Formalizing DIKW Architecture for Modeling Security and Privacy as Typed Resources

Yucong Duan[1], Lougao Zhan[1], Xinyue Zhang[1], and Yuanyuan Zhang[2](✉)

[1] College of Information Science and Technology, Hainan University, Haikou, China
duanyucong@hotmail.com, 2286345869@qq.com, yuexinaai@163.com

[2] College of Information Technology, Zhejiang Chinese Medical University, Hangzhou, China
zyy@zjtcm.net

Abstract. Currently the content of security protection has been expanded multiple sources. The security protection especially of the implicit content from multiple sources poses new challenges to the collection, identification, customization of protection strategies, modeling, etc. We are enlightened by the potential of DIKW (Data, Information, Knowledge, Wisdom) architecture to express semantic of natural language content and human intention. But currently there lacks formalized semantics for the DIKW architecture by itself which poses a challenge for building conceptual models on top of this architecture. We proposed a formalization of the elements of DIKW. The formalization centers the ideology of modeling Data as multiple dimensional hierarchical Types related to observable existence of the Sameness, Information as identification of Data with explicit Difference, Knowledge as applying Completeness of the Type, and Wisdom as variability prediction. Based on this formalization, we propose a solution framework for security concerns centering Type transitions in Graph, Information Graph and Knowledge Graph.

Keywords: Typed resources · Data, Information, Knowledge, Wisdom

1 Introduction

From the view of the forming process of a software requirement, individuals express their security and privacy concerns [15] in natural language specifications as a start. Currently the content of security protection has been expanded to multiple sources. The security protection especially of the implicit content from multiple sources poses new challenges to the collection, identification, customization of protection strategies, modeling, etc. We are enlightened by the potential of DIKW (Data, Information, Knowledge, Wisdom) architecture to express semantic of natural language content and human intention. Towards tackling the challenge of the complexity originating in the interleaving and association of crosscutting compositions in specification and models, we propose to categorize content objects and relationships uniformly as typed DIKW

H. Gao et al. (Eds.): TridentCom 2018, LNICST 270, pp. 157–168, 2019.
https://doi.org/10.1007/978-3-030-12971-2_10

content. But currently there lacks formalized semantics for the DIKW architecture by itself which poses a challenge for building conceptual models on top of this architecture.

We proposed a formalization of the elements of DIKW. The formalization centers the ideology of modeling Data as multiple dimensional hierarchical Types related to observable existence of the Sameness, Information as identification of Data with explicit Difference, Knowledge as applying Completeness of the Type, and Wisdom as variability prediction. Based on this formalization, we propose a solution framework for security concerns centering Type transitions in Graph, Information Graph and Knowledge Graph [13, 14]. We further categorize target resources of data and information according to their presence in searching space as explicit and implicit [2]. Corresponding security protection schemes are constructed according to the conversion and search cost differences corresponding to different types of resource expressions. These schemes support the design and provision of Value Driven [9] security solutions based on the differences of the conversion cost of different types of resources and the search cost after conversion.

Zins [4] illustrated the concepts of defining data, information and knowledge. Duan et al. [13] clarified the architecture of Knowledge Graph in terms of data, information, knowledge and wisdom. Chen et al. proposed visualization of data information and knowledge [3]. Work [1, 6] proposed to recover implicit information through abductive inference. Hundepool et al. propose to make useful inferences from groups while preserving the privacy of individuals [7]. Soria-Comas et al. [12] present ideas that privacy degree is in proportion to the amount of linkability. McSherry et al. [10] focus on sequential composition and parallel composition in composability properties.

2 Definitions in Typed DIKW Architecture

2.1 Existing Concepts of Content

To layout an uniform conceptual target of discussion, we define all content in a system description as resources regardless of whether it expresses static semantics or dynamic semantics.

Target of discussion:
$\text{Resources::=}<\text{content}_{static}, \text{content}_{dynamic}>$

Static content represents individual elements or facts, or expresses structural relationships among elements or structures. By structural relationship, we mean relationships (REL) of which the represented semantic meaning by them doesn't represent a meaningful temporal or asymmetric logical order. We denote structural relationships as reversible relationship, and denote temporal or asymmetric logical relationships as irreversible relationships.

