

Number game



–Experience of a European research infrastructure (CLARIN) for the analysis of web traffic

Go Sugimoto (Austrian Center for Digital Humanities)
CLARIN Annual Conference 2016, Aix-en-Provence, France
2016-10-27

Motivation -Raise Awareness

1. User centric R&D with marketing research
2. Open movement of R&D services
3. Present our situation
4. Recommend you to take actions immediately

1. Users and Research Infrastructure

“Research Infrastructures should have a policy defining how they regulate, grant and support Access to (potential) **Users from academia, business, industry and public services”**

(European Charter for Access to Research Infrastructures Principles and Guidelines for Access and Related Services. The European Union 2016)

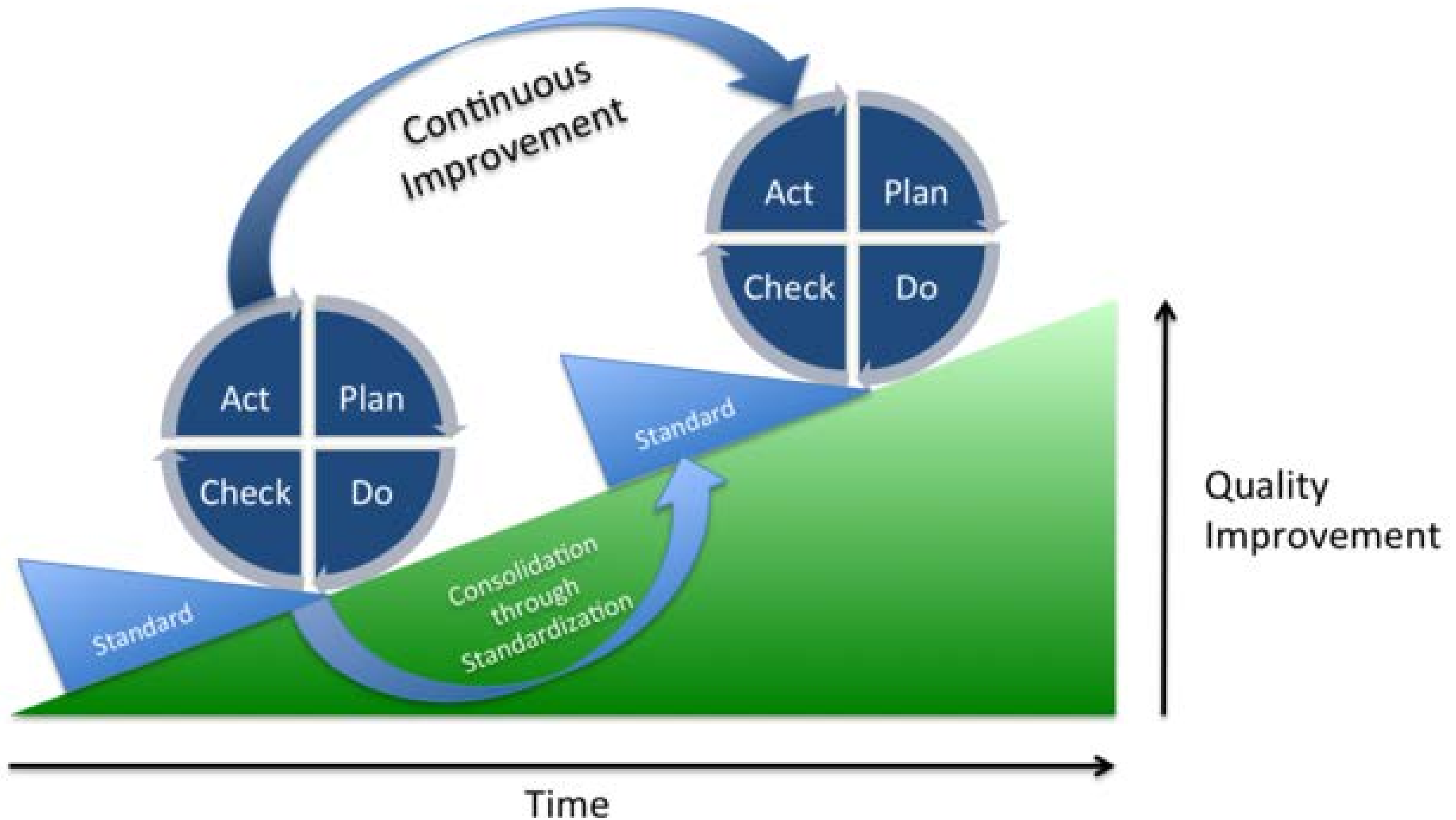
Previous studies

- Eckart et al., (2015)
- Goosen & Eckart, (2014)
- Haaf et al., (2014)
- Wynne (2015)

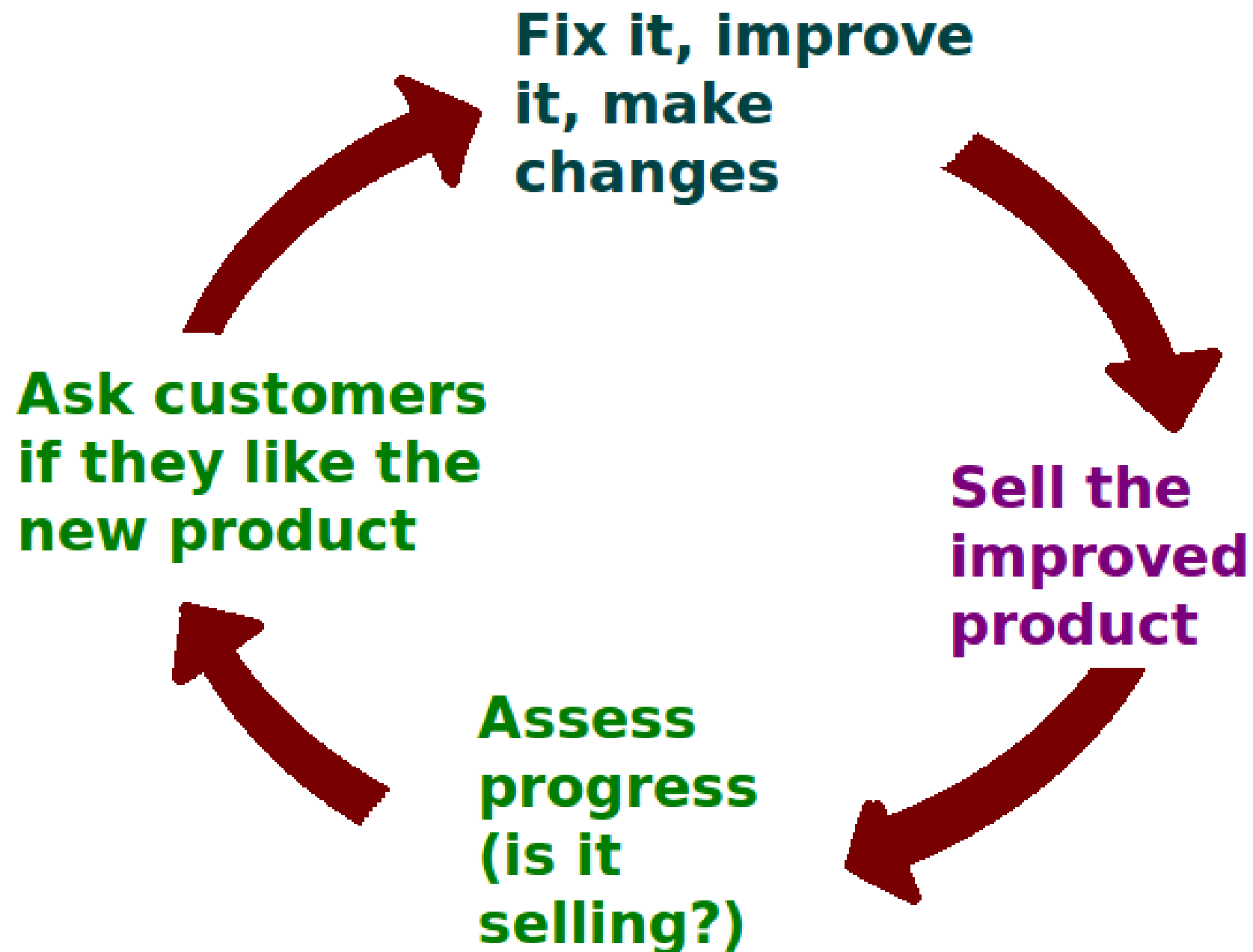


Jambo! Am I a user too?

