Why Is There Data?
By David Sisson and Ilan Ben-Meir

Introduction

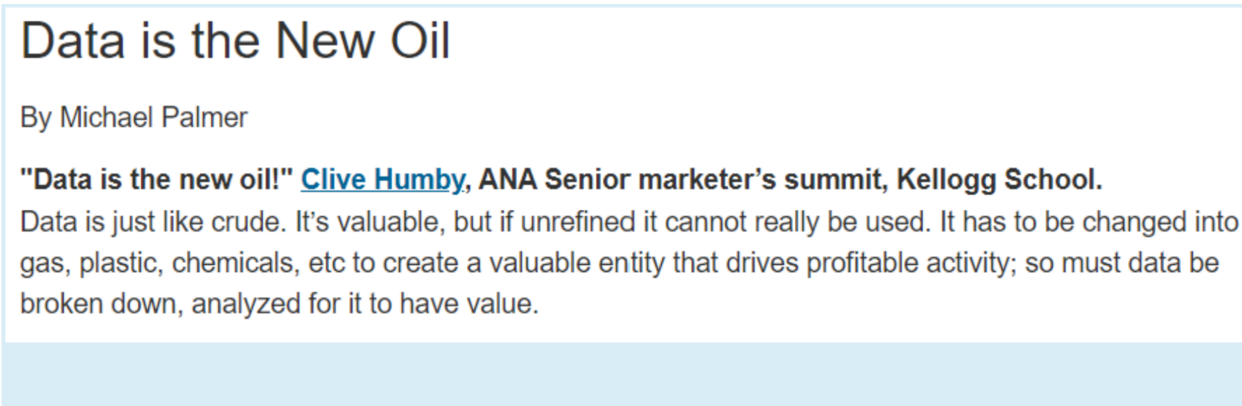
"Data is the new oil," Clive Humby declared in 2006.

Mr. Humby – in partnership with his wife, Edwina Dunn – had just successfully kickstarted the now ubiquitous brand loyalty card for Tesco, a supermarket chain headquartered in the United Kingdom, ushering in the age of "Big Data." Tesco's chairman at the time, Lord MacLaurin of Knebworth, was unequivocal about the success of Humby and Dunn's efforts, remarking that "[w]hat scares me about this, is that you know more about my customers in three months than I know in 30 years." Tesco rewarded Humby and Dunn for their accomplishment by finalizing its acquisition of the company the couple had cofounded in their kitchen, dunnhumby, for a reported £90M – so when Humby made his pronouncement about the value of data, he did so from a position of knowledge.

"Data is the new oil" has been a meme ever since. Like most such pithy statements, however, the meaning of this claim is determined by the context in which it is made – and Humby's full quote casts his remark in a somewhat different light than the abbreviated version. Although no publicly-available transcript of Humby's remarks exists, a blog post summarizing his comments presents them in (a potentially-paraphrased version of) their original context:

"Data is the new oil. It's valuable, but if unrefined, it cannot really be used. It has to be changed into gas, plastic, chemicals, etc., to create a valuable entity that drives profitable activity; so must data be broken down, analyzed for it to have value."

There is some ambiguity, at least in my mind, as to whether the full quote is properly attributed to Humby, or if data scientist Michael Palmer, who wrote the blog post, authored the fuller elaboration of Humby's initial claim (see Fig. 1). Whether one attributes the longer quote to Humby or to Palmer, however, the sentiment that it expresses is clear: In order for data to become *truly* valuable, truly *useful*, it must be **processed**.


Data is the New Oil

By Michael Palmer

"Data is the new oil!" Clive Humby, ANA Senior marketer's summit, Kellogg School.
Data is just like crude. It's valuable, but if unrefined it cannot really be used. It has to be changed into gas, plastic, chemicals, etc to create a valuable entity that drives profitable activity; so must data be broken down, analyzed for it to have value.

Figure 1: Michael Palmer quoting Clive Humby. Note only five words and an exclamation point are between quotation marks with attribution. There is some ambiguity in assigning credit for this full statement of the analogy between data between Humby and Palmer. Although the metaphor is clearly Humby's words.

The question animating this essay is thus a straightforward one: ***What sort of processing must data undergo in order to become valuable?*** While the question may be obvious, its answers are anything but; indeed, reaching them will require us to pose, answer – and then *revise our answers to* – several other questions that will prove trickier than they first appear. *Why is data valuable – what is it* ***for?*** *What* ***is*** *"data"? And what does "working with data" actually involve?*

Framing the Questions

The idea that data requires processing to become valuable is not new. More than sixty years ago, W. Edwards Deming (who, along with Joseph Juran, helped Japanese industry engineer Kaizen, it's quality-first business philosophy) said in an article published in the Journal of the American Statistical Association in 1942:

> Data are not taken for museum purposes; they are taken as a basis for doing something. If nothing is to be done with the data, then there is no use in collecting any. The ultimate purpose of taking data is to provide a basis for action or a recommendation for action. The step intermediate between the collection of data and the action is prediction.

Deming's words offer those who seek to understand the processes by which data is made valuable a powerful pointer in the right direction: The value of data is derived from data use in *making predictions* – and predictions are valuable insofar as they can be used as the basis for taking some action; if our predictions are accurate, then our actions will be more *effective* – which is to say, *the choices that we make will be more likely to result in the outcomes that we desire.*

At this point, it is possible to venture tentative first answers to some of our primary questions. Insofar as the value of data derives from the outcomes of the decisions that it helps us make, we can say that **what makes data valuable is its ability to help us reach a more accurate (albeit never perfect) understanding of the relationship between possible actions in the present and potential outcomes in the future that underlies any decision** – **put *very* loosely, data generates value by helping us to "improve the quality" of the predictions that structure the act of deciding on a particular course of action.**

This way of describing data's **purpose** – the *why* of data – also suggests an unconventional (but ultimately useful) working definition of "data" itself: **Data is the stuff out of which "better predictions" are made.** (We will return to the question of what "better predictions" actually *means* in due course.)

These tentative first answers are (at best) incomplete, but they go a long way toward bringing this essay's primary question into focus, helping us to see that one answers the question "***What sorts of processing must data undergo in order to become valuable?"*** by answering a much more approachable question: **"What is the process that transforms 'data' into 'better predictions?'"**

From "Data" To . . . "Wisdom"?

"Wisdom is borne of experience," the old saying goes – but what, exactly, *is* "wisdom"? At least colloquially, "wisdom" refers to something like "profound understanding," but the word also carries with it an implication of *calibration between expected and actual outcomes*. We call a choice "wise" if it leads to an expected and desired outcome. We call a choice "unwise" if, despite our expectations, it leads to an outcome we had hoped to avoid.

We can therefore say that "wisdom," as commonly defined, is closely related to "the ability to make better predictions" (even if we cannot yet say precisely *how*). The existence of this close relationship between "wisdom" and "better outcomes" serves the purposes of this essay well, because fields as diverse as Philosophy, Information Science, Knowledge Management, and Systems Theory have long shared an implicit epistemic hierarchy that begins with Data, ascends through Information and Knowledge, and culminates in Wisdom. This hierarchy, sometimes referred to as the Wisdom Pyramid (or simply as DIKW), thus offers a helpful outline that we can use to begin thinking through the process – or processes – of data processing.

