OVERVIEW

data.world

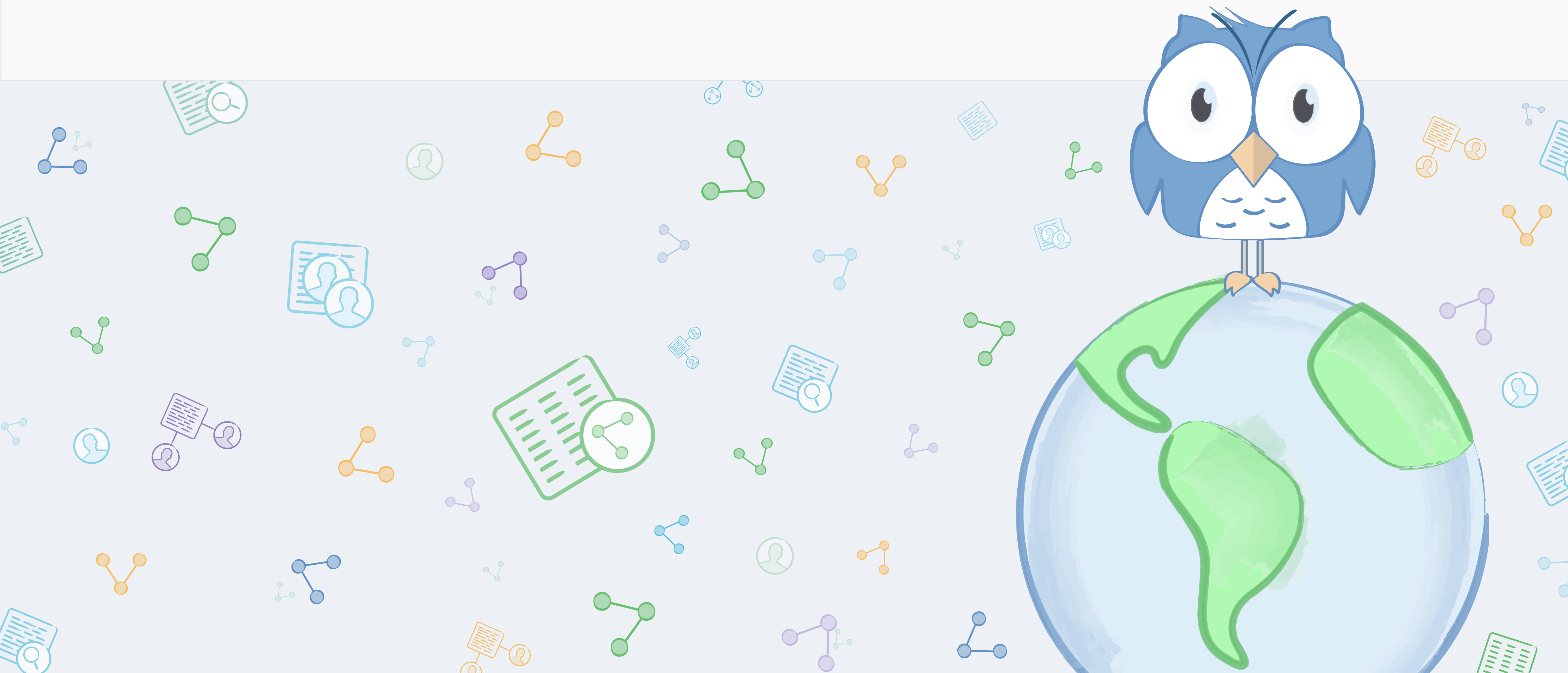## OUR PURPOSE

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

Our platform helps people who work with data solve problems faster by creating new ways to discover, prep, and collaborate.

data.world is the social network for data people.

data.world

OPEN DATA IMPACTS SOCIETY TODAY

2,400,000 websites in 1998

At least 7.5x as many open datasets today as there were websites at Google's launch

> **18,000,000** open datasets in 2016

Government:

- 1.8M open government datasets from the US, Canada, UK and Australia
- Data.gov has served millions of people over 8 years

Academia and research:

- Harvard Dataverse (73,000 datasets)
- Montreal Neurological Institute (first scientific institute to open up all results and data)

Business:

- The Weather Company (uses open government data in forecasts)
- Healthgrades.com (helps people choose providers with data from Medicaid, Medicare, state governments, more)

Citizen empowerment:

- Panama Papers (searchable database of 320,000 offshore companies and trusts)
- OpenSecrets.org (database of campaign contributions, lobbying expenditures, and more)
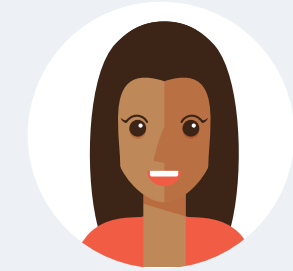
For more open data & data science facts, see our By the Numbers page

data.world

SOCIAL CHALLENGES

Will you share your data?

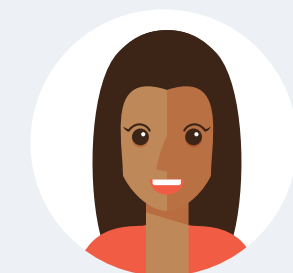*I am not sure that will be possible.*

But your work was published in Science, which requires you to share you data.

I saved it on a USB drive! It is in a box at home. I just moved. There are so many boxes! I forgot to label the boxes.

Later

When I opened up the data, it was in hexadecimal. I cannot read hexadecimal.

You asked me for the data and I gave it to you. You will need the program that created the file, Cytosynth. The company went bankrupt in 2007.

From "Data Sharing and Management Snafu in 3 Short Acts"

- Mindset, reputation, and capital don't yet reward proper implementation or deep collaboration

- Researchers rewarded for patents and published articles, not datasets or collaboration

- Hoarding mentality is still prevalent

- Reproducible research movement not living up to expectations

Would you rather conduct a three-hour onboarding session, or send someone a link?

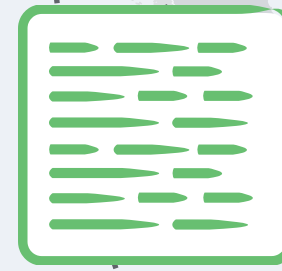**data.world**/data4democracy/election-transparency

WHAT WOULD IT TAKE TO SOLVE ZIKA FASTER?

Dr. Almeida
Recife, Brazil

Trying to ID a Zika-free *Aedes aegypti* replacement. Finds that another *Aedes* mosquito is Zika-resistant, but doesn't thrive in Recife's climate. Shares data via email with colleagues.

Hidalgo Flores
Mexico City, Mexico

Analyzing data from field agents after a week of prep. Finds evidence that Zika-resistant *Aedes vexans* mosquito thrives in warm coastal climates and overtakes *Aedes aegypti*. Shares data via FTP with colleagues. Posts data to Mexico's open data portal.
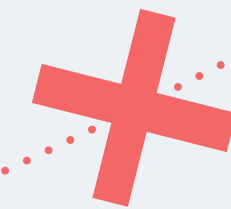
Flores has found the mosquito that Dr. Almeida needs, but Dr. Almeida won't know it until she attends Flores' conference session months later.

What kept this life-saving knowledge from reaching Dr. Almeida sooner?

data.world

THE NETWORK EFFECT

What if people around the world could look for other insights in the same data, post their findings for all to see, and discuss the data in context?
Networks leverage the wisdom of the crowd.

Hidalgo Flores
Mexico City, Mexico

Dr. Almeida
Recife, Brazil

NETWORK

What if Flores' data prep took an hour not a week?
Networks normalize data quickly and simplify transformation.

What if the coastal climate data in Dr. Almeida's file was automatically matched to the coastal climate data in Flores'?
Networks detect patterns instantly and new datasets increase the value of existing datasets.

What if Flores knew there was a place where his research would be seen and used, where his contribution would be recognized?
Networks extend reach and enhance discovery.

THIS IS DATA'S NETWORK EFFECT, AND IT CAN HELP US SOLVE EVERY PROBLEM WE FACE, FASTER.

data.world

data.world

WITH DATA.WORLD...

