Paid vs Volunteer Work in Open Source Projects

Dirk Riehle, Philipp Riemer,
Carsten Kolassa, Michael Schmidt
Friedrich-Alexander University

HICSS 47 – 2014-01-08
Research Question

1. How much commercial contribution (to open source projects) is healthy?

2. How much commercial contribution (to open source projects) is there?
Research Approach

1. Empirical assessment
2. Descriptive analysis
3. Exploratory modeling
4. Dimensional break-out
### Research Process

- **Raw data source**
  - Local database (ohloh.net, crawled sources)
  - Web services access (ohloh.net, sf.net, others)

- **Extraction (pre-processing)**
  - Database queries using SQL and SQL scripts
  - Uses Java for computationally heavyweight filters

- **Prepared Data (intermediate representation)**
  - Output of pre-processing stage for specific tasks
  - Aggregation speeds up analytical processing

- **Evaluation (analytical processing)**
  - Mines data for insight, hypothesis testing
  - Basic processing, Java programming, R-project

- **Analysis output and results**
  - Results of processing: averages, distributions, correlations
  - Presented as models, tables, graphs, charts, etc.
Data Source

Linux Kernel
- In git with clean timezone data
- Author/committer distinction
- From 2005 to 2011
- One project

Ohloh 2008 Snapshot
- All kinds of CMS, no TZ data
- Only committer information
- From 2000 to 2008
- 9192 projects in database
Methodological Challenges

- Time-zone data not available in Ohloh
  - Determined committer timezones for 8% of work using location data
  - Found no reason to question representativeness smaller sample
  - *Worked with two sets, known and extended committers*

- Cultural bias using Western definitions
  - Review of open source activity by time-zone shows Western dominance
  - *Our results are a lower bound estimate*
Overview Result (All Data Sets)

Linux Kernel Authors

Commits of Authors per Hour

Number of Commits (hourly)

Hours of a Norm Day

Linux Kernel Committers

Commits of Committers per Hour

Number of Commits (hourly)

Hours of a Norm Day

Ohloh Known Set

Commits of Committers per Hour

Number of Commits (hourly)

Hours of a Norm Day

Ohloh Extended Set

Commits of Committers per Hour

Number of Commits (hourly)

Hours of a Norm Day
Ohloh Extended Committer Set

Commits in Working Time of *extended Committers* (weekly)
Trend Line (All Data Sets)

Linux Kernel Authors

Commits in Working Time of Authors (weekly)

Ohloh Known Set

Commits in Working Time of known Committers (weekly)

Linux Kernel Committers

Commits in Working Time of Committers (weekly)

Ohloh Extended Set

Commits in Working Time of extended Committers (weekly)
About 50% of all work contributed to open source software projects has been provided Monday to Friday, between 9am and 5pm.
## Detail Aggregate Results

<table>
<thead>
<tr>
<th>Linux Kernel</th>
<th>Percentage of total commits made during working time</th>
</tr>
</thead>
<tbody>
<tr>
<td>author</td>
<td>45.00%</td>
</tr>
<tr>
<td>committer</td>
<td>51.36%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ohloh Projects</th>
<th>Percentage of total commits made during working time</th>
</tr>
</thead>
</table>
| known committer | 47.3%  
(min. 28.2%, max. 58.8%) |
| extended comitters | 55.4%  
(min. 36.5%, max. 59.5%) |
Linux Kernel Work Time Percentage

Distribution of Contributors per Perc. During Working Time

- **author**
- **committer**
Ohloh Work Time Percentage

Number of Contributors per Working Time Percentage

Number of Contributors

Perc. of Commits During Working Time

0% 20% 40% 60% 80% 100%

committer
### Work Time Percentage Data

<table>
<thead>
<tr>
<th>Working Time Work %</th>
<th>Volunteer (Spare Time) Work</th>
<th>Mixed</th>
<th>Paid (Working Time) Work</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0%</td>
<td>0.01%-5%</td>
<td>5.01%-94.99%</td>
</tr>
<tr>
<td>Linux Kernel</td>
<td>author</td>
<td>33.06%</td>
<td>0.35%</td>
</tr>
<tr>
<td></td>
<td>committer</td>
<td>11.59%</td>
<td>3.05%</td>
</tr>
<tr>
<td>Ohloh Projects</td>
<td>known committers</td>
<td>2.41%</td>
<td>1.21%</td>
</tr>
<tr>
<td></td>
<td>extended committers</td>
<td>7.04%</td>
<td>0.4%</td>
</tr>
</tbody>
</table>

- Note the inverse relationship Linux Kernel author / committer
Future Work

- Dimensional break-outs
  - By type of software (infrastructure, productivity, entertainment)
  - By time-zone and (implied) culture (what about China?)

- More current data sets
Thank you! Questions?

dirk.riehle@fau.de – http://osr.cs.fau.de
dirk@riehle.org – http://dirkriehle.com – @dirkriehle