HIGHER EDUCATION INSTITUTIONS AND THE GLOBAL ROLE OF FREE/LIBRE AND OPEN SOURCE SOFTWARE

A Report on Findings from the *FLOSS WORLD* Survey of Developing and Transition Economies

By

Paul A. David* and Joseph S. Shapiro**

**Oxford Internet Institute, All Souls College, Oxford and Stanford University*
paul.david@oii.ox.ac.uk

***London School of Economics and the Oxford Internet Institute*
j.s.shapiro@lse.ac.uk

First draft: March 25, 2007

Second version: May 6, 2007

Final version: 14 August, 2007

ACKNOWLEDGEMENTS

The research and analysis for this report was carried out as part of the *FLOSS WORLD* Project, led by the Maastricht Economic Research Institute on Innovation and Technology (MERIT) in the Netherlands, (http://www.FLOSSWorld.org/index.php) and funded as a Special Support Action under the EU's 6th Framework Research Programme. Launched in May 2005, *FLOSS WORLD* has involved the efforts of 17 partner institutions located in 12 countries around the world, in designing and implementing of a suite of surveys that for first time targeted software developers, business firms, government organisations, and institutions of higher education in eight developing and transition economies: Argentina, Brazil, Bulgaria, China, Croatia, India, Malaysia and South Africa. We are grateful for the careful and efficient assistance of Ellen Helsper in preparing the basic template for the *FLOSS WORLD* Survey of Higher Education Institutions (HEI), and for the cooperation of Sebastian Rahtz and Randy Metcalfe of the JISC's OSS- Watch Programme at the Oxford University Computing Service on whose experience in surveying UK HEI's use and policies concerning open source software our team was able to draw. Conduct of the survey itself would not have been feasible without the contributions of our partner research teams, who were responsible for adapting the master survey instrument to their respective "local" institutional conditions and languages, and for identifying individuals and institutions to whom the questionnaires were directed. No less vital were the contributions of colleagues at the three other collaborating EU research institutions that were vital to the conduct of this research project: MERIT and UNU-INTECH, the Universidad Rey Juan Carlos (Madrid) and FUNDECYT/Government of Extremadura, Spain. Their advice and critical comments on the basic survey design, and the technical support provided to maintain the questionnaires "online" were invaluable. Following the presentation of results from our preliminary analysis of the survey responses at the 2nd International *FLOSS WORLD* Conference (held in Brussels, 9-12 June 2007), we have had the benefit of the local knowledge and insights of many of our partners regarding the characteristics of the diverse institutions that responded to the HEI survey, as well as *FLOSS WORLD* of the circumstances surrounding the administration of the questionnaires which has contributed to the interpretation of the statistical findings. We wish to express our gratitude for the extra efforts made to transmit this information after the formal close of our colleagues' work on the project, and, in Appendix A6 have summarized material from two of the more extensive country reports on these matters.

1. Introduction

Policymakers in the private and public sectors, and researchers in the social sciences and software engineering, have sought to understand the dynamics and implications of the growth of Free/Libre/Open Source Software (FLOSS) by focusing primarily upon the motivations of the individuals that participate in its development and the conditions affecting the adoption of "open source" computer programs by business firms and government agencies.[1] Comparatively little attention has been given to examining the extent to which FLOSS is being created, released and applied within the Higher Education Sector (HES). Moreover, virtually all the previous empirical research in that vein has been concentrated upon FLOSS development and diffusion in economically advanced regions of the world. This report seeks to help fill the resulting "double-gap" in knowledge regarding the global role of "open source" software, by presenting the findings of a recently completed survey of administrative staff members and information technology managers at "universities" in seven developing and "transition" economies.[2]

The present study represents one part of a broader enquiry supported by the EU 6th Framework Research Programme -- the *FLOSSWORLD* project, the aim of which is to increase knowledge about the global phenomenon of Free/Libre and Open Source Software (FLOSS) development and adoption, thereby contributing to informed public policy and effective collaboration between the EU and developing countries. As a Special Action Project, *FLOSSWORLD* seeks to advance Europe's leadership in FLOSS development, to build a global constituency of policymakers and researchers, to enhance global awareness of FLOSS issues, and to strengthen the research communities in the participating regions. Toward those goals *FLOSSWORLD* has designed and carried out parallel surveys of government organizations, business firms, and individual developers, as well as universities in the selected group of countries: Argentina, Bulgaria, Brazil, China, Croatia, India, Malaysia, and South Africa. By highlighting the dynamics of open source software use in the universities of those countries, the present report facilitates comparisons of the role of FLOSS in a variety of organizations whose separate activities and interactions are of critical importance for the development of the capabilities of their respective societies and the well-being of their citizens. It examines inter-regional differences in the extent of the use of FLOSS in university teaching, research and administration, and the contributions of members of those institutions to open source software development. Its findings serve to highlights opportunities for improving the future role of FLOSS in the HES.

1.1 FLOSS in the HES: Some Motivating Issues

A growing body of research has pointed to the broader economic significance of the mode of production that has characterized the development of many of the most successful and extensively adopted FLOSS programs (e.g., the GNU Linux, Apache, Mozilla and Firefox, MySQL, OpenOffice). These typically decentralized, self-governing, trans-national

[1] Most of the systematic survey studies focus on the supply of open source software (FLOSS), inquiring into the characteristics, location and motives of the developers. Studies based upon extensive survey data include Robles et al. 2001; Ghosh et al. 2002; David, Waterman, and Arora 2003; Mitsubishi 2004. See David and Shapiro 2007 comparative details. The 2002 FLOSS Report by Ghosh et al., however, looked at business use and procurement policies as well. There are, in addition, numerous "case studies" of migration to FLOSS in business and non-profit organizations, although meta-analysis of their findings is generally difficult.

[2] The survey focuses on the main categories of higher education institutions, which for convenience we refer to simply as "universities". The latter term applies, therefore, to research universities, teaching colleges, research institutes, and other institutions that fall under the label of HEI -- with the caveat that some respondent institutions do not teach undergraduates, or do not conduct research, or in some other way do not fit the traditional conceptualization of a university. The higher education sector (HES) includes other tertiary educational organizations and programs that were not targeted by this survey, notably those specializing in continuing education and non-professionalized areas of vocational training.

community-based, and heavily volunteer-dependent collaborations utilize "open" peer-based coordination, continuous workflow and early and frequent code release practices – all of which present marked divergences from the mode of production that had come to be established in the "closed", proprietary sector of the global software industry. From the organizational standpoint, as well as from the uses to which these "open source" projects put the copyrights on the code contributed by developers, FLOSS as a system of allocating resources for production and distribution occupies a territory distinct from that governed by either the "visible hand" of hierarchical management or the "invisible hand" of the market. Its resemblance to prominent features of academic "open science" research collaborations has been remarked upon, but there are numerous respects in which FLOSS collaborations remain distinctive and far less institutionalized.[3]

Beyond its potentials as a paradigm for collaborative creation of a range of information products considerably broader than computer software, the FLOSS movement has come to be viewed as emblematic of a more general reorientation of the organization and conduct of many processes in the social and political sphere, one that may be particularly supportive of the production of public goods and participatory democracy (Benkler 2006). Rather strikingly, however, the roles that FLOSS may have in the sphere of Web-based skills acquisition, the support of conventional educational and training activities in computer programming and Web design, and the formation of human capital more generally, has not attracted equivalent attention.[4] Public discussions of the relevance of "open source" programs for e-learning at the tertiary level (at least, those carried on in English language sources) have been largely preoccupied with questions of patent rights and licensing cost of "course management" and e-learning support software (particularly those arising from the "Blackboard Learning System" patents, and the possible threats to FLOSS alternatives such as "Moodle," and MIT's "Saki Project").[5]

But, quite obviously, the potential contributions that the activities of FLOSS communities can make to university education and skill formation do not begin and end with the question of the costs to educational institutions of on-line course management software, and this is especially true when the situation of the developing and transition economies are considered. Universities (and educational institutions more generally) can be both significant contributors to, and beneficiaries from the development of FLOSS. The major educational and research missions of universities give them the potential to be powerful actors in this regard: they can prepare the users and developers of computer software to apply these tools as citizens, consumers, employees and entrepreneurs; they train researchers in scientific and technical fields that are becoming evermore reliant upon advanced digital information processing and retrieval technologies, and they are the institutional hosts for fundamental and applied research in the mathematical and computational sciences.

Furthermore, the policies of universities and the behaviors of their employees are of interest because the HES is a major employer and user of digital information and computer-mediated telecommunication resources, and the decisions made within these organizations about what hardware and software systems will be acquired and supported are shaped by

---

[3] For further discussion of the "open source" – "open science" relationship, see, e.g., Dalle, David, Ghosh and Steinmueller (2005); on open science institutions and norms among Internet-based research projects, see David, den Besten and Schroeder (2006).

[4] It is striking that the discussion of "educational instruction" by Benkler (2006: pp. 315, 327) is brief, focused on the limited capacity of electronic communications to transform university level (or other) face-to-face instruction, and confined to the MIT Open Courseware Initiative and the potentialities of employing such public domain strategies to create open platforms on which textbook authors and instructors can collaborate. Benkler (op.cit., p. 326) cites the South African project (Free High School Texts) as the "the most successful commons-based textbook authoring project, which is also the most relevant from the perspective of development."

[5] See discussion and links at http://en.wikipedia.org/wiki/Blackboard_Inc.

incentives and constraints that are not identical to either those in government agencies or in private, profit-oriented enterprises. Examining the use and development of FLOSS within HEI's, and the possible differences within the various functional divisions of the university (administration, teaching and research in the arts, and in the sciences) may be informative of more general questions about the ways in which different structures of organization and cognitive activity affect the balance of choice between proprietary software and FLOSS.

In addition, it must be recognized that HEI's are providers of high-speed internet connections with global sources of information that are likely to be far from ubiquitously available and therefore particularly important in many developing countries. Consequently, beyond questions concerning the extent to which universities and technical training institutes are offering instruction in the skills that would allow students and staff to modify and utilize open source computer code for their own uses, and to participate in collaborative projects on the Internet, one should consider how the policies of universities affect the access that students in developing regions can have to informal, experience-based learning interactions with global software development communities.[6] Opportunities to observe, passively follow email forum discussions, and learn how to elicit and absorb technical help from more skilled programmers and sophisticated users of specialized software packages, can be especially important in settings where such expertise is not locally available. Although it is sometimes argued that the lack of expertise in computer programming in a region's workforce constitutes a powerful reason for relying exclusively on "closed", user-friendly software packages from proprietary vendors who can supply external support and help for unsophisticated users, as a developmental strategy this is very short-run in its priorities. It limits opportunities for skills development that can transform the region's resource endowment, and it ignores the potentialities for knowledge transfers from the existing international community of FLOSS developers and users to accelerate that learning process.

In developing regions of the world, where young people are unlikely to have their own personal computers and high bandwidth telecommunications connections, educational institutions can provide key portals for self-initiated skill formation through contact with distributed "communities of practice." To the extent to which universities encourage students and staff members to avail themselves of those "learning resources," international FLOSS development communities in turn may adopt norms and procedures that accommodate and facilitate informal processes of skill acquisition – rather than dealing with neophyte developers in ways that rebuff an discourage individuals for whom such contacts could constitute an important source of knowledge. Whether the foundations that support the activities of some of the larger FLOSS development communities would be willing and capable of taking such a pro-active role in contributing to the formation of human capital, and the enhancement of the software skill proficiency and versatility of scientific and technical workers in the materially less well endowed societies would become a quite relevant question if the higher education institutions were effective in facilitating the connections from their side.

The foregoing considerations provide ample motivation for enquiring into the policies and practices of universities (in the developed and the developing regions) regarding FLOSS. Considering this, it is really surprising that so little empirical research has been focused on the subject. Previous to the study reported here, the only systematic survey-

[6] John Seely Brown, in a lecture entitled "Relearning Learning: Applying the Long Tail to Learning" delivered at MIT (in April 2007) elaborated on the changing face of learning in terms that emphasize the educational potentials of the kinds of distributed interactions with communities of practice that are envisaged here: "We learn through our interactions with others and the world", and there's no more perfect medium for enabling this than an increasingly open and organized World Wide Web....In a digitally connected, rapidly evolving world, we must transcend the traditional Cartesian models of learning that prescribe 'pouring knowledge into somebody's head'." For a summary and links to the lecture, see: http://www.checkpoint-elearning.com/article/3822.html.

based research addressing this range of topics (of which we are aware) is that carried out for UK universities and further education institutions by OSS Watch (2006).[7]

The data used in this paper have been extracted from the survey responses of two groups of university employees – administrative staff members and IT managers – totaling 446 individuals who reported on the conditions at 310 distinct HEIs. The first group, administrative respondents, includes Deans of Research, Vice-Provosts of the university, or other individuals in high administrative positions who must oversee operations of the university as an organization. The members of the group that we have labeled "IT managers" are occupants of a more varied assortment of university posts, including a Provost of Information Services, Vice-Provost of Information Technology, and other individuals (including those with teaching roles) who are responsible for managing the informational technology infrastructure of their institution. Universities are complex and highly variegated organizations, despite their outward similarities of purpose; they have correspondingly differentiated and complicated personnel structures whose members this survey did not explicitly seek to contact: a typical university employee works in a department or research institute that operates within a school, and is located on a specific academic university campus or in an urban facility. Moreover, schools and even entire university campuses may be only one local affiliate of a national university containing several such units. Adoption of FLOSS may differ within and between each of these organizational levels. In one university where individuals from several departments responded, we are able compare IT adoption policies and demonstrate the heterogeneity of FLOSS policies even within a single university. Although statistical averages and modal values are convenient in summarizing the data, considerable caution should employed in reaching for generalization about the way that FLOSS figures in work of the "typical" university community of any one of the countries in this survey, let alone in activities of the typical university student or academic workgroup in developing regions.

Conducting a study of open source software adoption by universities in economically less advanced regions, nevertheless, may yield further potentially useful insights regarding the important aspects of the development and growth process in a global context. One broad concern of economic growth policy is the formation of "absorptive capacity" in developing countries that will enable their producers to continue to identify, locate and successfully utilize scientific knowledge and technological information originating elsewhere, and particularly in societies where the scientific and engineering resource endowment is greater and average technical skill levels in the working population are higher. FLOSS itself is a technological artifact that is readily (and almost costlessly) transported, but whether it is "transferred" -- in the sense of being effectively absorbed into use – is a more complex matter. Studying the extent and pattern of adoption of FLOSS in developing and transition countries may therefore provide comparative measures of "absorptive capacity", as well as

[7] OSS Watch is the Advisory Service on free and open source software established by the Joint Information Services Committee (JISC) of the Research Councils in the UK, which is based in the Oxford University Computer Service. The OSS Watch 2006 survey is available at: http://www.oss-watch.ac.uk/studies/survey2006/survey2006report.xml. Also available are the findings of an informal survey (conducted by Barry Cornelius in April 2005) of Linux on desktop computers at Oxford University, and a May 2006 update presenting similar fragmentary data on FLOSS software use gathered from several other UK universities [http://www.oss-watch.ac.uk/studies/linuxdesktopssurvey.xml]. The OSS Watch 2006 survey employed an online questionnaire that extended a design piloted by a much smaller 2003 survey (also available on the OSS Watch website), and made use of many portions of the instrument developed for the FLOSSWorld HEI survey. Future research will therefore exploit the "developed country bench-mark" that this affords, offering direct comparisons with the responses reported here – subject, of course, to the variants in the wording of questions introduced by translation from the basic English version.

insights into possible mechanisms and institutional policies affecting the formation of capabilities for successful technology transfer in other areas.[8]

1.2 Organization of the Report and Overview

The presentation of our findings is organized as follows. Section two outlines the survey methodology and response rates. It emphasizes that the data represent a selected sample from predominantly technical universities in each of the seven countries. Within universities, the individuals who respond may be more interested in FLOSS than are the individuals who do not respond. So it would not be unreasonable to surmise that the picture created by our study describes the state of affairs among technology savvy individuals at the "leading edge" of universities' involvements with FLOSS, rather than the situation that is more typical in the HES as a whole.

Section three summarizes responses to most of the questionnaire items, and highlights the following five principle sets of descriptive findings:

- First, **reported average rates of FLOSS use and FLOSS development at universities vary substantially across countries, but FLOSS use lies in the range above 60 percent whereas FLOSS development typically lies in the range below 60 percent.** As might be expected, the reported use of FLOSS in these institutions is far more prevalent than development activity: whereas the national average proportions of universities that use FLOSS in some form fall in the range between 0.63 and 1.00, the corresponding range for the mean proportion that are reported to develop FLOSS lie in the range from 0.27 to 0.61 (save for handful of responding universities in Bulgaria, where reported rate averaged 0.93). Universities in Argentina and Brazil use FLOSS more extensively than do their counterpart institutions in other countries, especially those in China, among which the adoption of FLOSS is less common than the rest of the sample. Despite the prominence of the Indian Institute of Technology and the increasing role that information technology plays in India's economy, FLOSS does not have a comparatively prominent role in the countries universities of India. Overall differences between countries explain only 14 percent of variation in FLOSS use, so an institution's country says only about one-sixth as much about the institution's use of FLOSS as do the institution's other characteristics.
- Second, **views among administrators and IT managers** are not closely aligned **regarding the appropriateness of the share that their institutions' IT budgets** devote to software purchase and licensing fees, although there is a clear preponderance of opinion that budget shares in the range 0.20-0.40 "seem reasonable". A clear majority among the administrators are comfortable with software budget shares in the range up to .50; those that think otherwise appear to agree, on balance, that the shares in that range are too low rather than too high; whereas at each point in the budget range above .30 a clear majority of IT managers view their institutions' share as reasonable. Expectations that there will be a need to reduce expenditures on software are rather more closely aligned between administrators and IT managers at each level of the actual institutional budget range, but, perhaps not surprisingly, the consensus on the need for cuts that emerges among respondents whose institutional software expenditure shares are in the 0.30-o.60 range is considerably more pronounced in the case of the administrators.

[8] As the average university employee is likely to have a better educational preparation and higher skill levels than the average member of the work force, the implicit measures of absorptive capacity yielded by looking at FLOSS adoption and use in HEI's relate to a "best practice" rather an "average practice" concept of technology transfer.

- Third, **a substantial portion of universities ask technical job applicants about their FLOSS experiences** and give positive weight to such experience in making hiring decisions. This finding is consistent with, but by no means proves, the contention that an important motivation for developers to volunteer contributions to FLOSS projects is their expectation that such participation will improve their employment prospects in the software industry or professional careers by signaling their technical abilities and achievements.
- Fourth, **a majority of the survey respondents report that courses are offered at these institutions that would provide students with basic and advanced programming skills**, and courses in web management and advanced html appear to be no less frequently available. While this is important in initiating development of the abilities of graduating cadres entering employment to use and modify open source code, and permitting them to interact with and participate in the code development activities of international FLOSS projects, comparable skills development support generally is not afforded to university staff. Nor are these software skills course offerings equally extensive across the countries: whereas as average of 3 or more courses per institution is the country norm, with Bulgarian and Indian universities averaging more than 4 such offerings apiece, the average barely exceeds 2 courses per institution in Croatia.
- Fifth, **having a departmental policy for purchasing computer software which is clear or which supports FLOSS adoption is neither necessary nor sufficient for a university to attain high levels of prevalence in the adoption of FLOSS.** Many universities that extensively use FLOSS lack clear software policies, and many universities with IT "neutral" policies regarding procurement do not use FLOSS. **Such policies, however, when not anti-thetical to releasing FLOSS appear to be strongly associated with the reported presence of open source software development activity within the institution.**

Section 4 analyzes the correlates of reported FLOSS use, software adoption policy, open source software development activity, and IT personal hiring preferences, with the aid of two statistical methods – principal component analysis and multivariate regression. These techniques are explained briefly in the methodological appendix. It is found that cross-country differences explain only 15 percent of variation in the reported prevalence of FLOSS use. Knowledge of a university's country, in other words, says less about the university's use of FLOSS than does knowledge of other characteristics of the university. Survey data are relatively uninformative about FLOSS use, however, as data from the present survey explain only 20 percent of variation in FLOSS use across respondents.

