



Leverage from
the EU
2014–2020

European Union
European Regional
Development Fund

Thoughts on work-in-progress crowdsourcing metrics of digital collections

Tuula Pääkkönen

Liber 2015, London, UK – 24-26.6.2015

Information Systems Specialist

National Library of Finland /
Centre for Preservation and Digitisation



THE NATIONAL LIBRARY OF FINLAND

National Library of Finland



Centre for
Preservation &
Digitisation

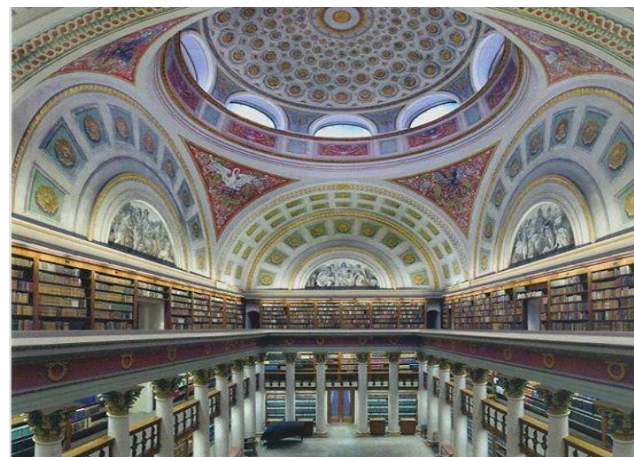
2015

2006

1990

1828

Imperial Alexander University



1707

Free deposit law

Turku Academy



Akatemiatalo. Akademibuset. The Academy. C.C. Gjørwell. 1815.

1640



THE NATIONAL LIBRARY OF FINLAND

Source: <http://www.kansalliskirjasto.fi/yleistieto/kirjastotietoutta/historia.html>

Crowdsourcing

- National Library of Finland launched the new crowdsourcing project #digitalkoot on 13.5.2014
- The idea was to let users to do clippings from the digitized collections
 - Cut interesting articles
 - Add metadata annotations
 - Collect own scrap book
 - Share these clippings via social media

Why Crowdsourcing?

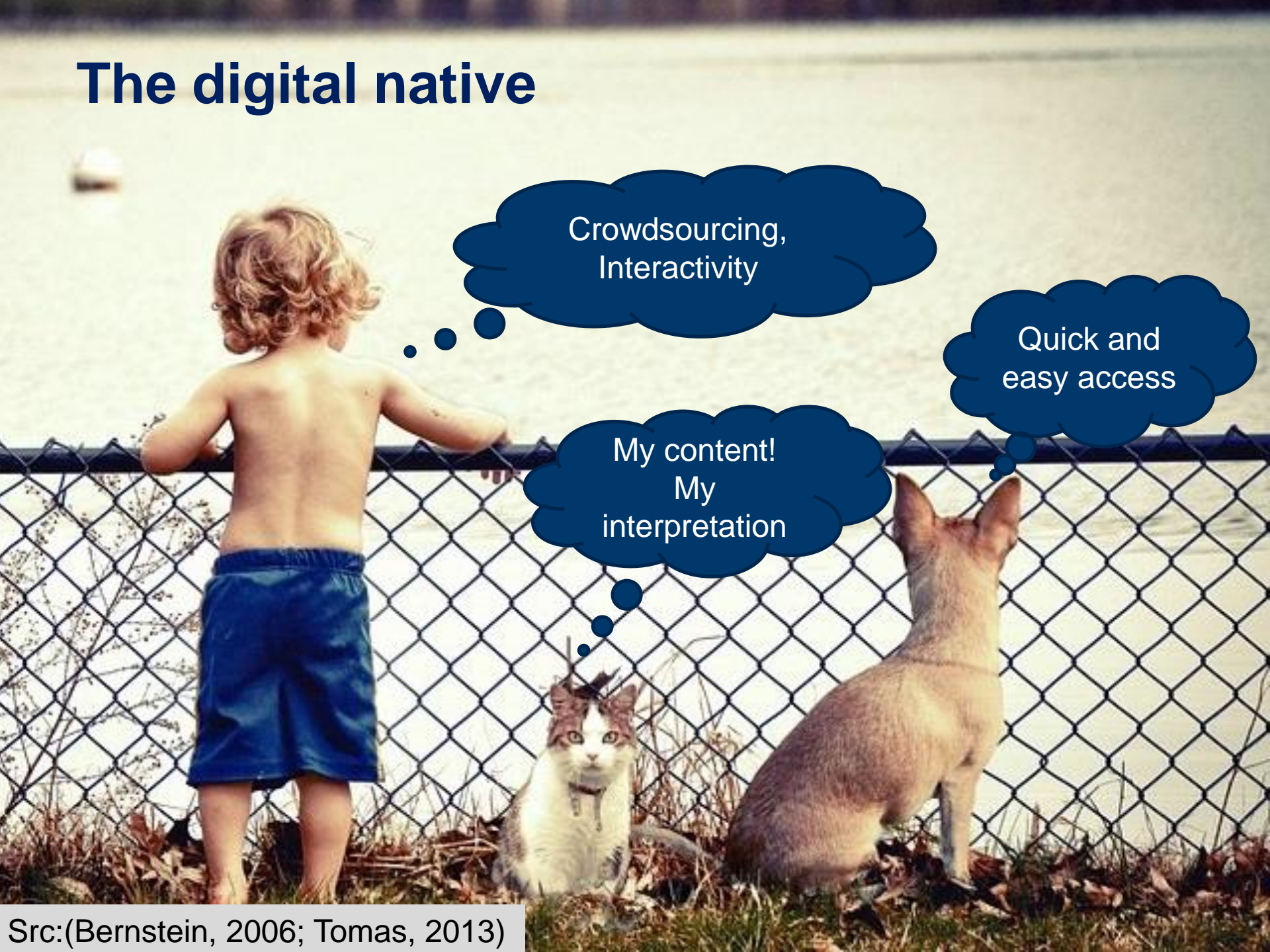
- Crowdsourcing = collaborating with undefined group/online community to complete certain tasks

Why?

