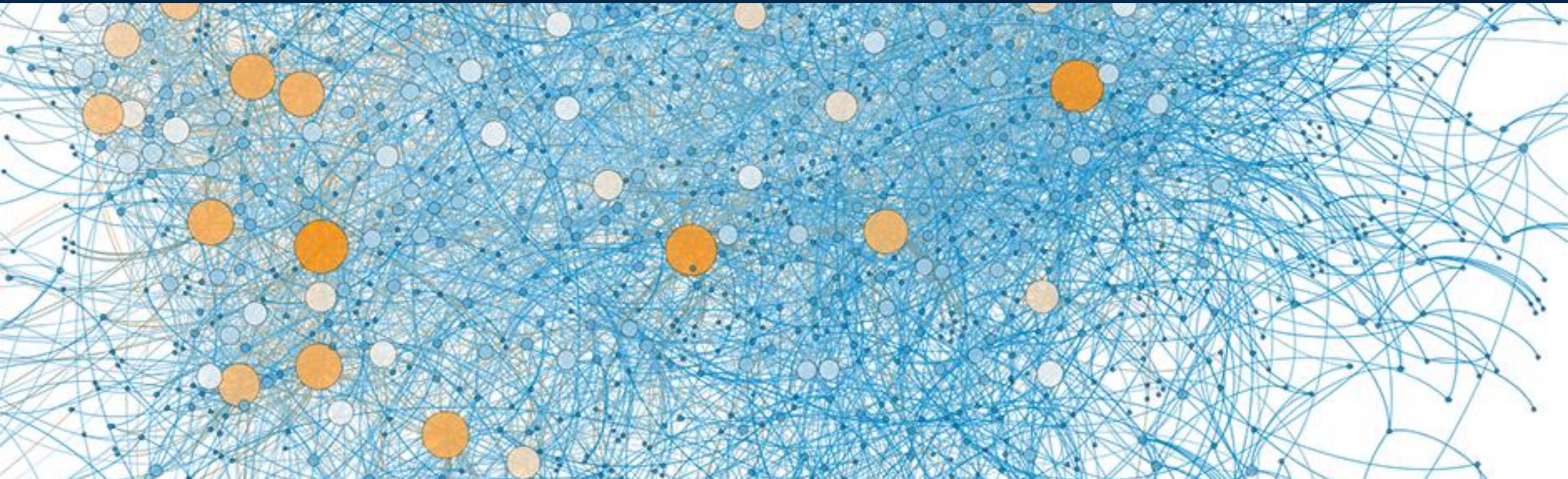


# Environmental Health ... In Context

Plus some obligatory network visualization eye-candy!

Sam Adams, Senior AI Researcher, RTI International



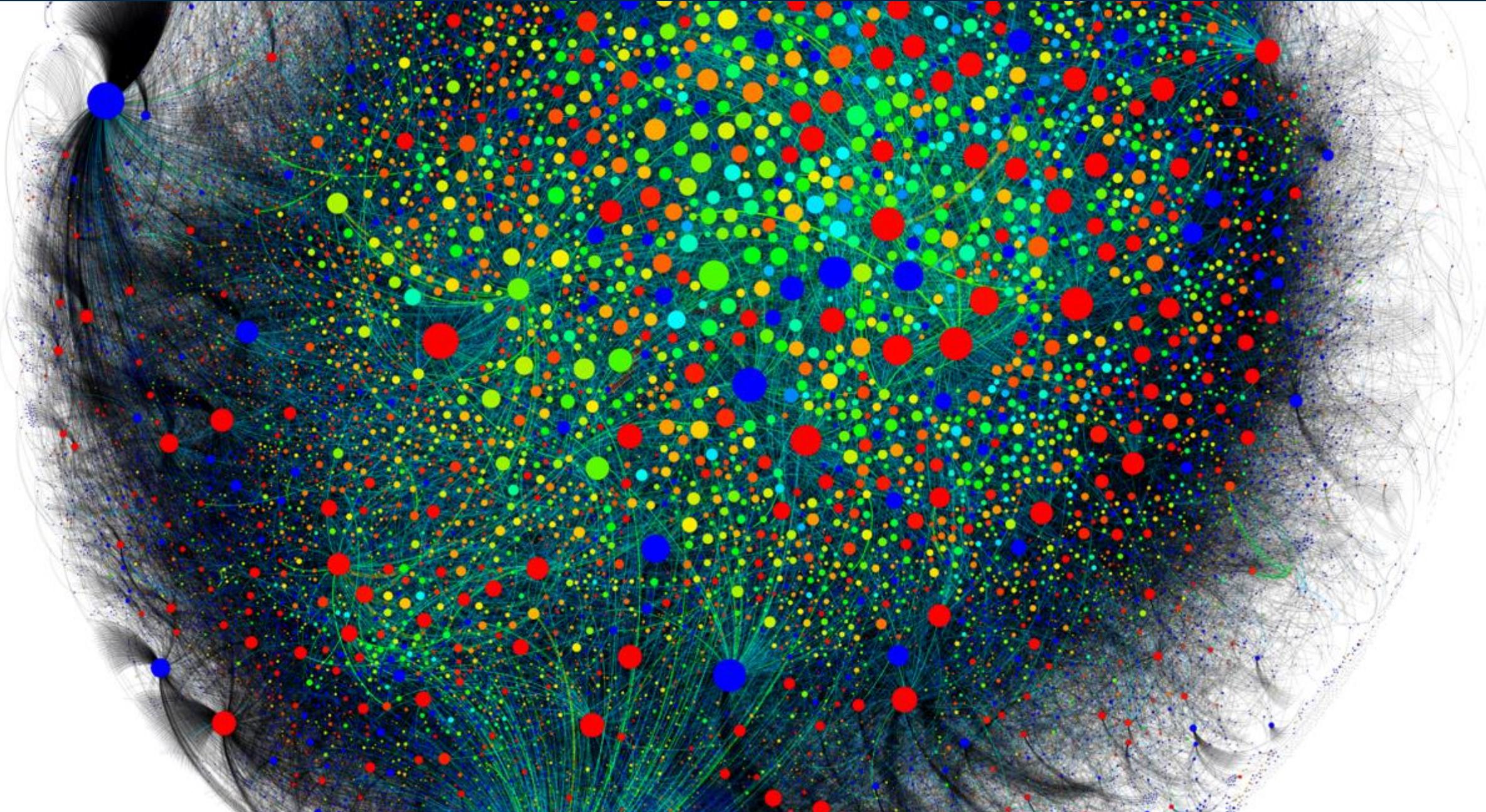
# (nearly) Everything I know about Environmental Health I learned in kindergarten

DISCLAIMER

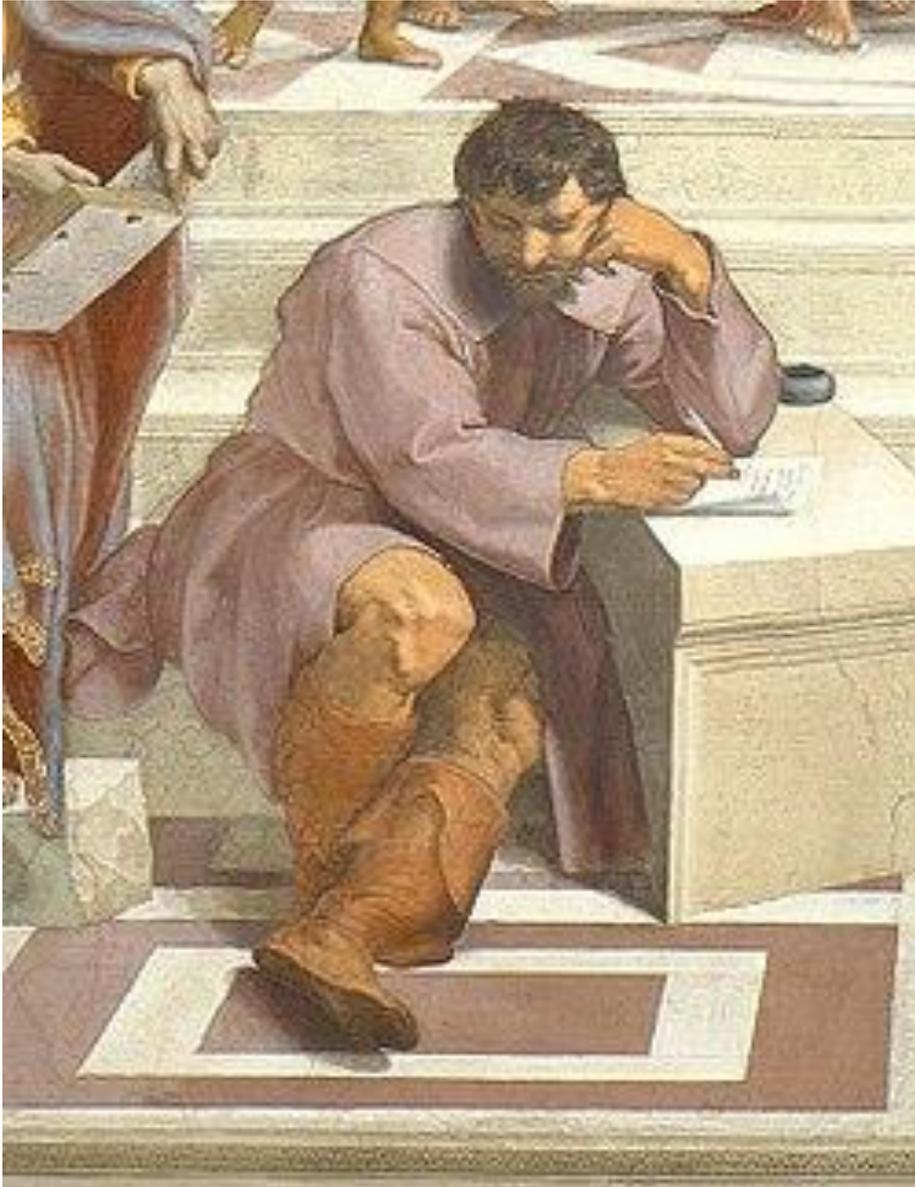


There's a germ in the blood  
Of the worm in the beak  
Of the bird on the egg  
In the nest on the branch  
On the limb on the trunk  
Of the tree in the hole  
In the ground in the woods  
By the meadow near a stream  
Running down to the river  
Flowing into the ocean  
With the sun on the water  
And the vapor in the air  
Rising up to the clouds  
Blowing over the land  
Chilling down in the front  
Coalescing to drops  
Falling out of the sky  
As the rain on the leaves  
Of the tree in the meadow  
Where the bird has a nest  
And the raindrops fall all-around...

In the environment, everything is connected... And its beautiful!



But this is not a new thought...



A wonderful harmony is created when we join together the seemingly unconnected

Heraclitus  
(535–475 BC)

What does this mean?

**40**

What does this mean?

**40°**

What does this mean?

**40°C**

What does this mean?

40°C



What does this mean?

40°C



What does this mean?



What does this mean?



What does this mean?



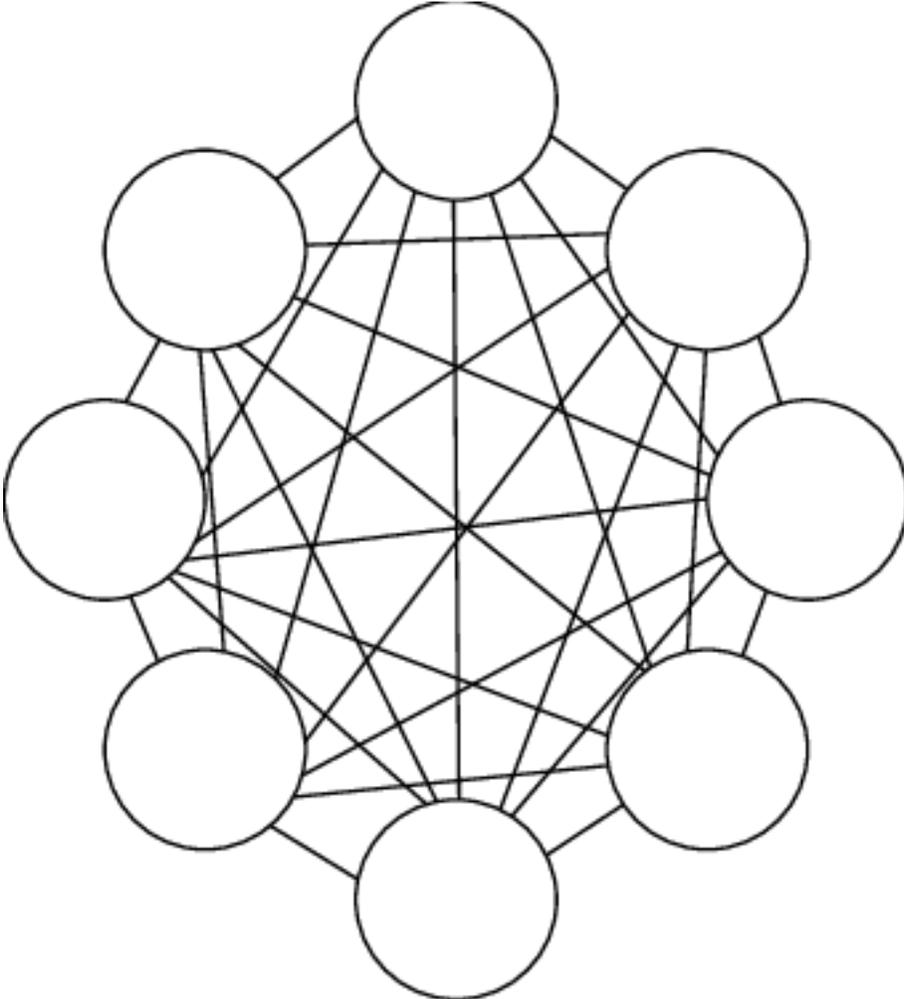
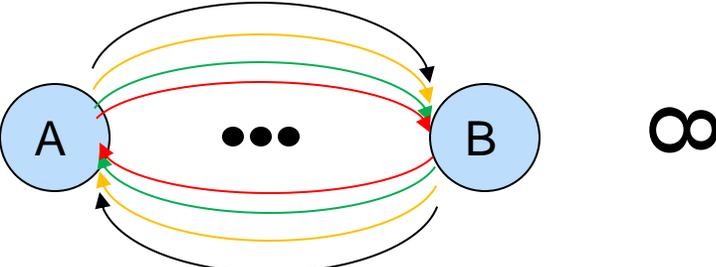
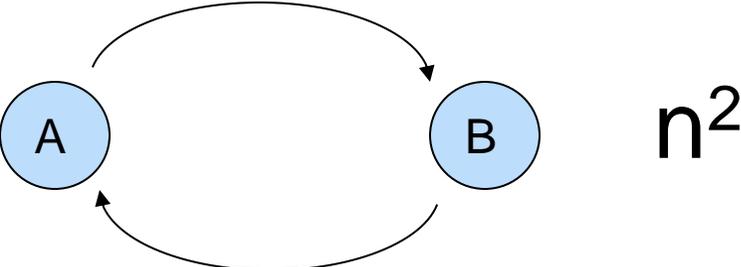
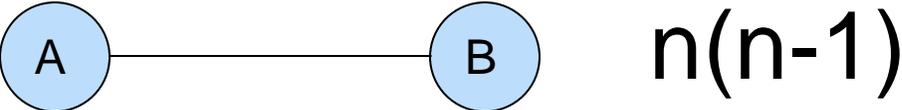
# What does this mean?



**Date: July 11, 1984**

# Connecting the dots

One piece of information is almost worthless in isolation.  
It is context that creates value, in science and in business.  
The connection is often worth more than the data it connects.  
The connection provides the meaning.  
Insight itself discovers or derives new connections.



