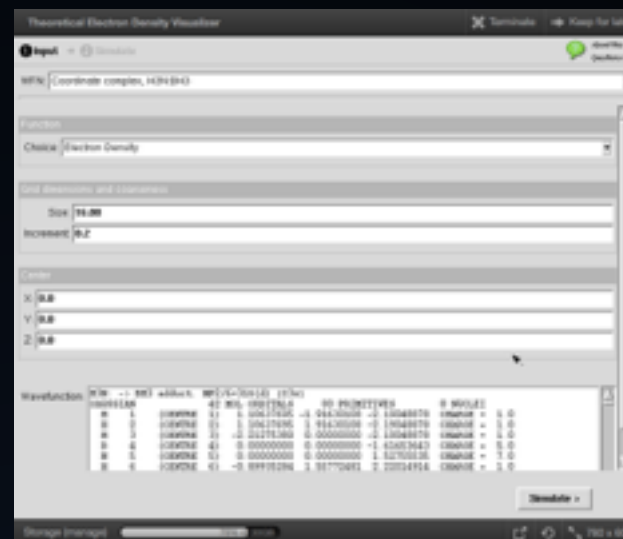
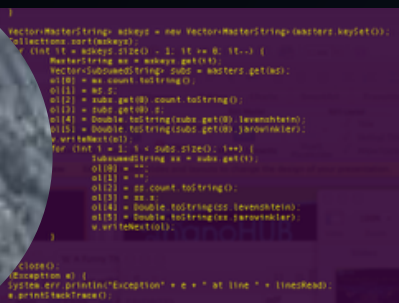
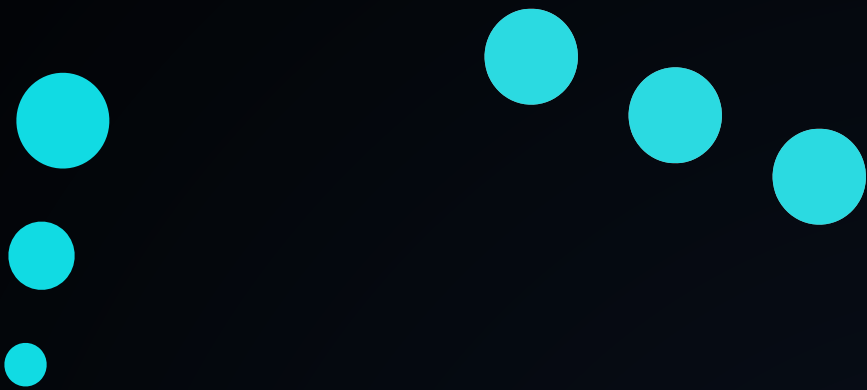




NANOHUB IN A DATA ECOSYSTEM

Michael Zentner


```
117     }
118
119     Vector<MasterString> wkkeys = new Vector<MasterString>(masters.keySet());
120     Collections.sort(wkkeys);
121     for (int it = wkkeys.size() - 1; it >= 0; it--) {
122         MasterString ks = wkkeys.get(it);
123         Vector<SubwordString> subs = masters.get(ks);
124         ol[0] = ks.count.toString();
125         ol[1] = ks.S;
126         ol[2] = subs.get(0).count.toString();
127         ol[3] = subs.get(0).S;
128         ol[4] = Double.toString(subs.get(0).levenshtein);
129         ol[5] = Double.toString(subs.get(0).jarowinkler);
130         w.writeNext(ol);
131         for (int i = 1; i < subs.size(); i++) {
132             SubwordString ss = subs.get(i);
133             ol[0] = "";
134             ol[1] = "";
135             ol[2] = ss.count.toString();
136             ol[3] = ss.S;
137             ol[4] = Double.toString(ss.levenshtein);
138             ol[5] = Double.toString(ss.jarowinkler);
139             w.writeNext(ol);
140         }
141     }
142     w.close();
143 } catch (Exception e) {
144     System.err.println("Exception" + e + " at line " + linesRead);
145     e.printStackTrace();
146 }
147 }
```



NANOHUB TODAY IS 350,000 ANNUAL USERS INTERNATIONALLY



370+ Simulation Tools
4700+ Other Resources
3000+ Online Presentations
540+ Teaching Materials



At No Cost: nanohub.org



WHAT WE DID NOT DO

Create the “Science”
Create / Enforce Coding Standards



WHAT WE DID DO

Harvested the “Science”
Generated an Audience
Helped People Explore Simulations in Interesting Ways
(Enabled Uncertainty Quantification)

A DATA ACTIVITY SPECTRUM



Create



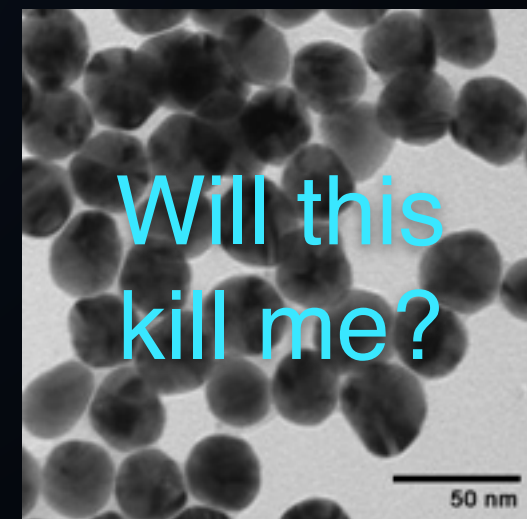
(Standardize & Store
& “Publish”)



Process



Analyze



Will this
kill me?

Answer



A DATA ACTIVITY SPECTRUM



Process



Analyze



Answer



A DATA ACTIVITY SPECTRUM



Process



Analyze



Answer

Cannot plan for every question...

Must let the human intervene...

Stored data not necessarily in best form to answer questions...



