



Overview & prototype demo

OA workshop
Chicago July 26, 2012

Dan Whaley
dwhaley@hypothes.is

Randall Leeds
tilgovi@hypothes.is

Our Mission

*To transform how we know
what's credible in the world
around us by enabling the
crowdsourced peer-review of
information everywhere.*

Why don't we have this yet?

Webklicker

SpinSpotter™
IS THAT TRUE?

Marginize
YOUR METAWEB

Shift
Space
.org

THE
AWESOME HIGHLIGHTER

Stickis
social + web + overlay

diigo

Kutano BETA

Bounce

A fun and easy way to share ideas on a website.

ire

NewsTrust

GOOZY

On
News
.Network

trailfire

ICorrect

Setting the record straight

thirdvoice

A.nnotate

Google
sidewiki

<http://bit.ly/existingannotation>

Why don't we have this yet?

- No effective peer-review model, if any.
- Not annotation based—i.e. many don't work on subparts of documents.
- No way to link into passages and annotations.
- Poor cold start strategies.
- Not standards based.
- Not open source.
- Not non-profit in structure.

What is Hypothes.is?

1. Open source (“Free”) & standards based.
2. Non-profit.
3. Enabling broad annotation of the world’s documents and other media.
4. Using a pseudonymous, occasionally anonymous identity model.
5. With an emergent & domain sensitive reputation model.
6. Coordinating with others to achieve long-term archiving.

FY 2012-2013 Funding



\$245k*



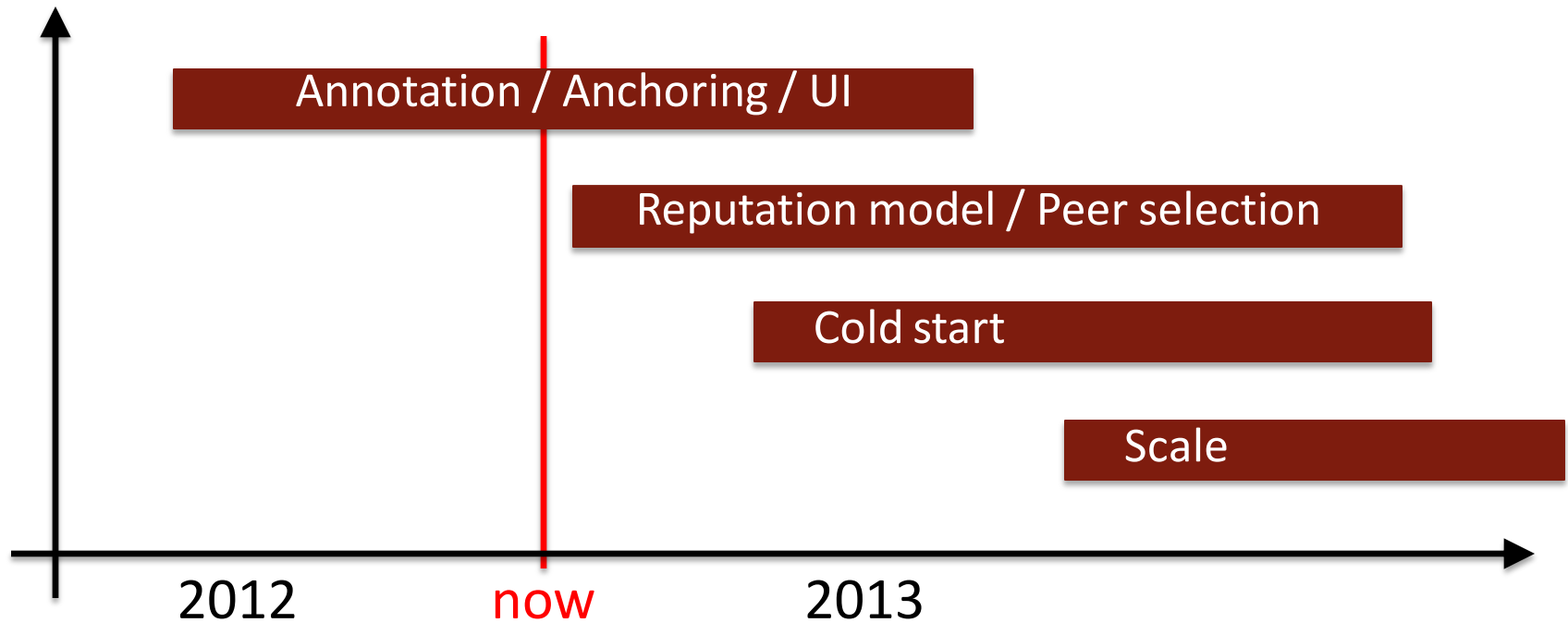
\$520k +
\$20k workshop

*includes matching gift and others

A few Grand Challenges

1. Effective reputation model (signal v noise)
2. Matching metamoderators (efficient use of experts)
3. Solving the cold start
4. Robust inter and intra-document anchors
5. Works well across the web and with a diversity of browsers, tools, etc.
6. Great UI