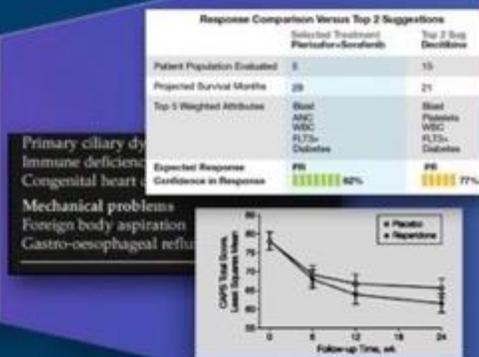
# Analytics and the Context Multiplier Effect

1x



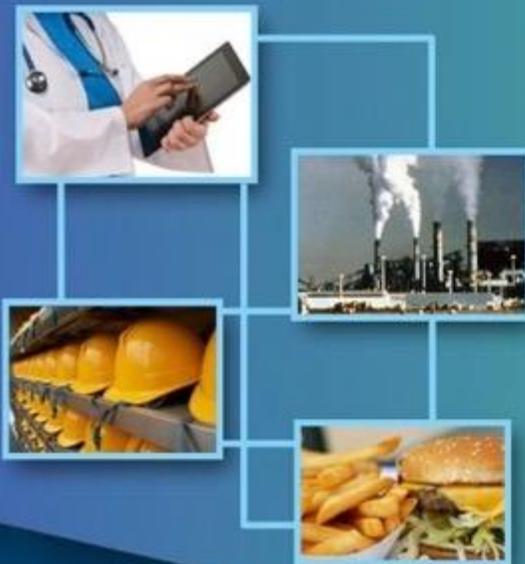
Raw Data

10x



Feature  
extraction metadata

100x



Domain linkages

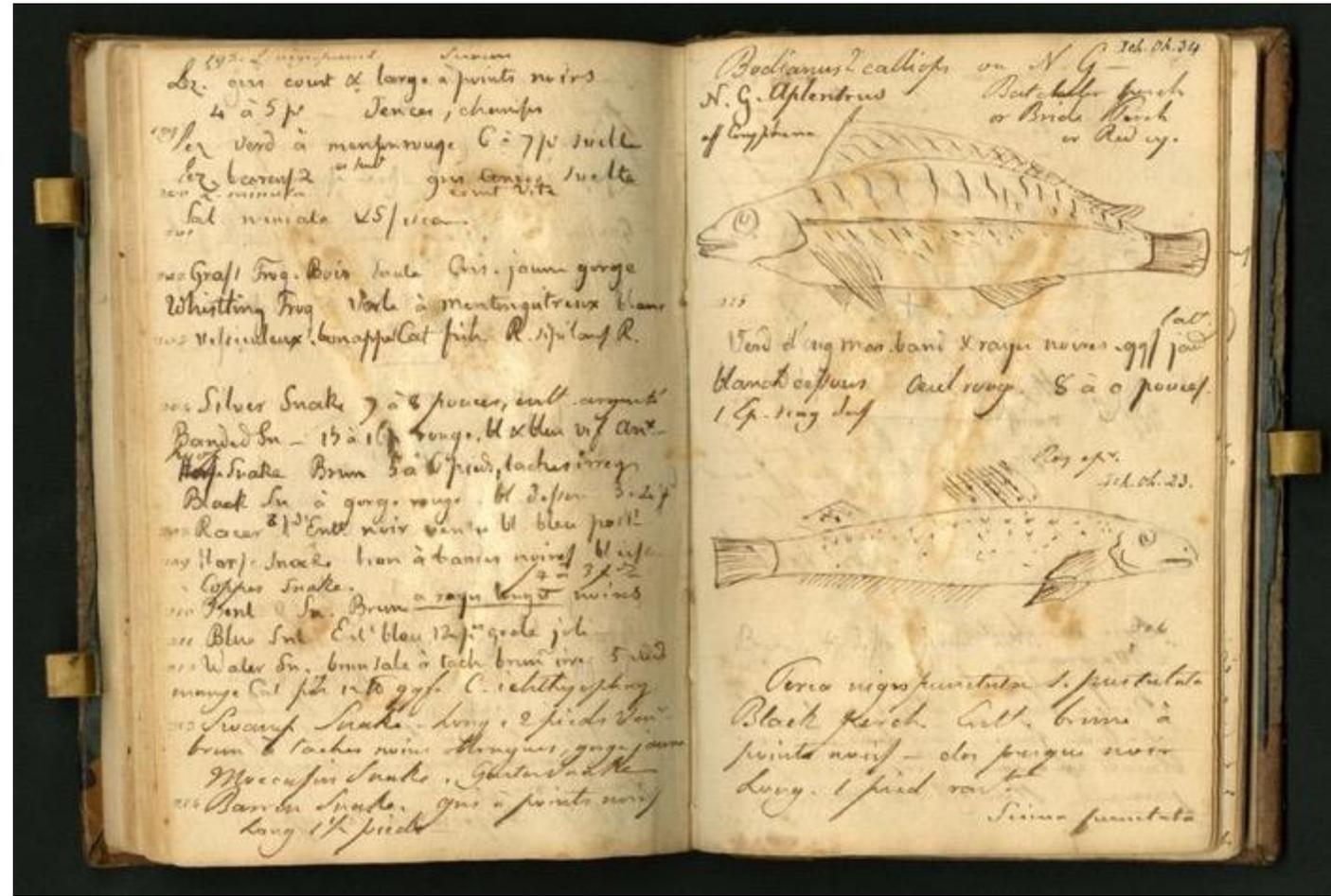
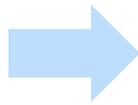
> 1,000x



Full  
contextual analytics

**So what does this all have to do with  
Environmental Health?**

# In the beginning, there was Observation and Description



# Sampling and Measurement



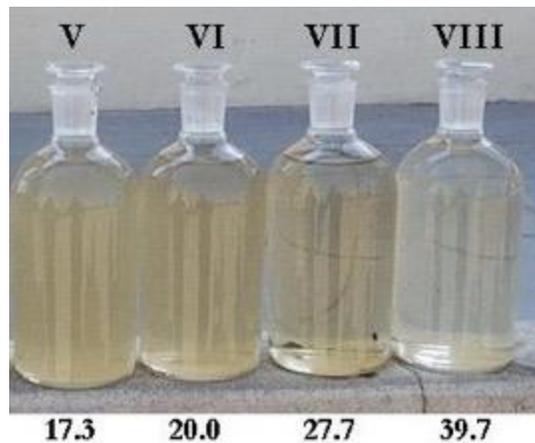
0 5.2 9.6 13.9



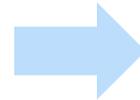
17.3 20.0 27.7 39.7

Distance from City (km)

# Analysis and Metric Standardization

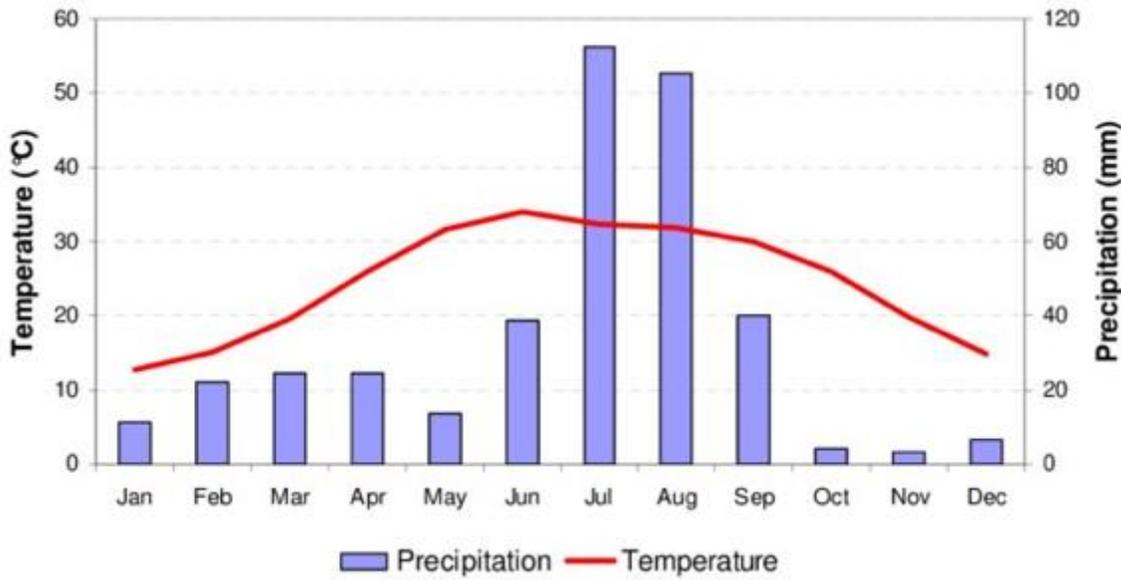
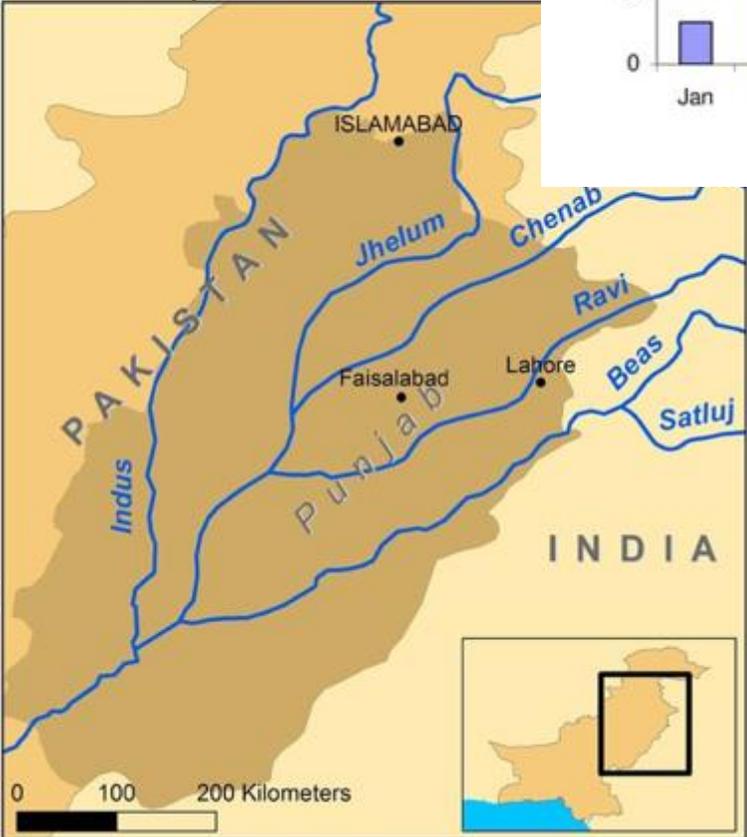


Distance from City (km)

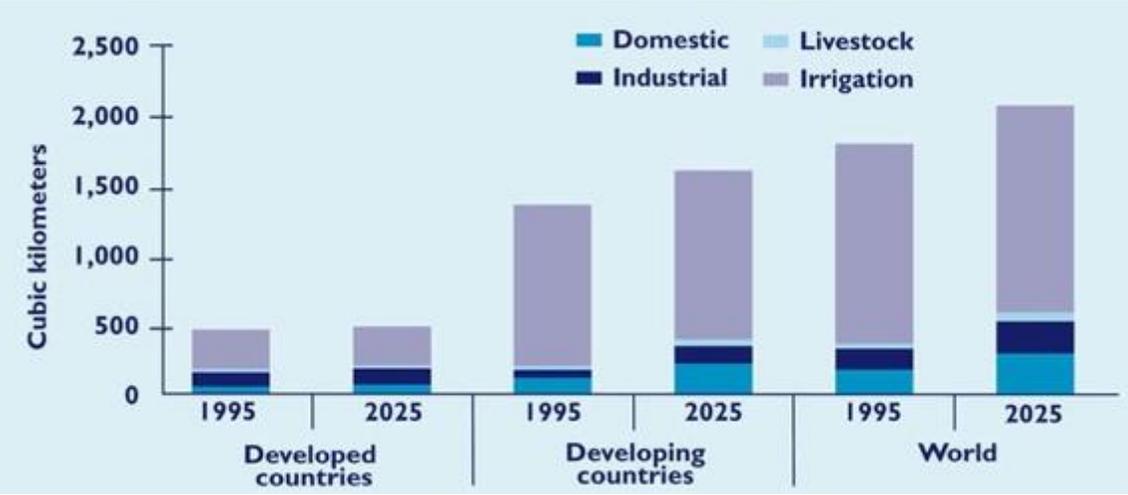
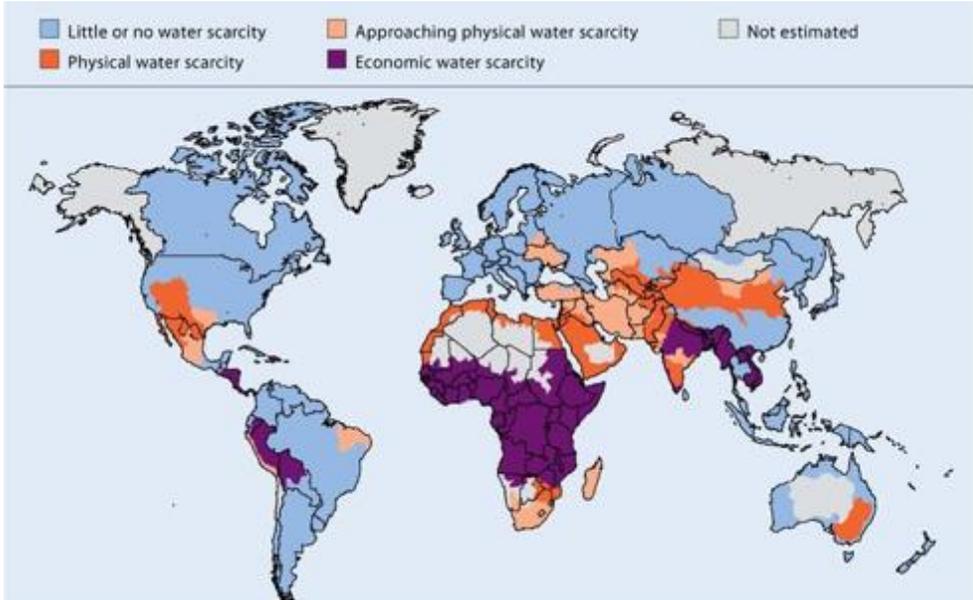
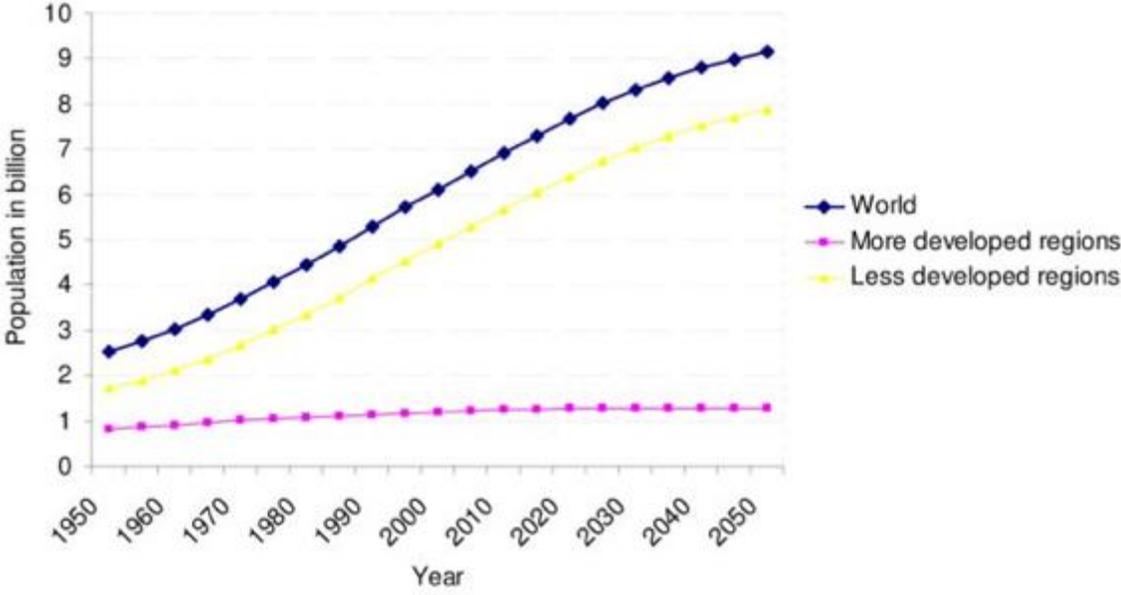