STORAGE VS QUESTIONS



STORAGE VS QUESTIONS



STORAGE VS QUESTIONS



A DATA ACTIVITY SPECTRUM

Interactive Visualization
& Exploration

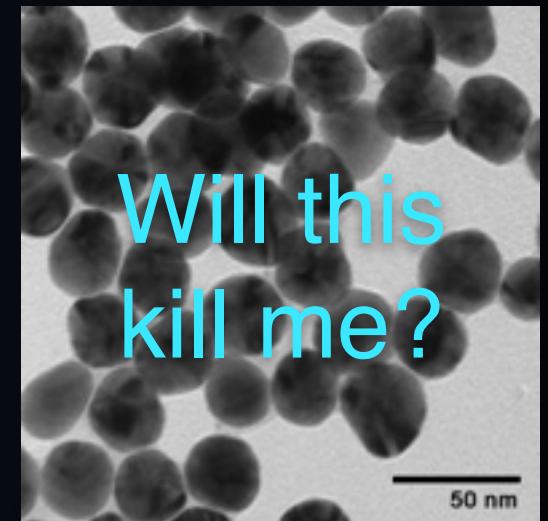
Simulation



Process



Analyze



Answer

Combination
(e.g. LD50 data)

Cannot plan for every question...

Must let the human intervene...

Stored data not necessarily in best form to answer questions...



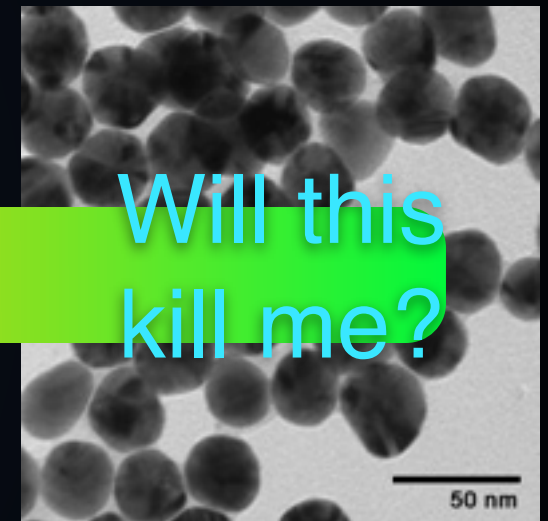
A DATA ACTIVITY SPECTRUM



Process



Analyze



Answer

Cannot plan for every question...

Must let the human intervene...

Stored data not necessarily in best form to answer questions...



A DATA ACTIVITY SPECTRUM



Create



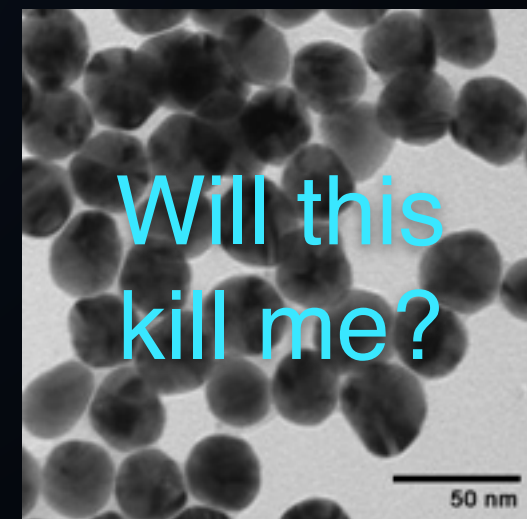
(Standardize &)
Store
(& “Publish”)



Process



Analyze



Will this
kill me?

Answer



MORE THAN ONE GLOBAL DATA STORE, FORMAT, EVEN SHAPE



MORE THAN ONE GLOBAL DATA STORE, FORMAT, EVEN SHAPE



DENSE DATA

Within Your Experiment, Your Knowledge is Somewhat Complete
(in your home you know what you own)

Sphere:	Gold				
Medium:	Water				
Graph type:	Cext/Csca/Cabs v. wavelength				
Scattering angle = 0					
Wavelength	Ref Index (real)	Ref Index (imaginary)	Cext	Csca	Cabs
400	1.087871431	1.446879408	2.04E-15	1.23E-16	1.92E-15
401	1.087514072	1.447303652	2.09E-15	1.33E-16	1.96E-15
402	1.087156632	1.447727993	2.09E-15	1.31E-16	1.95E-15
403	1.08679913	1.448152407	2.08E-15	1.30E-16	1.95E-15
404	1.086441546	1.448576918	2.07E-15	1.29E-16	1.95E-15
405	1.086083131	1.449000474	2.07E-15	1.28E-16	1.94E-15
406	1.085723491	1.449422573	2.06E-15	1.27E-16	1.94E-15
407	1.085363769	1.449844767	2.06E-15	1.26E-16	1.93E-15
408	1.085003987	1.450267031	2.05E-15	1.24E-16	1.93E-15
409	1.084644123	1.450689391	2.05E-15	1.23E-16	1.93E-15

← Dense



(SOMEWHAT) SPARSE DATA

In Your Immediate Community, Knowledge is Less Complete
(you know many of the things your neighbors own)

Liver Uptake

Brain Uptake

Gold



Sphere:	Gold				
Medium:	Water				
Graph type:	Cext/Csca/Cabs v. wavelength				
Scattering angle = 0					
Wavelength	Ref Index (real)	Ref Index (imaginary)	Cext	Csca	Cabs
400	1.087871431	1.446879408	2.04E-15	1.23E-16	1.92E-15
401	1.087514072	1.447303652	2.09E-15	1.33E-16	1.96E-15
402	1.087156632	1.447727993	2.09E-15	1.31E-16	1.95E-15
403	1.08679913	1.448152407	2.08E-15	1.30E-16	1.95E-15
404	1.086441546	1.448576918	2.07E-15	1.29E-16	1.95E-15
405	1.086083131	1.449000474	2.07E-15	1.28E-16	1.94E-15
406	1.085723491	1.449422573	2.06E-15	1.27E-16	1.94E-15
407	1.085363769	1.449844767	2.06E-15	1.26E-16	1.93E-15
408	1.085003987	1.450267031	2.05E-15	1.24E-16	1.93E-15
409	1.084644123	1.450689391	2.05E-15	1.23E-16	1.93E-15

