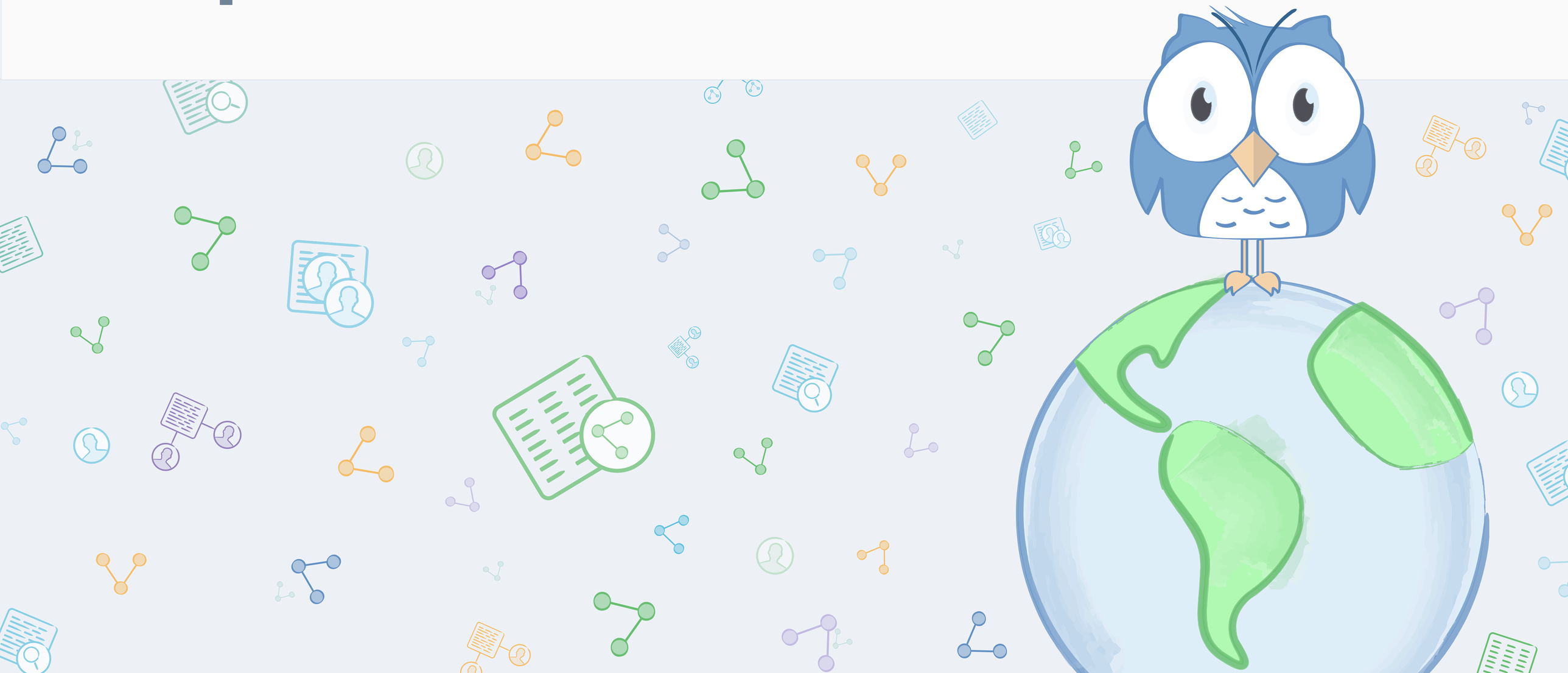
data.world

@OpenAustin



WE'RE ON A MISSION

We're building the most **meaningful**, **collaborative**, and **abundant** data resource in the world by dismantling the barriers between data and people.

data.world

A NEW KIND OF COMPANY

Benefit Corporation

- Expanded purpose includes public benefit
- Requires consideration of shareholders *and* stakeholders
- Flexibility to weigh public benefit in sale & IPO decisions

Notable Benefit Corporations

KICKSTARTER

BEN & JERRY'S

patagonia

LAUREATE EDUCATION INC

data.world

OUR PRODUCT

A data platform that helps people work together to solve problems faster by creating new ways to **discover**, **prep**, and **collaborate**.

data.world

OPEN DATA WANTS TO BE LINKED DATA

Because data is a social animal, too.

Jonathan Ortiz

September 19, 2016

There are a

HUGE NUMBER

of OPEN DATA SETS

TOO MUCH OF DATA'S GROWTH IS HAPPENING IN SILOS.

data.world

Only available as

DOWNLOADABLE
DOWNLOADABLE
DOWNLOADABLE

DOCUMENTS

data.world

OPEN DATA EXISTS IN MANY FORMATS

data.world

data.world

Few formats convey **MEANING** *about the contents in a way that can be* **SHARED** *and* **EXTENDED**.

