

*The cultivated person's first duty is to always be prepared to rewrite the encyclopedia.*

–Umberto Eco

# Wikipedia and collective problem-solving



**WIKIPEDIA**  
*The Free Encyclopedia*

Samuel Klein  
Wikimedia Foundation

Alan Turing Centenary  
Manchester, UK June 2012

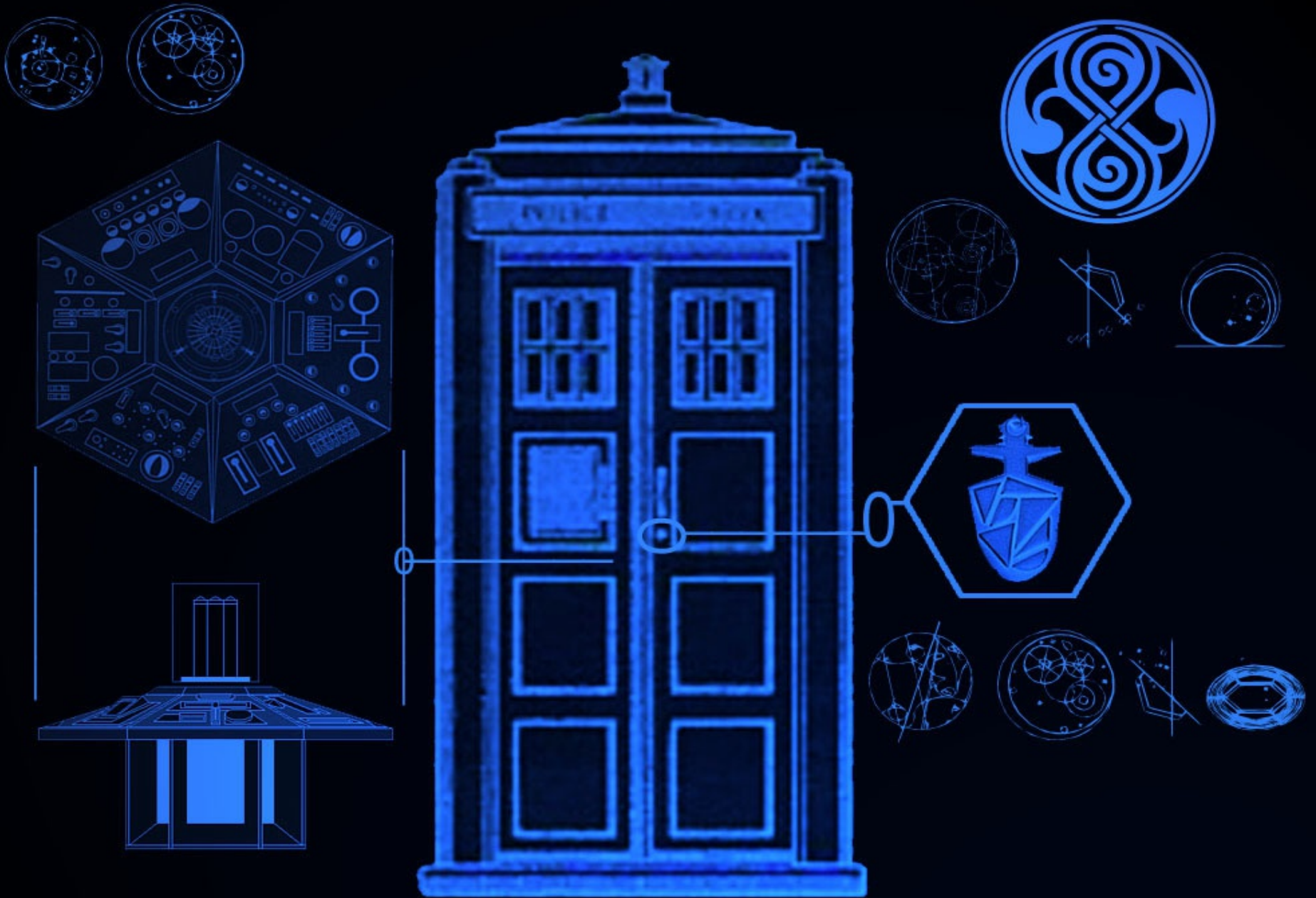
# Evolution of the large scale collaboration: a case study

- The encyclopedic drive & Wikipedia
- Current challenges for WP
- Future project domains

What else can be tackled this way?

