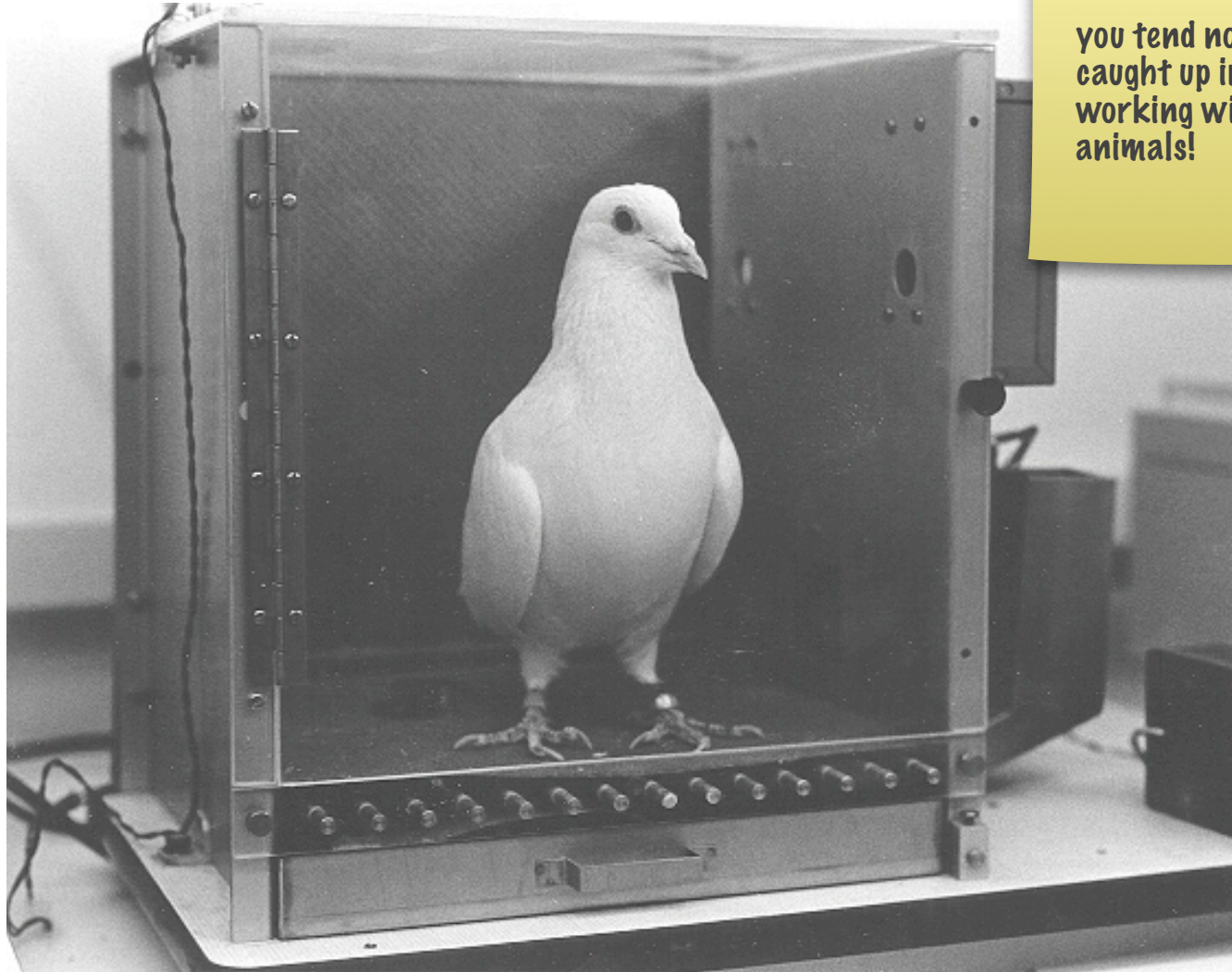


# The Stone Soup of Data

Jamie Taylor - Metaweb

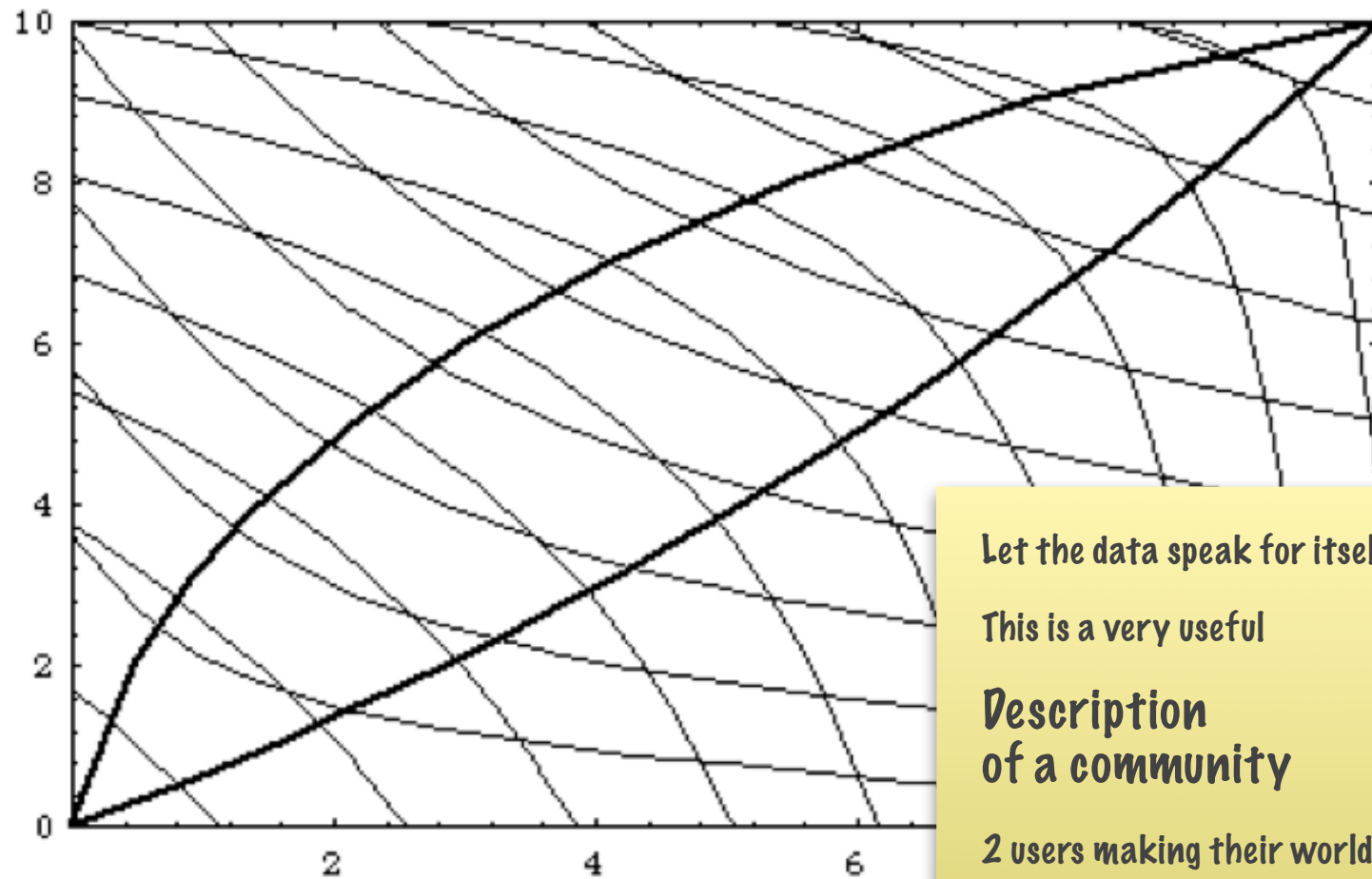
WWW2007





eschew mentalistic explanations

you tend not to get caught up in them when working with preverbal animals!



