



# On Decentralizing AI Data & Services

Trent McConaghy  
@trentmc0

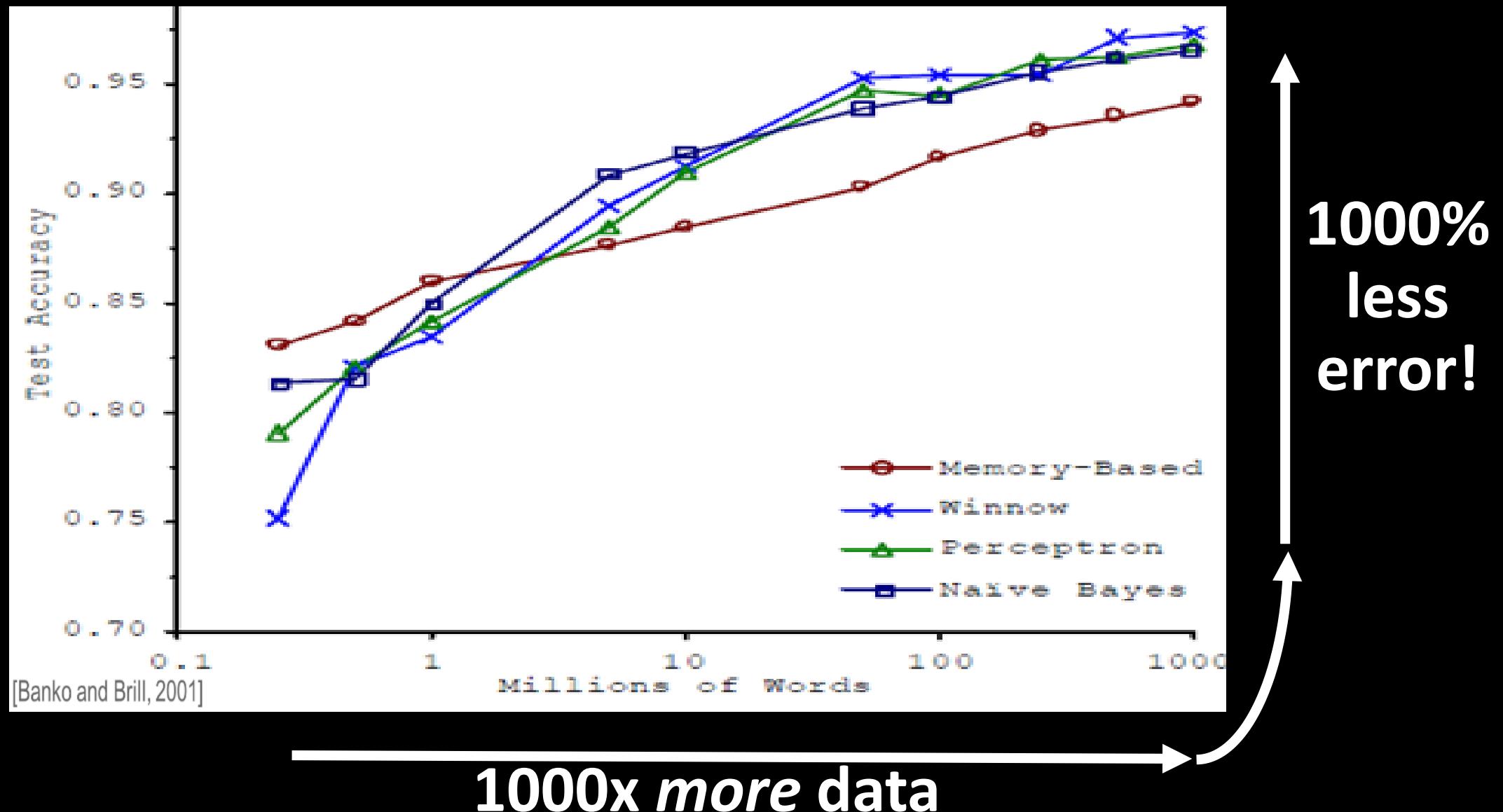
#Data  
#Incentives





# Audio radar

# The Unreasonable Effectiveness of Data



# The world's most valuable resource



Silo ++ data



++ accuracy



++ \$

Default incentive:  
hoard the data

**“Show me the incentive  
and I will show you the outcome.”**

-Charlie Munger

You can get people to do stuff  
by rewarding them with tokens.  
This is a superpower.





Change the  
incentives!

