the end user with health service. From the Fitbit providers perspective, ensuring data is secured and compliant is of highest priority. Regrettably, the consumer has a lack of understanding of the risks (Skierka, 2018, Banerjee et al., 2018) linked with some of the wearable devices or products. For example, some wearable devices have default passwords that can be found on public websites and cannot be altered (Government, 2019).

In this section, we evaluate the applicability of the 4I and demonstrate how the 4I framework can be applied by the Fitbit service providers to avoid unethical usage of data.
4) Establishing policies related to data retention, data protection, patch management, Fitbit device procurement, data sharing and management techniques such as data anonymization, obfuscation are identified in this phase of the framework.

2nd I in the 4I Framework:
The second phase **Insulate** includes the preventive actions taken to mitigate the risks identified in the previous stage

Technology can be used to implement the data protection policies related to healthcare devices. This can include preventive measures such as ensuring software is patched to the current version in accordance with the patch management policy. It can also include implementing the data management processes formulated in “Insulate” phase. For example, an agent can be installed at wearable user’s gateway or mobile application to ensure that data is passed to Cloud only if

- Latest firmware version is present in the IoT devices.
- Intended address to push data matches the hardware endpoint requirements such as Host IP address or Mac address
- Explicit Consent is recorded from customer
- Encrypted data is sent
- 3rd party software used is patched and upgraded.

3rd I in the 4I Framework:
The **inspection** phase is a combination of sophisticated real-time monitoring, auditing and reporting as performed by the software agent.

- A robust asset management software can map each Fitbit device all the way to the database where each record of data is stored in database or application server.
- For each Fitbit, security information and event management (SIEM) agent can scan the data records stored in the files or databases. The agent can flag a risk through automated alerts to the data governance team in case it finds non-encrypted records or inappropriate data access.
- With respect to data stored beyond the data retention requirement, the agent can check if any PII data is stored in fileservers or database in unencrypted form and may take remedial actions.

4th I in the 4I Framework:
In the **Improve** phase, continuous enrichment of the process is done to ensure that the operational process is continuously monitored and enhanced. For example, conducting a Third Party Vendor assessment to check cloud storage vulnerability and updating the Vendor Selection or SLA policy can be an improvement and outcome of this final phase of the 4I framework. Additionally, non-technical changes such as reworking the contracts to pass liability of data breach to the third party can be another consequence of the improvement phase. In nutshell, this case of Fitbit demonstrates the applicability of the 4I framework for data governance in IoT-enabled Digital IS ecosystem.

7 CONCLUSIONS
The effective and informed governance of data in IoT-enabled applications is a complex undertaking. Currently, there is no holistic framework exists to address the important research question: how to ensure data governance in IoT-enabled Digital IS ecosystems? This paper discusses the newly developed 4I framework that can provide the Governance coverage across the data lifecycle in IoT-enabled Digital IS ecosystem. The 4I framework is developed through analysis and review of existing scientific and practice-oriented literature related to IT, Data and Enterprise Governance within the context of IoT and Digital IS ecosystem. Data stewards can use the proposed framework to manage and define enterprise-wide guidelines, company rules, and data assets to deliver the essential data governance and quality. The initial applicability of the proposed framework is demonstrated with the help of a healthcare case study. We intend to conduct further detailed studies to enhance the 4I framework.

ACKNOWLEDGEMENTS
This research is supported by the Australian Government Research Training Program Scholarship scheme.

REFERENCES


Gartner 2016a. Data Risks in the Internet of Things Demand Extensive Information Governance.