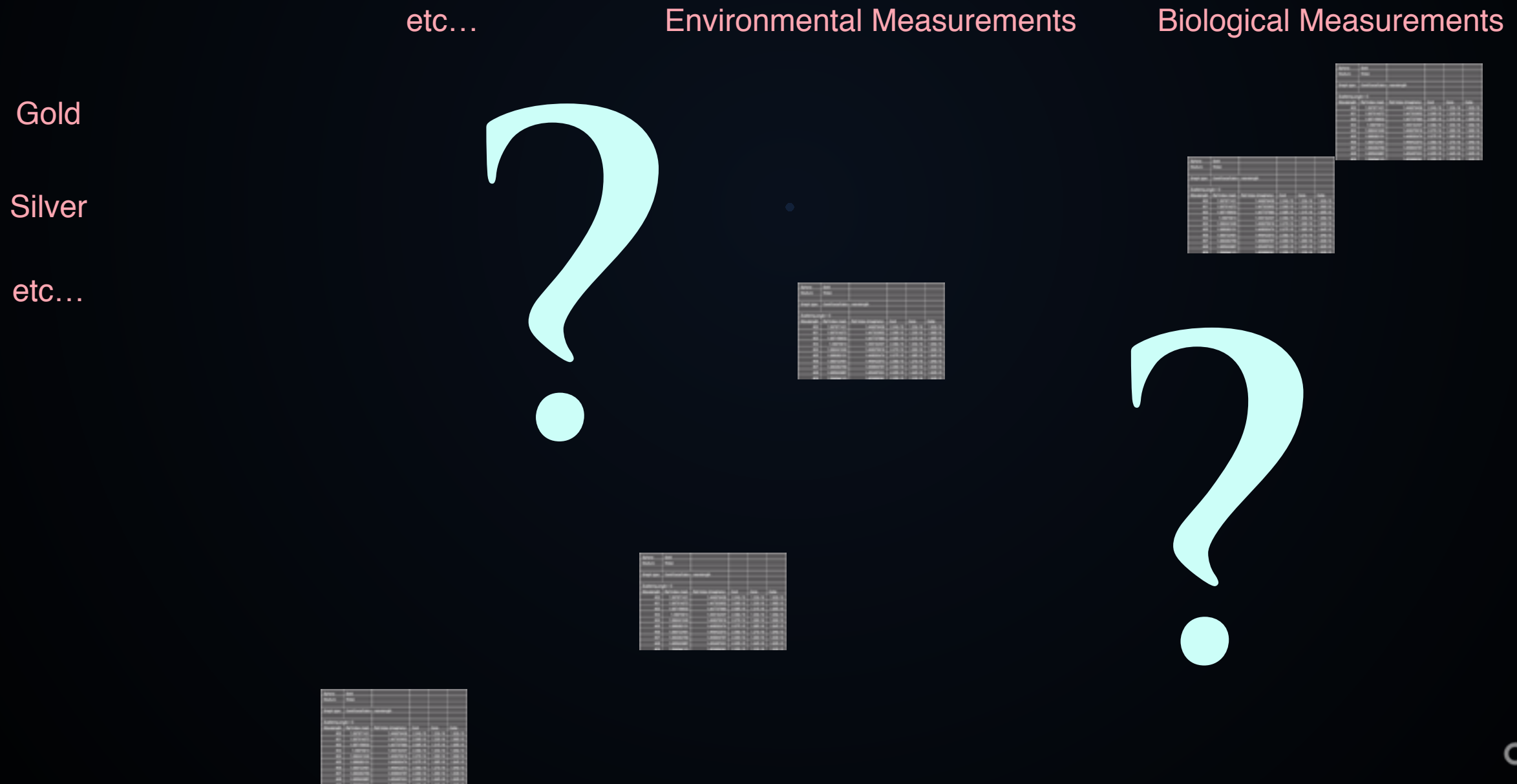
Silver

Sphere:	Gold				
Medium:	Water				
Graph type:	Cext/Csca/Cabs v. wavelength				
Scattering angle = 0					
Wavelength	Ref Index (real)	Ref Index (imaginary)	Cext	Csca	Cabs
400	1.087871431	1.446879408	2.04E-15	1.23E-16	1.92E-15
401	1.087514072	1.447303652	2.09E-15	1.33E-16	1.96E-15
402	1.087156632	1.447727993	2.09E-15	1.31E-16	1.95E-15
403	1.08679913	1.448152407	2.08E-15	1.30E-16	1.95E-15
404	1.086441546	1.448576918	2.07E-15	1.29E-16	1.95E-15
405	1.086083131	1.449000474	2.07E-15	1.28E-16	1.94E-15
406	1.085723491	1.449422573	2.06E-15	1.27E-16	1.94E-15
407	1.085363769	1.449844767	2.06E-15	1.26E-16	1.93E-15
408	1.085003987	1.450267031	2.05E-15	1.24E-16	1.93E-15
409	1.084644123	1.450689391	2.05E-15	1.23E-16	1.93E-15



(MORE) SPARSE DATA

In Your Extended Community, Knowledge is Less Complete
(you know many of the things your neighbors own)



(VERY) SPARSE DATA

In A Wide Community, Knowledge is Very Incomplete
(you don't know what most people on earth own)

