Data, Information, Evidence, and Knowledge:

A Proposal for Health Informatics and Data Science

Olaf Dammann, MD, SM

Dept. of Public Health and Community Medicine, Tufts University School of Medicine, Boston, MA, United States

Abstract

In this commentary, I revisit and modify Ackoff's data-information-knowledge-wisdom (DIKW) hierarchy. I suggest to de-emphasize the wisdom part and to insert evidence between information and knowledge (DIEK). This framework defines data as raw symbols, which become information when they are contextualized. Information achieves the status of evidence in comparison to relevant standards. Evidence is used to test hypotheses and is transformed into knowledge by success and consensus. As checkpoints for the transition from evidence to knowledge I suggest relevance, robustness, repeatability, and reproducibility.

Keywords: Data, Information, Evidence, Knowledge

Introduction

Data, information, and knowledge are central concepts in health informatics and data science. It is not always clear how authors define these entities and how they envision the transition from data to knowledge to work. In this commentary, I first review the knowledge/wisdom hierarchy proposed by organizational theorist Russell A. Ackoff in 1989 [1]. Second, I outline a modification of Ackoff's framework that does away with his notion of wisdom and makes room for evidence. I also discuss the transition process of from data to knowledge, with a focus on the transition from evidence to knowledge. I hope that the ideas summarized here will prove helpful to those in charge of knowledge generation in health informatics and data science.

Ackoff's Knowledge Hierarchy

Russell L. Ackoff (1919-2009) introduced what is now known as the *knowledge hierarchy* or *knowledge pyramid* (Fig. 1, left) in his presidential address to the International Society for General Systems Research (ISGSR) in 1988 [1]. He starts with the notion that wisdom is situated at the top of a hierarchy of types of content in the mind, followed by understanding, knowledge, information,

and data (Fig. 1; of note, Ackoff's original article does not have a figure, nor does it refer to pryamids.) He defines *data* as symbols that are properties of observables, and *information* as descriptions. The difference between the two is not structural, but functional, and information is inferred from data.

Ackoff discusses management needs in terms of information availability. He states that managers are usually confronted with an information overload and do not necessarily need more relevant information but less irrelevant information, a truism then and now. He defines *knowledge* as know-how that comes from learning, i.e., by instruction or from experience, and adaptation, i.e., the correction of the learned in accordance with new circumstances. This process requires understanding what error is, why error occurs, and how to correct it. Ackoff thinks that (1) information systems can be automated and generate information out of data, (2) that computer-based knowledge systems require higher-order mental faculties; "they do not develop knowledge, but apply knowledge developed by people", and (3) that wisdom adds value, endures forever, and will probably never be generated by machines.
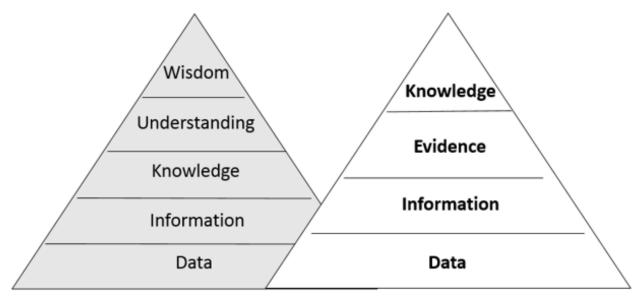

Figure 1. Ackoff's Knowledge (DIKW) hierarchy (left) and the DIEK modification proposed in this commentary (right)(reprinted from [3]).

Ackoff's hierarchy is often depicted as a pyramid (as in Fig. 1 in this article) with data at the bottom, information and knowledge above, and wisdom at the top. Probably for this reason, Jennifer Rowley uses the term "wisdom hierarchy". [2] Although she seems more interested in the wisdom part than in other components of the pyramid, the bulk of her 2007 paper on Ackoff's work is a summary of terminological definitions, of *data, information, knowledge, and wisdom*, as pulled from major textbooks used in information system and knowledge management education. Her review reiterates two opinions; first, her view that data, information, and knowledge are connected, one helping define the other, and second, her view about the organization of the hierarchy as such. The ways how the individual items in the hierarchy are converted and elevated to the next level is less well defined.

