

## The Kingdom of Education Data

Glynn D. Ligon, Ph.D.

### Intro

Real-time data work for making timely decisions by teachers and other educators. Validated, official statistics work for state reporting (and *EDFacts*), longitudinal analytics, growth models, research, etc. Does your state's SLDS<sup>1</sup> do both well? Is your data model founded upon providing data for decision making? Maybe your data model has been created to describe whatever data are being collected now—no matter how those data were defined, compiled, and offered up for use.

ESP has documented over 50 adjectives describing data (formative, summative, imputed, personally identifiable, discrete, categorical, etc.). Bottom line, we have to match the nature of our data with the intended use of them to ensure valid decisions.

What better way to examine our education data than to compare ourselves to the classification of the animal kingdom.

## The Kingdom of Education Data

Education IT pros are organizing our data using data models. Let's step back from the details of those data models for a moment. For example, the Common Education Data Standard (CEDS) has worked diligently to create nine domains with corresponding entities, categories, and elements for education agencies to model their systems around. However, I just reviewed their documentation, again. <https://ceds.ed.gov/publications.aspx> Something's missing.

The workmanship is fine. The impression given off is that a committee began right in the middle of a problem and is methodically working its way outward.

That's probably how most, if not all, large data models develop.

Let's go even broader.

For centuries, scientists have been arguing about classification of animals and plants--taxonomic ranking. Are there eight or nine classifications, multiple subclasses, and ancestral or hereditary hierarchies? The competitions for classification dominance predate Aristotle. This is the style of intellectual debate we should be having over education data. Not just whether there should be a category for PK12W rather than just K12.

So, here we go. I propose a rethinking of education data into a classification system that parallels that of the animal world. To begin this process, we must have some fundamental understandings.

1. We collect data about learning within the education universe to inform decisions.
2. The decisions to be made should determine the data we collect, manage, and provide to the decision makers.
3. Everything else is process<sup>2</sup>.

Our Human Classification

The most agreed-upon modern human classification has eight levels. These are typically shown as an inverted pyramid.

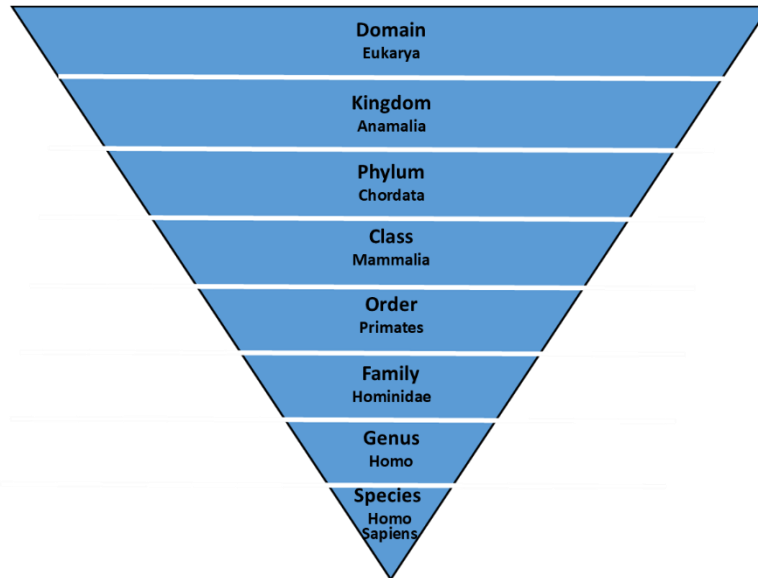


Figure 1: Classification System for the Animal Kingdom Shown for Homo Sapiens (Us)

Our Education Data Classification

Our vision for education data is quite different. There is a smaller domain and many more data elements. Our pyramid is not so tipsy.