Let the data speak for itself!

This is a very useful

**Description  
of a community**

2 users making their world better!

Very different than what you hear  
people saying when describing on-line  
communities.



Story of Stone  
Soup is told here.

# What is the story about?

Multiplying a scarce resource through community

This is an economic tale

community was altered

the rules of exchange were  
manipulated

Changed the Economy

community = economy

--Everyone better off--

# So far...

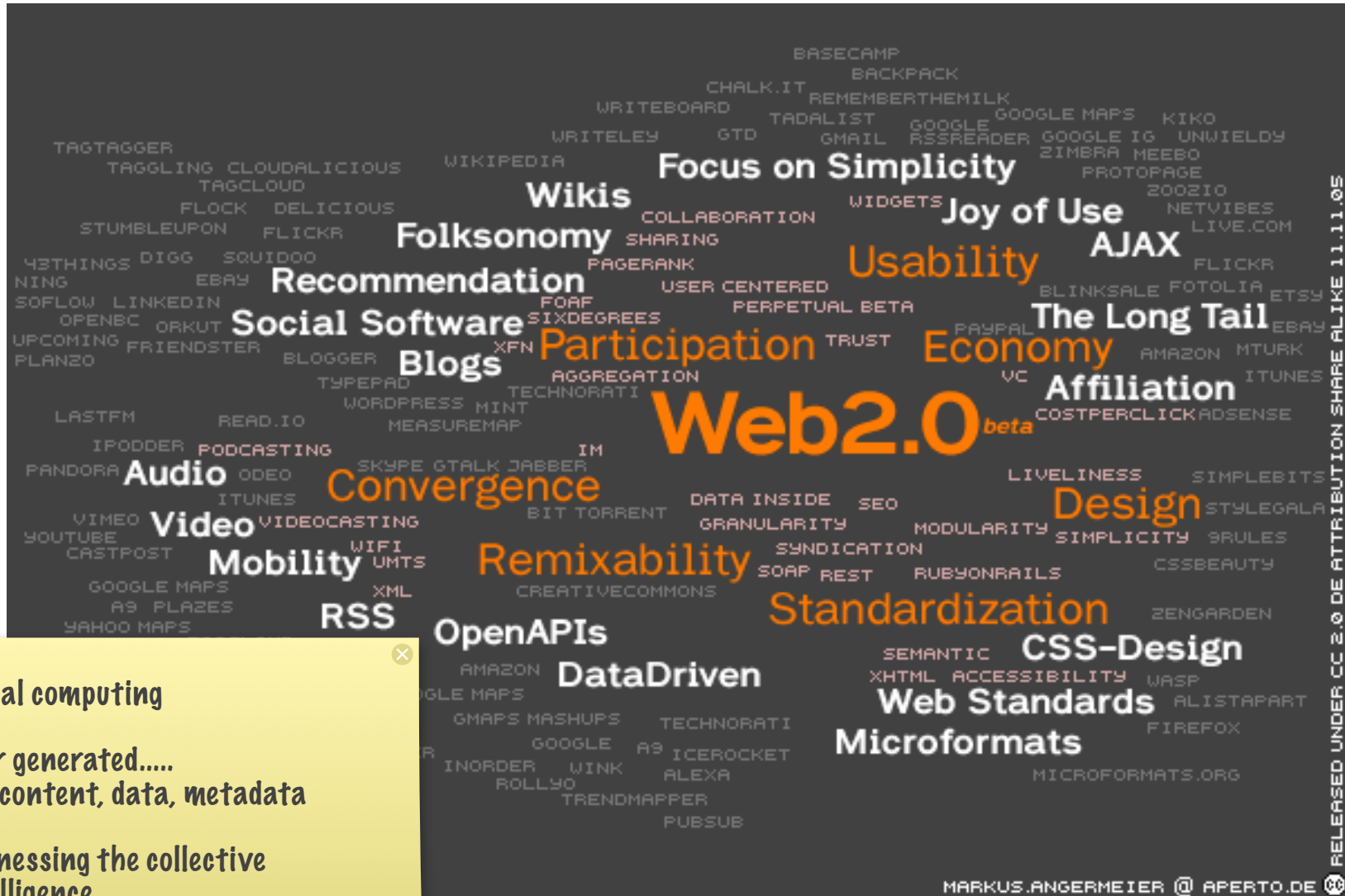
- Jamie is an economist in disguise
- Stone Soup is an economic tale
- Community is another word for economy

**To understand a community...**

...follow the flow of value.