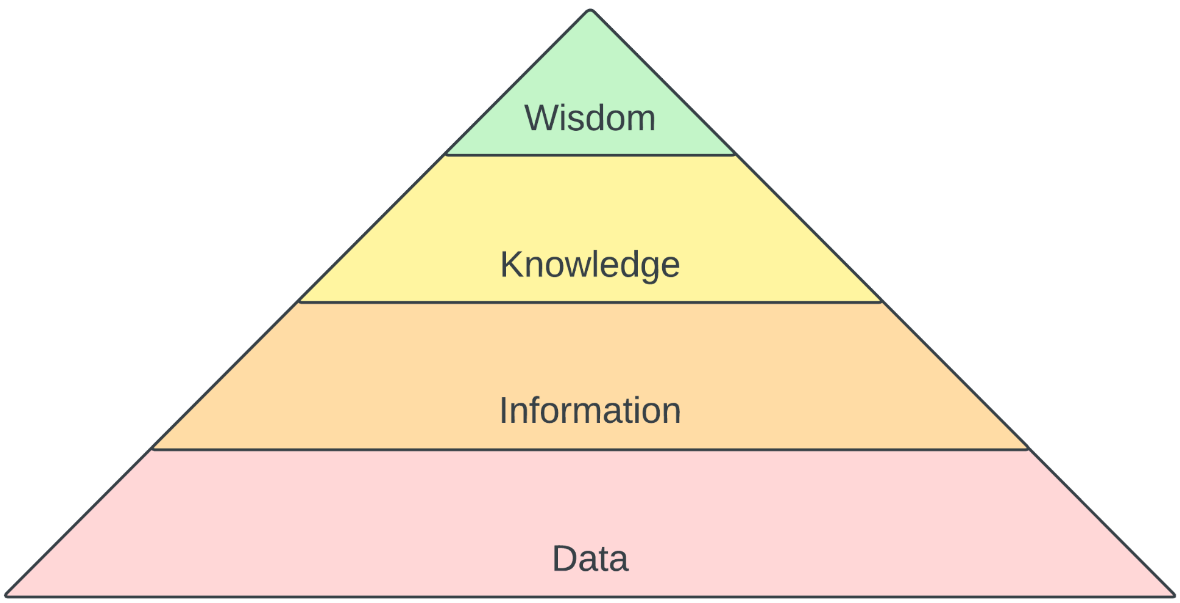


Figure 2: The "Wisdom Pyramid" showing the traditional, hierarchical arrangement of the terms "Data", "Information", "Knowledge", and "Wisdom" (abbreviated to DIKW).

The following section will therefore review the conceptual underpinnings of the Wisdom Pyramid, in order to examine the ways that this model – or something like it – can help us to answer the question of how "data processing" actually works.

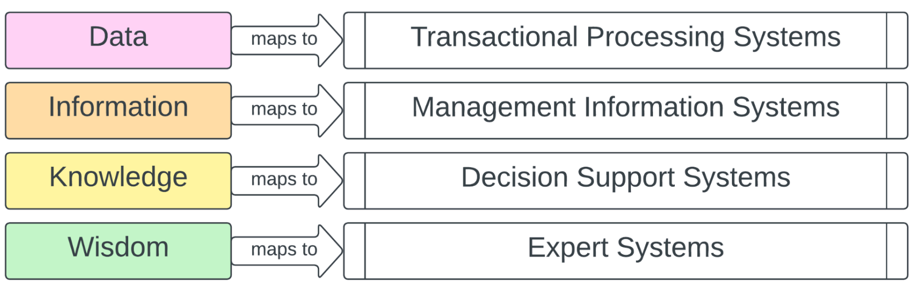
Within the discipline of Knowledge Management (at least), the formalization of the DIKW Pyramid is frequently attributed to Russell Ackoff's article "From Data to Wisdom," published in the Journal of Applied Systems Management in 1989 (Volume 9, pp. 3-9). It is worth noting that the full title of Ackoff's article is "From Data to Wisdom – Presidential Address to ISGSR, June 1988"; in other words, the attribution to Ackoff is based on an invited speech to a convention, rather than on a peer-reviewed academic paper.

Milan Zeleny, a contemporary of Ackoff, also discussed DIKW in an article published in 1987. According to Zeleny, the outcome of all of this processing is decision making (*i.e.*, coordination of action), echoing Deming's assertion that we gather data to provide a basis for action.

In a formal critique published in 2008, Martin Fricke concluded that DIKW might be better understood as a popular concept with illustrative utility, rather than as a formal theory. (The utility of DIKW can be brought more in line with Fricke's critique, however, if one defines "data" more generally than as "the output of a sensor," and if one views the DIKW framework as describing a set of interrelated but *not* strictly hierarchical processes.)

A 2014 paper on the interrelations between data, information and knowledge by David Williams makes a clear case for moving away from taxonomical models such as the DIKW pyramid, and instead adopting models with richer flows between concepts of data, information and knowledge. Williams echoes Fricke's criticism of limiting "data" to being the origin of the process of generating epistemic objects; both men cite Karl Popper's view that *hypotheses*, which determine what data one needs to collect, play a more basic role in this process than the data itself – which suggests that "knowledge" has at least some ground on which to contest claims of data's ontological priority. Williams concludes that DIKW is a simplification that is useful in highlighting certain aspects of a very complex flow. Specifically, data is more than just the base of the pyramid, and the relationship between data and knowledge is iterative.

Jennifer Rowley's article "The Wisdom Hierarchy: Representations of the DIKW Hierarchy" is the study of DIKW as a model that is most useful for the purposes of this essay. Rowley provides a review and analysis of forms DIKW has taken in textbooks in the fields of Information Management, Information Systems and Knowledge Management, and points out that some authors have also inserted "Understanding" and "Intelligence" into the DIKW hierarchy (at various levels). Rowley goes on to map each layer of the DIKW stack to various computer systems that have been developed to improve worker productivity; Figure 7 in her paper pairs each level of DIKW with a corresponding system type:



Rowley argues that automation is more useful toward the "data" end of DIKW, and that humans are necessary at the "wisdom" end, because wisdom is associated with doing the right things, which requires ethical judgment – a capacity that human beings possess, but automatons do not. Therefore, the "Expert Systems" that Rowley associates with wisdom are a class of decision-making systems that combine a set of rules for decision-making with the logic used to determine the situations in which a particular rule applies; these systems are designed to mimic the decision-making process of human experts, whom they relieve of having to weigh in each time a decision is required.

There are situations, however, in which an "Expert System" will be unable to render a decision – *e.g.,* when no rule fits the situation, when multiple (but conflicting) rules apply, or when the system simply cannot reach a decision within the available time – so human expertise is a required fallback. Decisions therefore remain anthropocentric, even when automated; in other words, people are still necessary at the wisdom end of the stack.

Ackoff shares Rowley's belief that wisdom necessarily involves ethics, noting in his 1988 address that "**Intelligence is the ability to increase efficiency; wisdom is the ability to increase effectiveness.**" In an interview with Phyllis Haynes, Ackoff (borrowing from management consultant Peter Drucker) elaborates on how he differentiates between these two terms, explaining that "Efficiency is doing things right; effectiveness is doing the right things." As we have suggested above, the measure of "wisdom" is *effectiveness* ("doing the right thing"), and both Drucker and Ackoff here mean "right" at least partially in the ethical sense ("right vs. wrong"). The measure of "intelligence," on the other hand, is *efficiency* ("doing the thing right"); here, Drucker and Ackoff mean "right" not in the ethical sense, but as a synonym for "operationally correct." Under these definitions, automated systems can arguably be considered "intelligent" – but such systems cannot be ethical, so they also cannot be "wise".

Rowley concludes her study of DIKW with a key insight:

> The [DIKW] hierarchy is only mentioned explicitly in a few books, but it is implicit in the definitions of data, information, knowledge and wisdom across all books. Typically information is defined in terms of data, knowledge in terms of information, and wisdom in terms of knowledge. However, there is less consistency in the description of the processes that transform elements lower in the hierarchy into those above them, and some consequent lack of definitional clarity.

Rowley's conclusion importantly identifies the dynamic aspects of DIKW: **There are processes that generate information from data, knowledge from information, and so on. Crucially, however, theorists have not yet defined these "transforms" between the layers of DIKW as clearly as we have defined the layers themselves.**

In short, DIKW **does not** equip us with a sufficient vocabulary to understand the "how" of data processing, but it **does** provide us with a basic model that we can then modify, dynamize, reorient, and otherwise elaborate upon, in order to get there.