Timeline



Annotation



+

Deep linking



+

Peer-review



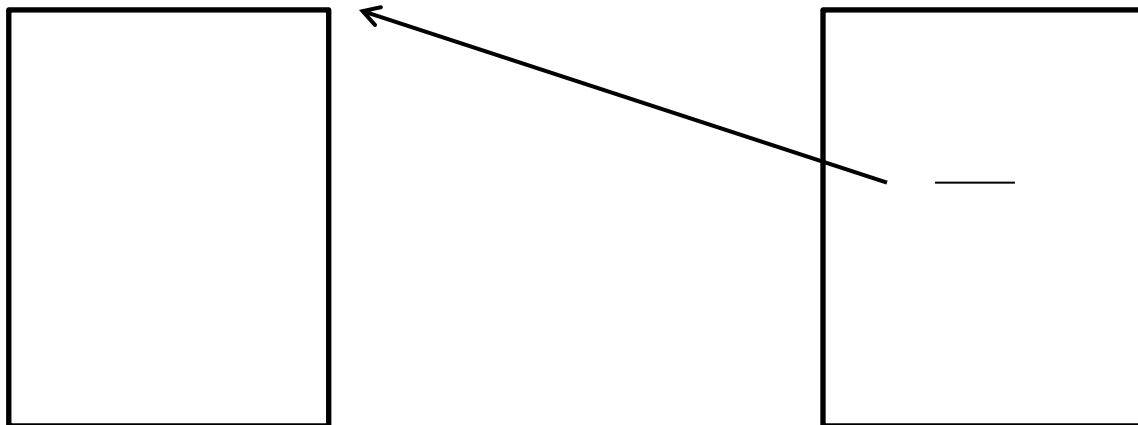
Annotation



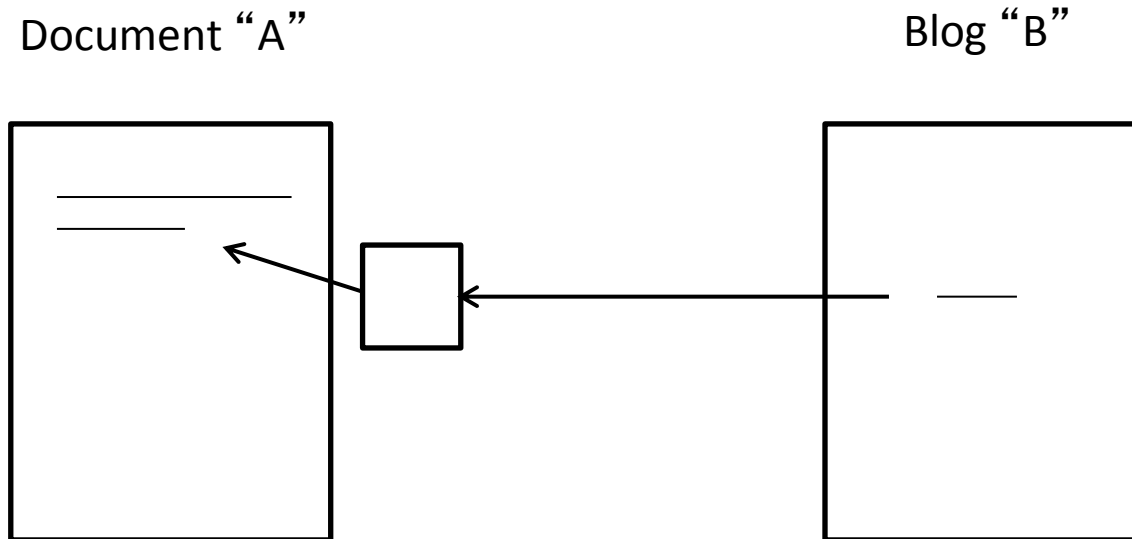