Section 5 summarizes the report's findings with a view to their possible bearing on higher education and science policies aimed at human capital formation involving IT skills and the absorption of technologies embedded in software. A statistical Appendix presents auxiliary tables, methods for statistical tests and regression analysis, and further discussion of possible selection bias in the survey.

2. Survey Populations and Respondents

2.1 Respondent Individuals

Local survey affiliates in each of eight countries identified potential universities as targets for direct email requests to cooperate by responding to the survey. In most cases the university administrators and information technology (IT) personnel were contacted by direct email and informed of the online (Web) location of the questionnaire(s).

The survey responses thereby obtained includes the answers supplied by a total of 446 individuals, just over half of whom are IT managers. National sample sizes vary: China, the country with the world's largest population, has only 54 respondents, while Malaysia has

128. Bulgaria and South Africa have small samples of 7 and 12 respondents, respectively, while Argentina, Croatia, and India have intermediate sample sizes. The samples from Brazil, Croatia, and India substantially over-represent IT staff, while the sample from Argentina most under-represents IT staff. These differences in proportion across countries have robust statistical significance, but they could indicate different true ratios of IT to administrative staff in these countries, or they could reflect varying response probabilities across individuals and countries.

Respondents had mean age of 39.5, with the average South African aged 47 and the average Bulgarian aged only 35. Large web-based surveys of FLOSS developers place their mean ages in the range between 27 to 29 (David and Shapiro 2007), so the HEI survey represents an older population, reflecting the likely situation of university administrators and IT staff members, compared with considerably younger typical entrants to labor markets and the students who have been found to represent sizable proportions among the population of FLOSS developers at large. The mean ages of the HEI survey respondents for the eight countries show variations, as has been indicated, and the same thing may be said in regard to their composition by gender, which ranges widely: females were entirely absent among the respondents from South Africa, but represented more and two-fifths of the (comparably small) number of Bulgarian respondents. The mean of the female proportions in the whole sample (0.183) approximates the average that is found among the other, more substantial national samples, but even for that part of the data set the country-to-country differences are considerable – ranging from 0.11 to twice that proportion.[9] Whereas the respondents' mean age varies significantly across countries, in the case of their gender compositions we are able only marginally to statistically reject the hypothesis that in each country the proportion of females among responding university officers is the same.[10]

On average the survey required 20.7 minutes to complete, or just over one minute per question. A few respondents took much longer (Figure 1). National language imperfectly predicts time required for completing the survey, as the average South African respondent spent 21 minutes,[11] but the average Bulgarian and Chinese respondents spent only 11 and 14 minutes completing the survey, respectively. Connection speed, time required to think of an answer to a question, and other factors likely played roles in determining the time required for the survey. Survey duration was right-skewed, with a few respondents taking nearly an hour.

2.2 Respondent universities

To give a clearer picture of the institutions represented in these data, Table 2 provides descriptive statistics indicating the size and scope of the respondent universities' educational activities the several countries. For several reasons, calculating response rates for this survey is not trivial. Surveyors contacted one or more people at each university, and sometimes in different departments or campuses of one national university. Sometimes the person who responded from a university differed from the person who was contacted in that university. Further, surveyors and respondents refer to one university in many ways – via

---

[9] It is interesting to note that the mean proportion of females among these respondents closely approximates the (20 percent) share of females among the employess of proprietary software firms surveyed by the FLOSSPols Project (2006): the latter far exceeds the share of females among FLOSS developers that respond to web-surveys, which is in the neighborhood of 3 percent. In the case of the FLOSS-US survey, only 1.6 percent of the approximately 1500 respondents to that survey question identified themselves as female (see David, Waterman and Arora (2003).

[10] The Appendix Section A2 explains the Pearson χ^2 and ANOVA F statistics reported in some tables.

[11] In this instance, as elsewhere, we report sample statistics calculated after discarding outlier observations that are obviously erroneous: one of the South African respondents appeared to have taken more than 20 hours to complete the survey, most probably because they had interrupted their work but remained connected to the server, which eventually recorded their log-out time.

acronym, nickname, departmental name only, department and faculty, or any combination of the preceding options in English and/or local languages. Table 2 presents an "institutional response rate": the numerator is the number of distinct universities which had at least one person answer the survey, and the denominator is this number of distinct respondent universities added to the number of distinct universities which were contacted but did not respond. The response rate statistics are reported therefore approximate the portions of the contacted universities that had at least one person fill out the survey.

Overall, forty-eight percent of the universities invited to participate actually submitted a response, but the response rate varied from a high of 77 percent in Malaysia to a low of 24 percent in India. Since the survey offers nearly no data on the non-respondent individuals or institutions, the data allow no testing whether respondent institutions differ in observable ways from non-respondent institutions.

This response rate led 310 distinct universities to respond, implying that the survey contains just below 1.5 mean respondents per university.[12] Malaysia had the largest sample of individuals, but Croatia has the greatest number of universities, with 71 different institutions responding. Bulgaria had only five institutions respond, and South Africa had only 9.

A surprisingly high portion of universities only sent responses from IT staff. The survey requested responses from administrative personnel, and also requested administrative respondents to designate an IT manager who could either supplement or replace the administrative response. Given this survey design, one would expect a majority universities to have an administrative respondent, and some to also have IT managers as respondents in those instances where the administrator was unable to supply answers to technical questions; in a minority of cases, one would expect only responses from an IT respondent. The data, however, contain more responses from IT personnel than one might expect: 46 percent of universities have IT managers only, 35 percent have admin respondents only, and 19 percent have both types of respondents. The breakdown of respondent type varies across country: Brazil, China, and Croatia have relatively more IT managers, while Argentina has a disproportionate number of administrator-respondents.[13]

One-fourth of the responding universities represent technical institutions, as one might expect given the survey's announced focus on IT software. Technical institutions are identified from university names—institutions like "Binary" or "Multimedia" university are identified as technical, while others are not. University naming may reflect demand for different types of education in different countries, and it is striking to note that over 60 percent of respondent universities in Argentina and India are technical, while no more than a third of any other country's sample is technical.

Respondents to this survey also represent large universities, with mean enrollment of 13,500 students, about three-fourths of whom are undergraduates (Table 2). This reflects the presence of a long right-hand tail in the size distributions of the respondents' undergraduate and graduate enrollments. University size correlates somewhat with country size, as China has the largest universities (22,300 undergraduates and 31,000 graduates) and Croatia has

---

[12] When a national university has campuses in several cities, Table 2 identifies these campuses as distinct institutions. The existence of measurement errors in the size data is suggested by Figure 2, where "age heaping" is clearly present in the size distribution of the responses to questions about the numbers of undergraduate and graduate students enrolled: note the local peaks on 10, 20, and 30 thousand undergraduates; there are indications of heaping at 5, 7,5(?), 10 thousand graduate students.

[13] These results may say more about the assignment of job-titles than about the comparative technical expertise of respondence designated as administrative and IT managers, respectively. In presenting survey results, however, we consider the possibility that on average there may be differences between the typical occupants of those positions, and so separate the responses from the two groups for statistical analysis.

the smallest (1,100 undergraduates and only 770 graduates).[14] Indeed, of the 14 institutions in this survey with the largest student enrollment, 7 are Chinese. Administrative and academic staff sizes are somewhat correlated with student size—Brazilian and Chinese universities have the largest staff sizes, and those in Croatia have the smallest . For each of these enrollment and staffing variables, data show statistically significant differences across countries. Given the sample of 310 distinct universities, Table 2 implies that in total this survey represents universities teaching about 4.2 million students (=310 x 13,410).

Since Malaysia has the largest sample of 128 individual respondents, we disaggregate the responses for the 48 universities on which they reported, in order to obtain a quantitative indication of the degree of within-country heterogeneity (Table 3). Although most institutions are represented by one or two respondents, quite a few of these universities have many more than that. From University Putra Malaysia there were 18 respondents, constituting 14 percent of the Malaysian sample. Multimedia University provided the second-largest sample, with 8 respondents. By any measure, university size has little relation to sample size, perhaps due to varying response rates but possibly may be caused by reporting error in university size measures.[15] As one may see from Table 3, the variations in institution size as indicated by the structure of student enrollments, faculty and administrative personnel are generally consistent, but positions in the rank-order by size shows no systematic association with the institutions' place in the ranking by number of individual respondents.

3. Survey results

3.1 Work Responsibilities of the Respondents

The responsibilities of these respondents defy simple characterization. Half of IT managers hold an executive position as a manager or department head, while only a third of IT staff conduct research as a part of their work (Figure 3). Administrators juggle even more tasks, with over half doing some kind of teaching and nearly three-fourths holding an executive position. Although Figure 3 suggests that IT staff face varying levels of responsibility, a rather more comprehensive picture is presented of the technical skill/experience levels of IT managers, based on their responses to a subsequent set of survey questions concerning 7 specific technical capabilities and managerial capacities required by their positions (Figure 4).[16] Between half and two-thirds of IT managers reported that they have skills in the listed areas or have responsibilities in these areas as part of their work: 52 percent reported skill in database administration, 65 percent reported skill in administration, and intermediate proportions reported skill in software development, programming, internet/web design, ICT/software teaching, and network administration.

Most respondents – administrative and IT – have a variety of responsibilities for IT practices, making them potentially good sources of information for understanding FLOSS use in universities (Figure 4). Although nearly two-thirds of respondents purchase software, only 42 percent administer information and communication technology (ICT) budgets, and only a

[14] Nearly all enrollment levels reported in the survey are rounded to the nearest hundred or thousand, and administrative and IT staff may not know their institutions' exact enrollment, so these statistics probably include more measurement error than other statistics do.

[15] A least squares regression of the number of survey respondents from a university on variables for that university's enrollment and staff size reveals no association larger than 0.0001, with no statistical significance for any parameter estimate. The existence of measurement errors in the size data is suggested by Figure 2, in which the presence of "age-heaping" in the size distribution of reported enrollment numbers already has been noticed.

[16] The item asks whether IT staff have "personal technical skills and/or responsibilities in your employment position," so responses may include all staff that have technical skills in an area, regardless of whether the respondent's employment requires use of the skill.

third design or approve software licensing agreements. Over half develop and/or implement institutional ICT policies. These tasks require expertise in several subjects – technical expertise for identifying and implementing required ICT systems, legal expertise for designing or approving software agreements, and fiscal expertise in designing budgets – so it is unsurprising that respondents fill some but not all of these tasks. From Figure 5 it is seen that the pattern in the proportions of the various skills and responsibilities are quite similar for the subsets of university administrators and IT managers.

3.2 Institutional Decision-making for Software and IT Expenditure Shares

Advocates of FLOSS and advocates of proprietary software each argue that the respective merits of their products will cause some classes of users, if not all users, to adopt them in preference to the alternative. In many universities, as is the case in many business organizations, a centralized procurement process determines the software that manages systems and appears on individual users' computers. Indeed, 69.3 percent of respondents to the HEI survey agreed that decisions regarding software purchases for their institution are "made on an institution wide basis." Regardless of the substantive attributes FLOSS or proprietary software products in particular contexts of application, it seems likely that the process by which universities choose software will influence the software they ultimately install. For instance, one might suppose that IT managers, if they had the decision to make, would select different software from that which would be chosen by financial or other managers in the ranks of university administration. Understanding the locus of departmental decisions about procurement of software, then, may help understand why some departments and, indeed, some universities do or do not adopt FLOSS.

IT managers clearly play the most important role in choosing software—nearly half of respondents identified their institution's IT manager as most important in this decision, and another fourth of respondents said that the individuals holding that position was second-most in importance (Figure 6). Users play an important role in 40 percent of institutions, whereas financial and other managers play less important roles, and external consultants have effectively no role to speak of. Panels A and B of Figure 6 show that there is a generally close agreement of views between members of the general university administration and IT manages concerning the locus of influence in decision-making on these matters; the proportion among IT managers that rate their role as dominant is only slightly bigger than the proportion among administrators who accord them top importance.[17]

Only half of respondents offered answers to questions on the composition of IT budgets, which could reflect lack of knowledge on these budgets, unwillingness to divulge this information, or other reasons. Respondents who estimated the composition of IT budgets indicated that a fourth of expenditure goes to software and license fees while a tenth goes to IT personnel (Table 4). Malaysia and India dedicated the greatest portions of their IT budgets to software and license fees, while Bulgaria and Croatia recorded the lowest proportional expenditure on software and license fees. Statistics on budget expenditure for software explicitly excluded expenditure on IT personnel, though the survey did not specify whether these data included contracted-in personnel, such as database consultants from external firms.

[17] There is only one notable point of disagreement in the assessments reported in Panels A and B of Figure 6. 29 percent of the IT managers say that financial administrators have either a top or second-tier importance in software purchasing, and 23 percent of them attribute that influence to "other management"; by contrast, 36 percent of administrators give "other manangement" personnel primarily or secondary influence, and only 18 percent of them say that financial managers are at the top or second position in these decisions. IT managers appear to have a less nuanced perception than do the administrators of exactly who it is – if not they – that have sway in software purchasing decisions. This is perhaps also reflected by the relatively larger frequency (17 vs 10 percent among administrators) with which top influence is attributed to "others" by the IT managers.

Opinions about the “reasonability” of the current levels of expenditures on software purchases and license fees also vary widely among institutions and countries. In India, 71 percent of respondents felt that expenditure on these items was reasonable, while among Croatian respondents only a third took that view, and nearly 60 percent of Croatians stated that expenditure on these items was too high. Views about the reasonability of software and license fees did not strongly correlate with opinions about the need to change expenditure on these items in the near future. In Malaysia, for example, only 24 percent of respondents felt that expenditure on these items was “too high,” while 69 percent identified a need to reduce expenditure on them in the next two years. In Croatia, by contrast, 59 percent of respondents thought that expenditure on these items was too high, but only 7 percent saw a need to reduce expenditure on them in the near future (Table 4).

By combining observations on the actual IT budget structure with those on the respondents’ opinions about the appropriateness of the fraction of their institutional or department IT budget that was devoted to software purchasing and license fees, it is possible to form a view of whether or not there exists a clear consensus on this question, and whether in the absence of funding shocks such a consensus is essentially stable or would generate pressures for substantial reallocations of IT budgets. Figure 7 (Panels A and C) presents these results: there is a clear preponderance of opinion that budget shares in the range 0.20-0.30 “seem reasonable”, but the views among administrators and IT managers on the appropriateness of the share of their institutions’ IT budgets that is devoted to are not perfectly aligned.[18] A substantial majority of the administrators (c. 60 percent) are comfortable with software budget shares in the range up to o.45-.50; those that think otherwise are on balance of the view that shares in that range are too low rather than too high. By contrast, as actual software expenditure shares rise from 0.10 to 0.30 an emerging majority among IT managers view the software share of their IT budget shares as being “reasonable”; this reaches the 60 percent level when the software budget share is 0.40, and thereafter it continues to drift slightly higher throughout the range of ever-larger software shares. The appetite for more software appears not to be easily sated among IT managers. Indeed, among those not content with their actual budget shares, the preponderant opinion favoring higher relative expenditures for software remains positive throughout the entire range from 0.30 to 0.80.

Perceived needs to reduce expenditures on software in the course of the coming years are more closely aligned between the two groups (Figure 7, Panels B and D). A clear majority of opinion supporting that forecast emerges among both the administrators and the IT managers whose budget shares are in the range from 0.30 to 0.60. But the strength of that majority is more pronounced among members of the administration, where more than 60 percent always see a coming need to reduce outlays for software and 75% express that opinion by the time one reaches software budget shares of 0.60. The conjunction of the two sets of views suggests that if it is not possible to meet currently desired rates of software acquisition with lowered expenditures vis-à-vis the overall IT budgets in these institutions, there will be very substantial discontent – especially in the ranks of IT managers.

[18] The survey questions for Administrators (Q.10) and Managers (Q.17) asked explicity for their views regarding “the share of software purchases and licsence fees in the total IT budget,” rather than about absolute levels of expenditures on those items. This had the advantage of “normalizing” the responses for each institution and avoiding a need to convert different currencies in order to aggregate across countries, especially as exchange rates may well not reflect purchasing power relatives in software and IT equipment. The drawback, of course, is that the question did not elicit information regarding whether or not the level of the funding for software is regarded to be adequate.

3.3 IT Strategy, Use of FLOSS, and Development of FLOSS

One motivation for undertaking this survey was to understand how and why the use of FLOSS varies across countries. Table 5 describes one potentially important element of software adoption: formal departmental IT policies or strategies. Fifty-eight percent of respondents indicated that their department had a formalized policy: Malaysians were most likely to have such a policy (92 percent prevalence), while Brazilians were least likely to have one (29 percent). Three-fourths of stated policies mentioned FLOSS as an option for procurement, giving a good basis for FLOSS use. About 90 percent of institutional IT policies in Argentina and 95 percent in Brazil, mention FLOSS.[19] Even more surprisingly, over a third of IT policies require procurement of FLOSS software if it is available.[20] These requirements are most common in Brazil, India, and Argentina, but relatively rare in China and Croatia and completely absent in South Africa. In those policies that did not mention FLOSS, a large portion (67%) either did not discuss software or did not specify any type of software. A small portion (15%) focus on proprietary software, and only 2 institutions – one in Croatia, and one in Malaysia – had policies that explicitly excluded FLOSS.

In short, the institutions in this survey are only moderately likely to have a formal IT policy, but most formal policies mention FLOSS – nearly all formal policies in Latin America do – and over a third of these policies require acquisition of FLOSS software if such acquisition is possible.

Although not all institutions have IT policies that mention FLOSS, many institutions and administrators use FLOSS. Nearly 90 percent of respondents indicated that their institutions use FLOSS, but use of FLOSS differs across groups within an institution (Figure 8). In only 9 percent of institutions do administrators use FLOSS “a great deal,” and in half of respondent institutions, administrators use FLOSS only “a little.” Teaching staff and non-science students are even less likely to use FLOSS, while 6 percent of teaching staff and 17 percent of computer science students use FLOSS “a great deal.” Among all these groups, computer science students show the greatest use of FLOSS. (The Lickert scales used for responding to these questions do not allow simple comparison across countries, but in Section 4 we develop a scalar index summarizing use of FLOSS in institutions and compare this index across countries.)

To simplify presentation, we only disaggregate FLOSS use by country for the “whole institution” sub-question, and not for the sub-questions focusing on specific academic groups, but we report the responses from administrators and IT managers separately (Figure 9). Again, Latin America has the greatest prevalence of institutional FLOSS use, with 98 and 97 percent of institutions in Argentina and Brazil, respectively, using FLOSS. FLOSS use is least common in China, with only 73 percent of Chinese respondents using FLOSS. A statistical test rejects the hypothesis that FLOSS use is equal across countries.