First Revision: From DIKW → DIKUD

The first step in our revision of DIKW consists of modifying the basic Pyramid model to account for some of the criticisms and relevant insights discussed above, as shown in Figure 2 (below). Our updated pyramid retains the terms “Data,” “Information,” and “Knowledge” from the traditional DIKW hierarchy shown in Figure 2 above, but follows the lead of others (including Ackoff, as Rowley notes in her review) in inserting a level identified with “Understanding” between “Knowledge” and the pyramid’s apex. Following Deming’s interpretation of the purpose and teleology – the *why* – of data, we remove “Wisdom” from its place at the top of the Pyramid in the traditional model, and replace it with “Decision”; this change is motivated in part by a desire to emphasize the dynamic and context-dependent nature of the data processing system being architected, and in part by reasons that we will return to at the appropriate time.

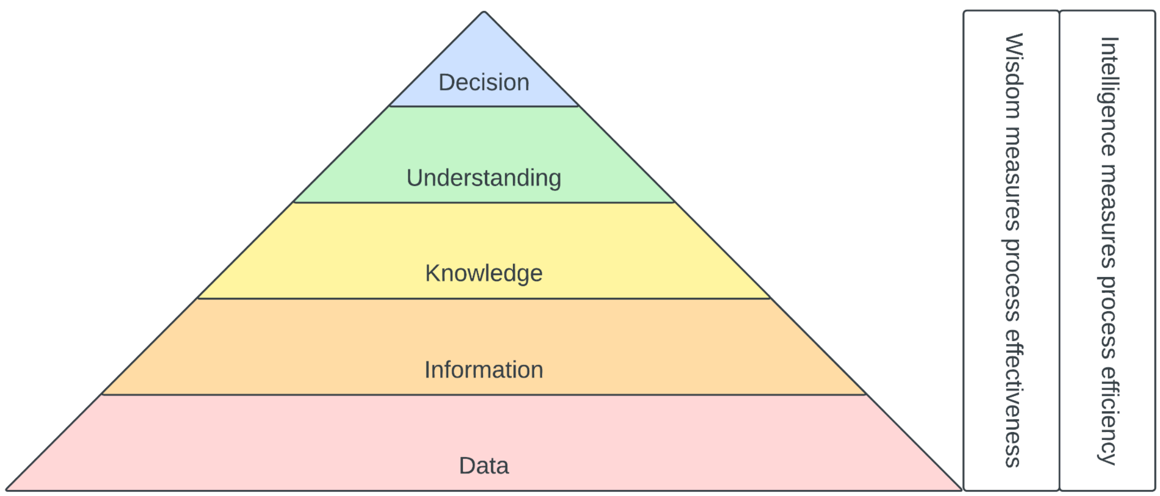


Figure 3: DIKUD modification of the DIKW Pyramid, including intelligence and wisdom as scales with which to measure process efficiency and effectiveness

It would be foolish, however, to jettison “Wisdom” from this ontology entirely. Therefore, taking inspiration from Ackoff and Drucker, our revision instead repurposes “Wisdom” as a Key Performance Indicator (KPI)-like scale to measure the *effectiveness* of the overall process of data processing; likewise, “Intelligence” is framed as a KPI-like scale to measure its end-to-end *efficiency*.

In this model, both “intelligence” and “wisdom” are defined in the context of the full processing stack running from “data” to “decision.” Wisdom and intelligence are also, and probably more often, defined in the context of an individual actor (*i.e.* “That person is intelligent, but that other person is wise!”). History, however, is replete with examples of anthropocentric viewpoints being forced to give way to viewpoints of higher diversity. We therefore suggest conceptualizing “Wisdom” and “Intelligence” not as *qualities of actors*, but as distinct (but complementary) *ways of measuring* or *evaluating* “how well” the overall stack is processing data into actual (or actuated) decisions – in short, as scales for the “rightness” of data driven decision making.

Second Revision: Refining DIKUD Definitions

Once we have determined the basic components of our revised “DIKUD model,” the next step is to *define* these components. Early in this essay, we ventured a tentative definition of data as “**the stuff out of which ‘better predictions’ are made.”** It is now time to revisit this definition.

If we take the hierarchies discussed seriously, then another candidate definition for “data” is **“the raw material from which information is derived.”** By the same logic, we can define “information” as **“the material from which knowledge is generated”**; “knowledge” as “**the material from which understanding is woven”**; and “understanding” as **“the emergent process through which a decision is reached.”** We can (temporarily) simplify the above definitions like so:

Table 1 – Definitions of DIKUD at a glance

Term	Definition
“Source” Data	**“Source” Data is serialized, but otherwise unprocessed, observation.** I.e. the output of a sensor; the “raw material” used by any data processing process.

Information	**Information is the result of processing “Source” Data.** Usually called “Data processing”, but will be called “Information processing for reasons that will become clear.
Knowledge	**Knowledge is the result of processing Information.**
Understanding	**Understanding arises from the processing of Knowledge.**
Decision	**Decision is the output of the process of Understanding.**
Intelligence	**Intelligence is a scale for the evaluation of processing *efficiency*.**
Wisdom	**Wisdom is a scale for the evaluation of processing *effectiveness*.**

The careful reader, however, may have noticed a bit of sleight-of-hand in the dynamical definitions of DIKUD offered in the table above: Each term is defined as the **product** of putting the previous term through some sort of **processing** – except for “Data” itself, which is defined as **“Unprocessed ‘Source Data’; i.e. the ‘raw material’ processed by any data processing process.”** In other words, unlike “Information,” “Knowledge,” “Understanding,” and “Decision,” Table 1 defines “Data” (i.e. “Source Data”) *only in terms of itself, and its **potential** to be processed into other forms.*

This semi-tautological “definition,” however, leaves open a crucial question: ***What is “data” in its own right?*** Earlier in this essay, we noted that Fricke’s critique of the DIKW model rests in part on a definition of data as “the output of a sensor”– a common intuition, and one related to the fact that some act of *sensing* (i.e. observation or measurement) is the *point of origin* for all “Data.” While this definition is not *incorrect*, it is meaningfully *incomplete*; it accurately reflects the role played by “Data” at the **start** of any *sense-think-act* paradigm, but misleadingly suggests that “Data” exits the stage once it has set the wheels of the plot in motion (so to speak).

We believe that “data” is better defined as “**the *serialized* output of a process of computation”** – in other words, **“the output of a process of computation, in storable or**

transmittable form." This definition of Data *includes* simple sensor measurements – a measurement being the serialized output of a sensor (e.g., a particular voltage across two points on the sensor, or a stream of measurements taken in succession), but it *also* includes a book, or a library of books. This definition of "Data" resolves the data-as-only-the-beginning criticisms, reflecting the fact that there is no *de novo* data – but also the fact that while "Source Data," "Information," "Knowledge," "Understanding," and "Decision" are all different *kinds* of epistemic objects, *"Data" is the material out of which **all five kinds** are made.*

"Source Data," "Information," "Knowledge," "Understanding," and "Decision" are best understood as ***modalities*** of data; **each names the *type* of "Data" *produced by* a specific *mode of (data) processing.*** (In the case of "Source Data," the "processing" in question is the process of "sourcing" that data, as we will explain in due course). Table 2 (below) integrates this revised understanding of data into our definitions of DIKUD.

Table 2 – Revised Definitions of DIKUD

Term	Definition
Data	**The serialized (storable or transmittable) output of *any* process of computation.**
Source Data[1]	**"Raw" data – *data in the modality in which it is input into a given data processing process, before it has been acted upon by that process.*** *In the context of a particular process, "Source Data" is "data in its initial/unrefined/unprocessed form."* Alternatively, the product of the process of *sourcing* data.
Information	***Processed* Source Data;** the product of *Information Processing*
Knowledge	***Processed* Information Data;** the product of *Knowledge Processing*

[1] (To avoid terminological confusion, we will hereafter use "Source Data" instead of "Data" to refer specifically to the first "D" in "DIKUD." We will retain DIKUD as an acronym, however, to avoid obscuring the fact that the initial input into a DIKUD system is "data," in the colloquial sense.)