$\text{REL::=}<\text{REL}_{reversible}, \text{REL}_{irreversible}>$
$\text{content}_{static}\text{::=}<\{\text{REL}_{reversible}\}^{*}, \{\text{element}\}^{*}>$
$\text{Structure::=content}_{static}$

Dynamic content represents temporal or asymmetric logical relationships which are irreversible. By irreversible, we mean that the reversed expression of the represented semantics marks a different semantic from the not reversed expression.

$content_{dynamic}$::=<$\{REL_{irreversible}\}^*$, $\{element\}^*$>

2.2 Typed Resources of DIKW

Towards processing the requirement content, we need to formalize the semantics of the content. We denote target semantic content with resources (RES).

RES::=<$content_{static}$, $content_{dynamic}$>

We category types of resource completely as Typed Data, Typed Information, Typed Knowledge and Typed Wisdom enlightened by the DIKW architecture as is shown in Fig. 1.

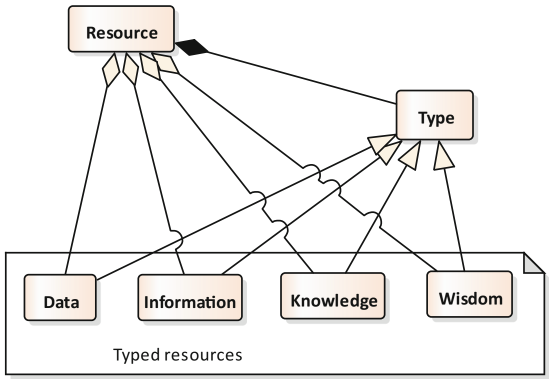

Fig. 1. Basic model of typed resources

$TYPED_{RES}$::=< Typed Data, Typed Information, Typed Knowledge, Typed Wisdom >

We elaborate the semantic of the elements of Typed Resources in the following sections.

Investigation on the formal semantic of Data, Information, Knowledge and Wisdom in DIKW architecture has long been posed as a big challenge for further investigations on the top of this architecture. Enlightened by the existence vs. identification level modeling perspective [5], empirically we summarized a conceptualization process for DIKW architecture as is shown in Fig. 2 with UML notations of Generalization, aggregation, composition and implementation.

The modeling starts from the observation of the real world which comprises objective observable existence of things. The focus centers the couples of "Entity vs. Relationship" as both objects of observations and elements for expression of observations, and "Identification vs. Notation" of which Identification marks the observed or

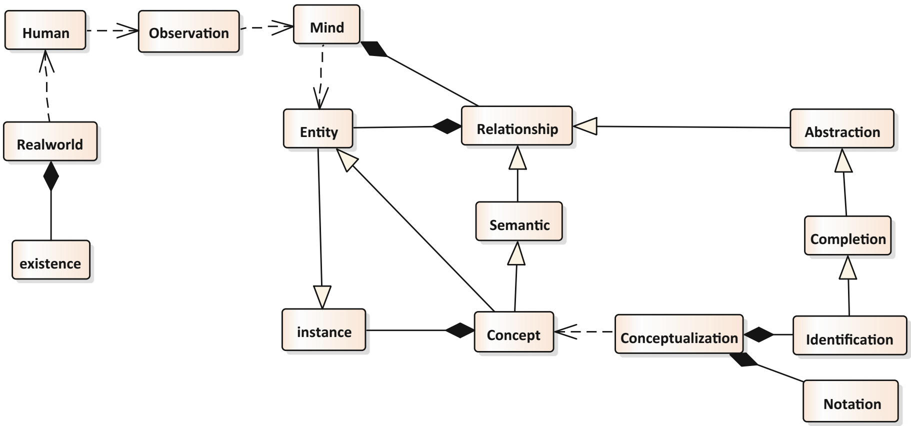

Fig. 2. Model of the conceptualization process

reasoned result of either sameness or difference, and Notation is used to explicitly give an exist symbol for an Identification. Identification on the sameness can be summarized to implement Abstraction. Semantic is based on relationship of entities which confirms to the intent of a human. Concept by its unity can be classified as Entity. Entity inherits instance since entity is judged instances in terms of the identity of its Unity.

