




# The Wikipedia

Dr. Luis Ibanez, Kitware

# © 2007 Luis Ibanez

- This presentation is Copyrighted by Luis Ibanez
- This presentation is distributed under the Creative Commons Attribution License 3.0:

<http://creativecommons.org/licenses/by/3.0/>

- You are free to Reuse 
- You are free to Remix 
- Provided that you give credit to the author 

This presentation was created using

## Open Source Software



Open Office copyright is jointly held by Sun Microsystems and Contributors.

The software is distributed under the GNU Lesser General Public License Version 2.1.

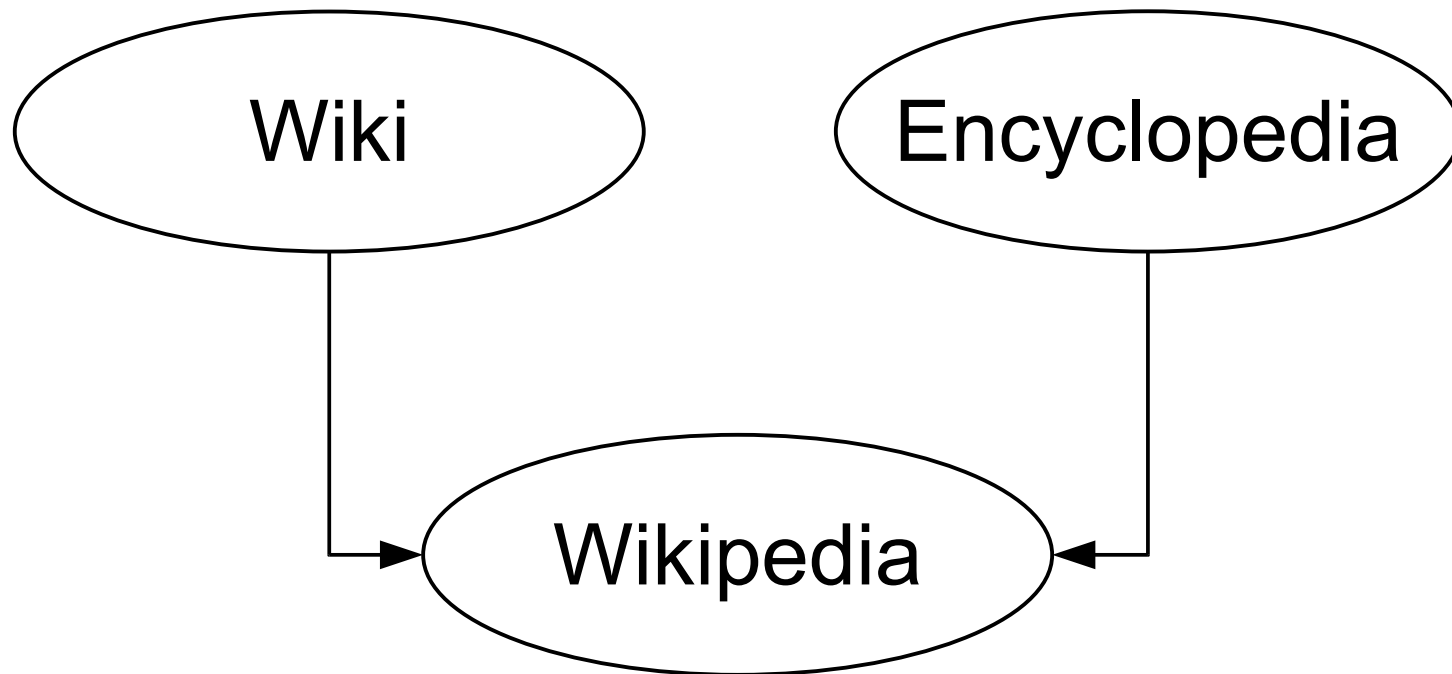


There is Hope...