Understanding	***The Processing of* Knowledge Data.** "An understanding" is never a static object, but always remains a process in its own right.
Decision	***Processed* Understanding;** the product of *the Process of Understanding.*
Intelligence	**A scale for the evaluation of processing *efficiency*.**
Wisdom	**A scale for the evaluation of processing *effectiveness*.**

While these definitions are obviously still incomplete, they make it apparent that **DIKUD is better understood as a *data processing system*, rather than as *a hierarchy or taxonomy of epistemic objects*.**

Figure 4 (below) offers a conceptual design for a DIKUD system in which "Intelligence" and "Wisdom" once again fulfill intrinsic and extrinsic KPI-like functions.

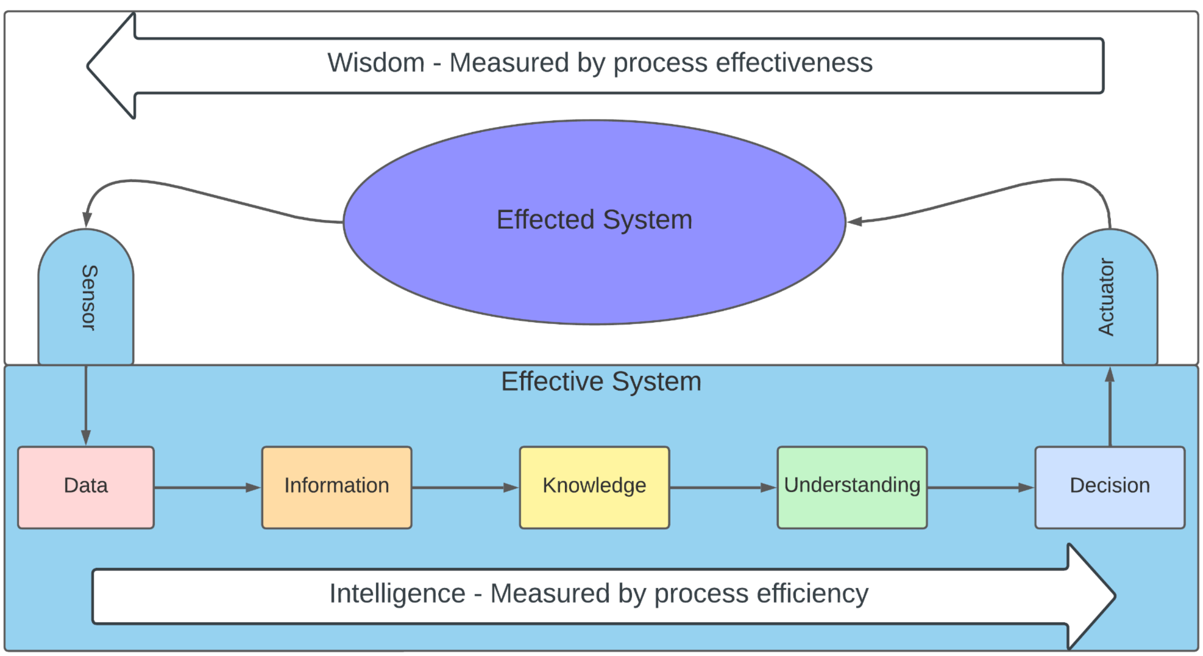


Figure 4: The DIKW hierarchy reconfigured as a DIKUD system, and applied as a dynamic function in an operational system with sensors and actuators.

The overall system consists of an **"Effective System,"** which represents the DIKUD Pyramid as an interconnected sequence of **processes** and the **products** of those processes, coupled with an **"Effected System"** from which the "Source Data" being processed by the "Effective System" is initially **sourced**, and back into which data fully processed by the "Effective System" is ultimately **sunk** (in the form of a "Decision" that has an *effect on* the "Effected System"). In this model, **"DIKUD" clearly represents the series** (or **pipeline**) **of *forms* or *modalities of data* through which "Source Data" progresses, as the "Effective System" processes data into decisions.**

In this view, it is apparent that "Wisdom" and "Intelligence" **do not belong** to the sequence of modalities that the "Effective System" processes data into. Rather, they are high-level **concepts** in relation to which that process's end-to-end performance can be assessed.

The nature of this relation, however, is less straightforward than it may initially appear. Previously, we suggested that "Wisdom" and "Intelligence" can be thought of as **KPIs** – metrics against which one can assess critical dimensions of a system's performance. By definition, however, KPIs must be computable, and neither "Wisdom" nor "Intelligence" are properties of a system that are susceptible to direct measurement, for the simple reason that neither term has a univocal and universally-accepted definition. Instead, both are what W.B. Gallie has called "essentially contested concepts" – concepts that do not "carry with them an assumption of agreement" as to their meaning (e.g., 'justice,' 'freedom,' or 'fairness'). Absent universal agreement about what "Intelligence" or "Wisdom" is, neither can be used as a KPI – but this does not mean that we cannot attempt to orient a system along these dimensions.

The fact that the concepts of "Wisdom" and "Intelligence" are "essentially contested" means that optimizing for either requires a multi-step process. Designers of DIKUD systems must first assert a specific **conception** of these (and any other) essentially contested concepts within the context of the particular system that they are designing, and then identify measurable and computable **proxies** for the conceptions that they have asserted. These proxies can then serve as KPIs, enabling the evaluation of the system's performance relative to the particular conceptions of "Wisdom" and "Intelligence" that have been asserted. At the same time, evaluators can refer back to the overarching **concepts** of "Wisdom" and "Intelligence" in order to assess how well the **conceptions** that have been asserted (and KPIs that have been identified) are, in fact, fit to the context at hand; that is, whether "improved performance" in terms of the chosen KPIs corresponds to **outcomes** that are subjectively better-aligned with evaluators' preferences.

Recalling Ackoff's framing, for example, one might identify the **concept** of "Intelligence" with the specific **conception** "Efficiency," which concerns **how closely a system's actual performance conforms to its design**, and the concept of "Wisdom" with the specific conception "Efficacy," which concerns **how well a system realizes the purpose for which it was designed**. Efficiency is a function of the process alone (the blue "Effective System" in the diagram above), while Efficacy is a function of both the "Effective System" and the "Effected System," because Efficiency is related to a process's inner workings, while Efficacy characterizes the relationship between a process and its intended outcomes.

While there are many potential "Efficiency" metrics, such metrics most often take the form of **rates of value production per unit cost**, e.g. transactions per second (for a database), or cost per 1000 tokens (for a large language model). "Efficacy" can also be measured in various ways; quality of service (QoS) metrics might include customer satisfaction ratings (or other stakeholder based assessments), and/or performance metrics (such as the out-of-sample r-squared value for a regression model).
Crucially, neither "Efficiency" nor any of its proxies is the same thing as "Intelligence" – and neither "Efficacy" nor any of its proxies is identical with "Wisdom." Rather, "Efficiency," "Efficacy," and all of their potential proxies are particular conceptions of these essentially contested concepts – specifications of what designers take them to

mean, in the context of the system in question. As conceptions of essentially contested concepts, however, neither “Efficiency” nor “Efficacy” can exhaust the possible meanings of “Intelligence” or “Wisdom”; a system that is not efficacious is probably also not one that we would generally consider “wise,” but there are certainly situations in which “Efficacy” and “Wisdom” diverge. In such situations, designers would be well-served by asserting an alternative conception of “Wisdom,” and identifying alternative KPIs against which the system’s performance relative to that conception can be evaluated.

