

Article

Not peer-reviewed version

---

# DIKW as a General and Digital Twin Action Framework: Data, Information, Knowledge, and Wisdom

---

[Michael Grieves](#) \*

Posted Date: 2 February 2024

doi: 10.20944/preprints202402.0129.v1

Keywords: Data; Information; Knowledge; Wisdom; DIKW; Digital Twins



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

WORKING PAPER

# DIKW As a General and Digital Twin Action Framework: Data, Information, Knowledge, and Wisdom

Michael Grieves <sup>1,2</sup>

<sup>1</sup> Digital Twin Institute; mgrieves@mwgvp.com

<sup>2</sup> University of Central Florida; michael.grieves@ucf.edu

**Abstract:** This paper will discuss Data, Information, Knowledge, and Wisdom, which is commonly referred to as DIKW. The DIKW Pyramid Model that is often referenced in both academics and practitioner circles will be discussed and shown to be faulty on several levels, including a lack of definitional agreement. A new DIKW framework will be proposed that focuses on what the DIKW elements do in the way humans think, not what they are by definition. Information as a replacement for wasted physical resources in goal-oriented tasks will be a central organizing point. The paper will move the DIKW discussion to the computer-based concept of Digital Twins (DTs) and its augmentation of how we can use DIKW to be more effective and efficient. This will especially be the case as we move to Intelligent Digital Twins (IDT) with Artificial Intelligence (AI).

**Keywords:** data; information; knowledge; wisdom; DIKW; digital twins

---

*Where is the wisdom we have lost in knowledge?  
Where is the knowledge we have lost in information?*  
T. S. Elliot, Choruses from the "Rock"

*God grant me the serenity  
to accept the things I cannot change;  
courage to change the things I can;  
and wisdom to know the difference.*  
Serenity Prayer

## 1. Introduction

Data, information, knowledge, and wisdom are fundamental for humans to deal with our existence in the world. However, these four words might have the most different definitions and interpretations in the English language of what they mean. I would like to start off by stating that I am under no illusions that I am going to correct what I see is the confusion, misuse, and, at times, contradictory use of the terms: data, information, knowledge, and wisdom.

I would contend that by the definitions I am proposing we use information when we are actually referring to data. We use data when we mean information. We constantly use data and information interchangeably. We seem to understand that knowledge is in a slightly different category. Wisdom is usually reserved for weighty moral matters and not applicable to everyday decisions.

In practice, we use the data, information, and knowledge pretty interchangeably. We can use data, information, and knowledge in three different sentences about the same thing, and no one would think to correct us. All three get lumped into a category of "stuff we know". When we use these words, in casual conversation, we understand reasonably well what we're talking about. No one knows well enough or considers it important enough to correct us when we use the terms in what a strict interpretation by my definition or many others would consider to be inaccurate.

I have long contended (Grievies 2000) that it's the physical hardware of our brains that allows us to do this because our mental hardware is highly associative and not discretely addressable (Kanerva 1988, Hely, Willshaw et al. 1997). As humans, we use data, information, and, to a lesser extent, knowledge interchangeably almost all the time. Data, information, and knowledge are neurologically associated and linked in our brains (Seung 2012). Thinking about one element immediately retrieves associated elements.

We would acknowledge that these words mean different things at the same time we are using them relatively interchangeably. We know that there's something about them that is different, but practically, because of how our neurological hardware operates, it is too confusing to make much of an effort at their differentiation.

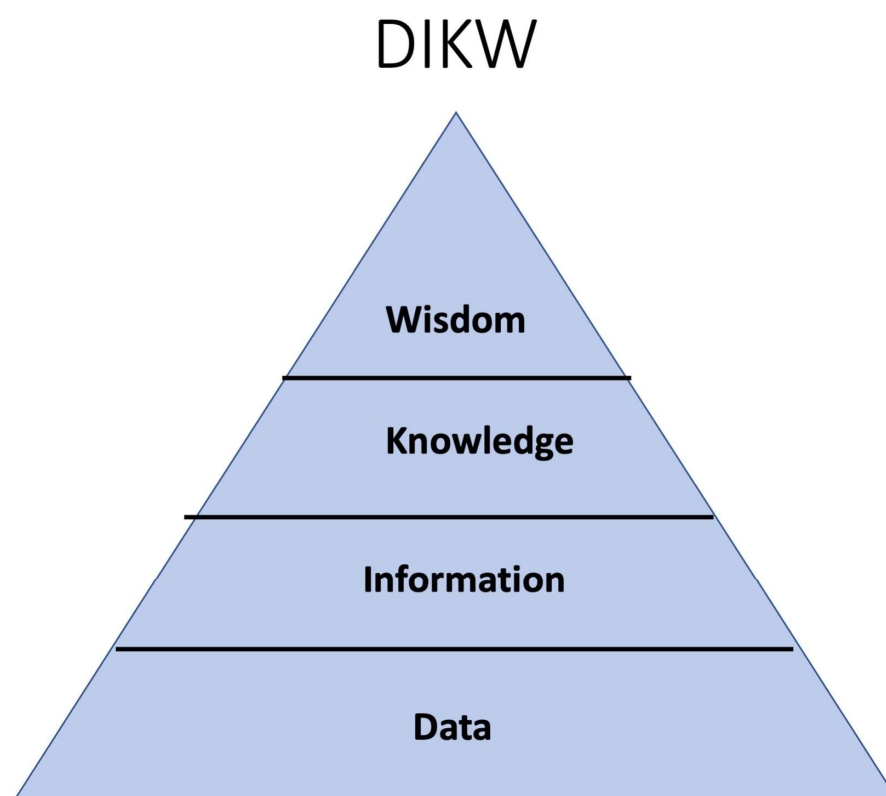
Wisdom seems to be in a different category. Wisdom has been mostly the purview of philosophers like Aristotle (Aristotle and Beresford 2020), for millennia. We really haven't considered that wisdom applies to the use of data and knowledge in ordinary decision-making using information.

We also have a very difficult time when we try to discriminate the terms even in academic literature. Periodically, the academic community makes an attempt at categorizing, defining, and ordering these different concepts. One of the attempts that is common and has even made it into the practitioners' community is the DIKW model.

## 2. DIKW Pyramid Model

In spite of some giving the poet of the opening poem excerpt, T.S. Elliot, credit for the DIKW pyramid model, most serious discussions start with a talk given by Professor Ackoff (1989). However, what gets left out is that his hierarchy had "understanding" between "information" and "knowledge." In spite of that, the hierarchy gets taken up by other academics and practitioners who attempt to articulate this hierarchy.

The DIKW model is shown in Figure 1. The figure represents a hierarchical model of data, information, knowledge, and wisdom. The general pyramid model hierarchical structure is where the lower level builds into the level above. Data makes up information. Information makes up knowledge. Knowledge makes up wisdom.



**Figure 1.** DIKW Pyramid Model.

The pyramid model also implies that the number of instances decrease or the value increases from the lower to the higher level. With the DIKW pyramid model, it's implied that value increases up the pyramid. Information is more valuable than data. Knowledge is more valuable than information. Wisdom is more valuable than knowledge. It's unclear if the DIKW model implies that the number of instances in each of these categories decreases. While pyramid models commonly imply that value increases while volume decreases as one move up levels, it's not obvious that this is the case with data, information, knowledge, and wisdom.

There are two core problems that undermine the DIKW pyramid as a useful concept. The first is that there is little or no consensus as the what the DIKW elements, data, information, knowledge, and wisdom, mean. The second problem is that there is no explanation of how these elements are of any use or create value. These two problems are related. It's impossible to create value when there is no agreement on what the individual elements are or what they do.

As academics have discovered when surveying the literature (Rowley 2007) and academics themselves (Zins 2007), the understanding of what these terms mean vary hugely and in many cases are so vague and incomprehensible as to be useless. Zins study of 45 top academicians produced 130 different definitions.