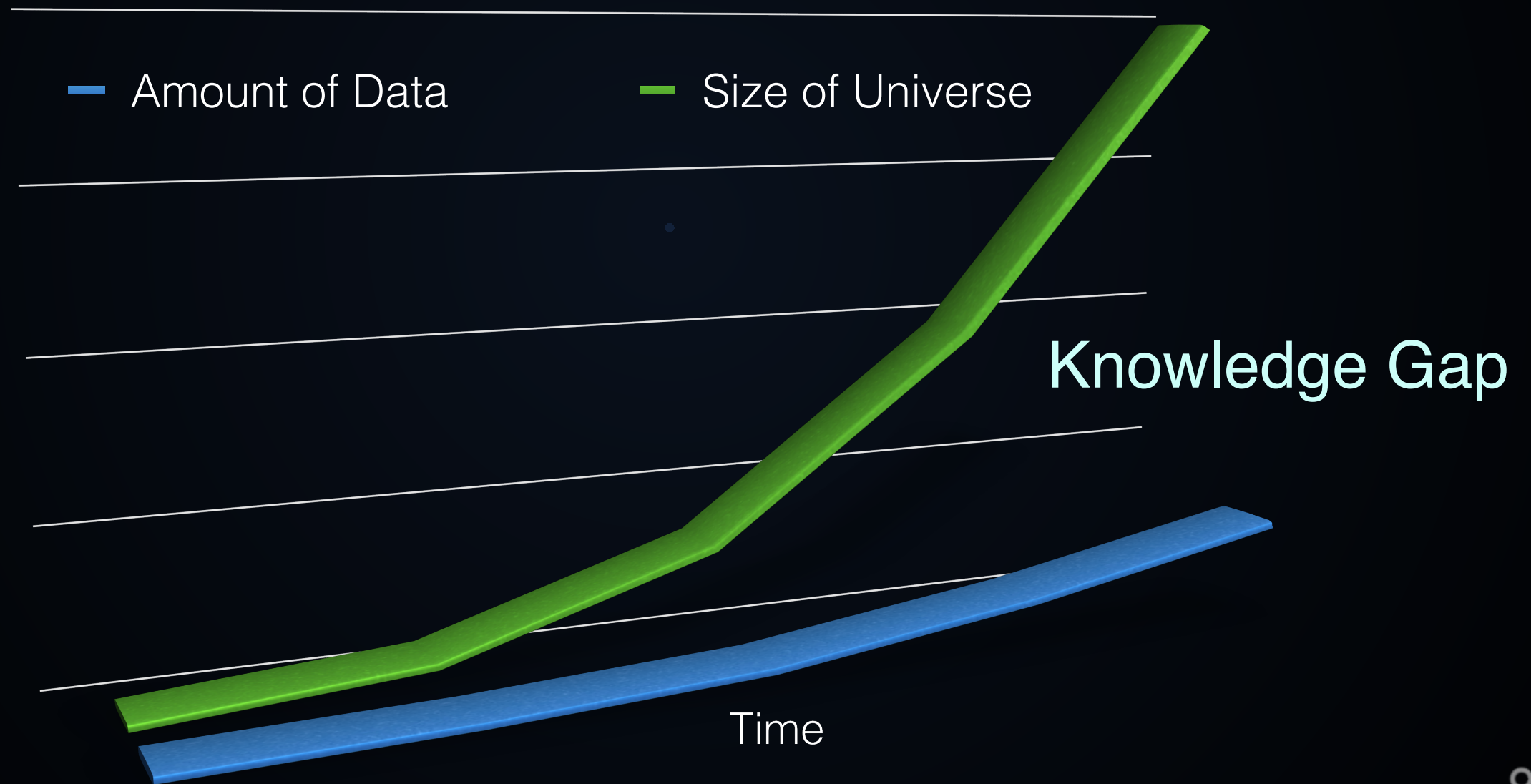
Electrical - Environmental - Medical - Physical - Financial - Interfacial - etc...

Wide
Variety
of
Materials



KNOWLEDGE VS UNIVERSE

The more we have, the less we know...



THE TROUBLE WITH TABLES...


Dense tables are
wonderful for sort,
search, and filtration

Run Meta Data				Inputs				Outputs			
User	Time	Tool	etc.	I ₁	I ₂	I ₃	etc	O ₁	O ₂	O ₃	etc

Sparse tables are
terrible for sort,
search, and filtration



EXPLORATION

Run Meta Data				Inputs				Outputs			
User	Time	Tool	etc.	I ₁	I ₂	I ₃	etc	O ₁	O ₂	O ₃	etc
											



EXPLORATION

How do we effectively explore a sparse data landscape?



IDENT: PARTNERSHIP WITH NANOMATERIAL REGISTRY

Interactive Data Exploration & Navigation Tool
for Nano Technology

ID	NMRI data
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	



IDENT

COLUMN POPULATION LOOKAHEAD

Nothing selected Use columns Custom column

Search

1	Absorbance:as a function of Solvent
1	Absorbance:as a function of Time
2	pH
3	Size:as a function of Light Exposure
4	Size:as a function of Temperature
6	Surface Area
8	Bulk Density (g/cm^3)
9	Specific Surface Area
1	Surface Charge
1	Electrokinetic Mobility
1	Zeta Potential (mV)

Bar Length Indicates Amount of Non-Null Data




IDENT

COLUMN OF INTEREST SELECTION

Nothing selected ▼ Use columns Custom column

288 total rows Edit filters

Table of chosen data

ID	Mean Diameter (nm)	Mean Hydrodynamic Diameter (nm)	DisplayName	
6			Ag	
7	7.2		Ag NP	
8	20.8	32	Ag NP	
9	29	39.6	Ag NP	
10	43.4	47.6	Ag NP	
11	41.9	43.1	Ag NP	
12	49.1	59.2	Ag NP	
13	53.5	57.1	Ag NP	
14	57.7	68.3	Ag NP	
15	8.2		Ag NP	
16	19.2		Ag NP	

Nanomaterials Registry Dataset

Visualize Content

Add a visualization +



IDENT

VISUALIZATION OF CONTENT

Before we take another step, we know what we will find.