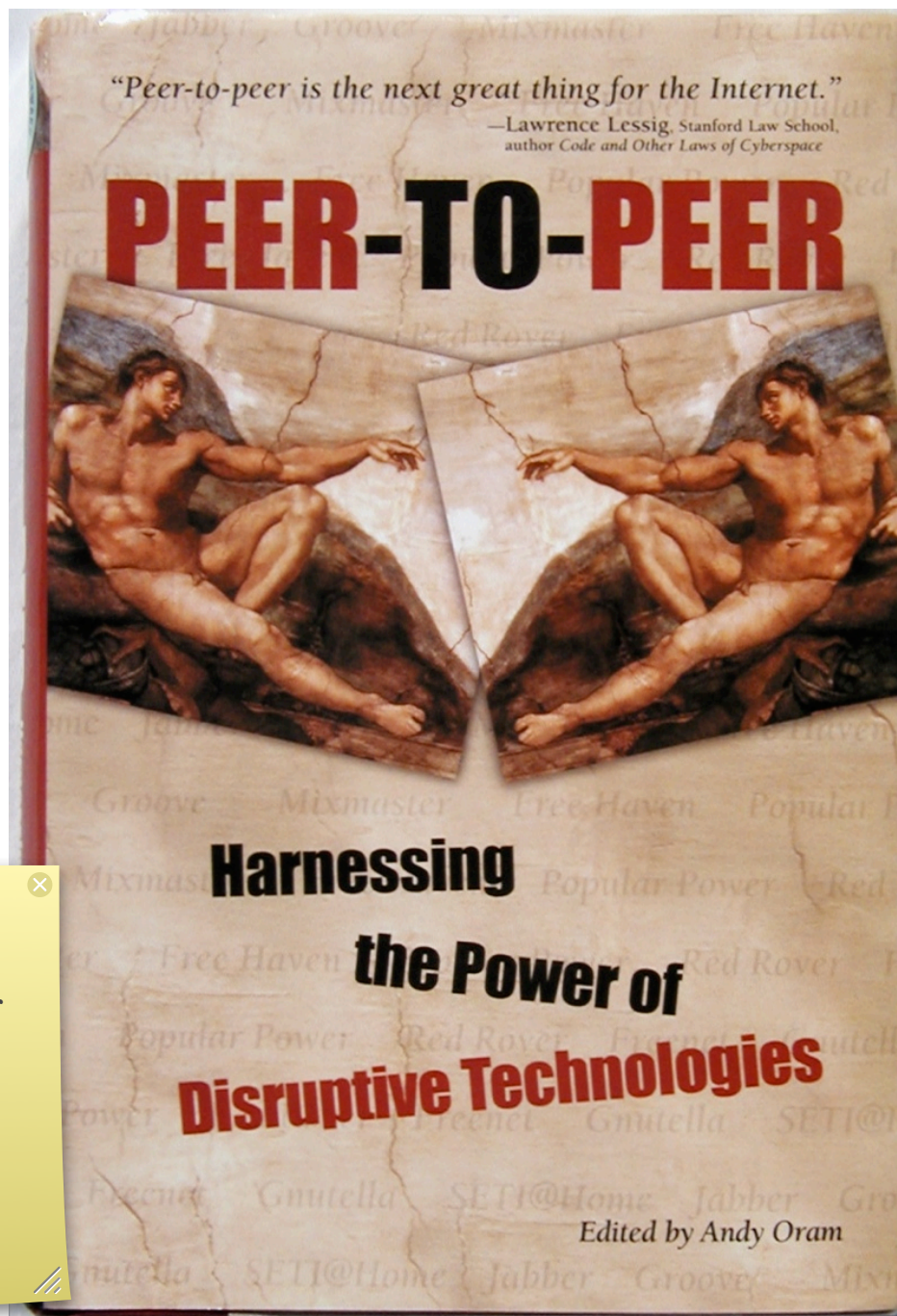


**Social computing**

**User generated....  
content, data, metadata**

**Harnessing the collective  
intelligence...**

**web 2.0 is about community**



“Peer-to-peer is the next great thing for the Internet.”

—Lawrence Lessig, Stanford Law School,  
author *Code and Other Laws of Cyberspace*

# PEER-TO-PEER



**Harnessing**

**the Power of**

**Disruptive Technologies**

Edited by Andy Oram




Not out of Date!

Technologies aren't hip, but message is about community.

Perfectly accurate today!

Read it again!

# Dan Bricklin: Cornucopia of the Commons

- Three ways to build a database
  - Organize manually 
  - Organize mechanically 
  - Get volunteers to build it 

Interesting examples

CDDB.....

Take community output  
and make it private.

THE  
WEALTH  
OF NATIONS  
ADAM SMITH

"I have never known much good done by those who affected to trade for the public good."

-Adam Smith, 1776

is clear: concentrate on what you can get from users, and use whatever protocol can maximize their voluntary contributions. That seems to be where the greatest promise lies for the new kinds of collaborative environments.


make DB Seeding compatible w/  
"Normal" User Activities & Profit from  
the Externalities... Seeding is A byproduct

It only took me  
half a decade  
to find people  
to do this with!

# Architecture of Participation

Open Source communities are “free markets for ideas”

-Tim O'Reilly



connection between  
Community and Economy

A yellow sticky note with a close button in the top right corner and a pencil icon in the bottom right corner. The text on the note is in a bold, sans-serif font.

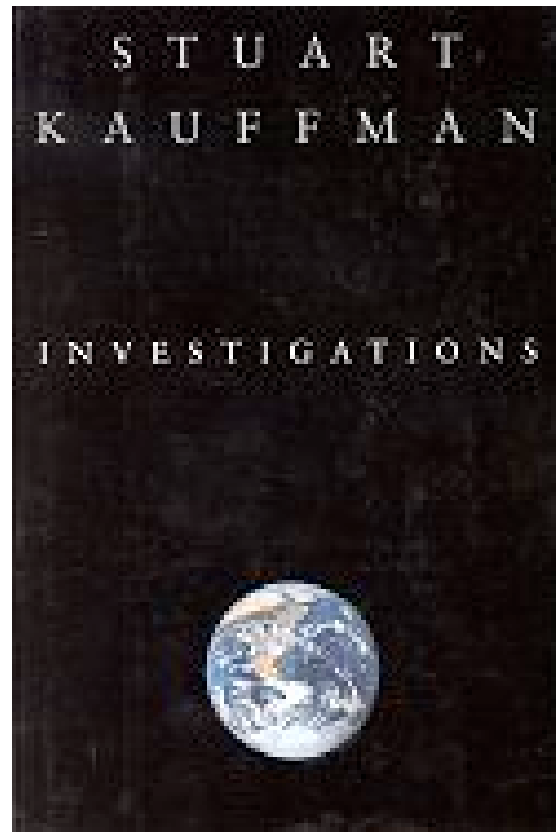
Tim cites this as an  
influence on his

Architecture of  
participation

# Architecture is Politics

-Mitch Kapor





**Taylor/Jordan 2000 work:**

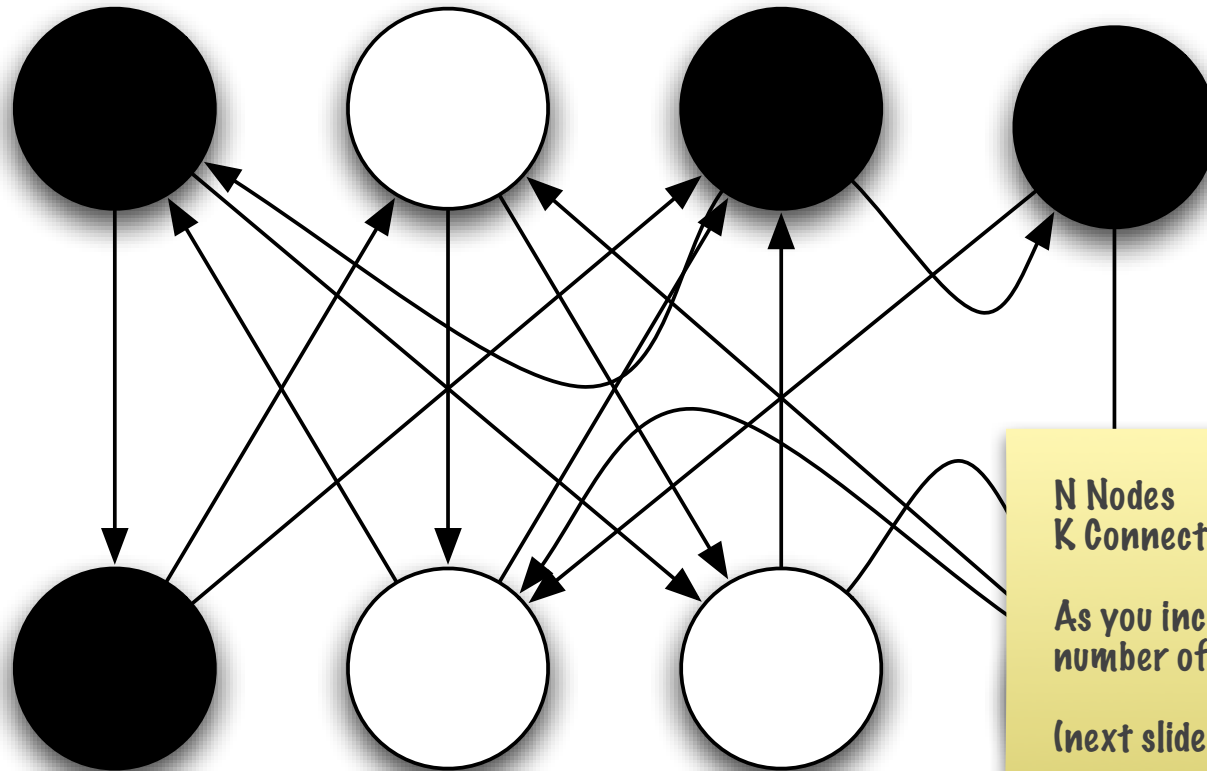
**Multiple NK Models,  
operating at different  
scales, describes why Linux  
didn't suffer the**

**Mythical Man Month**

**One of Stuart's many  
interesting works  
showing how**

**abstract models can shed  
light on basic phenomena**

# NK Boolean Networks



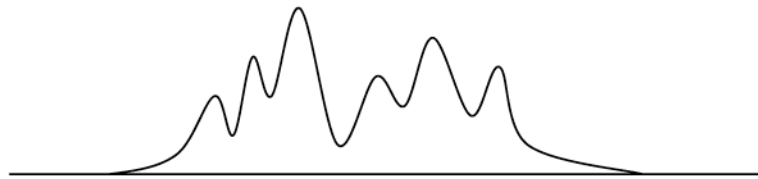
N Nodes  
K Connections

As you increase the  
number of connections...