The different understandings are also contradictory. Data is defined as "Data are patterns with no meaning" (Aamodt and Nygård 1995) and "Data items are elementary and recorded description of things, events, activities, and transactions"(K. and Laudon 2006). Another contradictory definition has data only coming from human observation, while another has data coming from any source (Zins 2007).

The definitions of information are more varied and inconsistent than data. One definition of information requires the "inform" to have the element of being "informed" as in a news bulletin, while other definitions treat the "inform" in the Aristotelean version of being simply "a form" (Machlup 1983). There is also the claim that information is something our minds need to consume in

the manner our bodies need to consume food in order to live (Miller 1983). A common view of information is that it is simply a transformation of data. Information is usually defined as a passive thing. In some definitions, it is simply organized data.

What adds to the confusion is information theory in communications (Shannon and Weaver 1949). There information is the amount of news or “surprise” in a communication channel between a sender and receiver. This connects to the “inform” part of information. While this is called information transmission, the diagram that Shannon uses refers to a “message”. While Shannon calls it an “information source” that the message is transmitted from, it is simply a string of bits or data that is being transmitted.

The definitions of knowledge seem equally varied and equally confusing. Knowledge is defined as more information or better information or collection of information (Zins 2007). Knowledge is reflected in the questions of “how” and “why.” Knowledge gets defined as tautologically as what the knower knows or sidesteps any definition by claiming that it’s one step above information and one step below wisdom. There is even the Zen-like, knowledge is ‘no-thing’ (contrary to “information-as-thing”). There clearly is no consensus as to how knowledge fits into a hierarchy of DIKW.

To sum it up there is a quote from one of the most impressive books on DIKW, “one thing is certain from the literature, no consensus, no clarity, but plenty of confusion” (Mortazavian 1983).

The other core problem is there is no explanation of how these elements create value. None of the definitions in the material cited here discussed what the value of data, information, knowledge, and wisdom was except in very vague ways. There are references to understanding, awareness, negotiating, and making decisions, but that is the extent of value specificity.

There is also the issue that there are two different contexts of DIKW elements in academic literature. One context is logico-conceptual that attempts to examine and explain how humans think. The other context is from a computer information systems perspective (Frické 2019). Neither of these contexts are successfully described by the DIKW pyramid model.

It should be clear at this point that the DIKW pyramid is fatally flawed as a workable concept. The DIKW elements themselves are clearly of interest and are related in some fashion. However, given that there is clearly little or no agreement as to what these elements are, the simple, hierarchical pyramid model doesn’t really describe what these elements are or how they produce results of value. There is much, much more complexity here that gets glossed over by this simplistic DIKW pyramid model.

The way these DIKW words have been defined and used have resulted in a Gordian knot that promises to resist all efforts to untangle it. This paper will not attempt to untangle it. It will cut through the knot by proposing a new, consistent framework. It will not change the way we casually use these words in everyday speech but may provide a framework for understanding how we need to look at digital-based information constructs, such as Digital Twins (DTs).

### 3. Models of Thinking

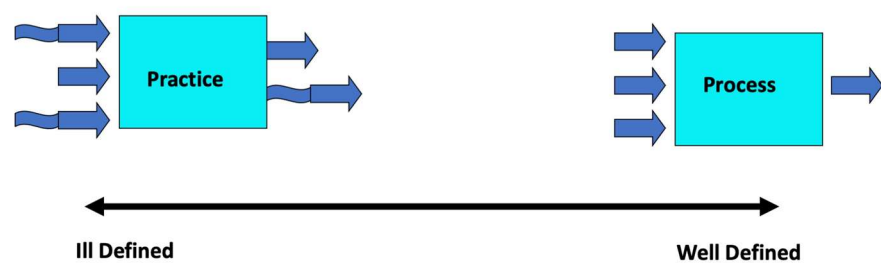
So, if we are going to discard the DIKW Pyramid Model, how are we going to define data, information, knowledge, and wisdom and their relationships with each other. It will be useful to step back and understand how we think about things and why.

The “why” is pretty important. The common things we humans think about is our goals and how to accomplish them. Much of human life is performing goal-oriented tasks. We have a goal. We think about how we can accomplish it. Since we don’t live in an environment of unlimited resources, especially time, we need to minimize the resources needed to accomplish our tasks.

Some of the goals are proactive – we would like to accomplish a new goal. Some of the goals are avoidance or remediation. We have a goal. We want to avoid a potential adverse event that will prevent us from accomplishing the goal. Worse, an adverse event has occurred, and we need to remediate the adverse event, or we will not accomplish the goal.

So how do we think about goal-oriented tasks like these? This section will explore two models that share a commonality. Those two models are System Thinking Model (STM), which is defined as System 1/System 2, and the Process and Practice (PnP) Continuum which is shown in Figure 2 and

that I introduced in the mid 2000s. The STM comes out of psychology (Stanovich & West/Kahneman), while PnP is a Product Lifecycle Model (PLM) and systems engineering perspective (Grieves 2006).



System Approach – Linking Inputs to Outputs

Figure 2. Process, Practice Continuum.

The left side of the PnP Continuum is a process version of systems. Process takes inputs, runs predefined routines that are invoked by those inputs, and produces an output or outputs. With processes, the routines will always produce every time the results that we desire. In a similar fashion, System 1 operates automatically and quickly, with little or no effort and no sense of voluntary control.

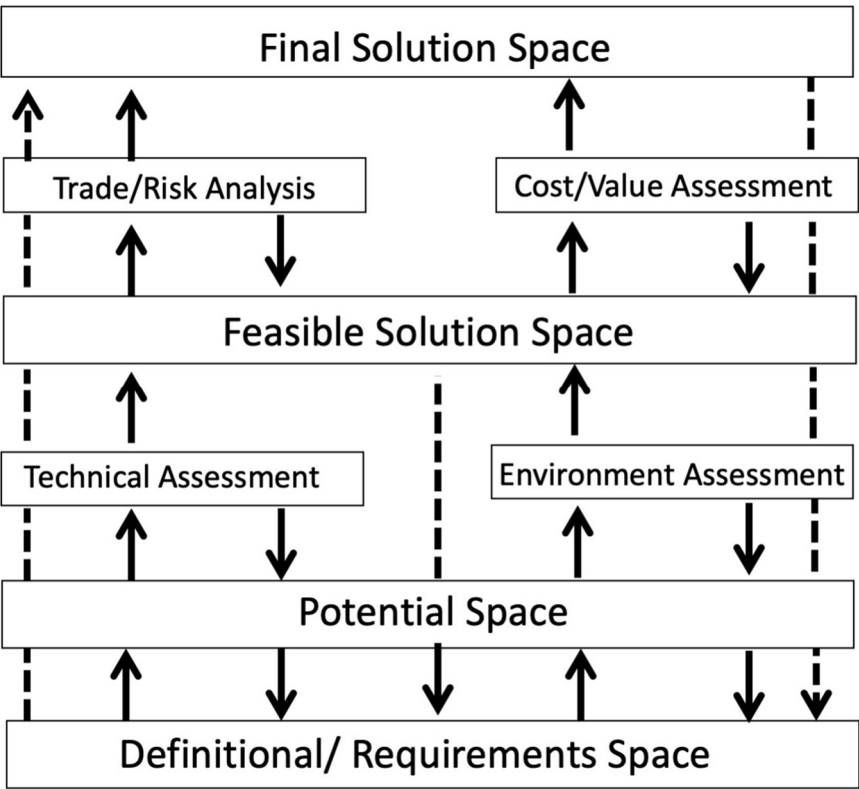
Process and System 1 thinking are fairly automatic. We take in inputs that trigger taking specific actions. We have predetermined that these actions given the inputs we have received will, in the worst case, accomplish our goals and in the best case will accomplish our goals using fewer physical resources than any alternative action.

