CLARIN Annual Conference 2016
Aix-en-Provence, 26-28 October 2016
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)
Development Cycle (PDCA)

Johannes Vietze [Link to Wikipedia page]
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)
Public health: The toxic truth about sugar

Overview of attention for article published in Nature, February 2012

SUMMARY
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

TWITTER DEMOGRAPHICS

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.
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

6612 to 5260 (-20.4%)
Year 1 Unique Visitors (Weekdays)

August 2014 - July 2015

Average 18.2

CAC 2014 Soesterberg
Year 2 Unique Visitors (Weekdays)

August 2015

- CAC 2015
- VLO3.3
- AVERAGE 14.4

July 2016

- VLO 3.4
- Google Indexing
- VLO 4.0
- DH
- ? 20th May
- ? 2nd Nov
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)

- 1 visit: 59%
- 2 visits: 9%
- 4 visits: 3%
- 3 visits: 4%
- 5 visits: 2%
- 6 visits: 2%
- 7 visits: 1%
- 8 visits: 1%
- 9-14 visits: 3%
- 15-25 visits: 3%
- 26-50 visits: 4%
- 51-100 visits: 5%
- 101-200 visits: 3%
- 201+ visits: 1%
<table>
<thead>
<tr>
<th>Download URL</th>
<th>Unique Downloads</th>
<th>Downloads</th>
</tr>
</thead>
<tbody>
<tr>
<td>catalog.clarin.eu</td>
<td>626</td>
<td>645</td>
</tr>
<tr>
<td>vlo.clarin.eu</td>
<td>114</td>
<td>120</td>
</tr>
<tr>
<td>catalog-clarin.esc.rzg.mpg.de</td>
<td>68</td>
<td>68</td>
</tr>
<tr>
<td>localhost:8080</td>
<td>40</td>
<td>41</td>
</tr>
<tr>
<td>dspace.library.uu.nl</td>
<td>27</td>
<td>31</td>
</tr>
<tr>
<td>beta-vlo.clarin.eu</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>clarin.phonetik.uni-muenchen.de</td>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td>/BASRepository/Corpora/ALC/CLARINDocu.zip</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>/BASRepository/Corpora/ALC/ses1006/0061006030_h_00.wav</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>/BASRepository/Corpora/ALC/ses1006/0061006030_m_00.wav</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>/BASRepository/Corpora/ALC/ses1008/0081008025_h_00.wav</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>/BASRepository/Corpora/ALC/ses1008/0081008029_h_00.wav</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>/BASRepository/Corpora/ALC/ses1008/0081008029_m_00.wav</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>/BASRepository/Corpora/ALC/ses1008/0081008030_h_00.wav</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>/BASRepository/Corpora/FORMTASK/CLARINDocu.zip</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>/BASRepository/Corpora/SC10/CLARINDocu.zip</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>/BASRepository/Corpora/SK-Home/doc_SKH.zip</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
VLO visits per country in Europe

UK-952
NL-1612
DE-3362
AT-1179
Percentage of VLO Unique Visitors in Internet Population

AT                  EE                                      NL
DK     DE                                 NO              SI    CH
Top 100 European Cities
CLARIN centers in Europe
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
Come to Poster&Demo 2 (tomorrow) for more stats (also on your request)
ACDH Visits

- 30+ min
- 15-30 min
- 10-15 min
- 7-10 min
- 4-7 min
- 2-4 min
- 1-2 min
- 31-60s
- 11-30s
- 0-10s

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)

Sessions: 1,643
Users: 925
Pageviews: 5,919
Pages / Session: 3.60
Avg. Session Duration: 00:03:36
Bounce Rate: 41.57%

% New Sessions: 48.20%

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 retweeted
Marko Robnik-Sikonja @MarkoRobnikS · 29. Sep.
@difier3 Congratulations! I am sure @CLARINERIC will soon have great user involvement.
<table>
<thead>
<tr>
<th>VLO</th>
<th>Year 1</th>
<th>Year 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Unique Visitors (per day)</td>
<td>18.2</td>
<td>14.4</td>
</tr>
<tr>
<td>Unique Page Views</td>
<td>16001</td>
<td>16192</td>
</tr>
<tr>
<td>Unique Outlinks</td>
<td>1105</td>
<td>987</td>
</tr>
<tr>
<td>Bounce Rate</td>
<td>52.4%</td>
<td>42.2%</td>
</tr>
<tr>
<td>Returning Visits</td>
<td>30.8%</td>
<td>46.6%</td>
</tr>
<tr>
<td>Term</td>
<td>Count</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td>geco</td>
<td>102</td>
<td></td>
</tr>
<tr>
<td>hzsk</td>
<td>89</td>
<td></td>
</tr>
<tr>
<td>german</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>dutch</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>corpus</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>russian</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td>french</td>
<td>33</td>
<td></td>
</tr>
<tr>
<td>null</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>treebank</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>Hausa</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>germanet</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>test</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>beeld en geluid kinderleukemie</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>hzsk webservices</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>terminology</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>english</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>format:audio/*</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>spanish</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>greek</td>
<td>20</td>
<td></td>
</tr>
</tbody>
</table>