DETERMINANTS CONSTITUENTS	Results expressed in mg/l unless specifically stated.	
	Purified Rainwater	WHO-QC Standard Recommended Limits
Conductivity ( $\mu\text{S}/\text{cm}$ )	148	Not Specified
Total Dissolved Solids (TDS @ 105 °C)	103	< 500
Total Suspended Solids (TSS)	BDL	Must be absent
Turbidity (NTU)	0.16	< 0.5 (Target < 0.2)
NTU after 0.2- $\mu$ lab filtration test	0.14	Shows filterable content
Colour (Pt/Co Units)	15	< 30 (Target < 20)
Odour (Instrumentation) (TON)	Not Done	< 5.0 (Target < 4.0)
pH	7.12	6.5 to 8.5
Alkalinity (Total)	29.3	20 to 200
Total Hardness (TDH as $\text{CaCO}_3$ )	3.73	20 to 200
Calcium (Ca as $\text{CaCO}_3$ )	3.44	10 to 200
Magnesium (Mg as $\text{CaCO}_3$ )	0.29	5 to 150
Aluminium (Al) ( $\mu\text{g}/\text{l}$ )	BDL	< 150
Copper (Cu) ( $\mu\text{g}/\text{l}$ )	7	< 200
Iron (Fe) Total ( $\mu\text{g}/\text{l}$ )	BDL	< 200 (Target < 20)
Iron (Fe – Dissolved) ( $\mu\text{g}/\text{l}$ )	BDL	Target < 20
Manganese (Mn) Total ( $\mu\text{g}/\text{l}$ )	BDL	< 50 (Target < 10)
Total Heavy Metal Content ( $\mu\text{g}/\text{l}$ )	BDL	< 10 (Target = 0)

# Adding Regional Hydrology Context

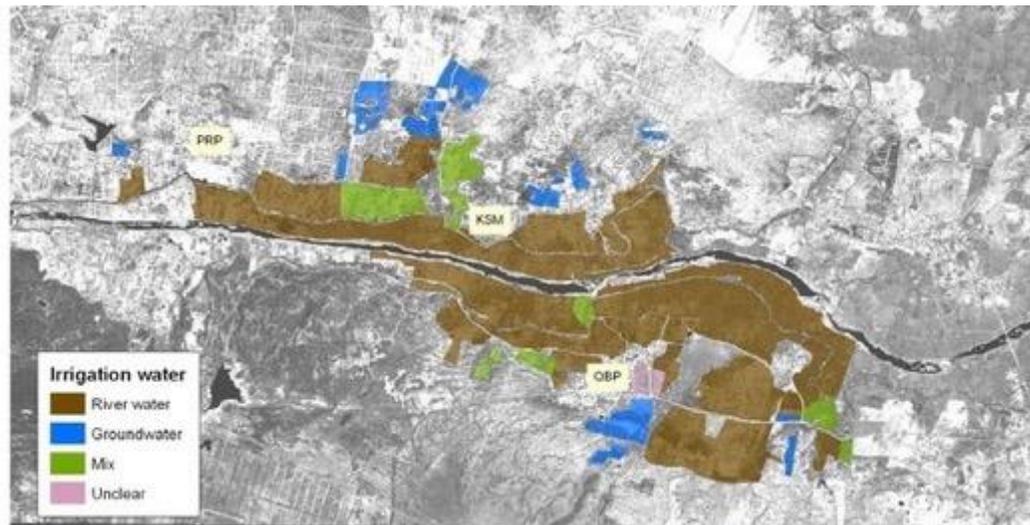


# Adding Global Water Abundance and Usage Contexts

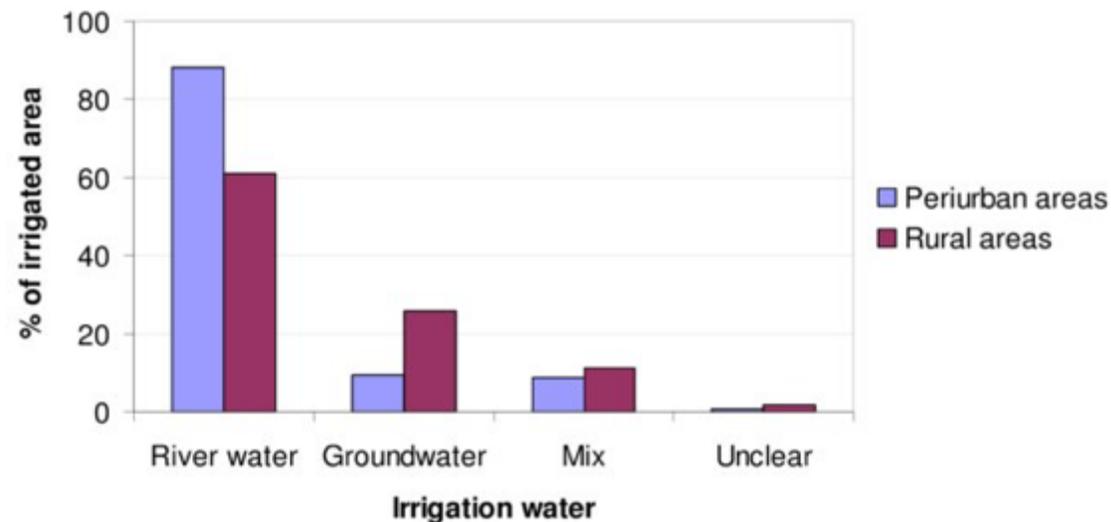
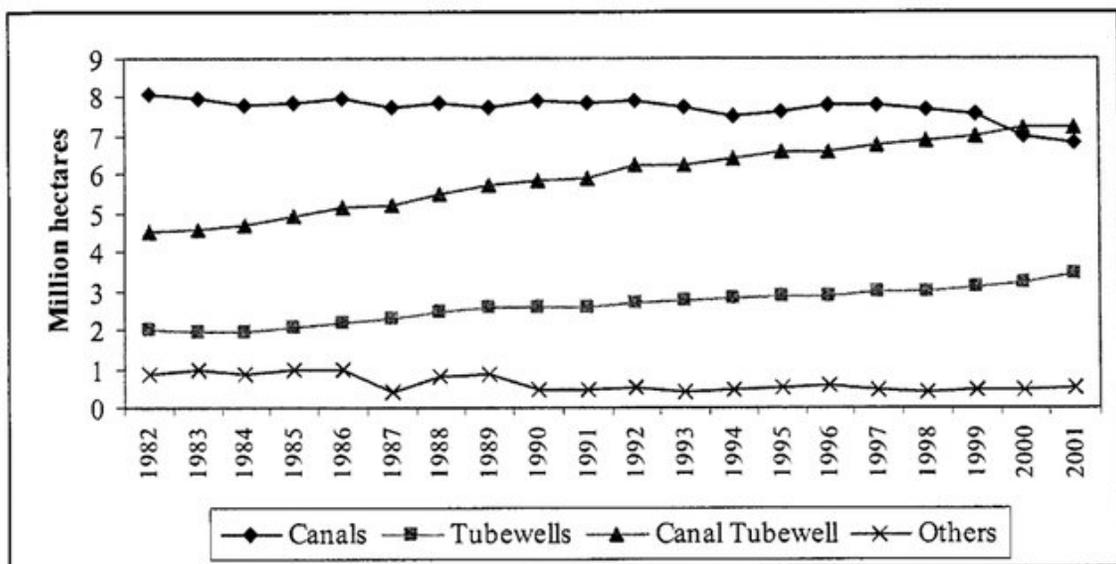
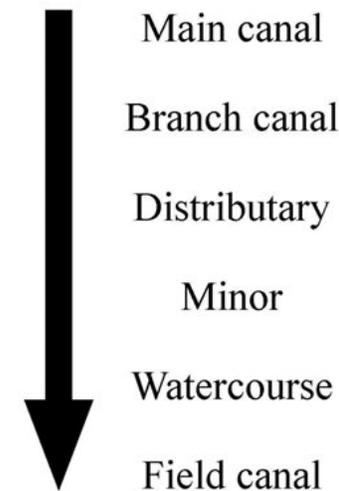


Weckenbrock, Philipp. (2010). Making a virtue of necessity: wastewater irrigation in a periurban area near Faisalabad, Pakistan : a GIS based analysis of long-term effects on agriculture.

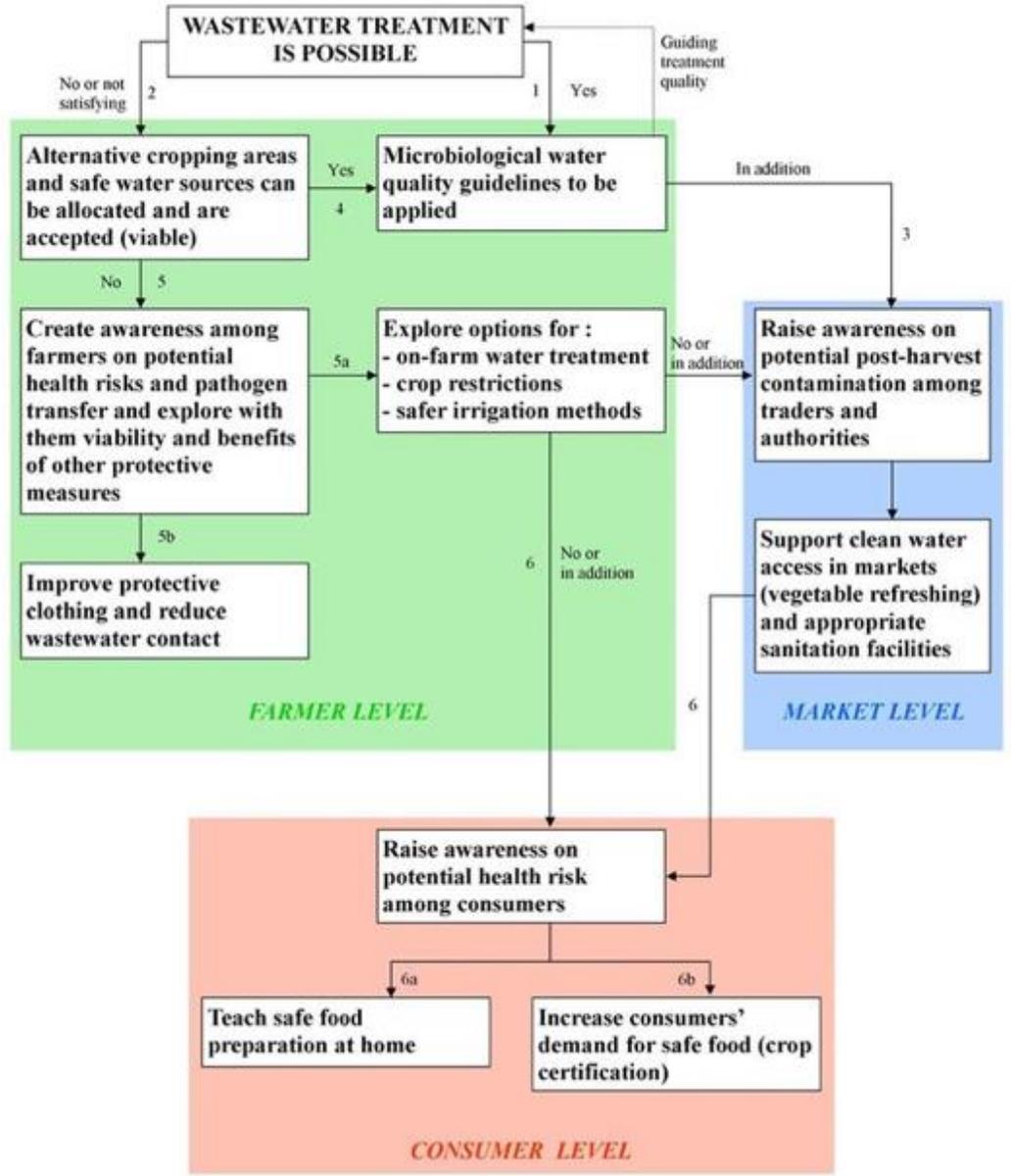
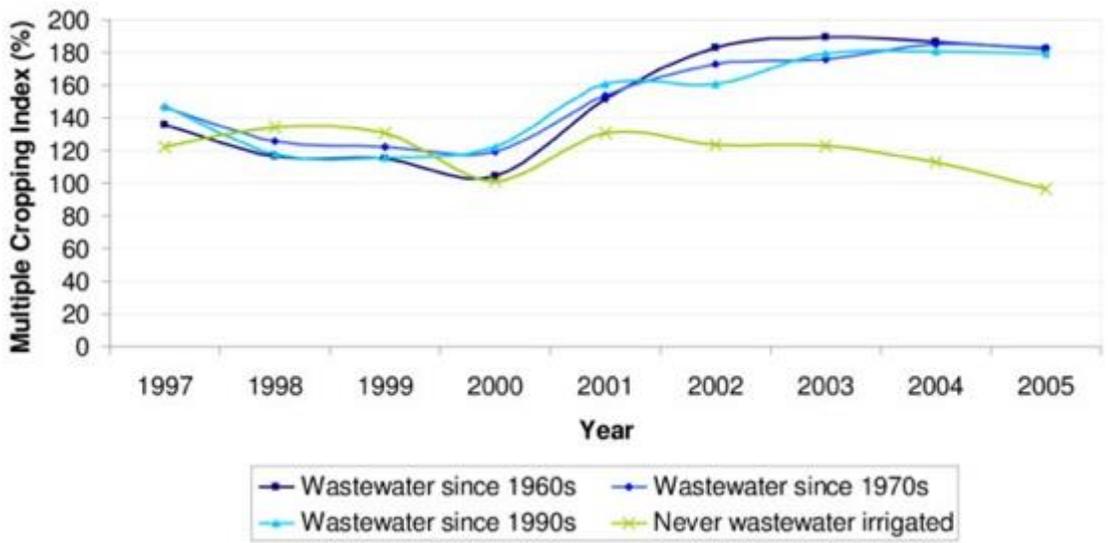
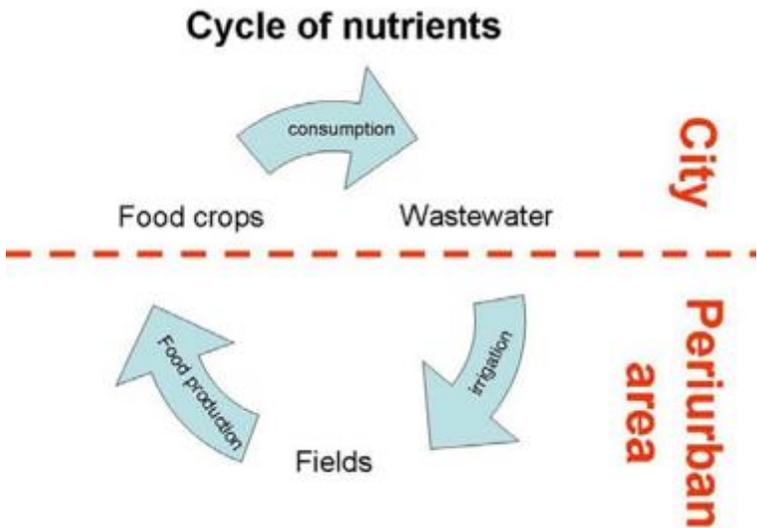
# Adding Irrigation Context



## Hierarchy of canals



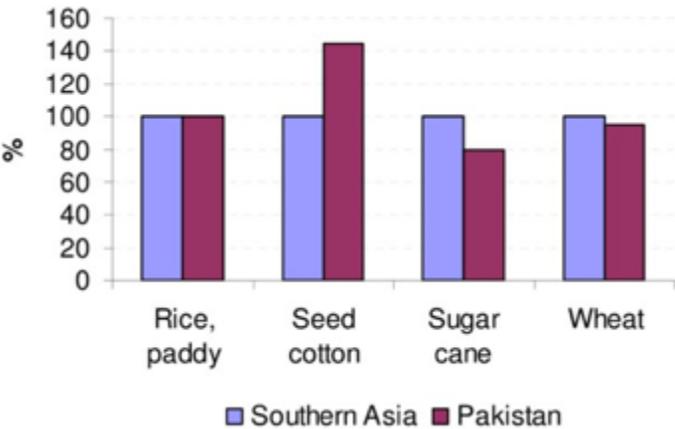
# Adding Waste Water Treatment Context



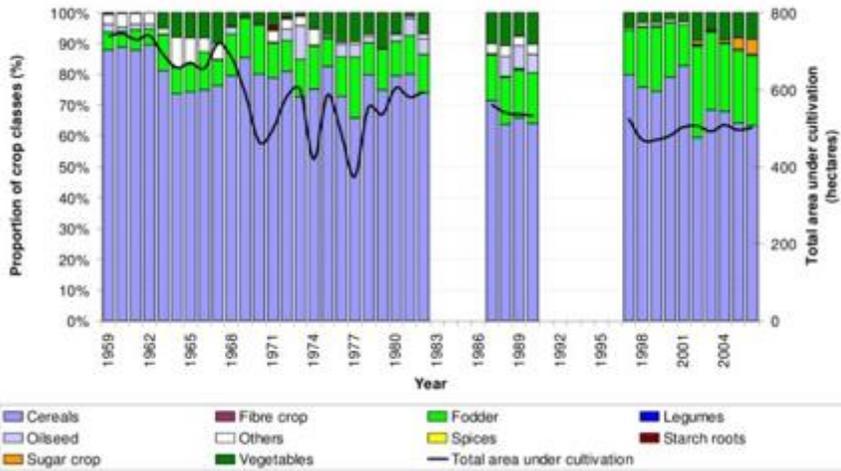
Weckenbrock, Philipp. (2010). Making a virtue of necessity: wastewater irrigation in a periurban area near Faisalabad, Pakistan : a GIS based analysis of long-term effects on agriculture.