Development Cycle (PDCA)



Cooperation between Development (software/services etc) and Marketing (Outreach/Dissemination/Communication/User engagement etc)



2. Trend for Measurable, Open, and Transparent Research & Development

Open Source, Open Data, Open Access, Open Science, ...
Open Evaluation

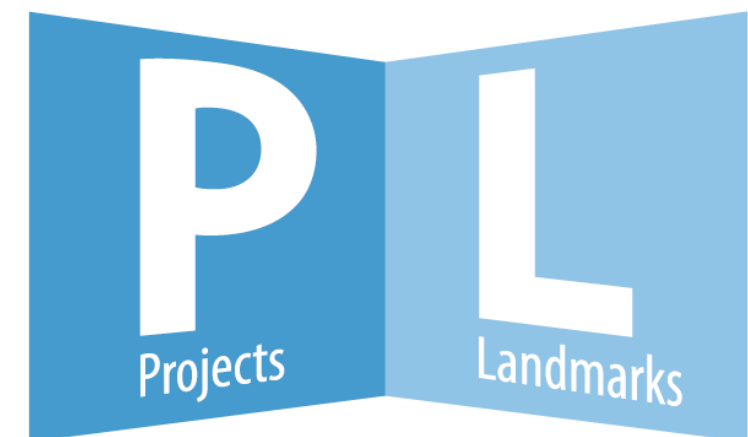
1) for users

2) for CLARIN:

Key Performance Indicators, Return Of Investment

“The ESFRI Landmarks need continuous support for successful completion, operation and upgrade in line with the **optimal management and maximum return on investment.**”

(ESFRI Roadmap 2016)



ROADMAP 2016

Public health: The toxic truth about sugar

Overview of attention for article published in Nature, February 2012



About this Attention Score

In the top 5% of all research outputs scored by Altmetric

MORE...

Mentioned by

- 14 news outlets
- 26 blogs
- 1606 tweeters
- 50 Facebook pages
- 1 Wikipedia page
- 18 Google+ users
- 1 video uploader

Readers on

- 515 Mendeley
- 6 CiteULike

SUMMARY

News

Blogs

Twitter

Facebook

Wikipedia

Google+

Video

Title Public health: The toxic truth about sugar
Published in Nature, February 2012
DOI 10.1038/482027a [↗](#)
Pubmed ID 22297952 [↗](#)
Authors Robert H. Lustig, Laura A. Schmidt, Claire D. Brindis, Lustig RH, Schmidt LA, Brindis CD

[↗](#) View

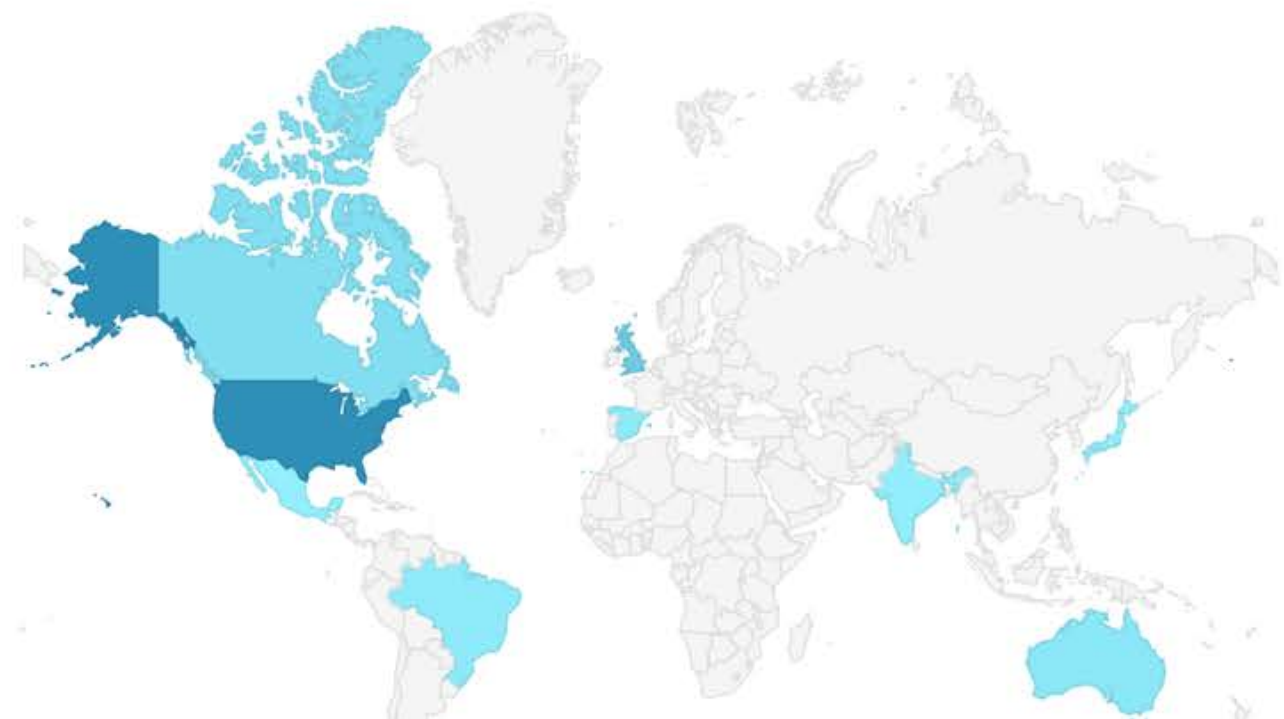
[✉](#) Alert

TWITTER DEMOGRAPHICS

MENDELEY READERS

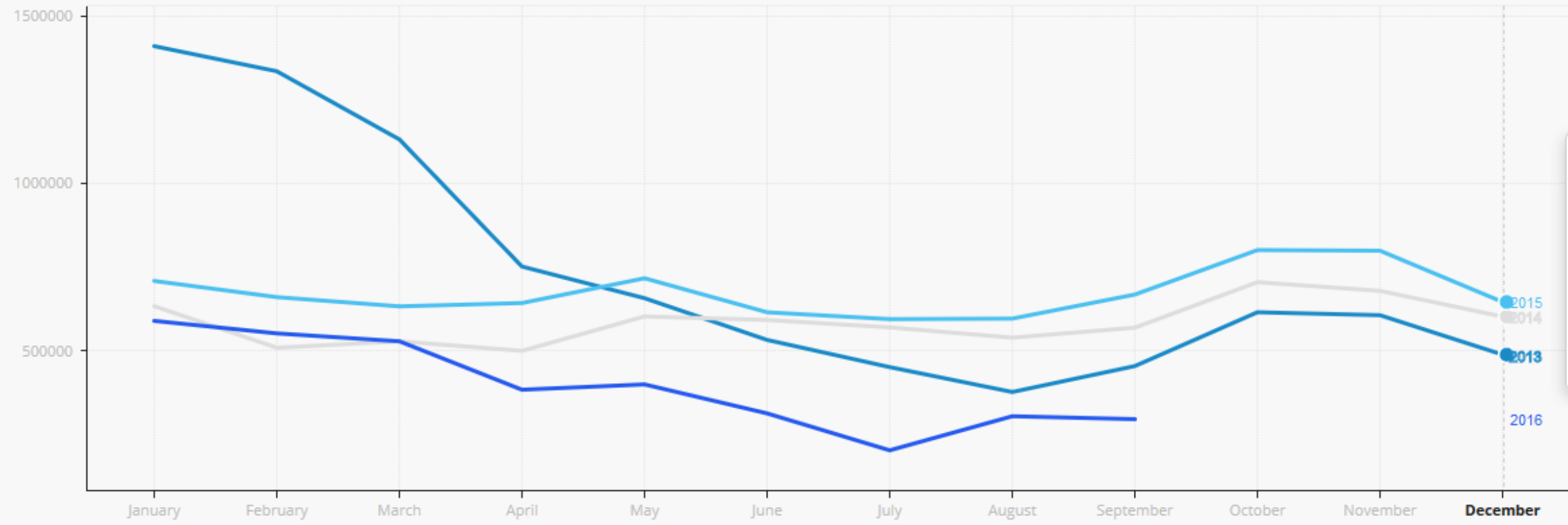
ATTENTION

The data shown below were collected from the profiles of 1,606 tweeters who shared this research output. [Click here to find out more about how the information was compiled.](#)



Altmetrics

Europeana Monthly page views of items

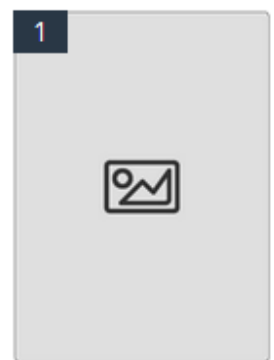


December	
2013	489,543
2014	603,182
2015	646,995
2016	NA

Total views in 2014 7,042,603	Total views in 2015 8,093,788 ▲ 14.93%	Total views in 2016 (till now) 3,578,559	Total click-throughs in 2014 1,939,062	Total click-throughs in 2015 1,853,953 ▼ 4.39%	Total click-throughs in 2016 (till now) 48,525
-----------------------------------------	-----------------------------------------------------	----------------------------------------------------	--------------------------------------------------	-------------------------------------------------------------	----------------------------------------------------------

TOP 25 ITEMS

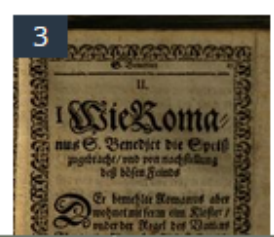
SHOW RESULTS FOR OF



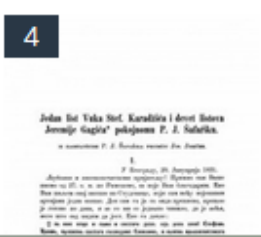
1 "Demonteerbare beeldhouwwerken in van Berrocal in Parijs, "" Hommage a Picasso
Views: 55332



2 Homicídios - Cabo mata militares em formatura no quartel da GNR da Ajuda
Views: 3146

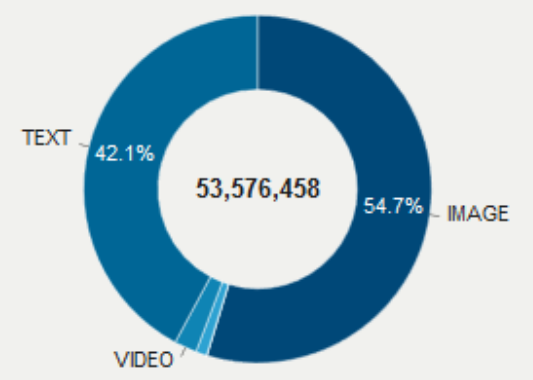


3 Vita & regula S. P. N. Benedicti. Latinae Ecclesiae Confessoris & Abbatis Ac Monachorum In Occidente Patriarchae Sanctissimi
...



