Asia Pacific Journal of Information Systems
Vol. 34 No. 1 (March 2024), 292-320

ISSN 2288-5404 (Print) / ISSN 2288-6818 (Online)
https://doi.org/10.14329/apjis.2024.34.1.292

Construction of a Digitally Represented Person by Personal Data: A Multidimensional Framework from an Inforg Perspective

Jinyoung Min[a], HanByeol Stella Choi[b], Chanhee Kwak[c], Junyeong Lee[d,*]

[a] *Associate Professor, Department of Industrial Security, Chung-Ang University, Korea*
[b] *Assistant Professor, Department of Management Information Systems, Myongji University, Korea*
[c] *Assistant Professor, Department of Artificial Intelligence Convergence, Kangnam University, Korea*
[d] *Associate Professor, Department of Management Information Systems, Chungbuk National University, Korea*

A B S T R A C T

The amount of data a related to a person is so substantial that it appears that a digital version of them can be built thereon. They are usually handled as personal information, and the attempts made to understand personal information have led to bundling and unbundling of various data, yielding numerous fragmented categories of personal information. Therefore, we attempt to construct a generalizable lens for a deeper understanding of person-related data. We develop a theoretical framework that provides a fundamental method to understand these data as an entity of a digitally represented person based on literature review as well as the concepts of inforg and infosphere. The proposed framework suggests person-related data consist of three informational inforg dimensions that can preserve the archetype of a person, form, content, and interaction. Subsequently, the framework is examined and tested through several analyses in two different contexts: social media and online shopping mall. This framework demonstrates the suggested dimensions are interrelated with certain patterns, the prominent dimension can determine the data characteristics, and the dimensional composition of data types can imply the characteristics of the digitally represented person in certain contexts.

Keywords: Digitally Represented Person, Informational Dimension of Personal Information, Inforg

Ⅰ. Introduction

We, as human beings, live in and through the online environment, where we connect and communicate with other people and purchase products and services. Our names, educational backgrounds, career histories, and even bio-information are retained online. Moreover, our observations and movements are recorded, leaving digital traces of our activities. The data that we leave online with or without noticing

*Corresponding Author. E-mail: junyeong.lee@cbnu.ac.kr

are sufficient to construct an online version of ourselves. Although this digital world can serve us, it can be used to identify us and even make us perform certain actions. On Facebook, a person's activity data are used to connect with people who may share the same interests but also to make them see certain advertisements. Through the analysis of the purchase history of its customers, Amazon gains a sense of their preferences and recommends products to buy. This abundant digital information of a person online and the subsequent use thereof has led to significant controversial issues such as which digital data of a person can be related to that person.

To solve these issues, researchers have delved into personal information and its articulation. Personal information can mean any information regarding an identifiable person in both a direct and an indirect sense. According to Article 2 of the European Union Directive (95/46/EC), for example, "personal data shall mean any information relating to an identified or identifiable natural person ("data subject"); an identifiable person is the one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity" (p. 3). The National Institute of Standards and Technology of the U.S. Department of Commerce defines personal information as personally identifiable information that can be used on its own or with other information to identify, contact, or locate a single person or to identify an individual in context. Because personal information is a very broadly defined term, researchers have made further efforts to articulate personal information. These efforts have mainly focused on creating the category of information that is perceived as personal with rather narrowly focused views on its identifiability and privacy. Examples include privacy concerns and the related category of personal information disclosure (Phelps et al., 2000), the category of the extent of sensitive personal information that is included (Markos et al., 2018; Xie et al., 2006), and the extent of the identifiability of personal information (Faja and Trimi, 2006; Xu et al., 2014). These views inevitably limit the following discussion on how to use or not to use identifiable information while avoiding potential harm to a person's privacy: they are useful in terms of what they have classified personal information for, but have failed to understand the person-related information from a perspective that can capture the overall image of a person as a digital entity.

Information technology maintains the records of a vast range and quantity of personal information; thus, the set of information can be used to construct a person as a digital entity even though individual information in the set cannot be linked to a particular person (Polonetsky et al., 2016). The development of data collection and analytic methods has further increased this tendency by changing the set and meanings of information that can be connected and attributed to a person. Therefore, in this era of the complicated information world, person-related information should be understood beyond identifiability and privacy and needs to be approached in terms of how digital information constructs a digitally represented person as a digital entity. This is accompanied by the understanding of personal information not as data that are bounded to a person's physical reality, but rather as digital data that go beyond that boundary under virtual materiality (Palfrey and Gasser, 2011). Moreover, Parkinson et al. (2018) suggested that the individual's digital information is not atomic, but has multiple layers and multiple owners. This means that information related to a person may consist of multiple aspects from the perspective of data characteristics: one aspect of information can exhibit only a certain facet of a digital entity, whereas another or the combination of aspects can reveal its different facets. Furthermore, multiple information may differ from one another but play similar roles in functions because they share the characteristic that a certain aspect is stronger than the other aspects. Therefore, to understand a person as a digital entity, it is necessary not only to articulate the personal information, but also to determine those fundamental aspects that determine the characteristics of information related to a person.

Therefore, we use the inforg and infosphere, which are concepts from information philosophy (Floridi, 2007, 2014). In his philosophical concepts of the digital world, inforg means the entire set of information related to the specific person, and infosphere represents digital reality, an informationally built environment encompassing all information entities and their properties, interactions, and relations (Floridi, 1999). The notion of inforg and infosphere will be used to construct a framework that can provide a fundamental method for theoretically and practically understanding the person-related data as part of a digitally represented person. The framework introduces three inforg dimensions, namely form, content, and interaction. Our approach also provides a means of understanding the data of a person as a digital entity by preserving the innate nature of a person as an archetype, thereby aiding in its application to various contexts.

In the remainder of the paper, we first provide a literature review on how person-related data are understood and handled, and then present the theoretical background that forms the basis for our suggested framework. Thereafter, we describe the details of the framework developed from the theory and previous literature. This is followed by attempts to present the insights that the framework can offer through several data analyses in two different contexts: social media and an online shopping mall. First, a correlation analysis is conducted to investigate the dimensional associations. Subsequently, clustering on multidimensional mapping is performed to examine the characteristics of a digitally constructed person in the particular context. Thereafter, the visualization of the data types and their clusters on a two-dimensional map is presented to demonstrate the role of the framework in context comparison. The paper is concluded with a discussion and an outline of the practical and theoretical implications.

Ⅱ. Literature Review: Data that Construct a Person as a Digital Entity

In an effort to construct a framework, we first explored previous literature to understand how existing studies viewed and attempted to capture the characteristics of the digital data of a person (or personal information). Most studies have examined and categorized various types of personal data. The classification criteria reflect how those studies understood the basic data characteristics of a person. <Appendix A> summarizes this review.

First, the digital data of a person were studied in terms of the privacy concerns_perceived by a person and the subsequent information disclosure (Beldad et al., 2011; Faja and Trimi, 2006; Jin, 2013; Knijnenburg, 2018; Markos et al., 2018; Milne et al., 2017; Phelps et al., 2000; Shibchurn and Yan, 2015; Xie et al., 2006). These studies mainly focused on determining the types of personal information that people are willing or unwilling to disclose. <Appendix A> summarizes the categories of the information that the researchers classified, and which information categories are the most and least susceptible to disclosure. Their efforts demonstrated that information that cannot solely be used to identify a person, such as demographic (Phelps et al., 2000) or lifestyle and entertainment-related information, can be easily disclosed, whereas information that can single a person out, such as health data (Jin, 2013), is not readily disclosed. However, the willingness to disclose information is not based on the data types, but contingent on the situations for disclosure (Xie et al., 2006). For example, the data of daily lives can be considered not to be private and risky (Jin, 2013), but community interaction data, which are likely to occur within daily lives, are considered as private and risky to disclose (Milne et al., 2017). Another example is that, although anonymous, certain private-self related information may be more sensitive to disclosure than personally identifiable information (Markos et al., 2018). Moreover, data such as tastes and preferences that are considered as the least private may reveal embarrassing information about a user when gathered and aggregated (Knijnenburg, 2018). Overall, these examples and findings suggest that even though the data types are similar, the extent to which people feel safe regarding their disclosure varies depending on the circumstances and how they are used together. Therefore, the willingness to disclose data cannot be the fundamental characteristic used to understand the data that construct a person as a digital entity.