Some datasets are available via

APIs

But those APIs don't generally have consistent **interfaces** *or* **patterns**…

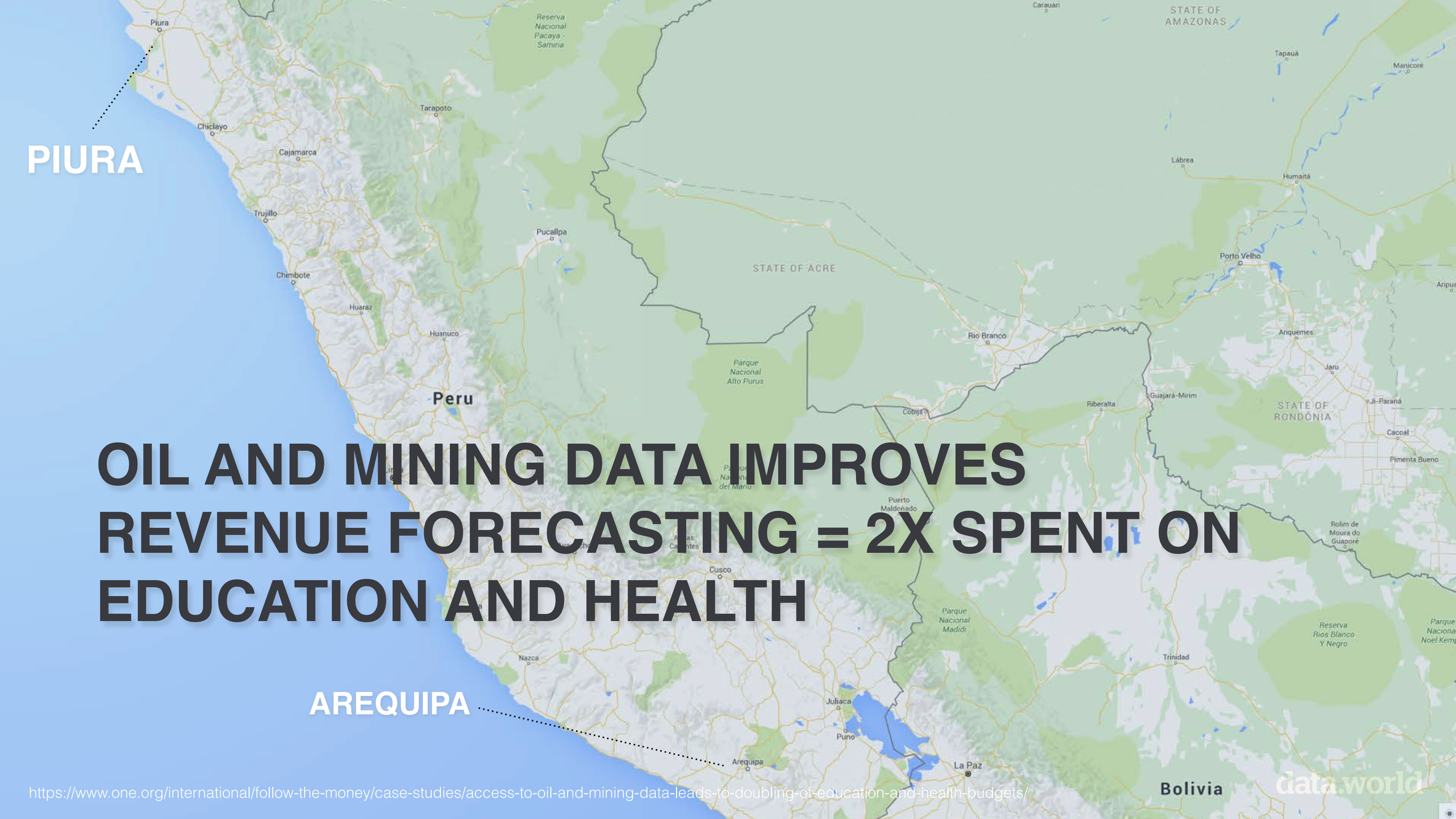

THEY LOAD IT IN
YOU PULL IT OUT
data.world

It is **GREAT** *that this open data* EXISTS

data.world

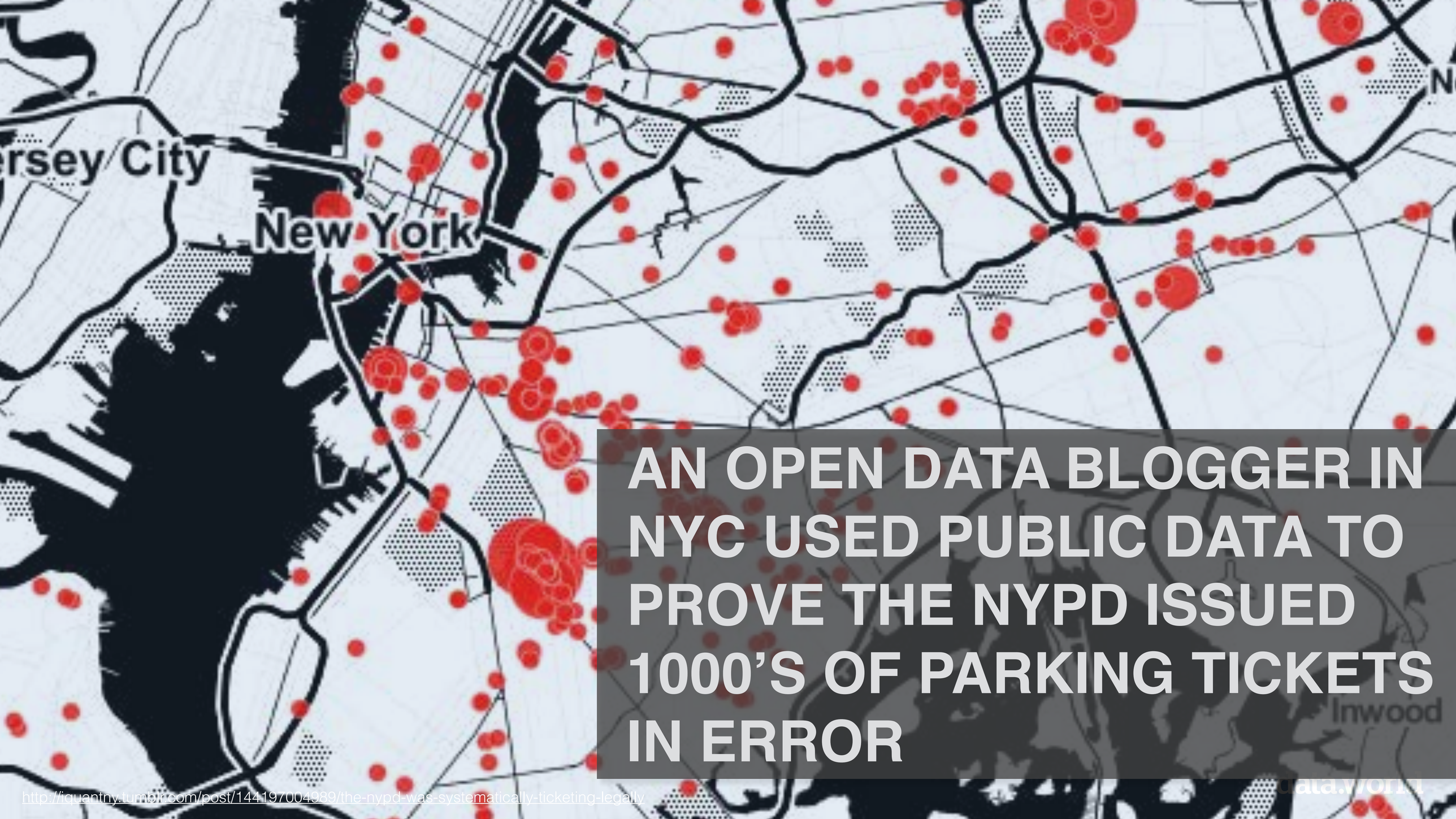
OPEN DATA FOR ALTERNATIVE RISK MODELS = $2B IN LOANS ACROSS 700+ INDUSTRIES

data.world



OIL AND MINING DATA IMPROVES REVENUE FORECASTING = 2X SPENT ON EDUCATION AND HEALTH

https://www.one.org/international/follow-the-money/case-studies/access-to-oil-and-mining-data-leads-to-doubling-of-education-and-health-budgets/



AN OPEN DATA BLOGGER IN NYC USED PUBLIC DATA TO PROVE THE NYPD ISSUED 1000'S OF PARKING TICKETS IN ERROR

http://iquantny.tumblr.com/post/144197004989/the-nypd-was-systematically-ticketing-legally