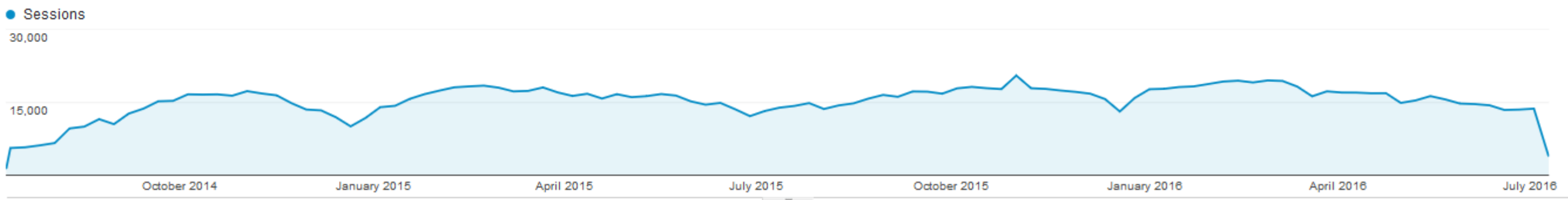
4 Jedan list Vuka Stef. Karadžića i devet listova Jeremije Gagića pokojnomu P. J. Šafariku
Views: 2112

TOTAL NUMBER OF ITEMS

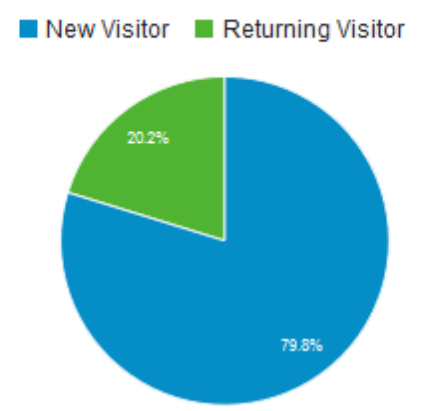


AVAILABLE FOR RE-USE





Sessions 1,611,542	Users 1,290,774	Page Views 6,754,088	Pages/Session 4.19	Avg. Session Duration 00:02:15
Bounce Rate 57.49%	% New Sessions 79.67%			



Demographics

- Language
- Country**
- City
- System
- Browser
- Operating System
- Service Provider
- Mobile**
- Operating System
- Service Provider

Country	Sessions	% Sessions
1. Netherlands	282,999	17.56%
2. France	170,237	10.56%
3. Germany	167,703	10.41%
4. Poland	166,764	10.35%
5. Estonia	133,940	8.31%
6. Spain	112,283	6.97%
7. Sweden	85,810	5.32%
8. Finland	54,181	3.36%
9. Denmark	44,296	2.75%
10. Belgium	43,759	2.72%

Archives Portal Europe (archival aggregator for Europeana)
 11 June 2014-11 June 2016

3. Let's Check - understanding our users

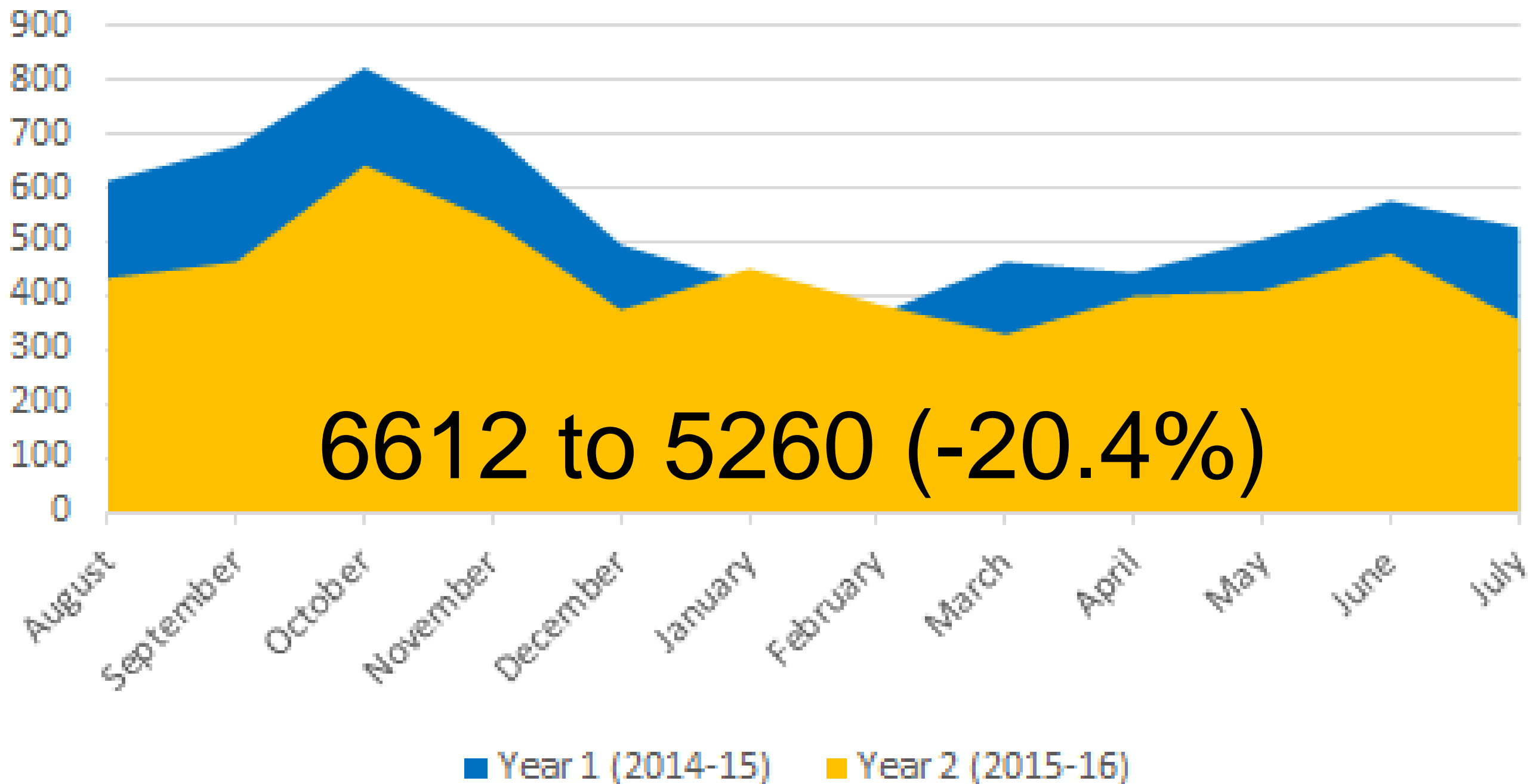
- Questionnaire (On website, e-mail, offline)
- Interview
- Observation
- Focus Groups
- User feedback /Help desk
- **Web Analytics (Website, FB, Twitter...)**
- etc