Second, the digital data of a person can be understood in terms of how well they can be used to identify a person (Chellappa and Sin, 2005; Faja and Trimi, 2006; Liu et al., 2005; Palfrey and Gasser, 2011; Polonetsky et al., 2016; Xu et al., 2014). The literature in this category has similar interests to the literature considering privacy concerns and information sensitivity, such as how comfortable users are in providing information, but focuses on how sensitively the system providers should handle the information when publishing it. The classification of data types in this research stream is determined by how strong their ability is to identify a person. There may be identifiable and non-identifiable information (Chellappa and Sin, 2005; Liu et al., 2005), according to the Federal Trade Commission. Identifiable information is information that "can be used to identify or locate an individual" (Federal Trade Commission, 2000, p. 9), whereas non-identifiable information is information that "taken alone (i.e., when not combined with other information), cannot be used to identify or locate an individual" (Federal Trade Commission, 2000, p. 46) and includes information that requires the use of sophisticated tracking technologies to identify an individual (Chellappa and Sin, 2005). Moreover, direct and indirect identifiers may exist (Polonetsky et al., 2016), and there may be multiple levels of identifiability of data types in accordance with privacy-preserving data publishing (PPDP) (Xu et al., 2014). PPDP was suggested for data collectors to be able to publish content while protecting the privacy of data providers (Fung et al., 2010). In these studies, similar insights to those of the first stream of the research were provided: the role and degree of the same data type for identifying a person may vary depending on how they are combined with other data types. Although this categorization is useful, it does not provide a fundamental understanding of which aspects of the data make certain data types identifiable and others not. Furthermore, information that is considered as non-identifiable information can easily be used to identify a person when combined with other data; therefore, it is necessary to delve deeper into the informational characteristics than simply categorizing the data, as the data type is the eventual realization and representation of these informational characteristic.

Taken together, the existing literature on personal information categories used narrowly tailored criteria and goals such as distinguishing identifiable information and privacy-sensitive information. These categories are useful for understanding the specific use of individual data for certain goals and contexts, but are difficult to use for understanding how the data represent a person in an overall sense and which aspects of the data are significant in achieving this. Therefore, an integrative framework is required that can embrace various views and provide a fundamental means of understanding the characteristics of personal data, while preserving the basic archetype of a person as a digital entity. In the following section, we introduce the theoretical bases for building the framework and explain the framework.

Ⅲ. Theoretical Bases of Framework: How Can a Person Be Represented in the Digital World?

3.1. Infosphere: Digital World We Live in

At present, information technologies, particularly information communication technologies, reconstruct our reality and lives with informational terms. People's careers are introduced as LinkedIn accounts. Their communications are recorded on chatting apps and social media as postings and replies. Thus, information technology transforms what it means by reality into the *infosphere* (Floridi, 2014). The infosphere can be understood minimally as "the whole informational environment constituted by all informational entities, their properties, interactions, processes, and mutual relations" (p. 41) and maximally as "a concept that can also be used as synonymous with reality, once we interpret the latter informationally" (Floridi, 2014, p. 41). Floridi (2014) suggested that, in this infosphere in which entities and agents are all informational, and thus, no physical differences exist among them, *processors* and *processed*, and their *interactions* are all equally informational. He also argued that the "infosphere will not be a virtual environment supported by material world. Rather it will be the world itself that will be increasingly understood informationally" (p. 49). Therefore, the infosphere can be synonymous with reality itself because our everyday life is becoming more of an informational one compared to the physical and material one that it used to be. In this world, stream of activities including processes, operations, and events are treated as information processes that generates a trail of information (Floridi, 1999). The inforg and infosphere concepts of Floridi have been applied to various domains. For examples, infosphere quality can be used for determining how well public health strategy is connected to solving threats and inequalities in public healthcare (Morley et al., 2020) and museums now have blurred online and offline boundary of exhibit by using technology, being infosphere of presenting reality informationally (Simone et al., 2021).

3.2. Inforg: A Digitally Represented Person

People are spending a substantial amount of time in the infosphere, digitally interacting with one another. They are the informational agent that also becomes an entity that produces information, becomes part of the infosphere, and affects the infosphere as it is the totality of those entities (Floridi, 1999, 2002a). In this infosphere, personal identities and communications are shaped by information technology. Others can infer who we are according to our postings on social media, and purchase history in online shopping malls. We comment on and reply to others' social media postings, receive product recommendations from others' purchases, and constantly communicate through digital means. Our identities, activities, and interactions have informational traces. As informational agents, we process our values, attitudes, likes, and dislikes in the infosphere, constructing the subjective informational environment (Floridi, 2002b). In this sense, we are informational organisms that are mutually connected and embedded in the infosphere, which are known as *inforgs* (Floridi, 2007, 2014). The notion of inforgs has arisen within the transformation of our material environment and physical agents into informational ones. Floridi (2014) argued that inforgs can be *de-individualized*, becoming "a kind of," and *reidentified*, being viewed as a specific crossing point of many "kinds of," and thus, can also be treated as commodities that may be the subject of a transaction on advertisement markets, as is now done on Facebook. Therefore, an inforg is the entire set of information related to the specific person that inhabits the infosphere. The definition provided by Micheli et al. (2018) of digital footprints— "the aggregate of data derived from the *digitally traceable behavior* and *online presence* associated with an individual" (p. 2)—suggests a similar idea. The concept of inforg can be further developed to datafication of humans as being represented as information can mean being fully integrated into the informational environment, the infosphere (Baelo-Allué, 2022). Also, Russo (2018) suggested that technology changes how we relate the infosphere and inforgs as machines can become other inforgs, and technologies can influence how we interpret reality. Furthermore, because the information that Jones (2010) described in his study of personal information does not include a person within itself, the concept of the inforg also implies that information that can be contained under the term inforg not only holds a person's own production of content, but also the result of interactions with other agents.

According to this theoretical basis and the findings from the literature review, we present the framework in the following section.

Ⅳ. Suggested Framework for Understanding Digital Data of a Person

Based on the concepts of the infosphere and inforg, we suggest the construction of the informational representation of a person as a digital entity. Because the concept of the inforg focuses on information organisms, and their connections and interactions in the infosphere, it embeds the idea that inforgs are distinguished from one another, act like organisms, and interact with one another and the environment (Floridi, 2014). This suggests that when a person is informationally represented and understood as an inforg, an inforg can be composed of the information that can distinguish it from the other inforgs, capture the unique containment of an inforg, and show how an inforg is communicating with its environment and other inforgs. The first type of information may be rather fixed characteristics that can define and identify a person in the real world, including his or her physical and material representation. This type of information constructs a form of an inforg that distinguishes one inforg from another, as the physical and genetic elements of an organism distinguish it from another. The second type may be unique and indeterminate thoughts and activities that are rather unfixed and change. Although the first type of information can exist without a person's efforts, as is the case with many physical characteristics, the second type of information cannot be manifested without a person's efforts and behaviors, as these are required for most coded thoughts and recorded activities. The third type of information indicates a person's connection with other people and the context within which the information resides. When it is represented informationally, it takes the form of linkages between inforgs. According to this inforg concept, we suggest a framework that views and constructs a digitally represented person as an inforg with three dimensions: *form*, *content*, and *interaction*. The following section describes the three inforg dimensions.

4.1. Three Inforg Dimensions that Construct A Person as A Digital Entity

4.1.1. Form Dimension of An Inforg

The form dimension distinguishes one inforg from the others similar to distinguishing one organism from other organisms, although it is done informationally. Therefore, the identifiable information of processors in Floridi (2014) can be revealed in this dimension. When physical agents are transformed into informational ones, the form dimension holds the basic information of the physical agents; thus, it mainly operates as identifiable information. The information categories that are classified as identifiers (Polonetsky et al., 2016; Xie et al., 2006), (personally) identifiable information (Faja and Trimi, 2006; Liu et al., 2005; Markos et al., 2018), or secure identifies (Chellappa and Sin, 2005; Phelps et al., 2000) and certain demographic information (Beldad et al., 2011; Knijnenburg, 2018; Milne et al., 2017; Shibchurn and Yan, 2015) reflect the concept of the form dimension. Furthermore, in the schema of Rumbold and Pierscionek (2018) for the articulated categorization of personal data, demographics, apparent human characteristics, and medical status and healthcare, the characteristics of an individual regarded may be the cases that form dimensional information operates strongly. Moreover, the identifiers and quasi-identifiers of Xu et al. (2014) exhibit a strong association with the form dimension, because these types of information describe the physical characteristics of an agent.

4.1.2. Content Dimension of An Inforg

The content dimension is what is held and created as the result of processing the information inside of the informational organisms without clear interaction with the infosphere, as is the case with the processing mechanisms of biological organisms that operate internally. Therefore, the direction of the created information is towards the inside but not the outside of an inforg. For example, the information categories classified as personal preferences (Milne et al., 2017), daily life and entertainment (Jin, 2013), preferences including books read and hobbies (Xie et al., 2006), purchase-related records (Markos et al., 2018; Phelps et al., 2000), tastes (Knijnenburg, 2018), interests, views and opinions (Shibchurn and Yan, 2015), and lifestyle interests and media habits (Phelps et al., 2000) are data types in which the contents of dimensional information are magnified.