# Wikipedia



# Wikipedia



# Wikipedia

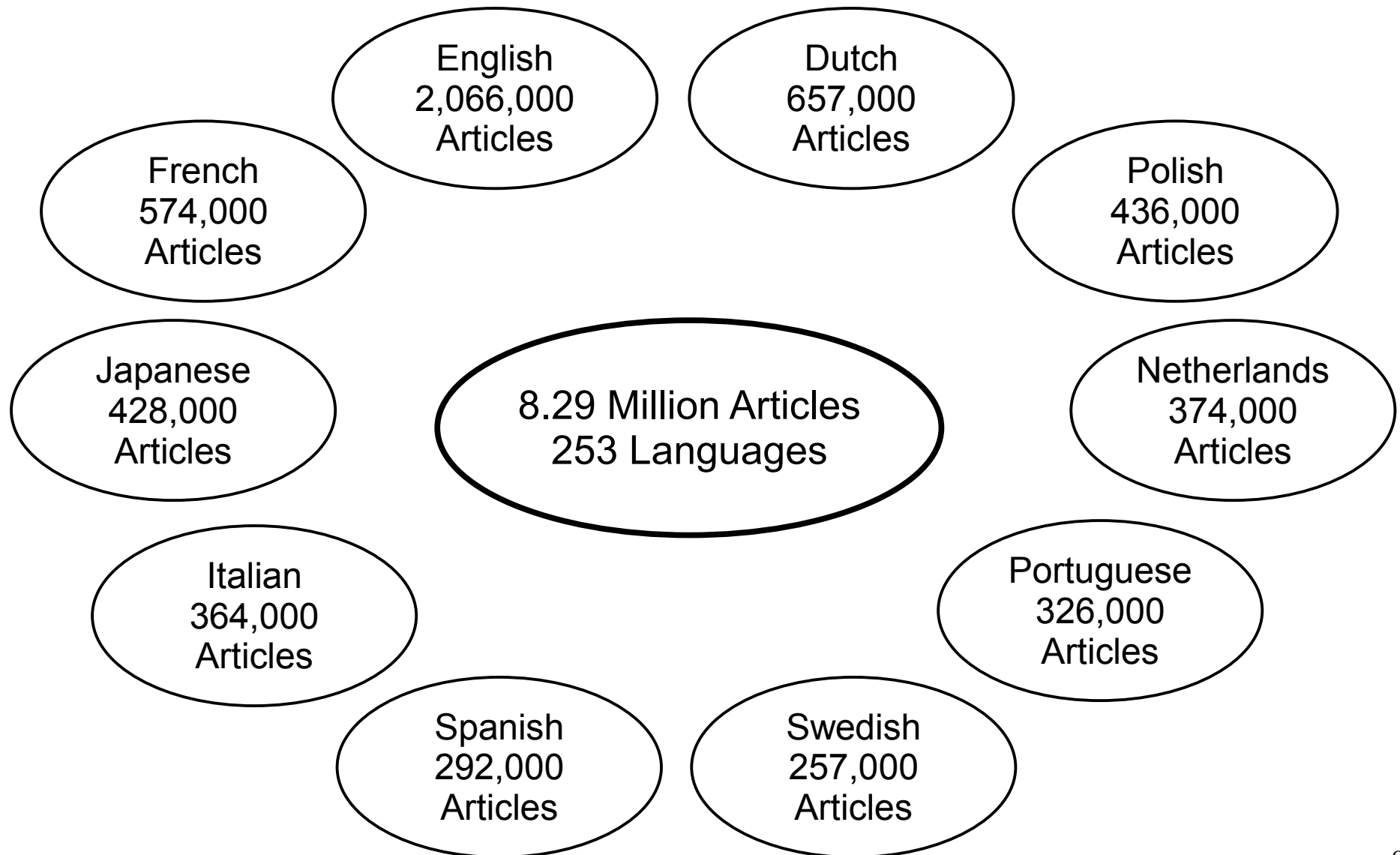
- Yahoo
- Google
- Microsoft Network (MSN)
- YouTube ●
- Windows Live ●
- Myspace ●
- Facebook ●
- **Wikipedia** ●
- Orkut ●
- Hi5 ●

[http://www.alexa.com/site/ds/top\\_500](http://www.alexa.com/site/ds/top_500)



● Web 2.0

# Wikipedias



# The Future of the WWW

...and non-commercial sites such as the **Wikipedia** have pioneered new collaborate styles of information sharing.



Tim Berners-Lee

**innovation will happen**

provided it has a platform of **open** technical standards, a flexible, scalable architecture, and **access** to these standards on **royalty-free** (\$0 fee patent licenses) terms.