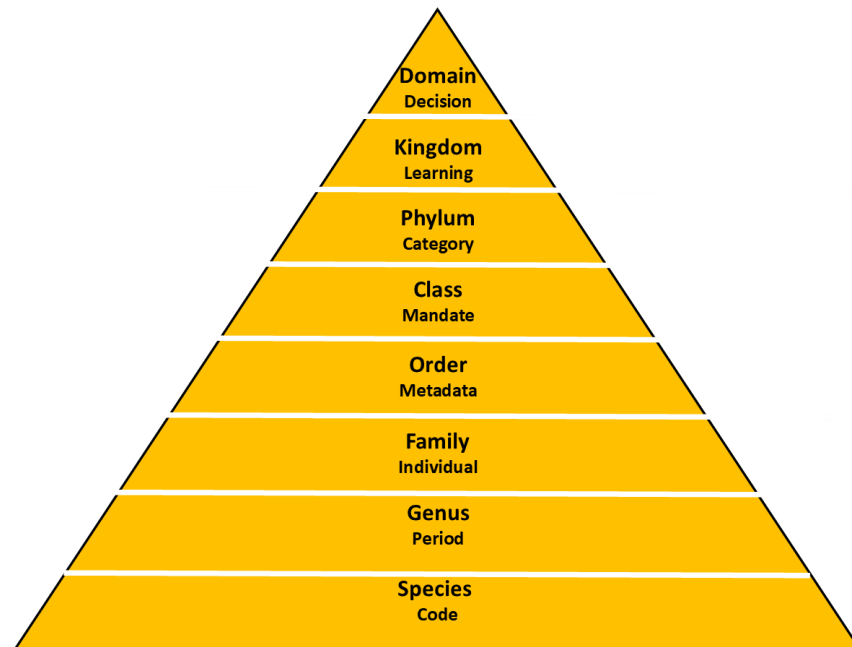


Figure 2: Classification System for Education Data

ESP Solutions Group has worked with most state education agencies and the U.S. Department of Education to create data dictionaries. We have compiled quite a collection of adjectives describing data. These adjectives are significant because each denotes a use, condition, purpose, readiness, or other state of data that is crucial for a data manager or decision maker to understand.

Classification (Taxonomy)	Humans	Learning/ Education Data	Definition of Classification in Learning/Education: Data Adjectives	CEDS Classification
Domain	Eukarya	Decision	<b>Decision Making:</b> Actionable	
Kingdom	Anamalia	Learning	<b>Learning/Teaching:</b> Education, Academic, Research	
Phylum	Chordata	Category	<b>Category:</b> Context, Subject	
Class	Mammalia	Mandate	<b>Mandate/Purpose:</b> Funding, Input, Outcome, Intervention	
Order	Primates	Metadata	<b>Content/Operational Definitions:</b> Deidentified, Transactional, Performance, Snapshot, Formative, Summative, Outcome, Process, Diagnostic, Historical, Longitudinal, Predictive, Lagging, Leading, Max-Yield, Meta, Conflicting, Extract	Domains
Family	Hominidae	Individual	<b>Entity (Individual, Person, Object, Owner):</b> Personally Identifiable, Confidential, Deidentified, Role, Demographic	Entities, Categories
Genus	Homo	Period	<b>Periodicity/Quality:</b> Realtime, Point-in-Time, Clean, Quality, Raw, Derived, Extract, Backup, Missing, Imputed, Obfuscated	
Species	Homo Sapiens	Code	<b>Type (Text, Code):</b> Backup, Continuous, Discrete, Categorical, Interval, Electronic, Deidentified	Elements

Figure 3: Education Data Classification Topology

In Figure 3, data are classified into our newly created taxonomy for education data.

Let’s look at student attendance as an example. The presence or absence by a student on a given date is a common data element in a data model and a database. Figure 4 aligns the data element “In Attendance” to each level of the Learning/Education Classification.

Generic Classification (Taxonomy)	Classification for Learning/Education Data	Definition of Classification in Learning/Education	Sample Data Element: “In Attendance” (Student Present or Absent on January 31, 2024)
Domain	Decision	Decision Making	Decision Making
Kingdom	Learning	Learning/Teaching	Learning/Teaching
Phylum	Category	Category	Management
Class	Mandate	Mandate/Purpose	Funding, Intervention
Order	Metadata	Content/Operational Definitions	Operational Definition
Family	Individual	Entity (Individual, Person, Object, Owner)	Student
Genus	Period	Periodicity	Single Date
Species	Code	Type (Text, Code)	Code

Figure 4: “In Attendance” Classified as a Data Element in the Learning/Education Taxonomy

### Comparison to CEDS

I mentioned earlier that CEDS’s data model used domains, entities, categories, and elements. Despite some overlap in terminology, these groupings do not span the range of our new Education Data Classification Topology. This perspective is consistent with the opinion that CEDS, as a data model, describes data as objects. To us, data are far more. Data are essential components of decisions. Data are building blocks for decision makers to follow their plan for action.

Yes, we need a data model. However, that data model must be within the much broader and more significant context of a comprehensive data classification system. That is how we maintain our focus on the true purpose for data.

### More Perspectives on Education Data

In the Optimal Reference Guide, Defining Data<sup>3</sup>, 2006, I discussed the then-current state of data and the future of data. That future is now. You may judge for yourself about my predictions. I’ll just say that I was going to revise the paper. Then I decided to let it continue to stand. We seem to be progressing at a deliberate pace.

One of my favorite sections is the Salsa Scale. As the Learning/Education Classification Taxonomy provides the high-level view of data, the Salsa Scale dips right into the granularity of data—the Species of data.

Here is the description of the Salsa Scale as described in the original paper.

After Vince Paredes (Vice President of Research and Development for ESP Solutions Group) championed the notion of enhancing the granularity of data within information systems, Barbara Clements (Chief Standards Officer, National Transcript Center, ESP Solutions Group) and I were having lunch debriefing from an NCE conference. Her lunch was to have included pico de gallo, a chunky mixture of vegetables and peppers. What she got was picante sauce, smaller bits in an almost liquid state. The ESP Salsa Scale was born. ...the granularity of pico de gallo allows analysis of the contents, whereas, the blending of ingredients in picante sauce hides the detail. The point of the salsa scale is that the more we blend our data and lose granularity, the fewer options we have to disaggregate the parts and understand what our students are really like. Barbara and I still debate whether ketchup or V-8 juice is the lowest end of the Salsa Scale, but we both agree we would not bother dipping a chip into either one to examine the contents.