Ultimately, it takes wisdom to assess how wisely one has conceptualized “Wisdom” and “Intelligence,” and intelligence to evaluate whether or not once done so intelligently. To use anything as a measure of itself, however, is to defeat the purpose of measurement; thus, the overarching concepts of “Wisdom” and “Intelligence” must be understood as uncomputable, undefinable values that exist outside of the context of any specific system – they are, instead, both the values that a particular DIKUD system’s KPIs seek to approximate, and the intrinsically-amorphous concepts against which the aptness of those approximations are evaluated.

Recalling Rowley's conclusions, Figure 4 uses arrows to indicate *where* “transforms” are needed, but it does nothing to define *what* those transforms are. Additionally, where Rowley mapped “Decision Support System” to the last stage of this system, the whole flow shown in Figure 4 can appropriately be called a Decision Support System – a *data-driven* one.

Having established a vocabulary of terms with specified interrelations, the next steps are to identify and define the “transforms” themselves – that is, the component sub-processes that must be wired together to form an end-to-end Source-Data-to-Decision processing pipeline. Before we can do so, however, it is first necessary to shift the focus of our model from **the modalities of data** (the various **products** of data processing) to **the variety of *modes of processing* that produce them.**

Third Revision: The Dynamics of Data

The traditional DIKW model presents the modalities of data as a hierarchy of epistemic forms – “Data” is processed into “Information,” which is processed into “Knowledge,” which is processed into “Wisdom.” We began our revision of this model by reworking the DIKW Pyramid around a DIKUD stack, then modeled that stack as a dynamical system – but both of these steps still leave us with representations of data processing that foreground the *outputs* of the component subsystems that make up the higher-order system that transforms input data into actuated decisions.

These outputs, however, ***are all still data*** *– they are simply data that has been put through different modes of processing***.** Each **mode of processing** takes a specific **modality of data** as its input, then acts upon or computes over that data in particular ways in order to output data of a different modality. In other words, “information processing” takes in “raw” Source Data, processes it, and outputs data in the modality of “Information.” By the same token, “knowledge processing” takes in data in the modality of Information, processes it, and outputs data in the modality of “Knowledge.” (As we will soon see, other basic structural features are also consistent across all modes of processing, even as each mode differs from the others in significant ways.)

To understand the “how” of data, it is therefore necessary to refocus our model of DIKUD-as-dynamical-system on its component **data processing sub-processes,** rather than on the kinds of epistemic objects (one might also say **“data products”**) produced *by* those processes. We can represent this shift of focus by “zooming in” on the “Effective System” depicted in Figure 4, but modifying it so that the “boxes” represent the **modes of processing**, and the arrows represent the **modalities of data** in which each process receives data from its predecessor and passes data to its successor, like so:

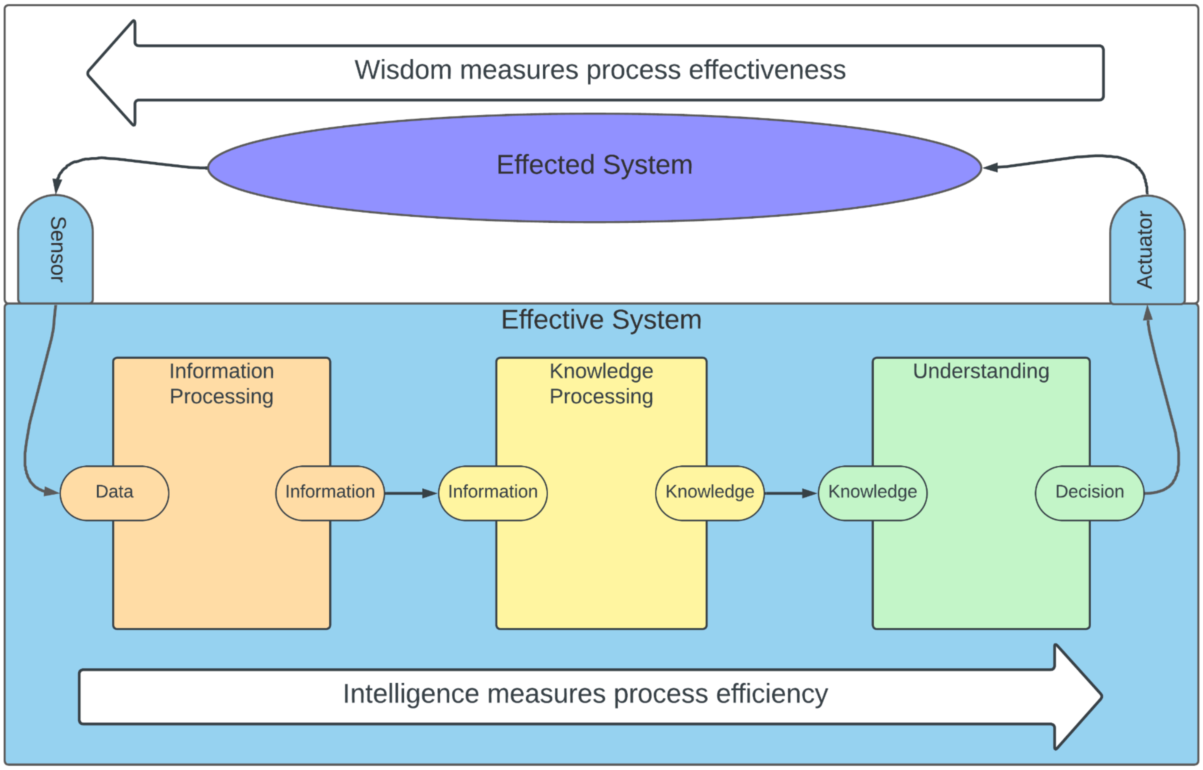


Figure 5: The DIKUD system expressed a series of transformations

Figure 5 represents DIKUD as a series of **processes**; the process of **Sourcing Data** takes some **signal** from the **Effected System**, **serializes** it, and outputs it into the **Effective System** in the form of **Source Data. "Information Processing"** takes this **Source Data** as its input, processes it, and outputs **Information**; "**Knowledge Processing"** takes in **Information**, processes it, and outputs **Knowledge**; finally, the **Process of Understanding** takes in knowledge, processes it, and ultimately produces an **Understanding**, which is translated into a **Decision** that has an actual effect on the **Effected System**.

At this point, all we have done is re-arrange the components of our model of a DIKUD system, but in doing so, we have brought our central question into its sharpest focus yet: **What happens *inside* the "black boxes" that process data from one modality into another? *What are the component sub-processes of each mode of processing? What characteristics differentiate the modes of processing from one another, and what do they all have in common?***

Fourth Revision: Defining the Transforms

“Sourcing” Source Data

We will begin our unpacking of the **modes of processing** by considering the tacit knowledge embedded in Natural Language. Source Data, as we have said above, is the “raw” data fed into the Effective System in our DIKUD model. This data is taken from the Effected System and input into the Effective System through the process of **“sourcing.”** Some act of **sensing** initiates the process of sourcing data – something about the state of the Effected System is measured or observed, then the result of this act of sensing is **serialized** (rendered storable or transmittable); it is this serialized output that feeds into the Effective System as Source Data.

Consistency of naming conventions across this section would suggest that we use the term “data processing” to refer to the process of sourcing source data from the Effected System and serializing it into a form that can be input into the Effective System just described. We have decided, however, to retain the term “data processing” to refer to the end-to-end processing of source data into decisions, or the general act of processing data from one modality into another. Thus, we will refer to the process of acquiring some signal from the Effected System, conforming that signal into the modality of Source Data for the Effective System (*i.e.* serializing it), and “publishing” it into the Effective System as **sourcing source data.** In reality, the boundaries between **sourcing source data** and the first steps of **information processing** can rarely be clearly drawn.

Information Processing

We will begin with the mode that is most familiar – **information processing**. Information processing is the process by which Source Data transforms into Information. The difference between (raw) “Source Data” and “Information,” as commonly understood, is that information is **meaningful**, while source data, by itself, is not.