Deep linking



Now

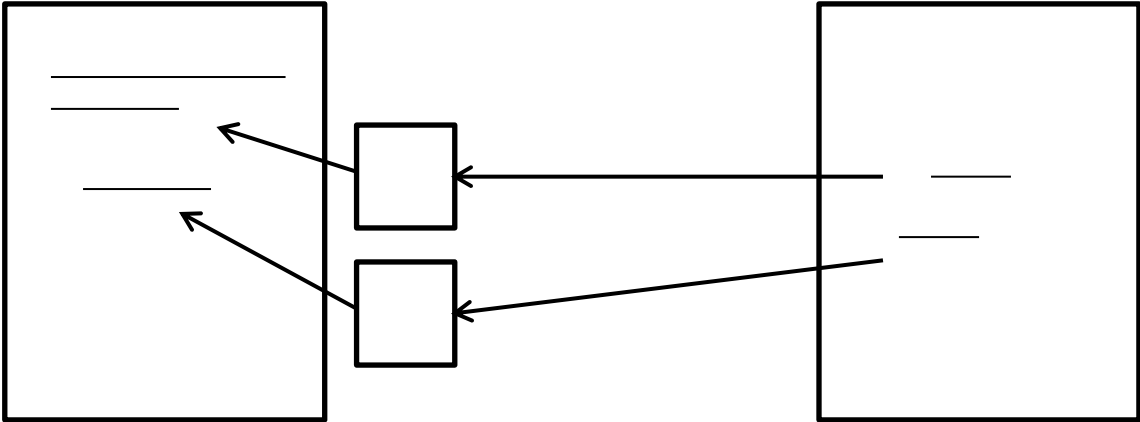


Coming soon



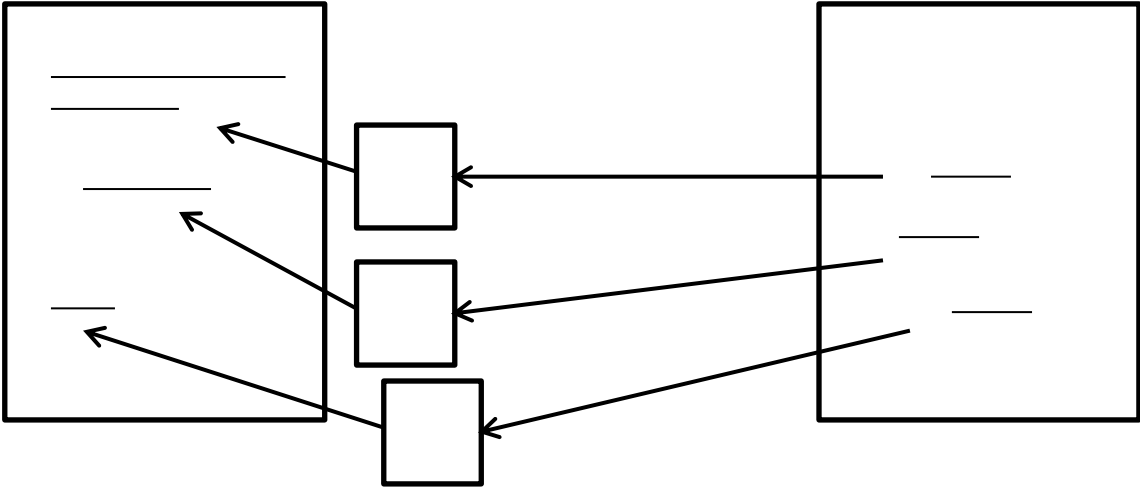
Document "A"