We cognitively defined things (THG) as covering elementary targets of observation of a human represented by HUMAN (hmn) at a given time of t. We denote a specific thing as THG (thg).

$$THG(thg)::=OBSERVATION_{HUMAN(hmn)}(t)$$

2.3 Meta Model of DIKW

Figure 3 shows the Meta model of DIKW elements based on extension of Fig. 2. Observations are conceptualized as with representations of identification of ID after a cognitive process of judgement on that whether this specific new thing of THG(thgN) is the same as or different from existing labeled thing of THG(thgE). If the evaluation of the sameness of the THG(thgN) through function of SameAs is positive, the ID of the THG(thgN) will be assigned with the ID of the THG(thgE). Otherwise a new identification will be created with a function of CreateID for THG(thgN).

ID(thgN)::=
(?SameAs(THG(thgE), THG(thgN)), ID(THG(thgE)), CreateID(THG(thgN)))

We define Data in our DIKW architecture as necessarily comprising an existence level prerequisite of an confirmed observation of existence, $Existence_{prerequisite}$ or EX_{pre} and a post-requisite of identification or label of ID_{pst}.

$$Data::=<EX_{pre}, ID_{pst}>$$

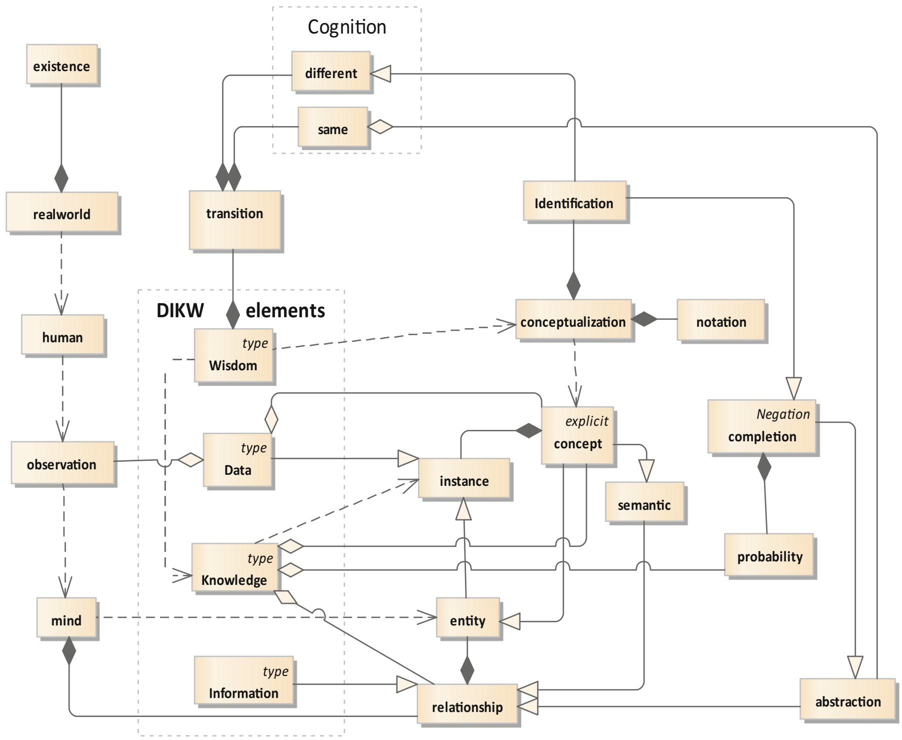

Fig. 3. Meta model of Data, Information, Knowledge and Wisdom

Here an observation means cognitively matched an output of mind thinking process either directly to existing concepts of THG(thg) or indirectly linked existing THG (thgN) as an evidence of the existence of a target of observation.