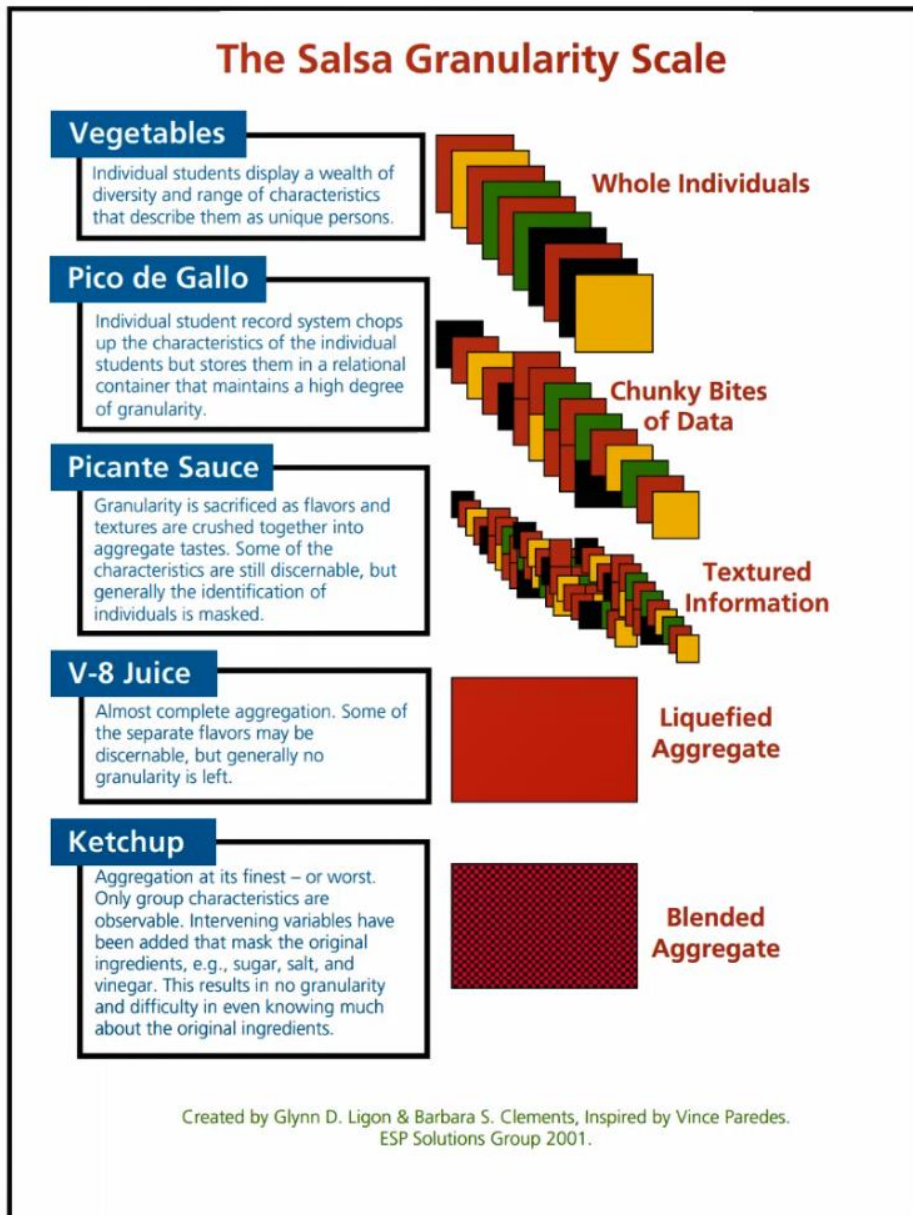


Figure 5: The Salsa Scale

### Final List of Essential Reminders about Data

Please keep in mind that this Education Data Classification Topology has not been vetted by our professional peers. This is not presented as anything more than a discussion starter. No wait, this is really more than that. This is a wake-up call to those working with data models. Please...

1. Remember, your data have to be organized around the decisions they are funded to inform.

Oh. There really aren't any other points to be made.

### References

1. SLDS State Longitudinal Data System

Systems developed by state education agencies with grants from the National Center for Education Statistics, Institute of Education Sciences. <https://nces.ed.gov/programs/slds>

2. American Productivity and Quality Center <https://www.apqc.org/>

Jack Grayson, Ph.D. founded APQC based upon the premise that successful companies achieve productivity through effective processes. The Malcom Baldrige National Quality Awards given by the National Institute of Standards and Technology (NIST), an agency of the United States Department of Commerce, are his legacy.

3. Defining Data, ESP Optimal Reference Guide, 2006, Glynn D. Ligon  
[http://www.espsolutionsgroup.com/espweb/assets/files/ESP\\_Defining\\_Data\\_ORG-1.pdf](http://www.espsolutionsgroup.com/espweb/assets/files/ESP_Defining_Data_ORG-1.pdf)