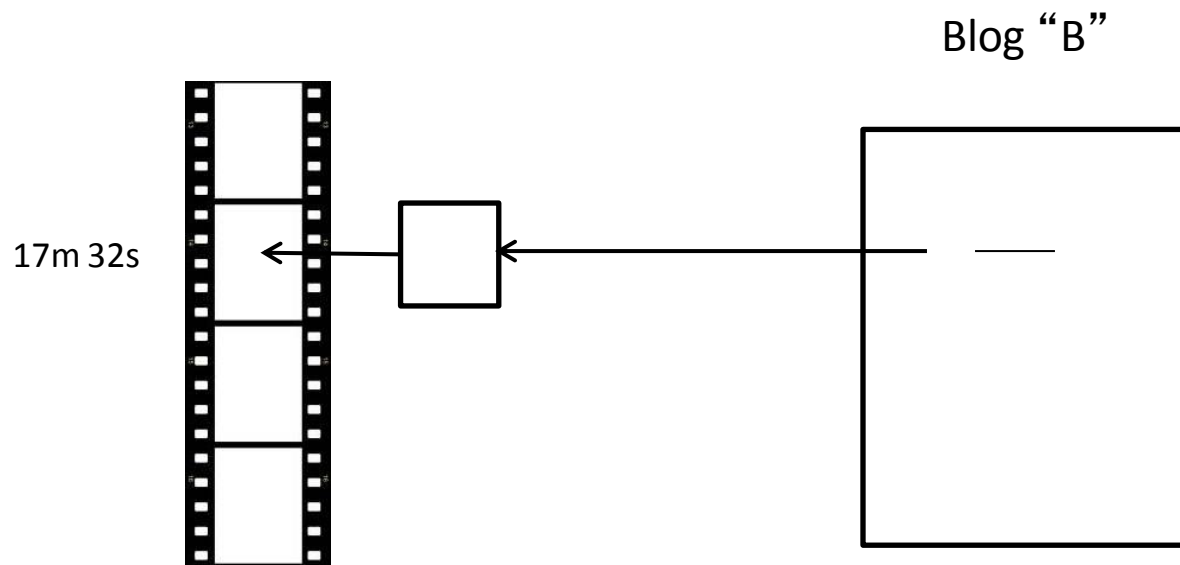
Blog "B"

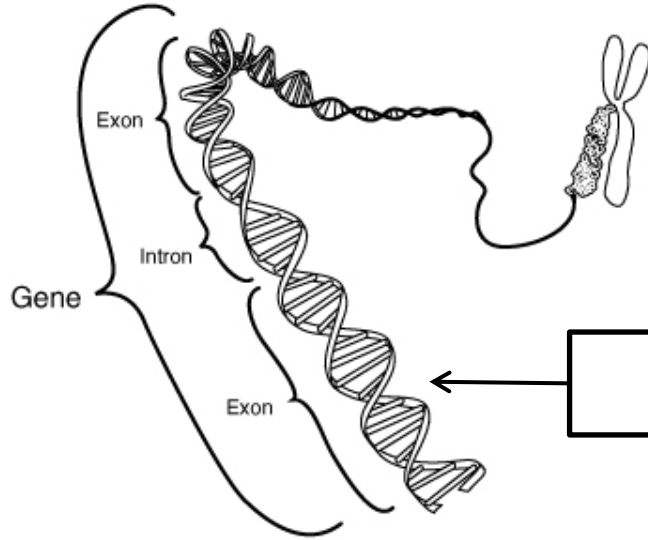


Document "A"

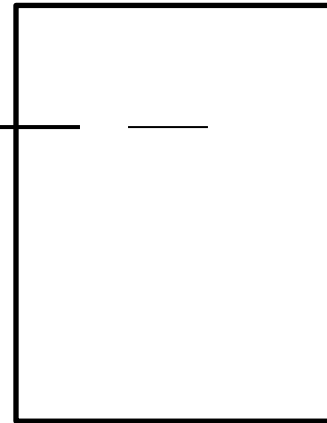
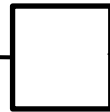
Blog "B"





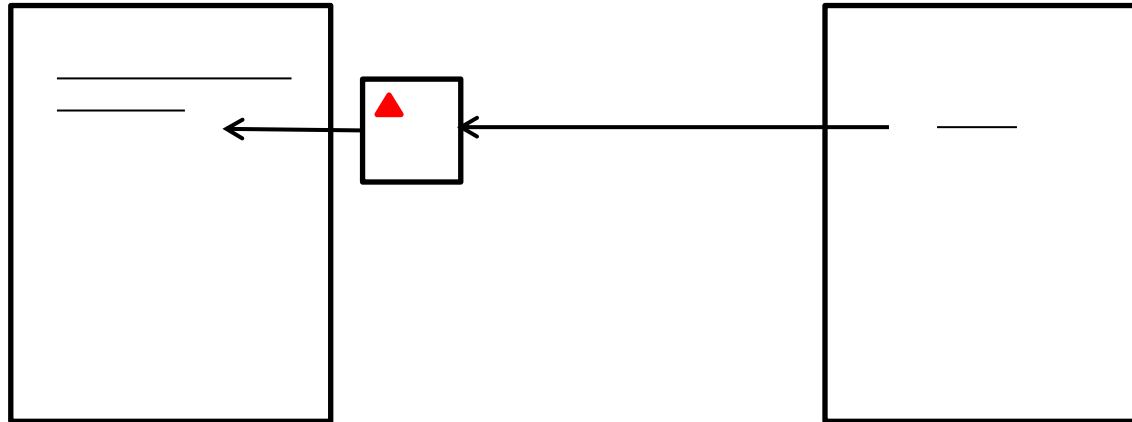


Blog "B"