- National treasures for all
- Retain existing users
- Get new users
- Dialogue



The digital native



Crowdsourcing,
Interactivity

Quick and
easy access

My content!
My
interpretation

Crowdsourcing task: clipping of article

Lauvantaina 16 p. Helmikuuta

KARJALATAR

Create a clipping



Create a clipping

...etta, „Bolle“ (6 m. 54 f.). Ruut
Nylanderin, Lemistä, „Hilpuri (6 m.
5. f.); Liuttin, Teistofesta, „Puijo“
(6 m. 13 f.); Finellin, Roffolasta,
„Bella“ (6 m. 13, f.). Ilma mainio,
10 paffastetta.

— Pitkällinen virka. Talonifantä
Adam Kouffu Lemin pitäjässä wuon-
na 1830 ifänsä kuoltua astui talossa
kuntomiehen virkaan. Nyt on hän jo
65 vuotta jota lesä kytännyt peltonsa
Ainoastaan yhtenä lesänä nuoruuden
ajalla keltatautia sairastessaan nautti
wapautta. Vielä tulewanakin lesänä
on hänen aikomus kytää jos Jumala
suo. Tosin ovat jo jäsenet jäykät
etenkin talwen aikana, eikä enää tule
työ niinkuin ennen, mutta ikääkin tu-
lee ensi lesäkuussa, jos siihen asti elää
79 vuotta. — U. S.

— Uusi kreikkalais-wenäläinen
ko. Senaatti on wahwistanut
tulliset kreikkalais-wenäläistä kirkkoa war-
ten, joka on aiottu salaneuwoos T.
Bernoffin kustannuksella rakennettawaksi
hänen omistamallensa Lintulan kirkolle
Kivennawan pitäjässä.

— Liikkeestä Merenkulkun yli
ei ainakaan vielä ole warmuutta ja-
noo U. S. Ei liene joo vielä hää-
... alluttuaan kirkonkone kaimuina.

nuorempana ollut merimiehenä mutta
16 vuotta ylsinomaan kerjännyt, kun
hänet otettiin kiinni, oli hänellä ra-
haa 1674 fr. 8 äyriä, joista 1660 fr.
tullasja. Niin pian kun hän oli saa-
nut 10 kymmen äyristä kotoon, waihtoi
hän ne hopeakuuniksi ja kun
oli saanut 20 kruunua waihtoi hän
ne kultarahaksi. Hänen tarkoituksen-
sa oli ollut, ilmoitti hän, saatuun
20,000 fr. kotoon „wetäytyä liikkes-
tä ja elää keroillaan“.

Sunnutai-lutemista.

„Eloa tosin on paljo, mutta
työwäkeä on vähä; rukoilkaa siis
elon Herraa lähettämään työwäkeä
eloonsa!“

Räiden Wapauttajamme sanain to-
tuuden tuntee jolainen ajattelewa
kristitty. Useimmat ihmiset seura-
kunnissamme mielellään kuulewat Ju-
malan sanaa ja tunnustawat, että
huolta olisi pidettävä kuolemattomas-
ta sielusta ja olisi päästäwä yhtey-
teen Jumalan kansa. Mutta ihmis-

oittaa, että
Jumalalle,
i ja maail-
... tämä tohittaa sitä että Herr...

Title *

Uusi kreikkalais-wenäläinen kirkko

Category *

Article or news item

Subject

Keywords

Lintulan luostari – Heinävesi

Kivennapa

Cancel

Save

<http://digi.nationallibrary.fi>

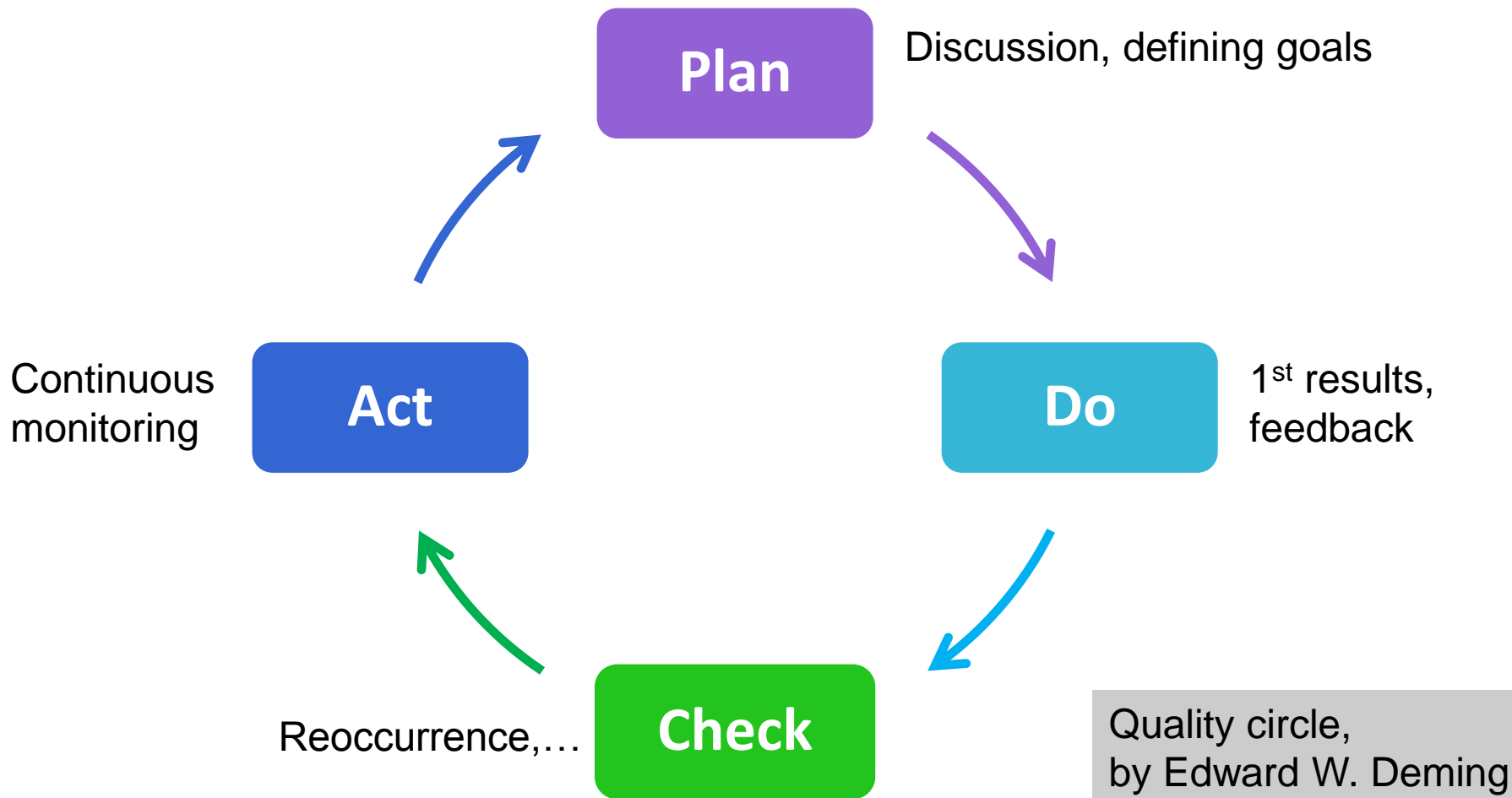
Kiinan ja Japanin
sota.