# Adding Agricultural Land Use Context

## Overview on Land Use in the Study Area

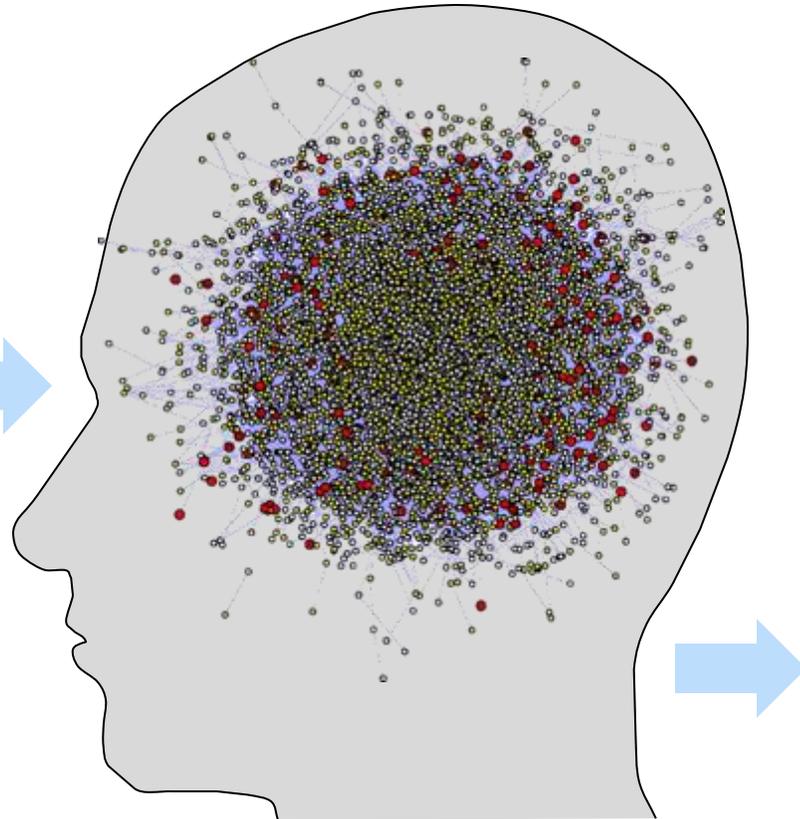
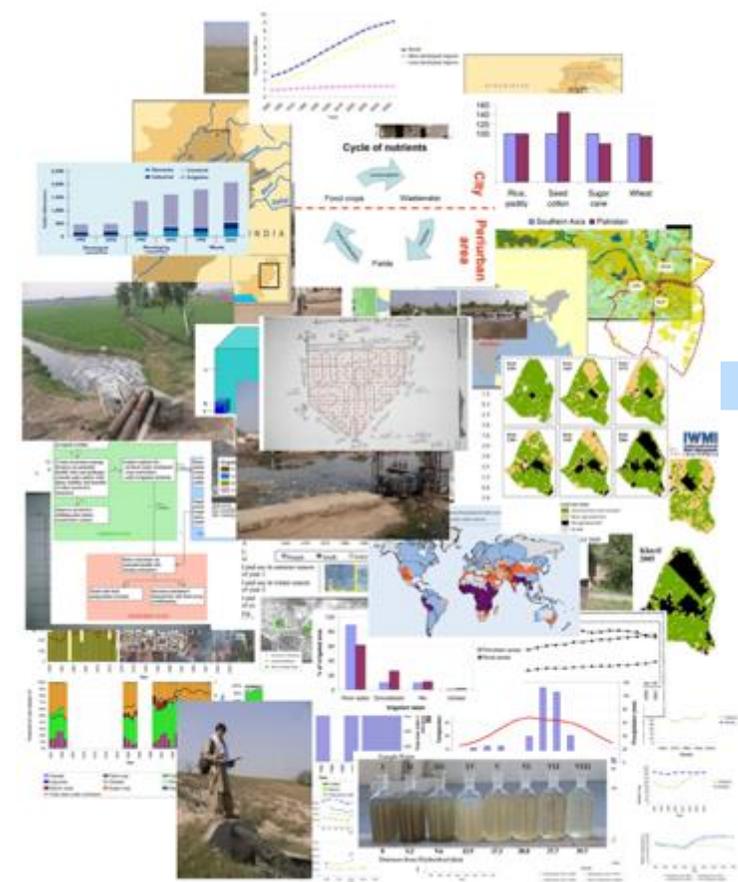


Village names:  
 PRP = Parvatapuram  
 KSM = Kachivani Singaram  
 QBP = Qutbullapur  
 MAM = Makta Anantharam  
 CRV = Chinnaraviralla  
 PLP = Pillaipalli





In the mind of the researcher... And eventually some kind of prose....



**Making a virtue of necessity –  
wastewater irrigation in a periurban area  
near Faisalabad, Pakistan**

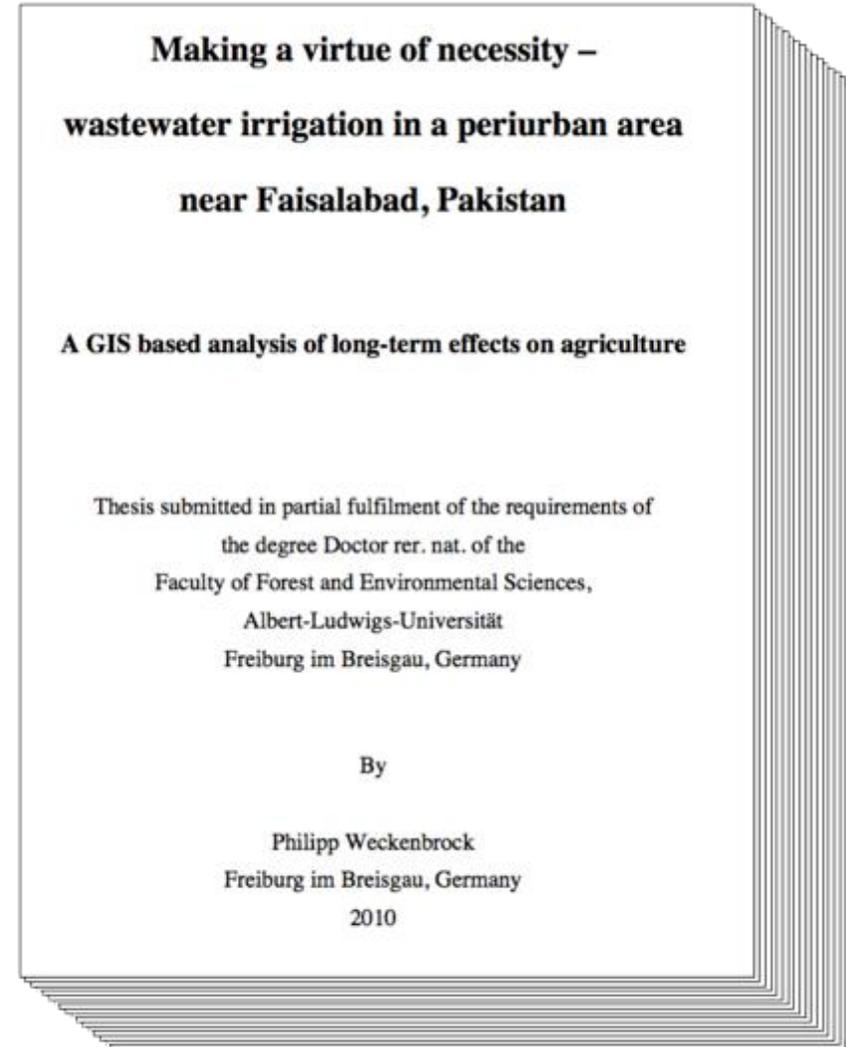
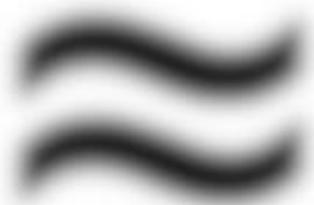
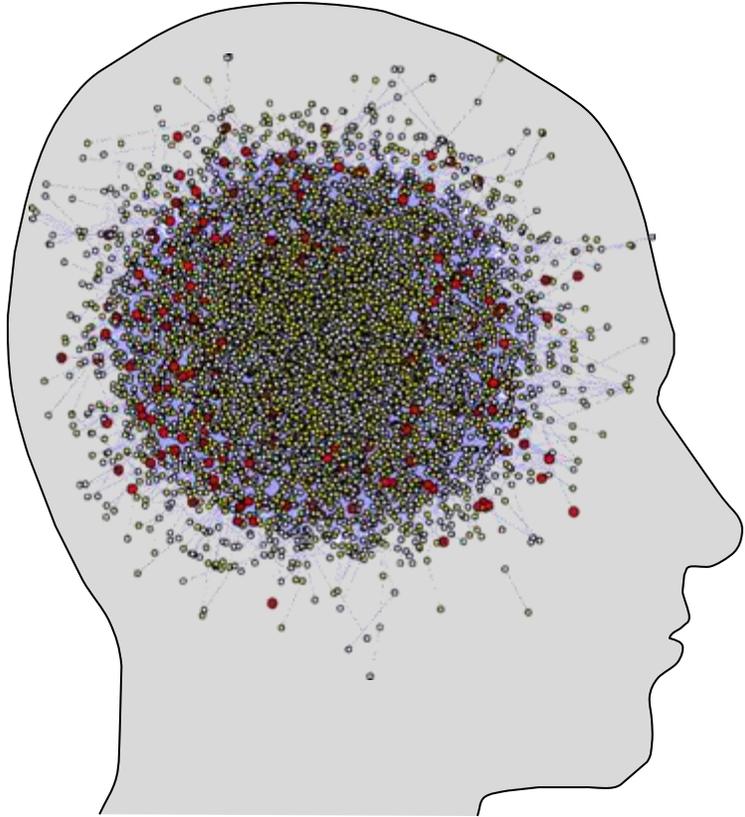
**A GIS based analysis of long-term effects on agriculture**

Thesis submitted in partial fulfilment of the requirements of  
the degree Doctor rer. nat. of the  
Faculty of Forest and Environmental Sciences,  
Albert-Ludwigs-Universität  
Freiburg im Breisgau, Germany