Document "A"

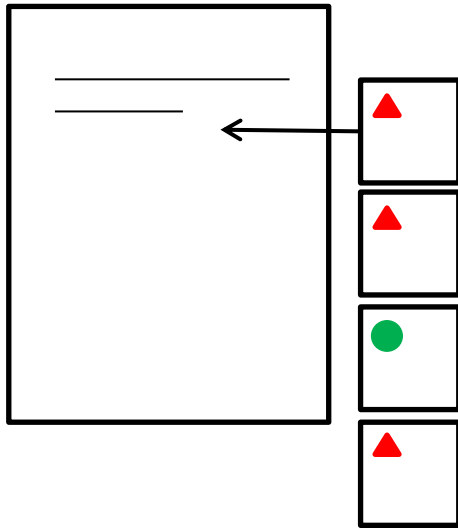
Blog "B"



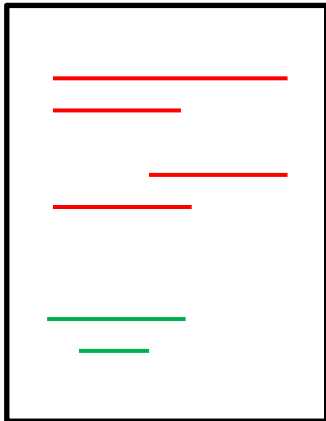
Challenge
Support
Neutral comment
Pose a question
Answer a question

Request expert review
Supply a correction
Joke
Provide a reference
Point to similar articles

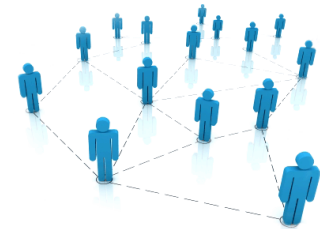
Document "A"



Document “A”



Peer-review



It doesn't have to be complicated

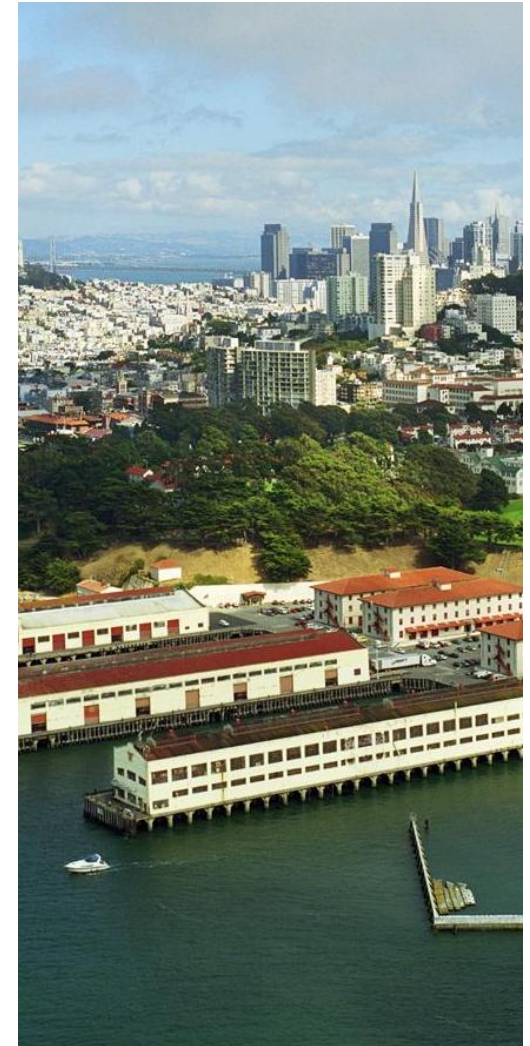
“I just say something, and they say ‘yeah, but have you considered X?’