- Tools for the next generation

# Housekeeping



# Encyclopédistes: saving the world?

- Louis de Jaucourt (Encyclopédie)
- W.C. Minor (OED)
- Farmbrough, Knapp (Wikipedia)



# TwenCen: World Brain

- 1895 – Permanent Encyclopedia\* (Otlet, cards)
- 1936 – World Brain\* (H.G. Wells, microfilm)
- 1945 – Memex\* (Bush, hypertext)
- 1965 – Xanadu\* (Nelson, transclusion)
- ...
- 1991 – World Wide Web (Berners-Lee, read-write)
- 1995 – WikiWikiWeb (Cunningham)
- 1999 – Free Universal Encyclopedia\* (Stallman)
- 2000 – Nupedia (Wales)

# 2001: ~~Nu~~Wikipedia



XANADU? Not yet...

But after a bit of hacking:

- Permanent links
- Transclusion
- Bidirectional links

...by his reason, ...  
...this singular passion ...  
...other animals, which is a ...  
...of mind, that by a perseverant  
...of delight in the continual and  
...indefatigable generation of  
...knowledge, exceeds the short  
...vehemence of any carnal  
...pleasure.

**WIKIPEDIA**  
The Free Encyclopedia

# Parallel work, 2000-2012

- 1999 – Everything2
- 2001 – H2G2 (Adams)
- 2007 – Freebase (structured data)
- 2012 – Knowledge Graph (Google)





# The Secret Life of Wikipedia

*Wikipedia*, *n.* A unique experiment in collaboration, derived from the Hawaiian *wiki*, "edited at high speed", and the Greek *παῖδh*, "by children"

# Differences from past efforts

## Flexibility + Ownership

## Automation + Copycat-friendly

Adaptation and forks welcome

## Minimal initial toolchain

## All actions reversible

Every revision kept

# Epic backlogs

## Backlog categories

- **General**
- **Administrative**
- **Assessment**
- **Stub sorting**
- **Template parameter**

## Lacking references

- Needs references: 234,508
- **Needs references (BLP): 762**
- Needs more references: 186,459
- **Needs more references (BLP): 58,367**
- Needs in-text citations: 43,688
- Needs reliable references: 39,043
- Has unsourced quotes: 734
- Has unsourced statements: 230,940

## Accuracy

- Uses broken or outdated citations: 99
- Uses incorrect citation syntax: 751
- Needs factual verification: 14,333
- Needs references list: 12
- Accuracy disputes: 6,155
- Neutrality disputed: 7,760
- Needs updating: 5,212

## Content

- **Copied and pasted: 57**
- Needs context: 1,755
- Needs coordinates: 163,885
- Uses improper non-free content: 135
- Uses in-universe perspective: 3,466
- Plot summary needs attention: 2,938
- Of unclear notability: 54,857

## Links

- **Has dead external links: 107,382**
- Needs external links cleanup: 2,480
- Needs to be wikified: 18,864

## Layout

- Needs sections: 147
- Needs expansion: 99,868
- **Needs to be merged: 13,050**
- Needs to be split: 151

## Style

- Needs clarification: 3,834
- Needs copy edit: 2,775
- Has peacock terms: 1,666
- **Uses promotional tone: 15,204**
- Has weasel words: 11,813