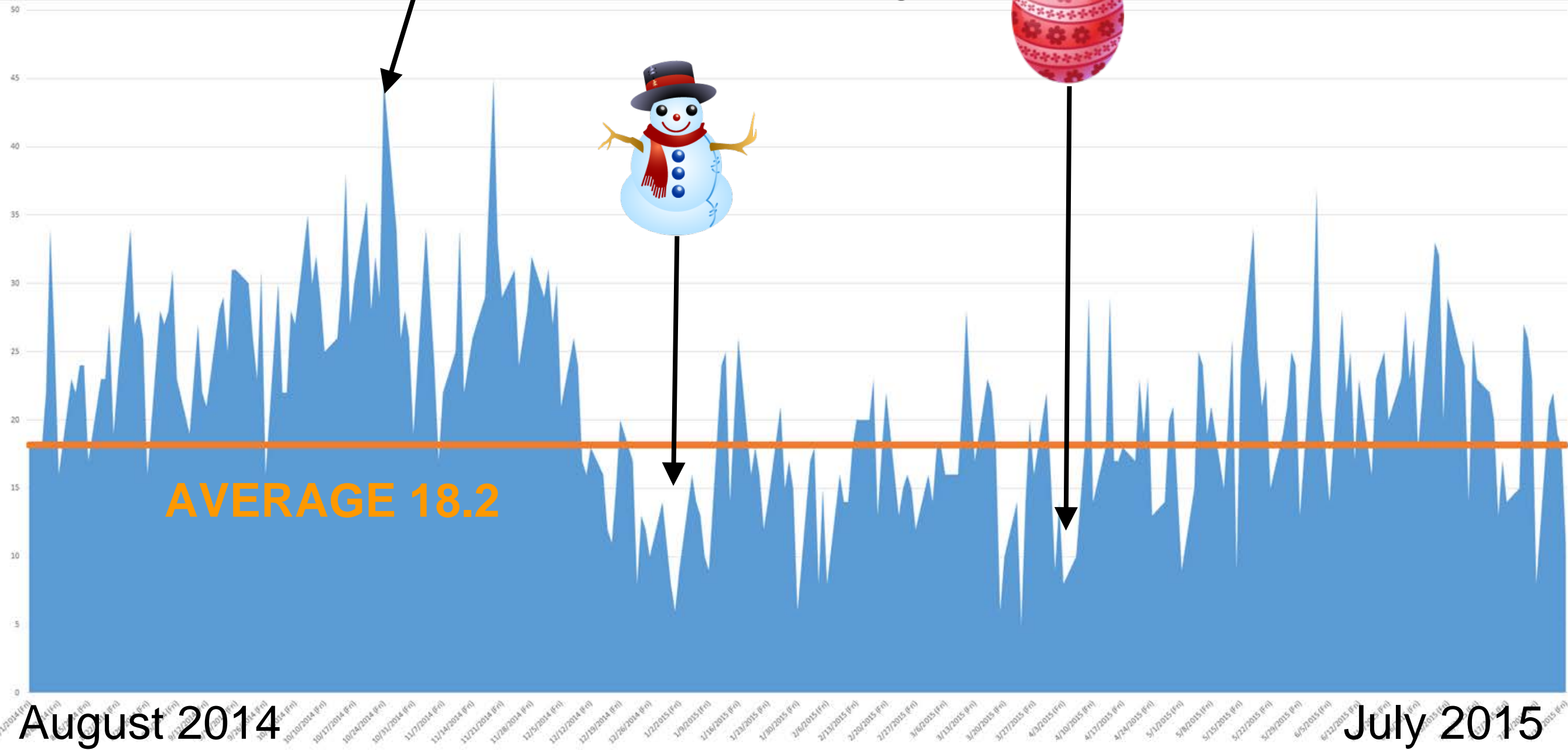
1 August 2014 - 31 July 2016 (2 years)

- * **Virtual Language Observatory (VLO)**
- * Weblicht
- * BAS
- * Federated Content Search
- * Identity Provider
- * Discovery Service

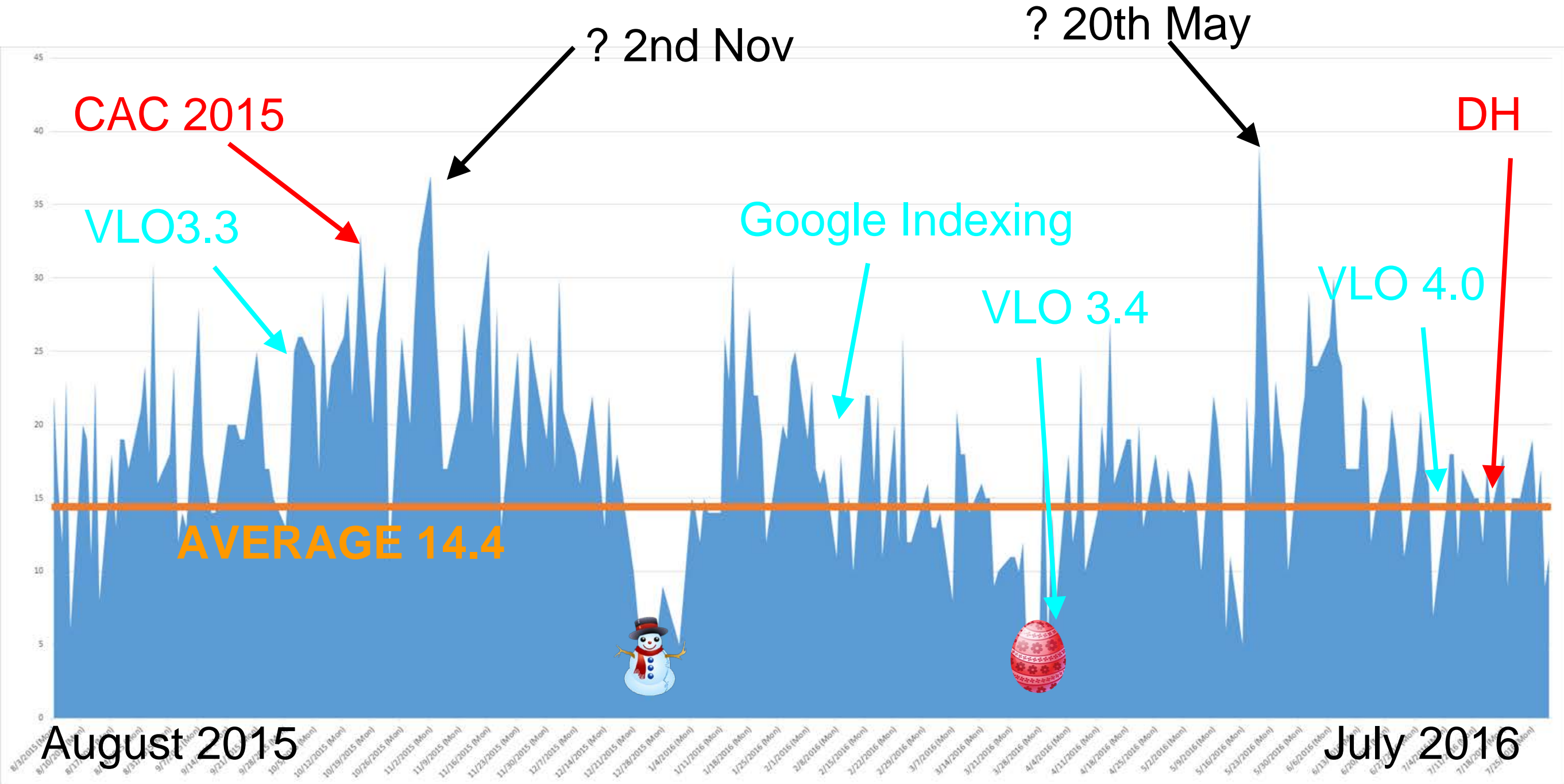
Virtual Language Observatory (VLO) unique visits



CAC 2014 Soesterberg



Year 1 Unique Visitors (Weekdays)



Year 2 Unique Visitors (Weekdays)

Google indexing to increase the web traffic

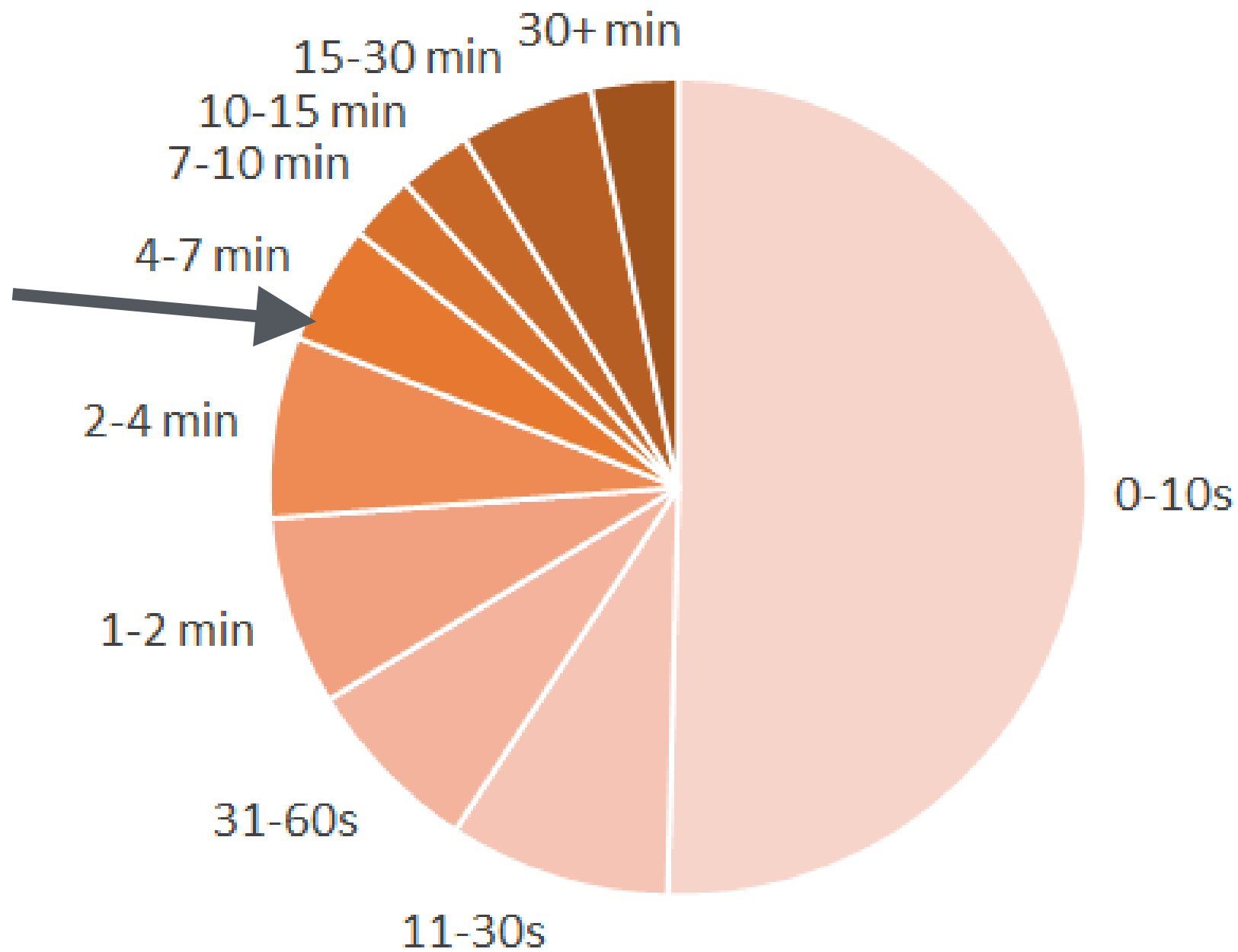
February 2016 ACDH generated sitemaps to crawl VLO
But, Google doesn't like them:

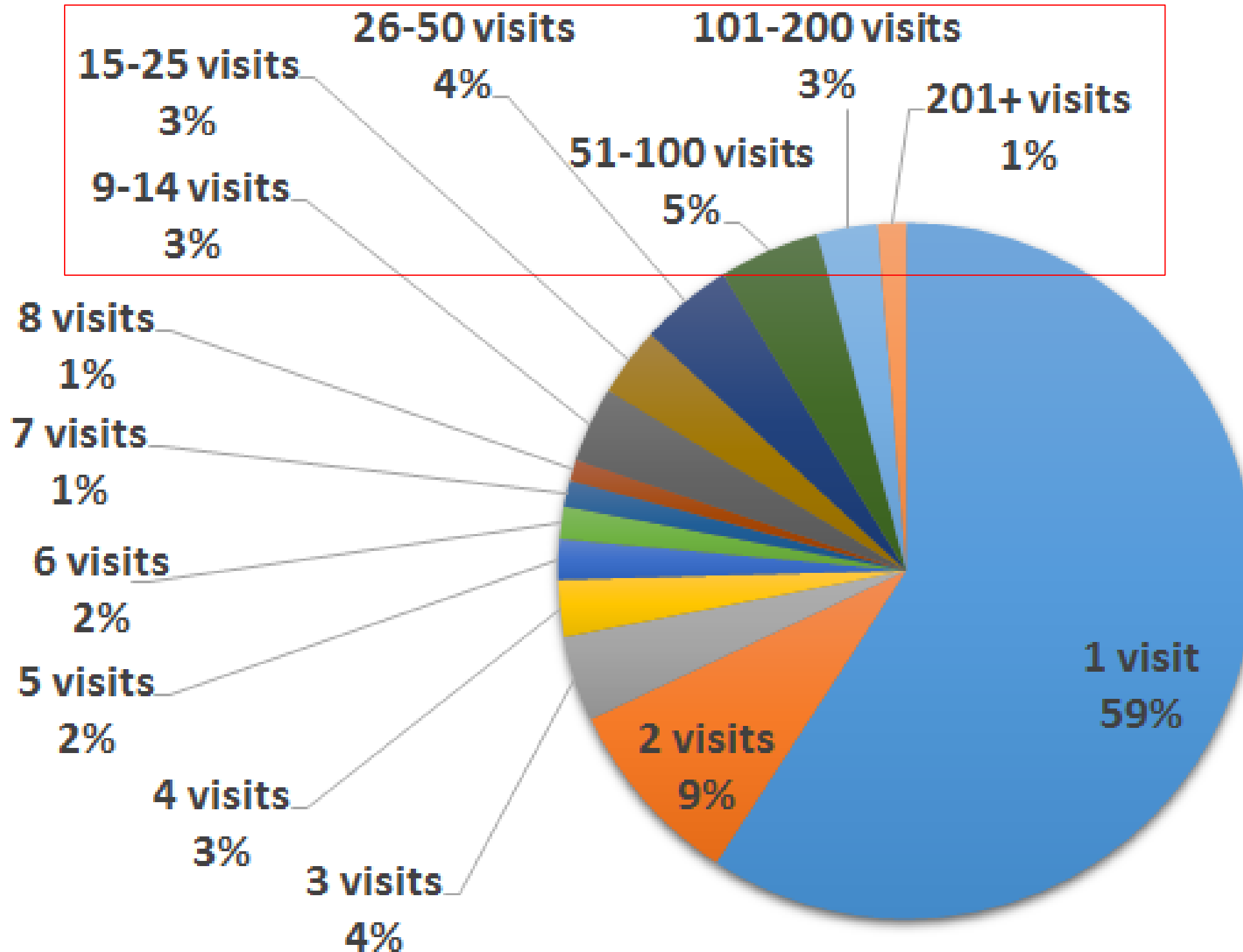
- 10098 indexed out of 881338 (sitemap)
- In total, 32172 indexed out of 881334 records (2016-10-10)

37% indexed and almost no impact on traffic

Visit duration (2 years)