The aforementioned statistics do not measure the intensity of FLOSS use within an institution, but among respondents there is general agreement that of the extent of FLOSS use should be increased for all academic groups (Figure 10). Overall, 95 percent of respondents believe that FLOSS use needs to increase in their particular institution, and the proportion within each of the countries does not vary notably from that high level (Figure 11): a statistical test fails to reject the hypothesis that the proportion holding that opinion uniform across all the countries.

Although Figure 8 showed that computer science students are more likely than other academic groups to use FLOSS survey, from Figure 10 it is seen that there was strong consensus among both Administrative respondents and IT managers that greater use of

[19] Only 4 Bulgarians responded to the question in column (3) of Table 4, and no Bulgarians answered the questions in columns (4) to (7).

[20] Note that column (3) of Table 4 only includes respondents who answered “yes” to the question in column (2), so *0.49*0.74 = 0.36.*

FLOSS was most relevant for computer science students: 88 percent of the former and 85 percent of the latter respondents thought that FLOSS use should be increased "much" or "a great deal" among computer science students. Strikingly, 87 and 83 percent of these respondents, respectively gave the same answers to the question in the case of the teaching staff. The proportion of IT managers that called for FLOSS use to increase "a great deal" or "much" among Administrative staff was somewhat higher (77 percent) than the proportion holding those views (69 percent) among the Administrative respondents themselves. There was close agreement in the strength of the two groups' consensus on the desirability of increasing the use of FLOSS among science students other than computer scientists, but Administrators were less inclined than IT managers to take the same position in regard to students outside the sciences.

The preceding questions have concerned respondents' reports on FLOSS use in their respective institutions. Since it is unlikely that they have precise knowledge of the prevalence of FLOSS use among particular groups (for example, non-science students), the foregoing data regarding their opinions is potentially subject to substantial divergences from the actual situation on which the respondents were commenting. Figure 15 offers a means of assessing the correspondence between opinion and reality in this matter: it shows the country-level prevalence of experience with FLOSS use among the survey respondents (administrators and IT managers combined). At an overall level, comparison with the data from Figure 8 suggests a reassuring level of consistency. According to Figure 15 the global proportion of respondent's reporting that they personally used FLOSS at work is 31 percent, whereas from Figure 8 it will be seen that administrators reported that 33 percent of the teaching staff and same percentage of their own colleagues were using FLOSS "a great deal" or "much, whereas IT manages gave 24 and 33 percent for the corresponding percentages.

There is, in addition, a reassuring degree of agreement in the two sets of observations if one uses them to rank-order the countries according to the degree of institutional use of FLOSS. For this purpose we can use the proportion reporting any level of FLOSS at work (from the compliment of the percentage reporting "None" in Figure 15), and the mean of the national proportions given by the administrators' and IT managers' who reported on institutional use of FLOSS (in Figure 9). The ordering of 4 highest countries on the latter measure ranking of countries is Bulgaria (100%), Croatia 89%), Brazil (76%), Malaysia (67%), and these 4 are also in the 5 top-ranked group on the measure constructed from Figure 9, which includes Argentina (99%) as well. Aside from the anomaly of Argentina's position, which is at the very bottom rank according to Figure 15's measure of any own use (31%). This gross agreement is only mildly reassuring, however, as even after omitting Argentina the rank correlation within the group formed by the top 4 is not perfect; and the same must be said in regard to the rank correlation between the bottom 4 (with Argentina again excluded) according to Figure 15, and the estimates from Figure 9: China comes lowest (at 69%) according to Figure 9, but on the basis of the estimates from Figure 15 it follows Malaysia and outranks India (53%) and South Africa (50%). Of course, it is quite possible that there are marked variations among these countries in the relationship between the extent of institutional use and the prevalence of FLOSS use among institutional employees who responded to the survey. The importance of student use, given variations in the relative importance of high- and low-use students in the sciences, could account for the looseness of the cross-country rank associations that have been examined.

The data based on "own use" also paint a picture of considerable inter-country variations in the circumstances of FLOSS use among personnel employed at HEIs. In Brazil, 76 percent of respondents use FLOSS at work, and about half of these respondents had used FLOSS outside work as well. In Croatia, 89 percent of respondents use FLOSS in some capacity, and 76 percent used it both privately and at work.[21] Two-thirds of Malaysian respondents had some personal experience with FLOSS, half of Indian respondents did, but among

[21] Only 3 Bulgarians responded to this survey item.

Argentinean institutions' respondents that proportion was only a one-third. A statistical test confirms that respondents in different countries have significantly different experiences in both the extent and circumstance of their use of FLOSS .

As important as these academic institutions' use of FLOSS may be in terms of its effects upon skill formation among their students, or in releasing IT budgets for other purposes, it is also relevant to consider the extent to which university-based developers in these regions are active as contributors to FLOSS production. After all, the prevalence of participation in the development of open source software among students is itself a indicator the degree to which expertise has been acquired in programming skills and understanding of software systems; and in the case of faculty it should be informative about the capabilities of the instruction that the institution is able to provide for those seeking to develop those skills. FLOSS development work can take many forms in the university setting : faculty and students might develop programs for their own use but then post these programs at an online code repository; students, staff or faculty, might contribute to one or another of the large FLOSS projects, such as Gnome and KDE which garner the most media attention, or they may collaborate to form small, team-based FLOSS development projects relating to their training, or create such software in the course of their individual research work, or participate in extensive multi-institutional scientific collaboration that develop highly specific FLOSS program to support their activities. This survey does not distinguish among these various forms of FLOSS development. Nor does it report code levels of software development activity such as numbers of "commits", lines of code contributed, average hours spent, average proportions of institutional personnel involved, nor any other measures of the intensity of participation in FLOSS production. Nonetheless, the survey provides a "first look" at the proportion of universities and of specific academic groups within these responding HEI's that engage in such activities.

Overall, approximately half of the institutions represented in the responses were contributing to FLOSS development (Figure 12). This statistic may reflect an average over time; it may reflect an average over individuals, or it may reflect a subjective judgment about the level of FLOSS participation necessary for a person to respond that the entire institution develops FLOSS.

Computer science students have the greatest involvement among all the academic groups listed in Figure 12, with one in ten institutions having computer science students that contribute "a great deal," and four in eight institutions having such students contribute "much" or more intensely to FLOSS. Half of institutions report that students in other sciences have little or no involvement in FLOSS, and three-fourths of institutions report that non-science students have little or no involvement in FLOSS. The extent of participation among administrative staff members falls somewhere between that of the computer science students and the other science students at these institutions. Teaching staff are nearly as likely as computer science students to contribute to FLOSS.

The finding that computer science students as a group contribute somewhat more frequently to FLOSS development than faculty members should not be particularly surprising. Firstly, since most computer science students develop software as part of coursework, computer science students face a low cost of putting homework-based FLOSS software online. Teaching staff, by contrast, may face greater obligations to develop teaching material and research papers, which may require substantial time investment to convert into FLOSS programs. Secondly, the responses regarding faculty do not relate specifically to members of science and engineering departments (except in those instances where the institution itself is small and quite specialized in those fields). Thirdly, the collaborative atmosphere of study in some universities could encourage students to work together with distant peers by developing FLOSS. Lastly, FLOSS developers emphasize the important roles of FLOSS in building programming skills and signaling ability to potential employers (see David and Shapiro 2007), and these human capital formation and competence signally motivations may have greater relevance for students, who may have yet to enter the labor market, than those already holding jobs on the teaching staff.

What may be rather more surprising, however, is that the prevalence of FLOSS development activity does not vary across countries in a way that is associated with the varying prevalence of FLOSS usage at these HEI's, at least not at the aggregate levels reflected in cross-country comparisons. Figure 8b showed that institutions in Latin America were most likely to use FLOSS, whereas institutions in China were least likely to do so. Figure 13 shows that the prevalence of FLOSS development in the two major countries in Latin America exceed the 0.50 mark, and exceeds the survey-wide average (0.47) level, but South Africa and Bulgaria also report very high rates of FLOSS development even though these two countries also have the smallest samples (6 and 12 responses to the relevant survey items, respectively). FLOSS development is least prevalent in Croatia's respondent universities, yet FLOSS use there was very widely reported by IT managers, and only somewhat less so by upper—level members of the administration.

3.4 Programming Courses offered

The level of FLOSS use in an institution may depend on the level of programming skills in that institution, and the presence of courses which teach programming skills may explain some of the variation in FLOSS prevalence across institutions and countries. Most institutions surveyed offer standard courses to students: introductory programming, advanced programming, and simple and advanced html. These high percentages partly reflect the fact that many respondents represent technical institutions and often computer science and engineering departments within technical institutions. Surveys of humanities faculty or liberal arts institutions would probably find far lower prevalence. The high prevalence of html courses may reflect market demand for programmers to develop sophisticated web pages, but only a third of respondent's institutions offer a shell scripting course. In general, these courses are more widely made available for students than for university staff (Figure 16), and this holds *a fortiori* for the more advanced software course offerings. For example, the average ratio of staff-to-student course in basic html is 0.63, but the ratio is 0.47 for courses in advanced html, and 0.50 for those in shell scripting.

To simplify comparison of these data across countries, we calculate the within-country mean of the total number of these courses that each institution offers. A statistical test confirms that total course offerings differ significantly across countries. The first bar of Figure 17, for example, shows that surveyed institutions in Argentina on average offer four of the seven options that are listed. Average course offers are most ample at the Indian institutions represented in this survey, where the average HEI offers nearly 5 of the 7 listed options: whereas in Croatia, by contrast, the average institution barely offers two of the courses mentioned. From Figure 17 China is seen to come next lowest to the bottom rank in the average number of course offerings per institution, and it was third from the bottom rank in the prevalence of university-based contributions to FLOSS production (Figure 13). In these two cases, a thin range of course offerings in programming, software and web skills, appears to go together with limited FLOSS development activity within the universities.

3.5 The Role of FLOSS Experience in Hiring Employees

Economic research on FLOSS development has debated the extent to which the interest in signaling skills or experience to future employers motivates developers to contribute to FLOSS programs (see Lerner and Tirole 2002, Ghosh et al. 20005, David and Shapiro 2007 and sources reviewed therein). Unlike surveys of FLOSS developers, the HEI survey is in part a survey of employers, since Universities hire some computer scientists and software developers as faculty and/or staff members. This provides an unusual opportunity to examine if non-business employers actually take FLOSS experience into account (whether positively or negatively) in their hiring decisions. Employer consideration of the FLOSS experience of job applicants as a positive qualification is a necessary but not sufficient condition to justify the view that developers voluntarily contribute to FLOSS development because they hope to thereby signal their software skills and related capabilities to prospective employers. Of course, to establish that signaling plays a role in developers' motives cannot in itself reveal the extent to which other motivating factors -- such as ideological commitment to "sharing", or intrinsic satisfactions and the desire to improve skills -- also influence the behavior of contributing FLOSS developers.[22]

Given these *caveats* about the interpretative conclusions to be placed on the data, the survey results confirm that university employers do give positive weight to the FLOSS experiences of job applicants. Overall, 54 percent of respondents report having asked job applicants about FLOSS experience during interviews (Figure 18). The questionnaire does not clarify whether all of the respondent's interviews have involved such questions, or whether the issue arose at only one interview. Only a fifth of these positions, however, involved leadership roles.[23]

At a minimum, then, one can say that the data militate against the view that university employers regard FLOSS experience as irrelevant in filing some faculty and staff positions. From Figure 19 it will be seen that 57 percent of respondents state that some positions in their institution require FLOSS experience, with FLOSS experience being most valued in Brazil and least valued in South Africa. There is in this regard as elsewhere a strong institutional emphasis on FLOSS in Brazil, asking job applicants about FLOSS experience is reported less frequently there than in Malaysia's HEIs (Figure 11c). At the opposite end of the spectrum, the practice of asking such questions about job candidates' experience with FLOSS is least prevalent in Bulgaria, Croatia, and South Africa, and a statistical test confirms that the cross-national differences in this regard are statistically significant.

The responses to a hypothetical job interview situation from administrators and IT managers further confirms the salience accorded to FLOSS experience in hiring decisions at these HEIs. The survey posed the following question to all the responding individuals:

> Suppose you were to face a choice between two prospective employees (person A and person B) with exactly the same level of formal qualifications but different experiences: Person A has proven experience developing an important component of a proprietary software product, as an employee of a proprietary software company. Person B has proven experience developing an important component of a free software / open source software product of equivalent complexity, as an independent participant of the developer community. Would you be more likely to hire person A or person B? *Please select one answer only*.

As Figure 21 indicates, the respondents overwhelmingly expressed a preference for the hypothetical FLOSS–experienced candidate to an alternative who had had experience with proprietary software development instead. Indeed, fewer than one in ten respondents said that they would give the job to the otherwise equally qualified candidate who had experience

---

[22] The FLOSSWorld survey of developers explores these motivations, and focuses in considerable detail on developers goals and expectations with regard to the acquisition and improvement of both software programming and a variety of skills and experience that may be acquired by participating in FLOSS development projects. Attitudes and practices of business employers in regard to their hiring criteria, which may be compared with those of the non-business (HEI) organization discussed here, were surveyed more-or-less at the same time in the same countries as part of the FLOSSWorld project. Comparison of the two perspectives, and the expectations of employment seeking developers in each country is a topic left for future research.

[23] Surprisingly, 18 percent of respondents state that they do not ask job candidates about FLOSS experience but that some positions in their institution require FLOSS experience. This statistic may arise because few positions require FLOSS experience and hence most interviews do not focus on FLOSS experience; because Universities learn about FLOSS experience from a candidate's resume rather than asking a candidate directly about FLOSS experience; or simply due to misreporting on one of the involved survey items.

in a proprietary company. The variations in the reported country proportions on this question around the overall proportion of 57 percent favoring candidates with FLOSS experience are not inconsistent with other indicators of pro-FLOSS sympathies: Brazil consistently displays the greatest support for FLOSS, with two-thirds of Brazilian respondents reporting that they would prefer the candidate with FLOSS experience. Argentina's respondents declare somewhat less preference for job applicants with FLOSS experience, although the low, 40 percent among them that declared a preference for FLOSS-experienced candidates (over one that only had worked for a proprietary software firm) did represent 0.8 of those respondents that expressed any clear hiring preferences at all. A statistical test cannot, however, reject the hypothesis that the portion of respondents who prefer the FLOSS–experienced job-candidates is the same across all these countries.

This majority preference for FLOSS experienced job-applicants may reflect a variety of factors. First, if respondents themselves contribute to FLOSS projects, they may share broad ideological sympathies with other FLOSS developers, and hence expect that a FLOSS developer would have a style of work or personality more similar to their own and would "fit in" with their colleagues and co-workers. Second, respondents may judge that even if two job applicants had worked satisfactorily on projects of equivalent technical complexity, the candidate with FLOSS experience would have received more criticism from peers and learned more about various aspects of the software development process than a person who had been employed as a programmer in a proprietary software company. Third, given that 57 percent of respondents state that some jobs in their institutions require FLOSS work (Figure 20), it is not surprising that 54 percent of respondents would display a preference for job candidates who have some FLOSS experience. A statistical test does not reject the hypothesis that the portion of respondents who prefer the FLOSS candidate is equal across countries. But the variations in the reported country proportions on this question are not inconsistent with other indicators of pro-FLOSS sympathies: Brazil consistently displays the greatest support for FLOSS, with two-thirds of Brazilian respondents reporting that they would prefer the candidate with FLOSS experience. Argentina's respondents declare somewhat less preference for job applicants with FLOSS experience, but, the low, 40 percent of them said they would give preference to a FLOSS-experienced candidate (over one that only had worked for a proprietary software firm) nonetheless represented 0.8 of those respondents that expressed any clear hiring preferences whatsoever.

In appraising the implications of these findings one should bear in mind the heavy representation of technical institutes in this sample of HEI's; also, the individuals to whom the survey was addressed were (in the case of those with IT management responsibilities) most likely to be hiring for positions in which software skills and experience of some sort would be an essential qualification. Nevertheless, it may be relevant to note that individuals who have experience as contributors in large FLOSS projects are likely to possess not only coding skills and familiarity with software tools, but a capacity to quickly grasp the organizational structure of large and complex community processes, and to interact productively with others in situations that allow individuals considerable autonomy while providing them with scant explicit management direction. That bundle of attributes is likely to be perceived by job interviewers to be particularly useful qualities for faculty and non-faculty employees in an academic institution.

3.6 Software use

A final set of questions, directed only to IT staff, asked respondents to list the software programs that the institution used on central server computers and on Desktop PCs. Windows remains dominant: 94 percent of respondents said that desktop PCs use Windows XP, and 67 percent of respondents indicated that central servers use Windows Server 2003. GNU/Linux has about equal popularity, with about 84 percent coverage of central servers and 55 percent coverage of desktop PCs. Sun's Solaris operating system is also popular for central servers. Presumably these universities have different servers running different operating platforms, explaining the high coverage statistics for Windows, GNU/Linux, and Solaris.

Some prominent FLOSS programs, such as FreeBSD and OpenBSD, were rarely used, with less than one in seven respondents using these on a central server or on a PC. In short, widespread support for and development of FLOSS programs at these institutions has not led these institutions to abandon use of Windows, or to pursue widespread adoption of many FLOSS programs besides GNU/Linux for PCs and central servers besides (Figure 22).

Responses to a subsequent question on whether a software program is used at all in an institution, however, reveal much more widespread use of FLOSS programs. Five in six respondents said their institution used Apache—a high rate of prevalence. Other major FLOSS programs also had achieved extensive penetration of the HEI market —83 percent of responding universities used GNU/Linux; 41 percent reported using GNOME; 48 percent used KDE; and other smaller programs had equally high prevalence (Figure 23). The difference between Figures 14 and 15 may arise because program use is highly varied within institutions, so although most FLOSS programs appear on at least one computer in each institution – giving the high rates of prevalence for Figure 23 – few FLOSS programs run on most PCs or servers in each institution, yielding the lower rates of prevalence seen in Figure 22.

The survey questionnaire's design allows for a simple examination of the extent to which survey respondents are aware whether or not the programs their institutions are using are FLOSS. An earlier survey question asked respondents to indicate whether their university used FLOSS, and the preceding questions ask IT managers whether their university uses specific FLOSS programs. Comparing these two responses provides an indication of the extent to which IT managers recognize either that some or none of the programs they use are FLOSS. Table 6 suggests that over 90 percent of IT managers correctly realize that at least some programs running on their computers are FLOSS. Only 8 percent of IT managers – 20 individuals – indicated that their university did not use FLOSS, but subsequently listed specific FLOSS programs running on their university's systems. Awareness of whether specific programs are proprietary or FLOSS, at least, is high.

3.7 Concluding Observation on the Within-Institution Diversity of Practices

The foregoing review of findings on the various aspects of university policies and practices affecting the extent of use of FLOSS, the provision of instruction in skills pertinent to creating software, staff technical capabilities and the hiring preferences that affect the availability personnel familiar with open source computer programs and the methods of creating them, should not be allow to promote the impression that there is substantial homogeneity of practices in all these respects through the universities from which the survey data have been drawn. Table 7 therefore should serve as a caution against slipping casually into thinking in those terms. Using the information supplied by multiple respondents who reported on the state of affairs as view from distinct departmental vantage points within a single institution, it reveals the existence of inter-departmental diversity of practice in every one of the dimensions itemized, save the first, which shows a reassuring consistency of reporting on the existence of an institution-wide IT policy. When considering the statistics formed by aggregating one or two observations from a number of separate institutions in a given country, it is therefore appropriate to bear in mind the likelihood that the variation that one finds within such a sample is just as likely, if not more likely to arise from the diversity of the places within those institutions from which the respondents have been drawn, as from inter-institutional differences. Or to put the point another way, differences in the relative representation within distinct HEI's of the various types of departments that appear in Table 7 are a likely source of observed inter-institutional variations in practices.