Wasta perjantaina sähötettiin
häistä että meriliinnoitukset
loitettu. Kiinalainen laim
joko kotonaisuudessaan häw

The background is a blurred screenshot of a library website. It shows a search bar at the top, a navigation menu, and several book covers or search results. The text is out of focus, but some words like 'Library' and 'Search' are faintly visible.

Digi.nationallibrary.fi – Metrics development

Process of Metrics Development



The Metrics

1

Quantity of clippings (by users & overall)

2

Keywords used in annotations

3

Feedback from users

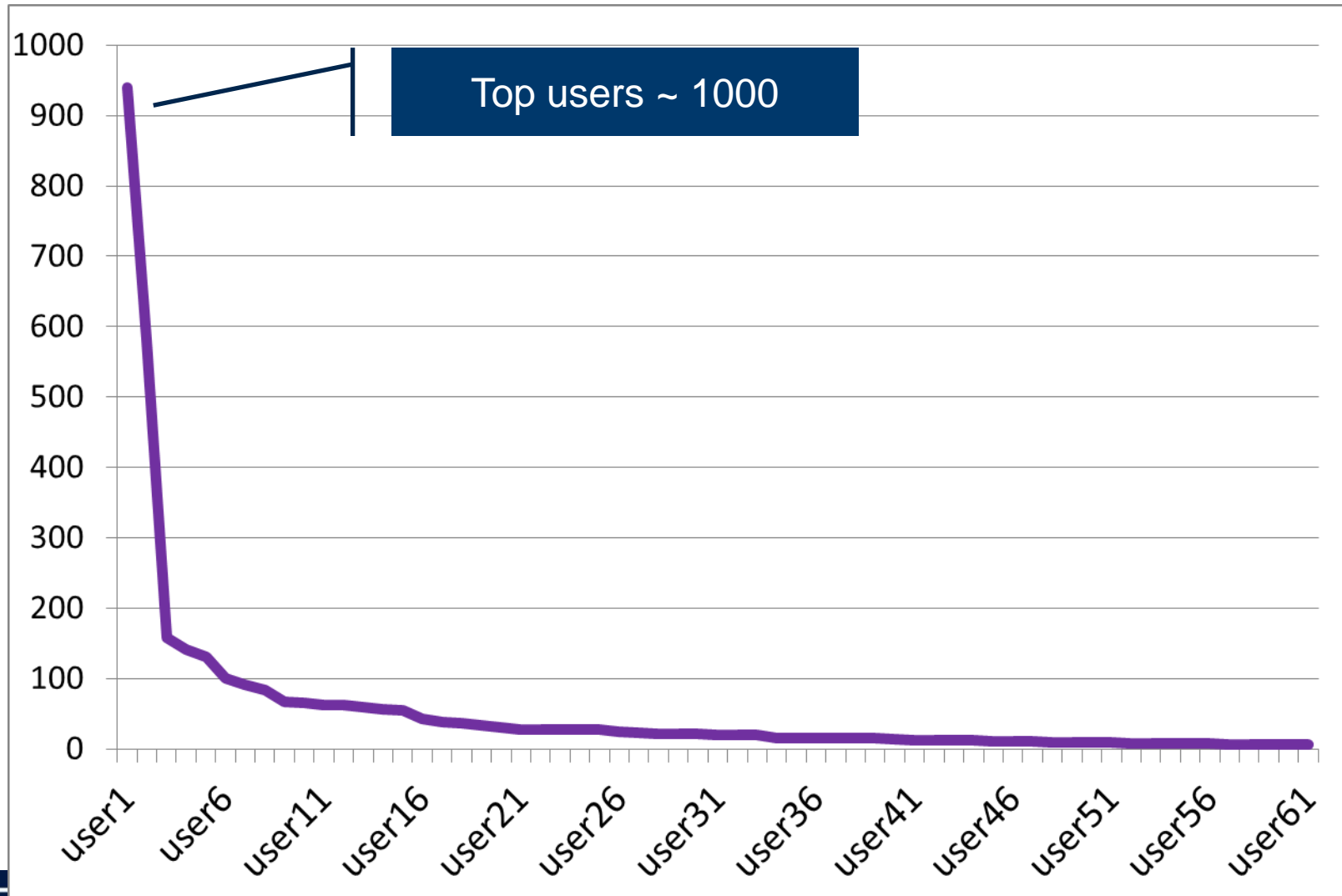
What happened?