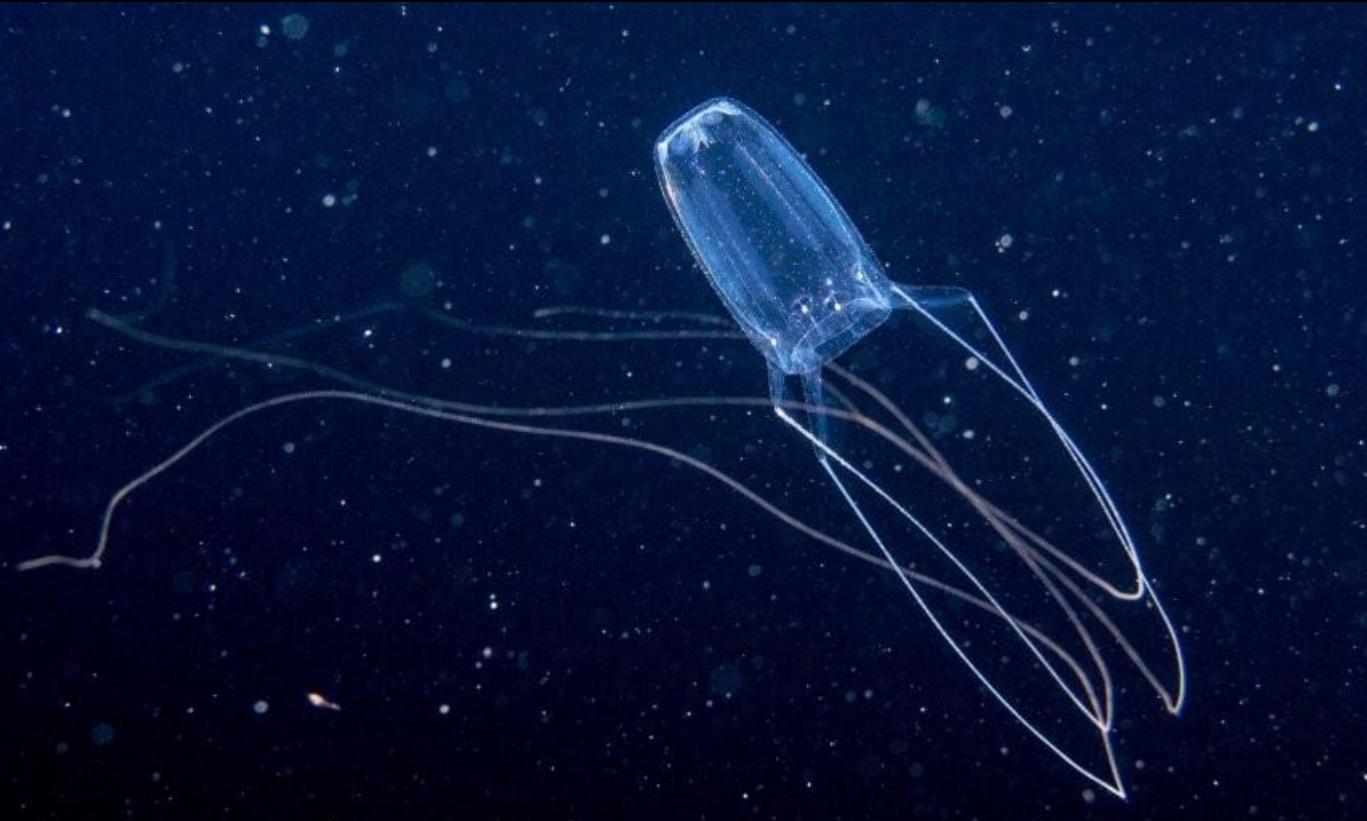
**Silo Pool** ++ data



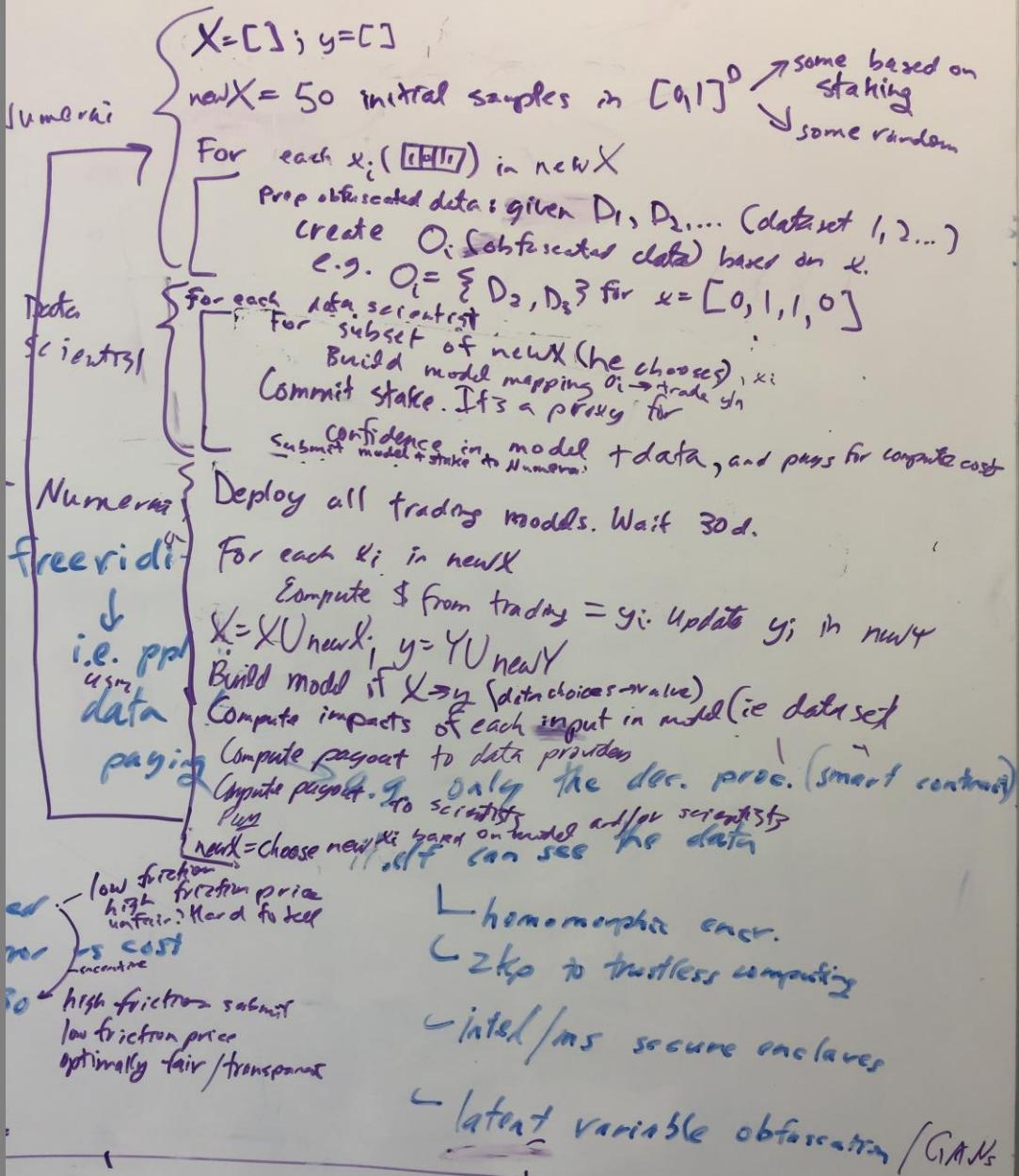
++ accuracy



++ \$

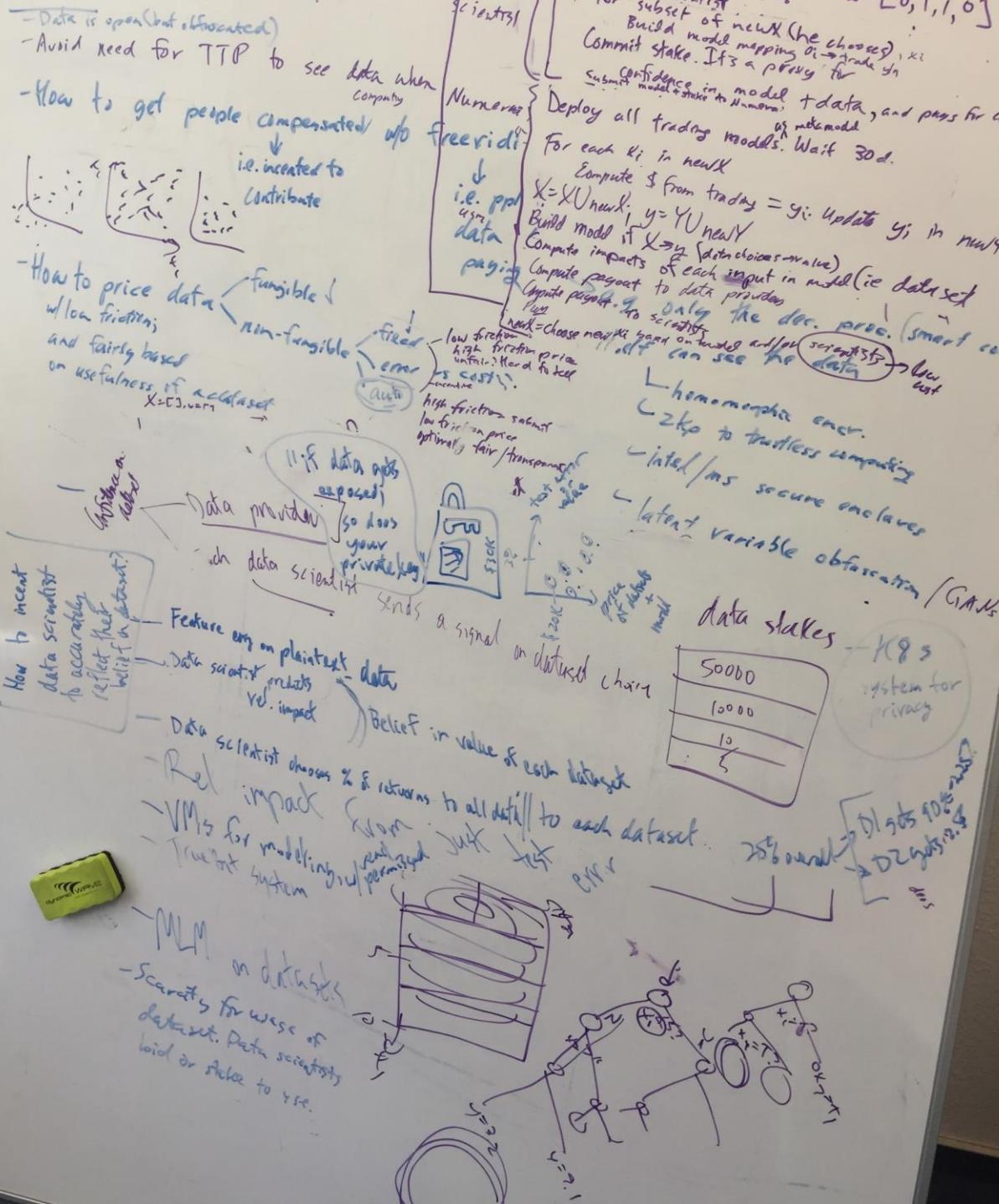


Early  
iterations



## Goals:

1. Healthy ecosystem → long term  
*initially*
2. Maintain ethics/values → governance

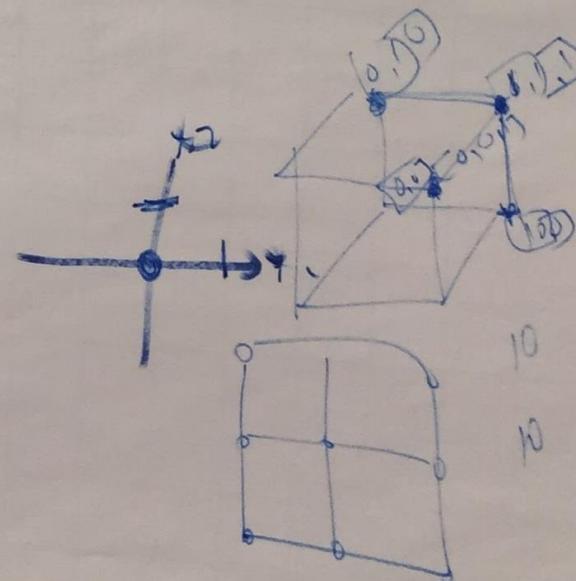
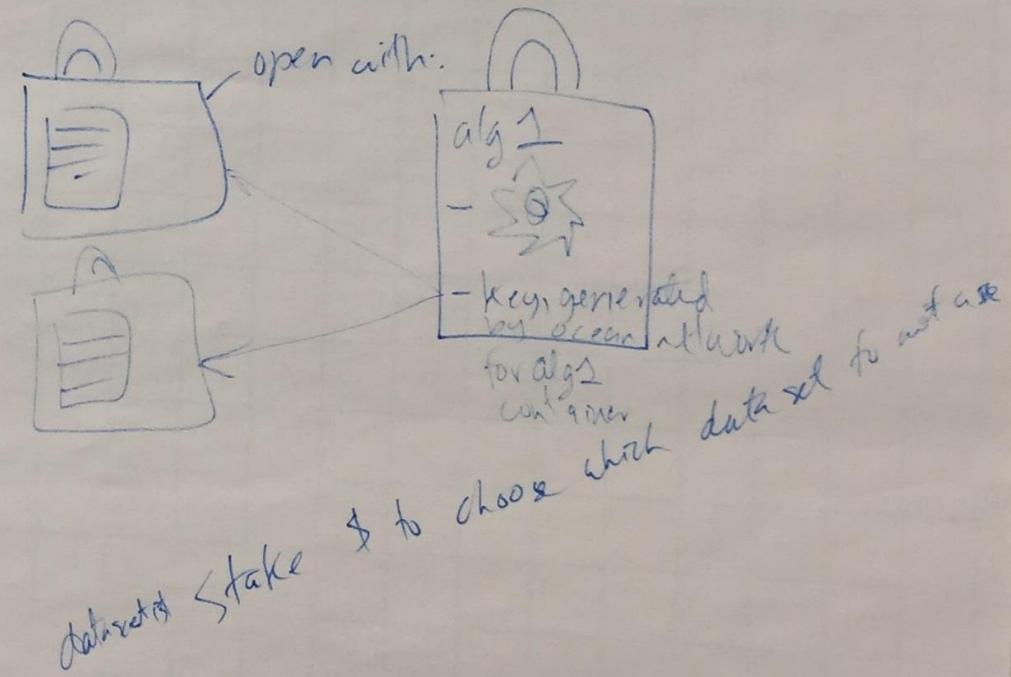


### Challenge

- who 1.1: how to ~~give~~ <sup>pay</sup> per dataset
- wif 1.1: feels wrong.

⑩ selected datasets  
⑩ models

1  
2  
3  
4  
5  
6  
7  
8  
9  
0



Data trading [Multi-level auction, user set price]

1. ~~DATA TOP 1K mbs~~ I have
2. Auction happens

highest bid → \$10K now  
 2nd-highest bid → \$6K in 2 mo  
 \$3K in 2 mo  
 \$1K in 3 mo

pre-set conditions based on top bid, #2 bid

When you get to #3 if "TOP 3 bids get data now; rest in 1-6 mos"

And: in 6 mos: data is set free.

Marketplace for obfuscating data

"Obfuscated data is only usable by Numerair itself"

= Stream / MLM  
 contents = data  
 - post contents, - post data,  
 get tokens, get tokens  
 - send tokens to update  
 - if others update  
 you get tokens

Pooling with ~~set price~~ supplier-set price

- 1. Supplier says total price \$20K.
- 2. Top bidders in pool get data now.
- 3. Rest get data later.

\$20K set by supplier

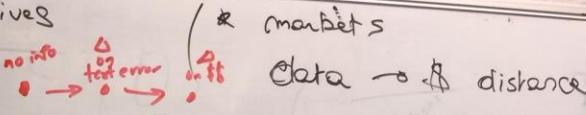
|       |
|-------|
| 10K   |
| 7K    |
| 5.32K |
| 1.1K  |
| 2K    |
| 1K    |

These folks get data now

others get data in 1 mo or 2 mo (data set free)

graduated DTs 1 mo, 2 mo, 3 mo

## Engagements / Incentives



- fix price up-front

- " " "

+ reputation (subjective)

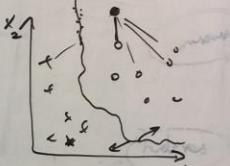
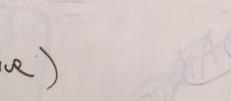
1 - You have \$50K stored in a wallet  
- 1/10000 unique wallet

- each key is a ~~secret~~ hash of ~~row~~ of data (HD wallet)

2. Able to compute a key from data is public

3. ∴ if my data is made public,  
then anyone can open my wallet & get \$50K

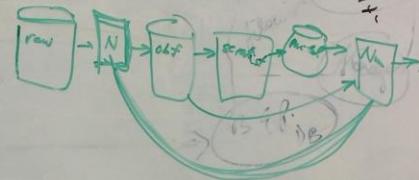
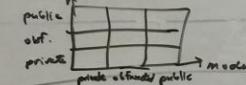
Shake



Model parents visible for data supplier

data visible for scientist

data private - model parents private

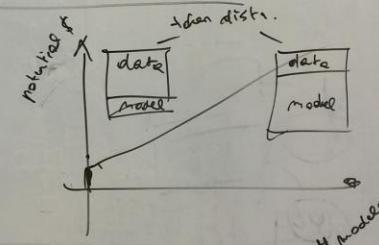


Co-owner of \$

or Home Markets?