# Wikipedia – Free Content



non-for-profit  
Wikimedia Foundation

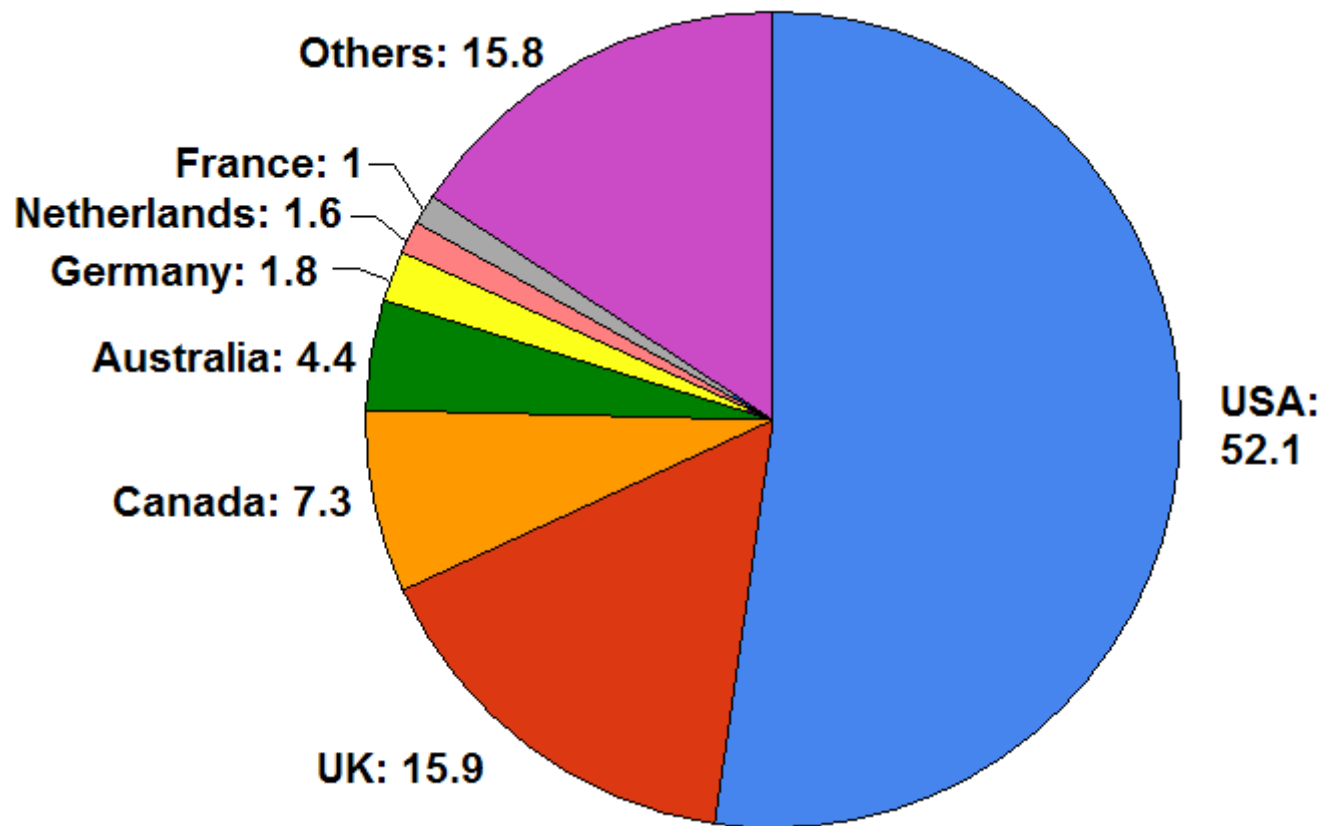
Less than 10  
full-time staff

Volunteers Collaborating

Some Language Versions  
carry full  
Free Content

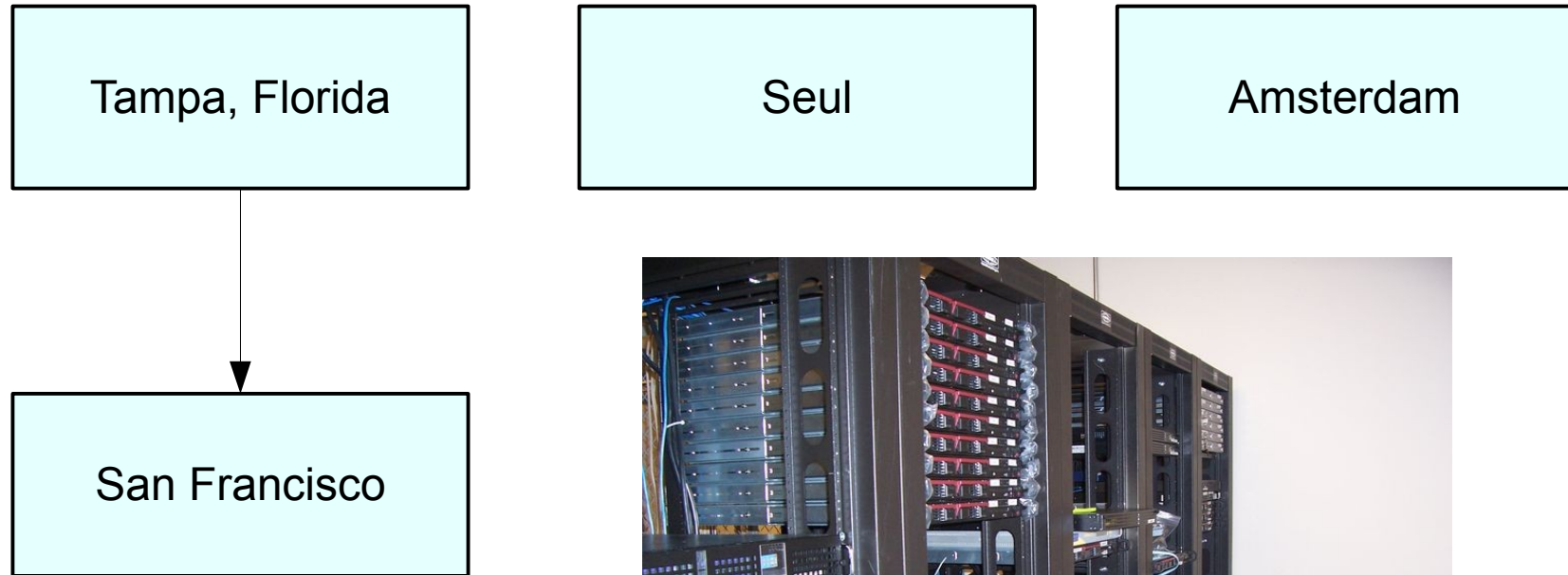
English Version  
carries some  
non Free Content

# Number of Contributors by Country



[http://en.wikipedia.org/wiki/Image:English\\_Wikipedia\\_contributors\\_by\\_country.png](http://en.wikipedia.org/wiki/Image:English_Wikipedia_contributors_by_country.png)