OBSERVATION(EX(THG(thg)))::=
LinkDR({THG(thg)}) OR LinkIDR({THG(thg)})
EX_{pre}::=EX(OBSERVATION(EX(THG(thg))))
ID_{pst}::=EX(ID(THG(thg)))

Based on this refined semantic of Data, we proposed the concept of Typed Data of D_{DIK} which is the foundation of our proposed Typed Resources (TR) of TR_{DIK} in the DIKW architecture. Our main innovation starts from proposing "Typed Data" as that modeling Data as defined purely by multiple dimensional related "Type (TR)" or "Class" which represents all shared recognizable commons of all subordinate instances of INSs.

TR_x::=ALL(OBSERVATION(THG(thg(x))))

By ALL, it covers unlimited amount of existing or not existing instances. Judgement on whether an INS belongs to a Type will decide the applicability of the operations and rules of the related Type. We denote Typed Knowledge as relying on the completeness of Types in terms of representing the consistency of the related instances and its interactions to infer the instance level or sub Type level associations.

Types are formed through given completeness semantic of the same things through the process of Wisdom.

TR_{DIK}: = <D_{DIK}, I_{DIK}, K_{DIK}>
D represents Data, I represents Information and K represents Knowledge for convenient description.
D_{DIK}::=<TR_x>

This definition of Data differs from existing modeling of data as instances or values in that Data is fully defined by its related or observed connected Types or Classes. For example, an observation of a cat is defined by its cognitively established connected types such as TR_{color}, TR_{size}, $TR_{sexuality}$, TR_{age}, etc. Our proposed ideology of shifting the modeling of Data from instance or values to purely Types brings a novel perspective as well on Typed Information of I_{DIK} and Typed Knowledge of K_{DIK}. Computationally the hierarchical extension of Types or Classes allows the adaptability of the attained precision and probability of correctness to be economically confirmed to the expected cost or investment from stakeholders, through planning the depth and scope of the to be traversed Type/Class hierarchy. This model of Value Driven tradeoff on Cost vs. Gain can be extended easily to graphs comprising of nodes in the form of Typed Resources.

Based on the semantic of multiple Typed dimensions, we can extend a value measure for D_{DIK} as typed frequency, TF_D. TF_D equals to the observed occurrence of D_{DIK} which is further refined to frequency values of its composing dimensional Types.

D_{DIK}::=<(Type,TF_D)>

We can further infer probability of D_{DIK}, represented Pr_D, based on TF_D through enforcing the probabilistic conditions.

The concept of Information is used to represent the Identifications ID of things THG(thg), originating in D_{DIK}, based on the judgement of the Sameness with the confirmation of the Difference.

I_{DIK}::=<$REL_{Difference}$(ID_x)>

The concept of Knowledge is used to represent the application of Completeness semantic which accompanies the semantic of Type as representing the Sameness set of Instances, in the form of Deduction which brings the patterns of Type level to Instance level or Induction which leverages instance level observation with Completeness through Abstraction to patterns of Type level. We count deduction only as a form of mechanism which doesn't require Typed Knowledge in the basic form.

K_{DIK}::=<$REL_{Sameness}$(Pattern(Induction, Deduction))>

The concept of Wisdom, W_{DIK}, is used to represent the modeling and reasoning of the variability and tendency of the change of Typed resources based on patterns or probability crossing cutting Types.

W_{DIK}::=<$REL_{Crosscut(Type)}$(Variability(Pattern, Probability))>

2.4 Connectives of Typed Resources

Based on above formalization, we propose the connectives of Typed Data, Information, Knowledge and Wisdom as is shown in Fig. 4. It further organizes the core concepts of Negation in relation to True vs. False originating in observation on Existence, Completeness in relation to Yes vs. No originating to judgement on Abstraction or Induction, Sameness in relation to quantity of frequency of Typed Data. A refinement of the connectives for Type/Class in terms of Object-Oriented attribute and method is shown in Fig. 5. A further extension of Fig. 5 is shown in Fig. 6 with added explanatory elements.