- **FINDING** the data you're looking for is **EASY**

- **DISCOVERING** data you didn't know existed is **EXCITING**

- **PREPARING** data is **PAINLESS** and **QUICK**

- **COLLABORATING** and **SHARING** is **SEAMLESS**, **SECURE**, and **SOCIAL**

- **COMBINING** open data with closed data is **SAFE** and **PRODUCTIVE**

Until today, doing any one of these things well would have required a custom data warehouse, millions of dollars, and thousands of development hours.

data.world is affordable enough to put on a personal credit card and starts helping new users solve problems faster within minutes of signing up.

WE MAKE OPEN DATA MEANINGFUL

data.world links datasets together using semantic technology. This gives concepts within data an independent existence, allowing people and machines to work with data without needing to learn everything about it first.

Linked data...

- Makes data discovery easier

- Makes data interoperable

- Helps people and machines unlock data's value faster

DATA CAN BECOME WELL DOCUMENTED OVER TIME

Joins from clinicaltrials.gov, cancer.gov & census.gov to examine cancer trials, mortality, incidence and demographics

Intelligence of dogs by breed

The Dollar-And-Cents Case Against Hollywood's Exclusion of Women

178,000 WWII air missions (digitized paper mission reports by location, mission and more than 60 other attributes)

The most comprehensive Product Hunt API dataset ever released

4 million Facebook posts from mainstream media sources

Complaints received by Bureau of Consumer Financial Protection about financial products and services from 2012-2016

Hate crime laws and data from the ADL and FBI, 1997-2015

Curated collection of 353 data visualization tools, resources and technologies

Using smartphones to collect movement patterns in Parkinson's patients

TOP DATASETS SHOW A SPECTRUM OF INTERESTS

USER STORIES

Enhancing data education and knowledge creation

Julie Schellberg uses data from data.world in her predictive analytics tutorial.

Vadim Markovtsev publishes metadata on 452,000,000 GitHub commits to data.world, and uses the dataset to prove the famous "Six Degrees of Separation" theory.
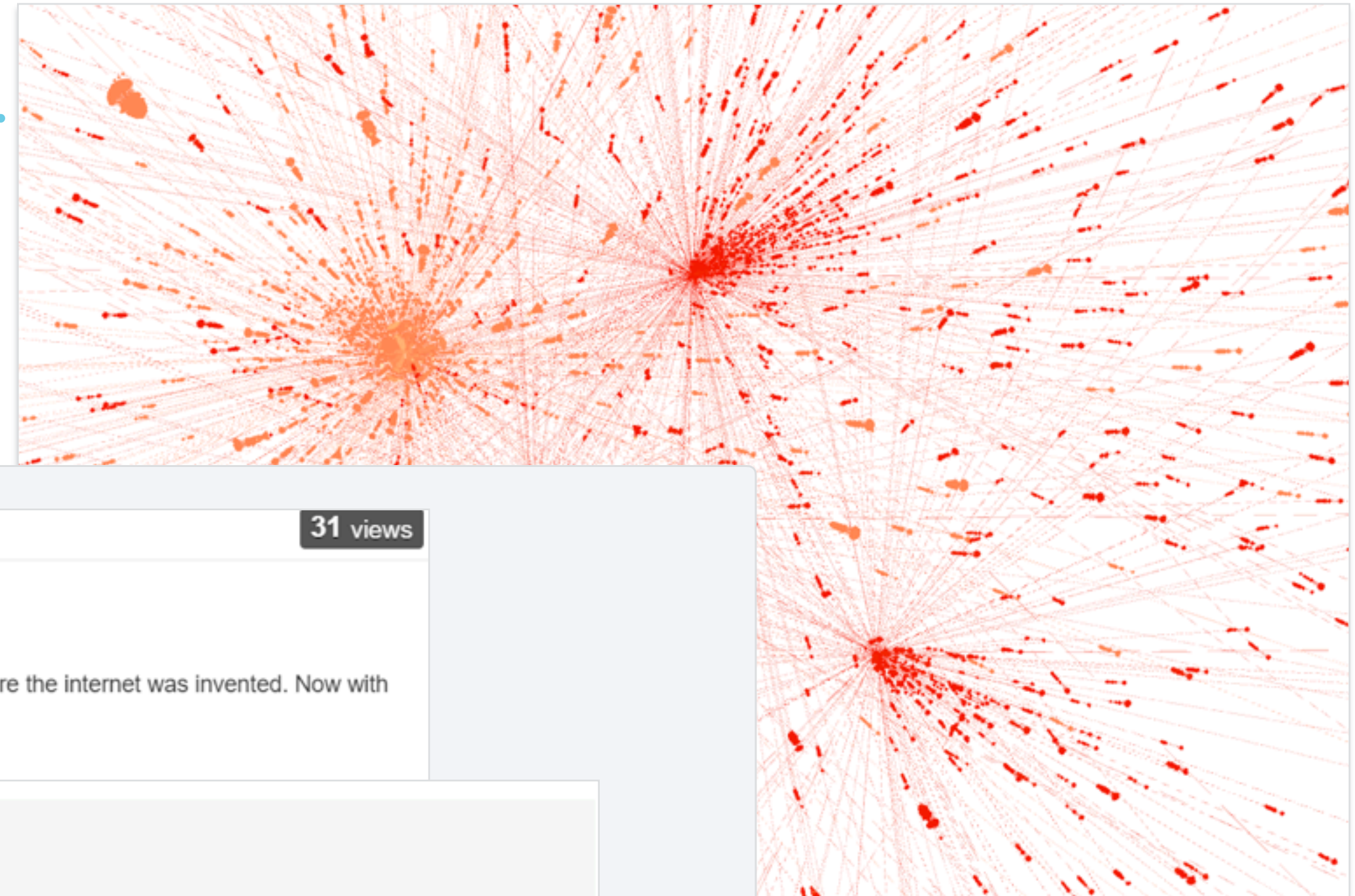
University of Texas CS professor Dr. Philip Cannata uses data.world extensively in his course Data Visualization.