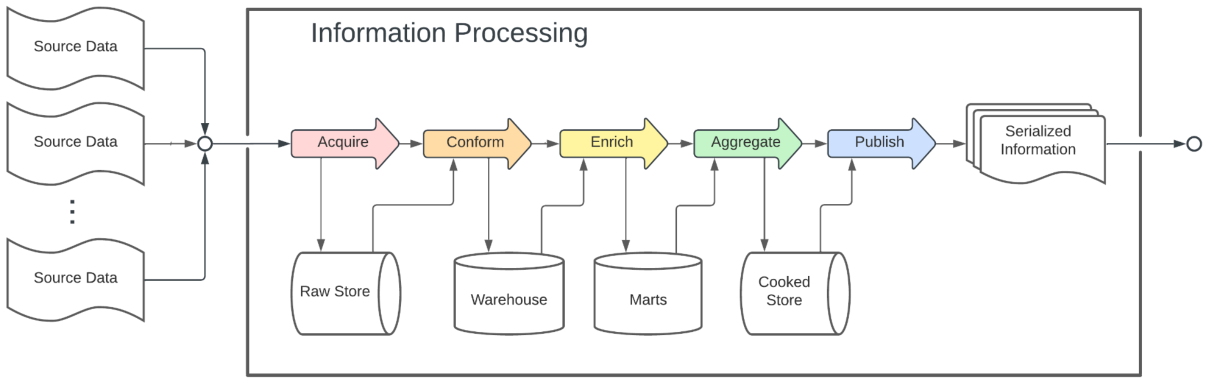


Figure 6 – Internal workings of an information processing system

One can therefore think of information processing as primarily a **sorting** or **contextualizing** operation – a fairly linear process of separating wheat from chaff, signal from noise, albeit one involving several sub-processes:

1. **Acquire** data
2. **Conform** it into measurements (values + dimensions) in alignment with local standards
3. **Enrich** it through JOIN operations
4. **Aggregate** it through GROUP BY operations
5. **Publish** the resulting information

Source data transitions into information fairly early in this process. The result of **conformation** is a data warehouse in which facts are stored, paired with dimensions to provide context; "facts in context" are already "meaningful data," such that the contents of a data warehouse of this sort can already be considered to be "information."

As epistemic objects (or "data products"; i.e. the outputs of data processing processes), source data and information can both be thought of as ***immutable*** – once created, their states cannot be changed or updated. The operations involved in processing source data into information are also ***idempotent*** – they will return the same result, no matter how many times they are run. Information can be either **true** or **false**, but the truth or falseness of a bit of information **does not** depend on the context in which one encounters it.

Source-Data-to-Information transformations are a mature part of the DIKUD process; in Rowley's nomenclature, “Information Processing” is the purview of Transactional Processing Systems and Management Information Systems. This mode of processing is heavily automated, and includes abstraction layers that hide the low-level details and improve the experience of both transform users and transform developers. Databases, data warehouses, data marts, data lakes, and data streams are all well understood infrastructural elements in transforming data into information.

Transforming source data into information is a past-oriented process. Source data, by definition, tells us about *things that have already happened* (as past-ness is a condition of something *having being observed/sensed*), and thus can only be processed into **information *about the past*** (including the very recent past, *i.e.* the “present” that has passed into the past by the time that one has finished observing it). It may be self-evident that **one cannot derive *information about the future* by processing source data**, insofar as information, as a modality of data, cannot *be* about the future. It is perhaps more surprising that, strictly speaking, one cannot derive ***information about the present*** from source data, either – for as we are about to see, the present is the domain of knowledge.

Knowledge Processing

“Knowledge processing” picks up where “information processing” left off. Just as source data is the raw material of Information, information is the raw material of knowledge. The *goal* of knowledge processing is to discover coherent associations among bits of information – patterns of association that are common across different views of the information in question, thus suggesting that the associations themselves have a kind of **objecthood**. The *result* of knowledge processing is the **formation** of “knowledge objects,” commonly referred to as “**concepts**.”

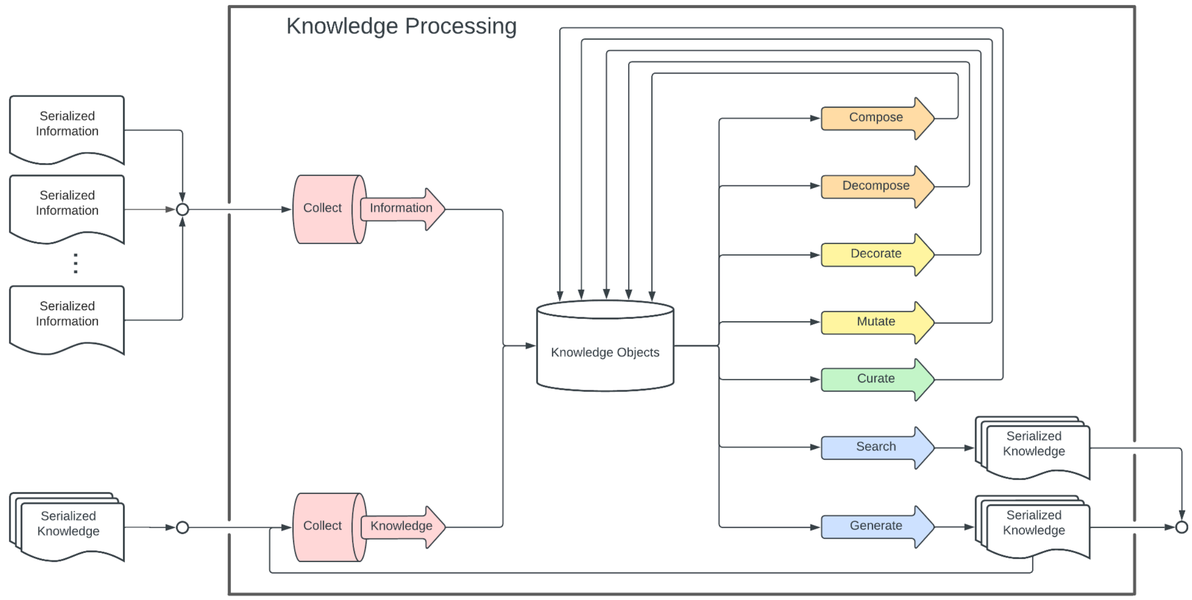


Figure 7 – The Conceptual Design for Knowledge Organization Infrastructure developed by BlockScience as an example of the Internal workings of a knowledge processing system

Knowledge processing should therefore be understood as primarily a process of **conceptualization** – **abstracting** from information into **internal representations**, and then **organizing** these abstractions into **concepts** or **a point of view.**

The automation of **Information-to-Knowledge transformations** is a work in progress. Knowledge graphs exist, but they are bespoke systems that have only low-level components in common. The configuration of these components for individual instances is largely performed by humans, with little help from automation tools.

In Rowley's nomenclature, automating the processing of information into knowledge is the purview of Decision Support Systems (DSS). We believe, however, that "decision support" is the purpose of the DIKUD system taken as a whole, and so prefer to avoid identifying any one of its individual subsystems with DSS at the expense of the others.

The essential operations of knowledge processing include, but are not limited to:

- Collection
- Composition
- Decomposition
- Mutation
- Decoration
- Search
- Generation

These operations exhibit a close correspondence with those at the heart of information processing:

1. Acquire → Collect
2. Conform → Decompose
3. Enrich → Mutate & Decorate
4. Aggregate → Compose
5. Publish → Search and Generate

Despite these structural similarities, knowledge processing differs from information processing in significant ways. In contrast to the linearity of information processing, the operations of knowledge processing are iterative, recursive, and self-referential; once information has been collected, the component processes of knowledge processing can occur in any order, any number of times. Knowledge objects (the products of knowledge processing) are also *highly* mutable – they can, and frequently do, change over time, and knowledge processing systems must account for this mutability. Furthermore, associating the same knowledge objects twice will not necessarily produce the same results, because attention matters; in other words, unlike information processing, knowledge processing is **not** idempotent, and it is possible to generate new knowledge objects from existing knowledge objects by way of continued processing.