- Over 17.000 clippings in newspapers
- Over 2000 clippings in journals
- Over 200 clippings in ephemera

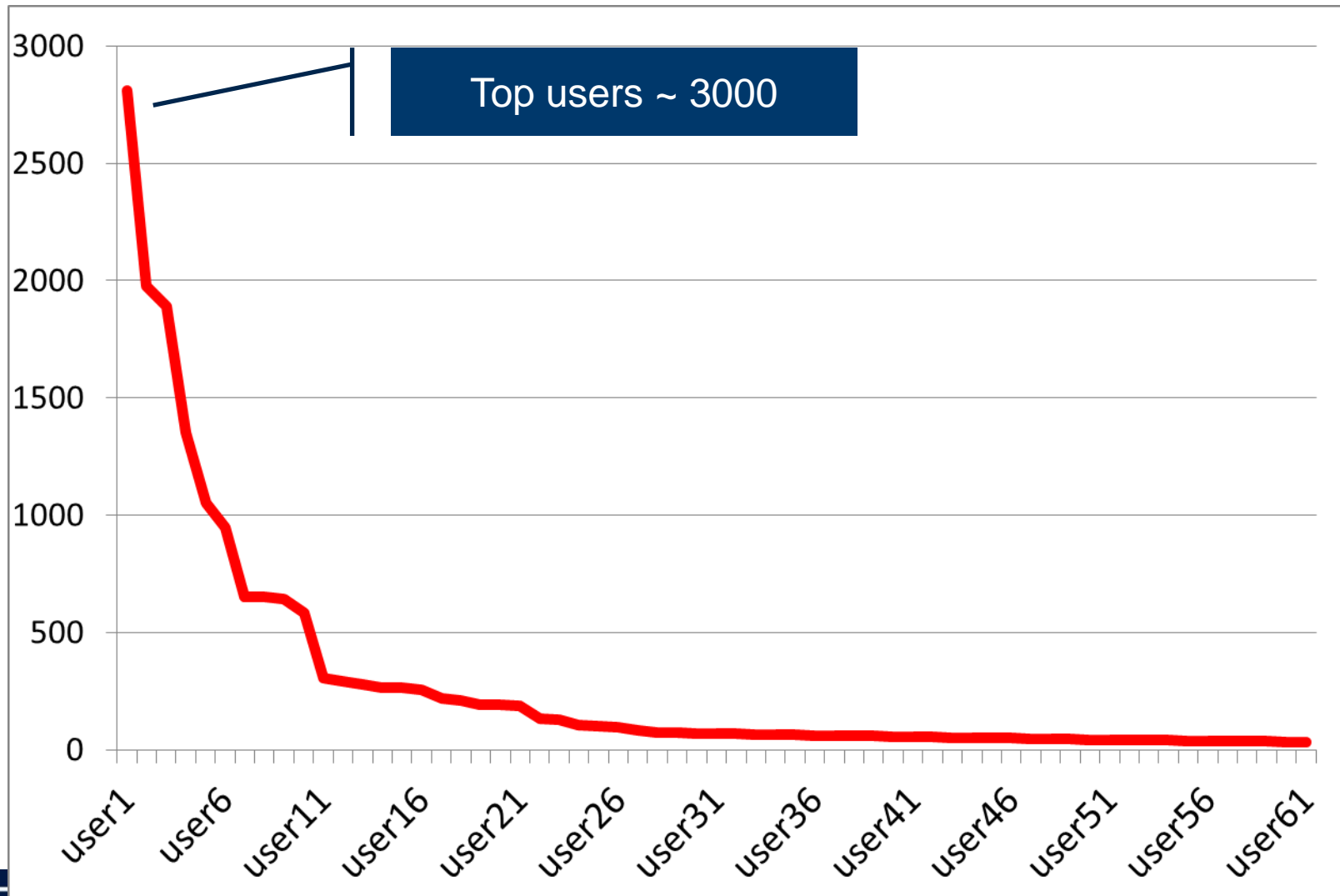
Can we see the 1% rule?

- The rule of thumb for an online community states:
 - 1% Creators
 - 9% Contributors
 - 90% The rest
- Source:
http://en.wikipedia.org/wiki/1%25_rule_%28Internet_culture%29
- In our case:
 - Hyperactive users, registered users, searchers