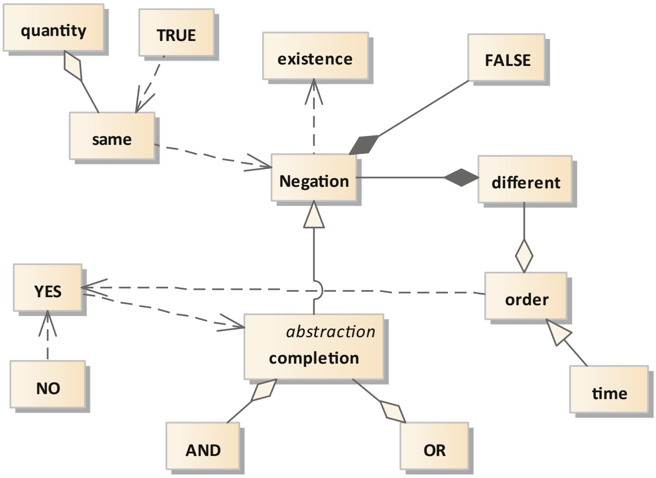

Fig. 4. Connectives of Typed Data, Information, Knowledge and Wisdom

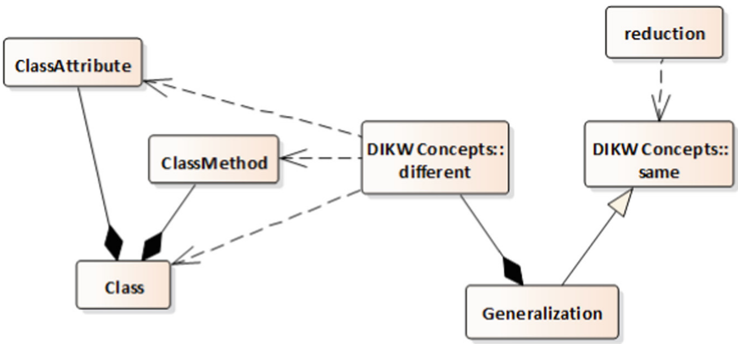

Fig. 5. Refinement of connectives

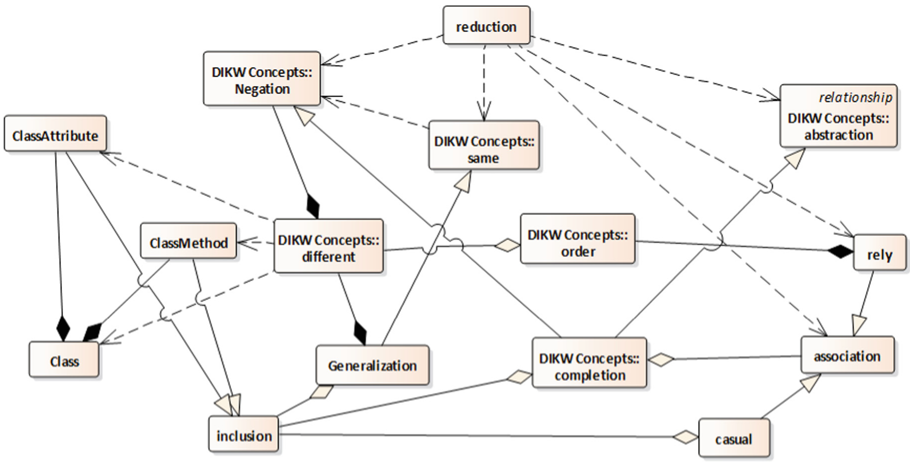

Fig. 6. Extended connectives of Typed Data, Information, Knowledge and Wisdom

3 Graphs of Typed Resources for Privacy Solution

Beside above formalization, we can easy the application of the DIKW architecture with manual modeling empirically as that: Data is not specified for a specific stakeholder or a machine and represents directly observed objects as isolation which only contains the shared common meaning of their necessary identifications. Information represents data or information which are observed or interacted directly or indirectly by human or things. Knowledge represents the abstracted data, information and knowledge in a limited or unlimited complete manner. Knowledge roughly maps to what Kant called Categories [8]. To manipulate the graphs in DIKW architecture, we need to mediate the bidirectional feasible transformations of resources among different types of Data, Information and Knowledge. Schemas [8] are proposed by Kant to cognitively mediate the cognitive objects/experiences mostly through logical reason and concretization in time or logical dimension. We have borrowed this term for specifying the transformation among resources with a focus on the type level implementation.