# Infrastructure

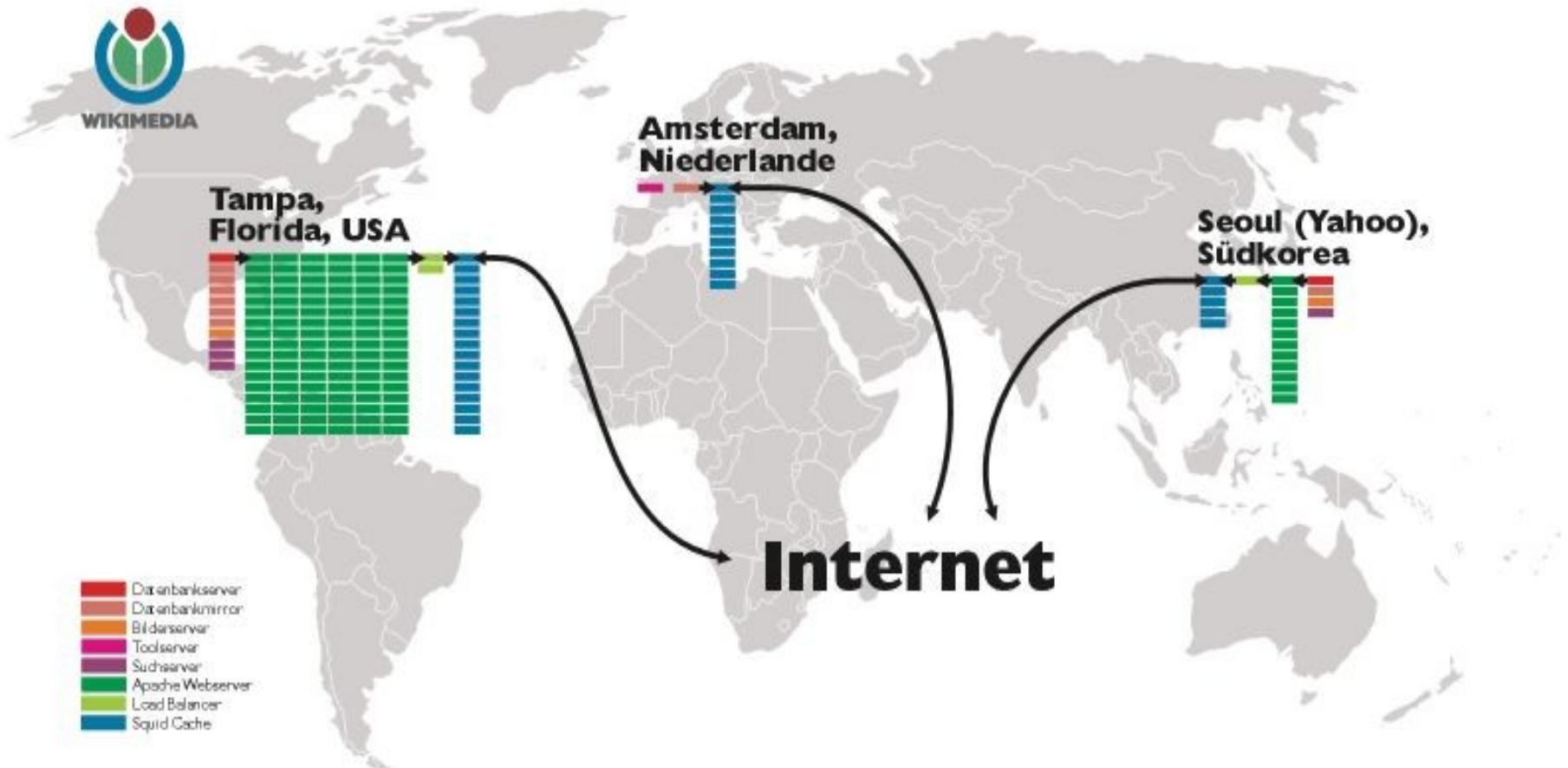


Dedicated clusters  
of Linux Servers



More than 100 servers, managing 10,000 to 35,000 requests per second

# Infrastructure



[http://meta.wikimedia.org/wiki/Wikimedia\\_servers](http://meta.wikimedia.org/wiki/Wikimedia_servers)