## File description

- **Copy to Wikimedia Commons: 8,435**
- Needs description: 5,063

## File quality

- Has inappropriate JPEG compression: 1,859
- Should be in PNG format: 451
- Has watermarks: 254

## Non-free file use

- Needs size reduction: 839
- Needs purpose specified: 256
- Publicity photographs needing rationale: 131

## File licensing

- **Has disputed copyright information: 1,560**
- Replacing placeholders: 846

## Meta

- **Biographies without "living" parameter: 64**
- Persondata without "name" parameter: 6,802
- Persondata without "short description" parameter: 436,245

## Quality content

- Featured article candidates: 49
- Featured article review: 8
- Featured list candidates: 39
- Featured list review: 9
- Good articles needing "oldid": 153
- **Unreviewed good article nominations: 232**

## Other

- Needs cleanup: 11,978
- Needs cleanup (disambiguations): 12
- Needs expert attention: 5,461
- Needs infobox conversion: 655
- Is orphaned: 175,723
- **Needs links to other pages: 227**
- Underpopulated categories: 1,386

Goes well with RecentChanges!

# Who Writes Wikipedia?

- The projects are like a federation of small towns: policies, civic bodies, newspapers.
- 1.2M people in 250 languages.
- Many copyeditors: more words have been deleted than remain
- Also bots and script-writers: automation supports half of all work done.

...making people happier! (hattip to Vardi)

# Legal structures

The Wikimedia Foundation + 40 national chapters:

- Donation-funded charities dedicated to preserving and supporting the projects
- Grew out of the projects once needed