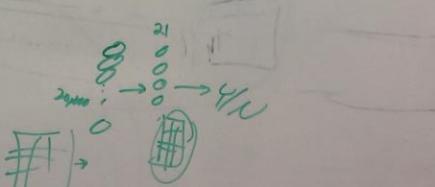
Tokenizing Access to Data Revenue, Fixed Supply

- Each dataset has its own tokens. Fixed supply
- "ICO a dataset" & is purchased for usage, it is split according to token ownership



Tokenizing Access to Data Itself

e.g. 100 tokens. You can access the data if you own the token.



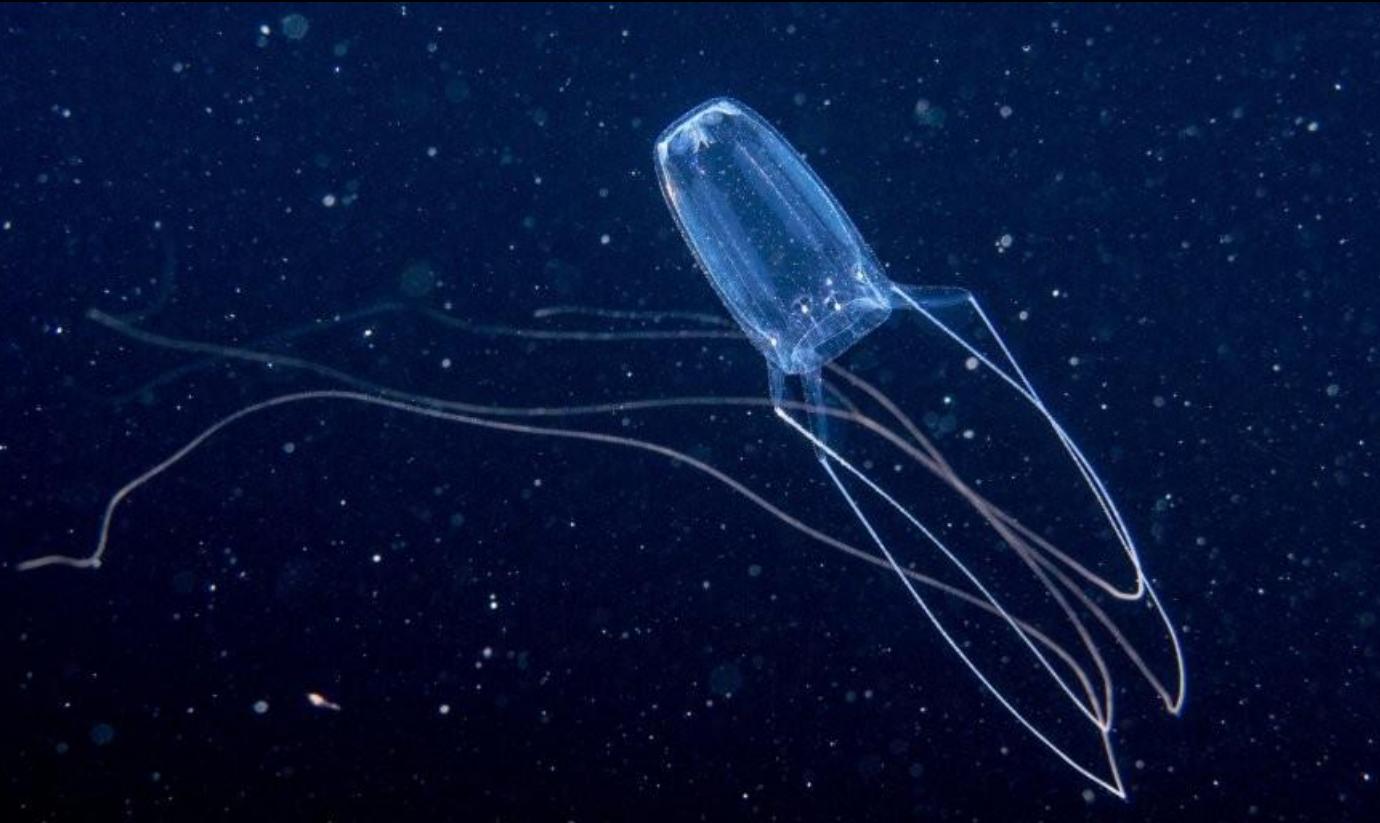
## Challenges

0. Deteriorate data.

1. How to ensure supplier gets paid w/o losing ability to get paid in future. "Free riding"  
"Privacy"  
"Copy vs title"

2. Friction in pricing) — overall price  
relative impact per dataset

| Static  | Dynamic dataset<br>(goes stale)  | Fungible → Non fungible   |
|---|--|---|
| <p><u>Signals</u></p> <ul style="list-style-type: none"><li>- stake in belief of <del>data</del> <sup>surprise</sup></li><li>- value of dataset <sup>surprise</sup></li><li>- by data scientist</li><li>- price asked by supplier</li><li>- price bid by scientist</li><li>- reputation of dataset</li><li>- supplier</li><li>- scientist</li><li>- boundaries for non-free-riding detection</li><li>- PNL <sup>value gained from dataset(s) in PNL</sup></li><li>- F <sup>e.g. Numerical models; insurance surveys</sup></li><li>- total value of network <sup>value per dataset</sup></li><li>- prediction market belief in value of dataset</li><li>- novelty of a dataset</li><li>- take a price of dataset</li><li>- scarcity demand</li></ul> | <p><u>Tools to address free riding</u></p> <ul style="list-style-type: none"><li>- Set the "free" after data stream to own</li><li>- Data provider doesn't care</li><li>- Licensing</li><li>- fine print <sup>combine</sup></li><li>- watermarking</li><li>- reputation</li><li>- provenance</li><li>- risk of litigation</li><li>- only the smart contract can see the data. Eg. docker + locks<br/>↓ MPC, zkSN</li><li>- If data set free, you lose your private key gets exposed<br/>Cloud storage of TTS</li></ul> | <p>Promote via ligendary rises</p> <p>Commercial<br/>Contract terms open source may<br/>impose restrictions</p> |



## Early iterations: *Flailing*

Can we  
structure this  
better?

# Realization: Tokenized Ecosystems Are a Lot Like Evolutionary Algorithms (EAs)!

| What                       | Tokenized ecosystem   | Evolutionary Algorithm  |
|----------------------------|---|---|
| Goals                      | Block reward function<br>E.g. “Maximize hash rate”                              | Objective function<br>E.g. “Minimize error”                                 |
| Measurement & test         | Proof<br>E.g. “Proof of Work”   | Evaluate fitness<br>E.g. “Simulate circuit”                                 |
| System agents              | Miners & token holders (humans)<br>In a network                                 | Individuals (computer agents)<br>In a population                            |
| System clock               | Block reward interval   | Generation  |
| Incentives & Disincentives | You can't control human,<br>Just reward: give tokens<br>And punish: slash stake | You can't control individual,<br>Just reward: reproduce<br>And punish: kill |

We can approach token design  
as EA design.

# EA Design



# Steps in EA (and Optimization) Design

1. **Formulate the problem.** Objectives, constraints, design space.
2. **Try an existing solver.** If needed, try different problem formulations or solvers.
3. **Design new solver?**

# 1. Formulation of an optimization problem

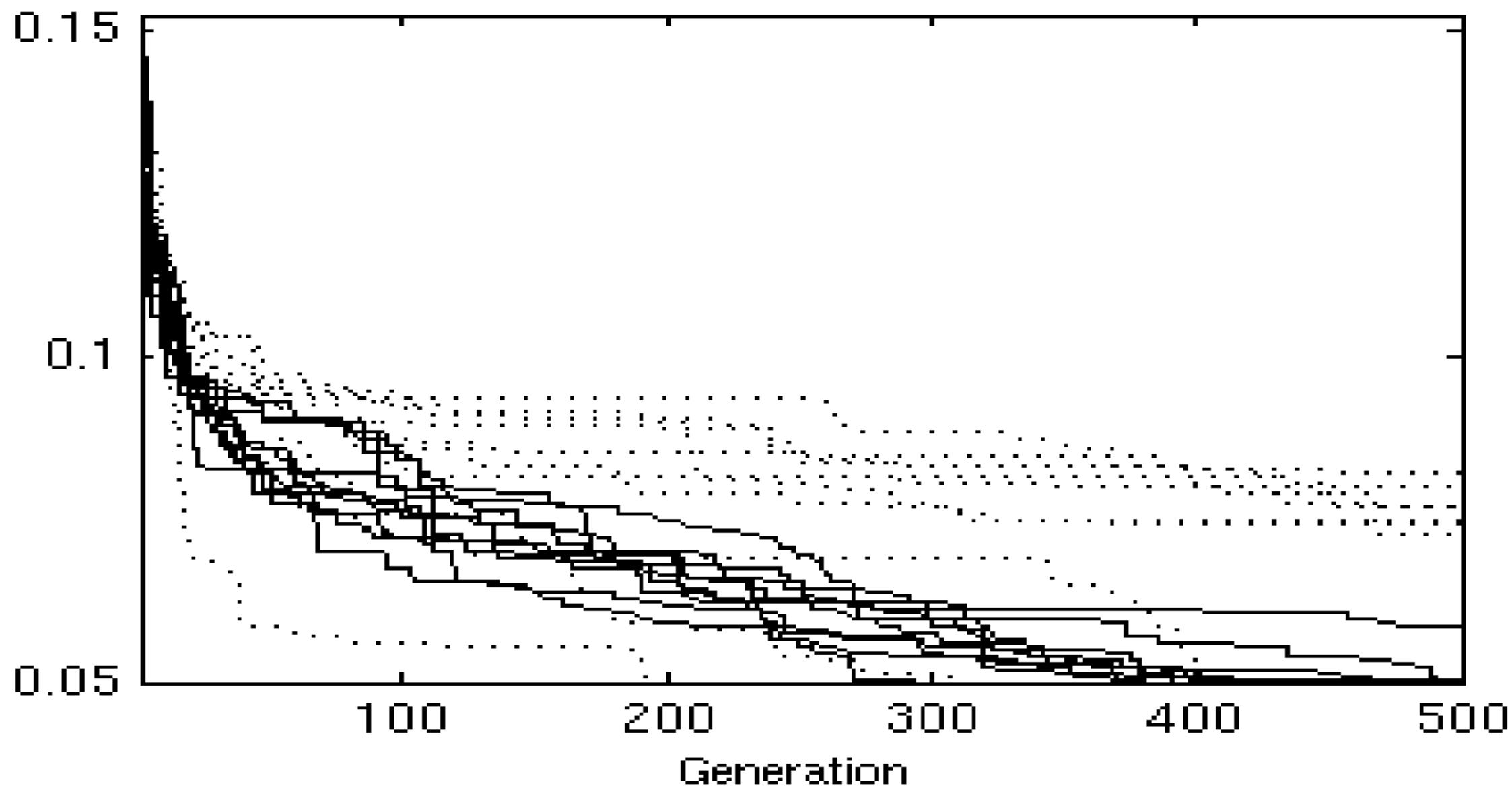
## Objectives & constraints in a design space