Mean Hydrodynamic Diameter (nm) ▲

Show column

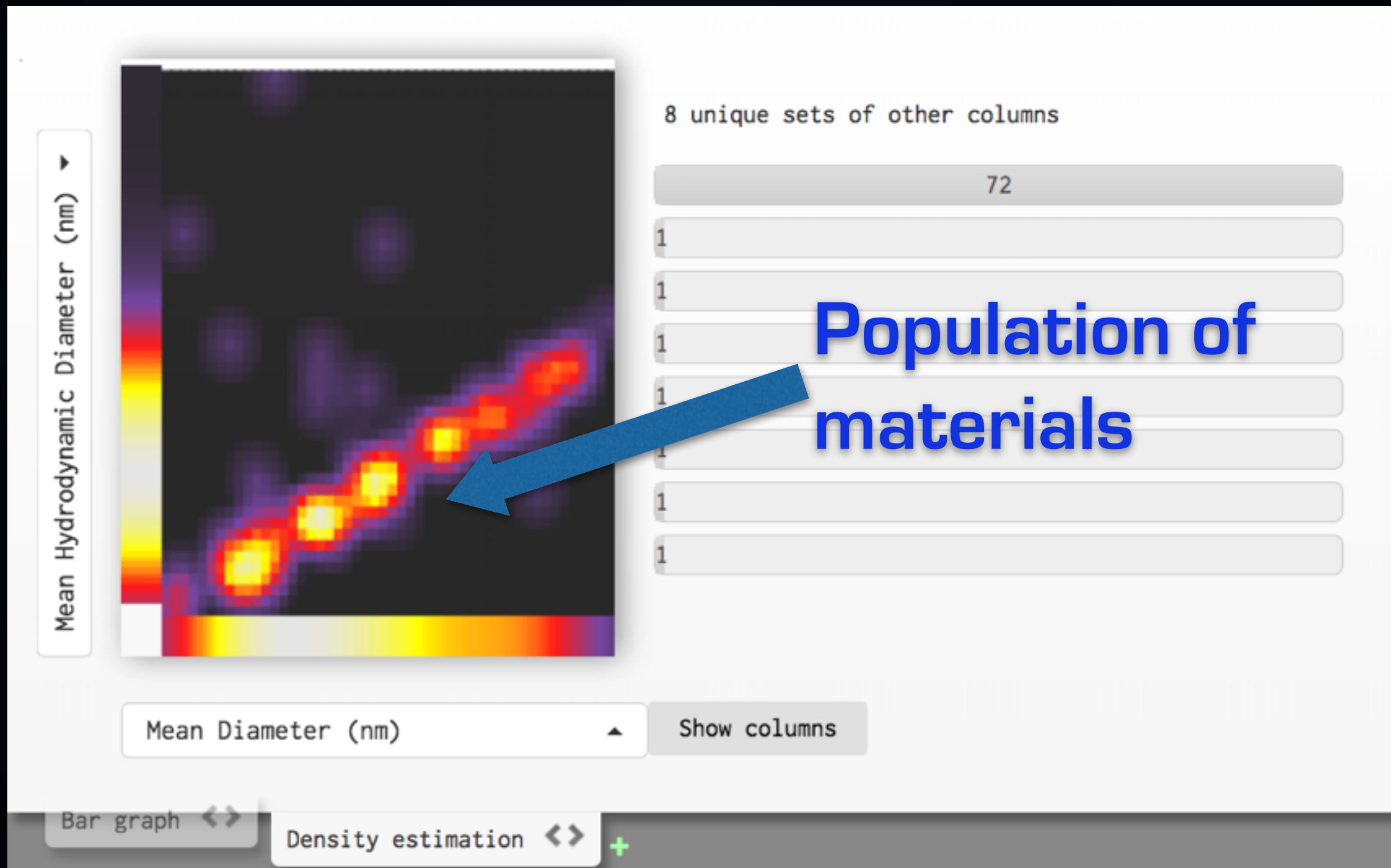
☒ Buckets: `Math.sqrt(count)`

Mean Diameter (nm) x

Bar graph <>

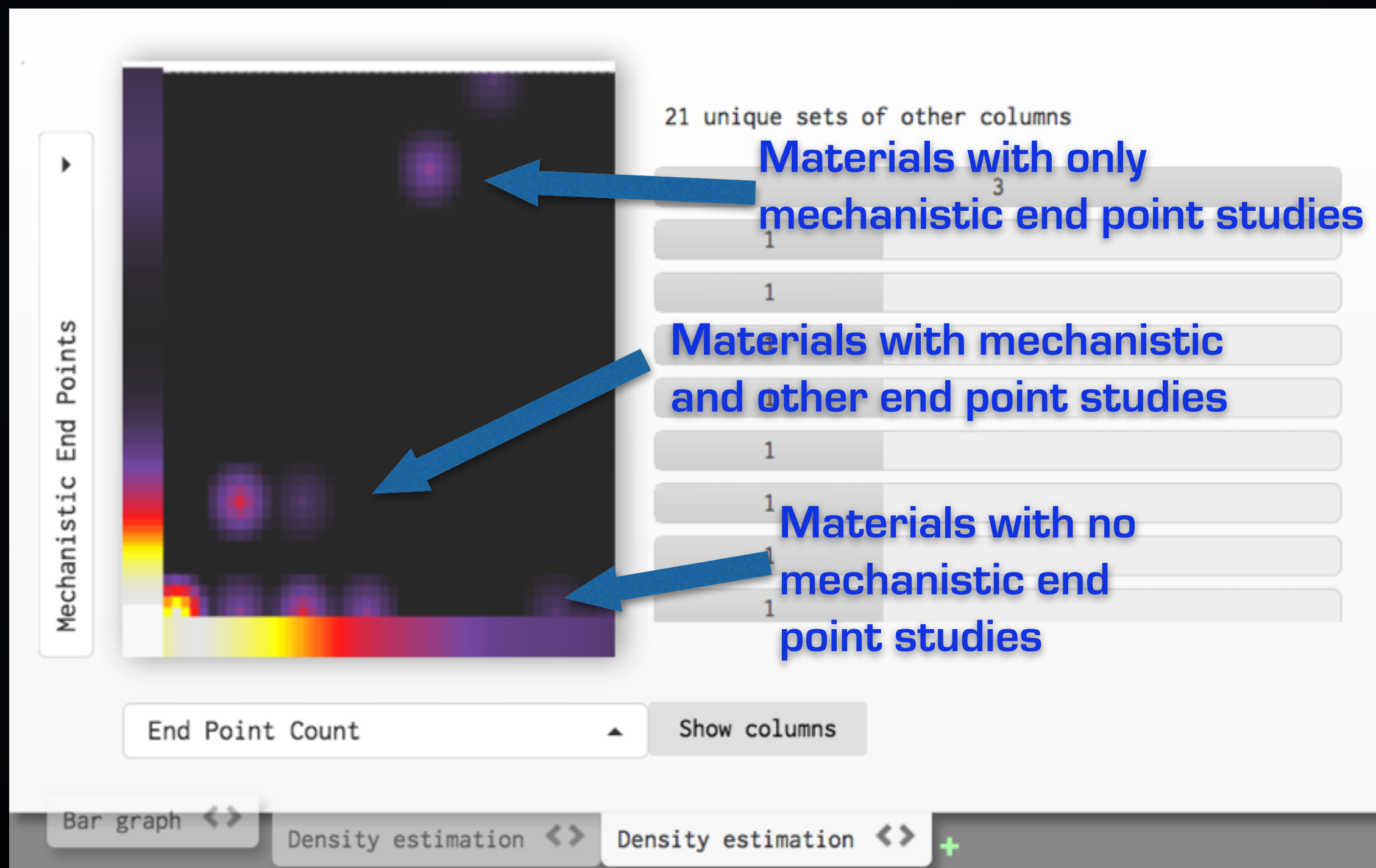


IDENT VISUALIZATION OF CONTENT



IDENT

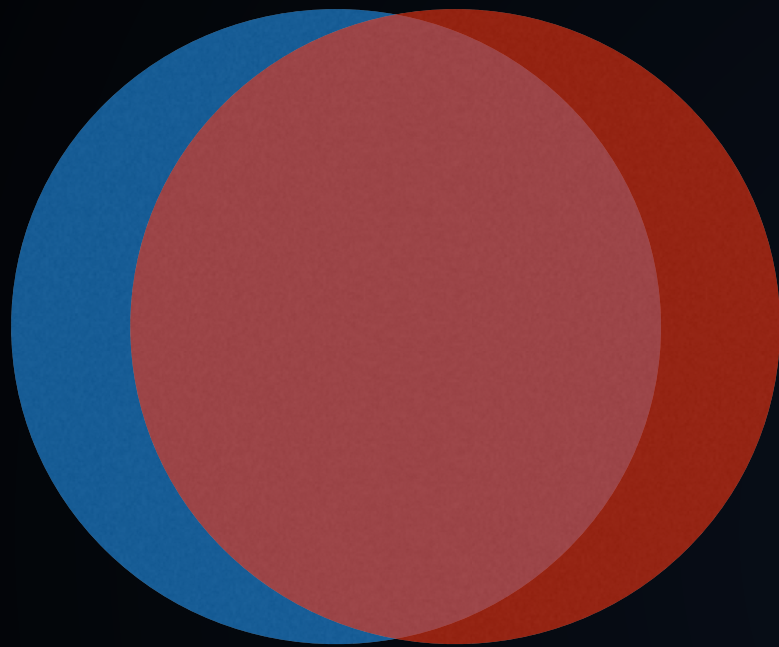