Mr. Wellington's analysis identified errors the department made in issuing parking summonses. It appears to be a misunderstanding by officers on patrol of a recent, abstruse change in the parking rules. We appreciate Mr. Wellington bringing this anomaly to our attention.

The department's internal analysis found that patrol officers who are unfamiliar with the change have observed vehicles parked in front of pedestrian ramps and issued a summons in error. When the rule changed in 2009 to allow for certain pedestrian ramps to be blocked by parked vehicles, the department focused training on traffic agents, who write the majority of summonses.

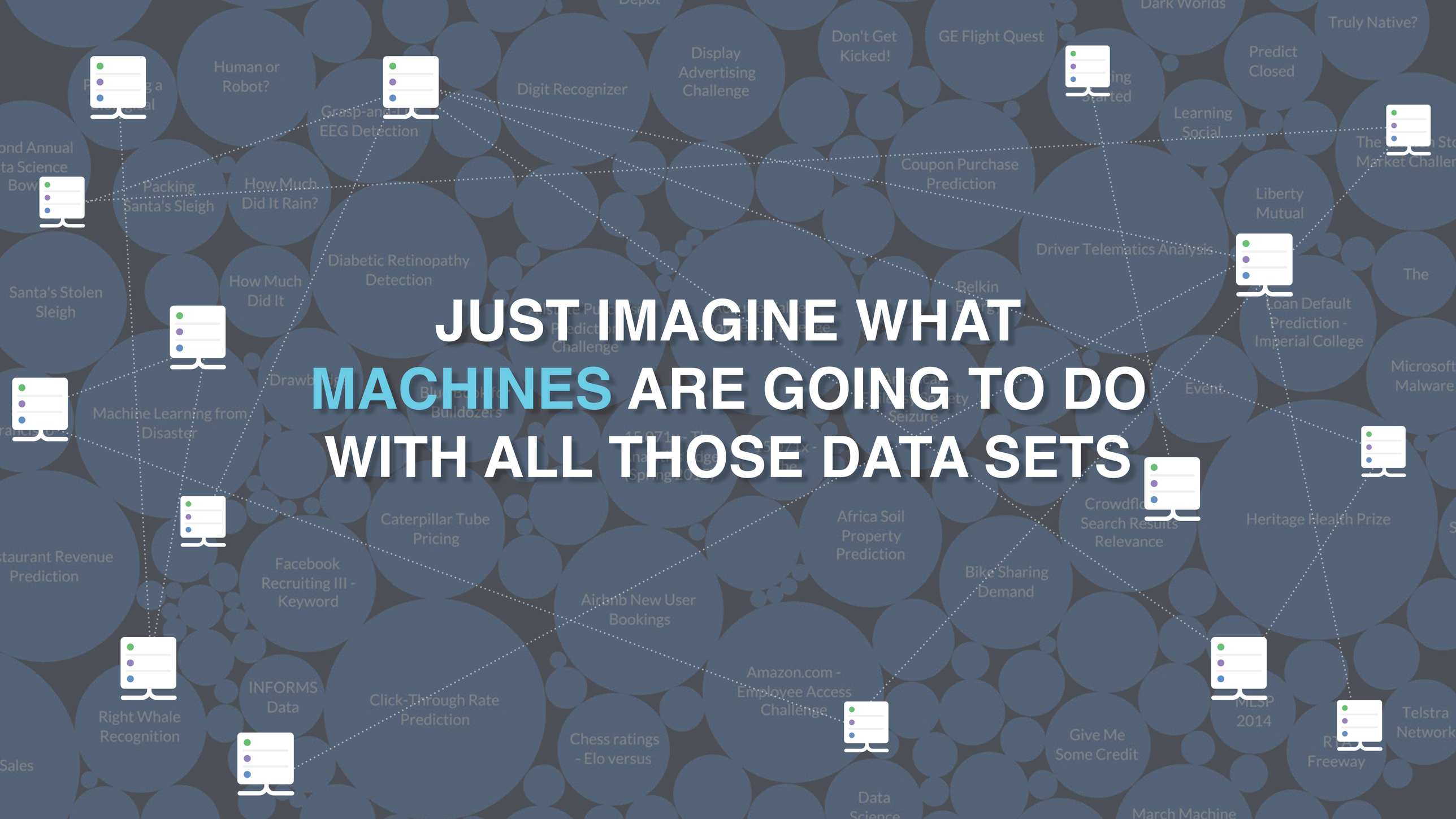
Yet, the majority of summonses written for this code violation were written by police officers. As a result, the department sent a training message to all officers clarifying the rule change and has communicated to commanders of precincts with the highest number of summonses, informing them of the issues within their command.