The right side of the PnP Continuum is practice. It is presented with inputs and needs to determine the appropriate routines to perform in order to obtain the desired outputs. Similarly, System 2 seeks to explain why and how we need to make some decisions more slowly as compared to System 1 thinking. System 2 allocates attention to the effortful mental activities that demand it, including complex computations. “The operations of System 2 are often associated with the subjective experience of agency, choice, and concentration” (Kahneman 2011).

Practice and System 2 thinking is deliberate and more time consuming. We receive inputs. However, while it may be associated with actions, we do not immediately select an action. We want to examine the inputs, think about alternatives and options, run predictions of outcomes. We exhibit conscious voluntary control.

I have proposed a Practice Model Methodology as shown in Figure 3 (Grieves 2014). This is the type of holistic approach that humans engage in for Practice and System 2. Human thinking differs greatly from the sequential approach that computers use. We would like to believe that we think in a deliberate and sequential way. We would describe populating this model from the bottom to the top, with each space being fully populated before we moved to the next level. In our description, the definitional/requirements space would be completely populated before moving up to the potential solutions space. That deliberate methodology would continue in an orderly fashion until we reached the final solution space.





**Figure 3.** Practice Model Methodology.

The reality is very different. While there is this general movement from bottom to top, humans are neither methodical nor orderly. As we expect regularity in the world (Popper 2014), we anticipate outcomes well in advance of sequentially arriving at them. A time-elapsd photography of this model in motion over time would show the different spaces being populated and reduced simultaneously across the different spaces.

As soon as the definitional/requirements space started to take shape, the other spaces, including the final solution space would start to be populated. As humans, we do this automatically. We start to think about interim steps and final solutions as soon as we start to formulate problems. As new data arrives for any of the spaces. we perform time evolved simulation outcomes of new information we create and/or of selected different information from our knowledge repositories. We then adjust all the other spaces to reflect that.

The issue with this approach is that we can close the final solution space too early. We engage in confirmation bias by interpreting new data in terms of the final solution space that we have already populated. However, there are methodologies, such as the Kepner-Tregoe Method, that attempt to prevent that. On the positive side, simultaneously populating these spaces allows us to take the intuitive leaps that innovation requires. Often this holistic approach presents us with a solution that a sequential approach might have filtered out because we can go back and adjust all the spaces including the definitional/requirements space to allow a highly innovative solution to emerge and provide functionality that we did not know we needed until we invented it.

**4. Defining and Discussing Information, Data, Knowledge, and Wisdom**

Human and non-human life, which for wont of a better term we will call “nature”, have two different approaches to existence. Nature has only one goal which is survival. Because it has

unlimited time and resources, nature tries all possible combination by genetic mutation and lets the environment select out the best options. Nature has no concern for individual life forms.

Humans don't have the luxury of nature's approach and have immense concern for individual life forms, specifically their own. My underlying premise is that most of human life is to accomplish a goal while minimizing physical resources: time (labor and elapsed), material, and energy. The idea of understanding what a thing is by understanding what the thing does or is for goes back to the ancient philosophers (Aristotle and Lord 2013). I therefore view these terms, data, information, knowledge, and wisdom, by what they do to accomplish this underlying premise, not by what the words mean. I don't view DIKW as a hierarchy, but rather as components of this action.

My DIKW framework is relatively simple, but it includes action elements:

- Data is a fact or facts about reality and the input to create information: we collect and process data.
- Information is the replacement of wasted physical resources: we create and use information based on data.
- Knowledge is the repository of data and potential information: we store data and information in knowledge repositories for future use.
- Wisdom is a selection mechanism of information to accomplish a particular task goal: we employ wisdom to determine what data and information from our knowledge repository to use for accomplishing our goals.

I contend that the DIKW framework proposed here is necessary to understand why Digital Twins (DTs) and digital transformation in general are having the impact on not only product manufacturers, but on all aspects of society. Understanding what data, information, knowledge and wisdom "does" versus what they "are" will help us be better equipped to design capturing the appropriate data, processing it into information, storing it as knowledge, and wisely selecting the appropriate information to use to be more effective and efficient in obtaining our task's goal.

#### 4.1. Data

The basic component of the DIKW framework is data. Data are declaratory statements of facts about reality. Reality consists of physical things in the physical world that have attributes. Examples of attributes are identity, location, dimensions, current state status, color, etc.

Data is obtained by sensing the physical world. Humans input and process data through their senses. I have pointed out that our highest capacity sense is vision (Grieves 2006) Page 15. Humans have created instruments that sense the physical world and translate that sensing into a framework we have developed to enhance understanding. Instruments can capture reality phenomena that humans are incapable of sensing, such as infrared waves.

Data always has context. There is no such thing as data without context. That is simply noise. A string of numbers is meaningless and would not be considered data unless we could put it into a context. While we may collect and store this, until we can decide that these inputs are a fact about reality, it is noise, not data.

Data is in two basic forms: raw and processed. Data may or may not need to be processed into a form that leads to creating information. Some data may have all we need in its declaratory fact about reality. The data statement, "the car coming to my intersection is at such a speed that it is unable to stop." There is nothing we need processed here.

On the other hand, there are data about reality that needs processing so that we can create a case. If we obtain the data points on an hourly basis of 76, 77, 78, 79, we need to know the context that these are temperature readings. The processing it requires is to fit a line to it and produce the next hourly data point of 80. The fact about reality is that the temperature is increasing one degree each hour.

One compelling explanation of how humans deal with data is that we expect to find regularities in our world. So, upon obtaining data, we immediately create a theory that attempt to explain how the data is consistent with some causal or correlated effect that meets our task goal. As discussed in the next section, we then create an action or actions, the final solution space, that will have that



predicted effect occur if positive or prevent that effect from happening if negative. The human brain does this with processes of which we have only a rudimentary understanding (Amunts, Rowald et al. 2022, D’Angelo and Jirsa 2022).

4.2. Information

As I introduced in my first book on Product Lifecycle Management (PLM) (Grieves 2006), one of the most obvious but least articulated observations about information is that it acts as a substitute or replacement for wasted physical resources, such as time, energy, and materials for a goal-oriented task. Tasks, by definition, are actions. So, using information substitutes or replaces wasting physical resources in the process of performing the task. Information must have an action component to qualify as information. Information might be immediately used or stored and saved for later use. Facts about reality that are novel and interesting to us but have no action component are simply data.

4.2.1. Example of information value

Figure 4 is illustrative of a simple example, but it is going to give you an indication of the kind of value that we create with information. I have a matrix that is a thousand by a thousand. On a regular basis a gold bar is randomly placed somewhere in that matrix. We have a robot retrieval machine in the lower left-hand corner or (0,0).

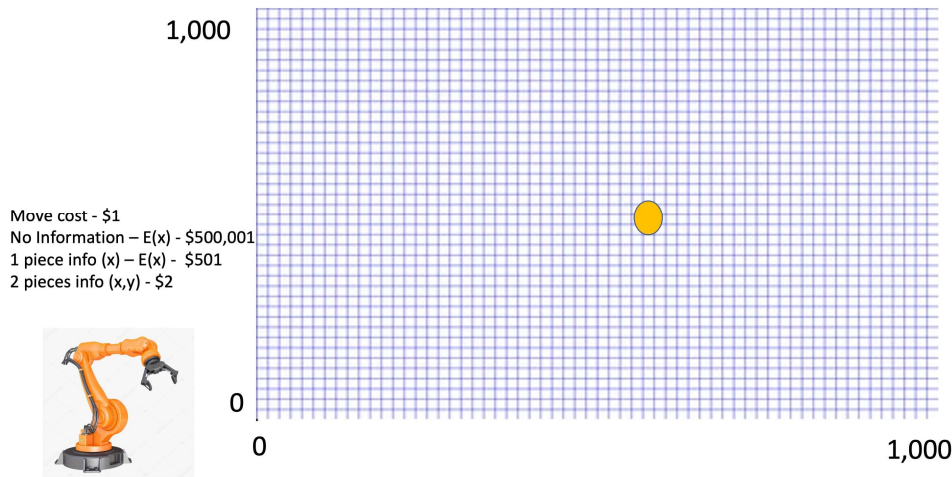


Figure 4. Example of Information value.