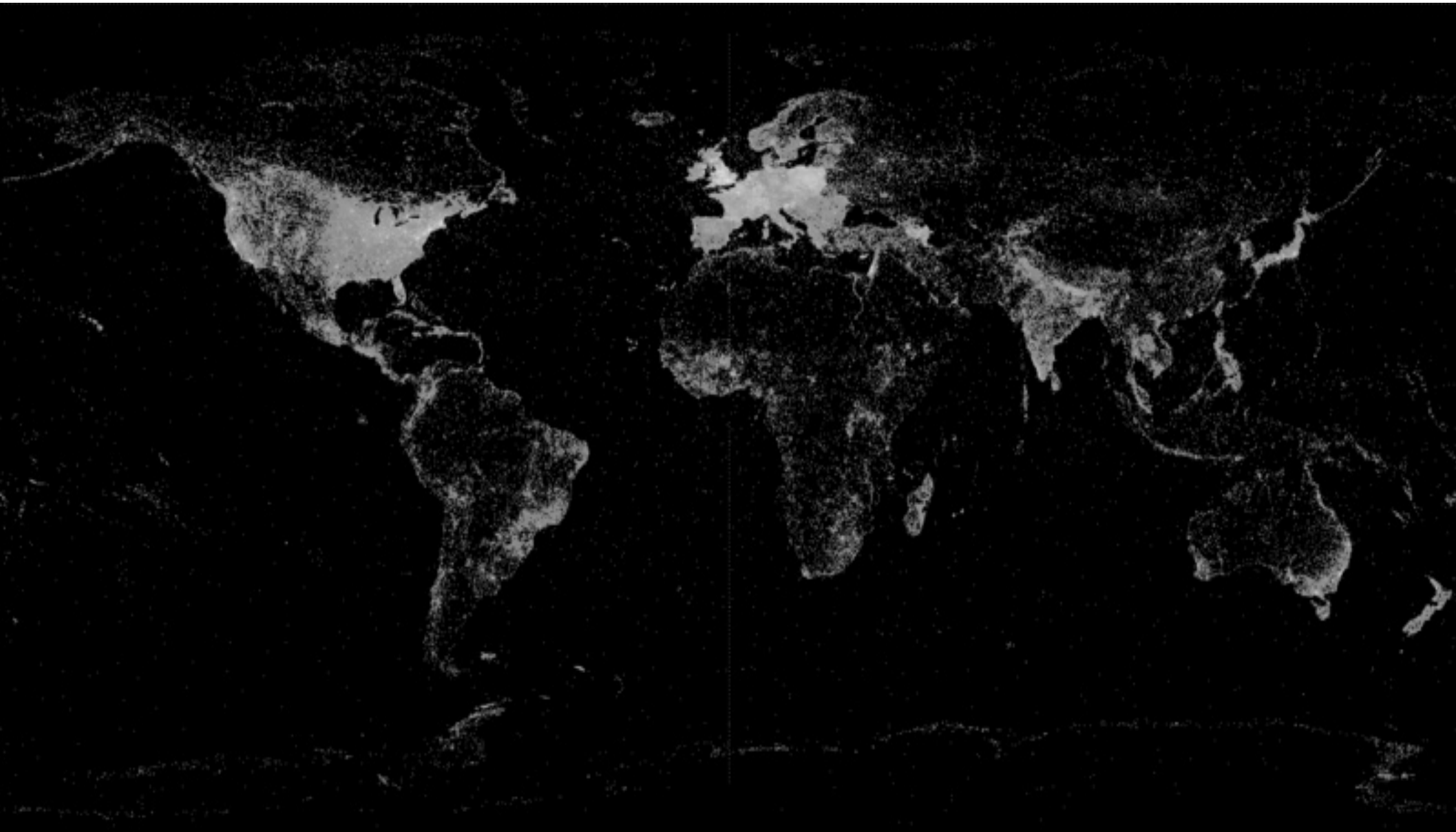
# Current Challenges and Problems

**Wikocracy** (n.): a mix of anarchy, despotism, democracy, meritocracy, and bureaucracy.

# Bootstrapping solutions to problems

- **Categories**  
(unrestricted, single-parent)
- **Layout templates**  
(now Turing-complete)
- **Quality: iterative metrics**
- **Truth disputes: Reducing to sourceability and neutrality**

# Systemic bias



Article distribution by geotag, 2011



# New types of knowledge



250+ languages, 40 chapters

# There is a deadline.



**Golden toad, extinct (1989)**

Image by Charles H. Smith, for the USFWS, via Wikimedia Commons: [http://commons.wikimedia.org/wiki/File:Bufo\\_periglenes1.jpg](http://commons.wikimedia.org/wiki/File:Bufo_periglenes1.jpg)





## Cologne city archive, 2009

Image from Frank Domahs via Wikimedia Commons: [http://commons.wikimedia.org/wiki/File:The\\_destroyed\\_sixstory\\_cologne\\_city\\_archive.jpg](http://commons.wikimedia.org/wiki/File:The_destroyed_sixstory_cologne_city_archive.jpg)

# Turing-complete templating system

```
{{#ifeq:{{{child|}}}|yes|<table class="infobox {{{bodyclass|}}}" cellspacing="5"
style="width:22em; {{#if:{{{float|}}}|float: {{{float|}}}; clear: {{#ifeq:{{{float|}}}|none|both|
{{{float|}}}}}; margin: 0 0 1em 1em;}} {{{bodystyle|}}}">
```

# Dependent on the open web



## Imagine a World Without Free Knowledge

For over a decade, we have spent millions of hours building the largest encyclopedia in human history. Right now, the U.S. Congress is considering legislation that could fatally damage the free and open internet. For 24 hours, to raise awareness, we are blacking out Wikipedia. [Learn more.](#)

### Make your voice heard



Facebook



Google+



Twitter

WIKIPEDIA  
~~WIKIPEDIA~~

SOPA blackout, Jan 18, 2012

# Solving new collective problems

What open problems – in science, culture, or society – can be solved with such systems?