4. Correlates of the role of FLOSS in an institution

Some of the discussion to this point has ventured speculations about the possible interrelationships among the responses to different topics covered by the survey questionnaire. To explore the data in a statistically rigorous and more systematic fashion, we may we first apply principal component analysis to a selection of an array of survey item responses, in order to construct five indices that capture the various main roles of FLOSS within higher education institutions. We can then compare the indices across the eight countries, and use regression analysis to measure the strength and statistical significance of association of these FLOSS indices with observable individual- and university-specific characteristics. We present and discuss these exploratory results as an essentially descriptive exercise, rather than a test of specific behavioral hypotheses or an effort to estimate underlying structural relationships.

4.1 Discussion of constructed indices

We construct the first index, which reflects the extent to which institutions use FLOSS, using six survey items (Appendix Table A1). Missing observations have been recoded for this purpose to have value zero, and we include indicators for non-response to reflect the recoding.[24] This first index increases as an institution's *use* of FLOSS increases. The index is based on one survey item, which asks whether the institution overall uses FLOSS, and five items which have Lickert scales for responses indicating the extent to which particular groups use FLOSS. For each of these last items, the response "not at all" has a negative scoring coefficient and most (though not all) of the "a little," "much," and "a great deal" responses have positive coefficients. Although these coefficients do not always increase in size as reported extent of use of FLOSS increases, the general pattern shows that increases in this index reflect increased use of FLOSS.

The second index reflects the extent to which institutions *develop* FLOSS (Appendix Table A2). The variable with the largest scoring coefficient is a binary indicator for whether any group in the institution develops FLOSS. The subsequent four items again question the extent to which particular groups in the university develop FLOSS. None have negative coefficients, but generally the smaller coefficients appear on the response, "Not at all," and the larger coefficients appear on the response, "A great deal."

The third index reflects the extent to which an institution's *software procurement policy strategy supports* or coheres with use of FLOSS. The three dominant contributors to the index are binary questions for whether the institution has a stated IT policy or strategy; whether the stated policy or strategy includes FLOSS as a procurement option; and whether the policy requires FLOSS. Again, more positive values of the index reflect more favorable policies towards FLOSS.

The fourth index measures the extent to which the institution has a *clear licensing policy for releasing software*. The largest coefficient appears on the first variable, which measures whether the institution has any policy developed by students or staff. The subsequent three questions have Lickert scales reflecting the extent to which the institution releases software under commercial licenses, FLOSS licenses, or for free. These coefficients do not display a clear pattern, so we interpret this index as reflecting the clarity of a license policy rather than the coherence of policy with a particular type of license.

The final index reflects the extent to which a respondent's institution *favors FLOSS experience in hiring* job candidates. Its four constituent questions – whether the institution

[24] Krueger and Zhu (2004) explain the rationale for this method in one context. The approach somewhat resembles that of the statistical mean-shift outlier model. We judge the use of indicators for non-response to be less subjective than imputation, and the potential bias of using indicators to be less severe than that of dropping all observations which fail to answer any relevant survey items.

asks job applicants about FLOSS experience; whether the institution has positions requiring FLOSS experience; whether such positions usually involve leadership; and whether, given a choice between candidates with proprietary and FLOSS experience, the institution prefers the FLOSS candidate – all have large positive coefficients.

4.2 Cross-country differences in FLOSS indices

We first consider the extent to which these factors correlate with each other—that is, the extent to which use, development, policy, and other FLOSS -related characteristics vary in different directions even within the same institution. Table 8, a matrix of pairwise correlations between these indices, shows that all but one of the pairwise correlations is positive, and half of them are statistically significant, implying that there is a notable amount of positive correlation across institutional characteristics related to FLOSS. But no correlation coefficient exceeds 0.39, and the mean pairwise correlation is 0.16, showing that these indices are not homogenous within institutions. The extent to which FLOSS is being developed in a HEI is the best single predictor of whether or not that institution has a clear policy for software licensing, whether the institution's policy supports FLOSS use, and whether FLOSS is considered in hiring job applicants. One realistic possibility is that administrators and academic staff who develop FLOSS become more supportive of using FLOSS on the institution's computers, hiring job candidates with FLOSS experience, and using permissive licenses to release software developed in-house. But another possibility consistent with Table 8 is that these different dimensions of institutional policy are not casually interrelated, but have a common latent cause: influential individuals in the institution may, due to their education, personality, or other reasons, be predisposed to develop FLOSS, prefer using FLOSS and hiring FLOSS-experienced job applicants, and it is their prior orientation that reflects itself in the policies of their institution. While it is not possible here to identify which of those possibilities underlies the pattern in the responses, one should note that the strongest correlation in Table 8 appears between having a clear licensing policy and the prevalence of FLOSS development. Whatever was the direction of the causal process that gave rise to this association in particular instances, it seems reasonable to suppose that in the sample as a whole both paths of influence had been present: university authorities who became aware that FLOSS development was being undertake in their institution would be inclined to set a policy about licensing, and the announcement of a clear policy permitting free and open source licensing of software developed with university resources would be likely to encourage the development of FLOSS by faculty, staff and students.

The indices for FLOSS use, development, and policy framework have statistically significant differences across countries, though the indices for type of software release and the role of FLOSS in hiring differ only marginally in that respect (Table 9). Consistent with the results noted in earlier sections of the report, Table 9 shows that Brazilian institutions are most likely to use FLOSS. HEI's in Bulgaria and Argentina also are likely to use FLOSS, whereas in Malaysia and South Africa that is somewhat less likely, and universities in China, Croatia and India – in that order – are still less likely to be found using FLOSS. Although the units of the within-country means of these indices do not have a clear interpretation, the large and somewhat surprising gap between India and other countries may reflect substantial differences in FLOSS use. Differences across countries explain 14 percent of the variation in FLOSS use, so while cross-country differences are important, within-country differences explain the great majority (86 percent) of the variation. To put that another way, if one wished to anticipate the extend of a randomly chosen institution's use of FLOSS, knowing the institution's national location would be only about one-sixth as informative as knowing the institution's other characteristics.

A somewhat similar pattern appears for the index reflecting development of FLOSS: universities in South Africa, Bulgaria, and Brazil are most likely to be developing FLOSS, those in Argentina, India, and Malaysia are somewhat less likely to do so, and HEIs in

China and Croatia are far less likely to be sites of FLOSS development activity. Given the correlation coefficient of 0.20 between the indices for FLOSS use and development (Table 8), the similarity of country rankings based on these measures is unsurprising.

We observe different cross-country patterns, however, in the measures of FLOSS-friendly IT procurement policy, type of software release, and the role of FLOSS in hiring. Malaysia and South Africa have the IT policies most conducive to FLOSS use, while Brazil, Croatia, and India have the least favorable policies. It is notable that Brazilian institutions have such high rates of using and developing FLOSS, despite the relative absence of policies supporting FLOSS use. Bulgaria and Malaysia have the clearest policies for releasing software, though the somewhat small Bulgarian sample gives little basis for inference, and we cannot reject the hypothesis that the clarity of licensing policy is equal in all countries. Brazilian institutions again put the greatest emphasis on FLOSS experience in hiring potential employees, while Argentine and South African institutions on average put the least emphasis on FLOSS experience when hiring potential employees.

Further enquiry into both the differences reported between use and development and the latter's association with clear university policy statements are in order before venturing anything resembling policy recommendations, if only because it is conceivable that in some part of these observations are artifacts of the survey instrument itself. University administrators – whose responses contribute to these averages – may have a view that is systematically different from that of IT managers as to what is entailed by "university development of software." When asked whether FLOSS is being "developed anywhere in your university" the administrative response may be colored by the idea of code development being undertaken as a condition of employment, or a formal requirement for students in some programs, and then being released under *corporate* copyright – whether of the conventional or the open source kind. IT managers, on the other hand, may be more likely to have in view the activities of students and faculty writing open source code as part of instructional course requirements or for their own research use, and not bothering to formally release it. The more formalized and clearly stated is university policy on such matters, the more likely it might be for administrators to recognize and acknowledge the existence of development activity within their institution. Lack that, however, administrators may systematically under-represent the extent of actual open source development that members of the university community are undertaking. Therefore, before making too much of either the statistical differences in between the prevalence of use of FLOSS and that of its development at these universities, or the association between the presence of explicit IT policies and the frequency of reports of university development work, some closer attention should be given to these issues of perception and interpretation on the part of the respondents. Nevertheless, some simple checks to see whether comparisons of the administrative and IT manager responses lend support to the form of bias suggested above do not find administrative personnel understating systematically reporting lower rates of FLOSS development than the IT manager counterparts.

4.3 Correlates of FLOSS Indices

To combine some findings of the preceding pages, we estimate regressions which measure the association of each FLOSS index with a respondent's sex, age, and identity as IT staff or administrator; with a university's reported enrollment and staff size, with the other indices; and with country indicators (Table 10). One can interpret the coefficients in these regressions as the association of a given explanatory variable with a constructed FLOSS index, controlling for other explanatory variables listed in the regression. For a few reasons, these regressions do not measure causal effects, but rather they provide a more careful way of measuring association than the correlations of Tables 8 and 9 have done.[25] Some of the regressions in Table 10 include university fixed effects (see Appendix Section 2 for details). Since the data include several respondents from most universities, these fixed effects regressions effectively control for all characteristics like location, management quality, and others which do not vary within a university.[26] Including university fixed effects to some extent reduced the bias due to omitted variables (including "unobservables"), and hence yields estimates of the coefficients that are likely to be closer to true values than are the least squares estimates. Unfortunately, including an indicator for each university both prevents inclusion of country indicators in regressions and substantially decreases precision by eliminating degrees of freedom. Given the relative merits of each type of estimator, both are presented in Table 10.

We find some association between the characteristics of the individual respondents and the role of FLOSS in the respondent's institution. Controlling for other variables, females are slightly less likely to report frequent use and development of FLOSS in their universities, though the association is not statistically significant. A similar role appears for IT managers, among whom the probability of reporting that their university uses or develops FLOSS is slightly lower probability than it is among administrative respondents, but here too the difference is statistically insignificant. Older respondents are less likely to report that their universities develop FLOSS, but more likely to say that their institution has a clear software licensing policy – which emphasizes that relationships which hold in the international and intra-national cross-section comparisons among institutions do not necessarily manifest themselves at the micro-level of individual respondents. The least squares and fixed effects estimates provide relatively similar sized coefficients on the age variable, though the fixed effects estimates have little precision, suggesting that the effect is not due to older people happening to work in universities which avoid development of FLOSS. Since contribution to FLOSS programs is widely found to be most common among young people (David and Shapiro 2007), these results hardly can be a surprise. For the same reason, the respondent's age itself, has no association with the likelihood of FLOSS being used in their teaching institution.

Some university characteristics, however, do exhibit a statistically significantly association with FLOSS use. Undergraduate enrollment has no large association with the role of FLOSS in a university, while universities with more graduate students or teaching staff are slightly more likely to develop FLOSS or to have supportive IT policies for using FLOSS. The FLOSS indices included as explanatory variables reaffirm the findings of Table 8, though in Table 10 the magnitudes of the associations are generally smaller since Table 10 controls for other university and individual characteristics while Table 5 does not. Table 10 also reaffirms some of the same cross-country associations that appear in Table 9, but these regressions explain only 6 to 29 percent of variation in FLOSS use and licensing across universities. Clearly this picture, while informative remains incomplete as many other factors influence FLOSS use, adoption, and license policy.

[25] Two realities temper causal inference from these regressions. First, numerous unobservable individual-, university-, and country-specific characteristics in each regression may correlate with both the FLOSS indices and other explanatory variables, thereby biasing results. Second, it is likely that the explanatory variables cause some change in the response variable, as the regression suggests, but also that the response variable causes change in the explanatory variables. Such simultaneity is difficulty to address, and forces interpretation of regressions as association rather than causation.

[26] Since reports of undergraduate enrollment, graduate enrollment, and staff size vary between respondents from a single university, for present purposes these variables do vary within a university.

5. Conclusions and relevance for policy

Free/Libre/Open Source Software (FLOSS) attracts interest from many researchers and policymakers because they seek both to explain the features of its unusual mechanism of production and to examine the potential insights of FLOSS for other individual and collective activities. Understanding how and why universities use and develop FLOSS is particularly important because universities prepare software developers for beginning work, train researchers in technical fields, provide high-speed internet connections for people in developing countries, and support much groundbreaking research in the mathematical and computational sciences. Furthermore, university employees may face different incentives than government and private sector employees do. Understanding how universities use and develop FLOSS may help understand the role that FLOSS plays in other institutions, and the policies that could affect FLOSS use in any institution.

This report, using a new survey of university staff members in developing countries, reaches a variety of conclusions which have some relevance for higher education and scientific policy in both Europe and developing countries. A simple but startling statistic is that although almost all universities use FLOSS, only half actually develop FLOSS. The peer nature of FLOSS production implies that FLOSS may provide free technical training, signaling value for obtaining employment, and flexibility to tailor software products from Europe and the U.S. to local languages and circumstances. Students and staff at many of the universities which do not develop FLOSS might investigate reasons for the absence of FLOSS development and mechanisms to address it. The data show strong association between having a clear policy for licensing out software developed in-house and the participation of students and staff members in FLOSS development activities. This relation seems sensible, since students and staff may be more likely to develop FLOSS programs if they know that their university will support release of these programs under a specific license scheme. Perhaps universities without a clear policy on licensing tacitly are supporting the use of Creative Commons or other similar licenses for outgrowths of research, but do not explicitly say so. Formalizing and publicizing such a policy could encourage increased development of FLOSS programs by students and staff, but the existence of a causal connection here remains to be established.

Several of the survey questions yielded responses that indicated a substantial consensus among administrators and IT managers that the proportions of departmental or university IT budgets allocated for software were reasonable when they remained in the range from 0.20 to 0.45, although opinions on this issue were by no means perfectly aligned between the two groups of respondents. The national average of HEI IT budget shares devoted to software lie in the 0.20-0.35 range (Bulgaria being distinctly on the low side at 0.125), so there is reason to think that there would be quite general discomfort with a serious effort to cut expenditures on software purchase and license fees, which a majority of the respondents – and especially those among the administrations – believe will be necessary in the near future. Differences among instructional and research programs from department to department, and among research groups within departments call for higher or higher or lower spending in this IT category and it is important that universities not adopt a one-size-fits all approach to setting budget norms, or imposing funding reductions. Nonetheless, encouragement of a systematic examination throughout these institutions of the potentials of migration to FLOSS to decrease expenditures on software and license fees could produce useful results in many departments.

The data also show that many universities ask technical job applicants about their FLOSS experiences and consider these experiences when hiring new technical staff. This finding supports the idea that one motivation for developer contributions to FLOSS is the goal of signaling ability to future employers. Using other survey data from employers, it would be useful to investigate support for this claim in other industries. If employers indeed consider FLOSS experience in hiring technical staff, it might imply an important role for FLOSS experience as part of a technical university education—universities, for example, could encourage students as an individual project or thesis to develop a module for a FLOSS project.

Finally, it has been found that many universities and departments in China, Croatia, and elsewhere do not offer standard programming courses that are potentially important for developers to learn skills necessary for making technical contributions to FLOSS programs. In part this finding may reflect a lack of awareness on the part of survey respondents from engineering schools or other non-computer-science departments of course offerings that do exist. But it may also reflect an opportunity for accreditation systems to include in their assessment criteria the question of whether a university's computer science programs adequately prepare students for technical work at the level that would be required to fully participate in FLOSS programs.

Bibliography

Benkler, Y. 2006. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. New Haven: Yale University Press.

Boston Consulting Group (BCG). 2003. *Boston Consulting Group/OSDN Hacker Survey*. Boston: Boston Consulting Group.

Dalle, J-M, and PA David. 2003. "The Allocation of Software Development Resources in `Open Source' Production mode." SIEPR Discussion Paper 02-27.

Dalle, J-M, PA David, RA Ghosh and W E Steinmueller. 2005. "Advancing economic research on the free and open source software mode of production," in *How Open Will the Future Be? Social and Cultural Scenarios based on Open Standards and Open-Source Software*, eds. M. Wynants and J. Cornelis, Brussels: VUB Press, 2005. [Preprint available at: http://siepr.stanford.edu/papers/pdf/04-03.html.]

David, PA, M den Besten and R Schroeder, "How 'Open' is e-Science?," in *Proceedings of the IEEE 2nd International Conference on eScience and Grid Computing*, Amsterdam, 4- 6th December 2006. (Oxford Internet Institute OeSS Project Working Paper, 10 July 2006, [available at: http://www.oii.ox.ac.uk/microsites/oess/papers.cfm]).

David, PA, and F Rullani. 2006. "Open source software development dynamics: project joining and new project generation on SourceForge." (October) SIEPR Discussion Paper, Stanford University, Sant'Anna Institute of Advanced Studies.

David, PA, and JS Shapiro. 2007. "Are 'open source' software developers at-large really representative of the participants community-based FLOSS projects?" Working Paper, OII Oxford and LSE (April).

David, PA, A Waterman, and S Arora. 2003. *FLOSS-US: The Free/Libre/Open Source Software Survey for 2003*. (September) SIEPR-Stanford Economics of Open Source Software Project [available at: http://www.stanford.edu/group/floss-us/report/FLOSS-US-Report.pdf].

Ghosh, RA, R Glott, B Krieger, and G Robles. 2002. "Free/Libre and Open Source Software: Survey and Study." International Institute for Infonomics, University of Maastricht, The Netherlands. Available at http://www.infonomics.nl/FLOSS/report/

Ghosh, RA, R Glott, B Krieger, and G Robles. 2005. "Free software developers: who, how, and why." In Soete, L, and B ter Weel. Eds., *The Economics of the Digital Society*. Edward Elgar.

Hars, A, and S Ou. 2002. "Working for free? Motivations for participating in Open-Source projects." *International Journal of Electric Commerce* 6: 25-39.

Hertel, G, S Nieder, and S Herrmann. 2003. "Motivation of Software Developers in Open-Source Projects: An Internet-based Survey of Contributors to the Linux Kernel." *Research Policy* 32: 1159-1177.

Krueger, A.B., and P. Zhu. 2004. "Another look at the New York City School Voucher Experiment." *American Behavioral Scientist* 47: 658-698.

Lerner, J, and J Tirole. 2001. "The Open Source Movement: Key Research Questions." *European Economic Review* 45(4): 819-826.

Lerner, J. and J. Tirole, 2002. "Some Simple Economics of Open Source," *Journal of Industrial Economics*, 50(2), 2002: pp. 197-234.

Lerner, J., and J. Tirole. 2005. "The Economics of Technology Sharing: Open Source and Beyond." *The Journal of Economic Perspectives* 19: 9-120.

Maurer, S.P., and S. Scotchmer. 2006. "Open Source Software: the New Intellectual Property Paradigm." NBER Working Paper 12148.

Mitsubishi Research Institute. 2004. "Free/Libre/Open Source Software Asian Developers Online Survey (FLOSS-ASIA)." Working Paper, Tokyo, Japan.

Open Source Software Advisory Service (OSS Watch). 2006. "OSS Watch Survey 2006: Report."