4.1.3 Interaction Dimension of An Inforg

The interaction dimension refers to the informational linkage with the infosphere, including other inforgs. Therefore, it is not only related to the inter-

<Table 1> Suggested Framework for Understanding Digital Data of A Person

Dimension	Conceptual Definition	Example Data Types from the Previous Literature
Form	The information that distinguishes one inforg from the others	identifiers (Polonetsky et al., 2016; Xie et al., 2006), (personally) identifiable information (Faja and Trimi, 2006; Liu et al., 2005; Markos et al., 2018), secure identifies (Chellappa and Sin, 2005; Phelps et al., 2000)
Content	Information created as the result of processing the information inside of the informational organisms without clear interaction with the infosphere	personal preferences (Milne et al., 2017), daily life and entertainment (Jin, 2013), tastes (Knijnenburg, 2018), interests, views and opinions (Shibchurn and Yan, 2015)
Interaction	The informational linkage with the infosphere, including other inforgs	Human-machine interaction category (Rumbold and Pierscionek, 2018), the context data (Knijnenburg, 2018), community interaction category (Milne et al., 2017)

action between the processors and processed but also includes the ones among processors and the ones among processed in the inforg and infosphere terms of Floridi (2014) . There is a direction within the processed information to or from the infosphere, which may be the person's replies and comments to other agents' postings on social media as well as the likes and replies from other agents on a person's postings. The information from the inforg to the infosphere usually includes the inforg actions to create linkages to the infosphere, but information from the infosphere to the inforg places an inforg in a passive position to be linked to the infosphere, as it is the result of other inforgs' activities. Examples include the human-machine interaction category of Rumbold and Pierscionek (2018), the context data of Knijnenburg (2018) that indicate a user's interaction with other users, and the community interaction category of Milne et al. (2017). The definition provided by Casemajor et al. (2015) of "passive participation," which is "engaging in a platform while being subject to processes of decision that happen outside of one's control" (p. 856), presents a similar concept. Micheli et al. (2018) also suggested that a person's digital footprint can be generated by other internet users. <Table 1> summarizes the conceptual definition of the three suggested dimensions and data type examples from the previous literature. <Appendix A> also shows how the category of data type in the previous literature can be explicated by the three dimensions of inforg framework.

Ⅴ. Applications of the Framework: Understanding Inforgs in Various Infospheres

The aim of our framework is to provide a better sense of the digitally represented person in a digitally constructed environment. Therefore, we attempted to apply it to several contexts and to determine the characteristics of the inforg, both in its own created setting and in comparison to inforgs in different contexts. For this purpose, we selected social media and an online shopping mall as the target contexts, as they can capture various aspects of people's digital lives, and thus, can provide a sense of the digitally represented person according to various angles. These two contexts vary in terms of how a person is revealed and handled in the context. First, social media bases its business on people and the various information types that they provide, and thus, social media basically presents a person as information and connects people. In this type of context, we can capture the highest degree of an inforg, the digitally represented person, and the interaction between them. Second, the primary interest of an online shopping mall business is not people per se, but transactions. However, online malls also gather various information types about a person as this information can help increase transactions, and thus, although not to the same extent as on social media, an inforg can be captured. Furthermore, the information characteristics in an online shopping mall differ from those in social media because the information is based more on materialistic activities; thus, different aspects of an inforg from the ones in social media can be captured.

To apply the inforg framework to contexts, we identified possible data types for each context. We created survey questionnaires that investigated the extent to which each data type is associated with the three inforg dimensions of the framework for the two different contexts. The participants were asked to rate how strongly each data type can be associated with the criteria provided. The form, content, and interaction dimensions were explained as follows: "the information of the *form* criterion consists of a person's real-world identity that makes the person distinguishable from others," "the information of the *content* criterion is contained inside of a person or a person's activities that can show his or her individuality such as thoughts, opinions, experiences, preferences, and personal activities," and "the information of the *interaction* criterion is linked to others and the environment that can show how he or she is associated with what," respectively. The answers were measured using a seven-point Likert scale ranging from "strongly disagree" to "strongly agree." (See <Appendix B> for sample question) They were then distributed through Amazon MTurk.

We collected responses from 203 for social media users and 202 online shopping mall users. We eliminated responses of 20 social media respondents and 16 online shopping mall respondents because their response times were less than 5 minutes, which was considered unfeasible for answering all the questions. In result, we used 183 and 186 responses for social media and online shopping mall, respectively, for further analysis. For respondents' demographic, 117 males and 66 females responded for social media context and 123 males and 63 females responded for online shopping mall context. The mean ages of respondents were 33.81 for social media respondents and 37.16 for online shopping mall respondents.

We conducted multiple analyses to provide insights into the inforg characteristics dimensions. First, to examine the suggested framework within multiple contexts, we calculated the correlations between dimensions to demonstrate how each dimension is associated with the other dimensions. Second, to apply the framework to the context and to determine how it can provide insights for understanding the characteristics of a digitally constructed person in the particular context, we visualized the data types and their clusters on a two-dimensional map. Third, to demonstrate how the framework can capture the differences in the contexts for constructing a digitally represented person, the results of the cluster visualization were compared and discussed.

5.1. Dimensional Associations of Framework

To investigate the associations among the dimensions, we calculated their correlations using the mean values of each data type. <Table 2> displays the correlation matrix along with the mean value and standard deviation of each dimension.

<Table 2> Correlation Matrix of Dimensions in Different Contexts

		Mean	S.D.	Form	Content
Social Media (n = 34)	Form	4.963	0.330		
	Content	4.842	0.350	-0.468**	
	Interaction	4.843	0.314	-0.683***	0.734***
Online Shopping Mall (n = 27)	Form	4.807	0.391		
	Content	4.807	0.359	-0.739***	
	Interaction	4.799	0.355	-0.784***	0.864***

The results for the social media context indicate that all three inforg dimensions were correlated with one another. However, their relationships were not always positive. The increased value of the form dimension is likely to be associated with the decrease in not only the content and interaction dimensions. As the form dimensions are linked with the identifiable information of physical agents, this reverse association suggests that the increase in the content or interaction dimensional information (or the combination of both) is likely to be related to the decrease in the form dimensional information. If a data type is designed to include a higher level of content and interaction dimensional information, it is unlikely to be designed to single out a person. This is in line with previous findings, whereby entertainment and lifestyle-related information was regarded as disclosable information (Jin, 2013; Phelps et al., 2000). Furthermore, the content dimension is strongly associated with the interaction dimension, suggesting that data types that are designed to include content dimensional information are likely to include interaction dimensional information, because such information is likely to form linkages to and from the content, and tends to exist as an embedded form.

The results can be explained by the characteristics of social media. Social media entice their users to provide information by offering communication, relationship management, and a self-presentation environment (Min and Kim, 2015). For this purpose, the use of social media starts with building a user profile, which operates as a basic unit to distinguish each user, and thus, it inevitably includes form dimensional information. From this starting point, users generate content and interact with other users; therefore, it is necessary for the related data types in social media to include content and interaction dimensional information heavily, and these are also likely to change over time. In this sense, high form dimensional information, which may include the gender, address, and mobile numbers, can be located low in the content and interaction dimensions, and high content and interaction dimensional information such as posts and replies can reveal low form dimensional information. Furthermore, the user profile, which can distinguish users, operates as an introduction of a self to other users and is created at the early stage of the use; therefore, most information in this profile is filled by social media users themselves, and is associated with the form dimension.

The online shopping mall context exhibited similar associations among the dimensions. This can be attributed to the manner in which personal data are collected and treated in online shopping malls. In online shopping malls, the various personal data types are gathered based on different purposes for data usage. For example, personal information that is required from service providers (e.g., credit card information and credit history) is collected for transaction purposes (Robinson, 2017), and the traces that are produced when consumers visit each page on the website (e.g., click data and cookies) are collected to provide a personalized shopping experience or product recommendations, whereas consumer data provided on a review (e.g., name, age, and review content) are disclosed for communication and sharing experiences with other reviewers. Likewise, consumers in online shopping malls disclose personal information to service providers in exchange for a better shopping experience (Chellappa and Sin, 2005) and publicly divulge information to communicate with other consumers (Chen and Xie, 2008). Therefore, personal data in online shopping malls represents all three dimensions of an inforg in interrelated way.