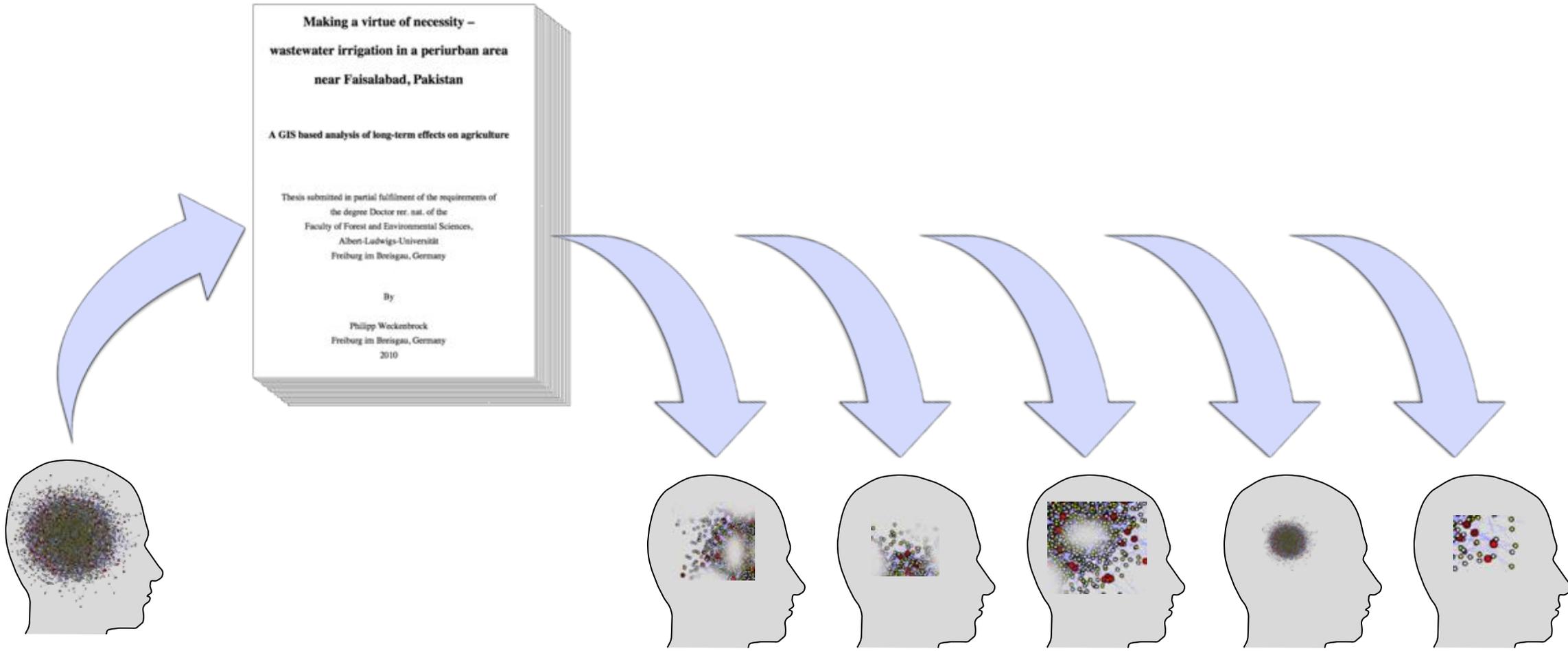
By

Philipp Weckenbrock  
Freiburg im Breisgau, Germany  
2010

Which is only a very rough approximation of the scientist's total understanding

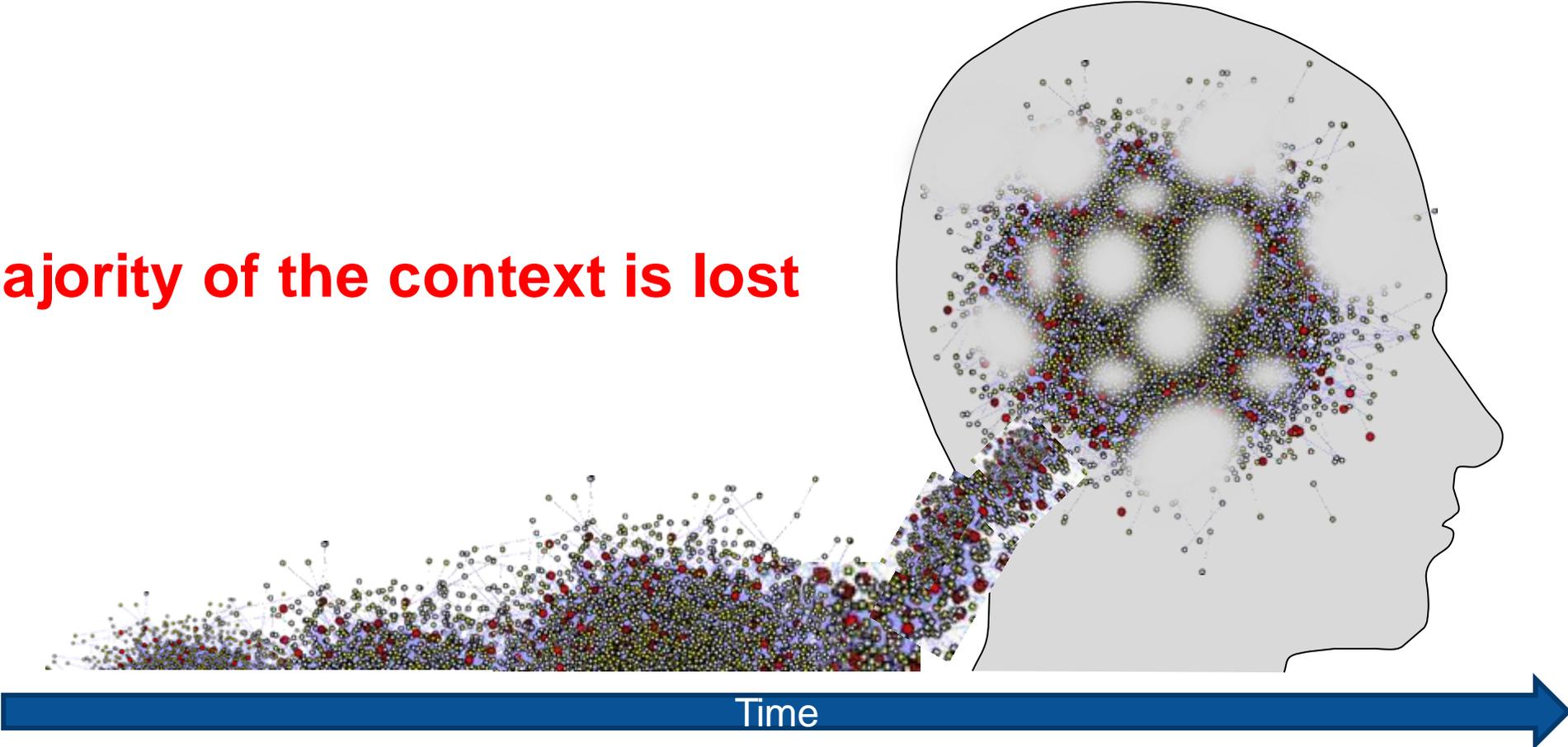


# How much of that context is ever truly communicated to the readers?



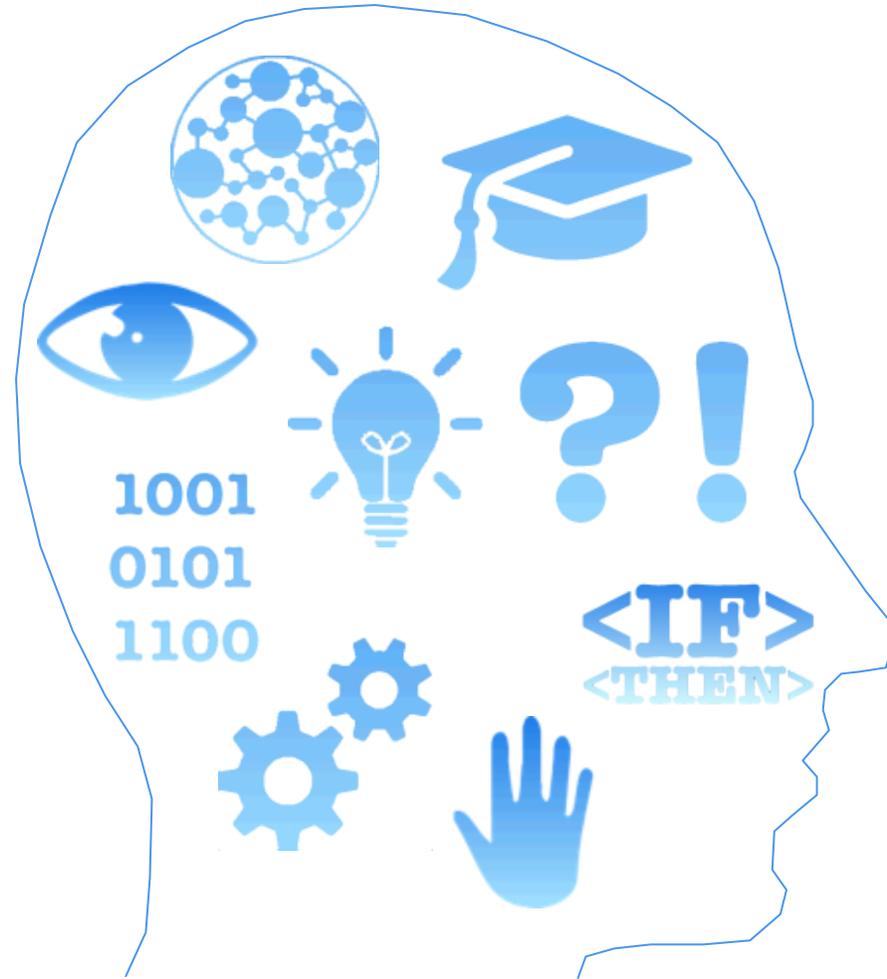
Even the original author rapidly loses details soon after the project

**The majority of the context is lost**



**So what does this all have to do with AI?**





1001  
0101  
1100

<IF>  
<THEN>

# The Megatrend of Computing, 1950's to 2020s

Platform

1950s



Electronics

1960s



Operating Systems

1970s



Database Systems

1980s



User Interface Systems

1990s



Business Process Systems

2000s



Internet WWW Mobile

2010s



Machine/Deep Learning

2015



Large Scale Knowledge Graphs

2020s



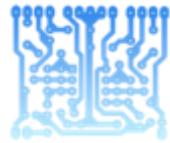
Machine Reasoning



Generalize, Standardize, Scale-up



Application



Pre-Software



Code Files  
Job Control

4GL Client-Server



Graphical User Interface

Application Server



Business Rules  
BPM XML HTML

Cognitive 1.0



Big Data/Analytics  
WATSON, UIMA

Cognitive 2.0



Curation Fusion

Cognitive 3.0



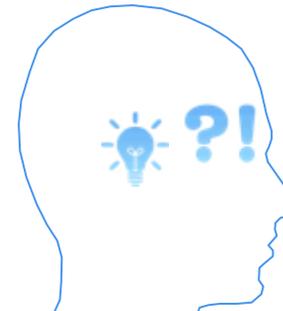
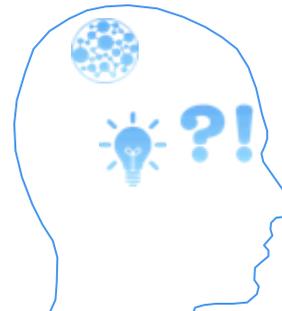
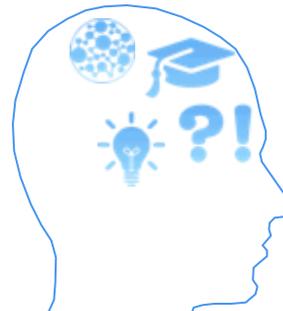
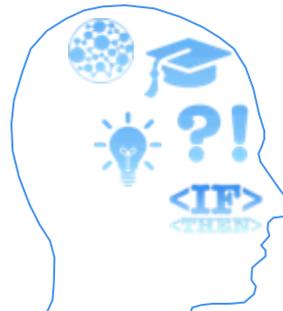
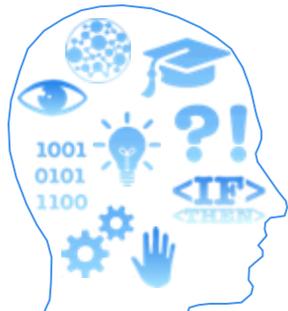
Semantics Ontology Inference



Programming effort required for new solutions significantly drops with each generation

User / Developer

Implement and Automate



Imagine and Innovate

# What's a knowledge graph?

A knowledge graph models information in the form of entities and the relationships between them

Knowledge graphs often contain elements for many different domains and modeling metalevels

- Data and Metadata
- Paradata
- Classes/Genotypes as well as Instances/Phenotypes
- Multi-temporal Versions
- Provenance and Permissions

Graphs (vertices and edges) can be representational as well as computational via traversals  
e.g., Bayesian networks, Spreading Activation Networks, etc.

Many of the world's largest organizations really heavily on knowledge graphs

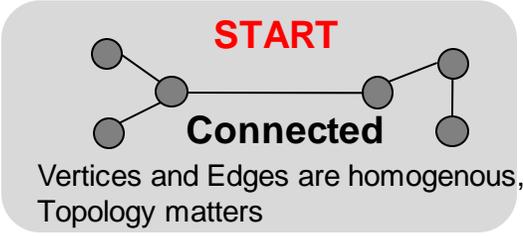
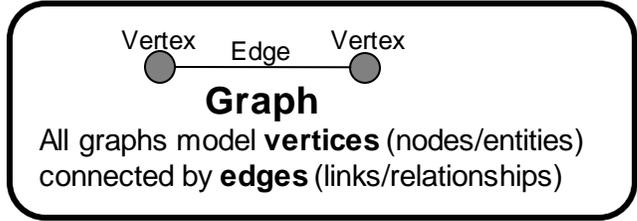


Google

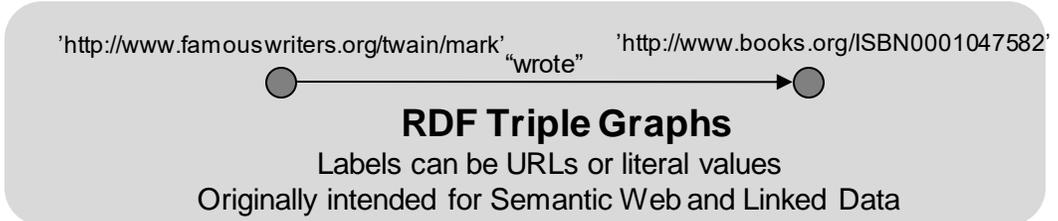
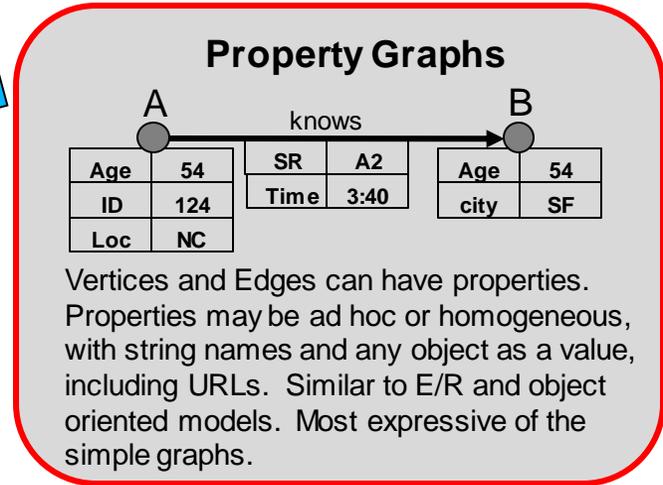
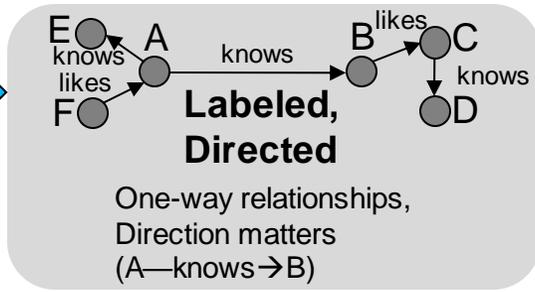
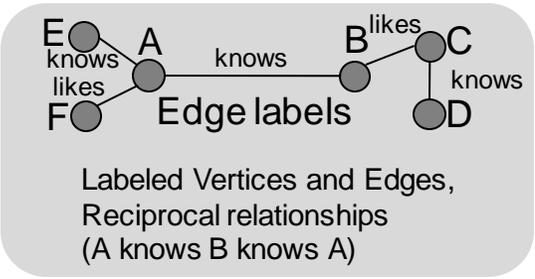
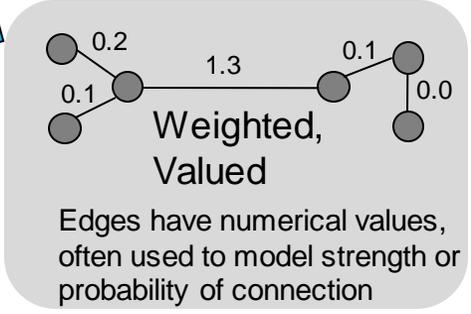
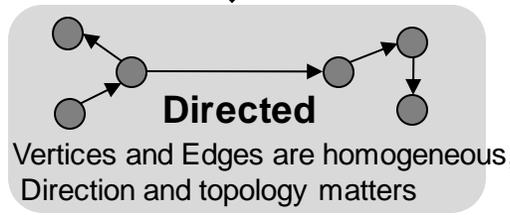
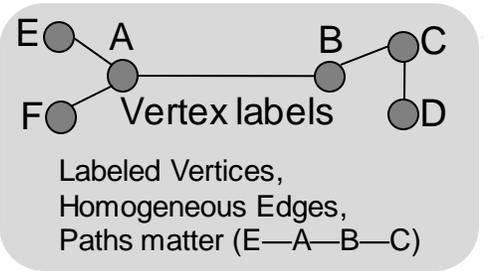
the user's age or age range,  
educational level or range,  
income level or range,  
language preferences,  
marital status,  
geographic location  
cultural background or preferences,  
defined groups  
psychographic information



# What do I mean by a graph?



Each more advanced graph subsumes the modeling capabilities of its predecessors



Semantic Web, (RDF, RDFS, SPARQL, OWL)

Apache Tinkerpop, Gremlin

Graph Databases are mature, stable technology and are readily available in your favorite cloud ecosystem, via opensource or numerous vendors