Thanks to this analysis and the availability of this open data, the department is also taking steps to digitally monitor these types of summonses to ensure that they are being issued correctly."

data.world

"*I was speechless.* **THIS is what the future of government could look like one day. THIS is what Open Data is all about.** *THIS was coming from the NYPD, who is not generally celebrated for its transparency, and yet it's the most open and honest response I have received from any New York City agency to date.* **Imagine a city where all agencies embrace this sort of analysis** *instead of deflect and hide from it.*"

JUST IMAGINE WHAT PEOPLE ARE GOING TO DO WITH ALL THOSE DATA SETS


Human or Robot?
Digit Recognizer
Display Advertising Challenge
Don't Get Kicked!
GE Flight Quest
Dark Worlds
Truly Native?
Predict Closed
Learning Social
Coupon Purchase Prediction
Packing Santa's Sleigh
How Much Did It Rain?
Liberty Mutual
Driver Telematics Analysis
Diabetic Retinopathy Detection
How Much Did It
Santa's Stolen Sleigh
Loan Default Prediction - Imperial College
Microsoft Malware
Machine Learning from Disaster
Event
Caterpillar Tube Pricing
Africa Soil Property Prediction
Heritage Health Prize
Facebook Recruiting III - Keyword
Bike Sharing Demand
Airbnb New User Bookings
Right Whale Recognition
INFORMS Data
Click-Through Rate Prediction
Amazon.com - Employee Access Challenge
Give Me Some Credit
Telstra Network
Chess ratings - Elo versus
Sales
Data
JUST IMAGINE WHAT
MACHINES ARE GOING TO DO
WITH ALL THOSE DATA SETS


But FINDING it
UNDERSTANDING it
and USING it can be a challenge
CSV
CSV

This process happens

OVER

AND OVER

AND OVER AGAIN

as each data user does it individually


So much HUMAN
EFFORT
is wasted on the
WORKING & REWORKING
of the
SAME DATA
XML
VII
XIX Data Science
data.world

The End

What is

LINKED DATA

?

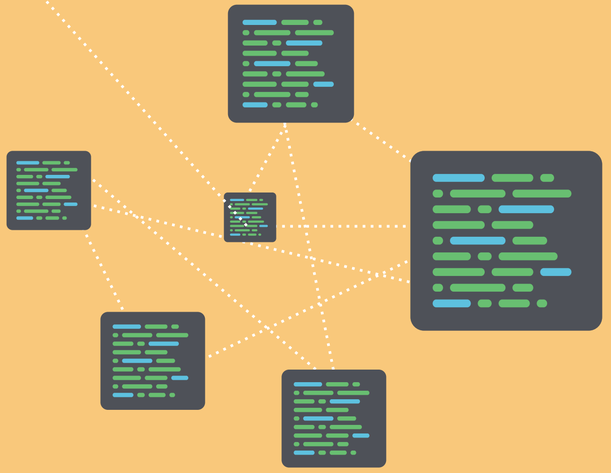
data.world

IMAGINE **RELEARNING WEB BROWSING**
FOR EACH NEW SITE YOU VISIT.

That's what it's like when data isn't linked.

It's applying the SAME *architecture as the* **WWW**

*of linked documents to...*DATA



data.world

First, break DATA *into* ATOMIC FACTS

(SUBJECT, PREDICATE, OBJECT)