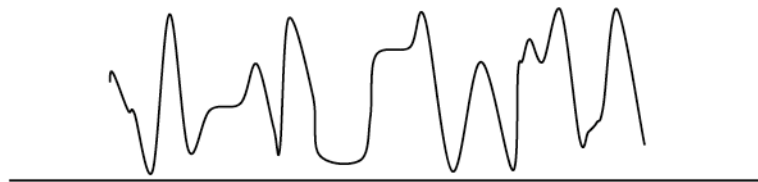
(next slide)



single-peaked



moderately  
rugged,  
correlated



randomly  
multi-peaked,  
uncorrelated

Creates tunable  
landscapes for  
optimization  
problems

"I couldn't do what I did with Linux for Windows, even if I had the source code. The architecture just wouldn't support it."

- Linus Torvalds as quoted by Tim O'Reilly

“Too much of the windows source code consists of interdependent, tightly coupled layers for a single developer to drop in a replacement module.”

-Tim O'Reilly

When designing systems for community participation you must choose a good political ~~architectural~~ model.

# Participatory Communities

Visitors  
Contributing



0.16%



0.2%



4.56%

The WP  
figure is not  
without  
controversy

Source: Hitwise, reported by Bill Tancer 4/25/2007 in Time Magazine

Web 2.0 has been successful in significantly broadening the amount material available to us, but reviewing the latest data reveals that we're still in the very early stages.

Suggesting.....

He thinks the relative % of visitors "participating" will grow!

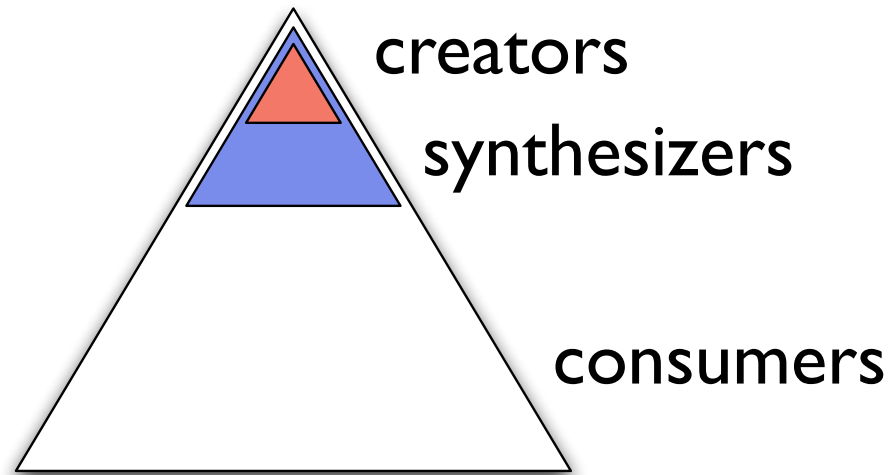
Popular Press....this is the impression people will walk away with.

(People will think we failed...since its unlikely to change!)