The algorithm's aim is formulated as a constrained multi-objective optimization problem

$$\begin{aligned} \text{minimize} \quad & f_i(\phi) \quad i = 1 \dots N_f \\ \text{s.t.} \quad & g_j(\phi) \leq 0 \quad j = 1 \dots N_g \\ & h_k(\phi) = 0 \quad k = 1 \dots N_h \\ & \phi \in \Phi \end{aligned} \tag{1}$$

where  $\Phi$  is the “general” space of possible topologies and sizings. The algorithm traverses  $\Phi$  to return a Pareto-optimal

## 2. Try an existing solver. Does it converge?



### 3. Design new solver

TABLE II  
PROCEDURE SANGRIAOPTIMIZATION()

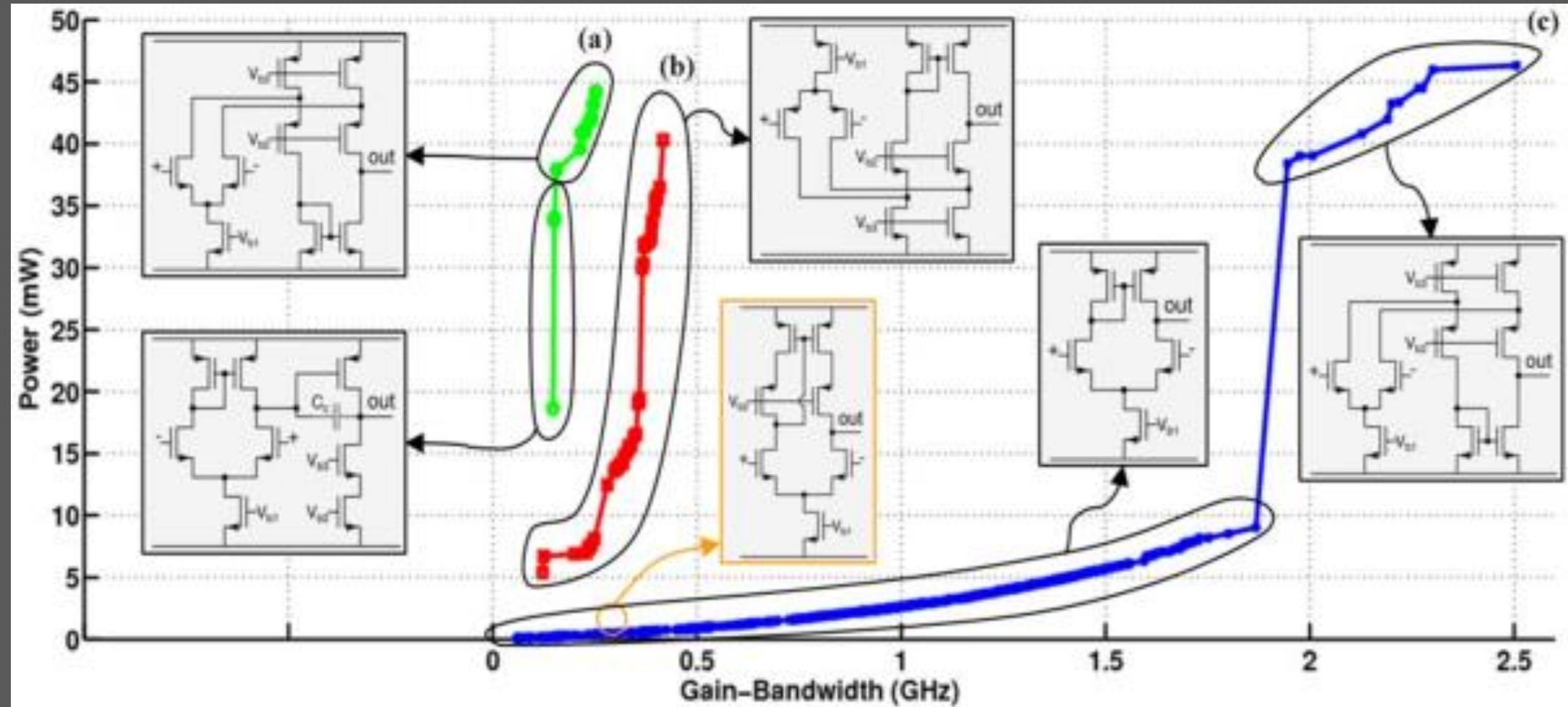
**Inputs:**  $D, N_a, K, N_L(k)$

**Outputs:**  $d^*$

1.  $N_{gen} = 0; P = \emptyset, P_{all} = \emptyset$
2. while  $\text{stop}() \neq \text{True}$ :
  3. if  $(N_{gen} \% N_a) = 0$ :
    4. if  $|P| < K$ :
      5.  $P_{|P|+1} = \emptyset$
      6.  $P_0 = \text{SpaceFillIndividuals}(N_L(k), N_D, D)$
      7. for  $k = 1$  to  $|P|$ :
        8.  $P_k = \text{SelectParents}(P_k, P_{k-1}, N_L(k))$
        9.  $P_{k,j} = \text{UpdateLocalOptState}(P_{k,j}, k), j = 1$  to  $|P_k|$
      10.  $P_{all} = \text{unique}(P_{all} \cup P)$
      11.  $P_{|P|} = P_{|P|} \cup \text{InnerOptimize}(P_{all}, D, k)$
      12.  $d^* = d_i$  in  $P_{all}$  with highest  $Y$  or  $Cpk$
      13.  $N_{gen} = N_{gen} + 1$
    14. return  $d^*$

and all individuals encountered so far in the search,  $P_{all}$ . Lines 2–13 are the generational loop, which repeats until stop.

# Example of a Successful Outcome



# Token Design as EA Design



# Steps in *Token* Design

1. **Formulate the problem.** Objectives, constraints, design space.
2. **Try an existing pattern.** If needed, try different formulations or patterns.
3. **Design new pattern?**

# 1. Formulate the Problem

(a) Ask

- Who are my potential **stakeholders**?
- And what do each of them **want**?
- What are possible **attack vectors**?

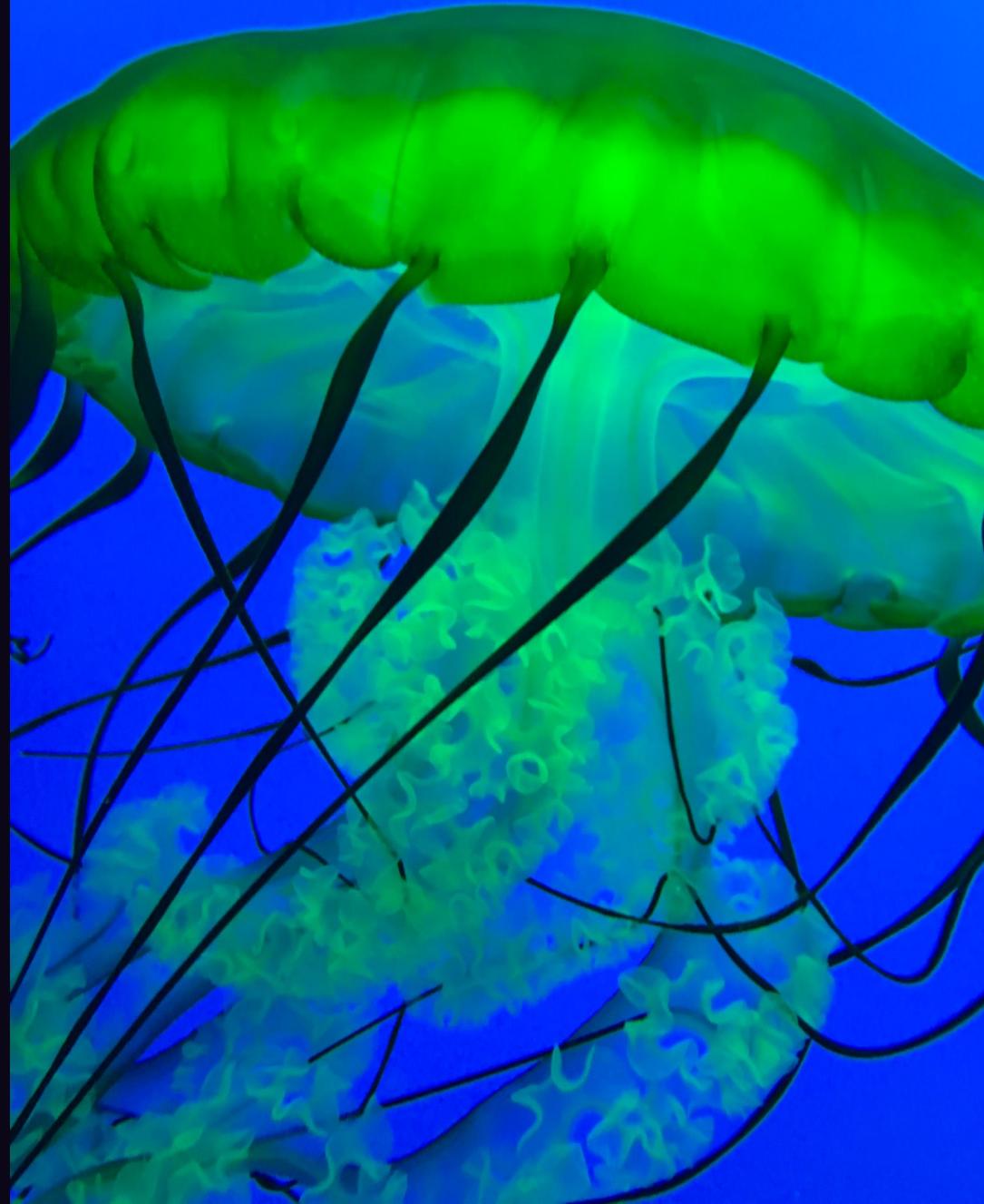
(b) Translate those into objectives and constraints.

## 2. Try Existing Patterns

1. Curation
2. Proofs of human or compute work
3. Identity
4. Reputation
5. Governance / software updates
6. Third-party arbitration
7. ...