The robot must go out and find that gold bar and bring it back. The cost to move this robot each time is one dollar (\$1). It doesn’t matter how far the robot moves, but every move costs a dollar. If there is no information about where that gold bar is, the expected cost since the robots has to inspect square by square over all the times is \$500,001. The costs are \$500,000 to search and find the gold bar plus one dollar to retrieves the gold bar and move back to (0,0). If I have no data about the location of the gold bar, the expected cost is \$500,001.

Information is to obtain the data of the gold bar location and move directly to that location and search from there if necessary. If I have one piece of data, the X or Y value, I can go to the X or Y coordinate and start my search along the other axis. The costs are reduced to \$501. If I have two pieces of data, both the X and Y, the robot can simply move to that location. My costs are reduced to \$2.

This demonstrates that there is potentially a huge opportunity for the use of information in this simple example. I will contend that that these opportunities on more complex tasks exist every day as we trade off information for wasted physical resources. In this case, I reduced my costs from \$500,001 to \$2 with information driven by two simple data points. However, without the ability to perform an action and become information, data points provide no value.

But operations cost is not the only factor here. There is also cost in terms of clock time. If it takes 10 minutes for the robot to search each matrix location, then on average it takes approximately 9 years and 6 months to find the gold bar. With information it takes 10 minutes. Information replacing wasted elapsed clock time may be far, far more valuable than any other physical costs. It can be the difference between a task goal being feasible and unfeasible.

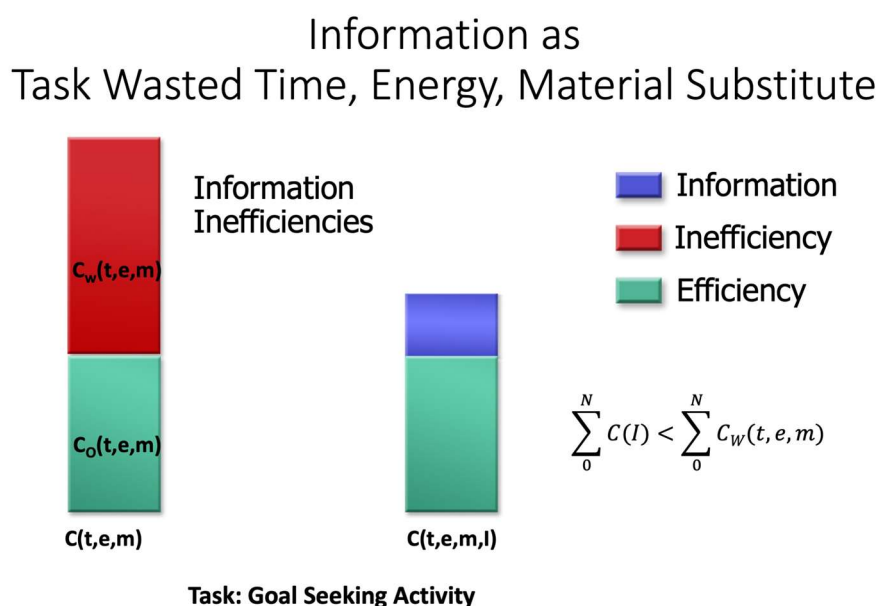
What is not included as costs in this example is instrumenting this matrix to detect if it has the gold bar. This would be an IoT device, and we would have to instrument a million places with a one-time cost being incurred. That cost needs to be included when we evaluate if we should incur it over all the times we search.

#### 4.2.2. Information as a substitute for wasted physical resources

More generally, most of human existence involves performing goal-oriented tasks minimizing the expenditure of physical resources needed to perform that task successfully. The physical resources we have at our disposal is time, both labor hours and elapsed time, energy, and materials. For any given goal-oriented physical task, we can divide the task into two categories of resource usage.

The first category of resource usage is the minimum expenditure of physical resources that if we were omniscient and omnipotent that we would need to perform the physical task. This category is the minimum of resources we would utilize to successfully complete the task if we knew the actions that we needed to take and that we could execute those actions perfectly. This category is always subject to constraints of what we will do (moral) and what we can do (physical and legal). The second category of resource usage is the remainder of the physical resources that we actually use to perform the task. These are wasted resources.

The left side of Figure 5 illustrates this relationship. The lower or green part of the bar represents the minimum expenditure of physical resources needed to complete a task. The upper part of the bar represents the information inefficiencies or wasted resources that are expended over and above that minimum expenditure of resources. The complete bar represents the total physical resources expended to complete the task.



**Figure 5.** Information as Task Wasted Time, Energy, Material Substitute.

We can measure the physical resources, time, energy, and material, expended on performing this physical task. However, we can't simply add units of time, pounds of material, and units of energy together. Because we live in a capitalistic society, we can aggregate the different types of physical resources by costing the different resources and aggregating the costs.

The right side of Figure 5 shows the role of information. The minimum expenditure of physical resources to perform the task efficiently and effectively does not change. However, information can substitute or replace the wasted resources. We said above that if we “knew” the actions we needed to take or not take, that is what we would do. The use of information is how we know what actions to take and not to take.

The issue we have with information is how do we cost it? We do not have units of information like we have for physical resources. In spite of having no unit of measurement, information has a cost. While we cannot measure those costs in units, we can quantify the costs of hardware, software, and labor costs to develop information.

The conditions under which this substitution of wasted physical resources by information holds true is indicated by the formula,  $C(I) < \sum C_w(t,e,m)$ , where  $C_w$  is the cost of the wasted resources in the upper left bar. This represents that the cost of information is less than the cost of wasted resources for all the times the task is performed.

In our opening example, the cost of information would be instrumenting, collecting, and maintain the IoT devices in the million locations. That would have to be compared against the time and expense of the robot performing unnecessary and wasted searches.

Figure 5 shows information replacing all the wasted resources. This is the ideal. This probably does not happen except in fully automated tasks. However, since the potential for wasting resources is infinite<sup>1</sup>, information can and has substituted for task wasted resources.

It is also important to understand that information is a non-rival good (Benkler 2006). Information is a resource that can be used over and over again. It is an asset and not an expense like the wasted physical resources this information replaces. However, for this to be the case, this information needs to be captured, organized, and reused. Some organizations neglect that. They believe that the physical product that they develop is the asset. However, the information is their true asset.

This idea of using information as a replacement for wasted resources is an obvious yet fundamental use of virtual/digital products in place of physical products. It has a role in both lean efforts and in innovation.

#### 4.2.3. The creation of information

If “Information is the replacement of wasted physical resources: we use information”, where does information come from. How do we create information and from what?

We use PnP and System thinking to create information. We start with a goal or goals for a task that we want to accomplish. To do that, we need to begin with data or facts about the reality that we are dealing with. As noted above, we use the ability to rely on the regularity of the natural world. These are causes and effects. We use that what we know of cause and effects to examine different causes to see if there are effects that will meet our goal. If we don’t have direct cause and effects, we look for correlations.

This methodology can be purely mental or can be physical. We can use trial and error to produce information. This is sometimes called the Edison method, because this is the method that Edison used to invent the electric light bulb.

We can also use intellectual methods. Case base reasoning (Kolodner 1992) is one such method as discussed in the next section, simulation is another method (Grieves 2022). We can also use intellectual methods. Case base reasoning is one such method as discussed in the next section, simulation is another method.