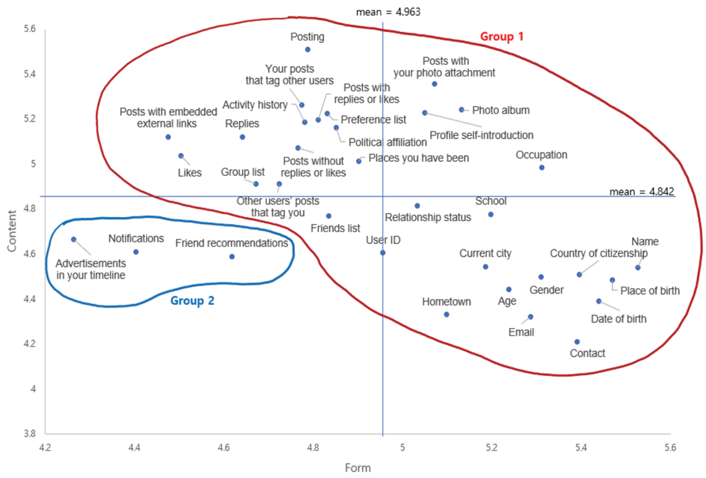
5.2. Clustering Data Types by Multidimensional Associations in Contexts

To demonstrate and interpret these multiple associations better, we visualized the results through two-dimensional mapping. Although the data type mapping of three inforg dimensions would be a form of a map on the three axes, it is difficult to view, understand, and interpret. Therefore, we used two-dimensional mapping schemes by drawing three scatter charts for each pair of inforg dimensions. Subsequently, we applied clustering algorithms to each mapping to understand the characteristics of inforgs better. We used density-based spatial clustering of applications with noise (DBSCAN) as the clustering algorithm. DBSCAN is an unsupervised classification algorithm that does not require the number of groups for clustering (Ester et al., 1996). DBSCAN requires two parameters, namely the starting distance of a group and the minimum number of samples that should be included in a group. To conduct conservative clustering, we selected the value of the former parameter to be directly below the value of bundling all data points as a single group at two decimal places, and the latter parameter as two data samples. Owing to this exploratory characteristic of DBSCAN, the DBSCAN result was provided not for a firm classification of the data types, but to aid in understanding the construction of an inforg in each context through the locational mapping and the presentation of the possible emergence of clusters.

5.2.1. Social Media Context

The 34 data types were located on the two-dimensional charts. The mean of each dimension was drawn to indicate the relatively high and low locations of the emerged groups. <Figure 1> depicts the associations of the data types in terms of the form and content dimensions. Two groups emerged through the DBSCAN clustering. Group 1 contains more than 90% of all data (n = 31, 91.2%). These are mostly data created by essential social media activities. Group 2 on the bottom left (n = 3, 8.8%) falls into the category of relatively low form and low content data types. These are notifications and advertisement in a user's timeline and friend recommendations, which are generated by social media platforms. The data types in such group may indicate that people perceive those data types as less useful for distinguishing a person from others as they are generated by a platform and do not provide distinguishable and deep information about a user.

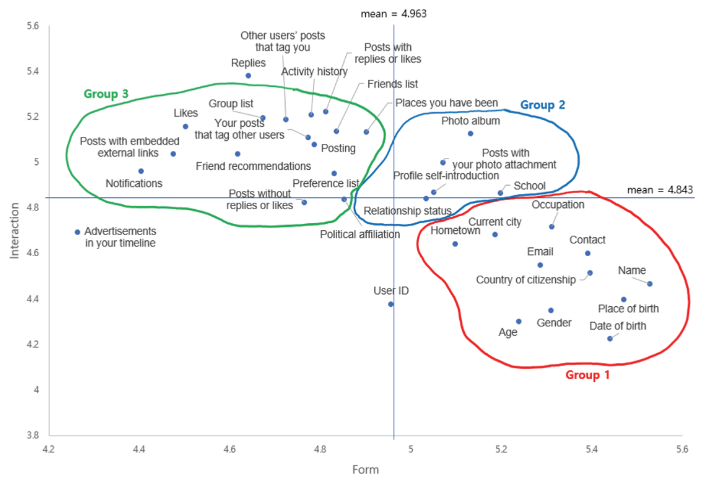
<Figure 2> presents the associations of the data types in terms of the form and interaction dimensions. Three groups emerged: Group 1 on the right can be characterized as relatively high form and somewhat



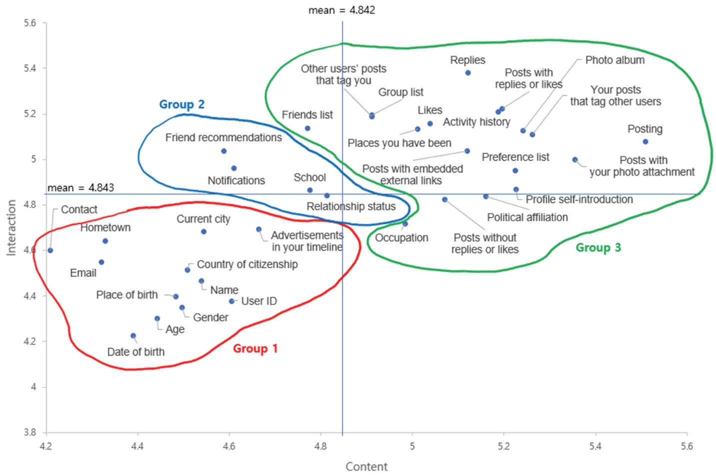
<Figure 1> Form and Content Dimensions in Social Media Context

low to medium interaction (n = 11, 32.4%), Group 2 in the middle is considered as relatively medium form and high interaction (n = 5, 14.7%), and Group 3 on the left is considered as relatively low form and high interaction (n = 15, 44.1%). The advertisements, replies, and user ID data types are not included in any groups. The data sets of Groups 1, 2, and 3 are likely to be related to the physical identity of a person, the activities through which the user identity is likely to be disclosed, and the various relational activities, respectively.

<Figure 3> presents the associations of the data types in terms of the content and interaction dimensions. Three groups were identified. Group 1 on the left can be characterized as relatively low content and low interaction (n = 12, 35.3%), Group 2 in the middle exhibits the characteristics of relatively low content and high interaction (n = 4, 11.8%), and Group 3 on the right can be considered as relatively high content and high interaction (n = 18,



<Figure 2> Form and Interaction Dimensions in Social Media Context



<Figure 3> Content and Interaction Dimensions in Social Media Context

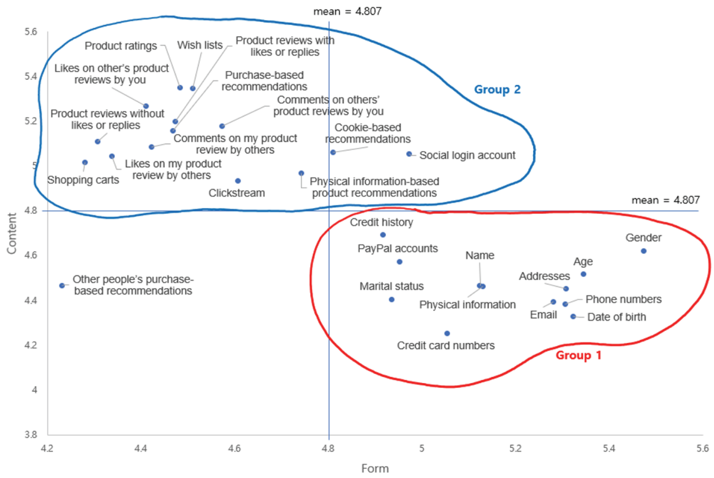
52.9%). The data set of Group 1 appears to have lower values for the content and interaction related information than that of Group 3.

The three two-dimensional maps reveal four noticeable characteristics of how social media data are characterized by the inforg dimensions. First, the information requested by social media platforms for personal profiles exhibit mostly high form but low content and interaction aspects, suggesting that such data types are strongly associated with the real-world identity of a person. However, although many posting-related data types are located high in the content dimension, several data types such as school and relationship status information exhibit relatively medium content and interaction characteristics, and medium-high form characteristics. This implies that there are data types that reflect either a person's physical identity or personal characteristics, but certain data types become the bridge between these two, projecting one's social identity and activities.

Second, regarding the form dimension, most data types are grouped together, although data types such as posting with a user's photo attachment and profile self-introduction, and data types such as likes and the group list exhibit differences in the form dimensional value, suggesting that the latter can also be used to identify a user, as the former can. Interestingly, each type of posting is located differently in the form dimension depending on what is embedded therein, suggesting that posting is inherently designed to include the multidimensional characteristics of an inforg.

Third, high content and interaction characteristics do not always originate from lengthy information. For example, likes are categorized as a relatively high content group, but liking is a simple and effortless activity that is represented only as several digits or emoticons. Therefore, the high content data appear to be obtained not from the quantity but from the possibility of various extractions and interpretations of information from the presented data.