And I say ‘Huh. Good point.’ ”

Ted Underwood / UIUC – Dept of English
commenting on the objectives of a good open
peer-review model

The Reputation Modeling Workshop

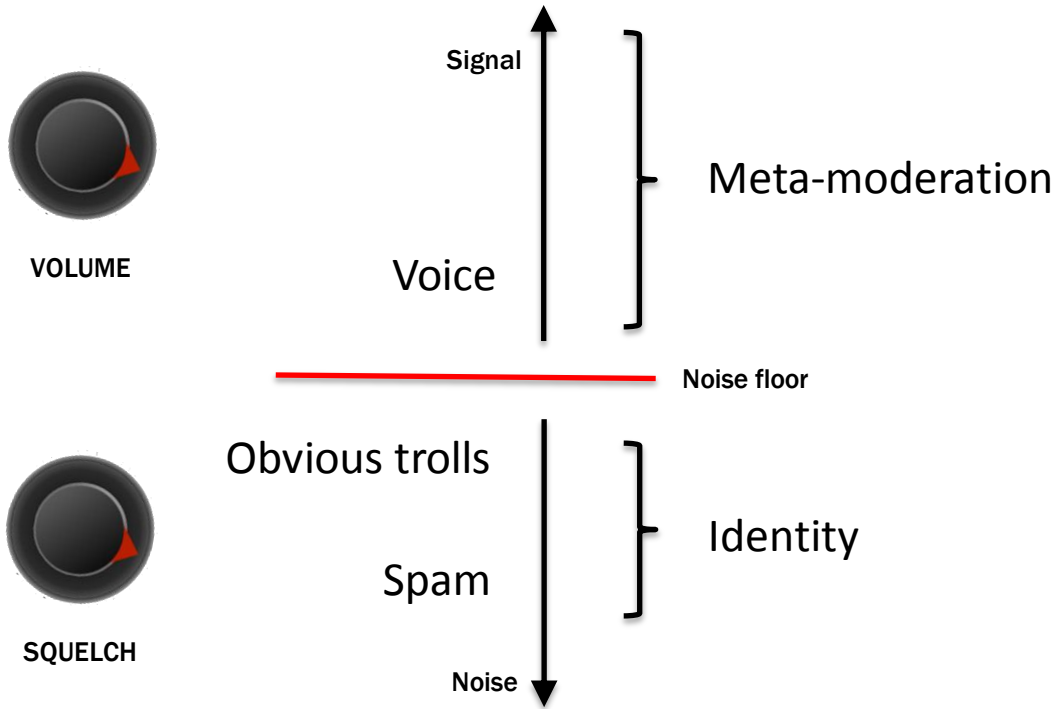
- Held Feb 22-24, 2012 in San Francisco
- Funded by \$20K grant from the Sloan Foundation's Digital Information Technology area
- Purpose: To bring together leading experts in reputation modeling to think through design challenges of enabling global annotation
- Over 50 attendees including:
 - Erik Martin / Reddit, GM
 - Jeff Atwood / Stackoverflow, Cofounder
 - Rob Sanderson / LANL
 - Radu Jurca / Google (Zurich)
 - Steven Tadelis / eBay
 - Dario Taborelli / Wikipedia
 - Roger MacDonald / Internet Archive
 - Paul Resnick / U Michigan
 - Jennifer Lin / PLoS
 - Craig Newmark / Craigslist



What we learned

- The identity model, and particularly how expensive new identities are is central to the dynamics of the reputation model.
- We need to balance direct participation w/ an objective control that should be semi-random and probably blind.
- It needs to be fun.