Data, Information, Evidence, Knowledge: *DIEK*

Although the Ackoff hierarchy has received much attention over the years, I strongly believe that in our current evidence-based environment some modifications are in order. First, in a book co-authored with philosopher of science Ben Smart, I suggest dropping the notion of wisdom because, first, the term is fraught with too much baggage from non-scientific context [3]. Second, Ackoff's definition of wisdom (the addition of value to knowledge that requires judgement) ignores the fact that judgement is needed at all levels of the hierarchy. More importantly, I do not think that wisdom adds much to the decision-making based on the hierarchy. Instead, I hold that knowledge deserves the position at the pinnacle of the hierarchy. Knowledge can be defined, in the context of medical and public health informatics and data science, as *predictive, testable, consistently successful belief*, if there is a causal connection between the facts represented by the data, information, and evidence on the one hand, and our beliefs on the other.
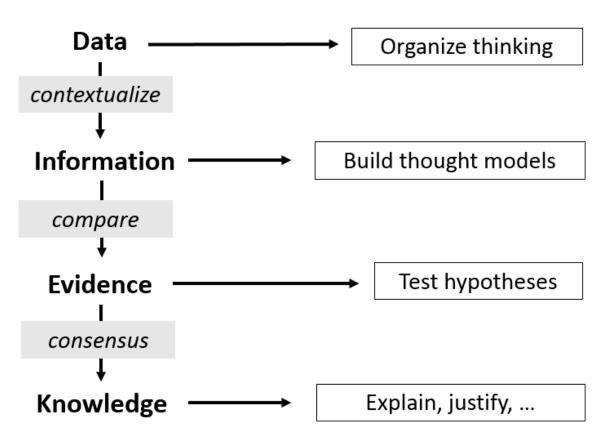
Data

In the context of Public Health Informatics, Mensah and Goderre define "data" as raw facts, statistics, context-free numbers [4]. I'd like to suggest that *data are symbols as retrieved, collected, or simulated* (Table 1). These include numbers resulting from measurements or from text-mining, images, sound recordings, survey results, simulations, and so on. They can usually be tabulated and depicted as graphs, or displayed as figures. More formally speaking, data are quantitative or qualitative values of variables. Figure 2 displays a framework for transitions from data to knowledge, and what the arrival at each new stage is good for.

Table 1. Explanations of what data, information, evidence, and knowledge are, and how they are produced, by whom, and why (modified from [**3**]).

Concept	What is it?	How produced?	By whom?	Goal?
Data	Numbers, Symbols, Text, Images, Sound recordings, Unit values	**Collected** from field research, database, measurements in experiments, from individuals, populations	Data Collector	Use as raw data or for information generation Storage, curation, retrieval
Information	Data in context	**Contextualization** by making data useful, and using them, for specific tasks	Informatician, informaticist, statistician, data scientist	Use as source for answering questions Storage, curation, retrieval
Evidence	Useful, contextualized information	**Comparison** with standards, reference values, reference information	Scientist, theoretician, philosopher, Interventionist, policy maker	Use for analysis and hypothesis-testing to support claims/hypotheses and decision-making
Knowledge	Evidence-based, (predictive, testable, consistently successful) belief	**Consensus** based on reasoning and discussion		Justification

Figure 2. Framework for the transition from data to knowledge (left) and what each level is good for (right) (reprinted from [3]).

Information is data contextualized

Mensah & Goderre further suggest that "information is the collection, aggregation, analysis, and presentation of data that provides understanding". [4] Although this definition describes how we arrive at information based on data, it does not tell us what information *is*. I think that *information is data in context*. Information is data that have been processed so it is clear what they are about. Once they are collected and contextualized, data *are* information. According to this view, all information is data, but not all data are information.

Evidence is information compared

Information thus conceived can give rise to evidence, which has been defined as "information bearing on the truth or falsity of a proposition". [5] Evidence is information that can be used to support a hypothesis by testing it. Thus, *all* evidence is information, but not all information is evidence. The comparison of information in support of competing conjectures helps define what counts as evidence that, in turn, generates the knowledge that a certain overarching claim is true.

Evidence is generated by comparing information to reference values or standards, which prepares the information for further analysis. In the context of public health, Brownson and colleagues have argued that

> *(f)or a public health professional, evidence is some form of data—including epidemiologic (quantitative) data, results of program or policy evaluations, and qualitative data—for uses in making judgments or decisions"* [6]

They describe three kinds of evidence in public health contexts: (1) the causes of illness and the magnitude of risk factors, (2) the relative impact of specific interventions, and (3) how and under which contextual conditions interventions were implemented [6]. We discuss the intervention-related part of these kinds of evidence in more detail elsewhere [7].