Fourth, friend recommendations and notifications are characterized as low form, low content, and high interaction information. These constitute information that is generated by the social media platform. They can be considered as not being eligible to identify a person and provide little content but offer connection information to other inforgs.

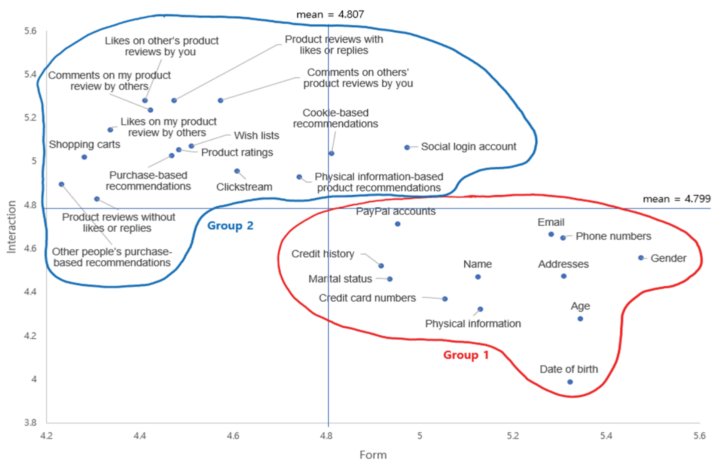


<Figure 4> Form and Content Dimensions in Online Shopping Context

5.2.2. Online Shopping Context

Regarding the form and content dimensions, the clustering results demonstrate that two groups emerged among the 27 data types: Group 1, the relatively high form and low content group (n = 14, 51.9%), and Group 2, the low form and high content group (n = 12, 44.4%), as illustrated in <Figure 4>. The data types in Group 1 are the basic information that are requested for the shopping itself and for a better shopping experience. Group 2 contains data types that are associated with the activities within online shopping malls, such as product recommendations or reviews.

<Figure 5> presents the two-dimensional representation of the form and interaction dimensions. Group 1, which is the relatively high form and low interaction group on the bottom right, includes basic information for shopping (n = 12, 44.4%) and Group 2, which is the relatively low to medium form and



<Figure 5> Form and Interaction Dimensions in Online Shopping Context

high interaction group on the top left, consists of shopping experience-related information (n = 15, 55.6%). Interestingly, the relative location of the groups on the chart is similar to that of the groups in the form and content dimensional chart.

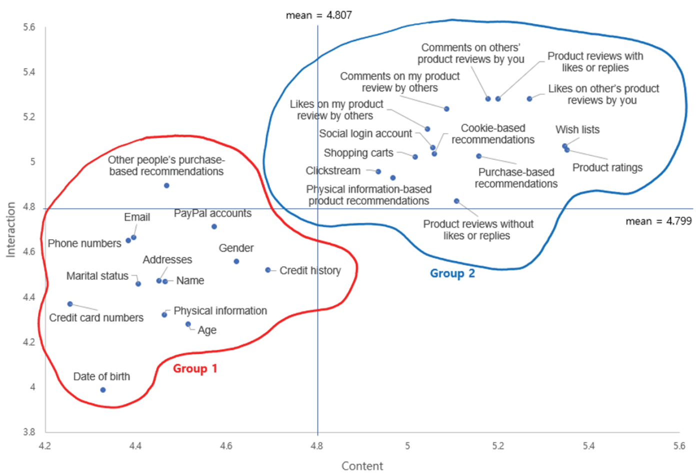
<Figure 6> indicates that basic customer information is clustered as Group 1 on the left, which can be characterized as relatively low content and low and medium interaction (n = 13, 48.1%), whereas the shopping experience and activity on a website related information is grouped as Group 2 on the right, which exhibits high content and high interaction characteristics (n = 14, 51.9%).

Through each clustering analysis of the inforg dimensions, the data types were clustered as two groups. The data types that were grouped together tended to be clustered again in the different dimensional combinations with an exception: other people's purchase-based recommendations had a different position in each clustering result. The group with basic customer information exhibits relatively high form, low content, and low interaction, whereas the group with shopping activity exhibits relatively low form, high content, and high interaction. These groups are clearly different in terms of the role (managing the shopping itself vs. helping indecision making), the place of creation (real world vs. shopping platform), and the complexity of the included information (simple and consistent vs. multiple layers of information with possible connections to other information).

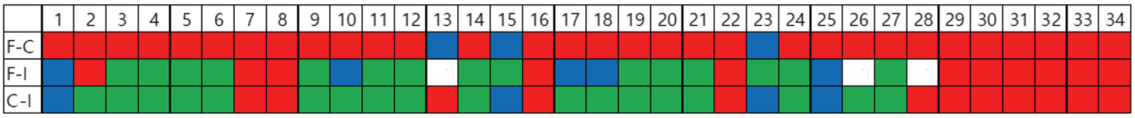
5.3. Comparing Clusters in Each Context from Inforg Perspective

To help understand how an inforg is characterized by the data types and their dimensional clusters in each context, we illustrate compact summarizations of the two-dimensional representations in <Figure 7> and <Figure 8>. Each box represents a data type. The color of the box indicates the group with the same color on the corresponding chart above. The three rows indicate the clustered groups in the two-dimensional clusters.

As illustrated in <Figure 7> and <Figure 8>, inforgs exhibit different characteristics in different contexts.



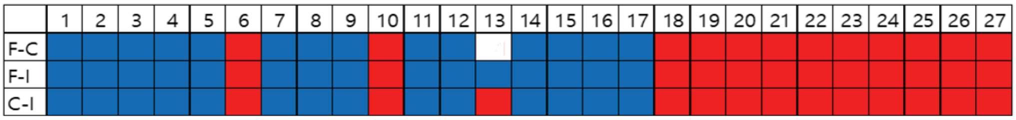
<Figure 6> Content and Interaction Dimensions in Online Shopping Context



Note: * F-C, F-I, and C-I represent the two associated dimensions, which are the form-content, form-interaction, and content-interaction dimensions, respectively.

** The numbers above indicate the data types, as follows: 1) Relationship status, 2) Occupation, 3) Posting, 4) Preference list, 5) Likes, 6) Friends list, 7) Current city, 8) Contact, 9) Posts without replies or likes, 10) Photo album, 11) Your posts that tag other users, 12) Posts with replies or likes, 13) Advertisements in your timeline, 14) Group list, 15) Friend recommendations, 16) Email, 17) Profile self-introduction, 18) Posts with your photo attachment, 19) Political affiliation, 20) Other users' posts that tag you, 21) Posts with embedded external links, 22) Age, 23) Notifications, 24) Places you have been, 25) School, 26) Replies, 27) Activity history, 28) User ID, 29) Country of citizenship, 30) Gender, 31) Hometown, 32) Name, 33) Date of birth, and 34) Place of birth.

<Figure 7> Clusters of Each Data Type for Inforg Dimensions in Social Media Context



Note: * F-C, F-I, and C-I represent the two associated dimensions, which are the form-content, form-interaction, and content-interaction dimensions, respectively.

** The numbers above indicate the data types, as follows: 1) Wish lists, 2) Shopping carts, 3) Purchase-based recommendations, 4) Likes on my product review by others, 5) Likes on others' product reviews by you, 6) Email, 7) Product ratings, 8) Comments on my product review by others, 9) Social login account, 10) Phone numbers, 11) Product reviews with likes or replies, 12) Clickstream, 13) Other people's purchase-based recommendations, 14) Comments on others' product reviews by you, 15) Cookie-based recommendations, 16) Product reviews without likes or replies, 17) Physical information-based product recommendations, 18) Marital status, 19) Addresses, 20) Physical information, 21) Credit history, 22) Credit card numbers, 23) Age, 24) PayPal accounts, 25) Name, 26) Gender, and 27) Date of birth.

<Figure 8> Clusters of Each Data Type for Inforg Dimensions in Online Shopping Mall Context

The data types in the online shopping context appear to be rather explicitly divided into two different groups: relatively high form, low content, and low to medium interaction (red-colored groups) versus relatively low to medium form, high content, and high interaction (blue-colored groups). This means that the data types that construct an inforg in this context are either identity-related data that may single out a person but do not provide rich information on what that person is like, and his/her association with the outside and others, or data that include rich information and a linkage to the outside of an inforg but are difficult to use to single out a person in the real world. Moreover, the data types clustered together tend to flock together consistently, even according to the different dimensional criteria, which means that data types that are considered to be similar regarding certain characteristic tend to be similar for other characteristics as well. This may suggest that the role and content of each data type in constructing an inforg are rather fixed, and the characteristics of each constructed inforg are not likely to be substantially different from those of other inforgs in that context.