BlockScience has been researching Knowledge Organization Infrastructure (KOI) through a Technology Research Pod initiative led by Orion Reed, with assistance from Luke Miller. Based on this research, Figure 7 (above) appears to be a flexible architecture that will get us from life, the universe and everything to a collection of concept-like objects that represent the universe and everything, and that will enable us to do something like *thinking about* life, the universe and everything.

If information processing is a finite process of sorting "meaningful facts" out of "raw" source data, knowledge processing is the potentially-infinite work of examining and cogitating over those facts, in order to tease out stable constructs from the relationships that emerge between them. One can think of this process as one of generalizing from specific information about "how x has worked in the past" to a perpetually present-tense sense of "how x works." Once again, one cannot, strictly speaking, have knowledge about the future**,** for the future, strictly speaking, *cannot be known*. **Knowledge processing, however, organizes and reorganizes information about the past into the *kind of epistemic object* (or *modality of data*) that can *subsequently* be processed into the only sort of insight into the future that is actually available to beings bound to linear time: *an understanding of the relationship between possible actions in the present and the likelihood of potential outcomes in the future.***

The Process of Understanding

While knowledge projects a view of the present from information about the past, the fact that this view has been abstracted exclusively from information about the past means that the present knowledge projects is never perfectly identical with the **specific present**, the moment that one actually occupies. Instead, "the present" produced by knowledge is a more generalized "perpetual present" – knowledge processing is more about extending the line of the past until it reaches the region of time that we think of as "now" than it is about mapping that region's terrain.

That process of mapping is the remit of the **process of understanding**. Our language winces at the phrase "understanding processing" because the temporality of the process of understanding is **circular** – and circles around the set of acts that make up the process of deciding upon a course of action.

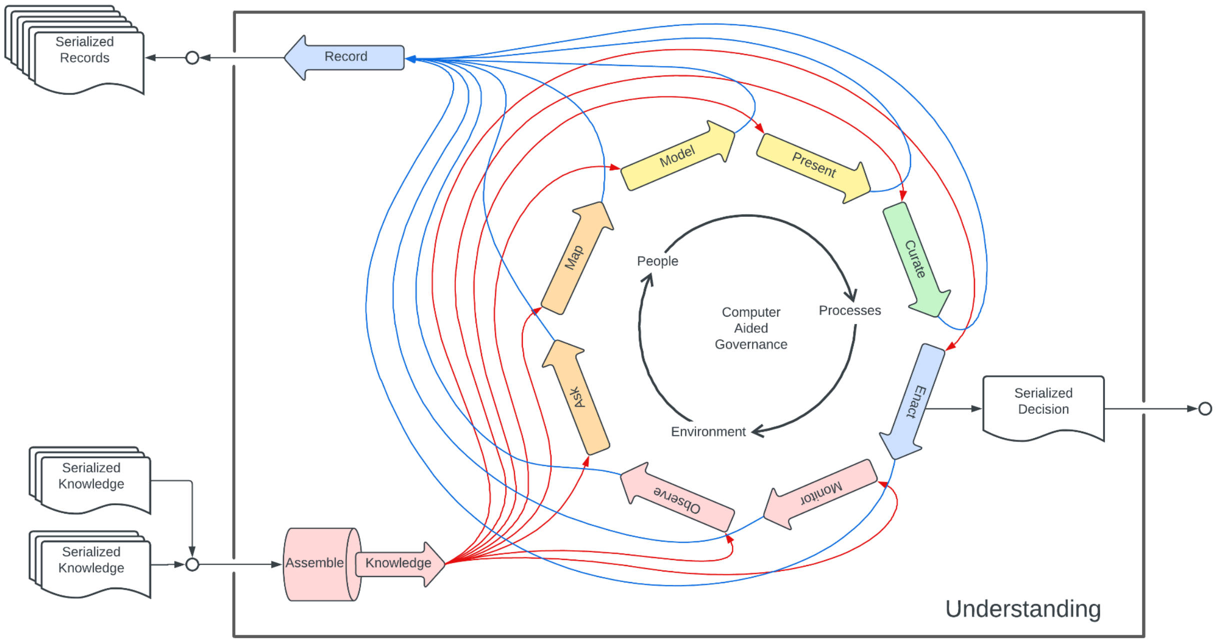


Figure 8: A schematic representation of The Computer Aided Governance Map developed by BlockScience as an example of the Internal workings of a system for developing understanding

The process of understanding **fuses** or **synthesizes multiple views of the perpetual present projected by knowledge into the best approximation available of "the situation at hand" – the specific present that we actually inhabit – and of the relationship between possible actions taken in that present and the likelihood of potential outcomes in the future.** The process of understanding thus operates in the spaces between the perpetual present of knowledge, the specific present of the now that we can know only in its passing, and futures in which that **specific present, rather than the perpetual present, will be the actual past.**

In other words, the process of understanding involves fusing multiple views on the perpetual present of knowledge into a shared view of the specific present – specifically, **the shared view of the specific present through which it is easiest to "connect the dots" into a line linking the observed past to the desired future.** Once this view has been constructed, it is inscribed back into the Effected System in the form of a **decision** premised on the best available understanding of the relationship between possible actions in the present and the likelihood of potential outcomes in the future.

How is this mapping of the present accomplished? There is, as yet, no widely-accepted answer, but BlockScience explores the process of understanding using the Computer-Aided Governance (CAG) Map represented schematically in Figure 8. The CAG Map was developed by our Governance Research Pod (whose members include Kelsie Nabben, Jeff Emmett, pseudonymous researcher Burrrata, Michael Zargham and others), and is evolving iteratively, in concert with our understanding *of* understanding. The operations that the CAG Map currently uses to model the process of understanding are:

- Monitor
- Observe
- Ask
- Map
- Model
- Present
- Debate
- Enact

As a model of the process of understanding – which is a process of fusing multiple viewpoints into a higher-dimensional model of the present and its probabilistic relation to the future than would be available from a single perspective – CAG is all about **communication.** Thus, each tool on the CAG Map aims to make effective communication more efficient, in the context of a decision.

CAG is currently implemented using a mix of tools and practices, which one can think of as "social tools." This broad selection of tools is necessary because CAG – and the process of understanding, more broadly – are all about the **specific context within which a decision is being made.** Not every tool is appropriate for every context, so CAG offers a flexible array of options. Integration across these disparate tools and practices is an area that is ripe for further exploration.

The CAG Map is built around an understanding that facilitating broader access to communicable knowledge is a powerful way to increase the efficiency of the process of understanding that transforms contending knowledges into actuated decisions (cf. Eric Alston's research on "Governance as Conflict"). A fancy data lake or knowledge graph would be out of place here; a calendar, docket, agenda, or even a to-do list is more than sufficient infrastructure to queue up "things that must be decided," and thus begin cycling through the process of understanding.

From this perspective, the process of understanding appears to be quite different from both information processing and knowledge processing – yet it still follows the same operational pattern of all data processing: acquisition, conformation, enrichment, aggregation, and publication.

1. Acquire ⟶ Monitor & Observe
2. Conform ⟶ Ask
3. Enrich ⟶ Map & Model
4. Aggregate ⟶Present & Debate
5. Publish ⟶ Enact

Decision

Like the process of sourcing data, the process of implementing a **decision** spans the border between the Effective System and the Effected System in a DIKUD architecture – and like the process of sourcing data, the process of implementing a decision is arguably better understood as a single action than as a process with meaningful inner workings. **Arriving at** a decision is the work of the process of understanding; once a decision has been reached, however, the process of implementing that decision *consists of* **actuating** the product of the process of understanding in question – which is to say, of taking some action within the Effected System that reflects the outcome of the data processing that the Effective System has performed. Inscribing the outcome of the Effective System's processing back onto the Effected System has an *effect on that system*, changing its state both directly and indirectly. This effect can (and presumably will) be measured as new source data is sourced from the Effected System into the Effective System and fed through the interrelated processes that make up "data processing" as an end-to-end endeavor.