Azure Cosmos DB



**COMPOSE**  
AN IBM COMPANY

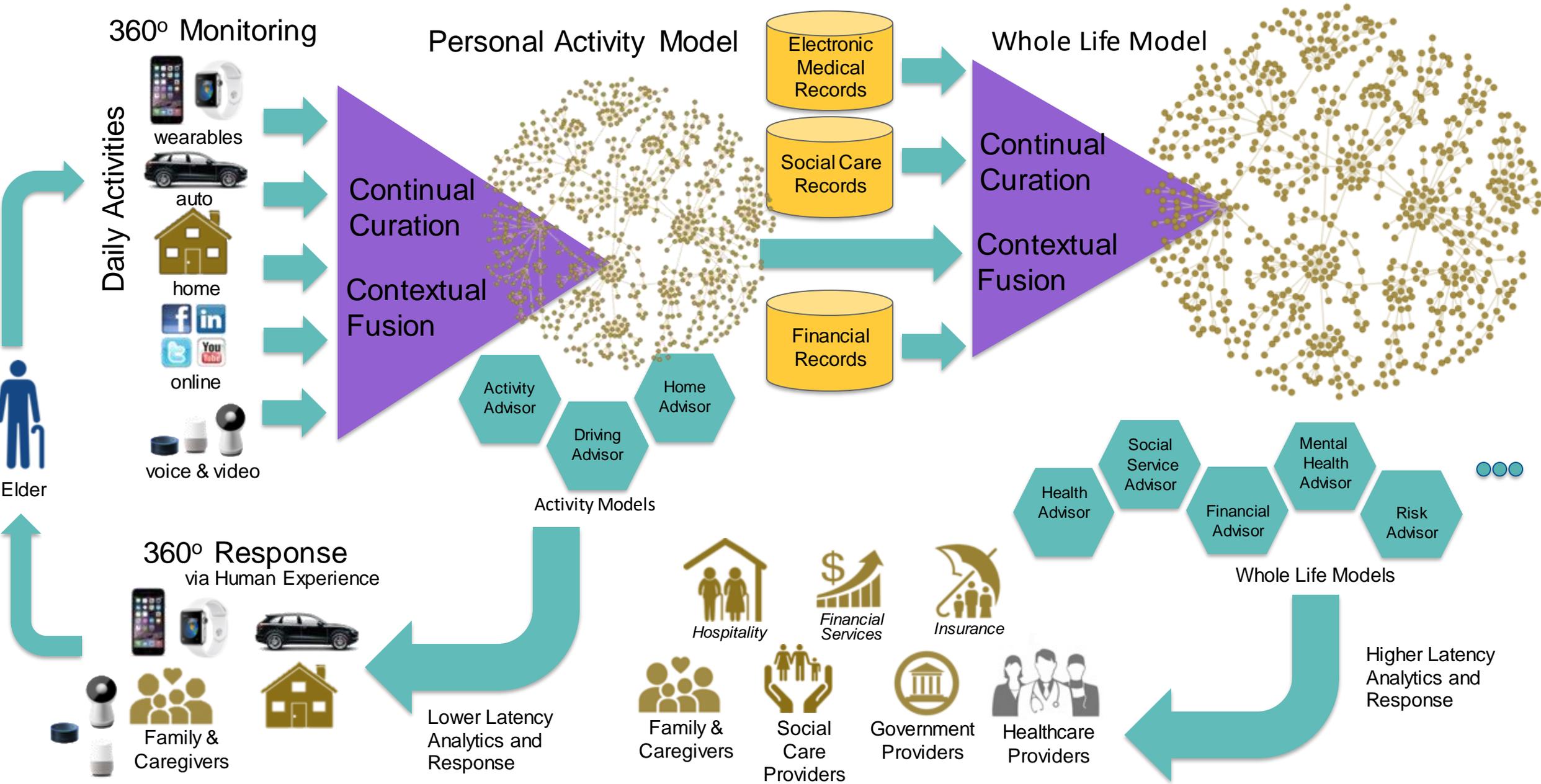


**DATASTAX**



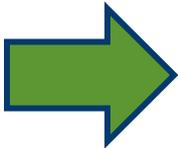
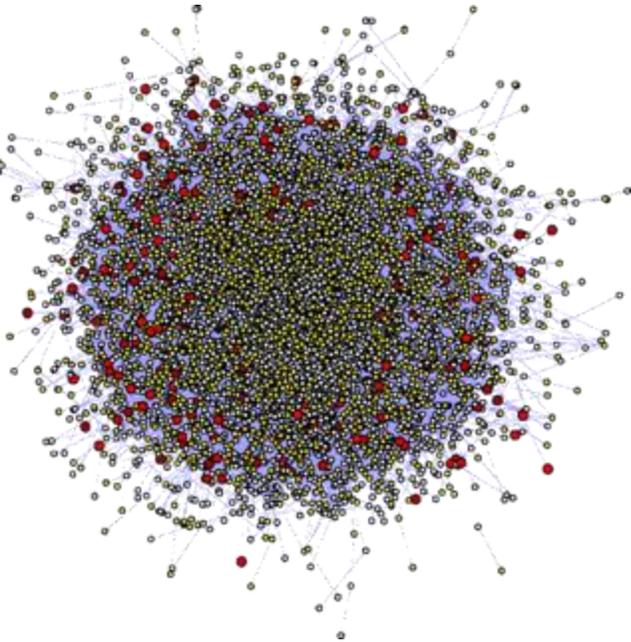
New graph database vendors arriving monthly

# Challenge 1: Continual Data Curation, Contextual Data Fusion

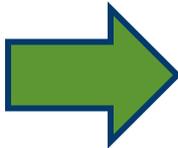
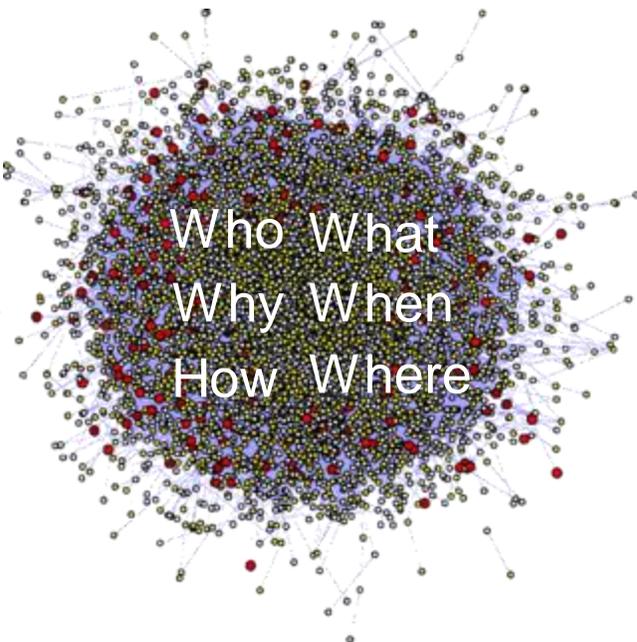


# Challenge 2: How do you get value from a hairball of integrated data?

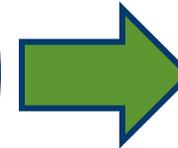
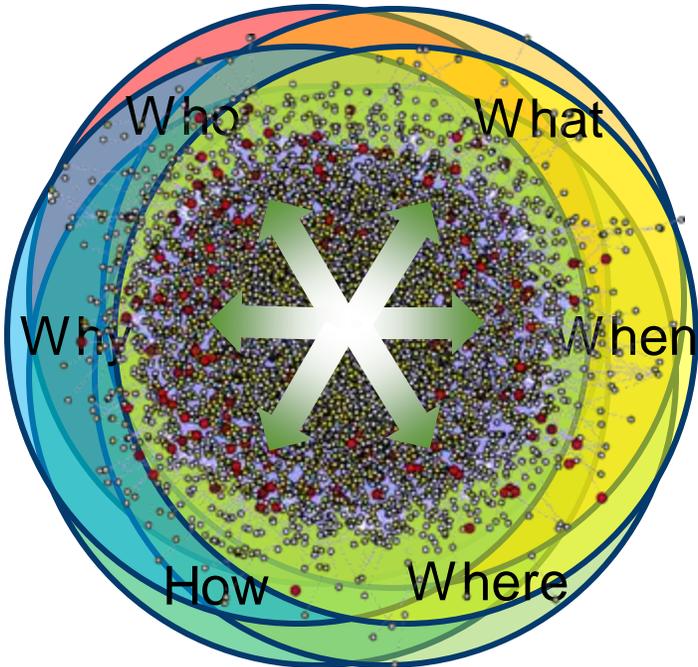
Integrated graph of all relevant data



Identify Basic Conceptual Categories

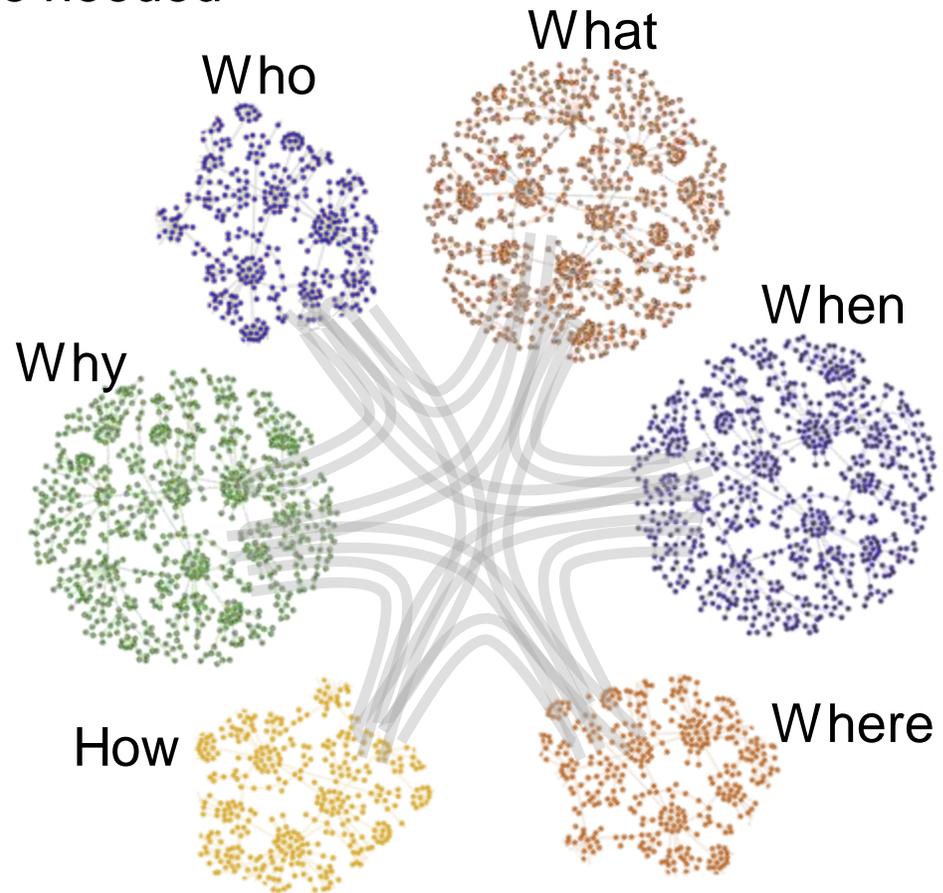
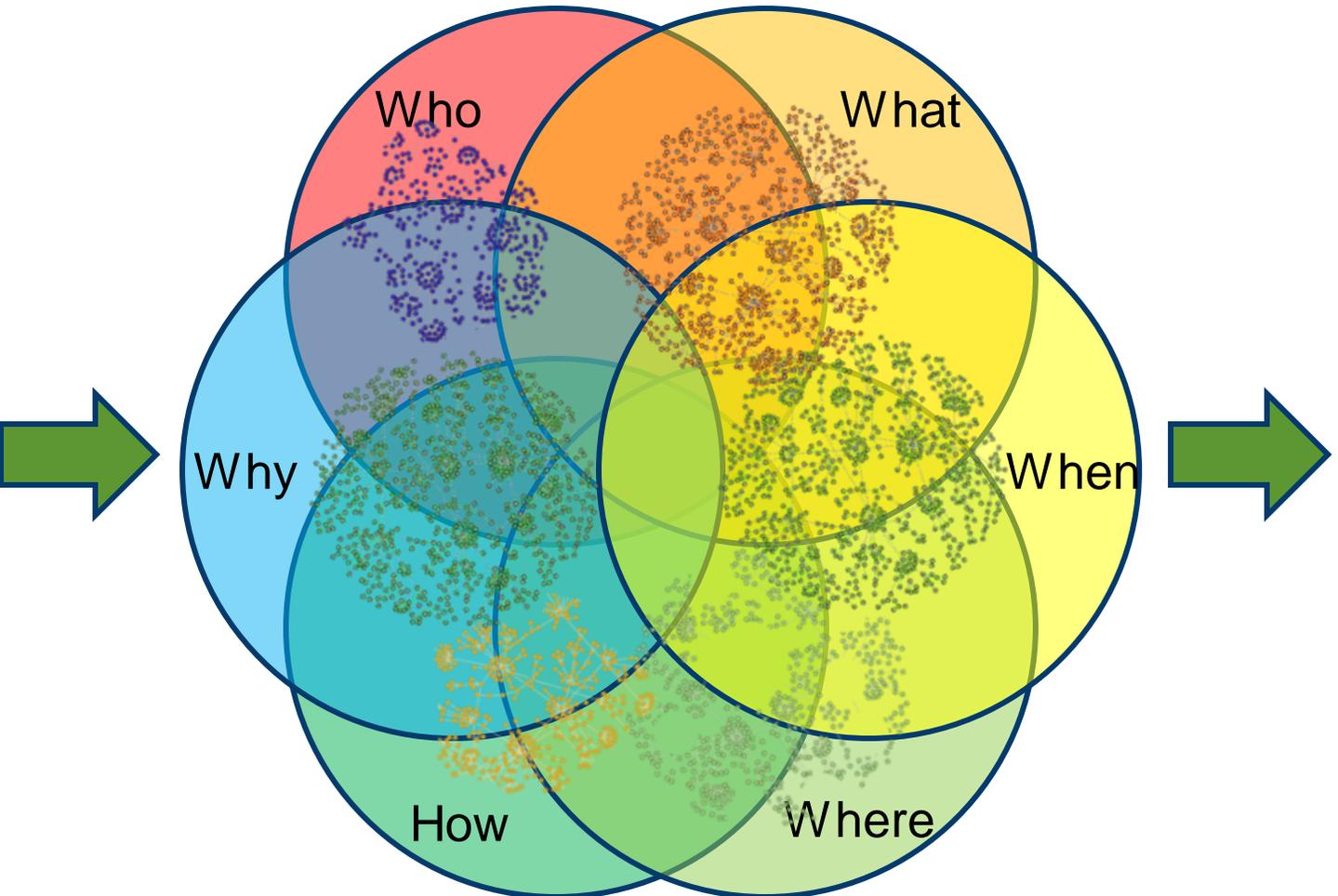