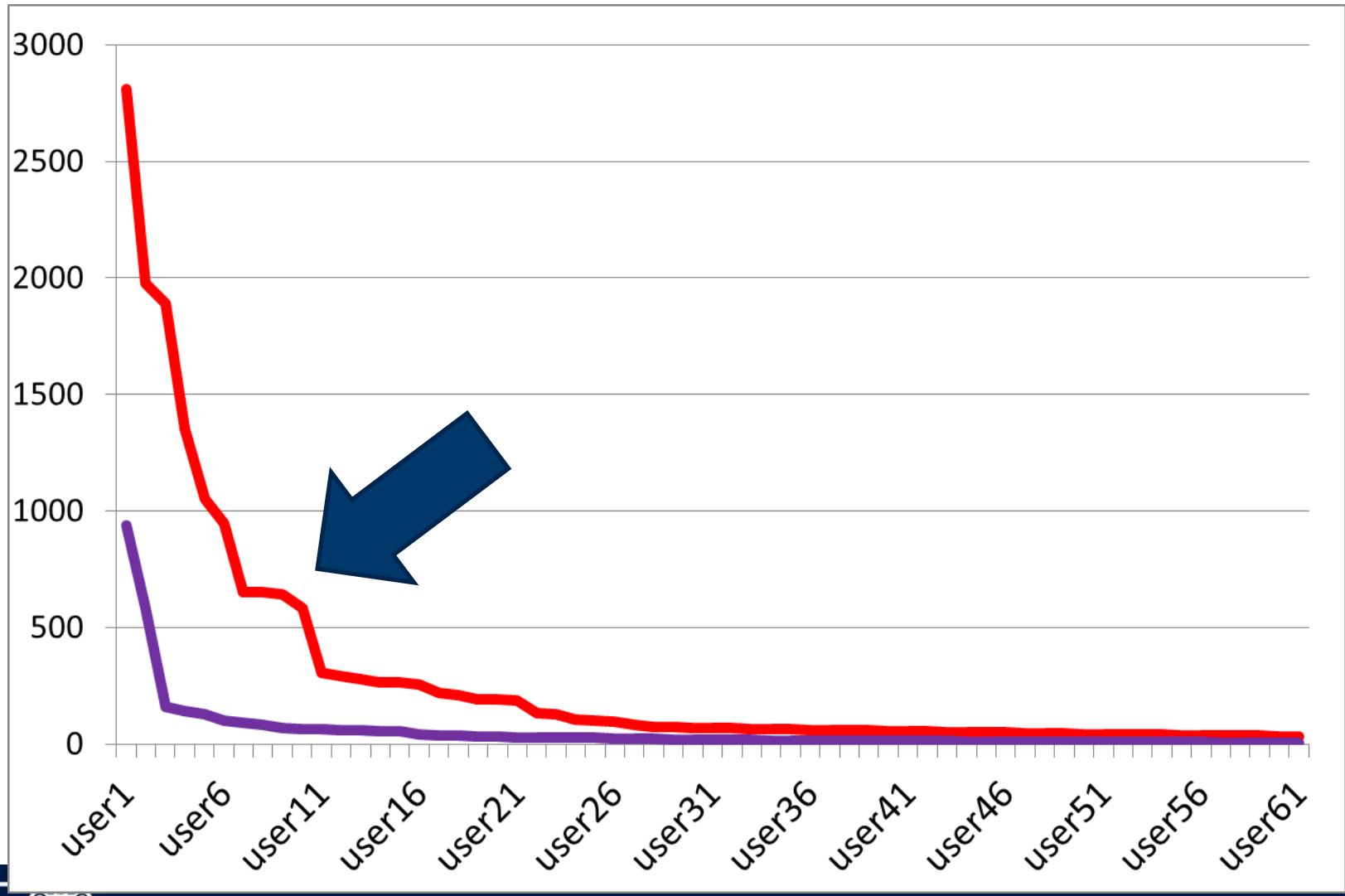
Number of clippings, first 6 months



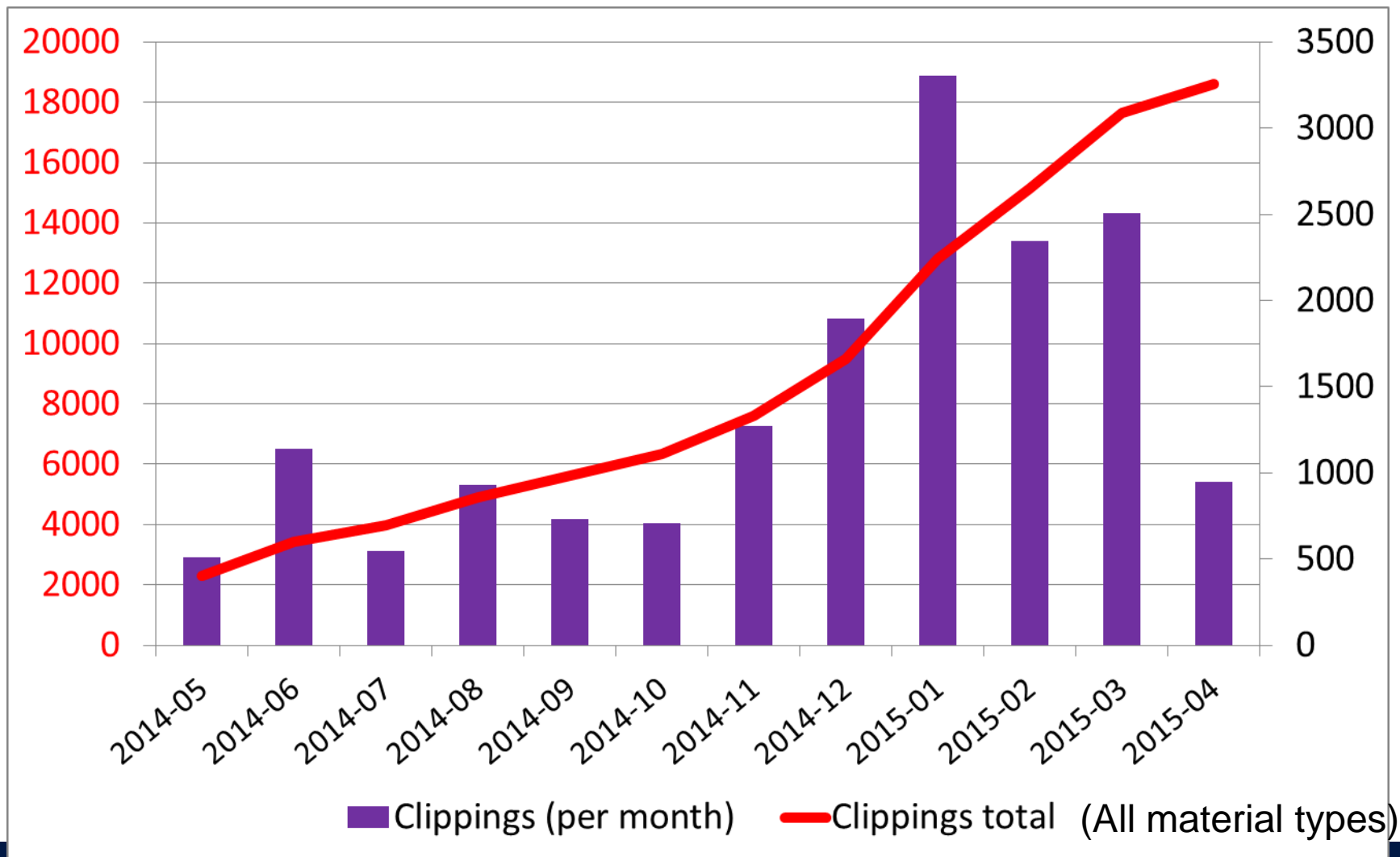
Number of clippings, 12 months



Comparison



Clippings Quantities over time

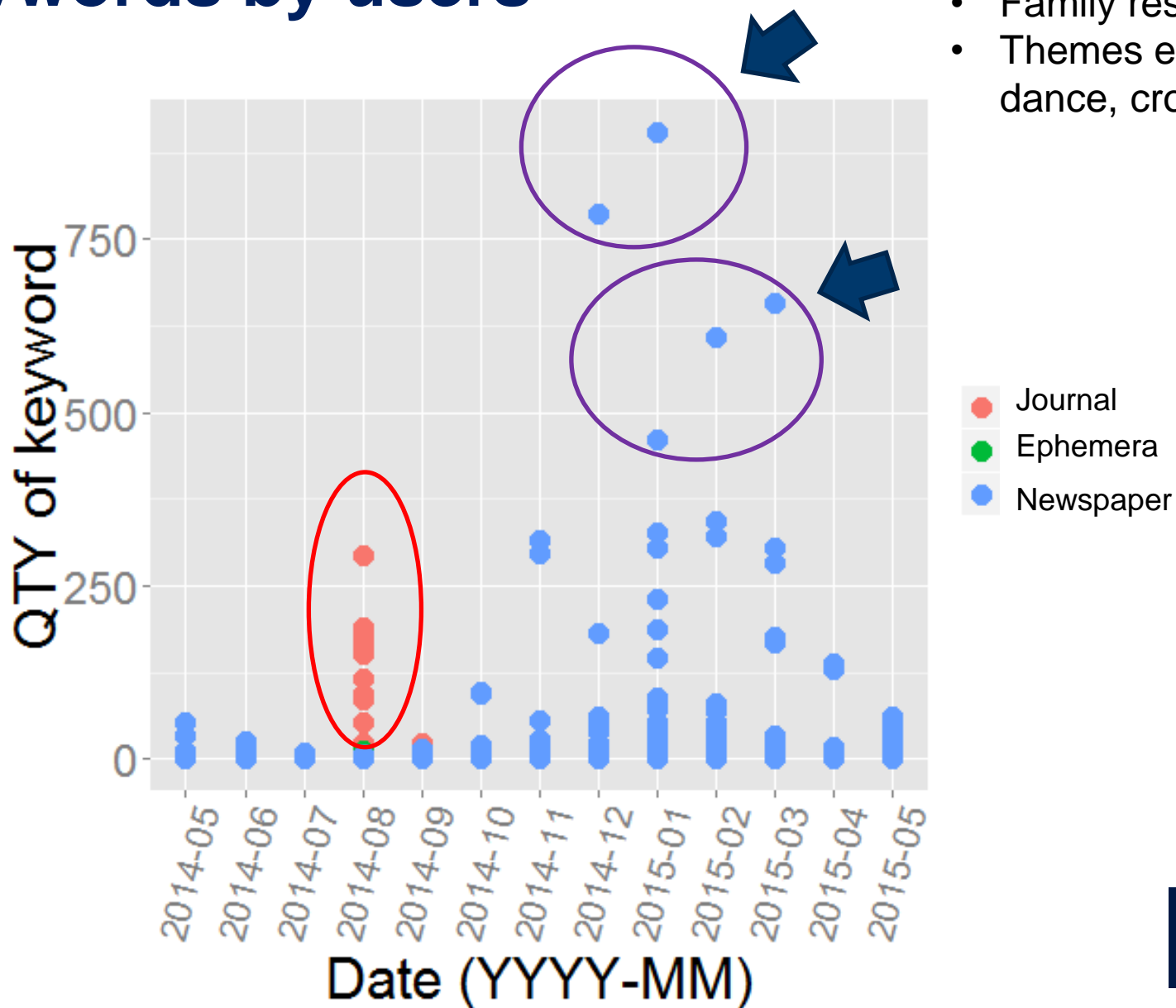


It's not just about clippings quantity

- Nearly 6000 different keywords used
 - ~25% has been picked from the general Finnish thesaurus
- 1300 different place names
 - ~42% from general Finnish thesaurus
- 16.000 keywords in total
 - for 19.000+ clippings (all material types))

Keywords by users

- Local historians
- Family researchers
- Themes e.g. car , dance, crosswords



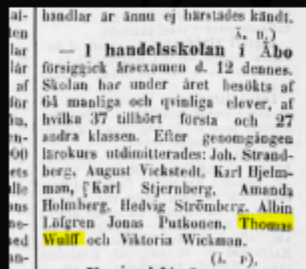
“Look, I’ve made this!”

Wulffin tarinoita

125-vuotisjuhlavuoden kunniaksi

Elämää Esplanaadilla