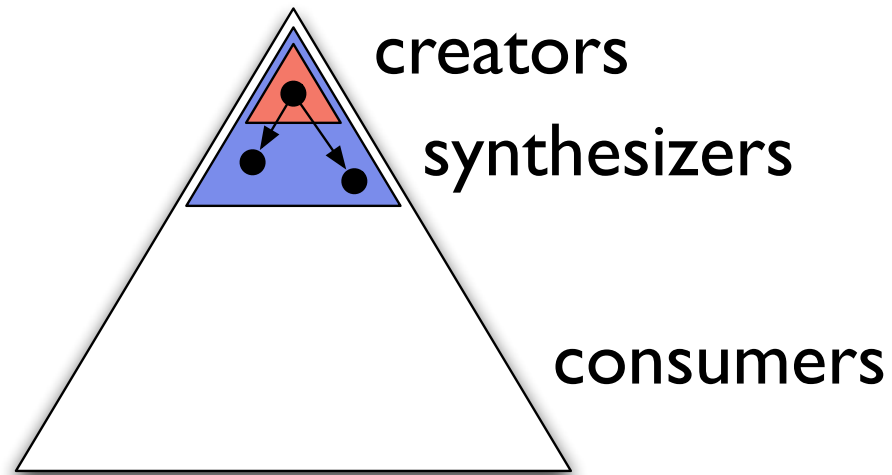
-Bill Tancer, Time Magazine, 4/25/2007



# Bradley Horowitz: Scalable Communities



# Bradley Horowitz: Scalable Communities



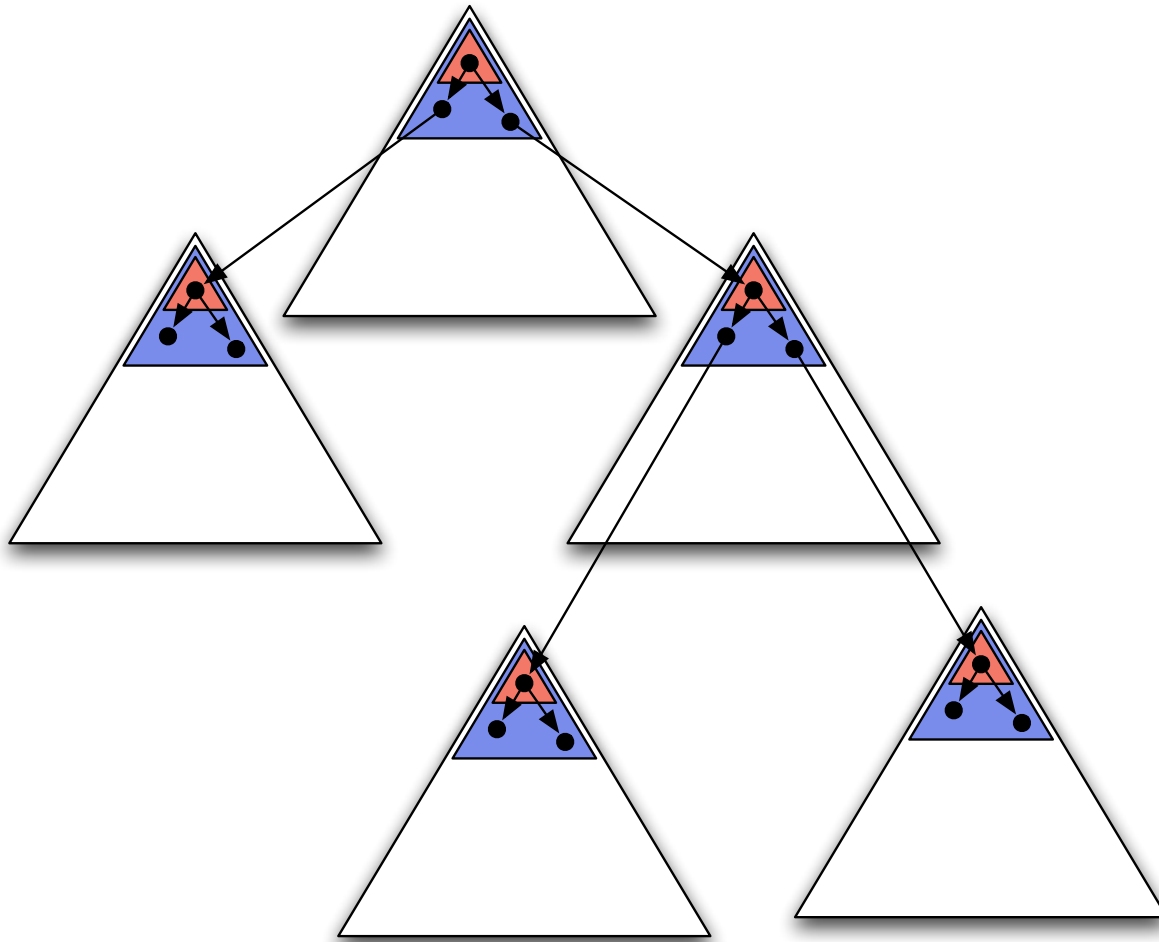
# Aaron Swartz

- Measure volume of content contributed
- Large number of users contributing large amounts of data, a low number of times

Both are correct!

Contributions at many different scales.

# Community at All Scales



# Econ 101 Refresher

- Agents have diminishing marginal returns
- Economic actions produce negative feedback
- Nice textbook equilibria
- We all live happily ever after



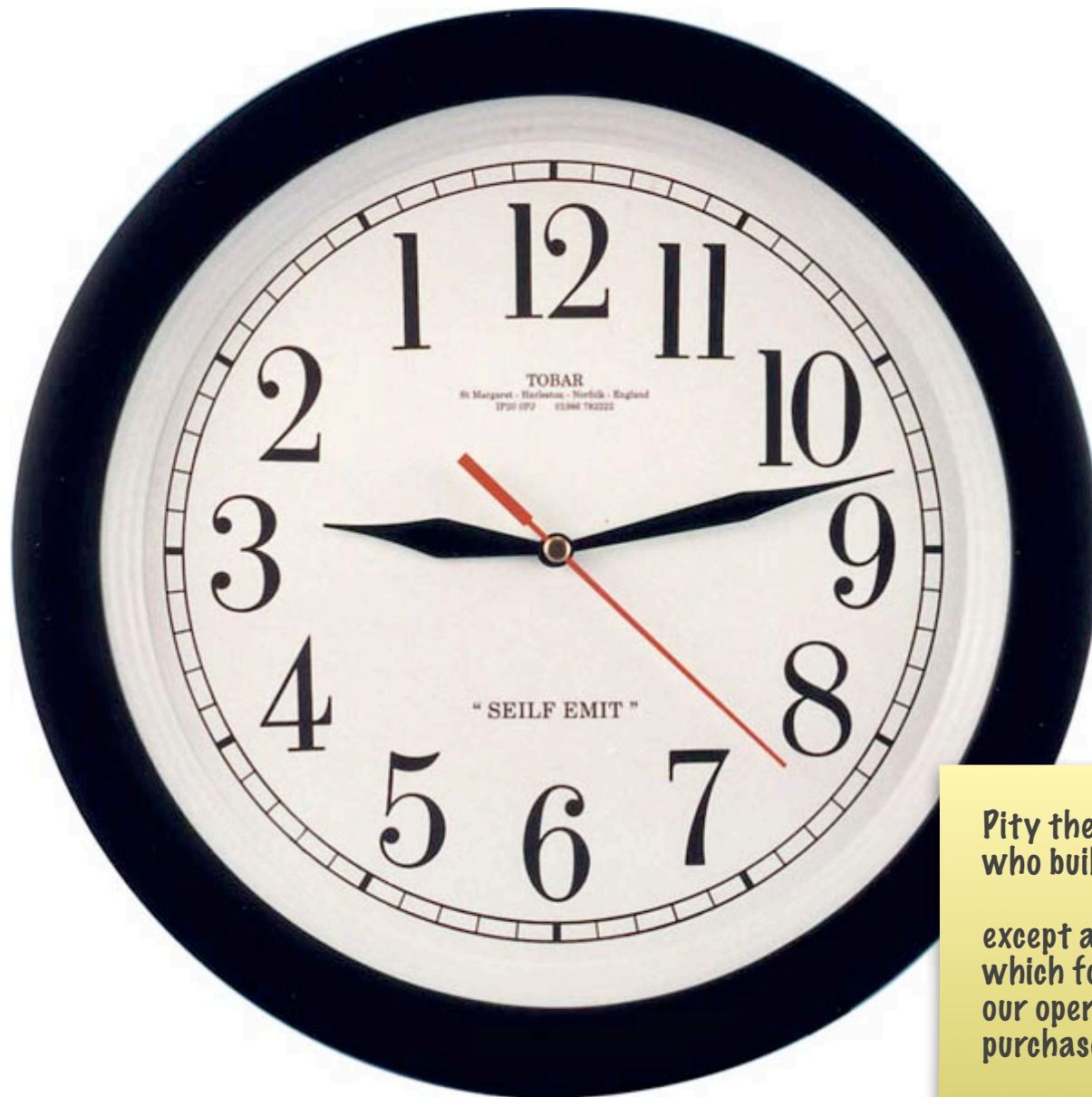
Contemplate your typical  
clock

We see them  
everywhere.

Didn't have to be this way.

Many early clocks were 24 hour clocks.