Schema "Data-Resource (Data, Information)": Data are observed by observers from outside world or from inside categorization on a set of resources, structured or not, which are given the conceptual unity as an entity, or on abstraction of information expressions which are exposed as temporal association among elements. Since resource elements can be abstracted upward or decomposed downward, the expressions of specific DG_{DIK} and IG_{DIK} are therefore intertwined based on the overlapping of the elements and their relationships. We propose to justify and predict the semantic meaning and semantic associations corresponding to resource element expressions based on the reasoning and calculation in a bottom up manner from composing elements of DG_{DIK} and IG_{DIK}.

Schema "Knowledge-Resource (Data, Information, Knowledge)": Knowledge here is either based on the probabilistic experience or based on reason on categories abstracted from directly observed resources or indirectly reasoned resources. A shared characteristic of both forms of knowledge is that they both demand a semantic identification of completeness regardless of whether the actual target resources which are the basis of conceptualization of related categories are limited or unlimited. The schema to enact knowledge on resources is either through temporally decomposing the content of the comprising categories in the knowledge expression as elements shared or can be related to elements in target resources, or through logical or probability reasoning first and decomposing and relating subsequently.

For construction of "Wisdom" related schemas, we adopt the intuition from Schopenhauer [11] to take wisdom as the balancing between reasoning and will for optimizing human long run goals. We omit the discussion on the schema of wisdom here.

We worked towards build "schemas" [8] for DIKW resources for privacy modeling and provision.

DIKWGraph. We specify the usually used concept of Knowledge Graph in three layers of Data Graph (DG_{DIK}), Information Graph (IG_{DIK}), and Knowledge Graph (KG_{DIK}).

DIKWGraph: = <(DG_{DIK}), (IG_{DIK}), (KG_{DIK})>.
DG_{DIK}. DG_{DIK}: = collection {array, list, stack, queue, tree, graph}.

DG_{DIK} is a collection of discrete elements expressed in the form of various data structures including arrays, lists, stacks, trees, graphs, etc. DG_{DIK} records basic structures of entities. DG_{DIK} records spatial and topological relationships with frequencies.

IG_{DIK}. IG_{DIK}: = $\text{composition}_{time}$ {D_{DIK}}.

IG_{DIK} comprises of temporal relationships based on D_{DIK} with specific scenarios. IG_{DIK} expresses the interaction and transformation of I_{DIK} between entities in the form of a directed graph. IG_{DIK} can record the interactions between entities including direct interaction and indirect interaction.

KG_{DIK}. KG_{DIK}: = $\text{collection}_{consistent}\{\text{Rules}_{\text{Statistic OR Logical}}\}_{category}$.

KG_{DIK} consistently accommodates either empirical statistical experiences expressed in terms of categories which represent the underlying elements as a whole or completely. Statistic in $\{\text{Rules}_{\text{Statistic OR Logical}}\}_{category}$ includes philosophy of Bayesian statistics. Samples of Bayesian classification algorithm may belong to a certain class based on probability.

Typed privacy resources. We formalize privacy resources in Table 1 and we give explanations in campus monitoring system. We define typed resources as a triad:

PR_{DIK}: = <PD_{DIK}, PI_{DIK}, PK_{DIK}>.

Table 1. Definitions of privacy resources

Type	D_{DIK}	I_{DIK}	K_{DIK}
Privacy	$\mathbf{PD_{DIK}}$:={Pdx_1,Pdx_2…Pdx_n}	$\mathbf{PI_{DIK}}$:={Pix_1,Pix_2…Pix_n}	$\mathbf{PK_{DIK}}$:={Pkx_1,Pkx_2…Pkx_n}

Explicit and implicit privacy resources. We further categorize target privacy resources of data and information according to their presence in searching space as explicit and implicit. Figure 7 shows our proposition of the semantic of Explicit as directly related to Existence of Typed Data by way of Entity, and Implicit as although mapping to relationships but not directly related to Existence of Typed Data by way of Entity.