---

<sup>1</sup> The information that a perpetual motion machine is not possible (task goal – create perpetual motion machine : action – don’t expend any resources ) still has not stopped the waste of resources trying to invent one. The information that the earth revolved around the sun and not vice versa was available for hundreds of years. That did not stop the waste of uncountable number of hours calculating the orbits according to the Ptolemaic theory. If a task is impossible to be do, the entire bar is red, i.e., all physical resources are wasted resources.

Two possible things happen with the information we create: we simply discard or forget the information, or we retain it as knowledge. Our ability, as humans, to retain information in our brains is not very well understood, is unpredictable, and haphazard. In the past, we attempted to retain information we deem important by capturing it physically such as papers and books. We now can capture information digitally.

#### 4.2.4. The use of information

We use information when we execute the routines or operations that that makes up information. Once we have identified the information that given the data and given the goal will reduce wasted physical resources by minimizing the use of physical resources. Given the type of task, this may mean we continually monitor the relevant data coming in and make adjustments as necessary. We will use both Processes/System 1 routines which will be almost automatic and Practices/System 2 methodologies which will be deliberate.

The efficiency and effectiveness of our use of information depends on our ability to execute. I have continuously said that the caveat for using information as a replacement for wasted physical resources is “if I were omniscient and omnipotent.” This means that I can successfully execute the actions. We need to select information not only on the value of the wasted physical resources it can replace, but also on the probability of our success in performing the action required. We might elect to use information that replaces less wasted physical resources but has a 100% probability of being executed successfully.

If we are unable to execute the actions sufficiently that are required by the information, we may have to readjust, which has its own costs. In addition, humans are subject to confirmation bias where they fit the data into a preconceived notion of what’s occurring and use the wrong information and perform the wrong actions. The examples of this are rife in human life.

#### 4.3. Fact and information differentiator

Under this framework, there is a differentiator as to whether something is data or information. Information involves data but requires a potential action that can replace wasted physical resources. If it simply a statement about reality, it is a fact. If action is attached to the data to substitute for wasted physical resources, it is information. What simply may be data for some individuals can be information for other individuals by the addition of actions. As a simple example, “A train derailment has shut down the Main Street railroad crossing” is data to most people. To the individuals who were about to use that crossing, adding the action, “so, take the Fifth Street railroad crossing to prevent being caught in a major traffic jam on Main Street” makes it information.

#### 4.4. Knowledge

Knowledge is a repository of facts and information. The premise that knowledge is a stock while information is a flow has merit (Machlup 1983). Information is action, while knowledge is a repository of information. Because we can count on regularity in the world, we can and want to reuse information that we have created before. We can take a situation where we had the identical set of data, execute the actions called for in the information we used to accomplish our goal, and be assured that we will have the same outcome. That means that another key aspect of information is the ability to store and reuse it in a repository. Those repositories are called knowledge.

Through much of history, humans were very much like nature. Their goal was survival. They pursued survival differently than nature, which simply tried all possible combinations and let the environment determine the survivors. Humans did have the ability to capture data and create information and to use information.

What humans had precious little of was knowledge – the ability to store the information that they created. They came equipped with their biological repository, their brains. But the ability to share information between individuals and especially across generations was extremely limited.

Until there was an ability to accumulate data and information beyond the life of an individual, survival was the goal that humans were pretty much limited to. Humans sought to create knowledge capabilities, first with oral techniques, then with written ones. While writing was invented by the Egyptians in about 4,000 BC (Russell 1945), it was not until the invention of the Gutenberg press that dramatically increased the ability to capture and disseminate data and information in the knowledge repositories of books. That was the state of the art for symbolic knowledge until the advent of digital computers.

However, there was another way humans captured and stored knowledge. This was in the form of machines. In fact, machines were called frozen knowledge (Boulding 1966). Machines are the repository of data and information. Machines substitute wasted physical resources with the information that is embedded in them. When machines have computers in them that can change their capabilities that could be considered “liquid knowledge.”

#### 4.4. Wisdom

While the word “wisdom” usually invokes wizened philosophers thinking universal thoughts, practical wisdom as illustrated by the Serenity Prayer at the beginning of this paper is a part of everyday life. However, that may be selling the philosophers short, because the Greeks even had a separate word for practical wisdom, *phronesis*. Wisdom is evaluated after the fact by a) the results the task has or hasn’t accomplished and b) the minimum amount of resources were used.

If the task was completed successfully, the actions taken were wise. If the task was not completed successfully, the actions taken were not wise. If the task was accomplished, but the amount of resources were not the minimum required, then we didn’t select or know the optimal information. The most important success criterium is accomplishing the task. While we would desire that the minimum amount of resources be expended, we have a great deal more latitude in determining what is acceptable waste. As the focus on “lean” manufacturing and other discipline functions show, we do continually try to reduce wasted physical resources.

So a priori, wisdom is a selection process. We will have a substantial amount of information in our knowledge repository that may or may not be substitutable for wasted physical resources for our goal-oriented task. Wisdom is a process that allows us to select that information which will accomplish our goal-orientated task while keeping wasted physical resources to an acceptable level.

Wisdom is a context dependent selection. We need to have the data about the environment we are presented with at the current point in time. We need to take that data at  $t_0$  and predict an outcome in the necessary amount of time we need results.

*Wisdom requires taking information candidates from our knowledge repository, predicting and/or simulating the effect the candidates would have in task accomplishment and waste reduction, and then selecting the best information candidate to execute. As humans, we do this constantly. We attempt to predict future outcomes on the basis of proposed actions we take. Unfortunately, we are cognitively (Simon 1996) limited, so we do not always do a good job, especially when situations are complex. This is where digital twins and their capabilities can assist us.*

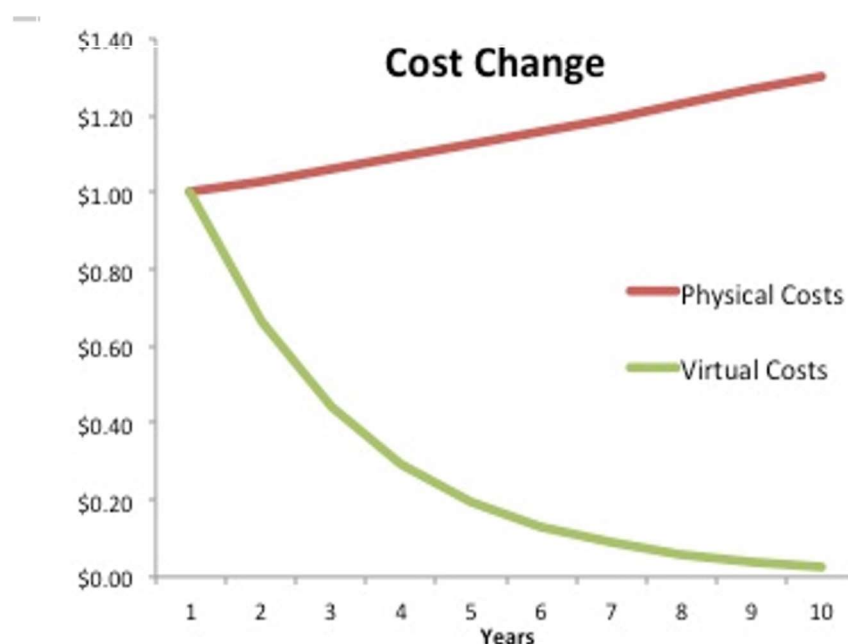
### 5. Digital Twins and Working in Digital Space

The advent and exponential advancements from the middle of the last century to now of digital computers increased computing and communication capabilities in an almost unfathomable way. We can now create a representation of our physical world and the objects in it in digital space. We can collect, process, and store data in amounts never even imaginable earlier. We can create and use information far more cheaply and effectively than in the past. We can build vast knowledge repositories, and deploy program algorithms that enable more effective and efficient task decisions..