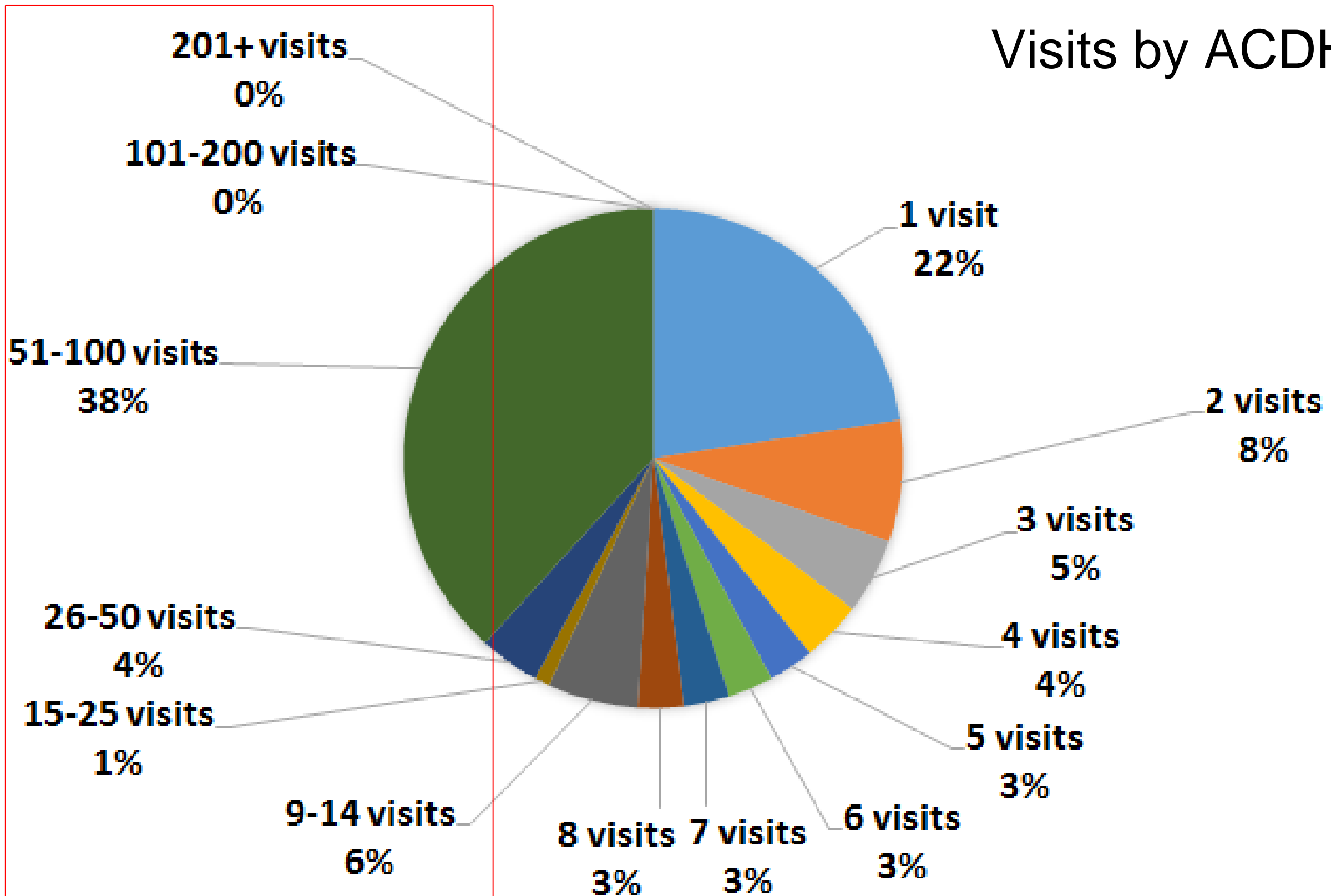
Average
4m 18sec





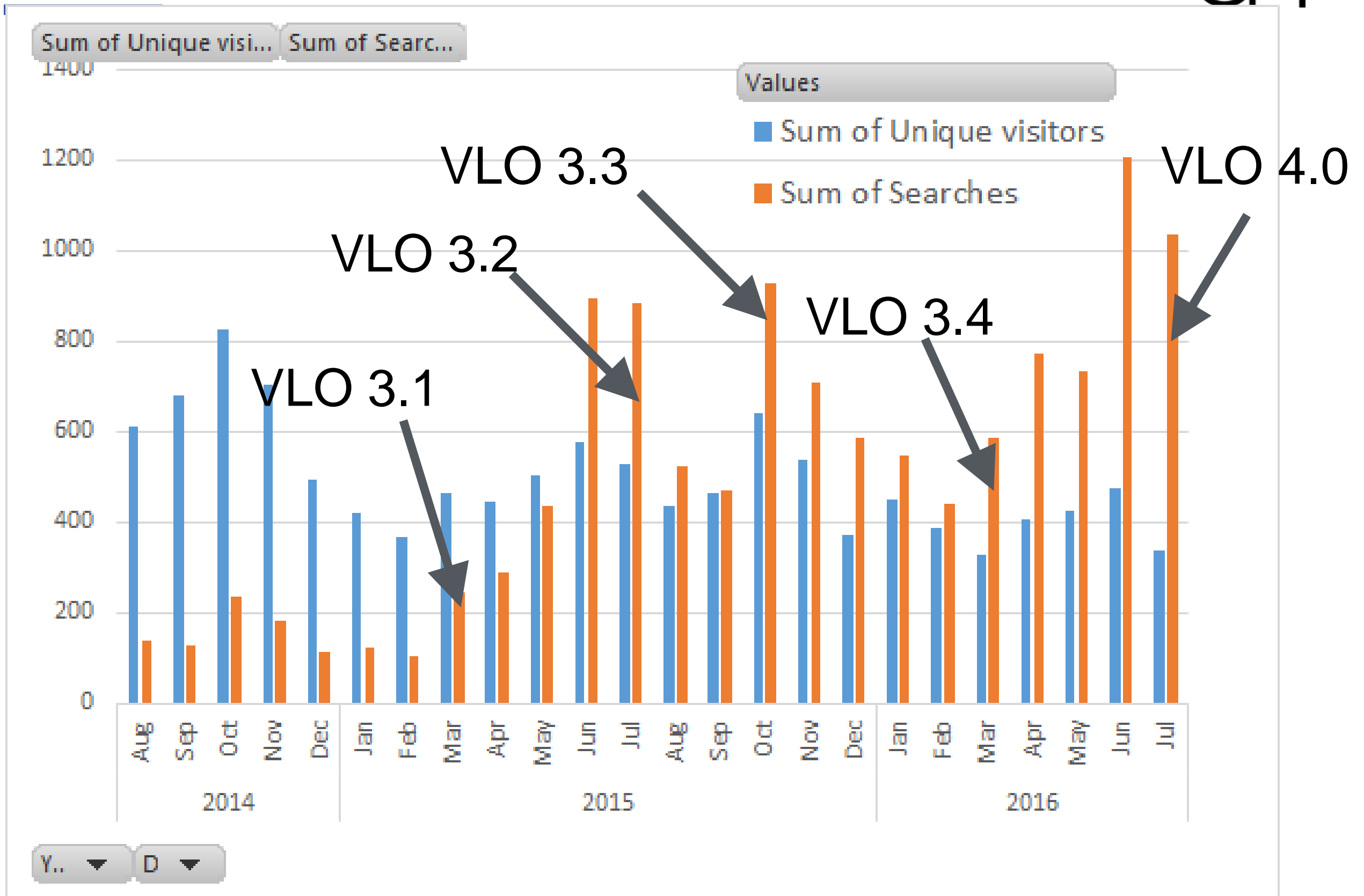
Visit frequency (2 years)