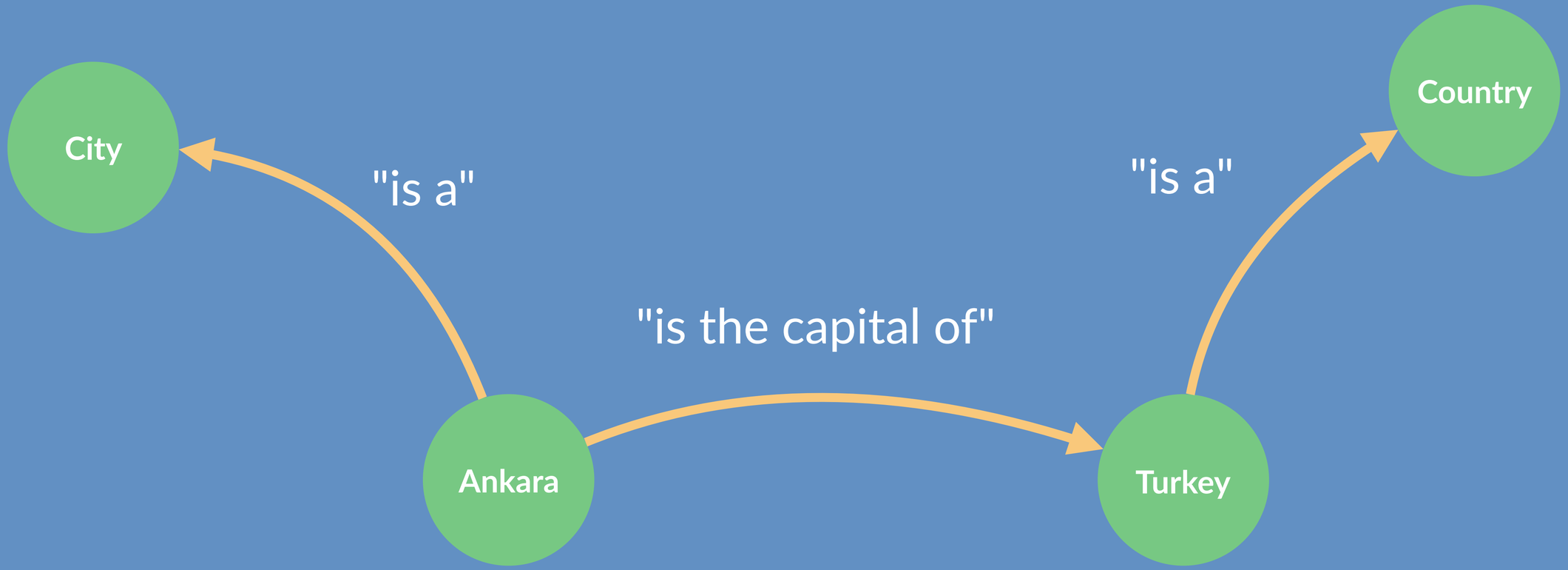
(Turkey, "is a", Country)

(Ankara, "is a", City)

(Ankara, "is the capital of", Turkey)

(SUBJECT, PREDICATE, OBJECT)

THE TRIPLE



data.world

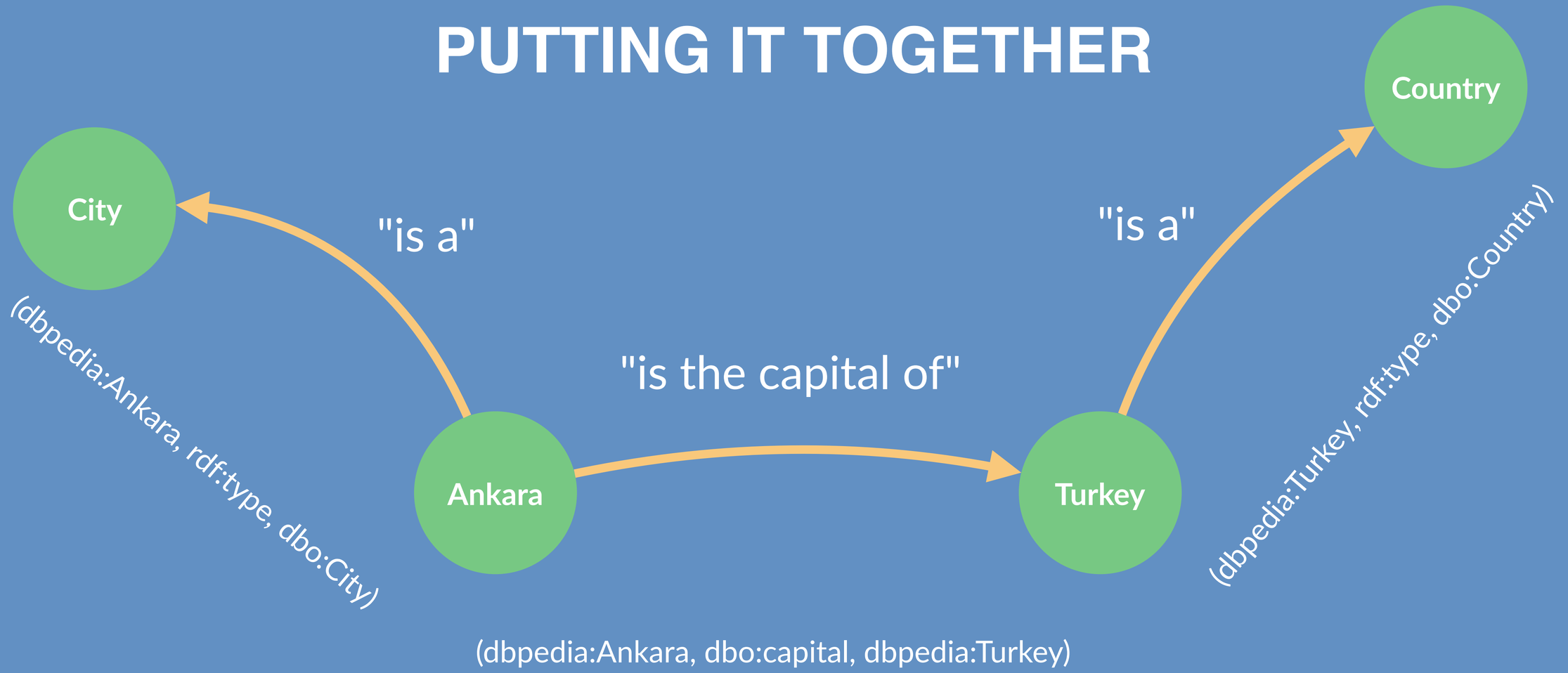
Refer to ENTITIES *and* RELATIONSHIPS
via **URIs** *so their* **MEANINGS** *can be discussed*