Rao, JNK, and AJ Scott. 1984. “On Chi-squared Tests for Multiway Contingency Tables with Proportions Estimated From Survey Data.” *Annals of Statistics* 12: 46-60.

World Bank. 2006. *World Development Indicators*. Washington, DC. The World Bank.

Appendix

Section A1. Scoring coefficients from principal component analysis.

Table A1. Scoring coefficients from principal component analysis for FLOSS use

Variable	*Scoring coefficient*
Use open source in institution	0.32
Use open source in institution (no response)	-0.24
Administration staff use OS: not at all	-0.21
Administration staff use OS: a little	0.26
Administration staff use OS: much	0.08
Administration staff use OS: a great deal	0.08
Teaching staff use OS: not at all	-0.30
Teaching staff use OS: a little	0.36
Teaching staff use OS: much	0.02
Teaching staff use OS: a great deal	0.02
Computer science students staff use OS: not at all	-0.20
Computer science students use OS: a little	0.21
Computer science students use OS: much	0.19
Computer science students use OS: a great deal	0.05
Other science students use OS: not at all	-0.29
Other science students use OS: a little	0.35
Other science students use OS: much	-0.07
Other science students use OS: a great deal	0.05
Non-science students use OS: not at all	-0.27
Non-science students use OS: a little	0.27
Non-science students use OS: much	-0.08
Non-science students use OS: a great deal	0.09
N	446

Table A2. Scoring coefficients from principal component analysis for FLOSS development

Variable	*Scoring coefficient*
Develop FLOSS anywhere in institution	0.45
Develop FLOSS anywhere in institution (no response)	-0.15
Administration staff (including IT staff) develop FLOSS: not at all	0.14
Administration staff (including IT staff) develop FLOSS: a little	0.24
Administration staff (including IT staff) develop FLOSS: much	0.14
Administration staff (including IT staff) develop FLOSS: a great deal	0.22
Teaching staff develop FLOSS: not at all	0.06
Teaching staff develop FLOSS: a little	0.28
Teaching staff develop FLOSS: much	0.15
Teaching staff develop FLOSS: a great deal	0.22
Students in computer science develop FLOSS: not at all	0.11
Students in computer science develop FLOSS: a little	0.28
Students in computer science develop FLOSS: much	0.20
Students in computer science develop FLOSS: a great deal	0.13
Students in other science develop FLOSS: not at all	0.04
Students in other science develop FLOSS: a little	0.25
Students in other science develop FLOSS: much	0.06
Students in other science develop FLOSS: a great deal	0.30
Students in non-science subjects develop FLOSS: not at all	0.04
Students in non-science subjects develop FLOSS: a little	0.15
Students in non-science subjects develop FLOSS: much	0.05
Students in non-science subjects develop FLOSS: a great deal	0.37
N	446

Table A3. Scoring coefficients from principal component analysis for institutional IT policy framework

Variable	*Scoring coefficient*
Institution has stated IT policy or strategy	0.46
Stated IT policy or strategy includes FLOSS as procurement option	0.46
Policy requires FLOSS rather than proprietary alternative	0.33
Strategy does not include FLOSS because: doesn't mention software	0.02
Strategy does not include FLOSS because: doesn't specify software type	0.03
Strategy does not include FLOSS because: focuses only on proprietary	0.01
Strategy does not include FLOSS because: explicitly excludes FLOSS	0.03
Strategy does not include FLOSS because: other	0.04
Institution has stated IT policy or strategy (no response)	-0.20
Stated IT policy or strategy includes FLOSS as procurement option (no response)	-0.47
Policy requires FLOSS rather than proprietary alternative (no response)	-0.45
Strategy does not include FLOSS because: (no response)	-0.06
N	446

Table A4. Scoring coefficients from principal component analysis for role of FLOSS in hiring

Variable	*Scoring coefficient*
Ask job applicants about FLOSS experience?	-0.32
Positions in company require FLOSS experience?	-0.49
Do these positions usually involve leadership?	-0.22
Hiring choice: prefer FLOSS experience?	-0.31
Ask job applicants about FLOSS experience (no response)	0.30
Positions in company require FLOSS experience (no response)	0.32
Do these positions usually involve leadership (no response)	0.49
Hiring choice: prefer FLOSS experience (no response)	0.28
N	245

Section A2. Averages across universities or individuals

For many survey items, the question of interest requires comparing across universities rather than across individual respondents. But the number of respondents per university varies depending on idiosyncratic factors. Averaging across individuals would give excess weight to universities where several individuals happened to respond. Averaging across individuals within a university, then averaging across universities, would discard the additional information that the data contain about universities where several individuals have responded. (Put another way, in a university where ten individuals respond, the data contain more information about the within-university mean than the data do in a university where only one individual responds, and inference should reflect the additional information.) As a further complication, in some questions we compare only IT managers across universities, in other questions we compare only administrative respondents across universities, and in others we compare all respondents across universities.

To properly make these comparisons, we construct and use weights as follows. For survey items which require comparison across individual respondents (for example, Table 1 or Figure 1), we use no weights and state in the table or figure notes that in the data each person receives equal weight. For survey items which require comparison across universities, we weight each observation by the inverse of the number of respondents from the observation's university. For example, if X_i is an indicator for whether person *i*'s university has a software acquisition policy, due to errors or other reasons, X_i may vary within universities. Then an estimator of the portion of *universities* which have a software acquisition policy is

$$\frac{1}{N}\sum_{i=1}^{N} X_i w_i \qquad (1)$$

where N is the total number of persons who responded to the survey item and w_i is the person-specific weight, i.e., the inverse of the number of respondents from person *i*'s university. In some estimates we further cluster standard errors within universities to reflect the fact that observations from the same university are not independently distributed.

To estimate a mean across universities for IT managers only, we estimate equation (1) but define N as the total number of IT persons who responded to the survey item, and define w_i as the inverse of the total number of IT managers from person *i*'s university.

Section A3. Statistical tests and regressions

We use three tests to investigate whether the distributions of responses to a given question item differ across countries, and whether they differ between administrative and IT managers . For continuous variables, we report the F-statistic from analysis of variance (ANOVA). This is numerically equivalent to estimating a regression while including an indicator for each country, then calculating from the regression the F-test that the countries all have coefficients of zero. The null hypothesis of the F-test is that a variable of interest has equal values across all countries. Rejection of this hypothesis (generally, p-value < 0.05 or <0.10) implies that the variable of interest has statistically significant differences across countries. The ANOVA test statistic is easily adapted to the use of weights.

The ANOVA F-test assumes that the involved variable is normally distributed. For unweighted estimates (i.e., across individuals rather than across universities) and discrete variables, we report a Pearson χ^2 test of independence across rows and columns. Let i index the I different rows, j index the J different columns, and n_{ij} represent a cell count. Define row and column sums as follows:

$$n_{i\cdot} = \sum_{j=1}^{J} n_{ij}$$

$$n_{\cdot j} = \sum_{i=1}^{I} n_{ij}$$

Further define $m_{ij} = \frac{n_{i\cdot} n_{\cdot j}}{n}$, where $n = \sum_i \sum_j n_{ij}$. Then the Pearson test statistic is as follows:

$$T=\sum_i \sum_j \frac{\left(n_{ij}-m_{ij}\right)^2}{m_{ij}}$$

Under the null hypothesis that the rows and columns are independent, $T \sim \chi^2_{(I-1)(J-1)}$. Rejection of this hypothesis (generally, p-value < 0.05) implies that the variable of interest has statistically significant differences across countries.

Pearson's test cannot be directly estimated when each observation has a weight, as occurs in estimates across universities rather than across individuals. But Rao and Scott (1984) generalize the Pearson test to allow weights. The test statistic resulting from their generalization has an F distribution under the null hypothesis of independence, and we report this F statistic in tables which require a test of independence (typically, whether a binary outcome has the same proportion in different countries) but which also require weights.

Section 4 uses least squares and fixed effects regressions to measure the association of several variables with FLOSS use, development, and three other FLOSS-relevant indices. We derive estimating equations as follows. For each person *p* in institution *i*, we observe the outcome Y_{ph}, a vector of individual-specific explanatory variables Z_{ph}, and a vector of explanatory variables X_h which is specific to a higher education institution. We first estimate the following equation using least squares:

(A1) $$Y_{ph}=\gamma_1 X_h+\gamma_2 Z_{ph}+\varepsilon_{ph}$$

The least squares estimates assume that the errors ε_{ph} are independently and identically distributed, with $\varepsilon_{ph} \sim N(0,\sigma^2)$.

In reality, however, the errors are not independent, since respondents from the same university will have above-average similarity. Clustering standard errors about universities would resolve this issue of inference.

But a second problem of consistency remains: the explanatory variables may correlate with university-specific unobserved variables. To see why, write the composite error term as the sum of two errors—a university-specific error, and an idiosyncratic error which varies across universities and individuals:

$$\varepsilon_{ph}=\mu_h+\eta_{ph}$$

If $cov(\mu_h,X_h)\neq 0$ or if $cov(\mu_h,Z_{ph})\neq 0$, which is likely if any unobserved university-specific factors correlate with the explanatory variables in the regressions, then equation (A1) will not provide consistent estimators of the parameter vectors γ_1 and γ_2. Fixed effects estimates include a dummy variable for each university μ_h.[27] Since these indicators capture any university-specific effects, the fixed effects equations cannot include the covariates X_h, which in our data means that the fixed effects estimates do not include country indicators. Since this loss is potentially traded off against potentially decreased bias, we present both least squares and fixed effects estimates.

Some estimates test whether an administrative and IT respondent from the same university will have different responses to a survey item. This test differs from ANOVA, or the Pearson tests because it eliminates any features common to universities, and retains only administrative-IT differences that persist within universities. To estimate this test, for each outcome, we estimate the following equation using least squares:

$$Y_{iu}=\theta_0+\theta_1 adm_{iu}+\sum_{u=1}^{U}\phi_u+\varepsilon_{iu}$$

For each individual *i* in university *u*, we regress the outcome *Y* on a constant, an indicator *adm* which takes the value one if the observation is an administrative respondent and zero otherwise, and on set of indicators φ_u, one for each university. Reported standard errors and p-values are robust to

[27] For simplicity, in estimation, the within-university mean of equation (A1) is subtracted from (A1), and the resulting equation estimated by least squares. This estimator is numerically equivalent to including an indicator for each university then estimating the equation by least squares.

heteroskedasticity. In tables and figures we report a t-test of the null hypothesis $H_0: \theta_1=0$. Failure to reject this hypothesis implies that within a university, the responses of admin and IT individuals are statistically indistinguishable.

Although some regressions have binary outcomes, we use the linear probability model rather than a probit or logit, for simplicity of estimation while including fixed effects. Similarly, since the likelihood function for an ordered probit is difficult to maximize while including fixed effects for over 300 universities, when the outcome is a Lickert scale or another ordered outcome, for this comparison of admin and IT managers we redefine the outcome to have value one for the most positive responses and value one for the most negative responses. In these situations each table explains the constructed definition of the response variable.

Section A4. Selection Bias.

Findings from this survey are subject to the standard caveats of possible selection bias. The survey obtained an average response rate of 50 percent, and lack of data on nonrespondent universities makes it difficult to know whether the data over-sample those individuals and universities with particularly strong positive or negative feelings about FLOSS. The universities in these data collectively teach about 3.7 million students, so the data certainly represent a large population of universities, though perhaps not a national population for any particular country. Nonetheless, one should note that characteristics common to the universities in this survey may not be common to all universities.

Probabilistic surveys or surveys with substantial nonresponse often estimate a weight for each respondent, where the weight represents the number of respondents in the true overall population that the particular respondent represents. Although the number of respondents per university varies, and although half of the universities contacted for the university did not respond, we do not construct survey weights for the data for a few reasons. First, the sample is not stratified, clustered, or probabilistic, so the survey design does not necessitate use of weights. Second, the original list of universities to survey was not based on a university census, so weights would only seek to ensure that statistics in this report represented the population of contacted universities, and not the population of all universities. Third, variation in the number of individuals who respond from each university is due to different universities having different staff sizes, different individual nonresponse, and a different number of individuals invited to participate in the survey. The absence of good data on these factors implies that statistics obtained using constructed sample weights might have more bias relative to true population values than the statistics presented in this report (which do not use weights) have. Fourth, since the data offer nearly no information on nonrespondent universities, we would have little basis for constructing sample weights which reflect a university's probability of responding to the survey. The preceding points suggest that estimating survey weights and using them to calculate statistics in this report might give more or less accurate results than the results obtained without such survey weights. Given that use of survey weights somewhat complicates explanations of results, and that FLOSSWorld targets policymakers, academics, the business community, and the FLOSS community, we judge that the report better serves its purposes without constructing and using survey weights.

Section A5. Principal component analysis

We use principal component analysis to construct four indices which measure broad characteristics of universities with regard to FLOSS. No single survey item fully captures a university's tendency to use FLOSS, develop FLOSS, maintain a FLOSS-friendly IT policy, or consider FLOSS in hiring, so we combine information from several survey items to derive indices measuring these underlying latent characteristics. Factor analysis methods can generally seek to reduce several variables in a dataset to one index. Principal component analysis – only one type of factor analysis – obtains factor scores which maximize the variability of the constructed factor. All factors are constructed to have mean zero and standard deviation one. The scoring coefficients report in Appendix section A1 represent the linear coefficients of each variable in creating the constructed factor. For example, a scoring coefficient of 0.2 on a particular covariate would indicate that this covariate increases the constructed factor by 0.2 standard deviations.

Section A6. Survey design

Countries used varying methodologies to identify and survey universities. A partial list of survey design follows.

Bulgaria: the total number of universities in Bulgaria is 45. The universities which have some reputation and are the preferred choice for study are not more than 12-15, and 15 universities were contacted by email and phone.

Croatia: Local affiliates in CARNet (Croatian Academic and Research Network) have two contacts with each member institution: a system engineer (technical) and a CARNet coordinator (administrative). Since Croatia has relatively few universities, and since departments within universities have financial independence for most purposes, including IT investments, affiliates contacted a variety of departments and not merely universities. The System engineer who is a CARNet technical contact was chosen as a “technical” respondent, while a vice dean for finance, investment, or business or strategic projects was chosen as an “administrative” respondent.

Malaysia: MIMOS collaborated with the Ministry of Higher Education (MOHE) for the survey. MOHE sent a circular / directive directly to all the Vice Chancellors from a selected university list via fax and email attaching both hard and soft copies of the survey forms. The directive requested the VCs to redirect the circular to all the Branch Campuses and relevant Faculties / Schools (e.g. Engineering, Computer Science, Admin, Library, Management etc). All the responses were directed back to MIMOS and were keyed in directly into the FLOSSWorld online template.

Table 1. Characteristics of respondents

Country	*Sample size* Number of respondent individuals	IT (% of total)	Admin (% of total	*Female (%)*	*Age (mean)*	*Wrote in name (%)*
Argentina	43	27.91	72.09	9.30	43.19 (10.01)	55.81
Bulgaria	7	42.86	57.14	42.86	35.43 (9.52)	85.71
Brazil	72	69.44	30.56	17.65	42.67 (9.06)	38.89
China	54	48.15	51.85	22.22	36.67 (9.75)	0.00
Croatia	83	68.67	31.33	11.11	38.03 (10.38)	100.00
India	47	72.34	27.66	23.40	40.17 (9.59)	44.68
Malaysia	128	42.97	57.03	23.02	37.91 (8.68)	0.00
South Africa	12	58.33	41.67	0.00	46.58 (8.63)	83.33
Total	446	54.71	45.29	18.26	39.52 (9.70)	38.57
N				438	431	446.00
Pearson $\chi^2(7)$		39.72		13.93		269.33
χ^2 p-value		0.00		0.05		0.00
F stat					4.67	
p-value					0.00	

Notes: see Appendix section 2 for explanation of χ^2 statistic and F statistics. Means are across individual respondents rather than across universities.

Table 2. Characteristics of respondent universities

	Institutional response rate	*Number of distinct universities*	*IT respondent only (% of universities)*	*Admin respondent only(% of universities)*	*IT and admin respondent(s) (% of universities)*	*Technical university (% of universities)*	*Undergraduate enrollment*	*Graduate enrollment*	*Teaching & research personnel*	*Administrative & support staff*
Argentina	0.64	36	0.25	0.67	0.08	0.61	12,379	4,758	1,188	290
							(22,703)	(9,947)	(2,240)	(668)
Bulgaria	0.42	5	0.40	0.40	0.20	0.20	4,415	875	484	395
							(1,674)	(177)	(233)	(359)
Brazil	0.31	57	0.63	0.26	0.11	0.04	11,248	2,086	1,144	1,122
							(8,721)	(2,559)	(1,278)	(1,335)
China	NA	50	0.96	0.04	0.00	NA	22,348	31,037	2,871	865
							(17,034)	(58,699)	(2,401)	(925)
Croatia	0.74	71	0.65	0.23	0.13	0.21	1,052	770	102	52
							(1,381)	(1,107)	(108)	(176)
India	0.24	34	0.62	0.26	0.12	0.65	3,424	1,431	574	515
							(10,291)	(2,714)	(2,199)	(1,266)
Malaysia	0.77	48	0.10	0.25	0.65	0.33	5,939	1,551	815	680
							(6,554)	(2,632)	(2,403)	(2,159)
South Africa	0.60	9	0.33	0.33	0.33	0.11	14,607	3,251	875	609
							(10,210)	(3,086)	(834)	(452)
Total	0.48	310	0.46	0.35	0.19	0.26	9,308	4,102	952	584
							(13,861)	(17,941)	(1,927)	(1,271)
N						446	301	252	298	310
Pearson $\chi^2(7)$						95.54				
χ^2 p-value						0.00				
F-stat			12.40	8.36	24.14		12.18	7.60	7.16	4.29
p-value			0.00	0.00	0.00		0.00	0.00	0.00	0.00

Notes: standard deviations appear in parentheses. Total is mean across universities rather than across countries. N presents number of respondents to each question. Institutional response rate is total number of universities which had at least one person respond, divided by total number of institutions contacted. Response rates are inexact due to various spellings, abbreviations, and informal names for Universities. The response rate is not calculated for the Chinese survey due to limited ability to match characters in the respondent and nonrespondent lists. Similarly, Chinese universities in administrative and IT surveys are not matched due to limited ability to match characters.