# Wiki Future

The problem with Wikipedia is that it only works in practice. – the Zeroth Law of Wikipedia



# Special collections



WLM 2011: Chaijna Monastery, Bucharest. (1 of 165,000)



# Revenge of the Semantic Web

[Denny](#) [My talk](#) [My preferences](#) [My watchlist](#) [My contributions](#) [Log out](#)



[Main page](#)  
[Contents](#)  
[Access the API](#)  
[Random page](#)  
[Donate to Wikidata](#)

[Interaction](#)  
[Help](#)  
[About Wikidata](#)  
[Community portal](#)  
[Recent changes](#)

[Languages](#)  
[Catalá](#)  
[Cesky](#)  
[Dansk](#)  
[Deutsch](#)  
[Eesti](#)  
[Español](#)  
[Esperanto](#)  
[Français](#)  
[Hrvatski](#)  
[Italiano](#)  
[O'zbek](#)  
[Complete list](#)

Article [Discussion](#)

Read

[Edit](#)

[View history](#)



[WikiTrust](#)



## Berlin [edit](#)

From Wikidata

Capital of Germany [edit](#)

**Also known as:** [City of Berlin](#) [edit](#) | [x](#)

Continent [Europe](#) [\[3 sources\]](#)

Country [Germany](#) [\[2 sources\]](#)

Population [3,490,445](#) [\[1 source\]](#)

[3,500,000](#) [\[2 sources\]](#)

[\[other values\]](#)

Calling code [030](#) [\[2 sources\]](#)

Mayor [Klaus W|](#) [\[0 sources\]](#)

Vehicle registration [Klaus Wowereit](#) [\[1 source\]](#)

Area [German politician](#) [\[2 sources\]](#)

Twin city [Klaus Wunderlich](#) [\[3 sources\]](#)

[\[new fact\]](#)

<a href="#">Klaus W </a>
<b>Klaus Wowereit</b> <i>German politician</i>
<b>Klaus Wunderlich</b> <i>German musician</i>
<b>Klaus Waldeck</b> <i>Austrian musician and former lawyer</i>
<b>Klaus Wagner</b> <i>German mathematician</i>
<b>Klaus Wagner</b> <i>Stalker of the British Royal Family</i>

Wikidata (Allen, Moore, Google), WikiCite (DPLA).

# Parallel efforts today

Mapping: **OpenStreetMap**

Citizen Science: **Zooniverse**

Survey photos: **Geograph**

Q & A: **Quora**

# Everyone has something to teach

- Genealogy
- Local knowledge  
Monmouthpedia →
- University class projects



# Wikipedia Education Program

<http://education.wikimedia.org>





# Seeking new collective problems

Echoes of activator-inhibitor systems:

“Trolls are the driving force of Wikipedia. The worst trolls often spur the best editors into creating a brilliant article...”

– Bachmann's Law

Are there pre-patterns as well? What open problems can be solved with these systems?

# Coda: Tomorrow's Hackers



# One Laptop per Child ~~Gamer~~

worldwide over 2.4 million children and teachers have no laptops





# Pervasive social problem-solving



Are gamers free labor? (Goble) As with the ACE: not for long. (Brooks)

Image by OLPC: <http://one.laptop.org/stories/peru-learning-how-learn>





Reading Wikipedia offline. Arahuary, Peru. (2007)

Image by OLPC: <http://one.laptop.org/sites/default/files/nic-on-a-boat.jpg>

*We can only see a short distance ahead, but we can see plenty there that needs to be done.*

– AMT





## Moving the Wikiball (Taipei, 2007)

Image by User:Everlong via Wikimedia Commons: [http://commons.wikimedia.org/wiki/File:Wikimania2007\\_wikiball\\_129.jpg](http://commons.wikimedia.org/wiki/File:Wikimania2007_wikiball_129.jpg)

**Many thanks** to: Andrei Voronkov, the University of Manchester + the Kurt Gödel Society  
**Material and images** from: Wikimedia Commons, Joseph Reagle's *Good Faith Collaboration*, Raul's Laws of Wikipedia, Benjamin Mako Hill + Phoebe Ayers