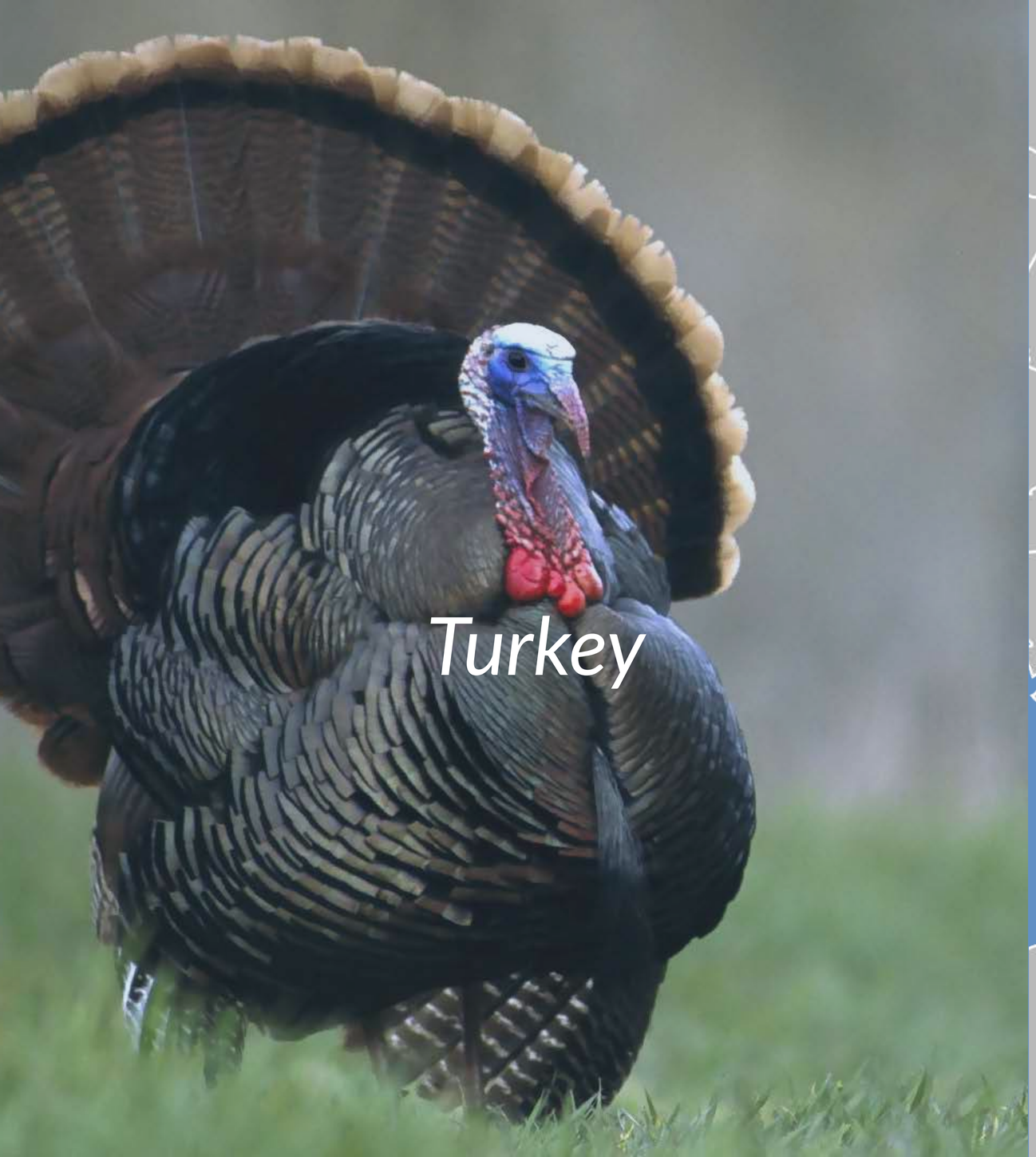
SUBJECT, PREDICATE, OBJECT

http://subject

http://object

PUTTING IT TOGETHER

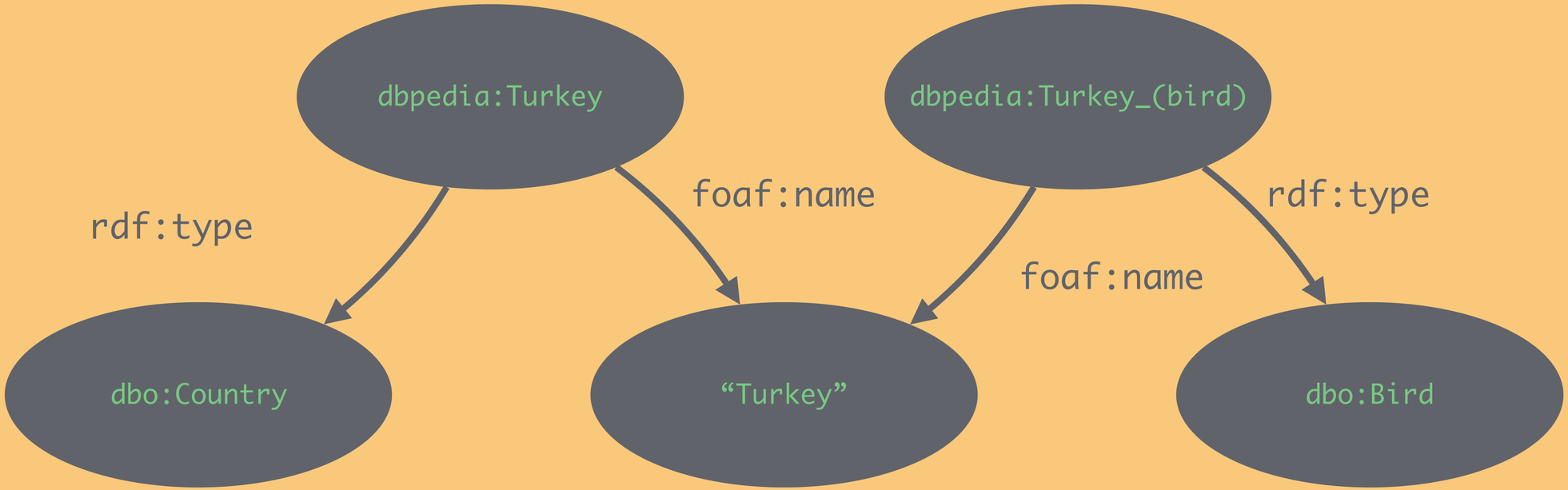



Turkey


Turkey

TURKEY *vs* TURKEY

```
(dbpedia:Turkey, rdf:type, dbo:Country)
(dbpedia:Turkey_(bird), rdf:type, dbo:Bird)
(dbpedia:Turkey, foaf:name, "Turkey")
(dbpedia:Turkey_(bird), foaf:name, "Turkey")
```



“AAA” Principal

ANYONE

Can say

ANYTHING