VISUALIZATION OF KNOWLEDGE



THE STUDY OF HOW PEOPLE STUDY DATA



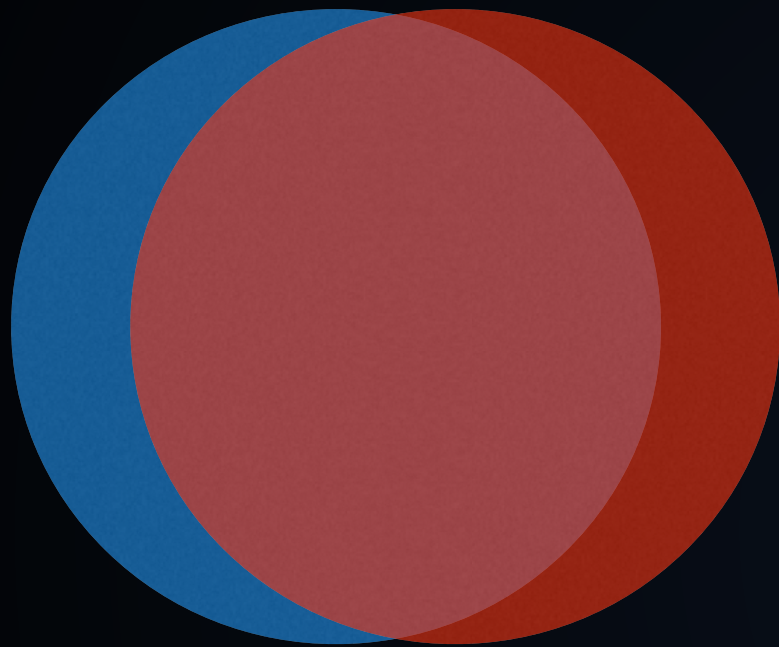
THE STUDY OF HOW PEOPLE STUDY DATA



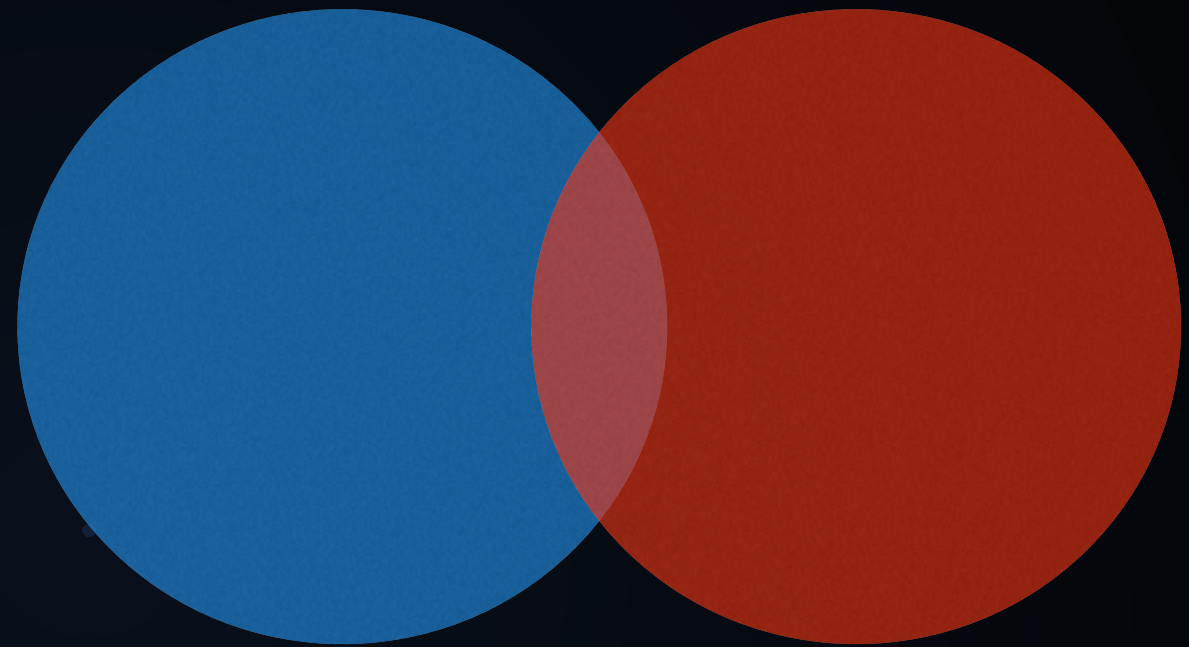
Small Digressions



THE STUDY OF HOW PEOPLE STUDY DATA



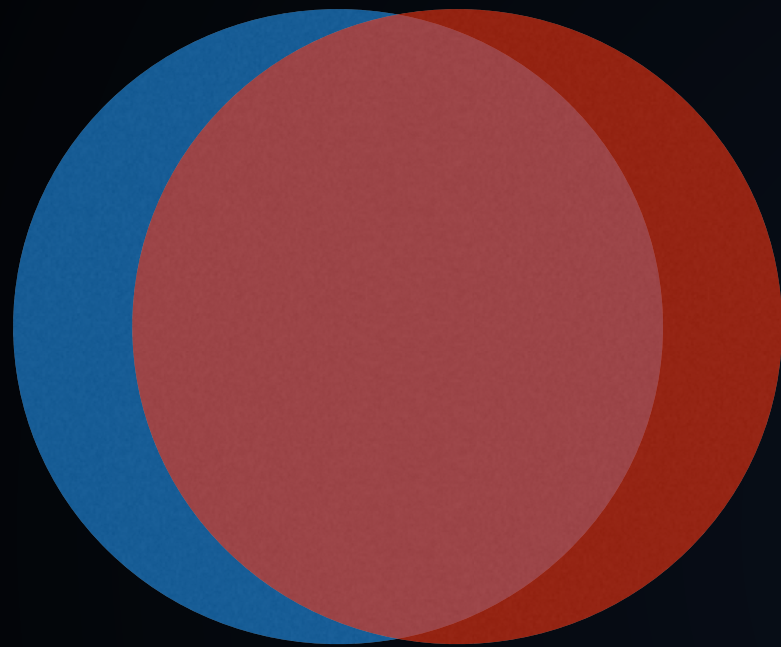
Small Digressions