Table 3. Sample sizes from Malaysian institutions

Institution	*Sample size*	*Undergrad enrollment*	*Graduate enrollment*	*Teaching & research staff*	*Administrative & support staff*
Binary University College	2	2,258	236	46	28
Curtin University	2	-	-	-	-
Help University College	2	-	-	-	-
Institute of Advanced Technology	1	-	65	-	51
International Islamic University Malaysia, Gombak	4	1,078	40	89	40
International Islamic University Malaysia, Kuala Lumpur	3	4,306	1,697	1,702	1,679
International Islamic University Malaysia, Selangor	2	1,100	125	90	30
International University College of Technology Twintech	1	400	-	30	3
Kolej University Islam Malaysia	2	2,500	-	131	10
Kolej University Kejuruteraan & Teknologi Malaysia	1	800	15	-	13
Kolej University Kejuruteraan Utara Malaysia	3	1,520	12	111	14
Kolej University Sains des Teknologi Malaysia	2	1,790	1,859	109	24
Kolej University Teknikal Kebangsaan Malaysia	2	4,776	302	475	-
Kolej University Teknologi Antarabangsa Twintech	1	400	-	30	3
Kolej University Tun Hussein Onn	5	2,700	20	127	49
Malaysia University of Science and Technology	2	-	-	23	23
Multimedia University	8	13,496	1,051	512	354
Open University	1	-	-	-	-
Polytechnic Sultan Hj Ahmad Shah	2	6,392	-	341	99
Polytechnic Sultan Salahuddin Abdul Aziz Shah	1	5,901	-	394	87
Polytechnic Ungku Omar	6	3,400	0	537	145
Sultan Zainal Abidin Islamic College	1	4,500	-	400	-
Swinburne University of Technology, Sarawak	2	850	208	74	59
University College Antarabagsa Sedaya	1	4,839	18	200	100
University College Sedaya International	1	4,839	18	200	100
University Kebangsaan Malaysia	4	18,686	7,315	-	-
University Kuala Lumpur, Bangi	2	-	1,300	170	60
University Kuala Lumpur, Gombak	2	1,500	1,500	150	50
University Kuala Lumpur, Kulim	3	713	0	57	41
University Malaya, Kuala Lumpur	3	-	-	-	-
University Malaya, Petaling Jaya	2	22,989	10,263	1,588	3,120
University Malaysia Sabah	2	1,234	60	90	50
University Malaysia Sarawak, Kota Samarahan	1	5,000	5,200	500	500
University Malaysia Sarawak, Kuching	1	-	-	-	-
University Malaysia Sarawak, Sarawak	2	2,840	110	360	360
University Pendidikan Sultan Idris	3	13,797	707	-	-
University Putra Malaysia	18	1,060	950	459	454
University Sains Malaysia, Georgetown	2	850	250	35	13
University Sains Malaysia, Minden	5	7,639	5,060	12	84
University Sains Malaysia, Penang	4	12,650	3,825	1,845	15
University Sains Malaysia, Pulau Pinang	2	760	224	32	49
University Technology Malaysia, Johor	2	4,000	500	2,000	1,000
University Technology Malaysia, Selangor	1	-	-	14,000	12,000
University Technology Malaysia, Shah Alam	3	11,250	2,750	1,146	520
University Technology Petronas	3	5,200	667	300	227
University Tun Abdul Razak	1	10,000	1,000	222	238
University Tunku Abdul Rahman	2	14,000	93	586	386
University Utara Malaysia	2	18,500	4,000	1,250	1,100
Total	128	6,634	1,664	632	502

Table 4. Mean expenditure on IT, software purchases, license fees, and personnel

	University software budget (Mean in PPP US$)	*Software budget per student (Mean in PPP US$)*	*Software purchases & license fees (as % of IT budget)*	*Personnel cost (as % of IT budget)*	*Software purchases as share of budget is: too high*	*Software purchases as share of budget is: reasonable*	*Need to reduce expenditure on software & license fees in next two years*
Argentina	1,018,272	111.66	23.71	20.20	0.15	0.64	0.59
	(1,248,041)	(189.96)	(29.40)	(27.57)	(.36)	(.48)	(.49)
Bulgaria	291,667	62.45	12.50	0.00	0.38	0.50	0.58
	(125,012)	(6.90)	(4.33)	(.00)	(.48)	(.50)	(.49)
Brazil	1,183,209	95.44	20.92	10.44	0.12	0.54	0.64
	(1,485,861)	(90.34)	(17.62)	(20.43)	(.32)	(.50)	(.48)
China	1,679,379	70.80	26.23	3.63	0.06	0.52	0.23
	(1,791,246)	(87.63)	(28.80)	(11.48)	(.23)	(.50)	(.42)
Croatia	58,771	49.79	16.88	3.92	0.60	0.35	0.09
	(68,439)	(62.92)	(25.18)	(11.29)	(.49)	(.48)	(.29)
India	856,771	424.78	31.18	14.22	0.11	0.72	0.49
	(992,069)	(409.44)	(23.46)	(28.67)	(.32)	(.45)	(.50)
Malaysia	1,591,794	311.99	33.24	10.26	0.28	0.57	0.70
	(1,801,621)	(477.09)	(20.60)	(15.32)	(.45)	(.50)	(.46)
South Africa	2,273,148	149.76	33.73	15.25	0.25	0.50	0.39
	(1,850,340)	(65.85)	(13.95)	(20.71)	(.43)	(.50)	(.49)
Total	1,124,106	167.86	24.99	9.21	0.25	0.53	0.45
	(1,527,296)	(293.40)	(23.99)	(19.55)	(.43)	(.50)	(.50)
N	182	114	199	214	340	340	373
F-stat	3.63	2.89	1.93	2.43	8.07	2.02	9.32
p-value	0.02	0.01	0.07	0.02	0.00	0.05	0.00

Notes: For first four columns, table entries are mean across universities within equal countries. standard deviations appear in parentheses. N presents number of respondents to each question. For continuous variables, F-stat is ANOVA-based test of null hypothesis that the variable in a given column has equal mean across countries. For binary variables, F-stat is Rao and Scott's (1984) implementation of a Pearson χ^2 test, which is necessary since the number of respondents varies by university. See statistical appendix for details. Mean year 2004 PPP exchange rate (most recent available) drawn from World Bank (2006). Software budget per student includes graduate and undergraduates combined.

Table 5. Presence of IT strategy, mention of FLOSS

		If yes in (1),	*If yes in (2),*	*If no in (2), why not?*			
	Institution has stated IT policy or strategy?	*policy mentions FLOSS as option?*	*policy requires FLOSS if available?*	*Doesn't mention software*	*Doesn't specify type of software*	*Focuses on proprietary software*	*Explicitly excludes FLOSS as an option*
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*
Panel A: Admin respondents							
Argentina	0.70	0.86	0.68	0.00	1.00	0.00	0.00
Bulgaria	0.50	1.00	0.00	n.a.	n.a.	n.a.	n.a.
Brazil	0.26	1.00	0.64	n.a.	n.a.	n.a.	n.a.
China	0.42	0.88	0.14	1.00	0.00	0.00	0.00
Croatia	0.35	0.85	0.20	n.a.	n.a.	n.a.	n.a.
India	0.45	0.75	0.33	0.00	1.00	0.00	0.00
Malaysia	0.91	0.62	0.43	0.30	0.39	0.03	0.08
South Africa	1.00	0.75	0.00	0.00	0.00	1.00	0.00
Total	0.60	0.76	0.42	0.27	0.45	0.08	0.06
N	168.00	101.00	69.00	22.00	22.00	22.00	22.00
F stat: countries have same mean	6.20	1.12	1.64	1.10	1.44	3.71	0.11
p-value	0.00	0.35	0.13	0.36	0.23	0.01	0.98
N	168	101	69	22	22	22	22
Panel B: IT managers							
Argentina	0.46	1.00	0.27	n.a.	n.a.	n.a.	n.a.
Bulgaria	1.00	1.00	1.00	n.a.	n.a.	n.a.	n.a.
Brazil	0.32	0.92	0.68	0.00	0.00	0.00	0.00
China	0.81	0.57	0.50	0.24	0.38	0.38	0.00
Croatia	0.41	0.77	0.31	0.20	0.40	0.20	0.20
India	0.30	0.83	0.78	n.a.	n.a.	n.a.	n.a.
Malaysia	0.91	0.49	0.39	0.11	0.54	0.11	0.00
South Africa	0.83	0.60	0.00	0.00	0.50	0.50	0.00
Total	0.52	0.70	0.45	0.14	0.46	0.21	0.04
N	222.00	109.00	73.00	30.00	30.00	30.00	30.00
F stat: countries have same mean	6.59	1.94	1.53	0.29	0.35	0.77	1.07
p-value	0.00	0.06	0.15	0.88	0.84	0.54	0.37
F stat: same IT-admin mean	2.26	0.84	0.38	1.53	0.05	3.30	0.03
p-value	0.13	0.36	0.54	0.22	0.83	0.08	0.87

Notes: Table entries are portion of respondents from each country that answered "yes" to the indicated question. N presents number of respondents from to each question. For binary variables, F-stat is Rao and Scott's (1984) implementation of a Pearson c2 test, which is necessary since the number of respondents varies by university. n.a. denotes no responses from a particular country to a particular question

Table 6. Believed and actual use of FLOSS

		Report use of a specific FLOSS program	
		No	*Yes*
State that institution uses FLOSS	*No*	0.01	0.01
	Yes	0.09	0.89
N	238		
F stat	10.17		
p-value	0.00		

Notes: Table entries are relative cell frequencies. Includes only IT managers , since admin survey did not include a question on whether desktop computers or servers use particular FLOSS programs. Figures are averages across universities. Columns are based on answers to questions 23-4, wherein IT managers list programs used on PCs, servers, and elsewhere in the university. Rows are based on answer to question 15, wherein respondents indicate whether the university uses any FLOSS programs.

Table 6b. Comparing reported FLOSS use between IT and Admin respondents

	Admin	*IT*	*Admin*	*IT*
	Response: not at all		Response: a great deal	
Administrators	20%	12%	5%	12%
Teachers	3%	10%	4%	8%
CS students	3%	5%	10%	23%
Other science students	18%	13%	12%	6%
Non-science students	36%	33%	4%	3%

Notes: Compare to Figures 8b and 8c

Table 7. FLOSS dynamics within a university

Department or institute within the university	*Life science institute*	*Social science institute*	*Engineering department*	*Different engineering department*	*Life sciences department*	*Graduate studies office*	*Life science institute*	*Life science department*
Institution has stated IT policy or strategy	1	1	1	1	1	1	1	1
Stated policy or strategy lists OS as procurement option	0	1	0	0	1	1	n.a.	1
Use OS in your institution	0	1	1	1	1	0	0	1
Develop OS in your institution	0	0	1	1	0	0.5	n.a.	0
Staff/students have developed a licensing policy	n.a.	n.a.	0	1	n.a.	1	n.a.	n.a.
Use FLOSS on servers or desktops	1	n.a.	n.a.	n.a.	1	1	0	1
Positions in university require OS experience	1	n.a.	n.a.	n.a.	1	1	0	0
Are these usually leadership positions	1	n.a.	n.a.	n.a.	1	1	n.a.	n.a.
Choice of hiring candidate: prefer proprietary	1	n.a.	n.a.	n.a.	0	0	0	0
Choice of hiring candidate: prefer FLOSS	0	n.a.	n.a.	n.a.	0	1	0	1
Choice of hiring candidate: wouldn't matter	0	n.a.	n.a.	n.a.	1	0	1	0
N	1	1	1	1	1	2	1	1

Note: n.a. indicates that no individual answered the indicated question. All responses are from a single university. Table entries are portion of individuals in each department (column) who answered "yes" to the question (row).

Table 8. Correlations of FLOSS indices

	Use FLOSS (1)		*Develop FLOSS (2)*		*FLOSS-friendly IT policy (3)*		*Value FLOSS in hiring (5)*
Use FLOSS	1.00						
Develop FLOSS	0.23	***	1.00				
FLOSS-friendly IT policy	0.08	*	0.15	***	1.00		
Value FLOSS in hiring	-0.12	*	0.31	***	-0.13	**	1.00

Notes: standard deviations appear in parentheses. *** represents statistical significance at 99% level. Each university has equal weight in calculations.

Table 9. Differences in FLOSS indices across countries

	Index: use FLOSS (1)	*Index: develop FLOSS (2)*	*Index: FLOSS-friendly IT policy (3)*	*Index: Value FLOSS in hiring (5)*
Argentina	0.61	0.28	0.62	0.02
Bulgaria	0.73	1.44	-0.10	2.24
Brazil	0.89	0.65	-0.33	-0.57
China	-0.71	-0.57	-0.19	-0.02
Croatia	-0.58	-0.71	-0.31	0.07
India	-0.84	-0.16	-0.57	0.31
Malaysia	0.48	0.38	0.80	0.28
South Africa	0.50	1.22	0.96	0.25
N	446	446	446	244
ANOVA F	11.23	5.61	4.38	2.18
Prob > F	0.00	0.00	0.00	0.04
R^2	0.15	0.08	0.07	0.06

Notes: standard deviations appear in parentheses. N presents number of respondents categorized on each index. Each index has mean zero across all observations.

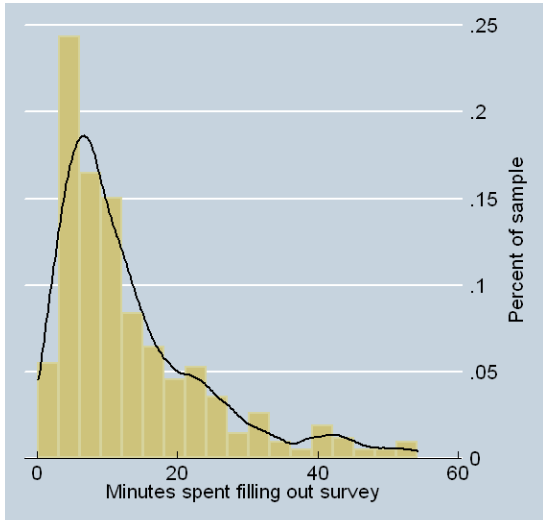
Table 10. Correlates of FLOSS indices, least squares estimates

Dependent variable	*Use FLOSS*		*Develop FLOSS*		*FLOSS-friendly IT policy*		*Value FLOSS in hiring*	
	(1)	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(9)*	*(10)*
IT respondent	-0.171	-0.013	0.026	0.09	0.117	-0.175	-	-
	[0.169]	[0.312]	[0.217]	[0.457]	[0.214]	[0.418]	-	-
Female	-0.377	0.054	-0.267	-0.716	-0.494	-0.432	0.205	-0.286
	[0.219]*	[0.560]	[0.256]	[0.748]	[0.260]*	[0.656]	[0.299]	[1.207]
Age	-0.018	0	-0.021	-0.007	-0.016	-0.026	0.025	-0.056
	[0.009]**	[0.023]	[0.012]*	[0.030]	[0.011]	[0.027]	[0.014]*	[0.088]
Undergraduate enrollment	0	0	0	0	0	0	0	0
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]**	[0.000]	[0.000]	[0.000]
Graduate enrollment	0	0	0	0	0	0	0	0
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Teaching & research staff	0	0	0	0	0	0	0	0
	[0.000]*	[0.000]	[0.000]*	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Index: FLOSS use			0.156	0.094	-0.033	-0.016	-0.04	-0.083
			[0.068]**	[0.259]	[0.077]	[0.198]	[0.062]	[0.377]
Index: FLOSS development	0.099	0.042			0.085	0.164	-0.191	-0.103
	[0.042]**	[0.116]			[0.055]	[0.128]	[0.053]***	[0.280]
Index: FLOSS-friendly IT policy	-0.021	-0.009	0.088	0.195			-0.082	-0.088
	[0.051]	[0.105]	[0.056]	[0.152]			[0.052]	[0.295]
Country: Argentina	0.003		-0.163		-0.245		-0.217	
	[0.252]		[0.431]		[0.407]		[0.603]	
Country: Bulgaria	0.13		1.186		-0.875		2.288	
	[0.335]		[0.560]**		[0.845]		[0.931]**	
Country: Brazil	0.292		0.113		-1.207		-0.955	
	[0.241]		[0.414]		[0.343]***		[0.447]**	
Country: China	-1.219		-0.882		-1.039		-0.435	
	[0.280]***		[0.382]**		[0.374]***		[0.528]	
Country: Croatia	-0.772		-0.721		-1.009		-0.52	
	[0.278]***		[0.350]**		[0.319]***		[0.474]	

Country: India	-1.204		-0.138		-1.236		-0.369	
	[0.389]***		[0.430]		[0.384]***		[0.532]	
(Reference Country: Malaysia)								
Country: South Africa	-0.131		0.84		-0.098		-0.182	
	[0.341]		[0.662]		[0.500]		[0.640]	
Constant	1.293	0.304	0.898	0.278	1.368	1.466	-0.545	2.257
	[0.380]***	[0.994]	[0.558]	[1.227]	[0.492]***	[1.115]	[0.726]	[3.301]
University fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
N(observations)	426	426	426	426	426	426	236	236
N(universities)	300	300	300	300	300	300	236	236
R^2	0.2	0.07	0.16	0.09	0.11	0.07	0.18	0.14

Notes: parentheses contain heteroskedastic-robust standard deviations clustered about 281 universities for estimates without fixed effects. Parentheses for fixed effects estimates contain heteroskedastic-robust standard errors. * significant at 10%; ** significant at 5%; *** significant at 1%

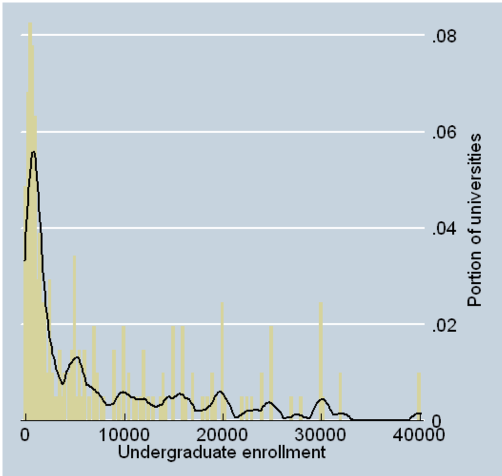
Figure 1. Time spent on survey



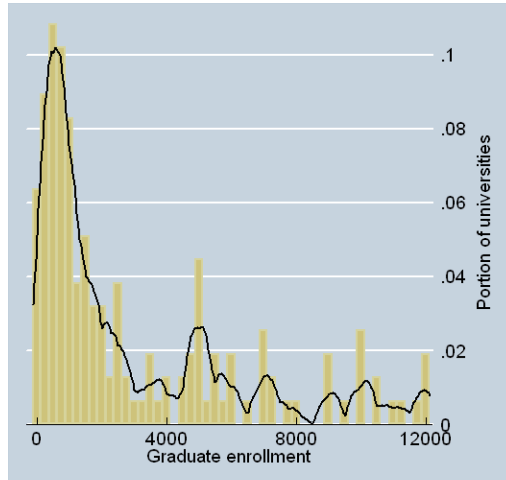
Notes: Line represents density estimated using Epanechnikov kernel evaluated at 300 points with bandwidth of 3. Time spent on survey has mean=20.68, median=10.60, and standard deviation=37.95. Bin width is 3. Graph shows averages across individuals.

Figure 2. University enrollment

Panel A: Undergraduate student enrollment



Panel B: Graduate student enrollment

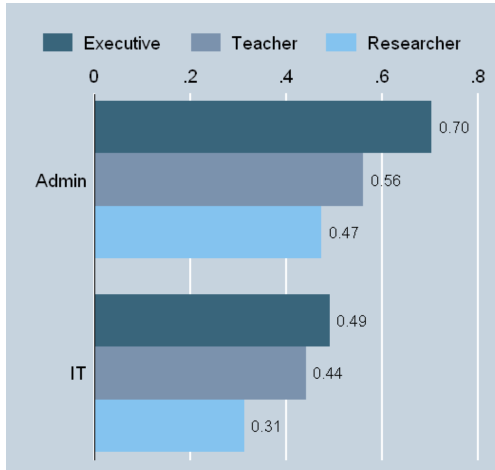


Note: lines represent density, estimated using Epanechnikov kernel evaluated at 300 points with bandwidth of 3.

Panel A: histogram has bin size of 250. Graph excludes the five universities which have reported undergraduate enrollment of over 40,000 students . Graph shows averages across universities. Undergraduate enrollment defined as maximum undergraduate enrollment reported among any respondent from a university.