## 2.1 Patterns for Curation

- **Binary** membership: Token Curated Registry (TCR)
- **Discrete-valued** membership: Stake Machines
- **Continuous-valued** membership: Curation Markets characterized by bonding curve
- **Hierarchical** membership: each label gets a TCR
- **Work** tied to membership: Proofed Curation Market
- **Non-fungible** tokens: Re-Fungible Tokens



# Case Study: Analysis of Bitcoin



# Bitcoin objective function

**Objective: Maximize security of network**

- Where “security” = compute power
- Therefore, super expensive to roll back changes to the transaction log



# Bitcoin objective function

**Objective: Maximize security of network**

- Where “security” = compute power
- Therefore, super expensive to roll back changes to the transaction log

$$E(R_i) \propto H_i * T$$

$E()$  = expected  
value

block  
rewards

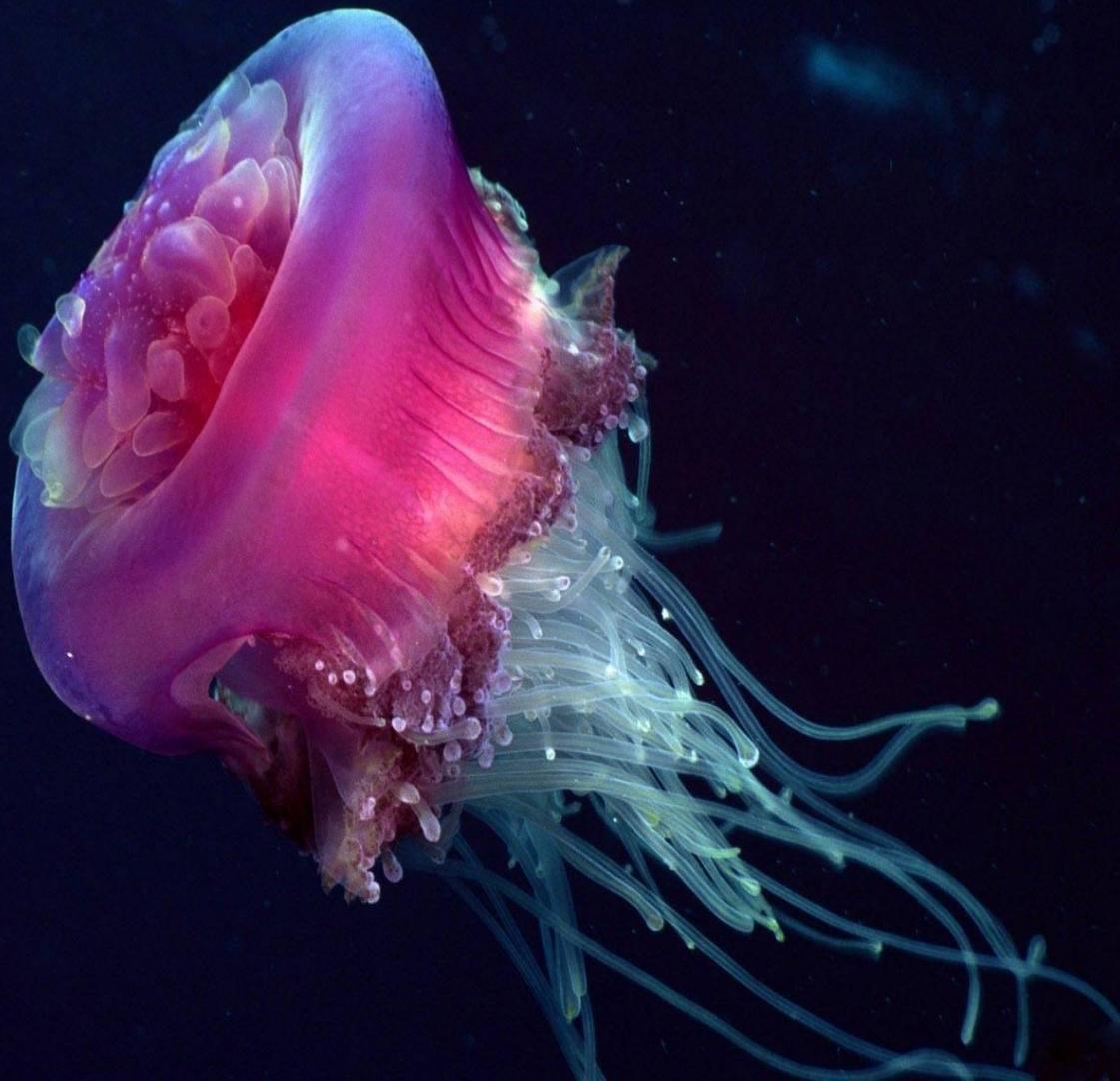
hash power of actor  
= contribution to  
“security”

# tokens (BTC)  
dispensed each  
block

# Result of Bitcoin's objective function:

People are maximizing security! = Maximizing electricity





# Case Study: Design of Ocean

# 1. Formulate the Problem:

## (a) Who are stakeholders? What do they want?

Key stakeholders in Ocean ecosystem

| Stakeholder  | What value they can provide                     | What they might get in return                   |
|--|---|---|
| <b>Data/service provider, data custodian, data owner</b>   | Data/service (market's supply)                  | Tokens for making available / providing service |
| <b>Data/service referrers, curators. Includes exchanges and other application-layer providers.</b> | Data/service (via a provider etc), curation     | Tokens for curating                             |
| <b>Data/service verifier. Includes resolution of linked proofs on other chains</b>                 | Data/service (via a provider etc), verification | Tokens for verification                         |
| <b>Data/service consumer</b>   | Tokens  | Data/service (market's demand)                  |
| <b>Keepers</b>   | Correctly run nodes in network                  | Tokens for chainkeeping                         |

# 1. Formulate the problem: (b) Translate into objectives and constraints

Objective function: maximize supply of relevant data

Token rewards if: supply relevant data

Token rewards if: supply data, and curate it

# 1. Formulate the problem: (b) Translate into objectives & constraints

**Constraints = checklist:**

- For priced data, is there incentive for supplying more? Referring?
- For priced data, good spam prevention?
- For free data, is there incentive for supplying more? Referring?
- For free data, good spam prevention?
- Does the token give higher marginal value to users of the network versus external investors? Eg Does return on capital increase as stake increases?
- Are people incentivized to run keepers?
- Is it simple? Is onboarding low-friction?

## 2. Try Existing Patterns

Some patterns:

1. Actor registry
2. Data registry
3. Actor registry + data registry
4. Data registry + free-as-in-beer data curation market.

Curation: Pay tokens to listen.

## 2. Try existing patterns: evaluate on objectives & constraints. None passed...

| Key Question   | 1 | 2 | 3 | 4 |
|--|---|---|---|---|
| For priced data: incentive for supplying more? Referring?  | ✗ | ≈ | ✓ | ≈ |
| For priced data: good spam prevention?   | ≈ | ✓ | ✓ | ✓ |
| For free data: incentive for supplying more? Referring?  | ✗ | ≈ | ✗ | ✓ |
| For free data: good spam prevention?   | ≈ | ✓ | ≈ | ✓ |
| Does token give higher marginal value to users of the network, vs external investors? Eg Does return on capital increase as stake increases? | ✓ | ✓ | ✓ | ✓ |
| Are people incentivized to run keepers?  | ≈ | ≈ | ✓ | ✓ |
| It simple? Is onboarding low-friction? Where possible, do we use incentives/crypto rather than legal recourse?                               | ✓ | ✓ | ≈ | ≈ |

### 3. Try New Patterns

Some patterns:

1. Actor registry
2. Data registry
3. Actor registry + data registry
4. Data registry + free-as-in-beer data curation market. Curation:  
Pay tokens to listen.
5. **Data registry + free data curation market. Curation: Stake  
tokens as belief in reputation. Auto CDN.**
6. **Actor registry + free&priced data curation market. Curation:  
Stake tokens as belief in reputation. Auto CDN. “Proofed  
Curation Market”**

### 3. Try new patterns: evaluate on objectives & constraints

| Key Question   | 1 | 2 | 3 | 4 | 5 | 6 |
|--|---|---|---|---|---|---|
| For priced data: incentive for supplying more? Referring?  | ✗ | ≈ | ✓ | ≈ | ≈ | ✓ |
| For priced data: good spam prevention?   | ≈ | ✓ | ✓ | ✓ | ✓ | ✓ |
| For free data: incentive for supplying more? Referring?  | ✗ | ≈ | ✗ | ✓ | ✓ | ✓ |
| For free data: good spam prevention?   | ≈ | ✓ | ≈ | ✓ | ≈ | ✓ |
| Does token give higher marginal value to users of the network, vs external investors? Eg Does return on capital increase as stake increases? | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Are people incentivized to run keepers?  | ≈ | ≈ | ✓ | ✓ | ✓ | ✓ |
| It simple? Is onboarding low-friction? Where possible, do we use incentives/crypto rather than legal recourse?                               | ✓ | ✓ | ≈ | ≈ | ✓ | ✓ |

# Objective: maximize supply of relevant data

- Reward curating data (staking on it) + making it available
- New pattern: Proofed Curation Market

$$E(R_{ij}) \propto \log_{10}(S_{ij}) * \log_{10}(D_j) * T * R_i$$

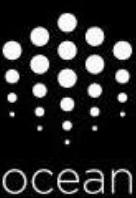
Expected  
reward for user  
 $i$  on dataset  $j$

$S_{ij}$  = predicted popularity  
= user's curation market  
stake in dataset  $j$

$D_j$  = proofed popularity  
= # times made dataset  
available

# tokens  
during  
interval

# From AI data to AI *services*



# Motivations:

- Privacy, so compute on-premise or decentralized
  - Data is heavy, so compute on-premise
  - Link in emerging decentralized AI compute

## **Objective function: Maximize supply of relevant *services***

**=reward curating *services* + proving that it was delivered**

$$E(R_{ij}) \propto \log_{10}(S_{ij}) * \log_{10}(D_j) * T * R_i$$

#Token  
Engineering



# Design of Tokenized Ecosystems

## From Mechanism Design to *Token Engineering*

Analysis:

Game theory

Synthesis:

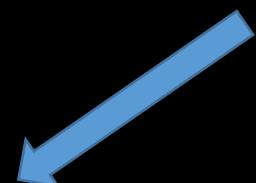
Mechanism Design



↓  
Practical  
constraints

Optimization Design

↓  
Engineering theory,  
practice and tools  
+ responsibility



**Token Engineering for Analysis & Synthesis**

# #TokenEngineering has grown into a global movement

ot secure | tokenengineering.net/community

site-name .wikidot.com Share on [Twitter](#) [Facebook](#) [Link](#) [Print](#) [Email](#)

Expert tip #1: Wikidot modules let you make smarter web sites

[main](#) [discuss](#) [edit this page](#) [view source](#) [history](#) [options](#)

## Community

Community: Events, Related Communities, more.

### TE Local Meetup Groups

(The actual meetup.com pages will typically have the name of the group in the URL.)