I have created a DXP called lending_club_loans (2) to accompany this post using data from Data.World on Lending Club loans. I have dabbled a bit with Lending Club myself, so I found this data set particularly interesting.

note ☆   31 views

Here's a link to the paper I was talking about last class

- The Irresistible Move towards Interoperable Database Systems - this is a paper I wrote and delivered at a conference in 1991 before the internet was invented. Now with data.world, most all of it is starting to come true.
- I put the link under the Week 1 class notes, however, because I wanted it to be next to the data.world link.

#pin

class_notes

⦿ Resolved   ⦾ Unresolved

Wendy He 16 hours ago
How cool!
"If you can call anyone in the world from anywhere in the world, shouldn't you be able to locate and manipulate any relevant data in the world from anywhere in the world?"

Here's another article on the data revolution. "Data.world sets out to solve a huge problem—one most of us haven't considered very deeply. The world is awash in data, but nearly all of it is confined by policy, storage constraints, or lack of discoverability."

🔶 **Philip Cannata** 7 hours ago   From this article:
"In short, data.world makes data discoverable, interoperable, and social. And that could mean an explosion of data-driven insights is at hand."

That's exactly what I was talking about and working on at MCC (Austin's first high tech software company) in the early 90s.

Reply to this followup discussion

A NEW KIND OF COMPANY

As a public benefit corporation, data.world is legally structured to pursue its public benefit mission and shareholder value with equal vigor through growth, IPO, and beyond. This rare, powerful corporate structure allows data.world to make a profound impact by democratizing access to the world's data and creating a sustainable archive. data.world's specific public benefit purpose can be found at data.world/benefit-statement.

data.world is also a Certified B Corporation, a prestigious designation awarded to over 2,000 companies globally that meet the highest standards of overall social and environmental performance, transparency and accountability and aspire to use the power of business to solve social and environmental problems.

Top-5 VC-backed Certified B Corporations in tech (by funding)

THANK YOU

Carolee Mitchell
carolee@data.world

data.world