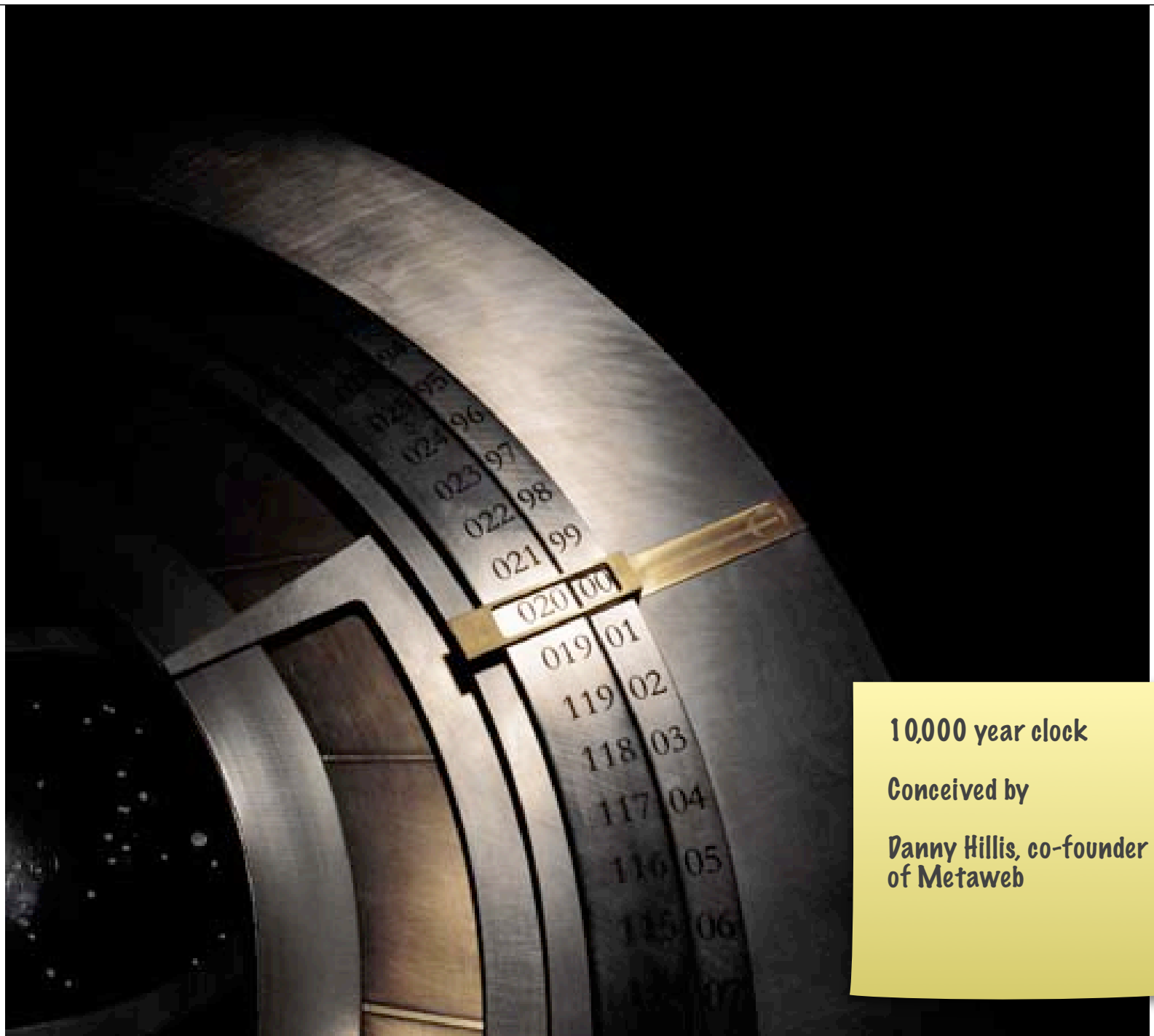




Pity the clock maker  
who builds these today

except as novelty items,  
which for some reason  
our operations team  
purchases





10,000 year clock

Conceived by

Danny Hillis, co-founder  
of Metaweb

# Brian Arthur

- Positive feedback amplifies economic forces
- Multiple equilibria
- Initial conditions can determine outcome
- The “optimal” outcome isn’t always found
- Homo Economicus might be dead

# Duncan Watts

## Cumulative Advantage: The Experiment

Network effects as an  
explanation for

VHS/Betamax

Largely Untestable....

...experiment published in  
Science

# Stone Soup Review

- Cumulative advantage
- Architecture for participation
- Community at all scales

**Cumulative Advantage:**  
-once people started to contribute....easier for others to contribute....positive feedback.

**Architecture for Participation:**  
-loosely coupled. If one contribution failed, others could succeed.

**Community at all scales:**  
-Different type of contribution  
-Individual contributions, family

# Stone Soup Effect

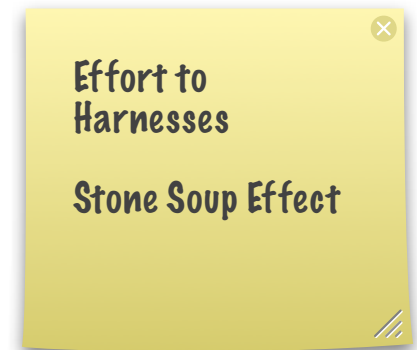
- Cumulative advantage
- Architecture for participation
- Community at all scales

Multiply a scarce resource through community

# freebase™



POWERED BY  
METAWEB™



# Freebase is a database

- Not a set of documents
- Has fine-grained structure
- Lets you ask unexpected questions
- It's machine readable

# Freebase is practical

- Solves problems people currently have
- Not an ontology or inferencing engine
- Optimized for real-time execution



# Freebase is universal

- Works across all languages
- Thousands of databases combined
- Everything is reconciled

## Bodybuilding



Bodybuilding awards	Mr Olympia 1970
	Mr Olympia 1971
	Mr Olympia 1972
	Mr Olympia 1973
	Mr Olympia 1974
Admitted steroid use	Jawohl!

## Movie database



Film performances	Hercules in New York
	Conan the Barbarian
	Predator
	Kindergarten Cop
	Terminator
	True Lies
	Jingle All the Way
	The Last Action Hero
	Total Recall

## Politics database



Political party	Republican
Offices held	Governor - California
Political Mentor	Sargent Shriver

## Arnold Schwarzenegger ▾

Discuss "Arnold Schwarzenegger"  Hide Empty Fields



◀ image 1 of 1 ▶

⌘ **Types:** [Person \(People\)](#), [Film Actor \(Film\)](#), [Film Producer \(Film\)](#), [US Politician \(Government\)](#), [Athlete \(danny's types\)](#), [Interesting People \(rubylabs's types\)](#), [Herkunft \(carlo\\_velten's types\)](#)

⌘ **Also known as** [Arnold Alois Schwarzenegger](#), [The Governorator](#)

⌘ **Gender** [Male](#)

⌘ **Date of Birth** [Jul 30, 1947](#)

⌘ **Place of Birth** [Thal, Austria](#)

⌘ **Country Of Nationality** [United States](#)

⌘ **Profession** [Actor](#), [Politician](#), [Bodybuilder](#)

⌘ **Spouse(s)** [Maria Shriver](#) - [Apr 26, 1986](#)

⌘ **Parents** [Aurelia Jadorny Schwarzenegger](#), [Gustav Schwarzenegger](#)

⌘ **Children** [Christopher Schwarzenegger](#), [Patrick Schwarzenegger](#), [Christina Schwarzenegger](#), [Katherine Schwarzenegger](#)

⌘ **Sibling(s)** [Meinhard Schwarzenegger](#)

⌘ **Height (meters)** [1.88](#)

⌘ **Weight (kg)** *double-click to add "Weight (kg)"*

⌘ **IMDB Entry** <http://www.imdb.com...>

⌘ **Party** [Republican Party](#)

⌘ **Sport** [Powerlifting](#), [Bodybuilding](#)

### 🕒 Page History

**Created by** [Metaweb](#) [Oct 22, 2006](#) 1:49am

**Last edited by** [carlo\\_velten](#) 4 days ago at 6:46am

### ⌘ Web Link(s)

<http://www.schwarzenegger.com>

### ⌘ Employment history

[Flex Magazine](#) - executive editor

[Planet Hollywood](#) - Co-founder

### ⌘ Education

*double-click to add "Education"*

### ⌘ Film performances

[Twins](#) - Julius

[True Lies](#) - Harry Trasker

[Total Recall](#) - Quaid

[The Terminator](#) - Cyborg t800

[The Running Man](#)

[The 6th Day](#)

[Terminator 2: Judgment Day](#)