- [TE Berlin](#)
- [TE Budapest](#)
- [TE Munich](#)
- [TE Toronto](#)
- [TE London](#)
- [TE Zurich/Zug](#)
- [TE Tokyo](#)
- [TE NYC](#)
- [TE Amsterdam](#)
- [TE Stockholm](#)
- [TE San Francisco](#)
- [TE Vancouver](#)
- [TE Vienna](#)
- [TE St Petersburg](#)

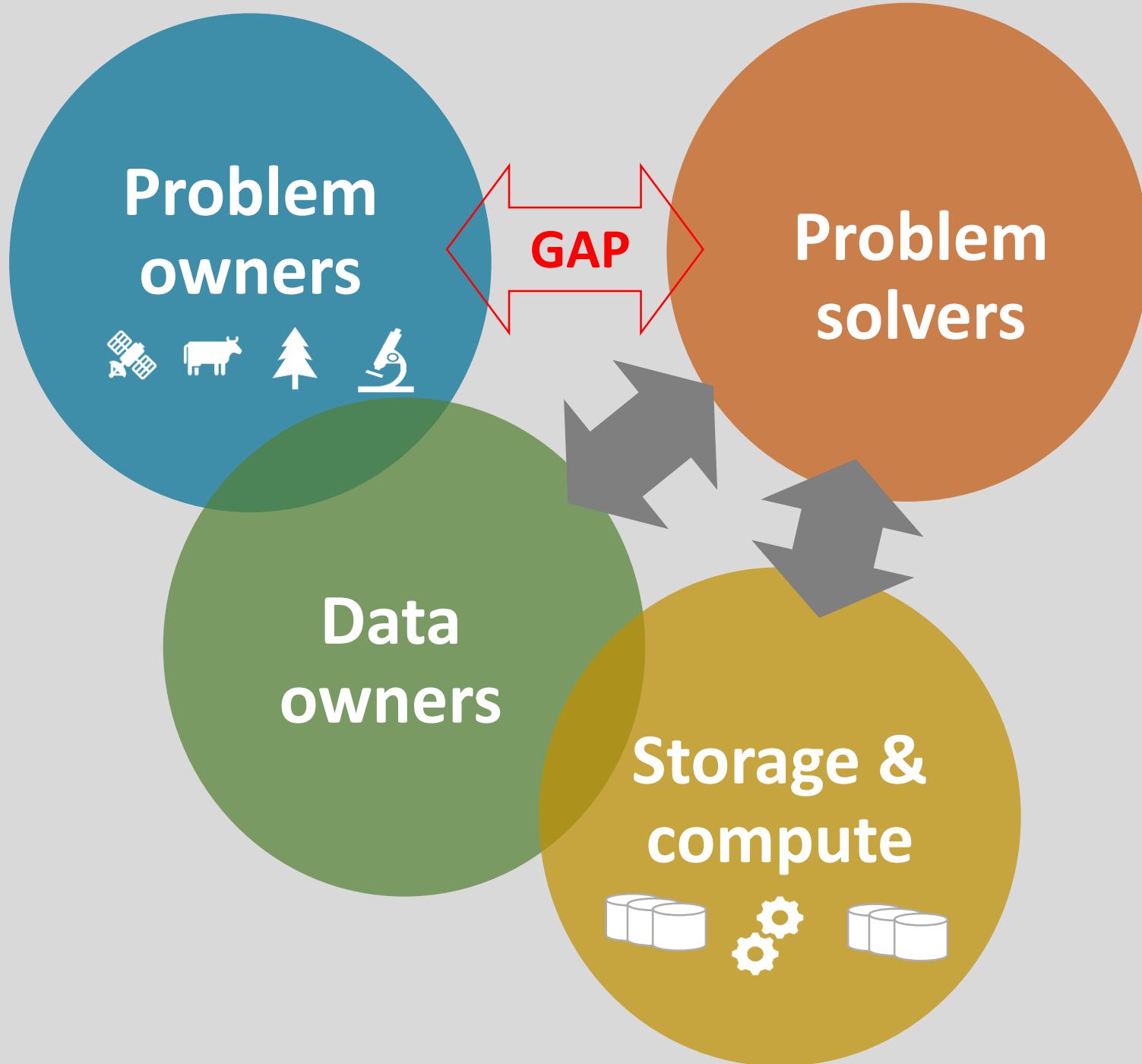
Wanna start your own TE meetup? Please do! :) You can find the details on the [Meetup.com](#) site.

The screenshot shows a Meetup.com group page for "Token Engineering Global". The page features a world map with red location pins marking various cities where local Token Engineering meetups are held. The pins are located in North America, Europe, and Asia. Below the map, the group's name is displayed as "TE Token Engineering Global" with a logo consisting of the letters "TE" in white on a dark square. To the right of the name, statistics are shown: 1,448 Members, 11 Groups, and 8 Countries. The top navigation bar of the Meetup website is visible, including links for "Start a new group", "Explore", "Messages", and "Notifications".



# Ocean, Towards a Decentralized Data & AI Stack



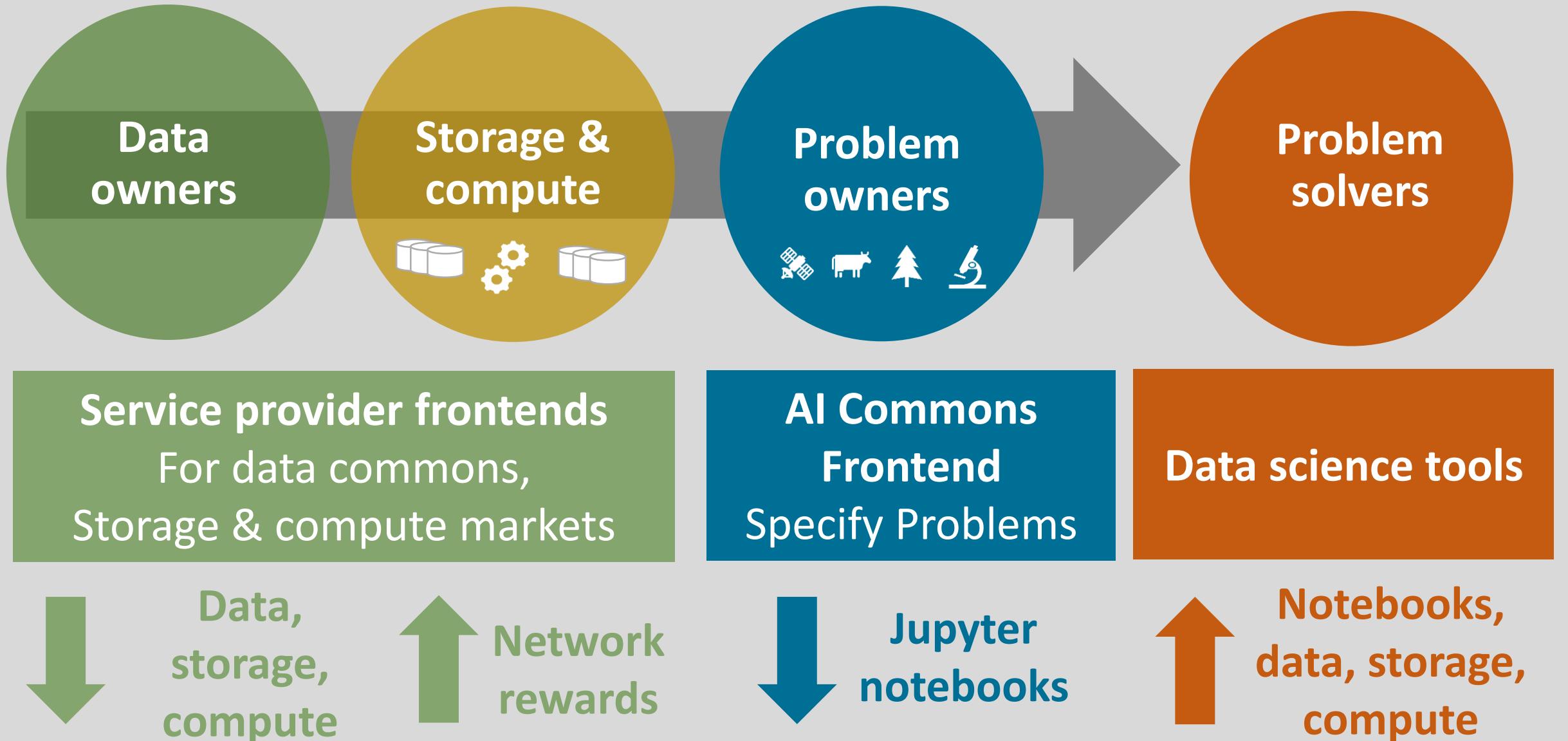




Connective tissue: public utility network



**Blockchain public utility network for the AI Commons**



**Public utility network incentivizing data:**

- Block rewards for data
- Privacy-preserving AI modeling

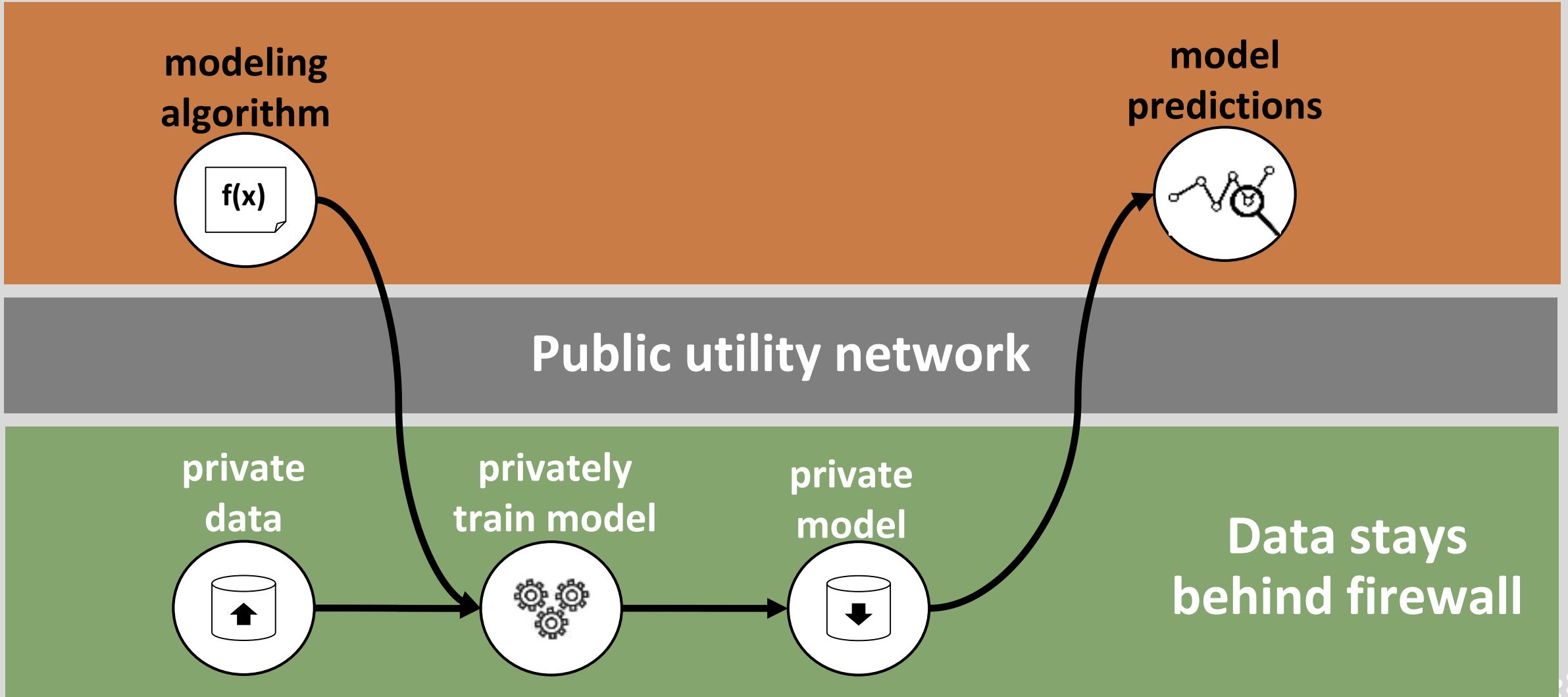
|                        |                        |                        |                        |                        |
|------------------------|------------------------|------------------------|------------------------|------------------------|
| Have data              | Have AI<br>(Want data) | Have data              | Have data              | Have AI<br>(Want data) |
| Have AI<br>(Want data) | Have data              | Have AI<br>(Want data) | Have AI<br>(Want data) | Have data              |
| Data marketplace       | DM                     | DM                     | DM                     | DM                     |
| DM                     | DM                     | DM                     | DM                     | DM                     |
| DM                     | DM                     | DM                     | DM                     | DM                     |
| DM                     | DM                     | DM                     | DM                     | DM                     |

Catalyze the commons *and* profit-making, via a substrate for a thousand marketplaces to bloom.

