A Dataset of Open Source Android Applications
Daniel E. Krutz, Mehdi Mirakhorli, and Samuel A. Malachowsky
http://androsec.rit.edu

INTRODUCTION
• Began by collecting open source Android apps from F-Droid
• Analyzed Git Histories and performed static analysis using
  o Androrisk: App vulnerability level
  o Sonar: Coding metrics
  o Stowaway: Over & Under Privilege level
• Collected 1,179 apps with 4,416 versions
• 435,680 Version Control commits
• Data is publicly available on website and through SQLite Database

APPROACH
• Collected data from F-Droid, open source Android website (APK & Git Info)
• Analyzed using various static analysis tools

WEBSITE
• Search for app ‘FBReader’
• Violations per KLOC - Per Version
  • Average Violations/KLOC = 113
  • Over & Under Privileged Rate
    • Under Priviledged Rate = 1.66, Average Over Privileged = 0.66
• Androrisk Scores
  • Average Androrisk Score = 29.66

DATA
• Overall Stats
• Write Your Own Query
• Aggregate Statistics

<table>
<thead>
<tr>
<th>Value</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apps</td>
<td>1,179</td>
</tr>
<tr>
<td>Versions</td>
<td>4,416</td>
</tr>
<tr>
<td>Committers</td>
<td>4,535</td>
</tr>
<tr>
<td>Commits</td>
<td>435,680</td>
</tr>
<tr>
<td>Versions</td>
<td>48</td>
</tr>
<tr>
<td>Committers</td>
<td>202</td>
</tr>
<tr>
<td>Commits</td>
<td>65,110</td>
</tr>
</tbody>
</table>

F-Droid

Semantically Organized Analytical Data