# MediaWiki



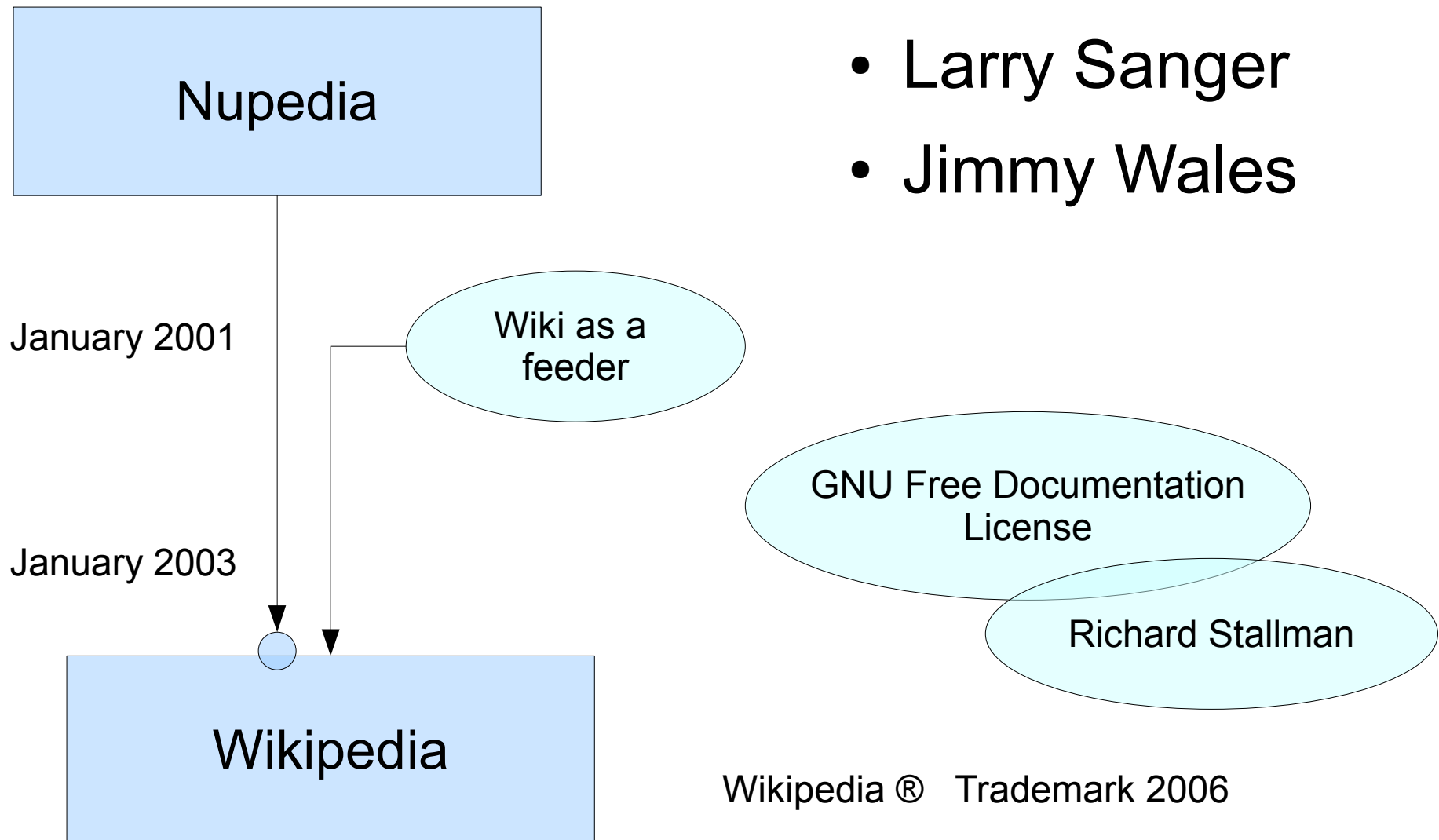
<http://www.mediawiki.org/wiki/MediaWiki>

# MediaWiki

- Written in PHP
- Built upon MySQL
- Licensed as GPL
- Page modifications are added to the database
- Easy page recovery in case of vandalism
- Manage image and media files
- Supports caching
- Coupled with Squid proxy server

# Founders

- Larry Sanger
- Jimmy Wales



# Wikipedia Size

2 Million Articles in 2006

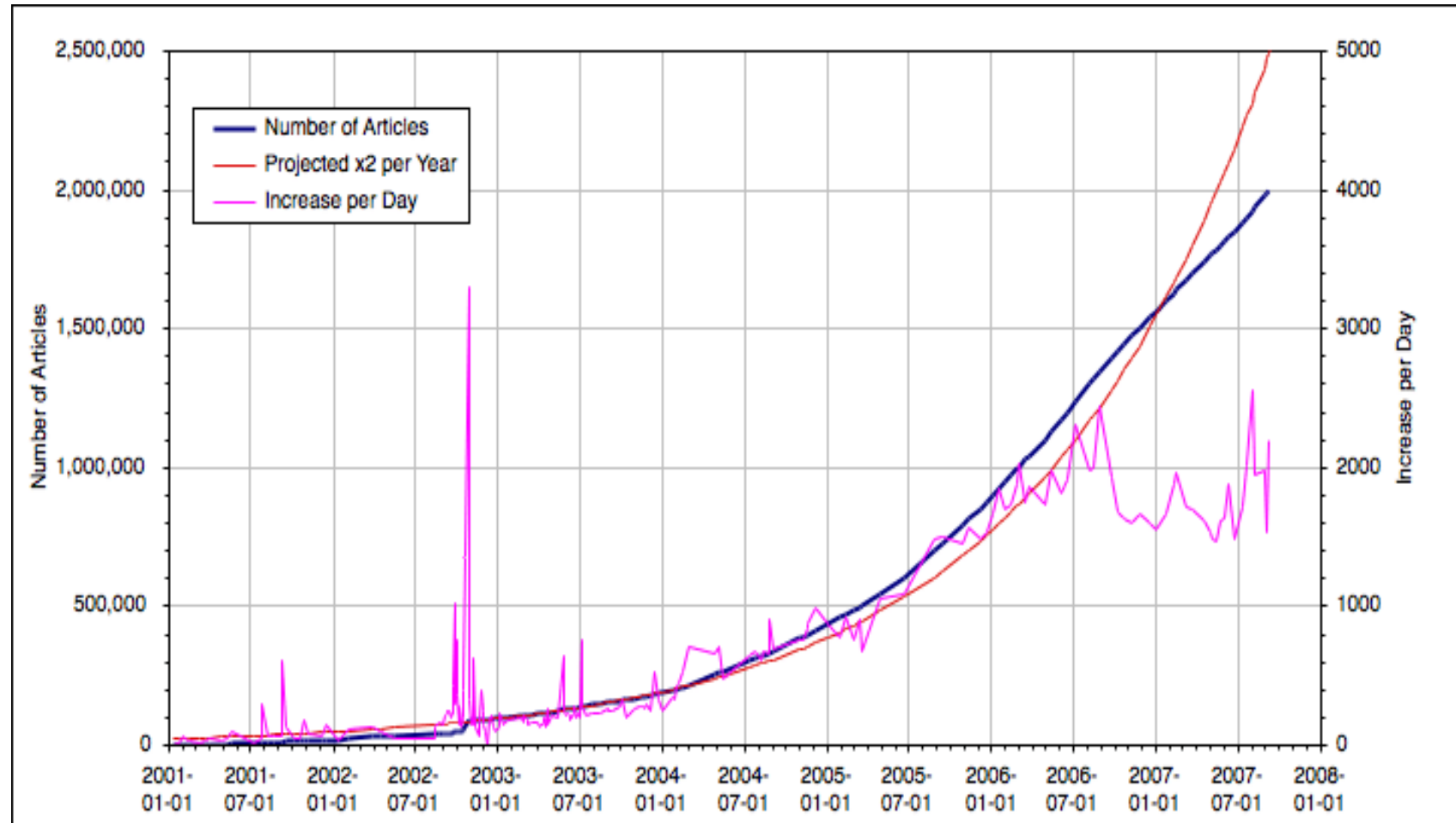
Yongle Encyclopedia (1407)