For tasks that are repeatable over and over again, replacing wasted physical resources with computing-based information is an attractive proposition. In fact, it becomes more viable over time. This is because information technology decreases at an exponential rate, while physical resource costs



increase at the rate of inflation. Figure 6 shows the impact of this over time (Grieves 2023). The caveat is that, for simple, one-time tasks, wasting resources through trial and error may be more economical.



**Figure 6.** Physical vs. Digital Costs.

This graph also shows that even if physical versus digital costs are currently the same, this changes very dramatically over time. For tasks that will be repeated continually over the years, the advantages of using information versus physical resources widens quickly and substantially.

This dynamic means that there are compelling reasons to move work from the physical world to the digital/virtual world. A major concept that enables this transition being rapidly implemented is the concept and application of Digital Twins.

### 5.1. Digital Twin Model

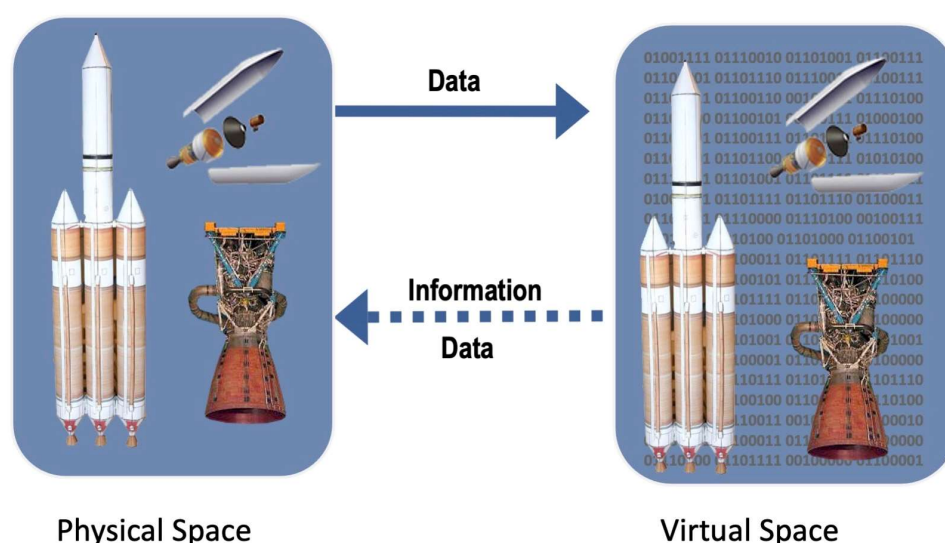
As of today, the model in Figure 7 is the accepted model of the Digital Twin. While definition may vary and vary widely, images usually show Digital Twin representations that are fairly consistent in the representation of physical space and products, digital space and products, and a two-way connection between them (Grieves 2011).

The commonly accepted Digital Twin Model is based on the one that I introduced in 2002 (Grieves 2002) and was the underlying premise of my work on Product Lifecycle Management (PLM) (National Academies of Sciences 2023). The model has been reduced to the one as shown in Figure 7 and consists of three main components:

- The physical product in our physical environment
- The DTs in a digital/virtual environment
- The connection between the physical and virtual for data and information



# Digital Twin Model



**Figure 7.** Digital Twin Model.

On the left side are physical products in physical space. These are the Physical Twins (PTs). On the right side are digital/virtual products, which we now refer to as Digital Twins, in digital/virtual space that is the environment that the counterparts of physical products operate in. The third element is the communications connecting the two spaces and products, with data from physical space and products populating the digital/virtual space and products, and data and information coming back from digital/virtual space and products to be used in the physical space.

The digital/virtual space on the right was originally referred to as the Digital Twin Environment (DTE). Subsequently, it is being called the Digital Twin Metaverse (DTM) (Grievess and Hua 2024 Forthcoming). It needs to be populated with the rules or laws of the physical universe or the subset of those rules that will support the use cases that specific DTs need to support.

## 5.2. Types of Digital Twins

There are three types of digital twins: the Digital Twin Prototype (DTP), the Digital Twin Instance (DTI), and the Digital Twin Aggregate (DTA). They are intended to span the entire lifecycle of the product. DTPs are developed and used in the product creation phase and persist through the follow on phases of build, operate/support, and dispose. DTIs are created in the manufacturing phase and are aggregated as DTAs for use in the operation and support lifecycle phase.

The DTP is the digital twin that comes into existence before there is any physical version of a product. While general models exploring aspects of a potential product or product class may have been created, the DTP does not exist until there is a decision to fund the development of a new product and begin work on the product. The ideal is that there will be no physical version of the product until all the details of a product's geometry, behavior, means of production, and lifecycle aspects have been fully defined. The ideal is to develop the product virtually, test the product virtually, manufacture the product virtually, and operate and support the product virtually, with atoms only being deployed when the product has been perfected (Grievess 2015).

The DTI originates when an individual version of the product is manufactured and assembled. The DTI is the as-built of the product, capturing all the relevant data about the product as the product is produced. The DTI will be linked to its individual PT for the life of the PT and will even exist beyond the PTs removal from service and its disposal. During the PT's life, the DTI will be updated

to replicate the data of the changes to the PT. At its ideal, any information about the PT can be obtained by issuing a query to its DTI. Practically, this will be dependent on the use cases.

AS the name implies, the DTA is an aggregate of all the data from the population of DTIs. The DTA will give a composite picture of the variation of DTI geometries and behaviors. A primary use of the DTA is to be able to calculate Bayesian based predictions for individual DTIs and to provide longitudinal learning for new products based on the collected performance degradation of older products. The structure of these DTA repositories will depend on use cases. Information will be generated both in the DTM or back in physical space. In physical space, human minds will be the producers of information.

## 6. Applying the DIKW Framework to Digital Twins

The discussion thus far has been oriented to DIKW as being the purview of human faculties and capabilities. However, limiting these concepts to human capabilities greatly limits the potential value of DIKW. Humans have limited memory and computing capability. Because human cannot share their brains, humans in the past had to resort to inefficient physical artifacts, such as books, to share data and information as their knowledge repositories.

Up until the middle of the last century, humans and their brains were the only computational and thinking “machines” in existence. The development of computers changed that. Assessing computer capability utilizing DTs against DIKW elements in reverse order, we can make the following statements:

- Wisdom – DTs can select information to accomplish a particular task goal from its knowledge repository.
- Knowledge – DTs either store data and information in their own knowledge repositories or can access data and information in other computer’s knowledge repositories.
- Information – DTs can recommend and even use information based on data as a replacement for wasted physical resources.
- Data – Data can be collected, processed and organized into DTs.

DTs uses all four DIKW elements: data, information, knowledge, and wisdom. In order to replicate physical objects and their physical environment, DTs need facts about reality. That is data. The data can be raw or processed. In some cases, DTs will receive a stream of raw data that it will use directly, process to use, or simply organize or store. Data that needs no processing because it is a fact about reality is complete enough that it can be acted on is also “processed” data.

As the DT model indicates, data comes from the PTs and environment of physical space on the left side. On the right side, data is collected, processed, and organized in the DTIs and the DT Metaverse. Some of the data will be processed into information by the rules set up in the DT Metaverse. This means that there is a potential action that can be taken as a substitute for wasting physical resources.

This makes the DTM and its DTs a knowledge repository. It is our stock of data and information that is required by our definition of knowledge. However, this knowledge repository provides no value unless it can be conveyed back to the physical environment. There are a number of ways that can be accomplished. This is indicated by the arrow from the right side to the left side with data and information.

1. Inquiries from the physical environment
2. Alerts to the physical environment
3. Commands sent from the DTI to its PT.