In general, *evidence is information that bears on the truth of a proposition compared to a standard.* According this definition, information becomes evidence only if it bears on the truth or falsity of the proposition that the gardener was indeed the murderer. Only if we can find *good evidence* that is coherent with this claim can we say that we have knowledge that he really is the culprit. Actionable knowledge is usually generated from coherent evidence from multiple independent sources of information [8]. If we refer to *evidence* as information that supports a specific proposition by bearing on its truth, evidence is context-dependent, because it becomes evidence only by virtue of being relevant as support for a specific proposition, and relevance is, by definition, a contextual concept.

Knowledge from evidence

The traditional tripartite concept of *knowledge as justified, true belief* goes all the way back to Plato [9]. Gettier argued in 1963 that the tripartite definition is not sufficient to constitute knowledge, in essence by offering two counterexamples in which some justified, true beliefs clearly do *not* count as knowledge [10]. Multiple strategies to defeat Gettier have been suggested [11]. In our present context, I think that knowledge consists of beliefs that

1. turn out to be predictive: predictions that are based on such beliefs turn out to be correct;
2. generate hypotheses that can be tested, and
3. ideas that lead to interventions that are successful,
4. for a long time.

In other words, I suggest that beliefs qualify as knowledge if they predict outcomes with satisfactory precision, if they can be translated into scenarios that put the belief to the test, and if actions based on such beliefs are consistently successful. In short, *knowledge is predictive, testable, consistently successful belief*. Indeed, this is exactly what we refer to some belief as being evidence-based. This is why *evidence*-based medicine and public health should actually be considered *knowledge*-based once the evidence has turned out to be predictive, is tested, and interventions have been designed and are consistently successful. Of course, the decision when

that point has been reached is not made by any one person, but by consensus [12,13]. Thus, all knowledge is evidence, but not all evidence is knowledge.

Are there checkpoints that support the decision to promote evidence to the level of *good* before we have seen the quality of its predictions, witnessed its testability, and received the good news that interventions based on such evidence are being consistently successful? Here is a collection of candidate checkpoints that I think allow us to proceed from evidence to knowledge. Since we ask this question with an intervention in mind, our query is not really what makes evidence so good that it is knowledge, but rather what makes evidence so good that it is *useful* in our context. Usefulness, in turn, is simply the possibility to use this knowledge in ways that turn out to help improve the health of individuals and populations. We need knowledge to justify action.

First, although this should go without saying, good evidence is *relevant* to the problem at hand. Consider this quote from the *Annual Review of Public Health*:

> *Legislators and their scientific beneficiaries express growing concerns that the fruits of their investment in health research are not reaching the public, policy makers, and practitioners with evidence-based practices. Practitioners and the public lament the lack of relevance and fit of evidence that reaches them and barriers to their implementation of it [14]*

If evidence is irrelevant, it isn't useful. The focus on usefulness is, yet again, motivated by the goal of health informatics efforts to inform decision making which leads to effective action. Second, good evidence is *robust*. This is what Broadbent has called the stability of a result, i.e., the characteristic of a theory or piece of evidence that it is (a) not soon contradicted by good scientific evidence, and (b) unlikely that it will soon be contradicted by good scientific evidence, if good research were to be done on the topic [15]. Third, good evidence is *repeatable* in the sense that similar data gathering and integration efforts lead to similar evidence repeatedly: "Repeatability concerns the exact repetition of an experiment, using the same experimental apparatus, and under the same conditions". [16] Fourth, good evidence is *reproducible*: "Reproducibility is … implementing the same general idea, in a similar setting, with newly created appropriate experimental apparatus". [16]

Conclusion

My version of the Ackoff hierarchy is based on what is being done to make such transitions possible, not what transitions represent or what happens when moving from one level to another, such as changes of meaning and value [17] or the physical, cognitive, and belief structuring when constructing data, information, and knowledge, respectively [18]. As Rowley's focus is on the relative paucity of explications of *wisdom*, mine is instead on the fact that the concept of *knowledge*, now at the top of the hierarchy, is not well defined either.

A similar model has been proposed by Richard Heller. In his model, accessing data yields information, appraisal of which yields knowledge. What is missing in Heller's model is the distinct role that evidence plays between information and knowledge. Neither in his book [19] nor in the

underlying paper [20] does he define evidence. However, in their 2002 publication, Heller and Page offer a list of statistical and implementation characteristics they see as methods with an appropriate population focus that can be aligned with the methods used in evidence-based medicine because the authors consider the entire process from data via information to knowledge to be *evidence-generating*.

