



# FREQUENTLY ASKED QUESTIONS

---

## **Who did this study and why?**

Red Hat, Inc., contracted with Georgia Institute of Technology (Georgia Tech) to create a tool to quickly compare and contrast countries based on their open source activity and environment. This work culminated in the Open Source Index (OSI).

Dr. Paul M.A. Baker, research director for the Center for Advanced Communications Policy, and Dr. Douglas S. Noonan, associate professor at the School of Public Policy, led the study. Nathan Moon, graduate research assistant, and Art Seavey of the Estis Group contributed to the study and these materials. From this framework the research team developed an Open Source Index assessing a country's status on open source, in a multi-dimensional manner, facilitating cross-country comparisons.

## **Why is my country not listed?**

In a perfect world the Index would be able to draw on a wide variety of data sets that were populated with identical global data. In practice, getting comparable data on every country in the world on such a wide variety of factors, especially related to open source, is virtually impossible. The fact that a country is not included in the Index does not mean that a country has any better or worse environment for open source. It simply means that there was not enough data available to be able to give that country a score.

## **What is the difference between Activity and Environment?**

Activity is an index which measures the amount of open source happening today. It tends to be made up of concrete factors, such as existing open source and open standards policies and number of OSS users or producers, such as Linux and Google Summer of Code.

Environmental factors are more speculative. Even a country that does not have a high degree of current penetration of open source may have a high number of internet users and information technology patents. These factors may indicate a favorable environment for open source software to take hold. Still, the correlation between a country's score on the activity and environment is quite high.

## **How did you come up with the score?**

Each country received two summary scores: one for Activity and one for Environment. Both Activity and Environment are measured on three dimensions, or sub-scores: Government, Industry, and Community/Education. Then, each dimension within Activity and Environment is measured along several quantitative indicators. Indicators for each dimension were selected by the researchers and averaged to produce each country's raw score. The countries' raw scores were then used to rank the countries in the list. If you would like to see a detailed explanation of how the scores were developed please [read the study](#).



### **Why is this study important?**

This study shows that open source is widespread and thriving. Still, very little empirical evidence exists for documenting and explaining why or how the open source model works. Scholars have examined the adoption of open source by national governments, via policy mechanism/regulatory approaches, but there is still not enough data showing us why open source succeeds in some places and not in others.

By 2001, Peru, Brazil, Argentina, France, and Mexico all had pending policy measures pending at the national level involving open source software. Other national and local-level efforts were also taken up in such countries as Germany, Spain, Italy, and Vietnam to establish official alternatives to the use of closed, proprietary software by government. The Open Source Index facilitates the objective comparison of those countries, which have the highest indicators of open source penetration, and whether common factors link those countries together.

### **How can I use this data?**

You will notice that the data sets and research paper from Drs. Baker and Noonan are available for download from this website. We hope that researchers, students, government officials, CIOs, and FLOSS advocates will use this data not just to examine the use of open source in their native countries, but also to encourage the advancement of research in this area to build more successful and robust open source environments. As a first version of the tool, there is room for improvement and we encourage you in the spirit of open source to improve upon it and share your findings with others.

The Open Source Index 2008 is available under the Creative Commons Attribution-Share Alike 3.0 United States (CC-BY-SA) license. Although you are free to copy, publish, and make derivative works of the Index and accompanying documents, you must attribute the work to "Open Source Index 2008, Red Hat, Inc." You may distribute the Open Source Index 2008 or your own modification of this work only under the terms of: (i) this License; (ii) a later version of this License with the same License Elements as this License; (iii) a Creative Commons jurisdiction license (either this or a later license version) that contains the same License Elements as this License; or (iv) a Creative Commons Compatible License.

### **How were the variables selected?**

Georgia Tech compiled over 700 variables that might fit into a slot, or slots, within the framework (See: "*What was the research process?*"). The final choice of what variable would make it into the framework balanced how many countries the variable covered and its closeness to measuring, versus just indicating, the concept. Other considerations included the type of measure it was, for example, a scale "from 1 to 5," or a ratio "gross domestic product"; but, ultimately, variable selection came down to the choice of the researchers at Georgia Tech. In some parts of the framework no measures were available. In others there may exist a "better" variable, but it did not cover more than a handful of countries.

### **How is the raw score calculated?**

To calculate a raw score, all variables have to first be transformed into the same units. To do this, the OSI calculates a z-score for each variable. A z-score shows how high, or how low, a country scores compared to the average of the countries in the list, measured in units of standard deviations. In other words, a z-score for a specific variable shows a relative score.

The z-score of each variable is then averaged in the respective dimensions to produce the raw score. So, for example, a country, like Australia, with a raw score of 0.61 in Government Activity is the average of the z-scores of the variables that measure or indicate procurement, policy and use. To produce the final raw score, all raw scores of the dimensions in the Environment, or the Activity Index are also averaged.

Sometimes indices apply different weights to different dimensions. The Open Source Index does not apply any weights, or rather applies a weight of "1" to all dimensions. However, with further research, it may become of interest to "weight" government measures as more important than say education measures. We encourage you to perform these experiments on your own.

Finally, the OSI keeps Activity and Environment raw scores separated. Depending on future research interests, others may want to calculate "Activity/Environment" or "Activity + Environment" and analyze the results.



### How are the ranks calculated?

The rank displayed is where from positions 1 to 75 a country's raw score lands. It's important to remember this does not necessarily mean country No. 70 ranks 70 in the world, but rather relative to the country set included in the Open Source Index.

### What does it mean if a country ranks low?

Quite literally, if a country scores low in a specific area then it means relative to the other 74 countries in the OSI, it doesn't have as much of, or doesn't have at all, a certain variable. The practical significance is something that must still be evaluated. This Index does not make any final causal statements about what causes or does not cause open source activity. It's a tool to start to find those relationships and build an evidence base.

### What was the research process?

This study began with an extensive review of the academic literature and professional and general media. Based on that and further analysis they developed a framework to measure open source. This framework has "slots" for discrete concepts, for example, academic use. At the same time, Georgia Tech, in turn with students from the open source class, performed a wide search for any and all existing empirical data that might fit within the framework. For more on variable selection see *"How were the variables selected?"*

Once the variables were selected, the team at Georgia Tech tested numerous ways to construct the index for robustness, some covering more countries than others, some using geometric means rather than just averages, and others with regression techniques. We chose to move on with the OSI presented here.

### How was the interactive map built?

This map interface is built using OpenLayers. OpenLayers is an open source JavaScript library for building map applications on the web. The data polygons were edited in QGIS and then converted to GeoJSON with command line utilities from GDAL/OGR. The country shapefile is provided by [thematicmapping.org](http://thematicmapping.org). To minimize file size, unused attribute fields were removed from the shapefile before adding the Open Source data. The country polygons are reprojected into Mercator by OpenLayers and then displayed over a background image. The polygon colors and transparency are all controlled in the HTML code. The background image is currently provided by Google Maps, but will soon be replaced by OpenStreetMap data. Flag icons are provided by [famfamfam.com](http://famfamfam.com). Resources for creating a map using this structure include:

<http://www.qgis.org>

<http://www.gdal.org>

<http://www.openstreetmap.org>

<http://www.osgeo.org>