Once a decision has been implemented, the end-to-end performance of the DIKUD system that produced that decision can be evaluated, as detailed earlier in this essay. The conception of "Intelligence" that we have previously asserted – "Efficiency" – posits an implemented decision as the terminal point of an open loop that begins with Source Data and runs through the remainder of the "Effective System." Therefore, KPIs meant to function as proxies for a DIKUD system's "Efficiency" (and, by extension, approximations of its "Intelligence") will measure the trade-offs that "Effective System" makes as it processes data into a decision (i.e. speed vs. cost, precision vs. speed, etc.).

By contrast, the conception of "Wisdom" that we have previously asserted – "Efficicacy" – understands the decisions implemented by the Effective System as the points at which the larger loop connecting "Effective System" to "Effected System" *closes*. Therefore, KPIs meant to function as proxies for a DIKUD system's "Efficacy" (and, by extension, its "Wisdom") will measure the impacts on the Effected System of the decisions implemented by the Effective System. Insofar as "Efficacy" conceptualizes the overall DIKUD system as a closed loop consisting of both Effective and Effected Systems, evaluating the "Efficacy" of a DIKUD system is meaningfully different from evaluating the "Efficacy" of any particular decision implemented *by* that system, because any implemented decision functions as both output of and input into the overall system.

As a particular DIKUD System implements more decisions, and is thus able to process more data concerning the outcomes of its previous decisions, the "Efficacy" of that

DIKUD System should progressively improve over time – but whether this *more efficacious* system is also *wiser* is something that can ultimately only be determined subjectively. Conversely, because “Efficiency” metrics are unaffected by the state or activity of the Effected System, there is no reason to expect either the “Efficiency” or the “Intelligence” of a DIKUD system to increase over time (absent further engineering).

From the perspective of those served by a DIKUD system, attempting to optimize for “Wisdom” is likely to produce “better outcomes” than attempting to optimize for “Intelligence” – but only if the conceptions that one asserts (and proxy KPIs that one identifies) for these “essentially contested concepts” are well-suited to the context at hand.

Fifth Revision: Bringing It All Together

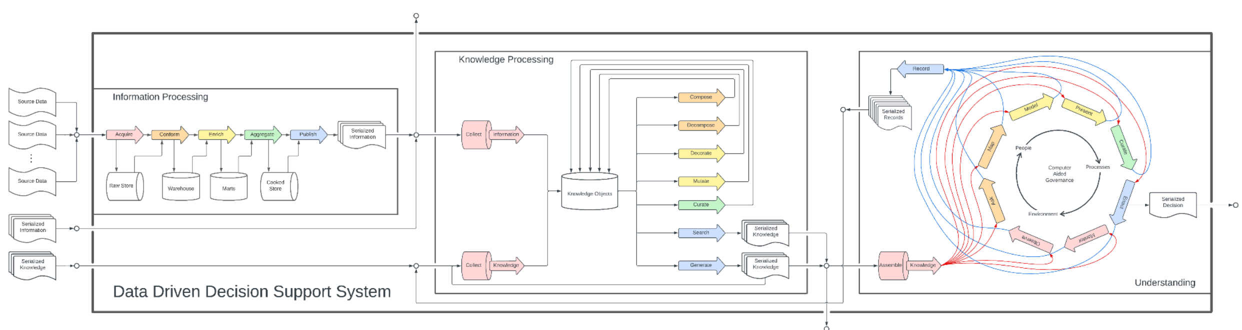


Figure 8: Combines Figures 6-8 into a data driven decision support system.

In the discussion of knowledge processing found in the previous section, we suggested that “Decision Support” is the purpose of the full data processing stack, and not just what are today referred to as “Decision Support Systems.” It is exceedingly difficult to change the commonly-accepted meaning of a well-defined term in common usage, however, so I propose instead the term **“Sensemaking Systems”** to describe systems that process data pursuant to the DIKUD architecture that we have described and analyzed in this essay.

R.J. Cordes (following the lead of Karl Weick, who wrote a book on the topic), defines "sensemaking" in a paper published in 2020: "Sensemaking is literally the act of making sense of an environment, achieved by organizing sense data until the environment "becomes sensible" or is understood well enough to enable reasonable decisions."

This essay's iterative revisions of the DIKW hierarchy of epistemic objects are meant to help identify and separate the concerns that those seeking to design a Sensemaking System will need to address. We have done so by suggesting conceptual subflows for such a system, and attempting to illustrate the operations and workflows that each of those subsystems should support, in order to function as the architecture requires.

Table 3 – Final Definitions of DIKUD (at a glance)

Term	Definition
Data	**The serialized (i.e. storable or transmittable) output of a process of computation.**
Source Data	**"Raw" data – *data in the modality in which it is input into a given data processing process, before it has been acted upon by that process. In the context of a particular process, "Source Data" is "data in its initial/unrefined/unprocessed form."***
Information	***Meaningful* Source Data;** Source Data that has been *sorted.*
Knowledge	***Organized* Information;** Information that has been *conceptualized*.
Understanding	***A Fusion of* Knowledges;** synthetizing Knowledges for application to a particular context.
Decision	***Actuated* Understanding;** Understanding that has been *implemented*.

The differences between these operations and workflows suggest different processing models for each subsystem – but although the concerns of each subsystem are **separable**, they are not **isolated**. Ultimately, the utility of a Sensemaking System is a function of the **harmonization** of its subsystems, not in their individual **optimization**. At present, **source data** ⟶ **Information transformations** are *heavily* automated through digitization and digital computation; the throughputs of these systems are such that working with their output streams is commonly referred to as "drinking from the firehose." The work of automating **information** → **knowledge transformations** is ongoing, and is one of the areas in which Large Language Models have the potential to power significant advances in the state of the art. **Knowledge** → **Decision** transformations (as we have defined **Understanding)**, however, remain overwhelmingly manual at the time of this writing; what automation tooling exists is limited to the standard office automation tools that automate some aspects of individual workloads (*e.g.*, word processors, spreadsheets, and emails), and scaling is achieved through bureaucracy.

Using Data for Collective Decision-Making Processes

The Era of Big Data saw the full automation of sourcing data and information processing, but has left us with a bottleneck at knowledge processing; the process of understanding, meanwhile, is still primarily reserved to centuries-old institutional hierarchies. As we enter the Age of Artificial Intelligence, and it becomes possible to automate an increasing proportion of the work done by sensemaking systems, it is critical that we focus less on aligning "intelligent" systems that focus on efficiency, and more on building *wise* systems – systems that more effectively help human beings meet their actual needs.

Data is very heterogeneous stuff, and therefore difficult to generalize about – but at the end of the day, it is the **only** medium through which observations about the past can be refined into present-tense knowledge and future-oriented insight. Ultimately, data exists so that there is *something* we can learn from, *something* we can think with, and *something* that we can use to anchor ourselves in a shared, specific present; “so that there is something to compute over,” in the parlance of our eternally-passing time.

Acknowledgments

This paper was authored by David Sisson and Ilan Ben-Meir, with content and edits from Michael Zargham and Jeff Emmett. It references ongoing research in the BlockScience Technology POD, particularly in reference to Knowledge Organization Infrastructure and its application to organizational decision-making.

About BlockScience

BlockScience® is a complex systems engineering, R&D, and analytics firm. By integrating ethnography, applied mathematics, and computational science, we analyze and design safe and resilient socio-technical systems. With deep expertise in Market Design, Distributed Systems, and AI, we provide engineering, design, and analytics services to a wide range of clients including for-profit, non-profit, academic, and government organizations.