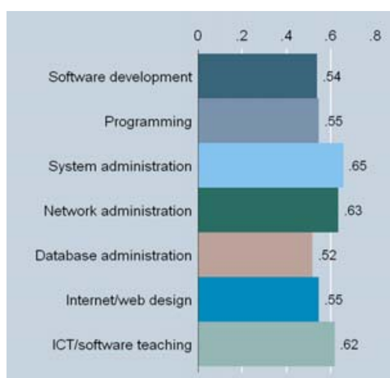
Panel B: histogram has bin size of 250. Graph excludes the four universities with reported graduate enrollment of over 12,000 students. Graph shows averages across universities. Graduate enrollment defined as maximum graduate enrollment reported among any respondent from a university.

Figure 3. Respondent responsibilities



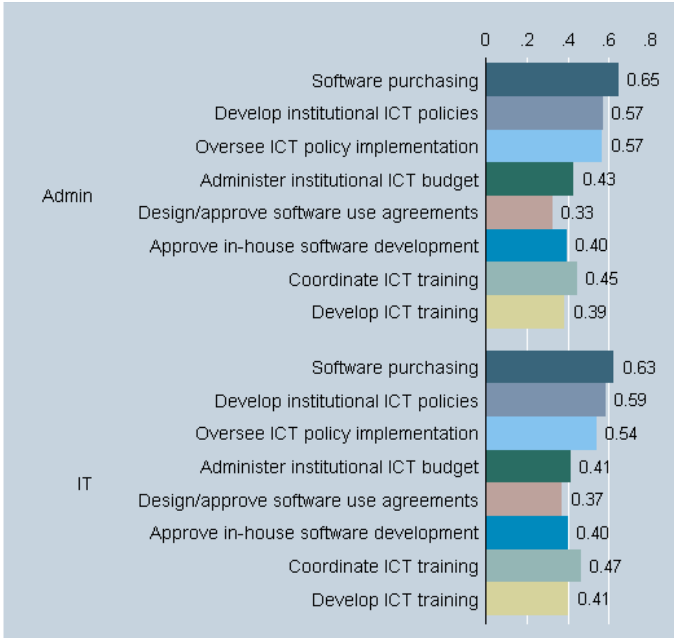
Notes: admin responses based on 196 individuals. IT responses based on 240 individuals. Graphs show averages across individuals. For the three responsibilities, respectively, a fixed effects regression comparing administrative and IT managers from within the same university shows that administrative respondents are 18.3 percentage points more likely to have executive responsibilities (p-value = 0.09), 10.6 percentage points more likely to have teaching responsibilities (p-value = 0.31), and 25.6 percentage points more likely to have researching responsibilities (p-value = 0.01). The statistical appendix explains the details of these estimates.

Figure 4. IT respondent technical skills or responsibilities.



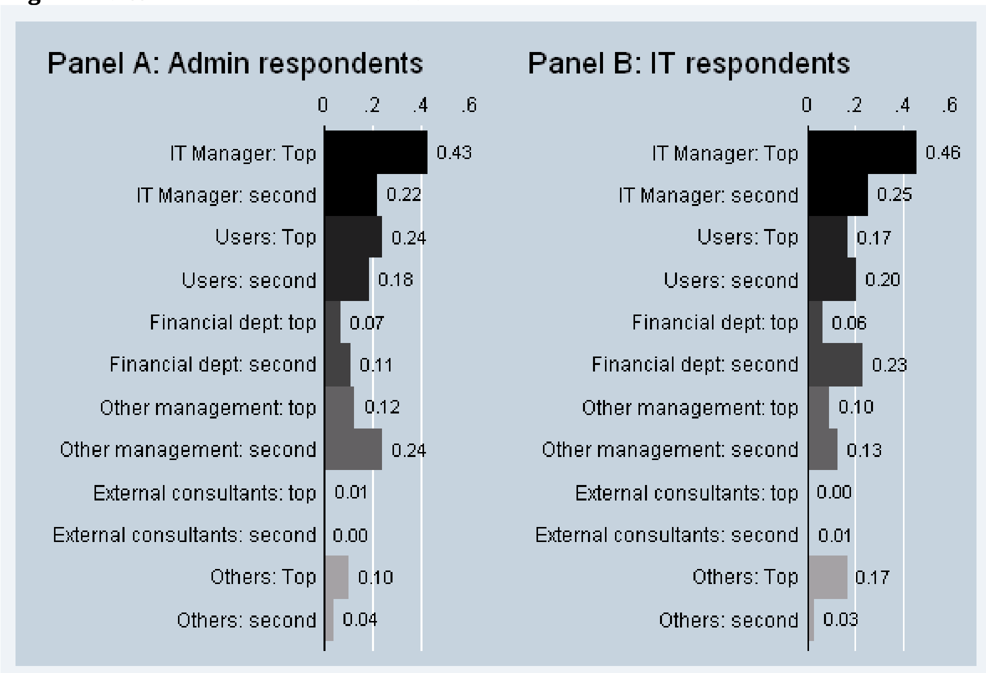
Notes: data based on 240 respondents. Graph shows averages across individuals. Only IT managers answered this survey item.

Figure 5. Respondent Administrative Responsibilities



Note: Admin statistics based on 202 responses. IT statistics based on 244 responses. Data show averages across individuals. For each of the eight responsibilities listed, at any level of statistical significance above 90%, a fixed effects regression fails to reject the null hypothesis that admin and IT managers from the same university have the same responsibilities. The statistical appendix explains the details of this estimate.

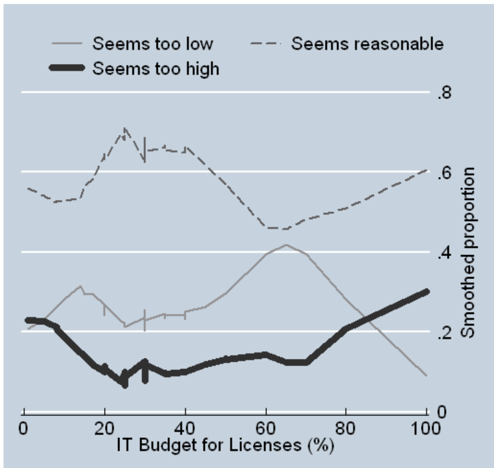
Figure 6. Who chooses software?



Notes: Panel A based on 173 observations; Panel B based on 218 observations. Graph shows averages across universities. For each item, a fixed effects regression fails to reject the null hypothesis that administrative and IT managers from within a university give the same responses. The statistical appendix explains the details of this estimate.

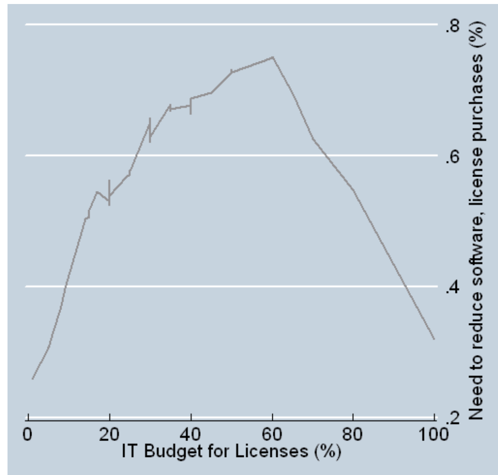
Figure 7. Perceived reasonability of budget.

Panel A. Admin respondents: Reasonability of expenditure on software & license fees



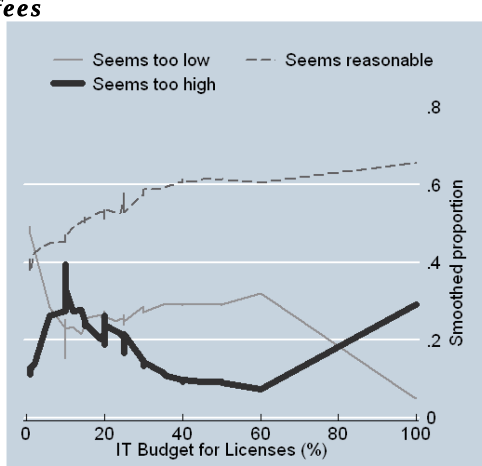
Based on 86 observations.

Panel B. Admin respondents: Need to reduce spending on software & license fees.



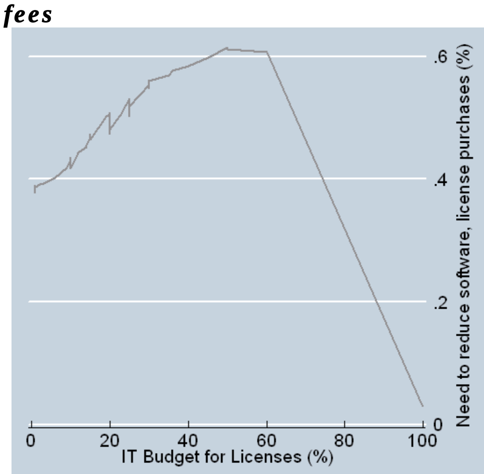
Based on 85 observations.

Panel C: IT managers : Reasonability of expenditure on software & license fees



Based on 102 observations.

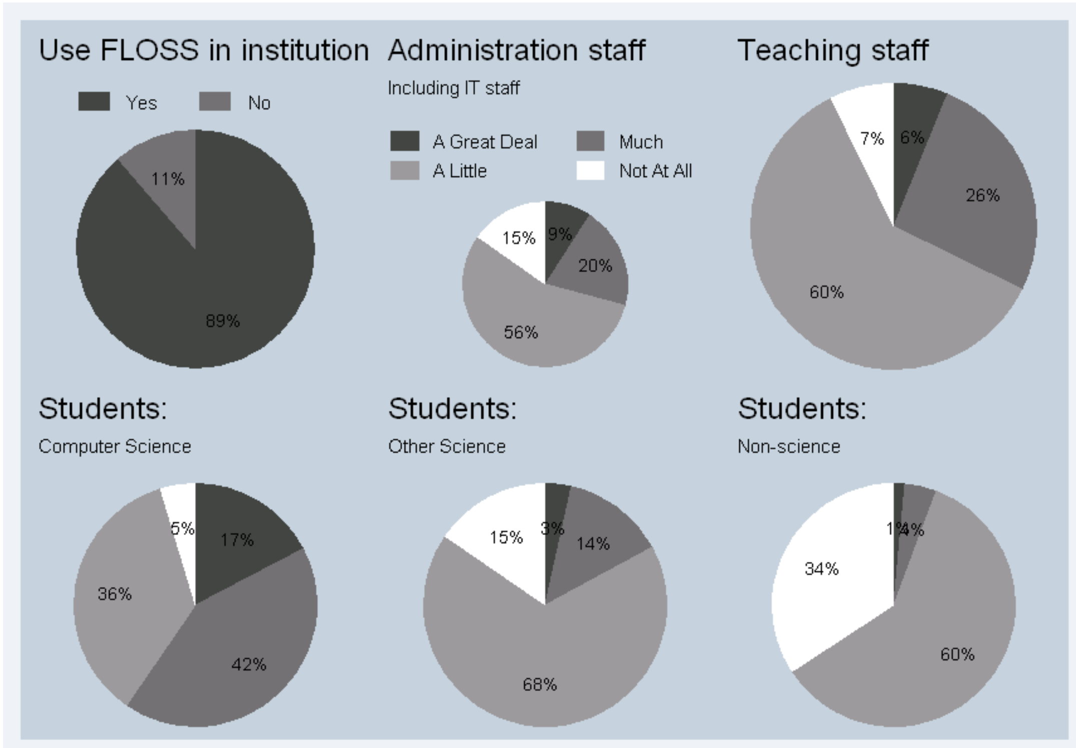
Panel D: Admin respondents: Need to reduce spending on software & license fees



Based on 102 observations.

Note: graphs show averages across individuals. Lines represent density function for each sub-population, estimated using Epanechnikov kernel evaluated at 300 points.

Figure 8. FLOSS use among administrators, teachers, and students
Panel A. All individuals



Panel B. Admin respondents

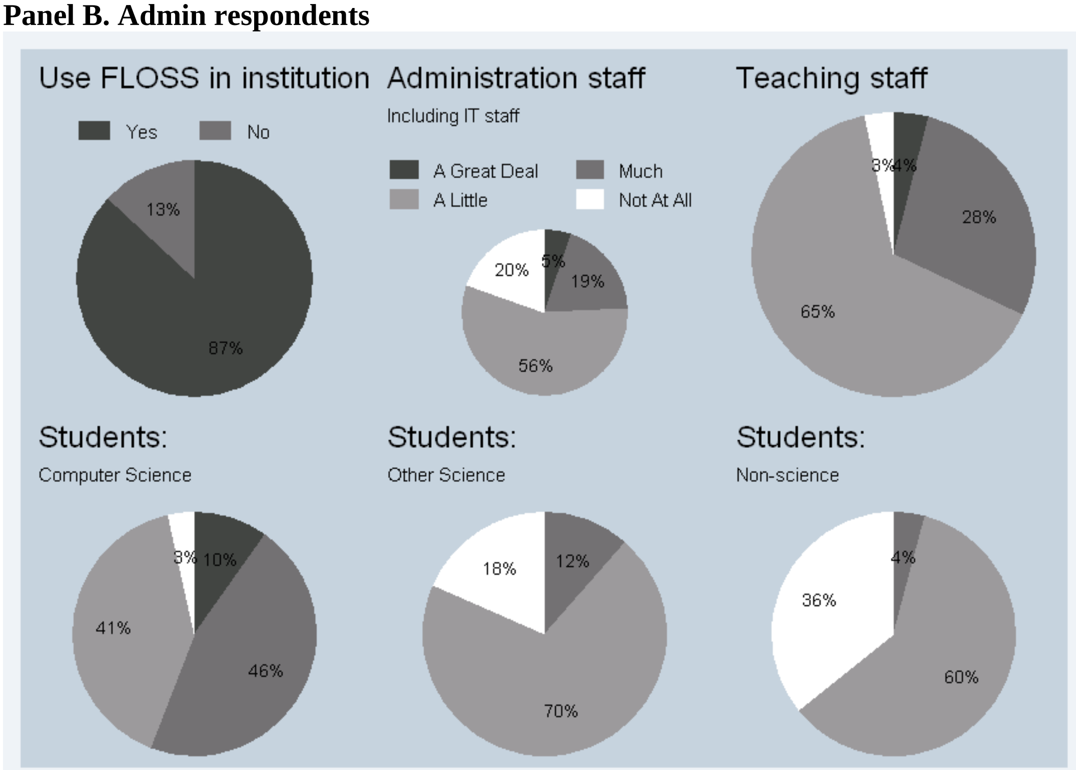


Figure 8, continued: Panel C. IT managers

.

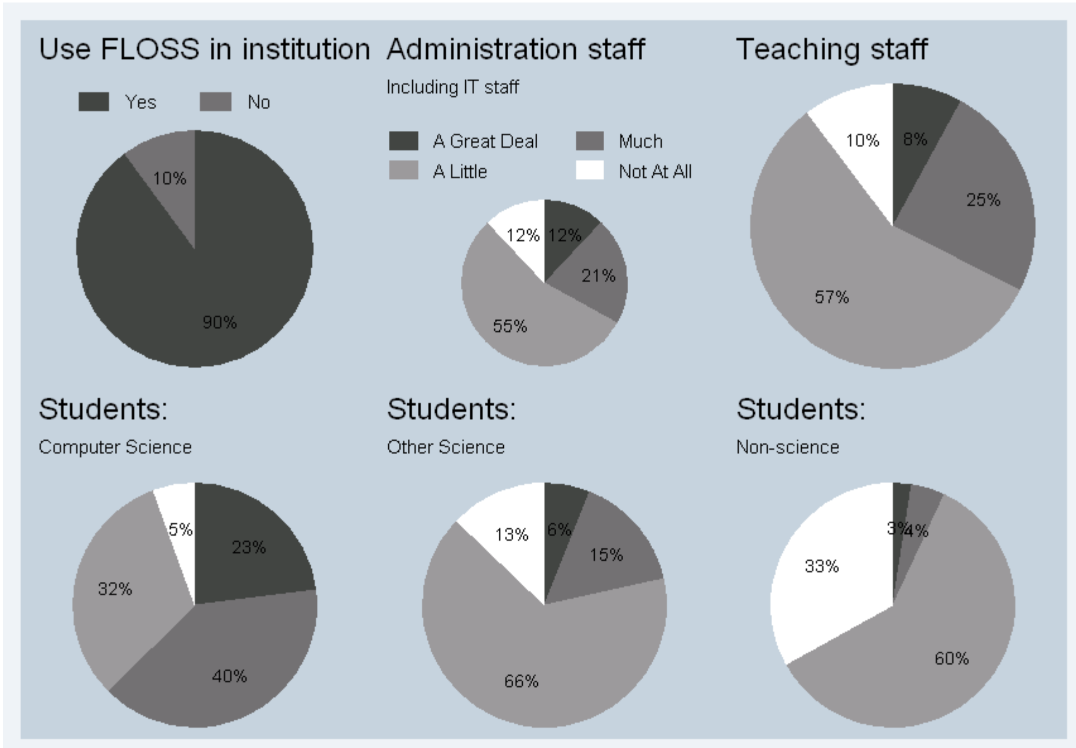
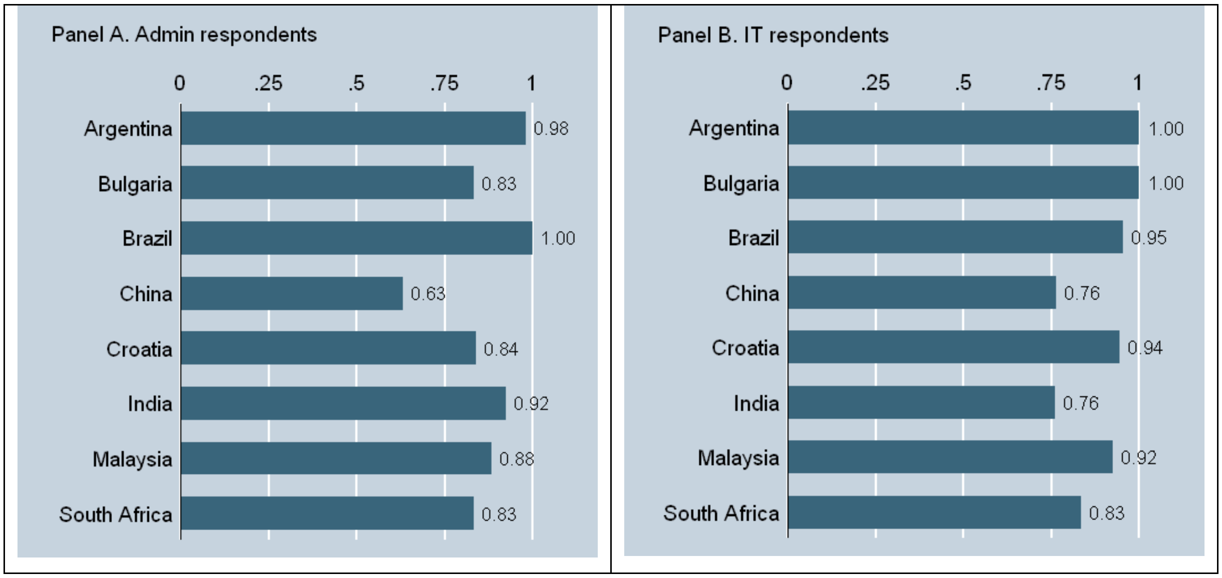