About

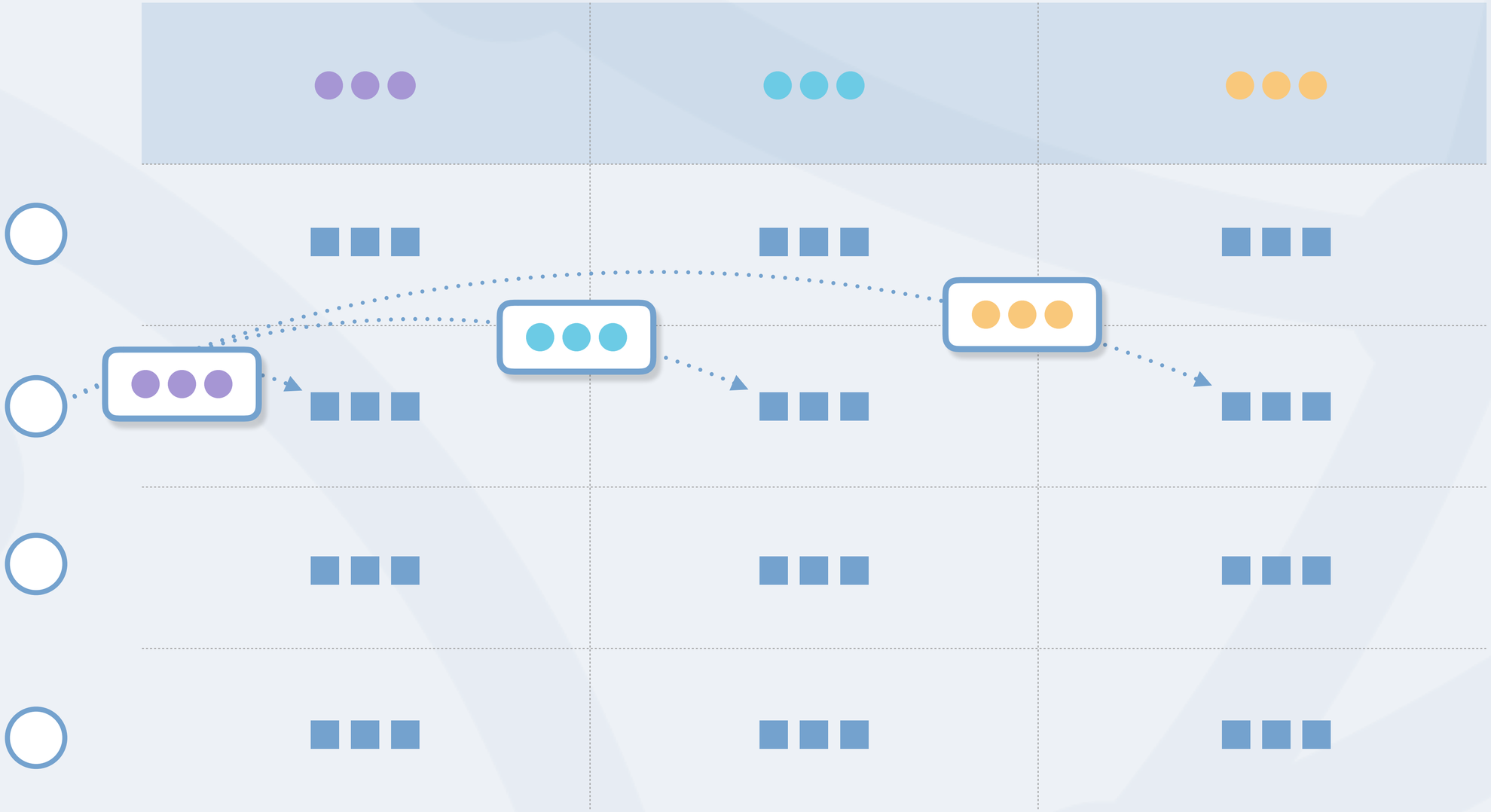
ANY TOPIC

YEA TRIPLES!

Triples are a universal format for representing facts - Any structured data can be mechanically transformed into triples.

data.world

TABULAR DATA AS A GRAPH



?

Why should you make your open data

LINKED

data.world

To make **DISCOVERY** *of your data easier*

To make your data **INTEROPERABLE**

To help the machines learn **FASTER**

The End ?