A close-up photograph of two hands, likely belonging to a woman, cupped together to hold a small bunch of white daisies with yellow centers. The hands are positioned in the center of the frame against a dark, textured background.

**Block rewards to incentivize data supply  
Add to data commons, make \$**

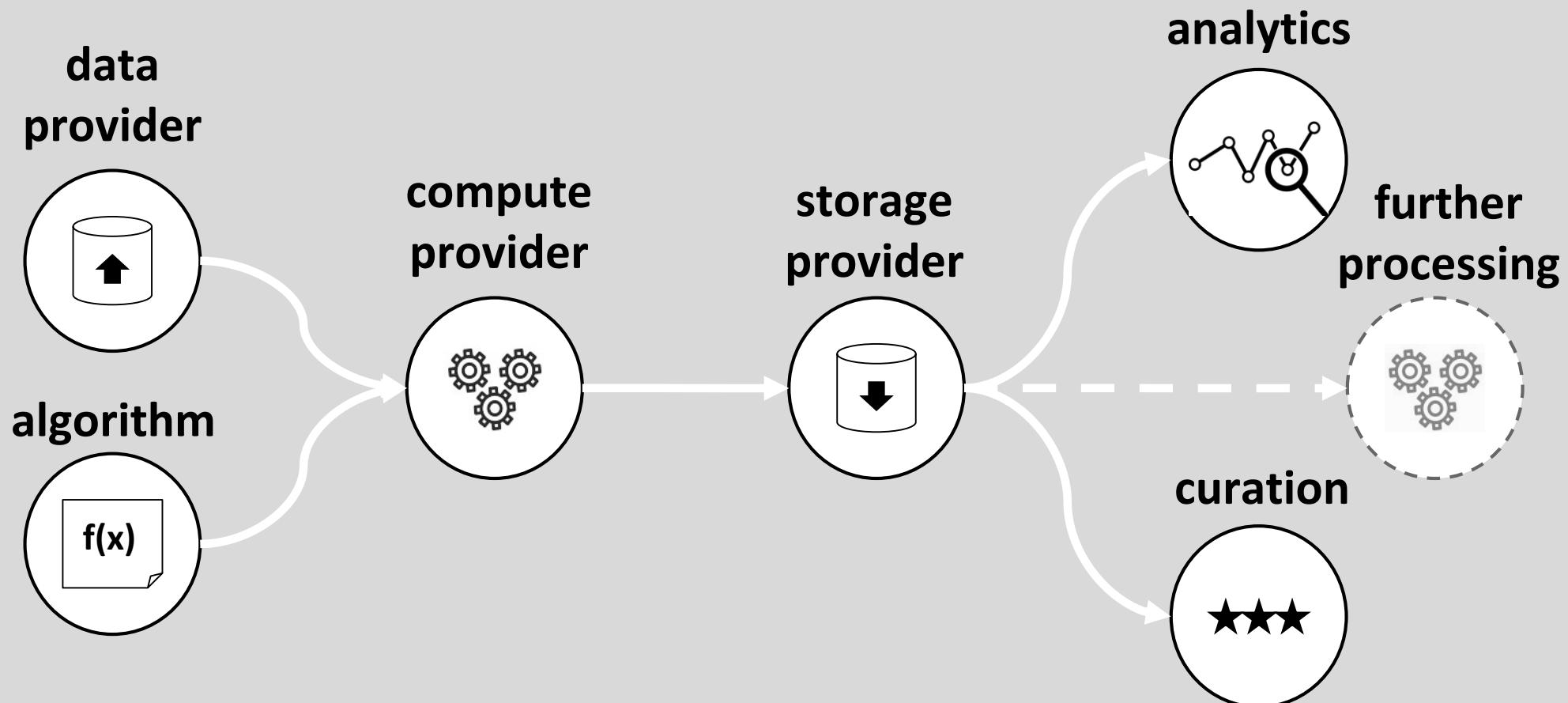
# Preserve privacy: bring AI compute to the data



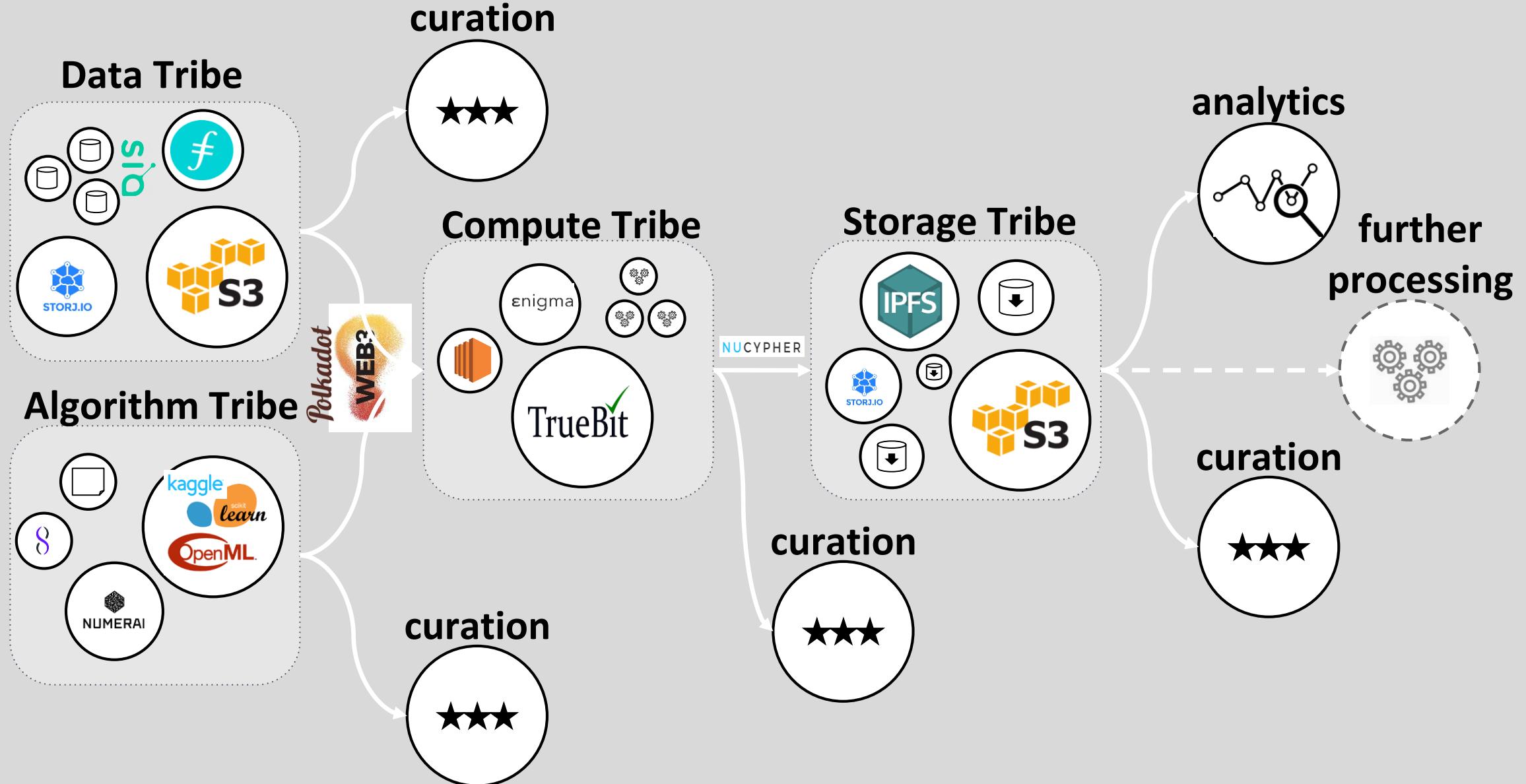
# Framings for Ocean



# “AI Compute Pipeline”

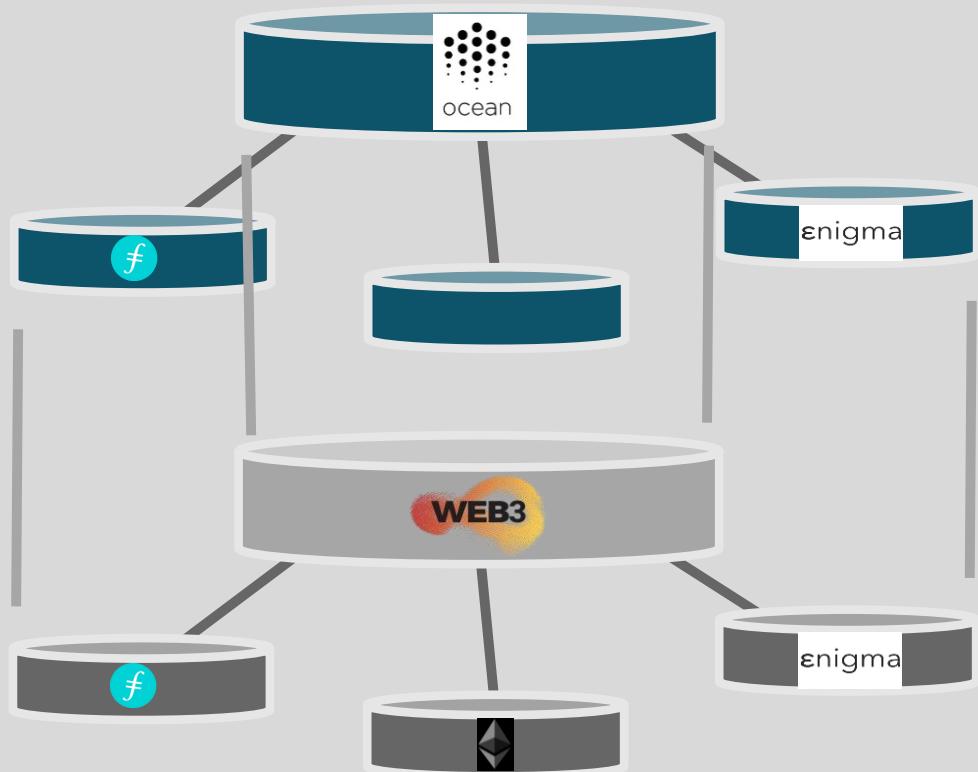


# “Inter-Service Network”

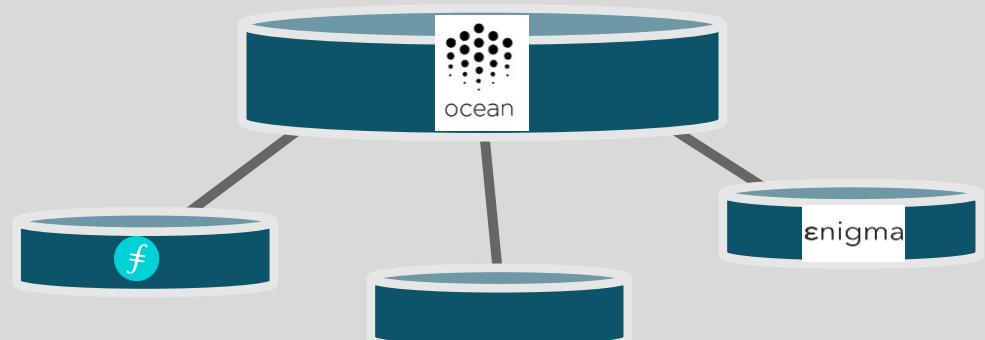


# “Network of Networks, for AI Stack”

Crypto perspective:  
Ocean is L2 on L1



AI perspective:  
Ocean is L1



# “SABRE for data”

SABRE is metadata/substrate for airline tickets. Ocean, for data itself.