Signal to Noise



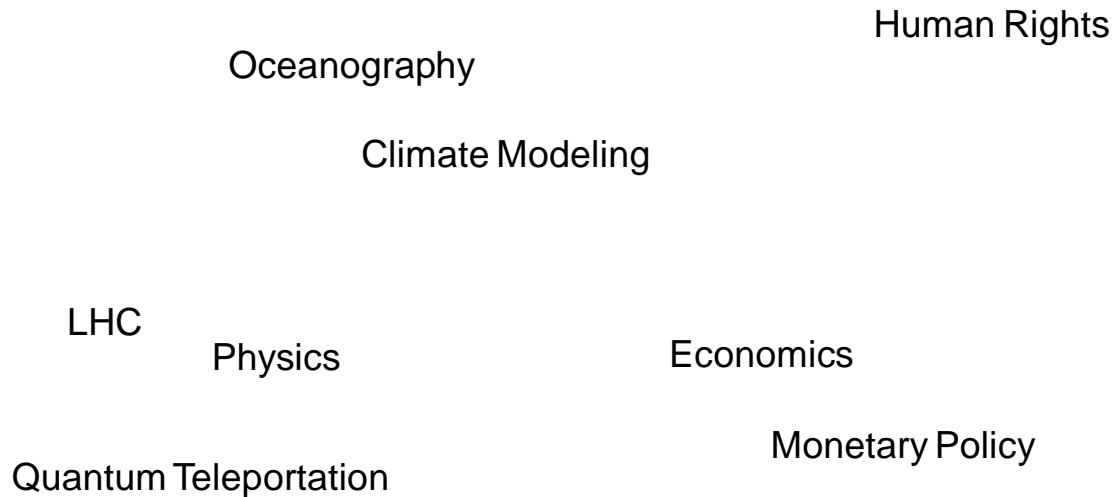
Design choices

- Pseudonymous identity, potentially gated through two factor authentication.
- Threaded, direct reply approach to moderation
- Both sentiment and score propagate up through threads
- A meta-moderated loop controls reputation
- Meta-moderation is blind and semi-random
- Meta-moderators are chosen by their reputation in the associated area. By domain proximity.
- Users can annotate for personal use, in either public or private groups, or in a global public channel.
- Standard social kit— following others, sharing, etc.

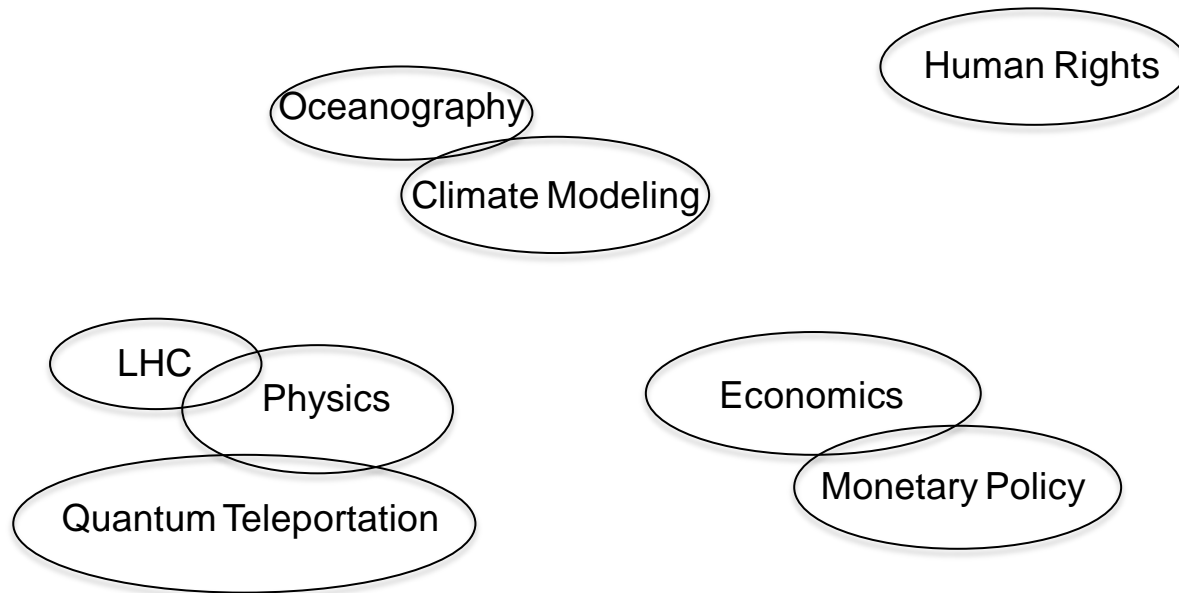
Two moderation loops

	Moderation	Meta-moderation
Opportunities	Come through replies in threads	Come via an activity stream
Participants	Involves people that are already at the site	Blends people between sites based on expertise
Reputation ...	Is leveraged	Is determined
Interaction Model	Direct response in threads	Blind, random
Qualifications	Anyone who has passed initial screen	Only people above a certain reputation threshold