Separate the graph into Conceptual Regions

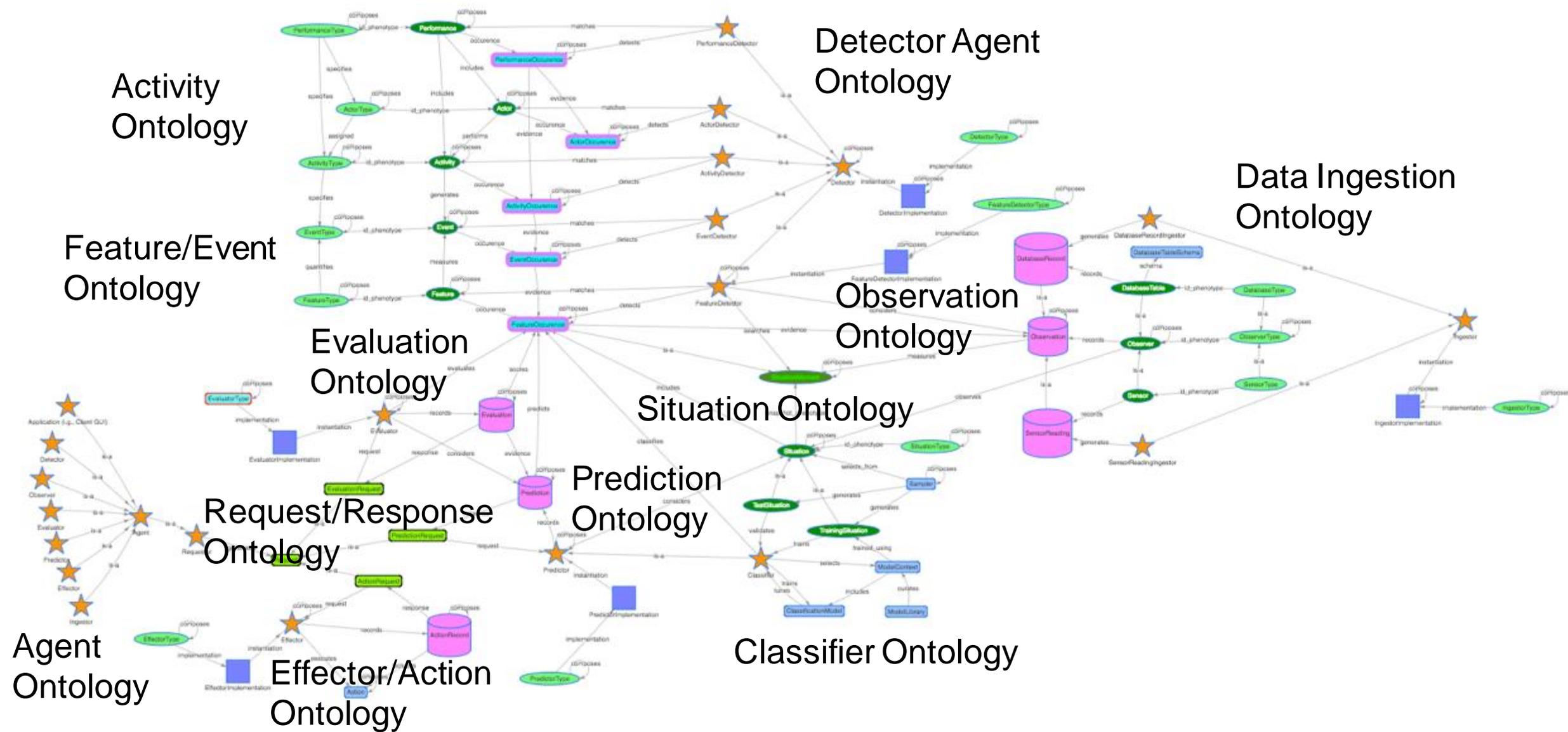


# From Hairball to Knowledge Graph

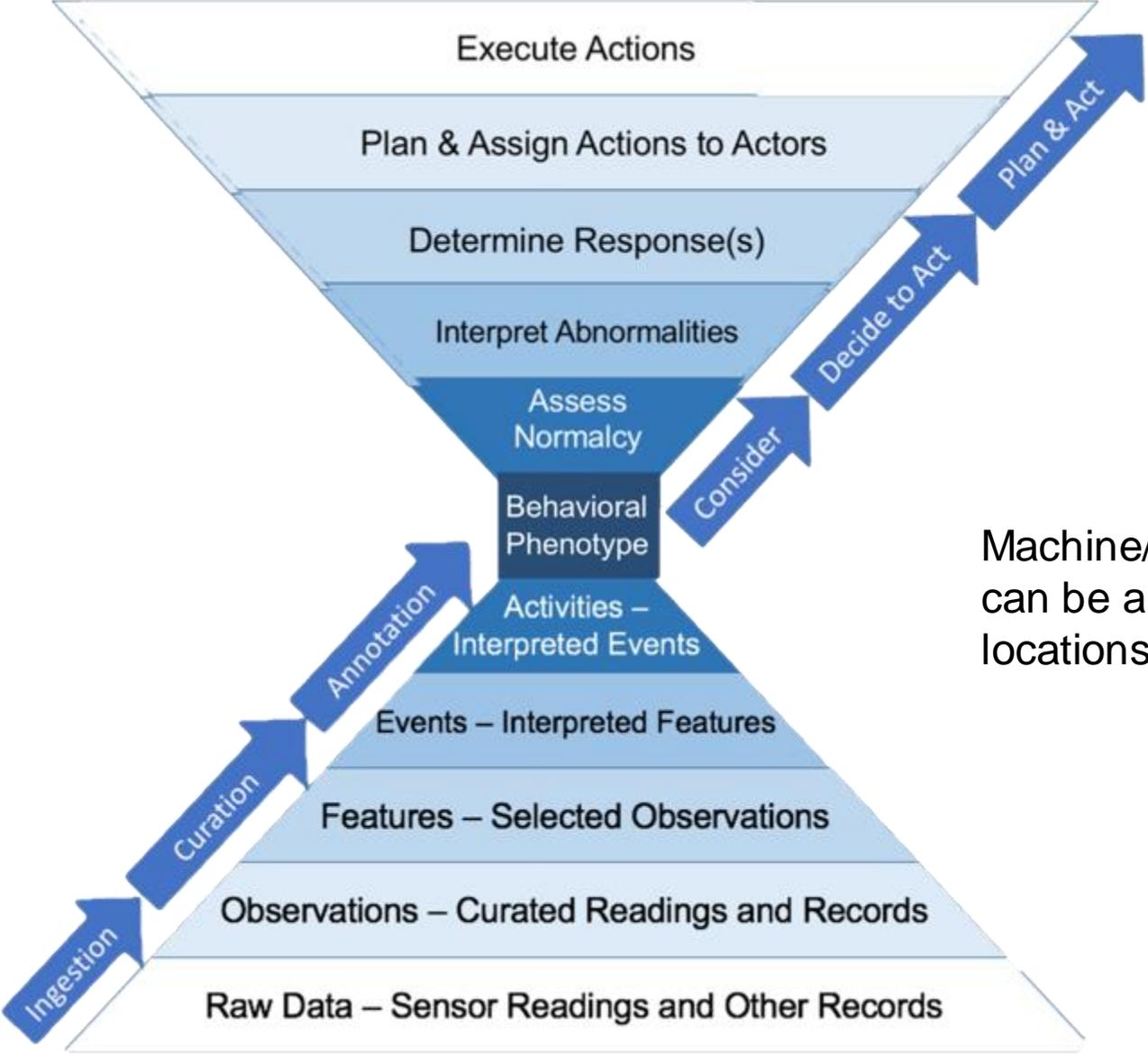
Apply/Develop Ontologies for each of these Conceptual Regions, transforming data where needed



# Blended Ontologies are crucial

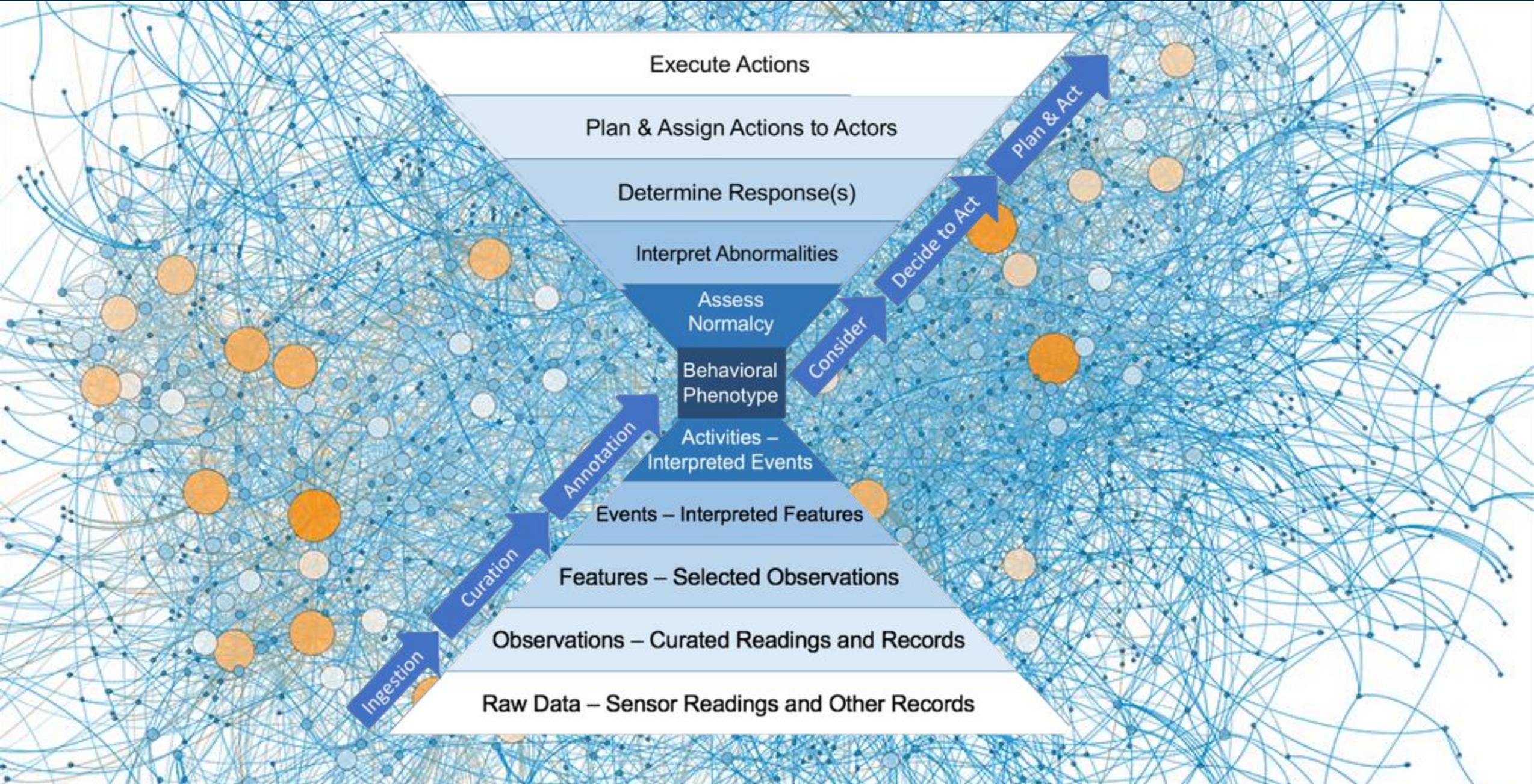


# Consider the entire process from measurement to action

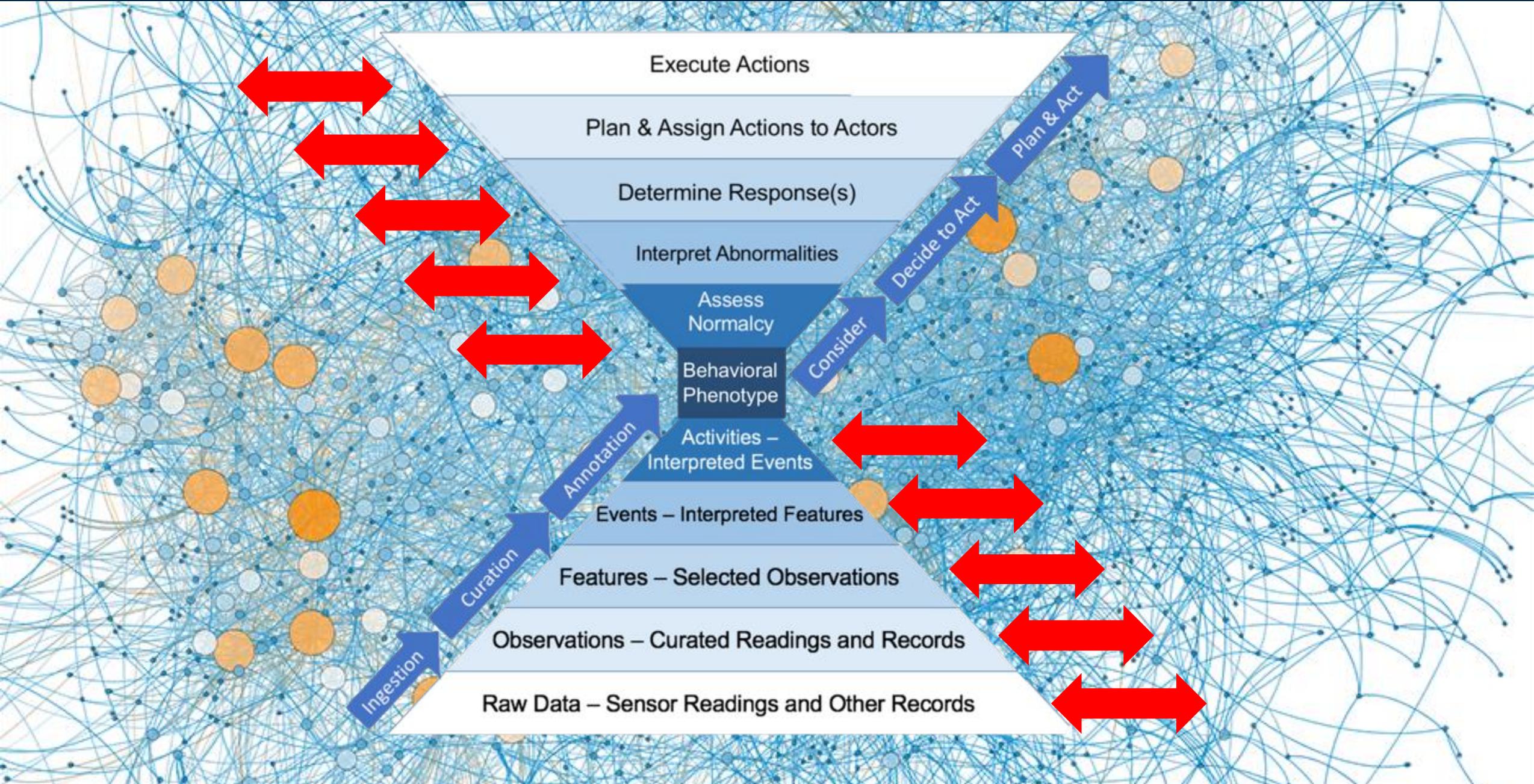


Machine/Deep Learning can be applied in many locations

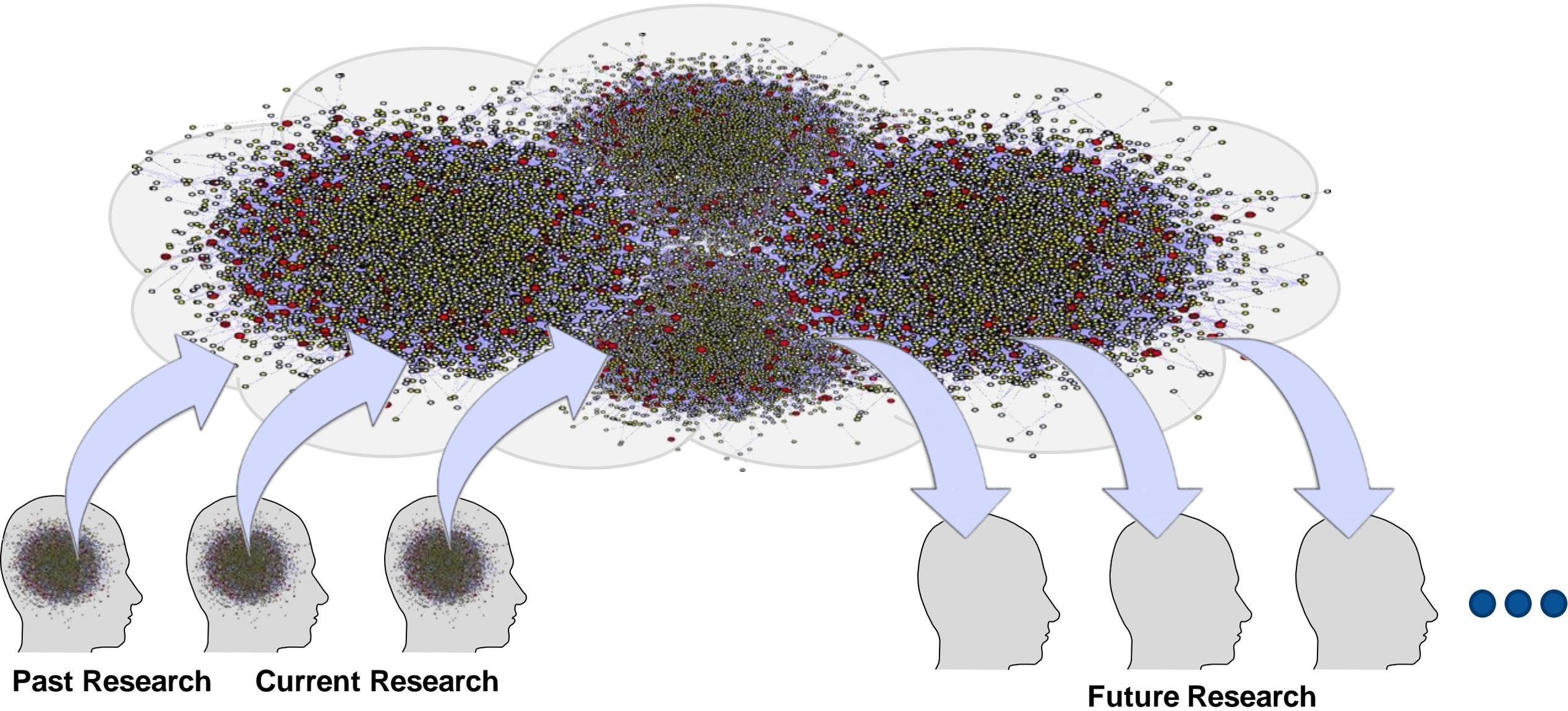
# All in the context of a comprehensive knowledge graph