Visits by ACDH

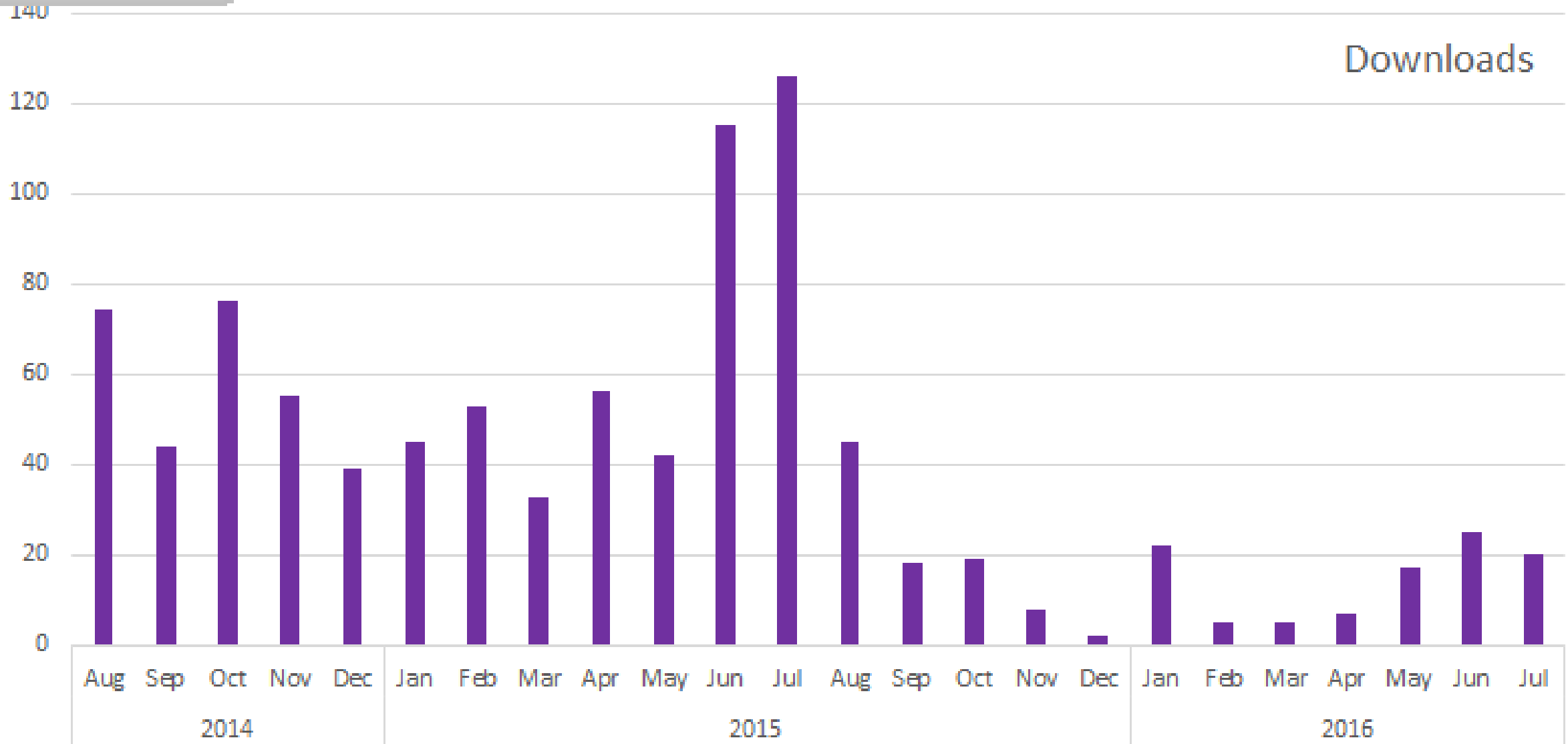


Downloads

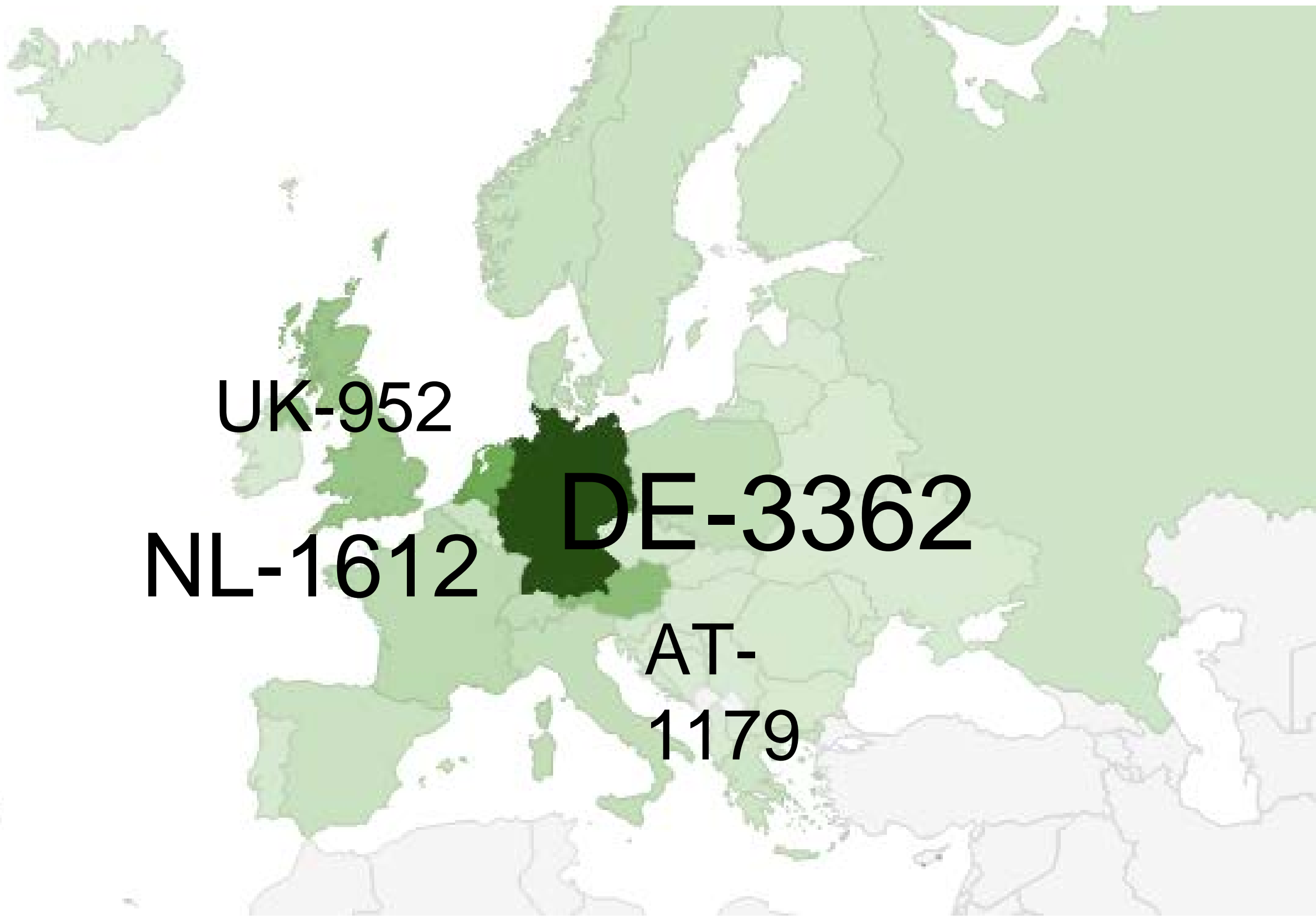
DOWNLOAD URL	UNIQUE DOWNLOADS	DOWNLOADS
catalog.clarin.eu	626	645
vlo.clarin.eu	114	120
catalog-clarin.esc.rzg.mpg.de	68	68
localhost:8080	40	41
dspace.library.uu.nl	27	31
beta-vlo.clarin.eu	16	16
clarin.phonetik.uni-muenchen.de	15	16
/BASRepository/Corpora/ALC/CLARINDocu.zip	2	2
/BASRepository/Corpora/ALC/ses1006/0061006030_h_00.wav	1	1
/BASRepository/Corpora/ALC/ses1006/0061006030_m_00.wav	1	1
/BASRepository/Corpora/ALC/ses1008/0081008025_h_00.wav	1	1
/BASRepository/Corpora/ALC/ses1008/0081008029_h_00.wav	1	1
/BASRepository/Corpora/ALC/ses1008/0081008029_m_00.wav	1	1
/BASRepository/Corpora/ALC/ses1008/0081008030_h_00.wav	1	1
/BASRepository/Corpora/FORMTASK/CLARINDocu.zip	1	1
/BASRepository/Corpora/SC10/CLARINDocu.zip	1	1
/BASRepository/Corpora/SK-Home/doc_SKH.zip	1	1



Sum of Unique Downl...

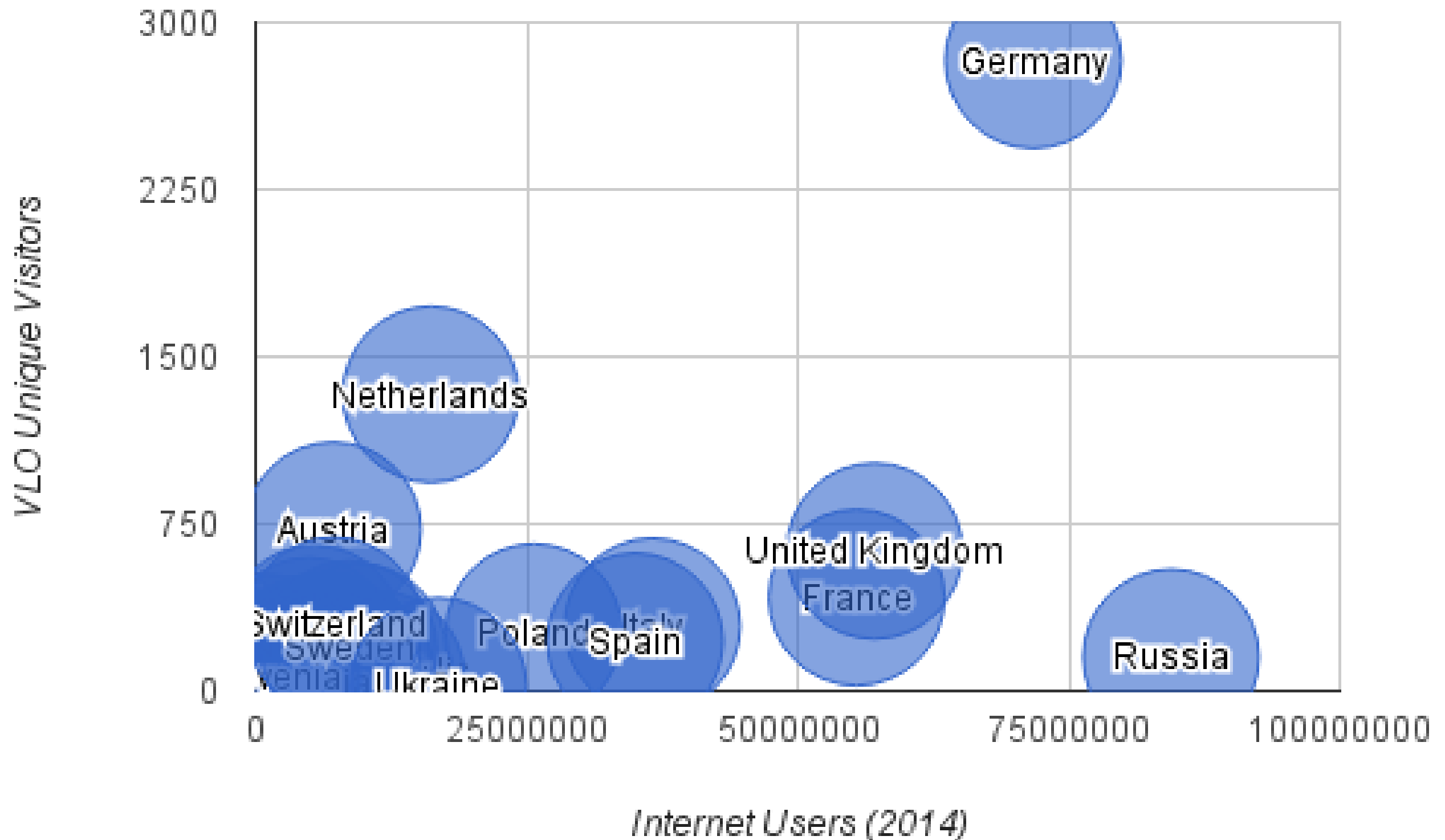


Y.. ▼ D ▼

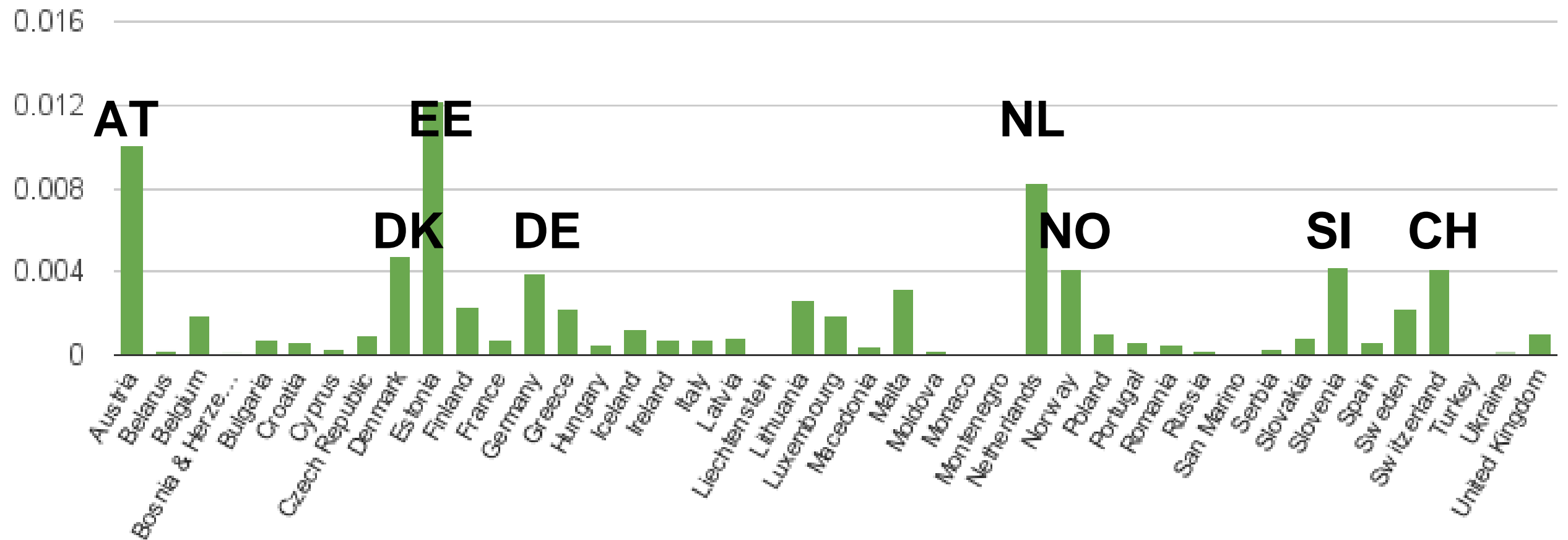


VLO visits per country in Europe

VLO Unique Visitors vs. Internet Users (2014)