### 6.1. Inquiries from the physical environment

The DTI responds to inquiries from the physical space. The request can come from a human or from the PT corresponding to the DTI. The DTI can access the knowledge repository of the DTA to look for correlations of actions that have happened in the population of all DTIs. The DTI can use Bayesian probabilities to give the inquirer an indication of the probability of actions that may occur

given the data the DTI and the DTA have. The DTI can also simply supply the data it has so that action is determined in physical space, or it can provide information recommending action based on the data conditions that are developed as a result of the inquiry.

Humans take complete control of actions under this scenario. Humans can add information from their internal knowledge repository, i.e., their brain, or from external knowledge repositories, i.e., books, standards, guides, etc. Humans can use the information recommendation from the DTI or develop their own information.

### *6.2. Alerts to the physical environment*

Based on the DTI monitoring data coming from the PT and the DTM, the DTI sends an alert to the physical space. The alert can be generated because the data is an indication of an anomaly itself. An example of that is data that an airbag in a vehicle has deployed.

The alert could also be triggered by multiple data points that are collected and assessed against data and information in the DTA knowledge repository. Based on Bayesian probabilities the incoming data indicates that there is a probability of a future anomaly occurring. An example of this is predictive maintenance that is triggered by current sensor readings correlating with previous component failures.

The alert can go to the PT or somewhere else such as a human or another system that is set up to collect such alerts. The alert can be in the form of data providing a fact about reality or it can be information with a recommendation of action to take to reduce wasted physical resources.

### *6.3. Commands sent from the DTI to the PT.*

Based on its programming, the DTI can send a command to its PT to invoke an action. This is obviously information. It is a human-not-in-the-loop so needs to be very deterministic. In this case, the information, knowledge, and any additional data needs to reside in the DT and DTM knowledge repositories.

## **7. Applying the DIKW Framework to Intelligent Digital Twins**

The Intelligent Digital Twin (IDT) was introduced in 2018 (Grieves 2018, Grieves 2022) to explain the role that AI would have in both assisting Digital Twins in their performance and in dealing with the increasing system complexity and emergent behavior of products themselves. The view here was that AI was not a replacement for humans but an augmentation for humans. IDT specifies four attributes for Intelligent Digital Twins as active, online, goal seeking, and anticipatory.

The characteristic of anticipatory requires that the IDT be constantly running simulations to look ahead into the future for its PT. It is intended to be a predictor for anomalies or potential failures that will hinder or prevent obtaining task goals.

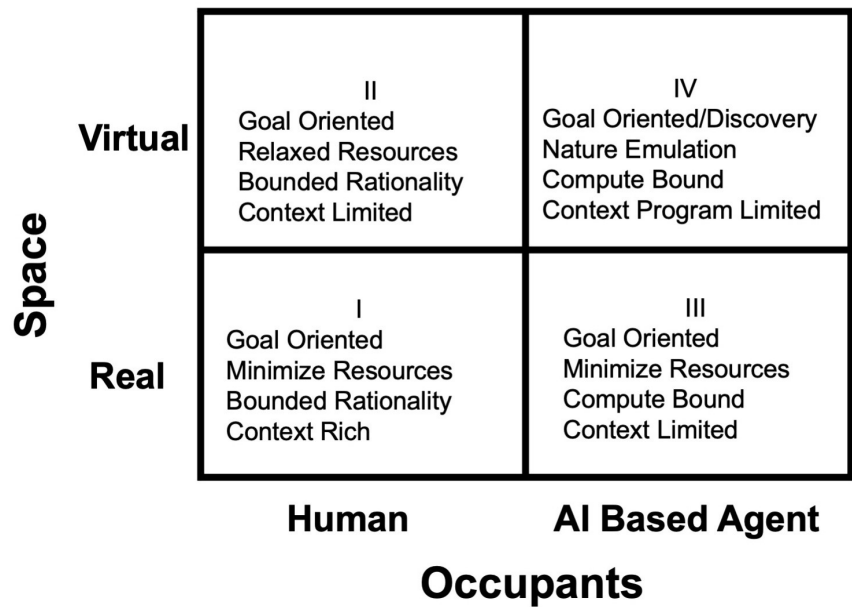
Computers can only do the above if they have been programmed by humans to do so. There has been claims about computers exhibiting emergent behavior. However, as I have pointed out, computer programs did not really exhibit emergent behavior (Grieves and Vickers 2017). Their programming was set so given the particular data and sequence, the program was always going to produce that behavior or output. We just didn't realize it.

Until programs had an ability to modify their programming or make other than if..then decisions, there would be no emergent behavior. That also means that the one thing missing from the DIKW list above is that computers could not create information. AI has those abilities and can create information.

When we refer to AI, as we do here, what are we referring to are computer agents that employ AI technology in order to search data, correlate it with data and information of subsequent outcomes with probabilities, and possibly create new information. I have long proposed that one of the characteristics of DTs is "Cued Availability" (Grieves 2006) pgs 91-93. Cued Availability can be described as an AI based agent assessing data coming in from the DT and its environment, assessing and simulating the possible states that could occur, and cueing us with information, i.e., the

probabilities of the future states that could occur and actions that we can take to complete our tasks successfully.

AI has unique and potentially powerful capabilities, as indicated by Figure 8, as shown in the matrix of human and AI based agent occupants of real and virtual spaces. The upper right quadrant of AI and virtual space is the natural habit of AI. As stated earlier, nature tries all possible combinations and has the environment select the best outcomes.

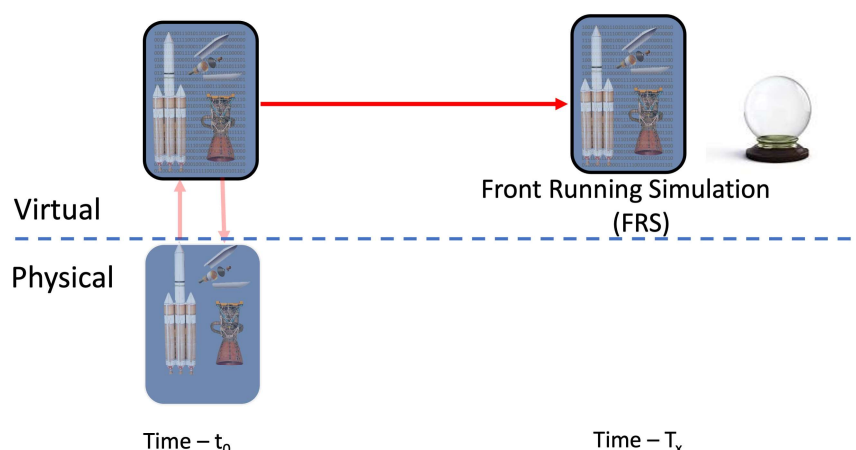


**Figure 8.** Occupants vs. Space.

Subject to compute capabilities, AI based agents can be very much like nature in developing a final solution space given the definitional/requirements of the task goal. AI can methodically work its way up the Practice Model Methodology in Figure 3. AI based agents can create an exhaustive potential solution space, do a technical and environmental assessment to derive a feasible solution space, perform trade/risk analyses and cost/value assessment, and come up with an exhaustive final solution space. It can apply Bayesian probabilities using the data and information from the DTA and provide the best alternatives to humans in real space.

AI based agents are time unconstrained in virtual space. This means that AI based agents can run time evolved simulations to determine, at least probabilistically, the outcomes of executing the information that the AI based agent has created and selected for the task goal as potential, feasible, and final solution spaces.

As Figure 9 shows, I have proposed that this capability as what I have called Front Running Simulation (FRS) (Grieves 2017, Grieves 2019). What FRS continually does is take in what we have identified as the relevant data on a continual basis. At every cadence of time zero, an AI based agent takes the data from its DTI and the environment. It then performs the Practice Model Methodology that also uses additional data and information from knowledge repositories it has access to. Wisdom is in the form of commands to PTs or Bayesian probabilities to humans who will then make the necessary decisions.