Record for 600 Years !!

# Wikipedia Size Comparison

	Articles (K)	Words (M)	Characters (M)	Words/art.
• <b>Wikipedia</b>	2,000	1,000	3,500	435
• Siku Quanshu	-	800	-	-
• Yongle Encyclopedia	-	370 / 770	-	-
• Enciclopedia Universal	1,000	200	1,000	-
• Gujin Tushu Jicheng	-	100	-	-
• Encyclopedia of China	80	126	-	1,580
• Enciclopedia Italiana	60	50	247	833
• Nationalencyklopedin	183	-	-	-
• Encyclopaedia <b>Britannica</b>	120	55	300	370
• Great Soviet Encyclopedia	100	21	200	570
• <b>Encyclopedie</b>	72	20	-	278
• <b>Microsoft Encarta</b>	70	40	200	200
• Encyclopedia Americana	45	25	-	556

# Number of Articles



[http://en.wikipedia.org/wiki/Image:EnglishWikipediaArticleCountGraph\\_linear.png](http://en.wikipedia.org/wiki/Image:EnglishWikipediaArticleCountGraph_linear.png)

# Wikipedia Statistics (English Edition)

- 2 Million Articles
- 175 Million edits by users
  - Average of 16 per page
- 5.7 Million registered users
- 1,390 Users have administrative tools

# Wikipedia - Essentials

- Wikipedia is not for sale.
- Non-for-profit.
- Free for everyone (learned from Free Software) **GFDL**
- 250 Languages (local chapters, volunteer translators)
- You can't change anything, only add to it. (MySQL)
- Quality Control: Editors
- Not an authoritative reference: (use critical thinking)
- Is a collection: contributions by unpaid volunteers
- For the long haul: at least 100 years from now.

# Wikipedia - Open Nature

- Collaboration of volunteers
- Consensus over Credentials
- Susceptibility to Vandalism
- Capability for Self-correction
- As accurate as other Encyclopedias
- Peer-reviewed
- Attention to Copyright and proper Licensing

# Wikipedia Accuracy

*“Internet encyclopaedias go head to head”*  
Jim Giles, Nature 438, 900 - 901 (2005).

- Entries on Science Topics were taken from Wikipedia and Britannica.
- Sent to domain experts (on blind study)
- 42 Entries tested
  - 4 Average errors in Wikipedia, 3 in Britannica
  - 4 Serious errors in both Wikipedia and Britannica
  - 162 factual errors in Wikipedia, 123 in Britannica
    - (but Wikipedia articles are 2.6 longer than Britannica)

# Wikipedia - Images

- Image self-pages (author, copyright, license)
- Over 2.5 million images
- Anybody can upload more images
- Serious copyright management (GPL, CC licenses)
- Vector images and Audio recordings
- Most image are stored in Wikimedia Commons
- Avoid repeated uploads
- You can use (free) images in your own work

# Cultural Freedom

- The freedom
  - **To use the work** and enjoy the benefits of using it
  - **To study the work** and to apply knowledge acquired from it
  - **To make and redistribute copies**, in whole or in part, of the information or expression
  - **To make changes and improvements**, and to distribute derivative works
- These freedoms should be available to **anyone, anywhere, anytime.**

# Users Access Levels

- **Stewards**
  - Allow users to change rights of others
- **Bureaucrats**
  - Create admins bots, move pages, rename users
- **Admins**
  - Block users, delete pages and history, protect pages  
rollback changes...
- **Registered**
  - create articles, upload files, move articles
- **All Users**
  - Create accounts, edit, read

# Wikipedia Fauna

- **WikiElf**
  - Works behind the scenes, infrastructure maintenance
- **WikiFairy**
  - Wiki editor who beautifies and standardize articles
- **WikiGnome**
  - User who makes small incremental improvements
- **WikiOgre**
  - Users who makes huge changes in articles.

# Wikipedia Fauna

- **WikiGremlin**
  - Creature that runs a Wikipedia website
- **WikiTroll**
  - Deliberate and intentional attempts to disrupt the usability of Wikipedia
- **WikiDragon**
  - Vast contributions. Creating entire articles.
  - Bold edits

# Wikipedia Special Forces

- Counter-Vandalism Unit
- New Pages Patrol
- Recent Changes Patrol
- Random Page Patrol

# Wikipedia Sister Projects



a multilingual free  
encyclopedia  
**Wiktionary**  
[ˈwɪkʃənəri] *n.*,  
a wiki-based Open  
Content dictionary  
Wileo [ˈwɪl kəri]



# Wikimedia Commons

- Media Repository
- Maintained by volunteers
- Material reusable across Wikipedias
- Freely licensed material
  - Photographs, diagrams, animations, music, spoken text, video clips...
- Mayflower (image search engine)
- 2.6 Million pages, 2 Million media files

# Wikisource

- Online **Library** of free content publications
  - Public domain, or
  - Freely available licenses
- Historical Documents
- Translations
- Examples
  - Bible, Tao Te Ching, Britannica 1911, Jules Verne, Grimm's Brothers Fairy Tales, Allan Poe.