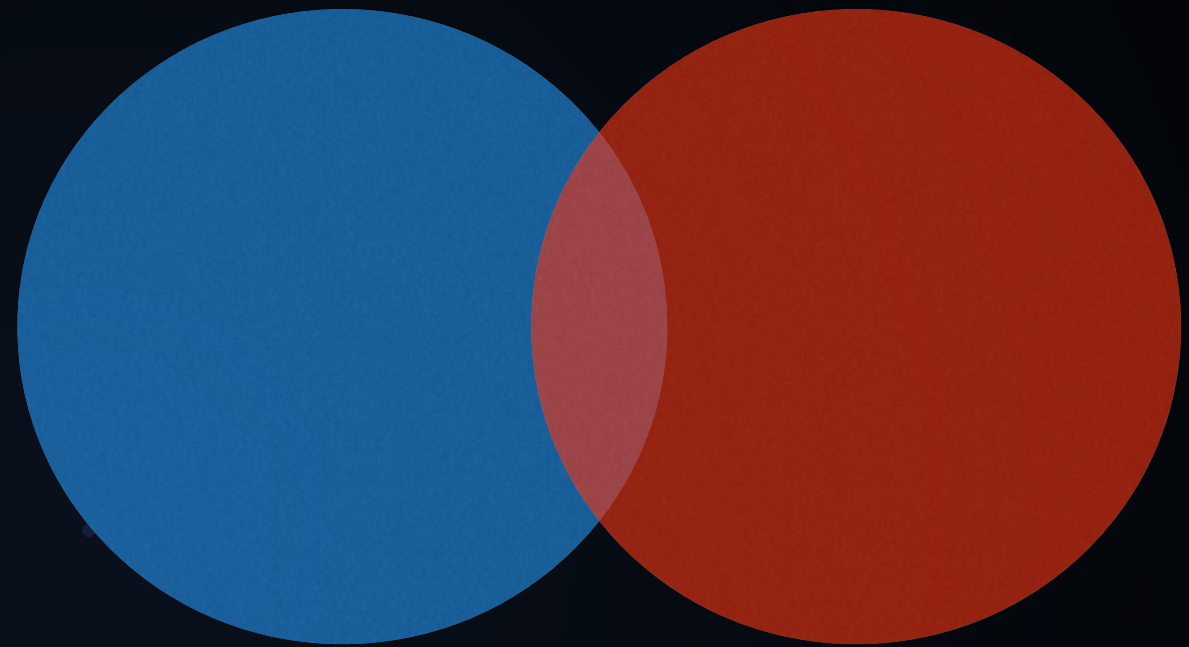
New Directions



THE STUDY OF HOW PEOPLE STUDY DATA



Small Digressions



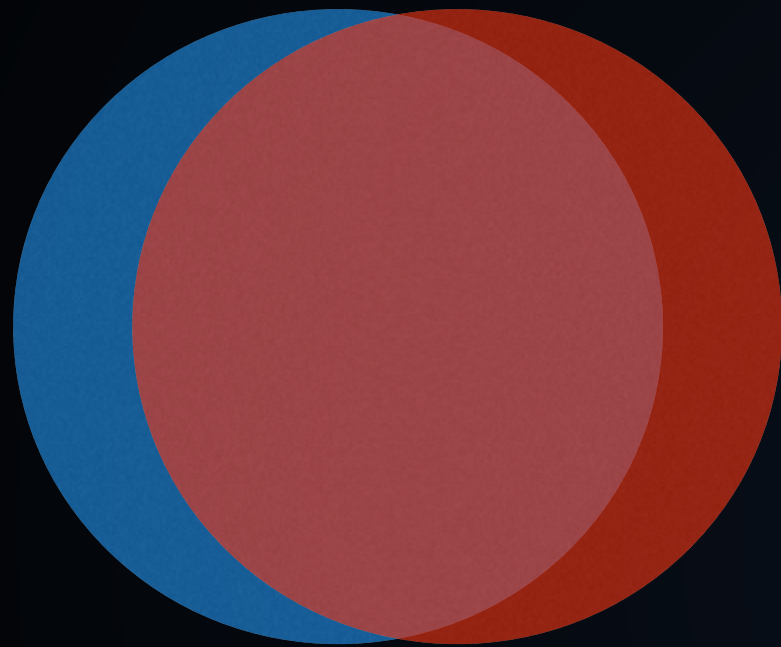
New Directions



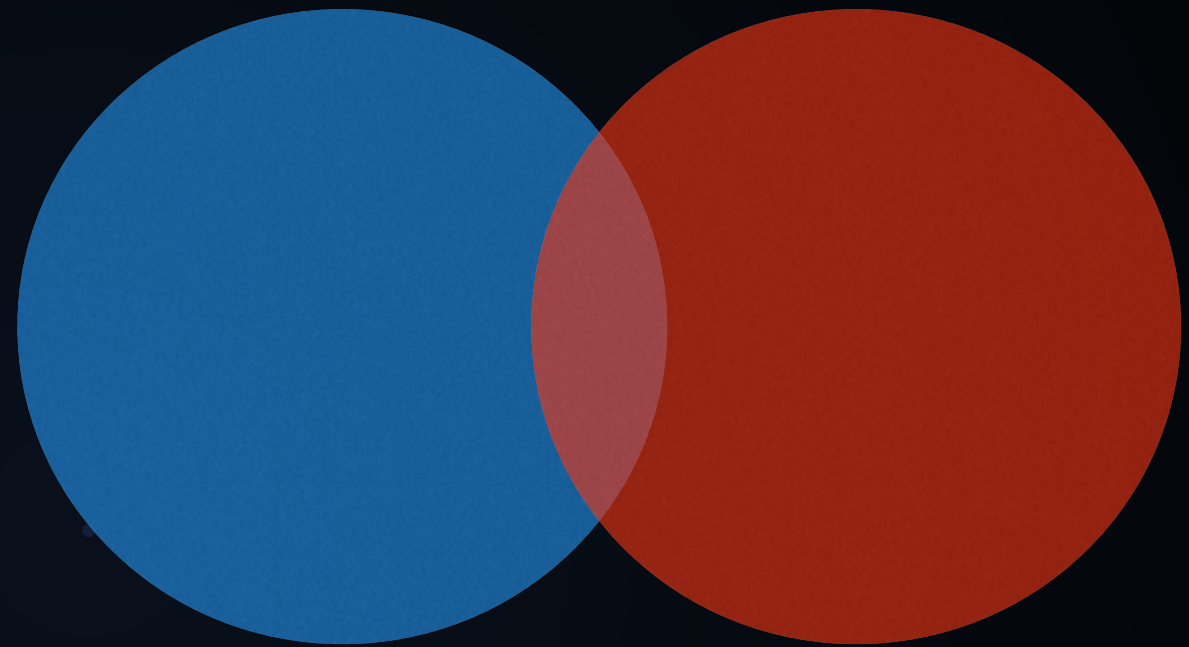
New Fields



THE STUDY OF HOW PEOPLE STUDY DATA



Small Digressions



New Directions



New Fields



?????





WHAT WE WANT TO DO

- Provide Interesting Ways to Explore Data
 - Provide Lensing Capability to Help People Combine Data
 - Study How People Study Data (social connections)
 - Find More Data Partners
- NOT become the master repository of data or ontologies