The data types in social media exhibit different characteristics. Data types that are grouped together appear to share certain dimensional characteristics, but when they are examined from a different view, they no longer tend to share the same characteristics. This may suggest that there are data types that may not simply be classified by the prevalent categories. Their existence implies that the context is complex, with various types of information, the actual containment of which may be very different from person to person, so that each inforg can be constructed significantly differently from other inforgs. Furthermore, those data types can become a bridge between the groups of data types that are rather fixed in their role of constructing an inforg, and thus, can make an inforg change and evolve over time. Examples of such data types in the social media context are photo albums, advertisements in the timeline, posts with a photo attachment, profile self-introductions, and replies. These are designed to contain abundant information that can accommodate various user needs. The context with these types of inforg characteristics can provide a very rich picture of the digitally represented person.

Ⅵ. Discussion

The results indicate that, because online activity is complex, even single data point is associated with multiple characteristics of the digitally represented person. Therefore, instead of classifying data types of personal information as separate categories, the dynamics of multiple characteristics that consist of each data type can be useful in understanding a person as a digital entity. Our framework provides useful insights that previous studies have not offered.

Our framework provides an integrative means of understanding the data that are related to a person. From the perspective of the archetype of a digitally represented person, data with strong form dimensional information tend to have low content and interaction dimensional information. That is, prominent processor information does not need to have processed or interaction information to construct a digitally represented person. The content and interaction dimensional information exists in and around the form dimensional information; thus, it is not the basic block of constructing an inforg. However, information with a high value for processed or interaction information causes the data set to be not simply a bundle of digital information, but a digital entity that embraces an extended and even evolved persona. Moreover, the interaction dimensional information operates as a link among nodes in a networked world, making the digital entity not only an independent and isolated persona, but also a connected social being. With the hints of these informational connections, we can gain a sense of the infosphere shape.

In more detail, the form dimension plays an important role in both constructing an inforg and in identifying a person. The form dimension defines the processor, thereby clarifying the boundary of a digital entity. Furthermore, its identifiability is expected because this dimension is closely linked to the physical identity of a person. The form dimension combined with the content and interaction dimensions also provides new insights. For example, in the social media context, the friends list is perceived to represent a person more than certain demographic information and even profile information. This suggests that even with low form information, high content and interactional information can operate as identifiable information, particularly in the context in which the network connection is enriched. Friends lists in social media and cookie-based recommendations in an online shopping mall are other examples. However, certain demographic information such as gender and age have previously been considered as easily disclosable information. The suggested framework explains the underlying reasoning: although information is highly associated with a person's identity, when it has low content and low interactional information, the value of its identifiability is decreased. By considering these factors together, our framework explains the findings of the previous literature. People are likely to disclose their daily life and entertainment related information (Jin, 2013; Phelps et al., 2000) because these are located low in the form dimension but high in the content and interaction dimensions. Thus, the likelihood of information disclosure depends on how high it is in their form dimension. However, the extent of the identifiability with the information can also be determined by the content and interactional characteristics of the information.

Moreover, the comparison of the inforg dimensional construction in different contexts can provide an understanding of how the representation of a person is built in that context: the role and characteristics of an inforg, which is the digitally represented person, are rather fixed in certain contexts. This means that we can make a limited and fixed conjecture of a person in that context.

6.1. Theoretical Implications

This study provides several theoretical implications. First, the suggested framework offers an integrative framework for understanding data as part of constructing the digitally represented person. Previous studies narrowly focused on certain aspects of the data, such as the identifiability and willingness of disclosure. In this manner, the data can be understood at the individual data level, and certain data can be considered as more important than others. With the suggested inforg framework that understands each data type as an element of building a person in online space, each data type has its own values and meanings in various manners. Considering the various roles of the data, we can extend the approach to use, treat, and protect data.

Second, the inforg framework provides a theoretical background for projecting an archetype of a person onto the data while preserving the role of each data type in that archetype. For example, the form dimension that represents the physical characteristics of a person in the digital reflection can operate as identifiable dimensional information. The content dimension that can capture the inside of a person can reflect the information that constitutes that person. The interaction dimension that contains a person's connection with the outer environment suggests that we can expand our understanding of a person by linking the data sets of other people and multiple environments. Therefore, the inforg framework offers a clear means of linking data to human beings.

Third, according to the inforg perspective, we can understand the particular manner in which the data set in a certain context reflects a person. For example, the data in the online shopping mall context is rather clearly divided into the data with strong form dimensional characteristics and the data with strong content and interaction dimensional characteristics, suggesting that the role of the data type is rather fixed in that context. However, social media data types are not as clearly separated as online shopping mall data, suggesting that a person is represented in a more complicated manner and that how the inforg can emerge is rather unpredictable.

6.2. Practical Implications

The suggested framework may also be useful to companies, relevant industries, and related institutes that deal with person-related information.

First, our framework can provide a basis for understanding the information of individuals that is handled by online platform providers. The personal information of users is used to provide more convenient and relevant services. However, there is also pressure to alleviate users' privacy concerns while using their information. Our framework can be useful to these providers for clarifying the standards and guidelines for handling personal information, considering the interrelated nature of the inforg dimensions: varying levels of user controls can be implemented depending on each data type and the associated levels of dimensional characteristics. For example, highly interaction-related control options such as limiting the exposure to certain audience groups can be provided to data types with high form.

Second, understanding personal information from the archetype of a person can aid in making the appropriate decisions in various circumstances and determining a suitable environment for dealing with person-related information. For example, with the rapid spread of Covid-19, many governments took different paths, and some of them used information technology intensively to stop the pandemic. They collected various personal information types, including not only symptom-related health data provided by individuals in quarantine but also locational data through the tracking of mobile device traces and credit card usage history (Goh, 2020; Huang and Simmon-Duffin, 2020; Zastow, 2020). In such an urgent crisis, the use of various means of and sometimes unlimited access to information on a person may be easily acceptable for the greater good, but what has once been accepted can be implemented more easily in different circumstances as well. Which personal information can we or should we be tolerant towards providing in such a situation? To what degree? Which personal information should we not sacrifice under any circumstances? To answer these questions, our framework serves as a reminder that now is the time to construct a digital archetype of a person and that we should consider the elaborate characteristics of each information type; thus, we should take into account the clear picture of how the digital information of a person is built and which consequences may be caused depending on their characteristics. Subsequently, more specifically, we can determine how to draw the boundary of the data collection and usage around the three inforg dimensions by asking questions such as: Should we collect data with strong interaction dimensional information without delinking the linkage? Should we use content dimensional information to identify the form dimensional information that the users have not provided? How much or how strong form dimensional information should we gather? Is it appropriate to combine weak form dimensional information with strong content and interaction dimensional information? Asking such questions from the perspective of the suggested framework will provide guidance for a better means of handling personal information.

6.3. Limitations and Future Work

We have provided a framework to understand a digitally represented person and demonstrated how it can operate in various contexts. The results exhibit several limitations and suggestions for future studies.

First, although we have attempted to apply our framework to various contexts including social media and an online shopping mall, which offered interesting insights, studying other cases such as the mobile device usage environment will be helpful for generalizing the results and gaining additional knowledge from the framework. Moreover, such applications to other contexts can enhance the understanding of our framework conceptually and practically. For instance, mobile devices are considered to be very personal, and various online and offline activities are performed as mundane activities. Thus, applying our framework to this context can provide insights into how online and offline activities are combined and become informational, understand the distinct role of inforg dimensions more clearly, and ultimately leading to the construction of an inforg with rich digital footage.

Second, the suggested framework can evolve to be applicable to non-person processors. Although the processor was limited to a person in this study, the agents of processing information can be entities other than human ones in the digital world. In particular, in the era of artificial intelligence, the processor that processes and connects information may be a machine. In this case, the archetype through which the framework should construct is not the physical identity of a person, along their minds and thoughts, and their connections to outside world, but the range and scope of the machine, its operation of information, and its connections to other machine or human inforgs. Our framework can aid in building the clear outline of the machine, thereby making it not only more understandable, but also subject to appropriate regulation and development in the right direction.

Ⅶ. Conclusions

Within the metaphor of a building (Brand, 1995), the form dimension is a skin of the digital data set. It encompasses the structure of the information and defines the identity of the data set. The content dimension consists of the substance that fills the interior. As a building has meaning when people actually use it and reside in it, content dimensional information makes the data set truly meaningful. The interaction dimension is similar to the electricity, water, and communication line of the building, preventing it from being isolated from the outside world and leveraging the interior substance and service of the building. The context becomes the site on which the building is constructed. As the design of the building should change depending on the site conditions, the context determines how the whole and ultimate construction of an inforg is shaped. As a building evolves over time, the inforg can start from its real-world identity but can emerge and change into something different that gains a life path on its own. First, we create them; then, they change and evolve, shaping the online version of ourselves, and eventually, we, in the offline reality, are shaped by them on a certain level (particularly interaction dimensional information). That is, the form, content, and interaction dimensions shape an inforg and the inforg reshapes them. Once live, the information set cannot be handled in the manner in which it was created. It has its own methods, is connected to many other different information types, has many more meanings than it once had, and becomes part of a bigger set of information, namely the infosphere. Our framework suggests a perspective for understanding the personal information and provides specific dimensions to handle them better in this sense; thus, the whole and integrative picture of a digitally represented person and the world in which it resides can be properly understood.