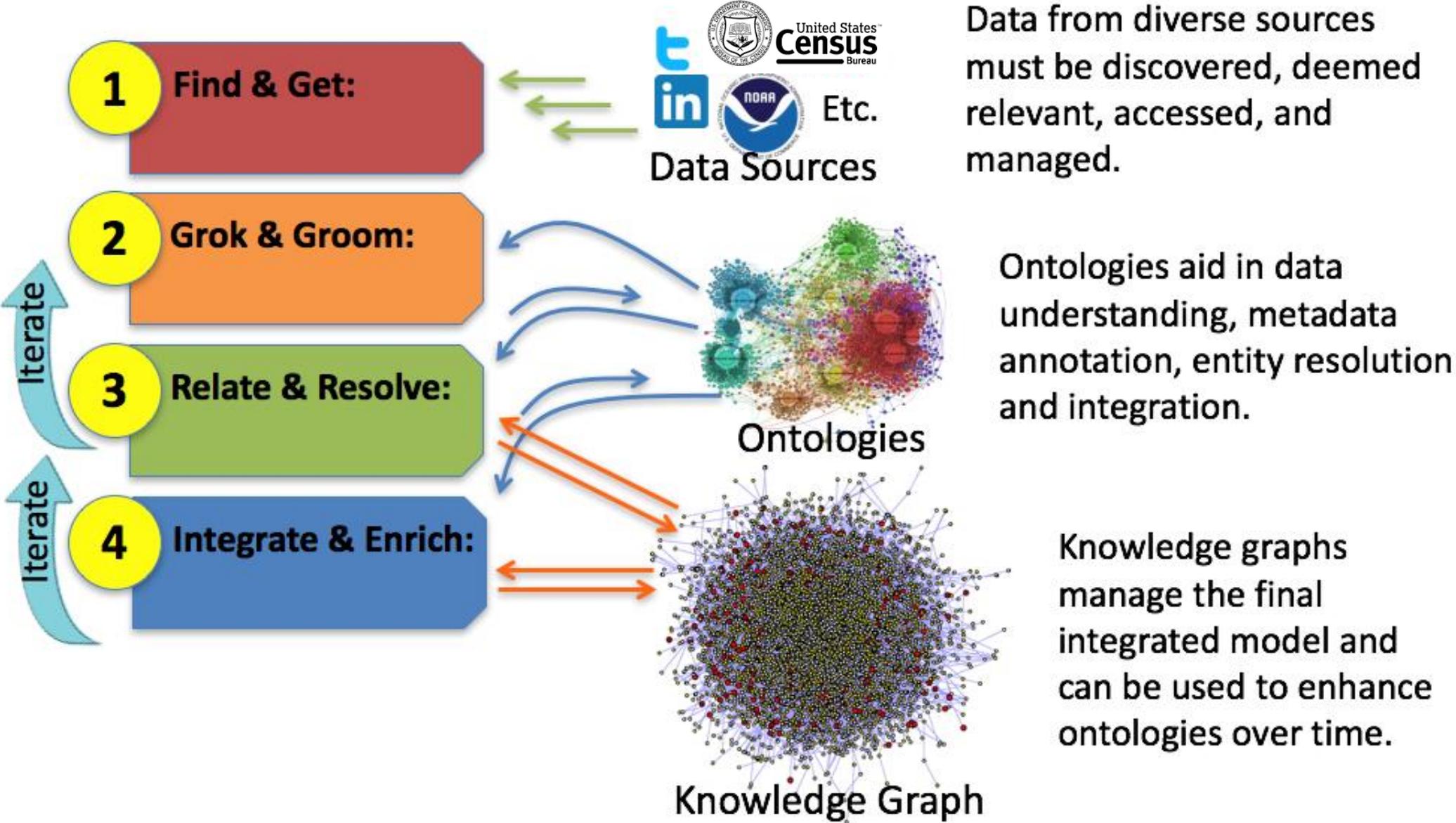
# With feedback... Knowledge scaffolds Deep Learning which extends Knowledge



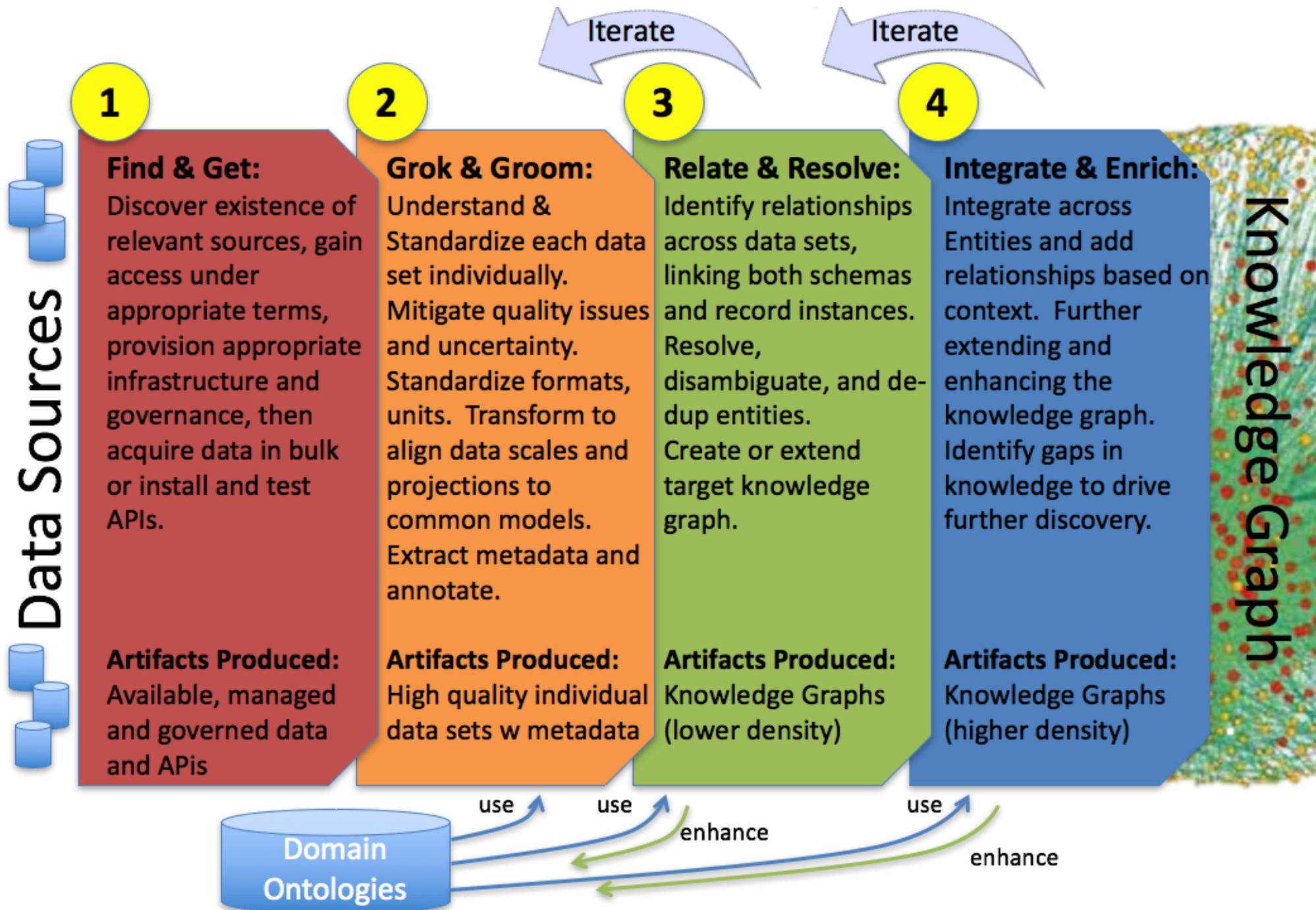
# What will it take to get to the Environment Health Knowledge Cloud?



# Start by understanding the process of data integration, curation, and contextual fusion



Even ignoring the knowledge graph target, nearly all of these activities occur in nearly every Data Science project of any complexity.

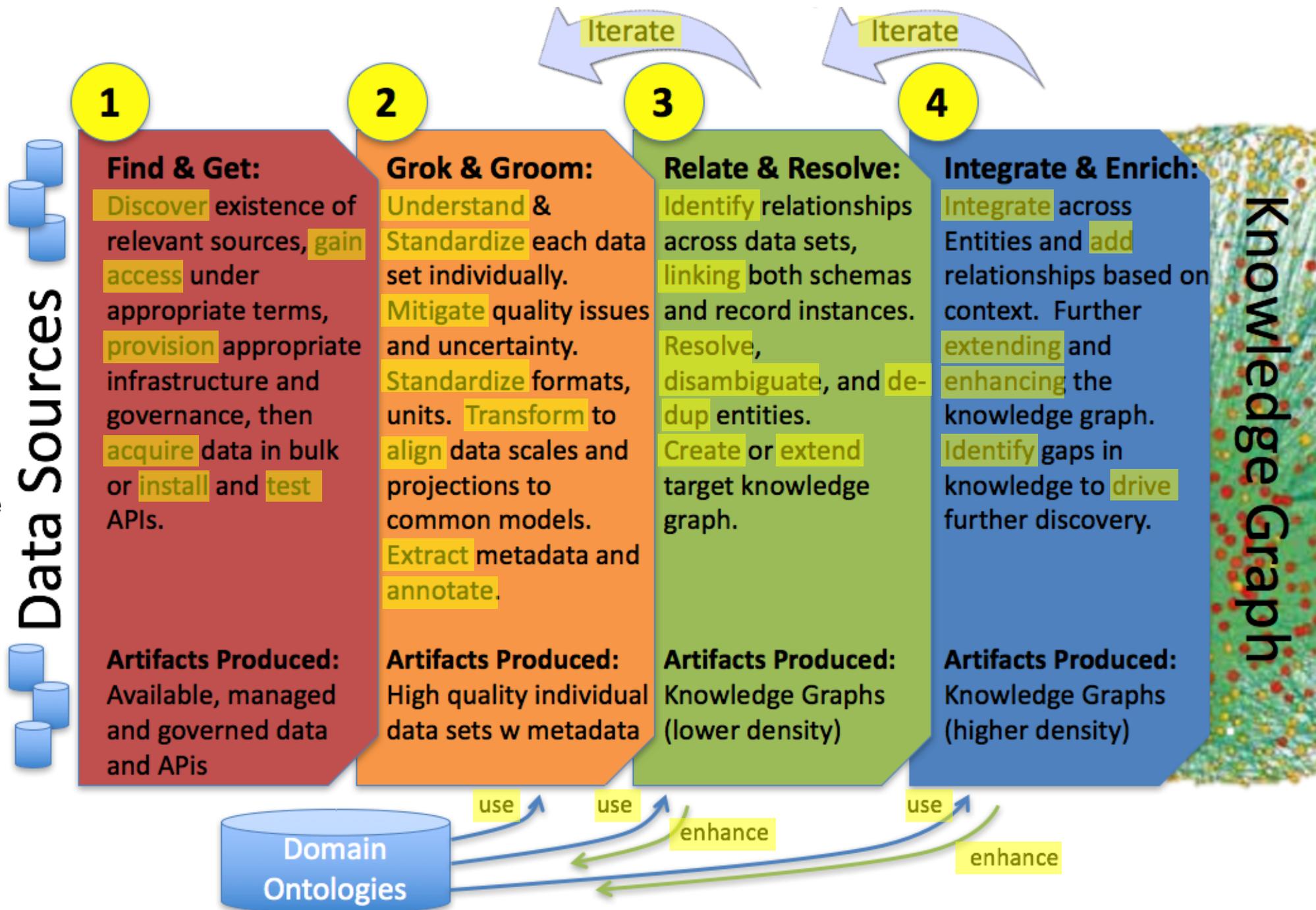


But most of the information created during the process is never captured.

Notice all the verbs

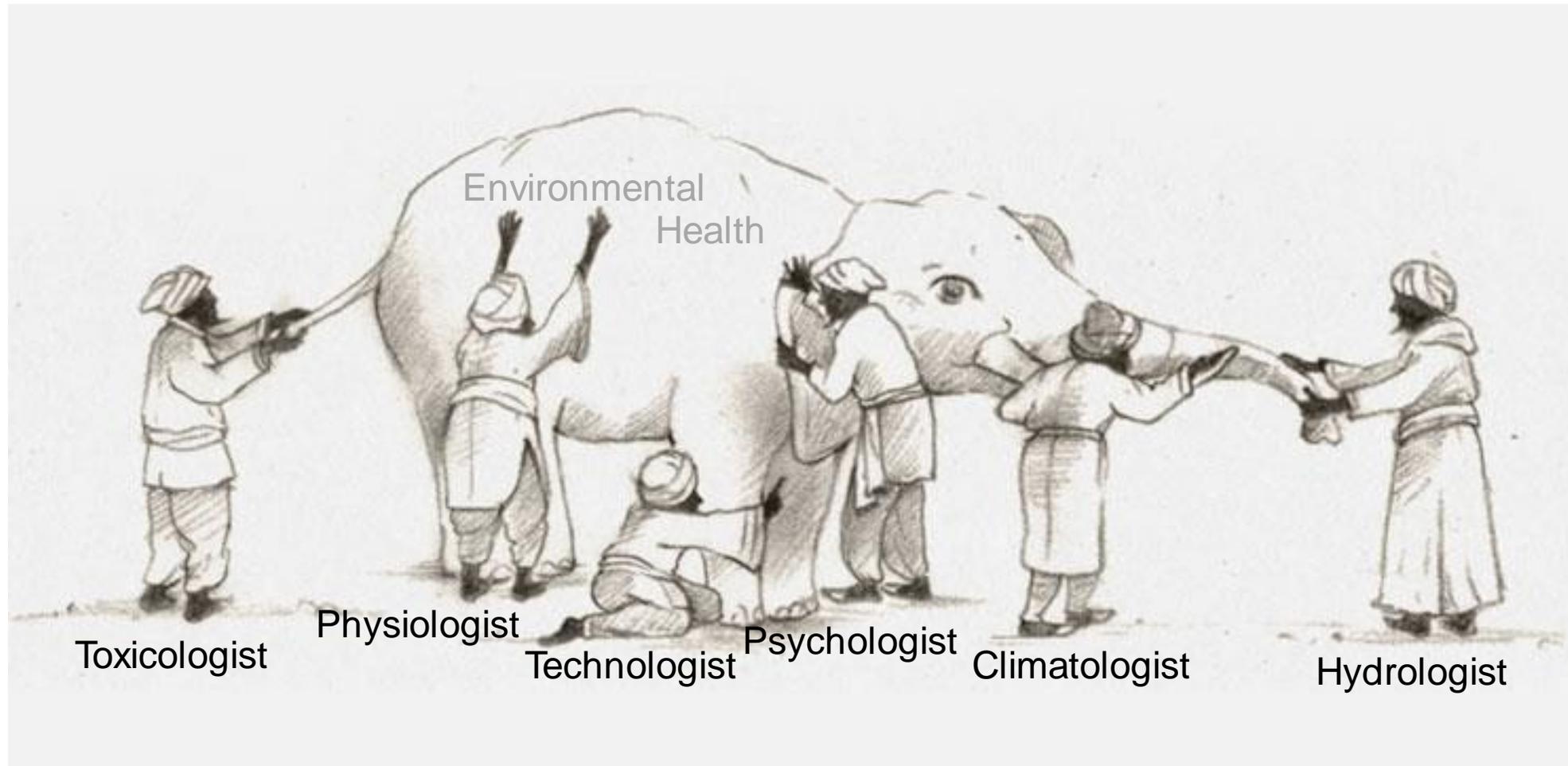
How much of the metadata, paradata, and process knowledge is captured and managed for sharing and reuse during or after the project?

That's the context.



# Challenges of knowledge/context sharing and access in large orgs/agencies

You must show how sharing increases value for all stakeholders.  
Adoption is always better driven by self-interest than by mandate.





delivering **the promise of science**  
for global good



Sam Adams  
sadams@rti.org  
919-485-2636