# Wikiquotes

- Free online compendium of Quotations
- 13,928 pages so far
- Categories
  - People
  - Proverbs
  - Films
  - TV shows
  - Literary works

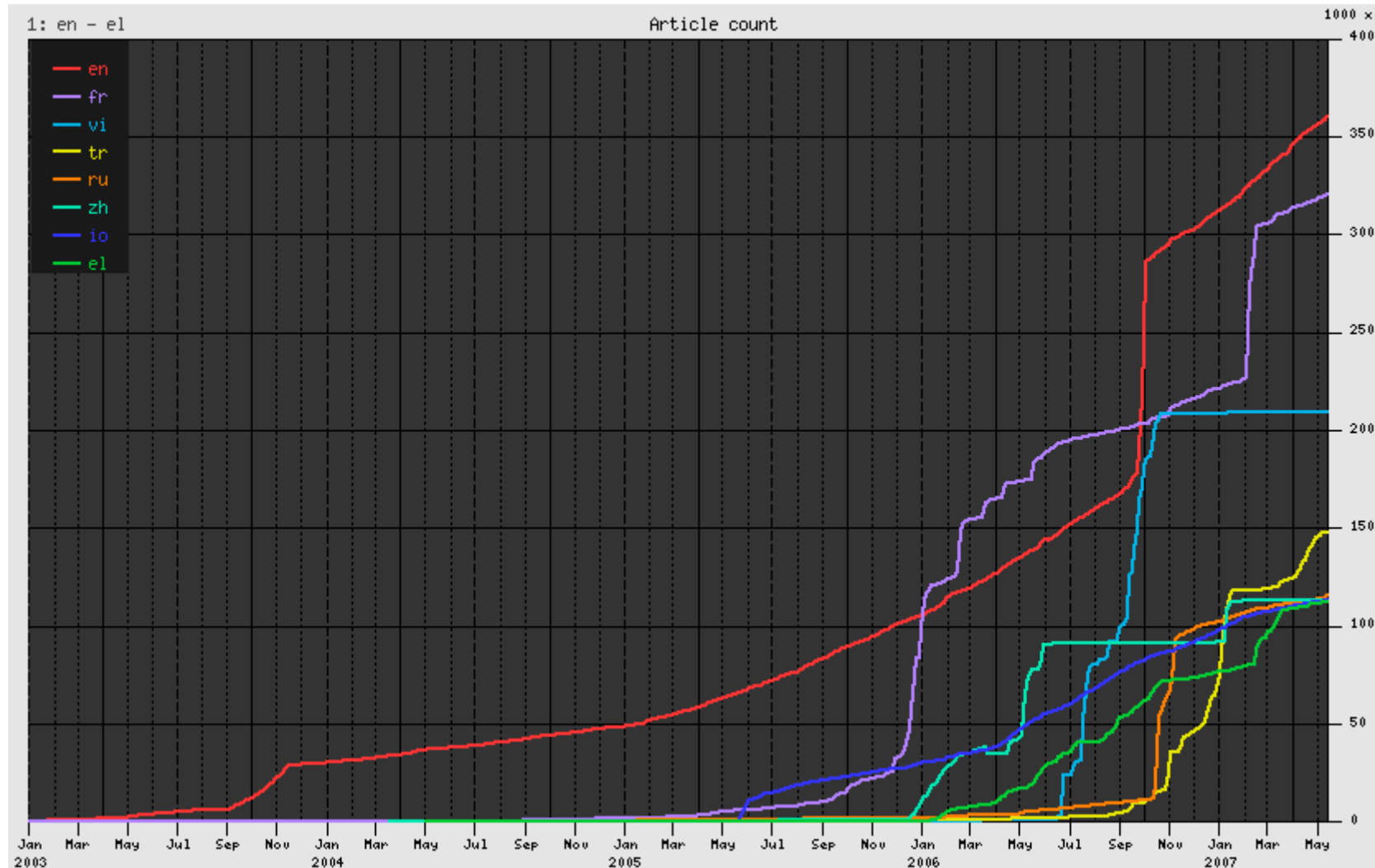
# Wikinews

- Free content news source Wiki
- Every story is written as a News reports (as opposed to an Article in the Wikipedia)
- Neutral Point of View Policy
- Started on December 2004
- By September 2007 it has 10,000 news articles
- Beyond Texts: Audio, Video
- Credibility Question
  - (but you can trust cable news... isn't it ?)