Acknowledgements

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2019S1A5A2 A03041910).

<References>

[1] Baelo-Allué, S. (2022). Technological Vulnerability in the Fourth Industrial Revolution: Don DeLillo's The Silence (2020). In *Representing Vulnerabilities in Contemporary Literature* (pp. 135-149): Routledge.

[2] Beldad, A., De Jong, M., and Steehouder, M. (2011). I trust not therefore it must be risky: Determinants of the perceived risks of disclosing personal data for e-government transactions. *Computers in Human Behavior, 27*(6), 2233-2242. https://doi.org/10.1016/j.chb.2011.07.002

[3] Brand, S. (1995). *How Buildings Learn*. London: Penguin Books.

[4] Casemajor, N., Couture, S., Delfin, M., Goerzen, M., and Delfanti, A. (2015). Non-participation in digital media: toward a framework of mediated political action. *Media, Culture & Society, 37*(6), 850-866.

[5] Chellappa, R. K., and Sin, R. G. (2005). Personalization versus privacy: An empirical examination of the online consumer's dilemma. *Inf Technol Manage, 6*(2-3), 181-202. https://doi.org/10.1007/s10799-005-5879-y

[6] Chen, Y., and Xie, J. (2008). Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management science, 54*(3), 477-491. https://doi.org/10.1287/mnsc.1070.0810

[7] Ester, M., Kriegel, H. P., Sander, J., and Xu, X. (1996). *A density-based algorithm for discovering clusters in large spatial databases with noise*. Paper presented at the Kdd.

[8] Faja, S., and Trimi, S. (2006). Influence of the web vendor's interventions on privacy-related behaviors in e-commerce. *Communications of the Association for Information Systems, 17*(1), 27. https://doi.org/10.17705/1CAIS.01727

[9] Federal Trade Commission. (2000). *Privacy online: Fair information practices in the electronic marketplace*. Retrieved from Washington, DC:

[10] Floridi, L. (1999). Information ethics: On the philosophical foundation of computer ethics. *Ethics and Information Technology, 1*(1), 33-52. https://doi.org/10.1023/a:1010018611096

[11] Floridi, L. (2002a). Information ethics: an environmental approach to the digital divide. *Philosophy in the Contemporary World, 9*(1), 39-45. http://dx.doi.org/10.2139/ssrn.3848486

[12] Floridi, L. (2002b). On the intrinsic value of information objects and the infosphere. *Ethics and Information Technology, 4*(4), 287-304. https://doi.org/10.1023/A:1021 342422699

[13] Floridi, L. (2007). A look into the future impact of ICT on our lives. *The Information Society, 23*(1), 59-64. https://doi.org/10.1080/01972240601059094

[14] Floridi, L. (2014). *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*. OUP Oxford.

[15] Fung, B. C., Wang, K., Chen, R., and Yu, P. S. J. A. C. S. (2010). Privacy-preserving data publishing: A survey of recent developments. *ACM Computing Surveys, 42*(4), 1-53. https://doi.org/10.1145/1749603.1749605

[16] Goh, B. (2020). China rolls out fresh data collection campaign to combat coronavirus. Retrieved from https://www.reuters.com/article/us-china-health-data-collection-idUSKCN20K0LW

[17] Huang, P., and Simmon-Duffin, s. (2020). White House Strips CDC Of Data Collection Role For COVID-19 Hospitalizations. Retrieved from https://www.npr.org/sections/health-shots/2020/07/15/891351706/white-house-strips-cdc-of-data-coll

ection-role-for-covid-19-hospitalizations

[18] Jin, S. A. A. (2013). Peeling back the multiple layers of Twitter's private disclosure onion: The roles of virtual identity discrepancy and personality traits in communication privacy management on Twitter. *New Media & Society, 15*(6), 813-833. https://doi.org/10.1177/1461444812471814

[19] Jones, W. (2010). *Keeping Found Things Found: The Study and Practice of Personal Information Management.* Morgan Kaufmann.

[20] Knijnenburg, B. P. (2018). Privacy in social information access. In P. Brusilovsky and D. He (Eds.), *Social Information Access* (pp. 19-74). Springer, Cham.

[21] Liu, C., Marchewka, J. T., Lu, J., and Yu, C. S. (2005). Beyond concern—a privacy-trust-behavioral intention model of electronic commerce. *Information & Management, 42*(2), 289-304. https://doi.org/10.1016/j.im.2004.01.003

[22] Markos, E., Labrecque, L. I., and Milne, G. R. (2018). A new information lens: The self-concept and exchange context as a means to understand information sensitivity of anonymous and personal identifying information. *Journal of Interactive Marketing, 42*, 46-62. https://doi.org/10.1016/j.intmar.2018.01.004

[23] Micheli, M., Lutz, C., and Büchi, M. (2018). Digital footprints: An emerging dimension of digital inequality. *Journal of Information, Communication and Ethics in Society, 16*(3), 242-251. https://doi.org/10.1108/JICES-02-2018-0014

[24] Milne, G. R., Pettinico, G., Hajjat, F. M., and Markos, E. (2017). Information sensitivity typology: Mapping the degree and type of risk consumers perceive in personal data sharing. *Journal of Consumer Affairs, 51*(1), 133-161. https://doi.org/10.1111/joca.12111

[25] Min, J., and Kim, B. (2015). How are people enticed to disclose personal information despite privacy concerns in social network sites? The calculus between benefit and cost. *Journal of the Association for Information Science and Technology, 66*(4), 839-857. https://doi.org/10.1002/asi.23206

[26] Morley, J., Cowls, J., Taddeo, M., and Floridi, L. (2020). Public health in the information age: recognizing the infosphere as a social determinant of health. *Journal of Medical Internet Research, 22*(8), e19311. https://doi.org/10.2196/19311

[27] Palfrey, J. G., and Gasser, U. (2011). *Born digital: Understanding the first generation of digital natives*. ReadHowYouWant.com.

[28] Parkinson, B., Millard, D. E., O'Hara, K., and Giordano, R. (2018). The digitally extended self: A lexicological analysis of personal data. *Journal of information science, 44*(4), 552-565. https://doi.org/10.1177/0165551517706233

[29] Phelps, J., Nowak, G., and Ferrell, E. (2000). Privacy concerns and consumer willingness to provide personal information. *Journal of Public Policy & Marketing, 19*(1), 27-41. https://doi.org/10.1509/jppm.19.1.27.16941

[30] Polonetsky, J., Tene, O., and Finch, K. (2016). Shades of gray: Seeing the full spectrum of practical data de-intentification. *Santa Clara Law Review, 56*, 593.

[31] Robinson, C. (2017). Disclosure of personal data in ecommerce: A cross-national comparison of Estonia and the United States. *Telematics and Informatics, 34*(2), 569-582. https://doi.org/10.1016/j.tele.2016.09.006

[32] Rumbold, J. M., and Pierscionek, B. K. J. B. d. r. (2018). What are data? A categorization of the data sensitivity spectrum. *Big Data Research, 12*, 49-59. https://doi.org/10.1016/j.bdr.2017.11.001

[33] Russo, F. (2018). Digital technologies, ethical questions, and the need of an informational framework. *Philosophy & Technology, 31*(4), 655-667. https://doi.org/10.1007/s13347-018-0326-2

[34] Shibchurn, J., and Yan, X. (2015). Information disclosure on social networking sites: An intrinsic–extrinsic motivation perspective. *Computers in Human Behavior, 44*, 103-117. https://doi.org/10.1016/j.chb.2014.10.059

[35] Simone, C., Cerquetti, M., and La Sala, A. (2021). Museums in the Infosphere: Reshaping value creation. *Museum Management and Curatorship,*

36(4), 322-341. https://doi.org/10.1080/09647775.2021.1914140

[36] Xie, E., Teo, H. H., and Wan, W. (2006). Volunteering personal information on the internet: Effects of reputation, privacy notices, and rewards on online consumer behavior. *Marketing letters, 17*(1), 61-74. https://doi.org/10.1007/s11002-006-4147-1

[37] Xu, L., Jiang, C., Wang, J., Yuan, J., and Ren, Y. (2014). Information security in big data: privacy and data mining. *IEEE Access, 2*, 1149-1176. https://doi.org/10.1109/ACCESS.2014.2362522