# It is all about UI

...or almost all about UI

When launching we discovered...not surprisingly

UI reveals the platform

- graph store isn't visible
- Service/Query tier isn't visible (to most)

UI makes or breaks  
All about Utility

# It is all about Client UI Features

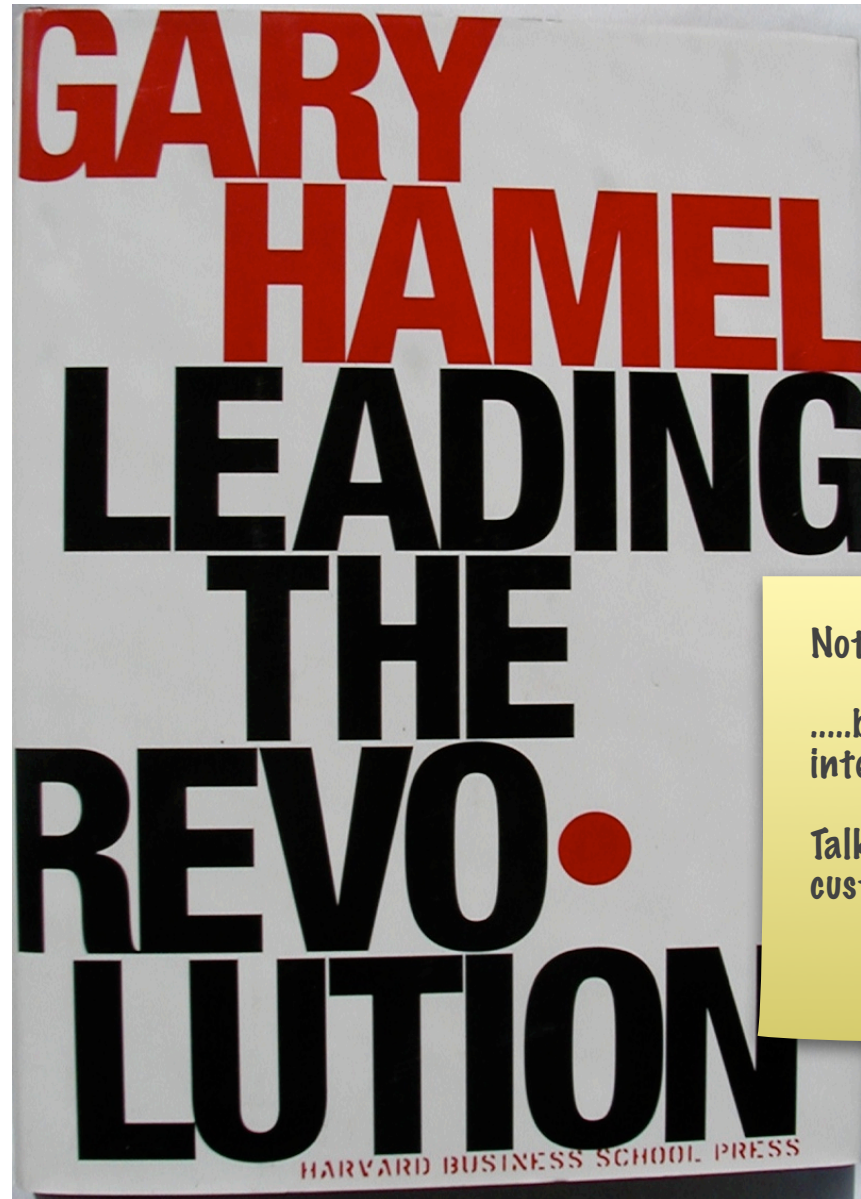
...or almost all about Client UI Features

People *\*really\** like our UI

People are making big contributions through it.

but we can/will do more to improve it...think of it like an experiment.  
(Like a garden!)

If community/data products are to succeed, we all need to focus on usability.



Not particularly scholarly

.....but useful and interesting.

Talks about serving your customers (aka community)

# What have loyalty programs shown us?

- Typical supermarket
  - Top quartile of customers spend 50x more per year than bottom quartile customers
  - Why then do we have express lanes?

# “Stones for Creators”

## ● Documentation

- Targets high value members of the community
- Invest appropriately





## Film

Compound Value Type

Type Key: [film](#) [edit](#)

Included types *add a parent type to include the properties of another type in all Film topics.*

### User Created Properties

[Add a New Property](#)



▶ [Initial release date](#) [edit](#) [Date/Time](#)

▶ [Directed by](#) [edit](#) [Film Director](#) *(links in as Films directed)*

▼ [Starring](#) [edit](#)

Property Key: [starring](#) [edit](#)

#### Expected Type

[Performance](#) [edit](#) *(links in as film)*

#### Advanced Options

Do not display in default template

#### Default orientation

Vertical List  Horizontal List

Restrict to one value

Display as disambiguator

Display as enumerated list

[Delete Property](#)

▶ [Produced by](#) [edit](#) [Film Producer](#) *(links in as Films Produced)*

### Topics

[Dudes](#)

[Vägen ut](#)

[What's Cooking?](#)

[Twin Peaks](#)

[The Simpsons](#)

[24](#)

[The Best of Everything](#)

[The Friend Who Walked the West](#)

[The Sun Also Rises](#)

[Barefoot in the Park](#)

[The Langoliers](#)

[Laurel Canyon](#)

[The French Connection](#)

**High Value  
Creators:**

**A community we  
are still working  
on servicing.**

# Responsiveness is critical

canonicalization

Speed is critical

Disambiguation

1) fly out

2) relevance

The screenshot shows a user profile on the Freebase website. The profile is for a user named 'jamie' who is a staff member. The profile includes a photo and several fields for user information: 'My Full Name', 'Birthdate', and 'Where I Live'. The 'Where I Live' field is active, showing a search dropdown with 'Fairfax' entered. The dropdown lists various locations including Fairfax County, Fairfax, Virginia, Fairfax, California, Fairfax, Iowa, Fairfax, Minnesota, Fairfax, Missouri, Fairfax, Ohio, Fairfax, Oklahoma, Fairfax, South Carol..., Fairfax, South Dakot..., and a 'Create New Location' option. A popup window is displayed for 'Fairfax, California', showing a map of California and text describing the town's location, population (7,500), and area (5.5 km²).

**freebase™**  
alpha

Home My Profile Types Developers Help

**jamie** STAFF

My Full Name double-click to add "My Full Name"

Birthdate double-click to add "Birthdate"

Where I Live

Fairfax|

Fairfax County Location, US County

Fairfax, Virginia Location, City/Town, US C...

Fairfax, California Location, City/Town, US C...

Fairfax, Iowa Location, City/Town, US C...

Fairfax, Minnesota Location, City/Town, US C...

Fairfax, Missouri Location, City/Town, US C...

Fairfax, Ohio Location, City/Town, US C...

Fairfax, Oklahoma Location, City/Town, US C...

Fairfax, South Carol... Location, City/Town, US C...

Fairfax, South Dakot... Location, City/Town, US C...

Create New Location

**Fairfax, California**

Location, City/Town, US City/Town

Fairfax is a town in Marin County, California, United States. The population is 7,500 at the 2006 census. Fairfax is located at 37°59' 10" N, 122°35' 34" W (37.986023, -122.592823)GR1. According to the United States Census Bureau, the town has a total area of 5.5 km² (2.1 mi²), all land. As of the censusGR2...

**About Me**

I did time in [Cambridge](#) the result of a search for a job in San Francisco reached me and I ended up in Cambridge.