# Wiktionary

- Free content Dictionary
- Available in 150 Languages
- Written collaboratively by Volunteers
- Wikisaurus (synonyms)
- Started on December 2002
- November 2006
  - 1.7 Million entries in 171 languages
- Many of the entries are created by “bots”

# Wiktionary – Growth per Language



[http://en.wikipedia.org/wiki/Image:Wiktionary\\_growth.png](http://en.wikipedia.org/wiki/Image:Wiktionary_growth.png)

# Wikiversity

- Free Learning Materials
- Five Languages
  - English, French, German, Italian, Spanish
- Host scholarly projects and communities
- Concept of “University of the World”
- Doesn't confer Titles (Degrees)
- Learning, Teaching and Researching

<http://en.wikipedia.org/wiki/Wikiversity>

[http://en.wikiversity.org/wiki/Wikiversity:What\\_is\\_Wikiversity%3F](http://en.wikiversity.org/wiki/Wikiversity:What_is_Wikiversity%3F)

# Wikispecies

- Free content catalog of all species
- Aimed at scientists
- Started in August 2004
- By May 2007 it reached 10,000 Articles
- Support for Taxonomy relationships
- Avoids duplication with Wikipedia

<http://en.wikipedia.org/wiki/Wikispecies>

[http://species.wikimedia.org/wiki/Main\\_Page](http://species.wikimedia.org/wiki/Main_Page)

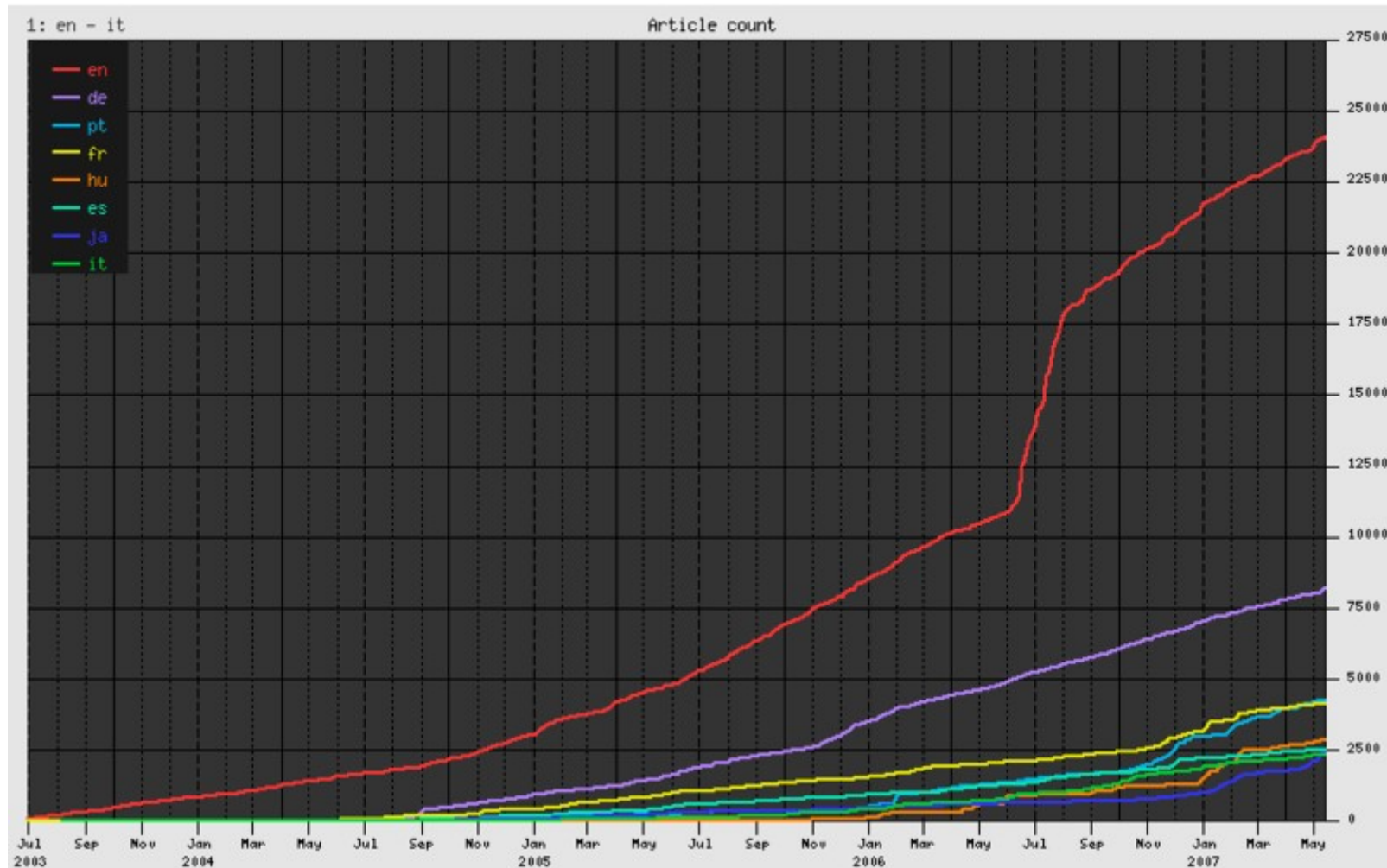
# Wikimedia - Meta-Wiki

- Coordination of all the Wikimedia Foundation Projects
- Administration
- Discussion about new and ongoing projects

# Wikibooks

- Collection of Free Textbooks
- Books directly written by contributors
- Self-publishing
- Started on July 2003
- Content available for continuous peer-review
- Anybody can edit them and improve them
- English version has 27,342 Modules

# Wikibooks - Growth



[http://en.wikipedia.org/wiki/Image:Wikibooks\\_growth.png](http://en.wikipedia.org/wiki/Image:Wikibooks_growth.png)

End