## Supply tickets:

Airlines



## Consume tickets:

Consumer  
travelers, biz ppl



|              |          |                                |                 |
|--------------|----------|--------------------------------|-----------------|
| Expedia      | Kayak    | 3 <sup>rd</sup> party reseller | Easyjet website |
| Travel agent | Hip munk | BA website                     | Travel agent    |

|              |          |            |              |
|--------------|----------|------------|--------------|
| Travel agent | Hip munk | BA website | Travel agent |
|--------------|----------|------------|--------------|

SABRE

## Supply data:

Biz's, NGOs,  
Governments



|             |    |    |    |
|-------------|----|----|----|
| Data market | DM | AC | DM |
| AI common   | DM | DM | AC |

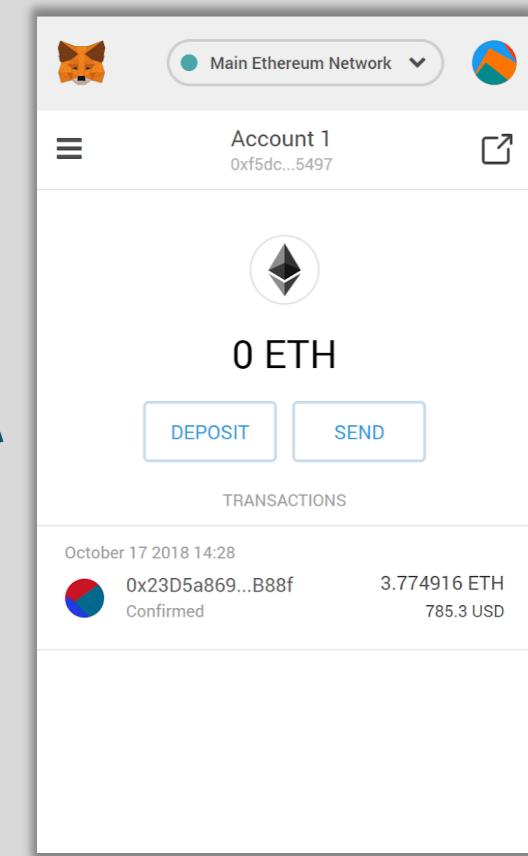
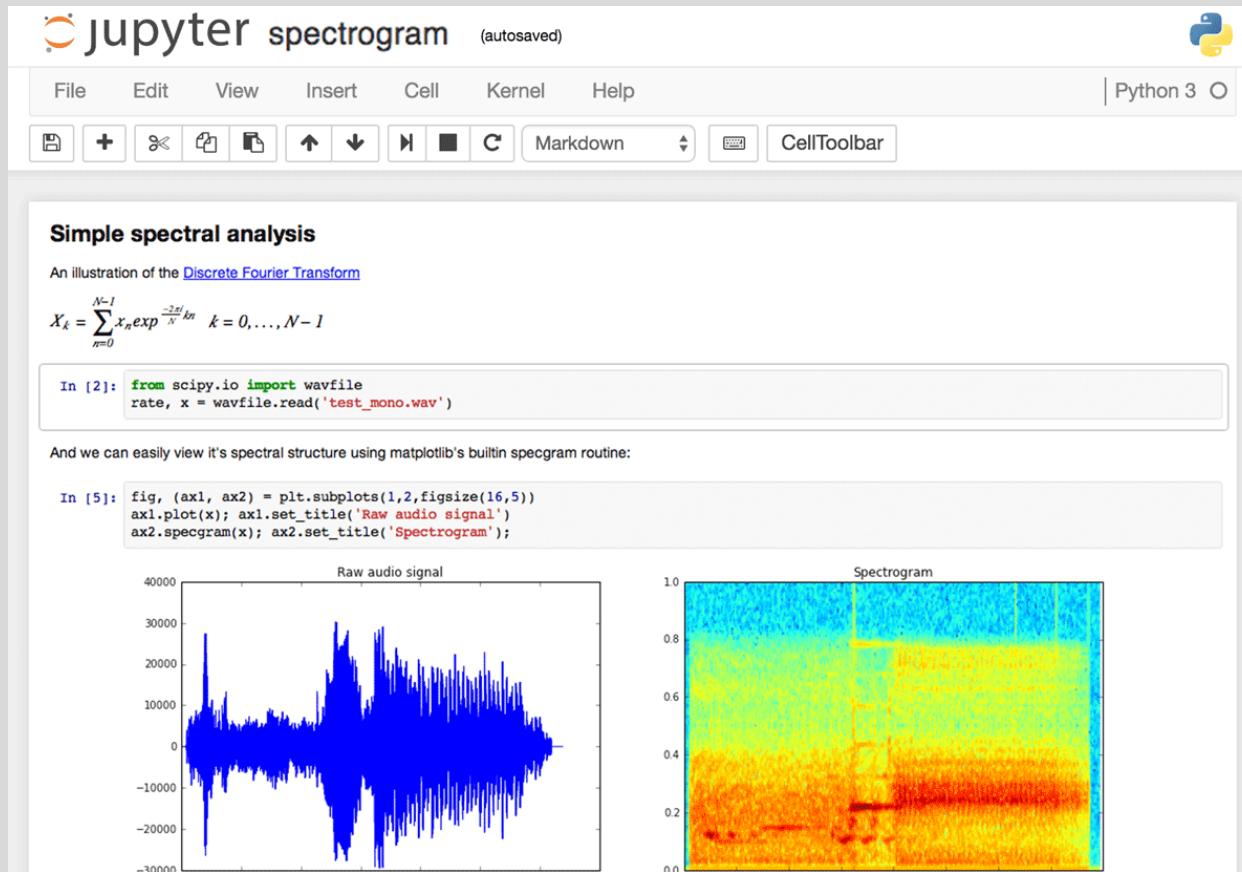


Public utility network



What does  
decentralized AI look like?

# Keep Data Science Tooling, Add Crypto



# Every dataset is an asset.

## Every Jupyter notebook is an asset (!)

Ocean ▶ KillTheHumans

5 Datasets

View

Login

Marketplace  
KillTheHumans

KillTheHumans is a marketplace to buy and sell datasets. In order to eradicate piracy, the publisher's name and data must be input correctly. Version 2.8.

1 Dataset 78XcYT9 Laser Target Practice Exercises

Publisher Name: Lazerdrone Ltd.  
Date Published: 03-08-2019

LAZ

2 Dataset 87ertHG Facial Recognition Data

Publisher Name: VoyeurDrone Inc.  
Date Published: 20-06-2018

VOY

3 Dataset 000x567 Thermal Sensor Data

Publisher Name: Predator Movement  
Date Published: 31-09-2022

KTH

4 Dataset 96zERC Human Locations based on IP

Publisher Name: SkyNet  
Date Published: 09-11-2020

KLO

5 Dataset 254phyY Human Biological Weaknesses

Publisher Name: The Overmind  
Date Published: 28-06-2016

REX

6 Add your dataset here ...

... and become part of the ecosystem

ocean

7845.12 OCN

11 Datasets

1 OCN = 0.12 ETH

Dominator

My Token Assets

|     |        |     |         |
|-----|--------|-----|---------|
| NTT | 102.10 | KIO | 55.00   |
| RED | 0.45   | OSC | 0.80    |
| INT | 45.38  | IOX | 1.00    |
| ZAP | 10.00  | BAR | 4523.12 |
| NTT | 100.00 | URD | 200.00  |
| AAC | 22.45  |     |         |

Dataset Selection and Token Supply vs Cost

| Dataset | Marketplace | #Name | Ranking   |
|---------|-------------|-------|-----------|
| 78XcYT9 | Marketplace | #Name | Ranking 7 |

+4.5%

Current: 9.20 OCN, 2.99k AAC

Future: 9.25 OCN, 3.04k AAC

+5.80 | Invest (+) or Withdraw (-)

Confirm Transaction



ocean

# New powers for data scientists

- **Way more data.** Data commons. Enterprise data without data escapes via on-premise compute.
- **Provenance** in data & AI training. Goodbye data honeypots.
- **More \$.** For generating data. For cleaning, labeling, feature engineering data. For algs. For curation.





<https://docs.oceanprotocol.com>

Ocean Protocol Documentation - [+](#)

[←](#) [→](#) [C](#) [🔒 https://docs.oceanprotocol.com](#)

 ocean

# Ocean Protocol Documentation

Learn about the components of the Ocean Protocol software stack, and how to run or use the components relevant to you.



**Core Concepts**  
Understand the fundamentals of Ocean Protocol.  
[Learn More →](#)

**Setup Guides**  
Setting up the Ocean Protocol components.  
[Learn More →](#)

**Tutorials**  
Browse tutorials for most common setup and development use-cases.  
[Learn More →](#)

 ocean



The **Manta Ray** notebooks provide a guided tour of Ocean Protocol in an interactive Jupyter Notebook environment. Start using Ocean Protocol with your own pre-configured and loaded cloud instance after login with your GitHub account.

This project is in alpha! Feel free to ask questions and post bug reports in [our Gitter channel](#). Notebooks are for tutorial and demonstration purpose only. Notebook instances may be periodically offline, and storage volumes will be purged.

JupyterLab Instance

Ocean Protocol Docu | Tools & Examples - C | Quickstart - Ocean P | Jupyter Notebooks - Manta Ray — Ocean | JupyterLab

Not secure | a1e6efedde7f611e8887b1225ec03c49-1029831930.us-east-1.elb.amazonaws.com/user/trentmc... ☆

File Edit View Run Kernel Hub Tabs Settings Help

Launcher

Name Last Modified

mantaray.jupyter 2 minutes ago

Notebook

Python 3

Console

Python 3

Other

\$\_-

=

ocean

## QUICKSTART

## Run & Try Everything

## MARKETPLACES

## Set Up a Marketplace

PUBLISHERS

## Publish Data or Services

KEEPERS

### Run a Keeper

```
git clone https://github.com/oceanprotocol/barge.git  
cd barge/  
  
. ./start_ocean.sh --latest
```

Seeing the dolphin means it's working:





# Conclusion

# Conclusion: On Decentralizing AI Data & Services

- Goal: spread benefits of AI, by equalizing opportunities to access data
- Token design  $\approx$  EA design.
  - So, approach token design as EA design!
  - Helped a *lot* for designing Ocean. #TokenEngineering.
- Decentralized AI gives data scientists new powers
  - *Way* more data. Privacy-preserving compute.