On the Geographic Allocation of Open Source Software Activities

Sebastian von Engelhardt, Department of Economics, Friedrich-Schiller-University Jena, Jena, Germany

Andreas Freytag, Department of Economics, Friedrich-Schiller-University Jena, Jena, Germany

Christoph Schulz, Department of Business Service and Infrastructure, Carl Zeiss AG, Jena, Germany

ABSTRACT

This article contributes to research on the geographic origin of open source software (OSS) developers by analyzing the geographic allocation of active OSS developers and OSS activities. Based on data from the SourceForge Research Data Archive, the authors exploit information about developers' IP address, email address, and indicated time-zone. This enables them to assign 94% of all registered users in 2006. As proxy for activity the authors use information about the number of posted messages. Thus they provide a detailed picture of the world-wide allocation of OSS activities. Such country data about the supply-side of OSS is a valuable stock for both cross-country studies on OSS and country-specific research and policy advice.

Keywords: Computer Science, Geographical Location, Open Source Activities, Open Source Software, Social Science

INTRODUCTION

Open source software (OSS) is developed by communities that include hobbyists as well as companies, and the source code—the human-readable recipe—is 'open'. This means that everyone has access to, and the right to read, modify, improve, redistribute and use the source code. Thus, OSS appears to be a case of a "private provision of a public good" (Johnson, 2002). As the community is often described as global, OSS seems to be a digital public good with a truly globalized private provision.

However, apart from anecdotal evidence for the internationality of certain OSS project teams, the question remains how global the OSS community actually is and how the supply side of OSS differs among countries. This has motivated researchers to study the geographical allocation of OSS developers. It turns out that the most OSS developers come from North America and Europe. This result is quite consistent regardless of the method used. The methods to gather information about the geographic origin of OSS developers can be broadly distinguished into two approaches. Some studies are based

on survey-data, while other work is based on specific data drawn from code of certain OSS projects such as credit files, mailing lists or data from platforms like SourceForge.

Robles et al. (2001) provide a combination of both types of data collection. In Ghosh (2006); David et al. (2003) and Ghosh et al. (2002) one can find survey-based information about the origin of OSS developers. Lancashire (2001) provides information about the world-wide distribution of Linux and Gnome developers, based on data collected from the Linux Credit file and in case of Gnome developer-contact information from the project's web-site. The most recent research dealing with the geographic origin of OSS developers is Gonzalez-Barahona et al. (2008), who provide a worldwide picture of OSS developers, weighted by population, internet users and GDP. Gonzalez-Barahona et al. (2008) build on Robles and Gonzalez-Barahona (2006). Robles and Gonzalez-Barahona (2006) use information about the email addresses of registered users and the indicated time-zone to assign developers at SourceForge in 2005 to their countries. However, they were unable to assign 25% to countries, because of the combination of a generic (non-country specific) Top Level Domain like .com with the country unspecific time zone GMT. Robles and Gonzalez-Barahona (2006) develop methods to estimate the geographic allocation of this 25%.

Our work is inspired by Gonzalez-Barahona et al. (2008) and Robles and Gonzalez-Barahona (2006), but proceeds along two lines: First, we do not have to estimate any geographic origins, since we can directly assign 94% of all developers registered at SourceForge in 2006. We make use of relevant information obtained from email, time zone and the Internet Protocol addresses. Combining these, we are able to assign 1.3 million developers to their countries without the need to estimate geographic origin. We cross-checked the results delivered by the different methods in order to have an indicator for the validity of each of our methods. Second, we provide information about how active each developer is. With individual data about the number of posted messages we have a good proxy for activity. We can thus distin-

guish active from non-active (but nevertheless registered) developers, and we are able to show the worldwide allocation of OSS *activities*. Information about activities are important, since members of the OSS community differ in their effort levels, numbers of contributions etc. (see e.g. David & Rullani, 2008).[1] With the active developers and activity, our study can show a more accurate geography of the supply side of OSS development.

The Need for Accurate Data About OSS Activities

Software-development is an important part of each country's ICT-Sector. But without information about OSS activities the picture of the software industry's supply side remains incomplete. For example, workforce and human capital data typically count only the paid labor force and thus ignore the—for the most part unpaid—OSS developers. On the other hand, adding the number of OSS developers to the number of paid jobs is also incorrect for two reasons. First, not all registered developers are active: at SourceForge, only every fifth member was active in 2006. Second, some OSS developers have jobs in the software-sector which would lead to double-counting. Taking the OSS-activity-level into account can, together with the numbers of the paid software development, provide a more accurate picture.

Data including the number of active OSS developers and their activity level by country are important for both, policy makers and businesses. Government competition policy as well as support for OSS development, OSS-based business models and OSS-based start-ups should be based on knowledge of the national human capital and capacity for OSS. This information is also useful for e.g. entrepreneurs contemplating an OSS-based start-up, firms planning to implement an OSS-based business model, etc.

Additionally, a better and more complete picture of software-development across countries can improve cross-country studies, including aspects of the digital divide. Finally, our dataset can be used for analyzing country-

13 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/geographic-allocation-open-source-software/78323

Related Content

Agreements Treaties and Negotiated Settlements Database
Marcia Langton, Odette Mazeland Lisa Palmer (2007). *Information Technology and Indigenous People (pp. 266-271).*
www.irma-international.org/chapter/agreements-treaties-negotiated-settlements-database/23563

Impact of Normality Pressure on Acceptance of Mobile Phone Technology: A Case of Employees of SMEs
Renatus Michael Mushi (2022). *International Journal of ICT Research in Africa and the Middle East (pp. 1-11).*
www.irma-international.org/article/impact-of-normality-pressure-on-acceptance-of-mobile-phone-technology/311413

Language as a Medium for Bridging the Gap Between Philosophy, Media, and Development
Odey Simon Robert (2022). *Handbook of Research on Connecting Philosophy, Media, and Development in Developing Countries (pp. 26-41).*
www.irma-international.org/chapter/language-as-a-medium-for-bridging-the-gap-between-philosophy-media-and-development/304259

Equality, Diversity, and Access in Digitalized Teaching in Higher Education
Ranson Sifiso Gwalaand Pfano Mashau (2024). *Accessibility of Digital Higher Education in the Global South (pp. 105-131).*
www.irma-international.org/chapter/equality-diversity-and-access-in-digitalized-teaching-in-higher-education/334635

eLearning X.0: Are Learning Management Systems Ready?
Mohamed E. Edrees (2014). *International Journal of Technology Diffusion (pp. 1-13).*
www.irma-international.org/article/elearning-x0/120503