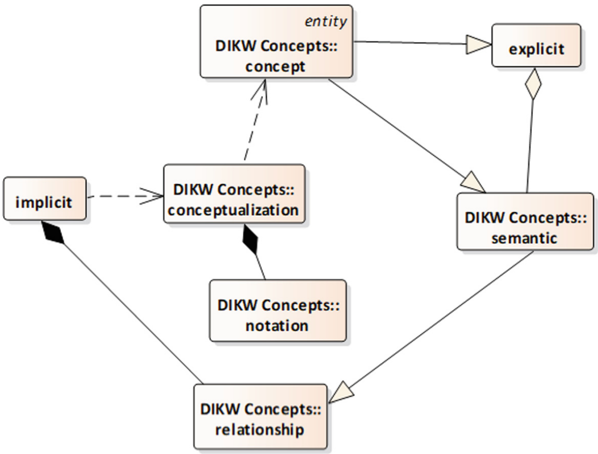

Fig. 7. Semantic of explicit vs. implicit

We proposed to list mark the cost of Type level transition for per unit of source content as is shown in Table 2.

Table 2. Atomic cost of transforming unit resource

	D_{DIK}	I_{DIK}	K_{DIK}
D_{DIK}	$SUnitCost_{D\text{-}D}$	$SUnitCost_{D\text{-}I}$	$SUnitCost_{I\text{-}K}$
I_{DIK}	$SUnitCost_{I\text{-}D}$	$SUnitCost_{I\text{-}I}$	$SUnitCost_{I\text{-}K}$
K_{DIK}	$SUnitCost_{K\text{-}D}$	$SUnitCost_{K\text{-}I}$	$SUnitCost_{K\text{-}K}$

We categorize typed implicit and explicit privacy resources in DIKW Graphs. We conclude an interpretation table shown as Table 3.

Table 3. Preparation of description terms

Term	Interpretation	Term	Interpretation
○	Explicit data resources	□	Explicit knowledge resources
●	Implicit data resources	■	Implicit knowledge resources
◌	Transformed D_{DIK}	⬚	Transformed K_{DIK}
△	Explicit information resources	⟶	Connection in resources
▲	Implicit information resources	- - ->	Existed resources
⛛	Transformed I_{DIK}	→	Add new resources

Figure 8 shows the flow chat of creating requirement specification modeling in DIKW Graphs. Users provide requirement specifications. We optimize requirement specifications through analyzing feasibility of target explicit and implicit privacy resources (A), enhancing consistency of explicit and implicit resources (B), eliminating redundancy of explicit and implicit resources (C), and enhancing completeness (D).

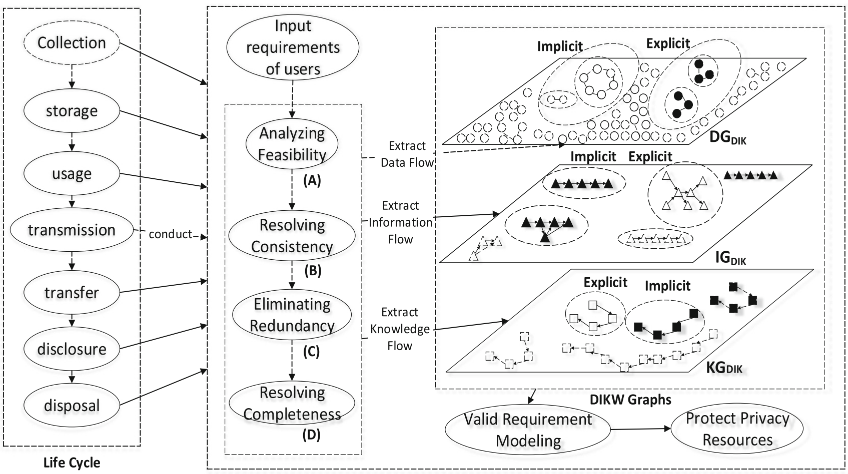

Fig. 8. Flow chat of creating requirement specification modeling in DIKW graphs.