Percentage of VLO Unique Visitors in Internet Population





Barcelona (1)

Top 100 European Cities

4. Recommendations for marketing

- Do more serious **User evaluation** esp outside CLARIN
- (Re-)do basic **marketing analyses** (SWOT, Marketing Mix, Segmentation, Positioning, etc)
- Do promotion (and communication) **strategically**
- **Agile marketing** -Measure stats before&after an event/promotion/press release/campaign etc
- Monitor and report the stats **regularly over time**

Recommendations for management

- User & market analysis, Outreach, and Infra services go **hand-in hand with PDCA** (esp tech-development alignment with user engagement)
- Formulate CLARIN **overall strategies** based on marketing research
- Consider **Open Evaluation** for measurable & transparent policies and operations for better credibility
- Start from small actions **now**, not later

Number game –Experience of a European research infrastructure (CLARIN) for the analysis of web traffic

Go Sugimoto

ACDH-OEAW

Vienna, Austria

go.sugimoto@oeaw.ac.a

t

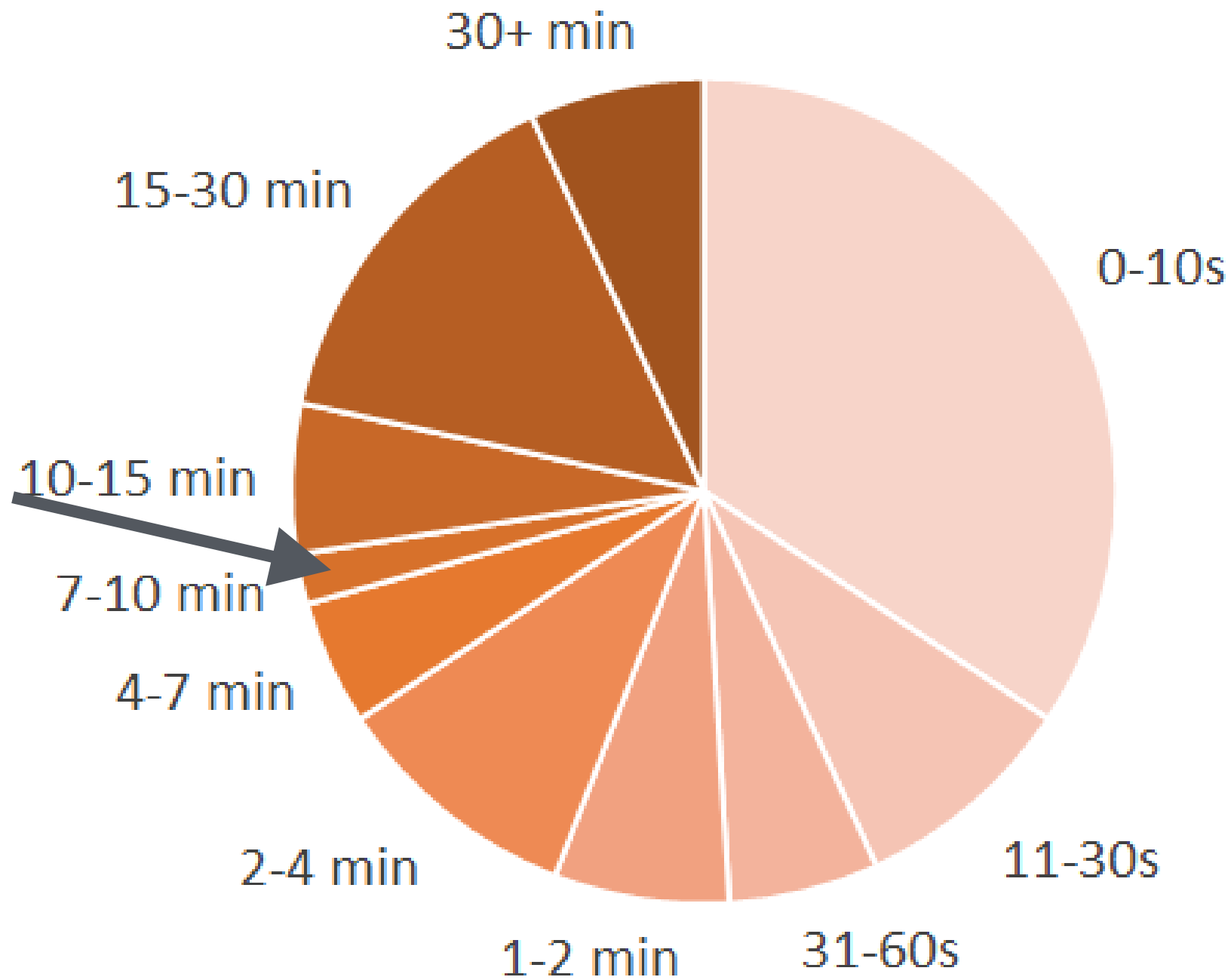
go.sugimoto@oeaw.ac.at

Come to Poster&Demo 2 (tomorrow)
for more stats (also on your request)

demonstrating the value of academic and research infrastructure in its own right. However, despite the emphasis on the users in the principles, it is evident that there is a lack of user evaluation in CLARIN. The most recent publication (Eckart et al., 2015) reports on the user behaviour of Virtual Language Observatory (VLO), one of the flagship services of CLARIN, but it does not offer any previous literatures on the subject. Although there are some user evaluations, they seem to be limited in the form of internal community feedback (Goosen & Eckart, 2014. Haaf et al., 2014). Wynne (2015) conducted a personal analysis on a various types of users and target domains, but there seem to be several contradictions on the target users and his conclusions are a bit of overstatement without providing proper evidences to prove them. This paper is intended to present objective statistics and make assumptions and conclusions in a more scientific manner. This is a vital step for CLARIN to confront with, because the question now is whether the CLARIN community will continue developing its infrastructural services without adequately and systematically assessing its use, or not. It is also

ACDH Visits

Average
8m 42sec



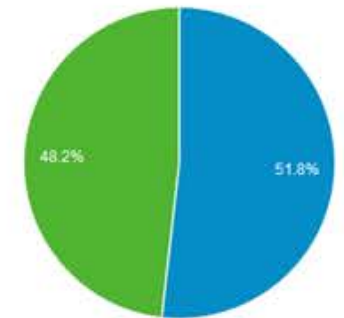
Science of Web Analytics

- What does Web Analytics record? Is it trustworthy?
- Can we compare the stats from different websites?
- Are the tools same?
- How can we use the stats properly or effectively?

Google Analytics vs Piwik (in ACDH)



Returning Visitor (blue) New Visitor (green)



Report

- 1,655 visits
- 4 min 14s average visit duration
- 42% visits have bounced (left the website after one page)
- 4.2 actions (page views, downloads, outlinks and internal site searches) per visit
- 0.61s average generation time

- 6,333 pageviews, 4,419 unique pageviews
- 1 total searches on your website, 1 unique keywords
- 50 downloads, 48 unique downloads
- 520 outlinks, 480 unique outlinks
- 247 max actions in one visit



CLARIN ERIC

@CLARINERIC

A European Infrastructure Program for the Social Sciences and the Humanities

Utrecht, the Netherlands

TWEETS 241 FOLGE ICH 110 FOLLOWER 326 GEFÄLLT MIR 15

Tweets Tweets & Antworten Medien

CLARIN ERIC hat retweetet



Marko Robnik-Sikonja @MarkoRobnikS · 29. Sep.

@dfiser3 Congratulations! I am sure @CLARINERIC will soon have great user involvement.



CLARIN ERIC

@ClarinEric

Home

About

Photos

Likes

Events



Like Message Share More

People

0 People Talking About This

88 Total Page Likes
0% from last week

0 New Page Likes
0%

— This week
— Last week

VLO	Year 1	Year 2
Average Unique Visitors (per day)	18.2	14.4
Unique Page Views	16001	16192
Unique Outlinks	1105	987
Bounce Rate	52.4%	42.2%
Returning Visits	30.8%	46.6%

geco	102	germanet	31
hzsk	89	test	30
german	50	beeld en geluid kinderleukemie	22
dutch	45	hzsk webservices	22
corpus	45	terminology	21
russian	44	english	21
french	33	format:audio/*	21
null	32	spanish	21
treebank	32	greek	20
Hausa	31		