Programming Knowledge We Can Rely On

Arie van Deursen
@avandeursen
Open Source is Everywhere

Projects with the most contributors

- MICROSOFT/VSCODE: 15K
- FACEBOOK/REACT-NATIVE: 8.8K
- NPM/NPM: 7.6K
- ANGULAR/ANGULAR-CLI: 7.4K
- TENSORFLOW/TENSORFLOW: 7.3K
Collaboration with Pull Requests

As a developer I offer a coherent set of changes to you, the project owner.

“I take responsibility for this change; Are you willing to integrate it?”
Learning from the trenches

• Does testing slow you down?
  – *It speeds up integrator’s work* (ICSE 2014)

• Is “async/await” useful in C#?
  – *Yes, but often used incorrectly* (ICSE 2014)
GHTorrent: *Data As A Service*

Scalable, query-able, offline mirror of data from GitHub REST API.

- Data since 2012
- 16 TB json in MongoDB
- 8 billion rows in MySQL
- 2 GB *per hour* collected

- 350 users in > 200 institutions
- 200+ papers
- 3 data mining challenges
- Confirmed commercial uses by Microsoft, Deloitte, Blackduck
- Support gifts from Microsoft

*Georgios Gousios, Diomidis Spinellis, Martin Pinzger, Arie van Deursen, Andy Zaidman, Margaret-Anne Storey, Alberto Bacchelli.*


GHTorrent dataset, pull-based software development, integrator and developer perspectives.
Programming Knowledge We Can Rely On

Arie van Deursen

@avandeursen