Olipa kerran erittäin tunnollinen Wulff. Hän työskenteli 1880-luvulla ja konttoristina. Liike oli kuuluisaa jota kutsuttiinkin Edlundin kulmak



Ilmoitus Morgonbladet-lehdessä 17.6.1878



Ehdotus sukupuuttoon metsästetyn #majava'n palauttamiseksi @hameensanomamat vuodelta 1883. [digi.kansalliskirjasto.fi/sanomalehti/bi ...](https://digi.kansalliskirjasto.fi/sanomalehti/bi...) #digitalkoot



Follow



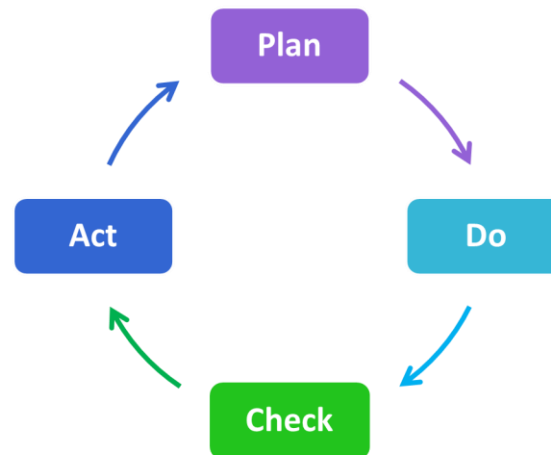
... 22-vuotiaana, vuonna 1878, ... kassaan. Helsingissä vallitsi yleinen ... ja asukasluvu oli kasvussa. Thomas todisti historiallisten palkkojen ja rakennusten syntyä. Hän oli paikalla, kun Runebergin patsas paljastettiin. Hän näki omin silmin hotelli Kämpin rakennusvaiheet. Hän asui ja työskenteli kaasulyhtyjen ja hevossaitiotien ajan Esplanaadilla, eikä ehkä vielä silloin aavistanut, että hänenkin nimensä tulisi painettua historiankirjoihin.