I blame this turn of events on a Fourth Grade teacher who taught me [BASIC](#) on a [CDC Cyber 70](#) (I am a Minnesotan through and through.) In college, my affliction became worse when I learned assembly on a [VAX 11/750](#). And my destiny sealed, when, in graduate school, I had my own [PDP-8](#) (it was the '90s you know) and I was compelled to write lots of [SKED](#).

# Small changes have a huge impact on behavior

## USER PROFILE

jamie STAFF



image 1 of 1

- My Full Name *double-click to add "My Full Name"*
- Birthdate *double-click to add "Birthdate"*
- Where I Live

Fairfax|

Fairfax County Location, US County

Fairfax, Virginia Location, City/Town, US C...

Fairfax, California Location, City/Town, US C...

Fairfax, Iowa Location, City/Town, US C...

Fairfax, Minnesota Location, City/Town, US C...

Fairfax, Missouri Location, City/Town, US C...

Fairfax, Ohio Location, City/Town, US C...

Fairfax, Oklahoma Location, City/Town, US C...

Fairfax, South Carol... Location, City/Town, US C...

Fairfax, South Dakot... Location, City/Town, US C...

[Create New Location](#)

### Fairfax, California

Location, City/Town, US City/Town

Fairfax is a town in Marin County, California, United States. The population is 7,500 at the 2006 census. Fairfax is located at 37°59' 10" N, 122°35' 34" W (37.986023, -122.592823)GR1. According to the United States Census Bureau, the town has a total area of 5.5 km² (2.1 mi²), all land. As of the censusGR2...

Small changes - big impact

an edit this page.  
ur account & invitations

### About Me

I did time in Cambridge the resu  
Francisco reached me and I ent

I blame this turn of events on a Fourth Grade teacher who taught me BASIC on a CDC Cyber 70 (I am a Minnesotan through and through.) In college, my affliction became worse when I learned assembly on a VAX 11/750. And my destiny sealed, when, in graduate school, I had my own PDP-8 (it was the '90s you know) and I was compelled to write lots of SKED.

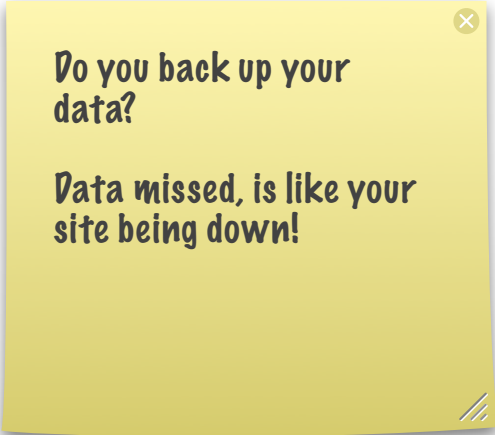
# “Data Stones”

- Seed Data
  - Wikipedia, Music Brainz, Government data
  - Books, Restaurants, Companies
  - Early users: Gene, Proteins, Taxa
- Schema
  - 700 Types



# Every assertion missed is data lost

## Low cost ways to participate.



**Do you back up your  
data?**

**Data missed, is like your  
site being down!**



[Previous](#) | [Next](#)

**Thu**

03.08.07



Tim  
O'Reilly

## Freebase Will Prove Addictive

Danny Hillis' latest venture, [Metaweb](#), is about to unveil its first product, the aptly named [freebase](#), tomorrow. While freebase is still VERY alpha, with much of the basic functionality barely working, the idea is HUGE. In many ways, freebase is the bridge between the bottom up vision of Web 2.0 collective intelligence and the more structured world of the semantic web.

Early visitors to the site might be inclined to dismiss it as an early candidate for [deadpool](#). After all, is there really room for what at first glance appears to be a duplicate of wikipedia and the [Open Directory Project](#), another site that purports to collect the world's information in one place?

**THE Alpha Geek**

**Great review.**

**Really understands what we are trying to do.**

In playing with the application since, I've found that some of these work in theory, but not yet in fact. For example, as I added myself as the CEO of O'Reilly Media, the system should have also updated the entry for O'Reilly (person), but it didn't.

Not Tim's fault. The capability of the system \*appears\*

INCONSISTENT

Predict + Control Behavior

Perfectly good reason it behaves this way.

....In theory, the entry for Sebastopol, California should have changed to reflect a new company located there, but it didn't.

But that's the way it is supposed to work...

In addition, two of the companies I added as O'Reilly subsidiaries aren't really subsidiaries, which suggests wholly-owned companies, but affiliates.

....I don't have the privileges to create a field in a company record, "Affiliated companies," but presumably, if it takes off, Metaweb will promote community members to the level where they can make these kinds of edits.

Doesn't he know you can prototype the changes...

If the UI doesn't make it easy...

Its as though the

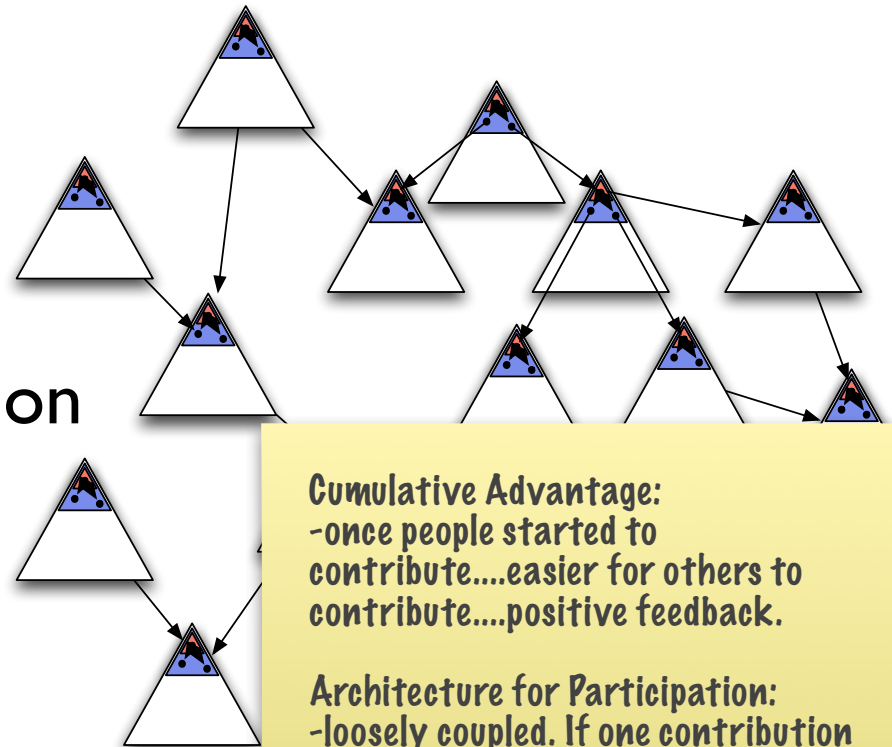
**CAPABILITY** Doesn't EXIST

Platform supports it, but user (Tim) didn't know that.



# Freebase as A Stone Soup of Data

- Cumulative advantage
- Architecture for participation
- Community at all scales



**Cumulative Advantage:**  
-once people started to contribute....easier for others to contribute....positive feedback.

**Architecture for Participation:**  
-loosely coupled. If one contribution failed, others could succeed.

**Community at all scales:**  
-Schema Designers, Developers, Domain Administrators, Individual Contributors.