I should stress that knowledge isn't something out there for us to discover. Instead, knowledge is *made*. In this commentary, I have outlined a framework that builds on Ackoff's knowledge-hierarchy, in which data give rise to information, which leads to knowledge and finally wisdom. My version of the model drops the notion of wisdom, because it is too imprecise a notion to be useful in a health science context. Instead, I suggest to insert the notion of evidence into the inferential sequence between information and knowledge. Data are used mainly as raw material for information generation. When these data are put into context, they yield information that may be useful as evidence. Based on such evidence, knowledge is generated. Knowledge is evidence-based belief that is predictive, testable, and consistently successful, as judged by consensus among stakeholders. I hope that this proposed modification of Ackoff's framework will contribute to the progress of health informatics and data science.

References

1. Ackoff RL. 1989. From data to wisdom. *J Appl Syst Anal*. 16, 3-9.

2. Rowley J. 2007. The wisdom hierarchy: representations of the DIKW hierarchy. *J Inf Sci*. 33(2), 163-80. https://doi.org/10.1177/0165551506070706

3. Dammann O, Smart B. *Making Population Health Knowledge*, in *Causation in Population Health Informatics and Data Science*. 2019, Springer Nature: Cham, Switzerland. p. 63-77.

4. Mensah E, Goderre JL. *Data sources and data tools*, in *Public health informatics and information systems*, J.A. Magnuson and P.C. Fu, Editors. 2014, Springer: London. p. 107-131.

5. Audi R. *The Cambridge dictionary of philosophy*. 2nd ed. 1999, Cambridge; New York: Cambridge University Press. xxxv, 1001 p.

6. Brownson RC, Fielding JE, Maylahn CM. 2009. Evidence-based public health: a fundamental concept for public health practice. *Annu Rev Public Health*. 30, 175-201. PubMed https://doi.org/10.1146/annurev.publhealth.031308.100134

7. Dammann O, Smart B. *Integrating Evidence*, in *Causation in Population Health Informatics and Data Science*. 2019, Springer Nature Cham, Switzerland. p. 99-115.

8. Dammann O. 2018. Hill's Heuristics and Explanatory Coherentism in Epidemiology. *Am J Epidemiol*. 187(1), 1-6. PubMed https://doi.org/10.1093/aje/kwx216

9. Ichikawa JJ, Steup M. *The Analysis of Knowledge*. The Stanford Encyclopedia of Philosophy 2013 03/07/2014]; Available from: http://plato.stanford.edu/archives/fall2013/entries/knowledge-analysis/.

10. Gettier J. 1963. Is justified true belief knowledge? *Analysis*. 23(6), 121-23. https://doi.org/10.1093/analys/23.6.121

11. Lucey KG. *On knowing and the known: introductory readings in epistemology*. 1996, Amherst, N.Y.: Prometheus Books. 437 p.

12. Fleck L. *Genesis and development of a scientific fact*. 1979, Chicago: University of Chicago Press. xxviii, 203 p.

13. Solomon M. *Making medical knowledge*. First edition. ed. 2015, Oxford: Oxford University Press. xiii, 261 pages.

14. Green LW, Ottoson JM, Garcia C, Hiatt RA. 2009. Diffusion theory and knowledge dissemination, utilization, and integration in public health. *Annu Rev Public Health*. 30, 151-74. PubMed https://doi.org/10.1146/annurev.publhealth.031308.100049

15. Broadbent A. *Philosophy of epidemiology*. New directions in the philosophy of science. 2013, Houndmills, UK: palgrave macmillan.

16. Feitelson DG. 2015. From repeatability to reproducibility and corroboration. *Oper Syst Rev*. 49(1), 3-11. https://doi.org/10.1145/2723872.2723875

17. Chaffey D, Wood SJ. *Business Information Management: Improving Perfomance using Information Systems*. 2005: Prentice Hall.

18. Choo CW. 1996. The knowing organization: How organizations use information to construct meaning, create knowledge and make decisions. *Int J Inf Manage*. 16(5), 329-40. https://doi.org/10.1016/0268-4012(96)00020-5

19. Heller RF. *Evidence for population health*. 2005, Oxford; New York: Oxford University Press. xii, 126 p.

20. Heller RF, Page J. 2002. A population perspective to evidence based medicine: "evidence for population health. *J Epidemiol Community Health*. 56(1), 45-47. PubMed https://doi.org/10.1136/jech.56.1.45