Data can enjoy a

"NETWORK EFFECT"

data.world

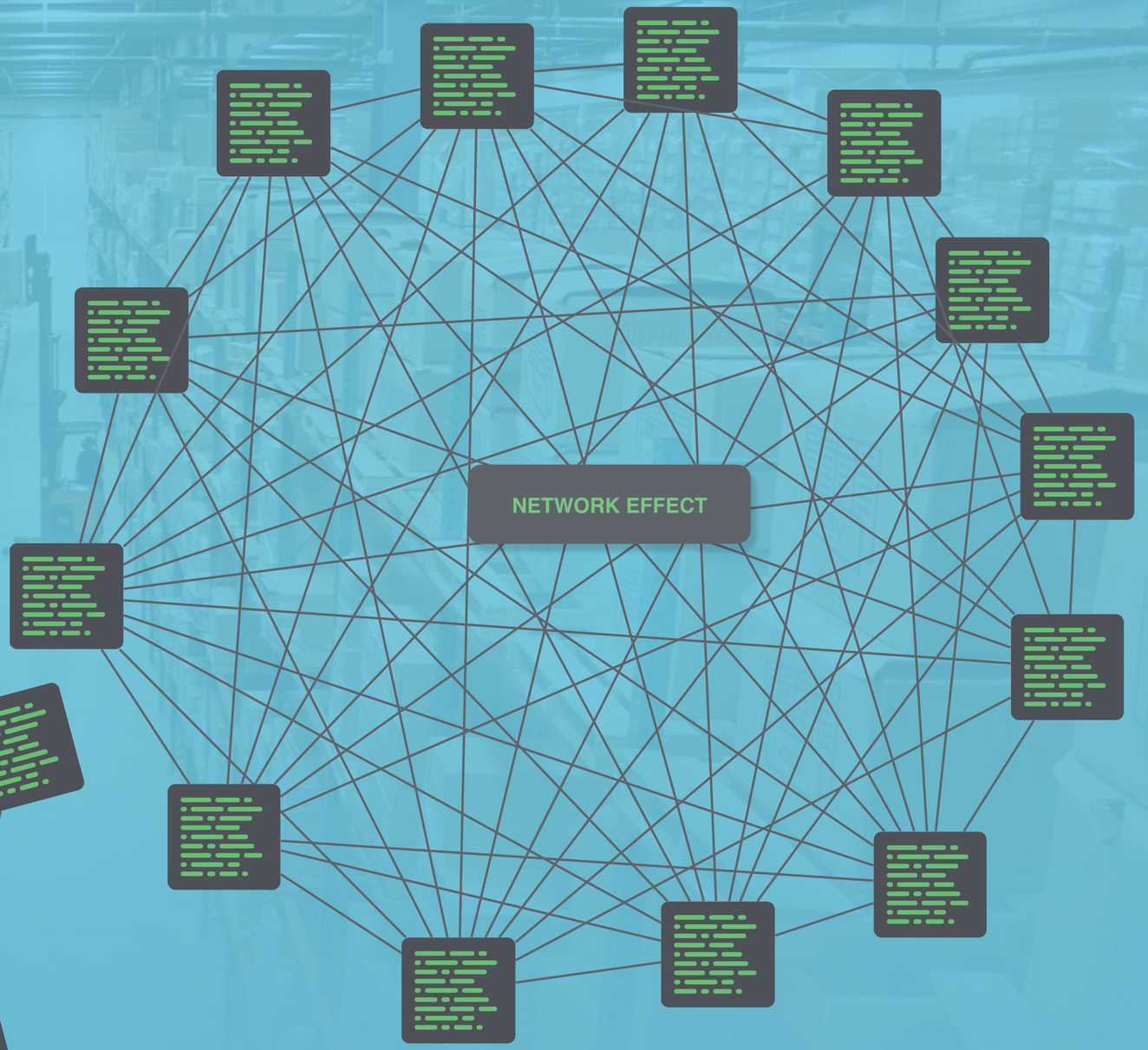
Each dataset that is added to the network

INCREASES the incremental VALUE

of every data set in the network

data.world

DATA NETWORK



data.world

LINKED DATA

is about publishing data as ATOMIC FACTS

and using **UNIVERSAL IDENTIFIERS**

to refer to concepts and relationships, so we can agree upon the SEMANTIC MEANING *of data.*

data.world

Your

OPEN DATA

wants to be

LINKED DATA

data.world


So the PEOPLE and MACHINES who are using that data to solve HUMANITIES BIGGEST PROBLEMS can leverage the sum of accumulated knowledge as effectively as possible.
Human or Robot?
Digit Recognizer
Display Advertising Challenge
Don't Get Kicked!
GE Flight Quest
Dark Worlds
Truly Native?
Predict Closed
Learning Social
Crasp-and-Lift EEG Detection
Coupon Purchase Prediction
Liberty Mutual
Driver Telematics Analysis
Packing Santa's Sleigh
How Much Did It Rain?
Diabetic Retinopathy Detection
Santa's Stolen Sleigh
Belkin
Imperial College
Microsoft Malware
Machine Learning from
Bulldozers
Seizure
Heritage Health Prize
Relevance
Property Prediction
Bike Sharing Demand
Facebook Recruiting III - Keyword
Airbnb New User Bookings
Amazon.com - Employee Access Challenge
Right Whale Recognition
INFORMS Data
Click-Through Rate Prediction
Chess ratings - Elo versus
Give Me Some Credit
MLSP 2014
Telstra Network
Freeway
data.world

The time to make your

OPEN DATA

accessible as

LINKED DATA

is

NOW!

data.world

The End

for real