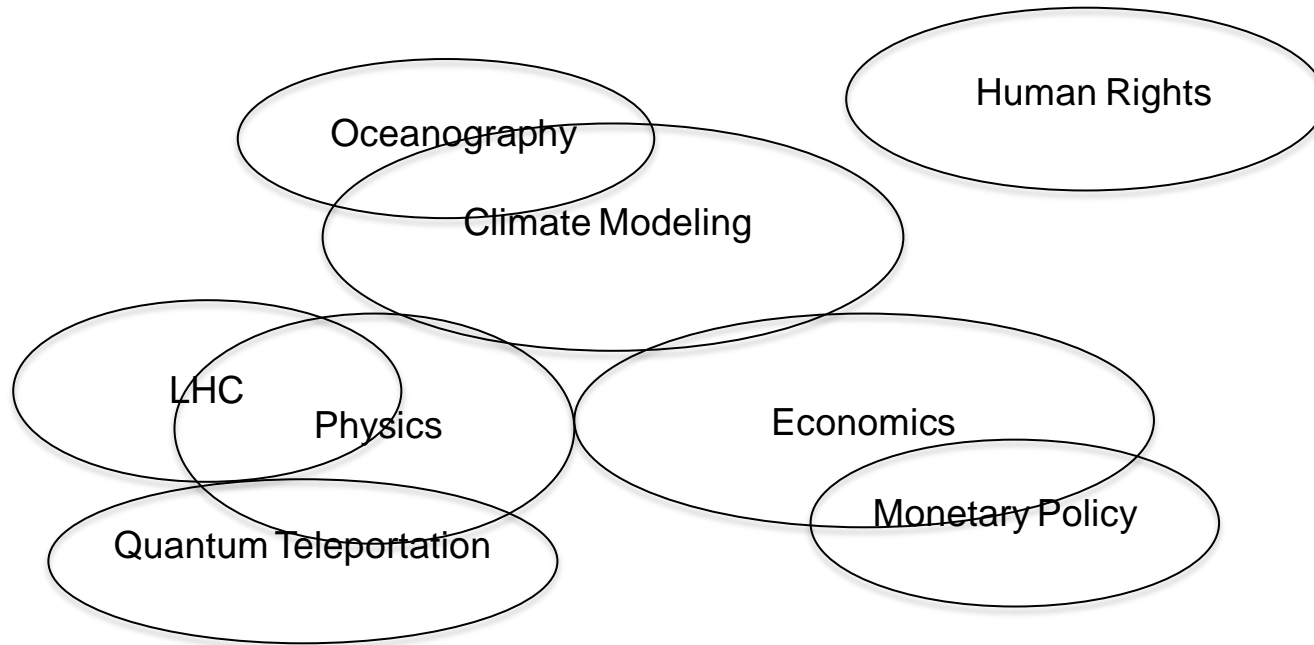
Meta-moderation matching through domain proximity



Meta-moderation matching through domain proximity



Meta-moderation matching through domain proximity



A diversity of interfaces

Plug-ins



URL

<http://hpt.is/TzLv>

Deep links / embeds



Website



API



Internet Archive partnership

- Will store annotations permanently
- Domain assignment agreement
- Will cache reference copies of annotated pages permanently (for versioning)
 - Exploring same relationship w/ Common Crawl
- Wants to enable annotation on their
 - Wayback machine (version history of internet)
 - Books
 - New video project
- Early, active & enthusiastic OA participant

Use case examples

- Gov't documents
 - Legislation, regulations, FOMC minutes
 - Scientific papers
 - Post-publication review
 - Terms of service, Privacy policies, EULAs
 - Facebook, Yahoo, OnStar, etc
 - News / Blogs
 - Books / Manuscripts
 - Video / Audio
 - Data (scientific, public,
 - Corporate documents
 - 10K, 10Q, analyst call minutes
 - General content
- HTML
 - PDF
 - EPUB
 - Video
 - Audio
 - Images
 - Manuscripts
 - Data
 - Software
 - Spreadsheets & other application-specific document formats

The cold start

(Selection of launch domains)

- Stable (easy to anchor to)
- Longevity (at least months)
- Single sources (canonical URLs)
- In the public interest
- Trending
- Content is not highly subjective
- No commenting system already
- Large pool of quality annotators
- No paywalls
- Annotation makes a difference

Use case examples

- News / Blogs
- Corporate documents
 - 10K, 10Q, analyst call minutes
- **Gov't documents**
 - Legislation, regulations, FOMC minutes
- **Scientific papers**
 - Post-publication review (arxiv, PLoS)
- **Terms of service, Privacy policies, EULAs**
 - Facebook, Yahoo, OnStar, etc
- General content
 - Books, PDFs, video, etc

Anchors

- Inter-document
 - How to create robust anchors that survive document changes in structure and content
 - Strategies for calculating confidence metrics and graceful failure modes
- Intra-document
 1. Different document formats (HTML / PDF / Text) should show same annotations (related to Canonical URL issue)
 2. Handling pagination vs. single page (related to reflowable text)
 3. Pure content anchoring (annotating Hamlet, or the King James bible wherever it is)?

DEMO