4 Conclusion

In this work, we proposed a formalization of the semantic of the DIKW (Data, Information, Knowledge, Wisdom) architecture. The formalization centers the ideology of modeling Data as multiple dimensional hierarchical Types related to observable existence of the Sameness, Information as identification of Data with explicit Difference, Knowledge as applying Completeness of the Type, and Wisdom as variability prediction of typed Data, Information, Knowledge and Wisdom. Based on this formalization, we modeled the security and privacy content as typed resources of Data, Information, Knowledge and Wisdom. Then base on the difference of the Type transitions among different Typed resources and the difference of the processing cost of converted expressions of target content, we proposed a solution framework which permits Value Driven application of protection functionalities and quality in accordance with the planned investment from stakeholders.

Acknowledgement. We acknowledge Hainan Project No. ZDYF2017128, NSFC under Grant (No. 61363007, No. 61662021, and No. 61502294), Zhejiang Province medical and health science and technology platform project No. 2017KY497. *refers correspondence.

References

1. Appelt, D.E.: FASTUS: a finite-state processor for information extraction from real-world text. In: Proceedings of IJCAI 1993, pp. 1172–1178 (1993)
2. Pearson, P.D., Hansen, J., Gordon, C.J.: The effect of background knowledge on young children's comprehension of explicit and implicit information. J. Literacy Res. **11**(3), 201–209 (1979)
3. Chen, M., et al.: Data, information, and knowledge in visualization. IEEE Comput. Graph. Appl. **29**(1), 12–19 (2008)
4. Zins, C.: Conceptual approaches for defining data, information, and knowledge. J. Assoc. Inf. Sci. Technol. **58**(4), 479–493 (2014)
5. Duan, Y., Fu, X., Hu, Q., Gu, Y.: An ontology definition framework for model driven development. In: Gavrilova, M.L., et al. (eds.) ICCSA 2006. LNCS, vol. 3983, pp. 746–755. Springer, Heidelberg (2006). https://doi.org/10.1007/11751632_81
6. Grishman, R.: Information extraction: techniques and challenges. In: Pazienza, M.T. (ed.) SCIE 1997. LNCS, vol. 1299, pp. 10–27. Springer, Heidelberg (1997). https://doi.org/10.1007/3-540-63438-X_2
7. Hundepool, A., et al.: Statistical Disclosure Control. Wiley, New York (2012)
8. Kant, I.: Critique of Pure Reason. Cambridge University Press, Cambridge (1998)
9. Shao, L., Duan, Y., Cui, L., Zou, Q., Sun, X.: A pay as you use resource security provision approach based on data graph, information graph and knowledge graph. In: Yin, H., et al. (eds.) IDEAL 2017. LNCS, vol. 10585, pp. 444–451. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-68935-7_48
10. Frank, D.: McSherry: privacy integrated queries: an extensible platform for privacy-preserving data analysis. Commun. ACM **53**(9), 89–97 (2010)
11. Schopenhauer, A.: The world as will and representation. Kenyon Rev. **20**(3/4), 44–46 (1998)
12. Soria-Comas, J., Domingo-Ferrer, J.: Big data privacy: challenges to privacy principles and models. Data Sci. Eng. **1**(1), 21–28 (2016)
13. Duan, Y., Shao, L., Hu, G., Zhou, Z., Zou, Q., Lin, Z.: Specifying architecture of knowledge graph with data graph, information graph, knowledge graph and wisdom graph. In: International Conference on Software Engineering Research, pp. 327–332. IEEE (2017)
14. Aamodt, A., et al.: Different roles and mutual dependencies of data, information, and knowledge—an AI perspective on their integration. Data Know. Eng. **16**(3), 191–222 (1995)
15. Tarr, P., Ossher, H., Harrison, W., Sutton Jr., S.M.: N degrees of separation: multi-dimensional separation of concerns. In: International Conference on Software Engineering, vol. 3, pp. 107–119. IEEE (1999)