Thomas Wulff nuorena

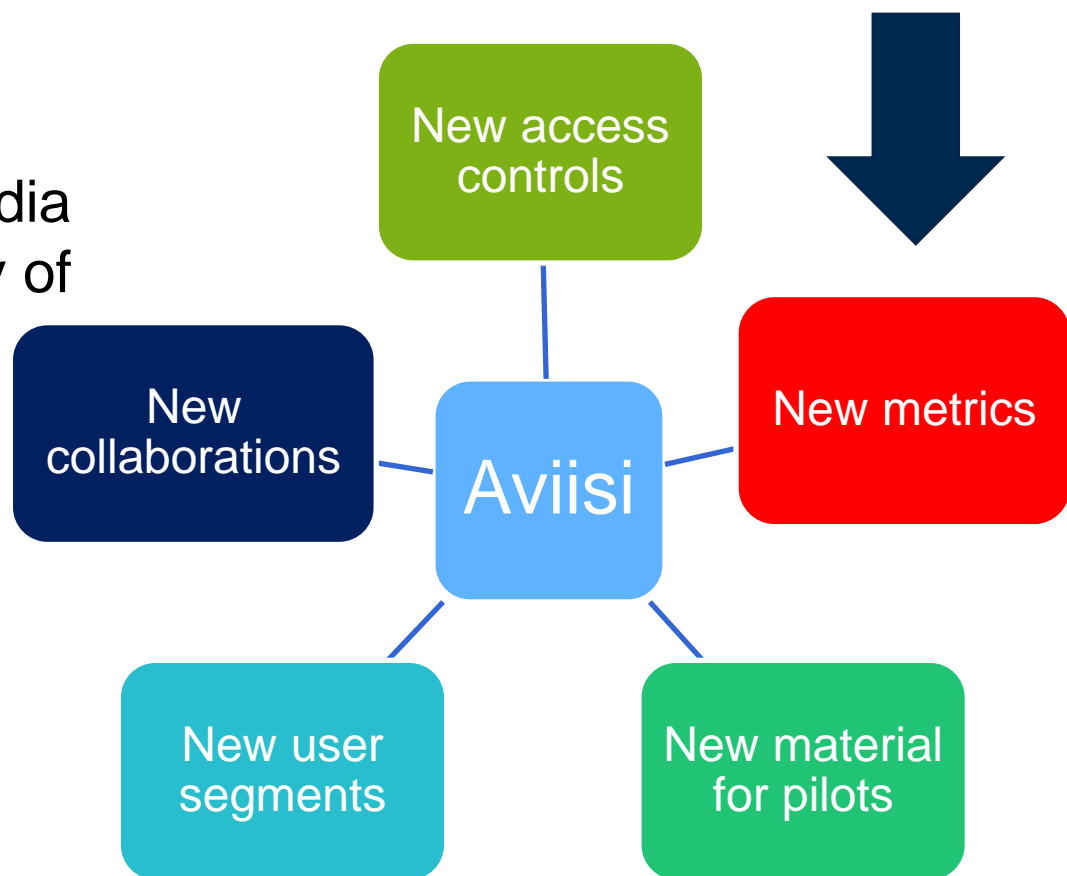
Source: <https://wulffinkulma.wordpress.com/1880-luku/>

More plans for metrics...



Newspapers to new purposes - Aviisi

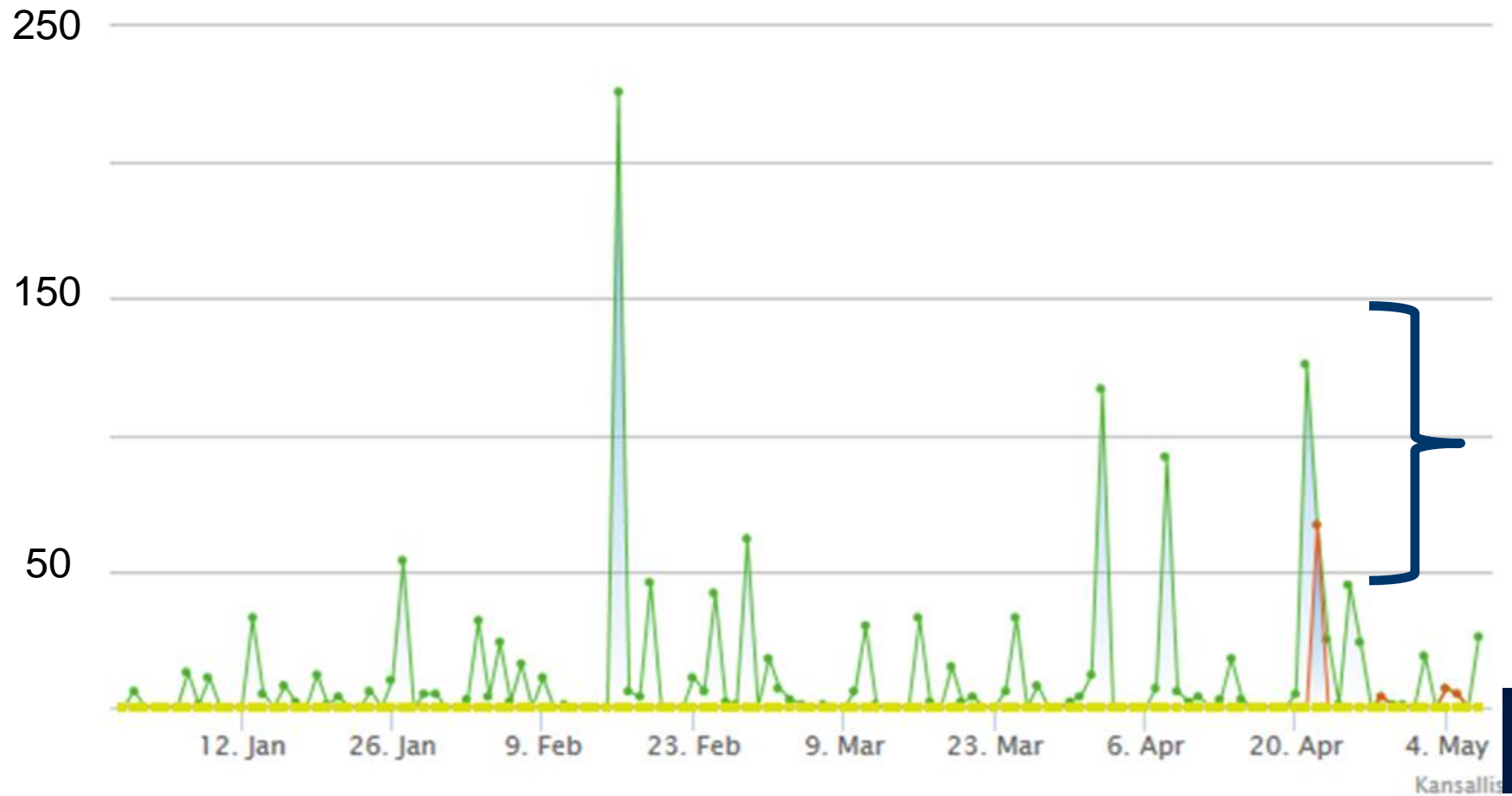
- Digitized collections to new usage purposes
- Collaboration with news media houses and National Library of Finland
- Page views+clippings