[38] Zastow, M. (2020). South Korea is reporting intimate details of COVID-19 cases: has it helped? Retrieved from https://www.nature.com/articles/d41586-020-00740-y

<Appendix A> Literature Review on Data of A Person

Authors	Category	Examples of Data	Form	Content	Interaction
Beldad et al. (2011)	Publicly accessible personal information	Name, postal address, telephone number	V		
	Demographic information	Age, ethnicity, gender	V		
	Highly confidential personal information	Email address, health-related information, mobile phone number	V		
	Personally identifying information	Email addresses, name, phone number, social security number, etc.	V		
Faja and Trimi (2006)	Non-identifiable information	Age, name of the school, preferred books		V	
	Identifiable information	Email, phone number, address	V		
Jin (2013)	Daily lives and entertainment	Favorite foods, movies		V	
	Social identity	School, occupation, group memberships		V	
	Competence	Intelligence, social skills		V	
	Socio-economic status and education	Education, political affiliation		V	
	Health	Mental and physical health	V		
Knijnenburg (2018)	Tastes and preferences	Explicit tastes or preferences		V	
	Implicit feedback (automatically collected from users' behavioral traces)	Clicks, purchases			V
	Context data (automatically collected, but concerns behavior that is not directly representative of users' tastes and preferences)	A user's interaction with other users, location			V
	Demographic data	Ranges from very innocuous (e.g., one's age and gender) to very private (sexual, medical, or financial topics)	V		
Markos et al. (2018)	Personally identifiable information (PII) – private-self	Medical information, DNA sequencing, credit card number	V		
	PII public self	social media profile, picture		V	
	Anonymous private self	Underwear purchased, hygiene habits, diary/journal		V	
	Anonymous public-self	Music collection, clothes purchased, family vacations		V	

<Appendix A> Literature Review on Data of A Person (Cont.)

Authors	Category	Examples of Data	Form	Content	Interaction
Milne et al. (2017)	Basic demographics	Place of birth, race	V		
	Personal preferences	Online screen name, religion		V	
	Contact information	Voice print, mobile number		V	
	Community interaction	Family/friend contact information, social network profile			V
	Financial information	Credit card number, credit score	V		
	Secure identifiers	DNA profile, social security number	V		
Phelps et al. (2000)	Demographic	Age, marital status, occupation	V		
	Lifestyle	Ways to spend leisure time		V	
	Purchase-related	Recent credit card purchases	V		
	Personal identifiers	Name, address, social security number	V		
	Financial	Annual household income	V		
Xie et al. (2006)	Identifier − general	Mobile number, homepage address	V		
	Identifier − financial	Credit card number and expiry date	V		
	Demographics − general	Date of birth, job title	V		
	Demographics − financial	Income	V		
	Demographics − preferences	Possessions, holiday destination hobbies		V	
Chellappa and Sin (2005)	Anonymous information (information gathered about page visits, without the use of any invasive technologies)	Machine's IP address, browser version and type, operating system	V		V
	Personally non-identifying information	Age, date of birth, ZIP code with no address, interests and hobbies, cookies		V	
	Personally identifying information	name, address, social security number	V		
Palfrey and Gasser (2011)	Digital identity (data elements that are divulged online to third parties)	Any information of oneself that comes up when Googling his/her name	V		
	Digital dossier (any personally identifying data associated with an individual)	One's medical record disclosed to a few designated parties of individuals	V		
Polonetsky et al. (2016)	Direct identifiers	Names, social security numbers, basic contact information	V		
	Indirect identifiers	Basic demographic information such as age and gender	V		

<Appendix A> Literature Review on Data of A Person (Cont.)

Authors	Category	Examples of Data	Form	Content	Interaction
Xu et al. (2014)	Identifiers (attributes that directly and uniquely identify an individual)	Name, mobile number	V		
	Quasi-identifiers (attributes that can identify an individual with the links of other data)	Gender, age, ZIP code	V		
	Sensitive attributes (attributes that are likely to be concealed by an individual)	Disease, salary	V		
	Non-sensitive attribute (attributes other than the above three data types)				
Jones (2010)	Own/controlled by me	Email messages, files on computers		V	
	About me	My credit/medical history, web history			V
	Directed towards me	Phone calls, advertisements			V
	Sent (provided) by me	Emails, tweets			V
	Experienced by me	webpages, articles I have read			V
	Relevant (useful) to me	Perfect vacation, job if I could find the right information		V	
Pangrazio and Selwyn (2019)	Data that users provide to devices/systems	Self-tracking information, social media data			V
	Data that devices/systems extract from users	Geolocational data from a user		V	
	Data that devices/systems process on behalf of users	Visualization, analytic pages of user activities			V
Rumbold and Pierscionek (2018)	Non-personal data	Data that have no connection to an individual or group of individuals			V
	Human-machine interactions (any human-machine interaction that can log human behavior)	Driving patterns or browsing history			V
	Human demographics, behavior, thoughts, and opinions		V		
	Readily apparent human characteristics (unprotected human characteristics that are readily apparent to the human senses without any aids)	Facial and body images		V	
	Readily apparent human characteristics (protected)	ethnic group, pregnancy	V		
	Medical or healthcare data	Genetic data, diagnoses	V		

<Appendix A> Literature Review on Data of A Person (Cont.)

Authors	Category	Examples of Data	Form	Content	Interaction
Schneier (2010)	Service data (data that users provide to a social networking site to use it)	Legal name, age	V		
	Disclosed data (what users post on their own pages)	Blog posts, comments on one's own digital space		V	
	Entrusted data (what a user posts on other people's pages)	Comments, replies on another's digital space			V
	Incidental data (what other people post about a user)	Other's posts on other's digital space about an individual			V
	Behavioral data (data that the site collect about a user's habits by recording what he or she does and with whom he or she does it)	Web history, music playlist		V	V
	Derived data (data about a user that is derived from all the other data)	Inference of an individual's preference using the composition of friends		V	

<Appendix B> The Sample Questionnaire

In the questionnaire, we first explained form, content, and interaction dimensions (as provided in Table 1 and the table below). Then, we asked participants to rate how strongly each data type is associated with each dimension. Since we repeated the same question for each data type (34 data types for social media and 27 data types for online shopping malls), only the sample question is provided using the "name" data type in the social media context. The answers were measured using a seven-point Likert scale ranging from "strongly disagree" to "strongly agree."

Explanations
FORM: The information of the form criterion consists of a person's real-world identity that makes the person distinguishable from others. CONTENT: The information of the content criterion is contained inside of a person or a person's activities that can show his or her individuality, such as thoughts, opinions, experiences, preferences, and personal activities. INTERACTION: The information of the interaction criterion is linked to others and the environment that can show how he or she is associated with what.

Questions (The below three questions were asked for each data type)
To what extent do you agree with the following statements? *Name* you use or see on the social media is associated with FORM *Name* you use or see on the social media is associated with CONTENT *Name* you use or see on the social media is associated with INTERACTIVE

◆ About the Authors ◆

Jinyoung Min

Jinyoung Min is an associate professor at the College of Business & Economics at Chung-Ang University, South Korea. She received her Ph.D. in Management Engineering at the Korea Advanced Institute of Science and Technology (KAIST). Her research interests include data privacy, social media, and algorithmic automation. Her research articles have been published in academic journals, including *Computers in Human Behavior, Journal of the Association for Information Science and Technology,* and *International Journal of Information Management.*

HanByeol Stella Choi

HanByeol Stella Choi is an assistant professor in Department of Management Information Systems, Myongji University. She received her Ph.D. in Management Engineering from the College of Business at KAIST. Her research interest includes business analytics, privacy, information security, and societal impact of information systems. Her research has been published in academic journals including *Journal of Management Information Systems, Decision Support Systems* and *Security Journal*.

Chanhee Kwak

Chanhee Kwak is an assistant professor in the Department of Artificial Intelligence Convergence. He received his Ph.D. in Management Engineering at KAIST. His research interests include data analytics and human interactions with IT artifacts. His research has been published in academic journals including *Journal of Management Information Systems, Journal of Business Ethics, International Journal of Information Management,* and *Journal of Knowledge Management.*

Junyeong Lee

Junyeong Lee is an associate professor in Department of Management Information Systems at Chungbuk National University, Korea. He received his Ph.D. in Management Engineering at the KAIST. His research interests include collective dynamics and human behavior in information systems. His work has appeared in academic journals including *Journal of Management Information Systems, Journal of the Association for Information Systems, Journal of Business Ethics, Strategic Organization, International Journal of Information Management,* and *Communications of the ACM.*

Submitted: September 18, 2023; 1st Revision: November 20, 2023; Accepted: January 7, 2024