**Figure 9.** Front Running Simulation.

## 7.0. Conclusion

DIKW, data, information, knowledge, and wisdom, are fundamental to our existence as humans. The DIKW pyramid model, where data is a subset of information, which is a subset of knowledge, which is a subset of wisdom is attractive, but doesn't stand up to scrutiny as a conceptual model. However, the core problem is that there has been no definitional agreement of DIKW. The focus of definitions has been almost all on trying define what these elements are and not on what we do with them in how we think.

How we think is explained by Process/System 1 and Practice/System 2 concepts. We think and act automatically using predetermined mental routines, and/or we think deliberately, although not in a sequential fashion.

The paper has proposed these definitions that centers around information being a replacement for wasted physical resources in goal-oriented tasks.

- Data is a fact or facts about reality and the input to create information: we collect and process data.
- Information is the replacement of wasted physical resources: we create and use information based on data.
- Knowledge is the repository of data and potential information: we store data and information in knowledge repositories for future use.
- Wisdom is a selection mechanism of information to accomplish a particular task goal: we employ wisdom to determine what data and information from our knowledge repository to use for accomplishing our goals.

While our mental hardware is completely different from computer hardware, we can apply DIKW to Digital Twins. DTs allow us better abilities to deal with DIKW in terms of access, efficiency, and effectiveness. With the addition of AI based agents, we enable an Intelligent Digital Twin which will be more like nature in trying all possible combinations through simulation to find the highest probably of task success. The ideal will be Front Running Simulation (FRS) that will be our assisting agent taking the data from the physical world at every  $t_0$  and giving us the best information from the knowledge repositories to help is wisely select the most probable course of action for our task goals.

**Conflicts of Interest:** The author declares no conflict of interest.

## References

- Aamodt, A. and M. Nygård (1995). "Different roles and mutual dependencies of data, information, and knowledge — An AI perspective on their integration." *Data & Knowledge Engineering* 16(3): 191-222.



- Ackoff, R. L. (1989). "From data to wisdom." *Journal of applied systems analysis* 16(1): 3-9.
- Amunts, K., A. Rowald, S. Petkoski, M. V. Sanchez-Vives, M. Axer, G. De Bonis, C. Pennartz, J. Bjaalie, A. Brovelli and J. Fousek (2022). The coming decade of digital brain research-A vision for neuroscience at the intersection of technology and computing, Computational and Systems Neuroscience.
- Aristotle and A. Beresford (2020). *The Nicomachean ethics*. London UK, Penguin Books.
- Aristotle and C. Lord (2013). *Aristotle's Politics*. Chicago, The University of Chicago Press.
- Benkler, Y. (2006). *The wealth of networks : how social production transforms markets and freedom*. New Haven Conn., Yale University Press.
- Boulding, K. E. (1966). "The Economics of Knowledge and the Knowledge of Economics." *The American Economic Review* 56(1/2): 1-13.
- D'Angelo, E. and V. Jirsa (2022). "The quest for multiscale brain modeling." *Trends in Neurosciences*.
- Frické, M. (2019). "The knowledge pyramid: the DIKW hierarchy." *Ko Knowledge organization* 46(1): 33-46.
- Grieves, M. (2000). *Business is war: An investigation into metaphor use in Internet and non-Internet IPOs* EDM diss., Case Western Reserve University.
- Grieves, M. (2002). *Completing the Cycle: Using PLM Information in the Sales and Service Functions* [Slides]. SME Management Forum, Troy, MI.
- Grieves, M. (2006). *Product Lifecycle Management: Driving the Next Generation of Lean Thinking*. New York, McGraw-Hill.
- Grieves, M. (2011). *Virtually Perfect: Driving Innovative and Lean Products through Product Lifecycle Management*. Cocoa Beach, FL, Space Coast Press.
- Grieves, M. (2014). Process, Practice, and Innovation. *Global Innovation Science Handbook*. P. Gupta and B. Trusko, McGraw Hill Professional: 159-172.
- Grieves, M. (2015). "Can the digital twin transform manufacturing." from <https://www.weforum.org/agenda/2015/10/can-the-digital-twin-transform-manufacturing/>.
- Grieves, M. (2017). "White Paper: Driving Digital Continuity in Manufacturing." Retrieved August 1, 2017, from <http://research3.fit.edu/camid/research.php>.
- Grieves, M. (2018). The Evolution of the Digital Twin. *IM+io Best and Next Practices*. Germany, AWSI Publishing: 66-69.
- Grieves, M. (2019). Virtually Intelligent Product Systems: Digital and Physical Twins. *Complex Systems Engineering: Theory and Practice*. S. Flumerfelt, K. Schwartz, D. Mavris and S. Briceno, American Institute of Aeronautics and Astronautics: 175-200.
- Grieves, M. (2022). Don't 'Twin' Digital Twins and Simulations. *EE Times*.
- Grieves, M. (2022). "Intelligent digital twins and the development and management of complex systems." *Digital Twin* 2(8).
- Grieves, M. (2023). "Digital Twin Certified: Employing Virtual Testing of Digital Twins in Manufacturing to Ensure Quality Products." *Machines*.
- Grieves, M. and E. Hua (2024 Forthcoming). Defining, Exploring, and Simulating the Digital Twin Metaverses. *Digital Twins, Simulation, and Metaverse: Driving Efficiency and Effectiveness in the Physical World through Simulation in the Virtual Worlds*. M. Grieves and E. Hua, Springer.
- Grieves, M. and J. Vickers (2017). Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems. *Trans-Disciplinary Perspectives on System Complexity*. F.-J. Kahlen, S. Flumerfelt and A. Alves. Switzerland, Springer: 85-114.
- Hely, T. A., D. J. Willshaw and G. M. Hayes (1997). "A new approach to Kanerva's sparse distributed memory." *IEEE transactions on Neural Networks* 8(3): 791-794.
- K., L. and J. Laudon (2006). *Management Information Systems: Managing the Digital Firm*. Upper Saddle River, N.J., Pearson Prentice Hall.
- Kahneman, D. (2011). *Thinking, fast and slow*. New York, Farrar, Straus and Giroux.
- Kanerva, P. (1988). *Sparse distributed memory*, MIT press.
- Kolodner, J. L. (1992). "An introduction to case-based reasoning." *Artificial intelligence review* 6(1): 3-34.
- Machlup, F. (1983). Semantic quirks in studies of information. *The Study of information : interdisciplinary messages*. F. Machlup and U. Mansfield. New York, Wiley: 641-671.
- Miller, G. (1983). Informavores. *The Study of information : interdisciplinary messages*. F. Machlup and U. Mansfield. New York, Wiley.
- Mortazavian, H. (1983). On System Theory and Its Relevance to Problems in Information Science. *The Study of information : interdisciplinary messages*. F. Machlup and U. Mansfield. New York, Wiley: 641-671.
- National Academies of Sciences, E., and Medicine. (2023). Foundational Research Gaps and Future Directions for Digital Twins. Washington, DC, National Academies of Sciences, Engineering, and Medicine.
- Popper, K. (2014). *Conjectures and refutations: The growth of scientific knowledge*, routledge.
- Rowley, J. (2007). "The wisdom hierarchy: representations of the DIKW hierarchy." *Journal of information science* 33(2): 163-180.
- Russell, B. (1945). *A History of Western Philosophy*. New York, NY, Simon and Schuster.



- Seung, S. (2012). Connectome : how the brain's wiring makes us who we are. Boston, Houghton Mifflin Harcourt.
- Shannon, C. E. and W. Weaver (1949). The mathematical theory of communication. Urbana,, University of Illinois Press.
- Simon, H. A. (1996). The sciences of the artificial. Cambridge, Mass., MIT Press.
- Zins, C. (2007). "Conceptual approaches for defining data, information, and knowledge." Journal of the American society for information science and technology 58(4): 479-493.