<http://blogs.helsinki.fi/digiaviisi>

Metrics for news media & library

Page views



To the actual pages

ISSN	Title	Views	Date	No	Page	Link
1458-2252	Kotkan Sanomat	2	27.12.1883	52	1	Linkki-SAN
1458-2252	Kotkan Sanomat	2	19.01.1887	3	1	Linkki-SAN
1458-851X	Mikkeli	1	13.03.1905	30	4	Linkki-SAN
1458-2252	Kotkan Sanomat	1	17.12.1903	145	1	Linkki-SAN
0356-1623	Länsi-Savo	1	12.01.1965	10	2	Linkki-SAN
0356-1623	Länsi-Savo	1	12.01.1965	10	5	Linkki-SAN
0356-1623	Länsi-Savo	1	26.08.1985	228	3	Linkki-SAN
1458-6592	Savo	1	28.04.1910	48	3	Linkki-SAN
1458-2481	Mikkelin Sanomat	1	28.04.1910	46	3	Linkki-SAN
1458-851X	Mikkeli	1	13.12.1892	00	1	Linkki-SAN

TEXT SEARCH CLIPPINGS TITLES PUBLISHERS

Titles / Pohjantähti : raittiin opiskelevan nuorison äänenkannattaja / 1910 / 01.01.1910 Pohjantähti : raittiin opiskelevan nuorison äänenkannattaja no 1
/ Article: Oi, muistatko vielä sen virren?

Oi, muistatko vielä
Eino Leino.

Hartaasti. (S. O. N. R:n laulukirjan n:o

p
Oi, muis-tat-ko vie-lä sen vir-ren, J

mf
Kun yö li-ki ik-ku-nan liik-kui,

mf
Se an-toi rin-ta-han rau-haa,

p
Jos muis-tat vie--lä sen vir-ren, Nyt lau--la,

Oi, muistatko vielä sen virren?

Like Share 0

Tweet 0

W BIB TEX

Publication
01.01.1910 Pohjantähti : raittiin opiskelevan nuorison äänenkannattaja no 1 s. 8

Category
Poem, song, sheet music

Subject
Human interest

Authors
Eino Leino, Oskar Merikanto

Show page

PUBLIC COMMENTS

No comments.

Login to comment

Source: <http://digi.kansalliskirjasto.fi/aikakausi/binding/909896/articles/116316>

Further metrics development

- Researcher use is especially interesting metric
- But what to measure?
 - Citations?
 - Actual filesystem level access of newspaper pages?
 - Amount of research projects?

The emergence of Socialist terminology in Finnish newspapers 1895 – 1910

Research Question
What are the key concepts of Finnish socialism from 1895 to 1910? What is specific in this political language? How does the vocabulary of socialism change in time and space?

Data
Our primary source is the newspaper and periodical corpus of the National Library of Finland. The corpus includes all the Finnish newspapers and magazines from 1820 to 1910.

Methods
We analyze keywords by computing word frequencies. Socialist keywords are detected by finding the words that are more frequent among socialist newspapers compared to the whole corpus of newspapers.

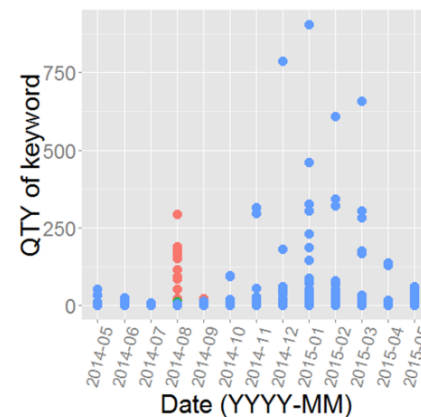
Picture source:

<http://dayofdh2015.uned.es/fnewspapers/2015/05/19/dhh-15-the-key-concepts-of-socialism-in-finnish-newspapers-1895-1910/>

In Conclusion

- Collect data for metrics from early on
- Crowdsourcing requires time
- Recognize the value that online community members create (Cothrel, 2000)

It's FUN to see what the users' come up with!



Tuula.Paakkonen@helsinki.fi

Thank You!

Tuula.Paakkonen@helsinki.fi

Feedback : <http://bit.ly/1GhDvsV>

<http://digi.nationallibrary.fi>



THE NATIONAL LIBRARY OF FINLAND

References

- Bernstein, J. 2006. 'Born Digital, Not Yesterday: Next-generation Web User Seeks Interactivity', *Econtent*, 29, 4, p. 14, Education Research Complete, EBSCOhost, viewed 22 March 2015.
- Joseph P. Cothrel, (2000), "Measuring the success of an online community", *Strategy & Leadership*, Vol. 28 Iss 2 pp. 17 – 21, <http://dx.doi.org/10.1108/10878570010341609>
- Tomas, K. 2013. Virtual Reality: Why Magazines Should Adopt a Mobile-First Publishing Strategy. *Publishing Research Quarterly*, 29(4), 301-317. doi:10.1007/s12109-013-9330-7