Use FLOSS in institution
Yes
No
10%
90%
Administration staff
Including IT staff
A Great Deal
Much
A Little
Not At All
12%
12%
21%
55%
Teaching staff
10%
8%
25%
57%
Students:
Computer Science
5%
23%
32%
40%
Students:
Other Science
13%
6%
15%
66%
Students:
Non-science
4%
33%
60%

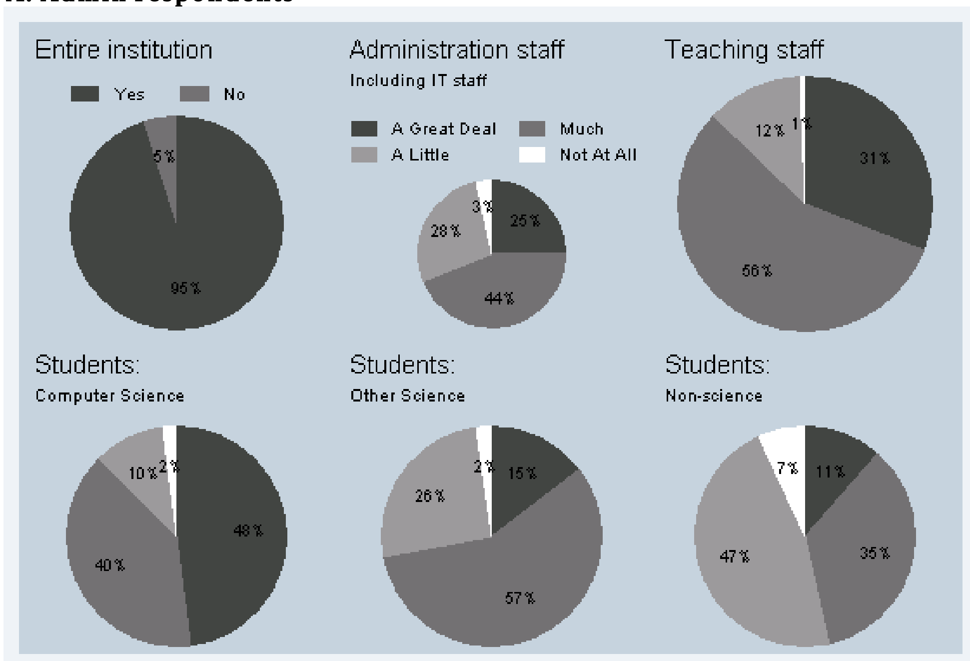
Figure9. Institutional FLOSS use, by country



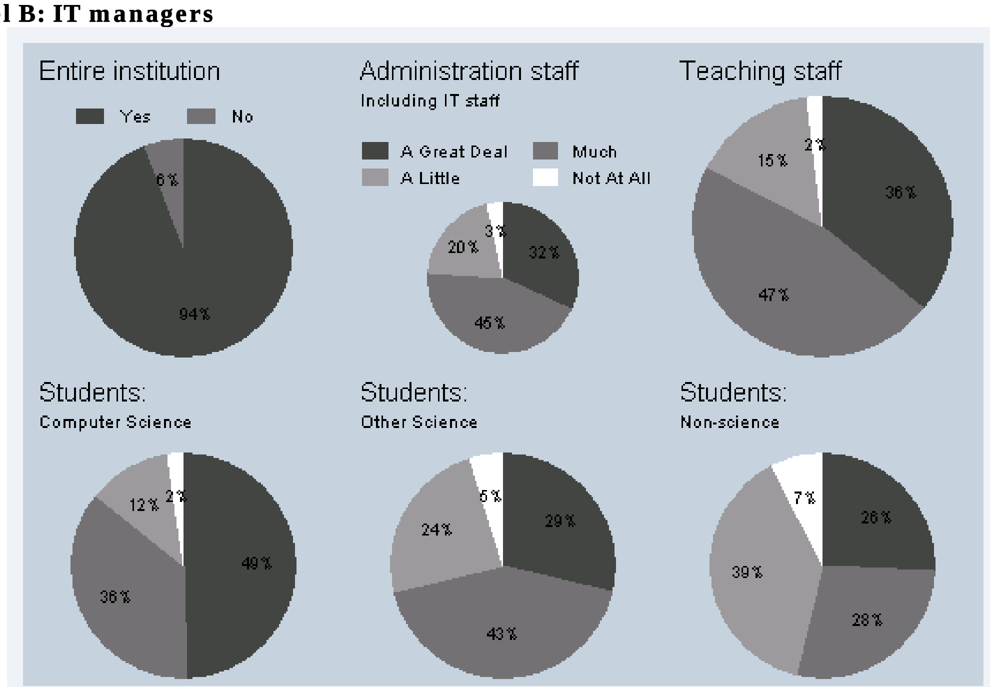
Panel A: 183 observations. Panel B: 238 observations. These graphs show averages across universities. A Rao-Scott F test rejects the hypothesis that admin and IT managers have the same distribution across countries with p-value=0.00. The statistical appendix explains the details of this test.

Figure 10. Extent to which FLOSS use should increase among administrators, teachers, and students.

Panel A: Admin respondents

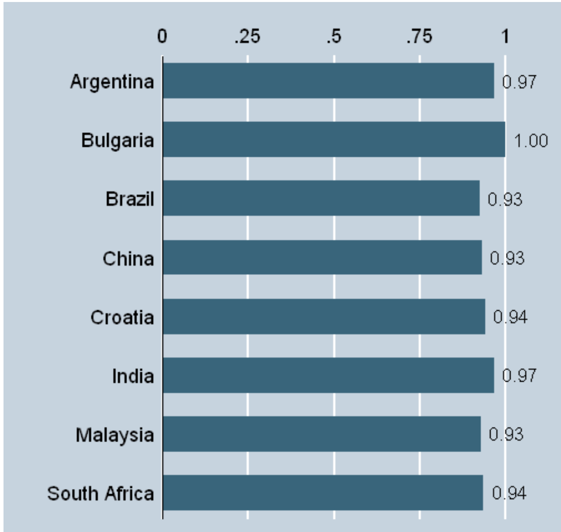


Panel B: IT managers



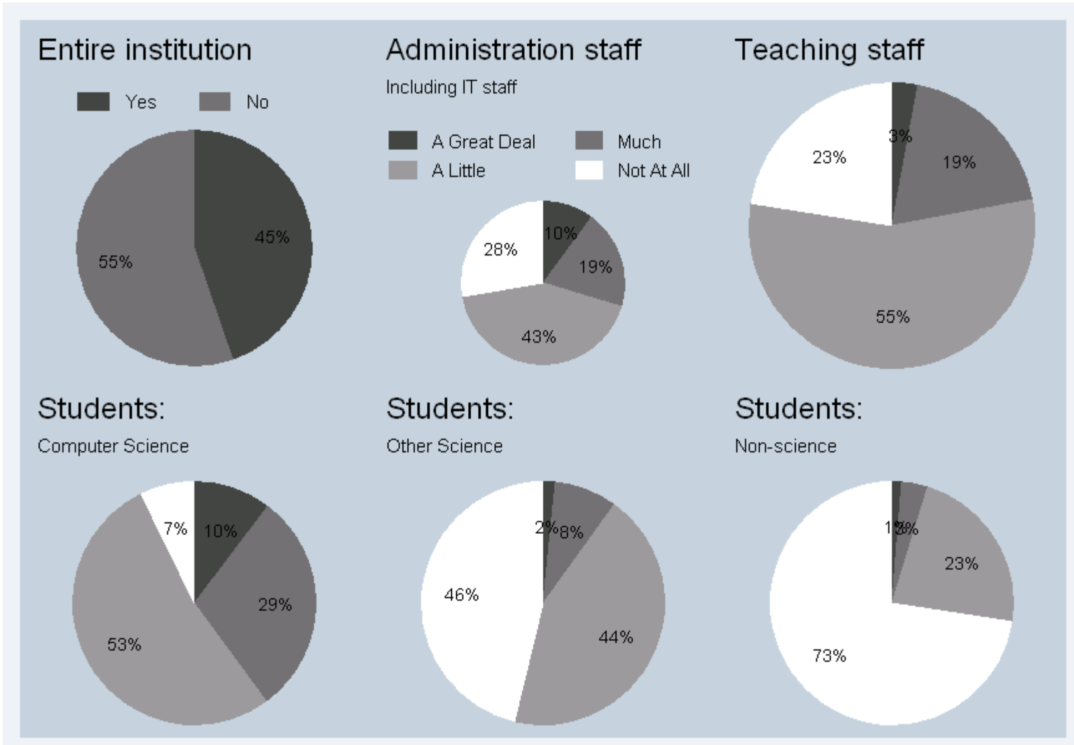
Panel A based on 165 individuals and Panel B based on 225 individuals, though not all individuals answer all sub-questions. Graphs show averages across individuals. Fixed effects regressions for each survey item fail to reject the null hypothesis that admin and IT managers from a university have the same response. The statistical appendix explains the details of these estimates.

Figure 11. Need for FLOSS use to increase in entire institution



Notes:Based on 390 observations. Mean=0.94. Graph shows mean across universities. A Rao-Scott F test fails to reject the null hypothesis that responses are the same in every country (p-value=0.97). The statistical appendix explains the details of this test.

Figure 12. Extent to which administrators, teachers, and students develop FLOSS
Panel A. All individuals



Panel B. Admin respondents

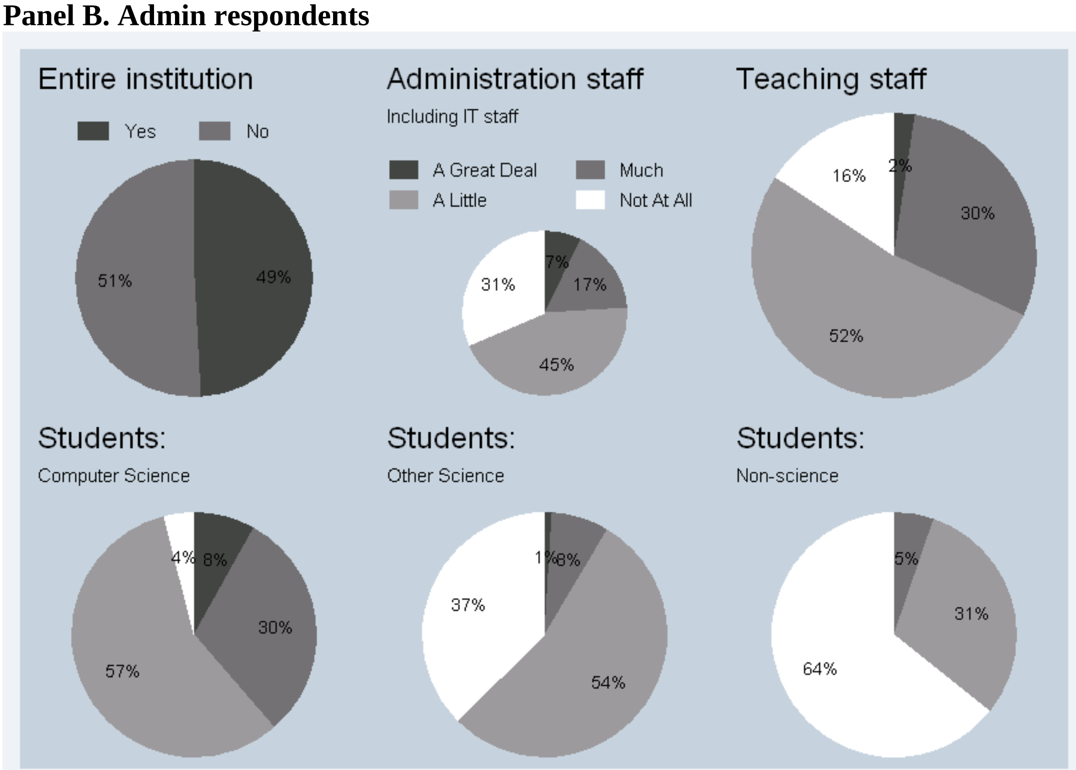
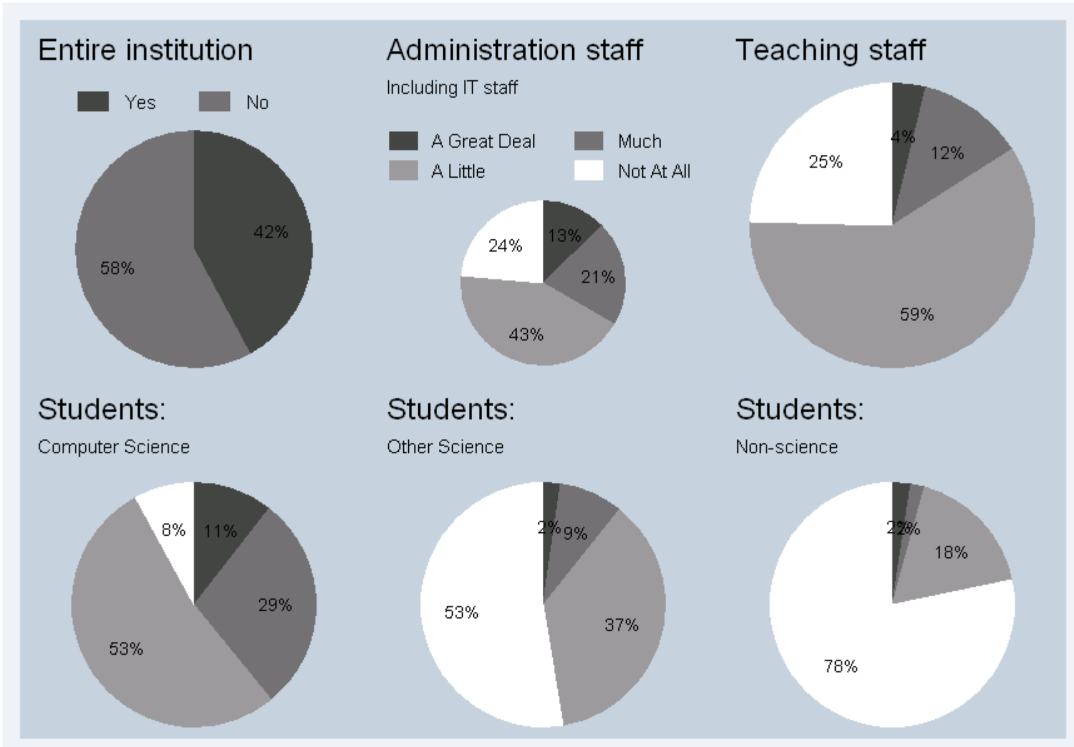
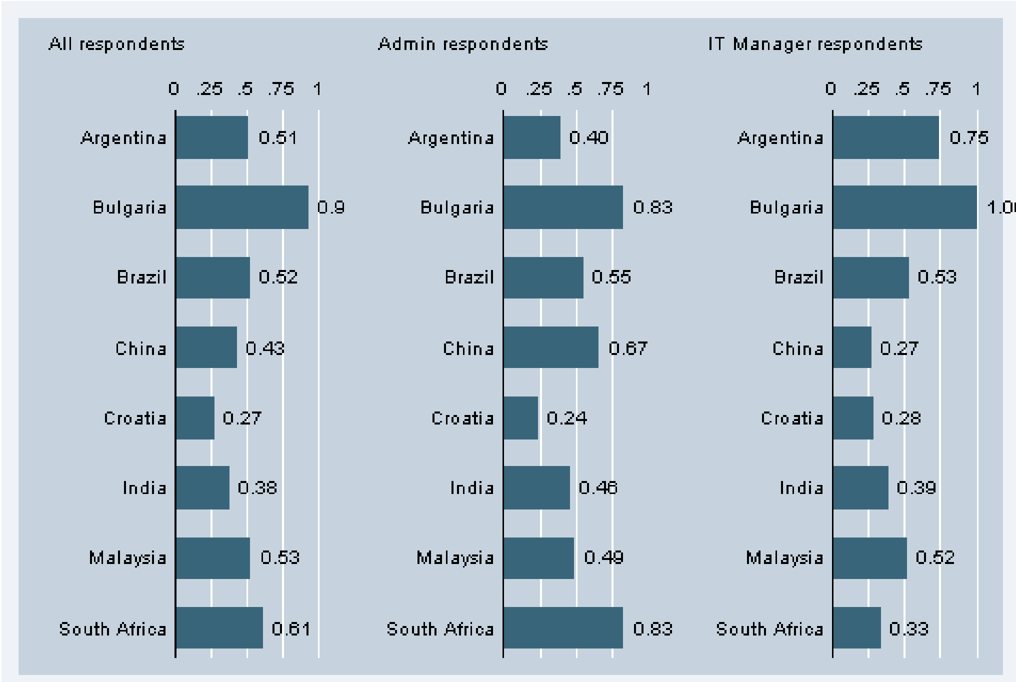


Figure 12, continued: Panel C. IT Managers



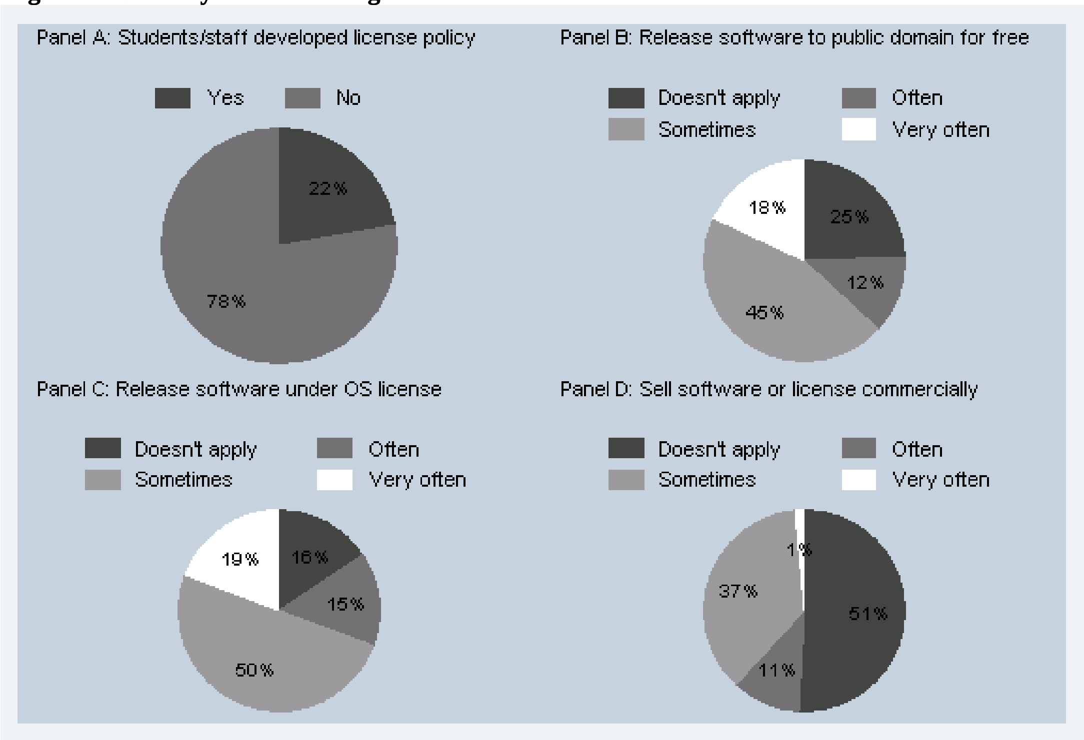
Graphs based on 403 responses, though some individuals do not answer all sub-parts of the question. Graphs show mean across universities.

Figure 13. Do you develop FLOSS anywhere in your institution?



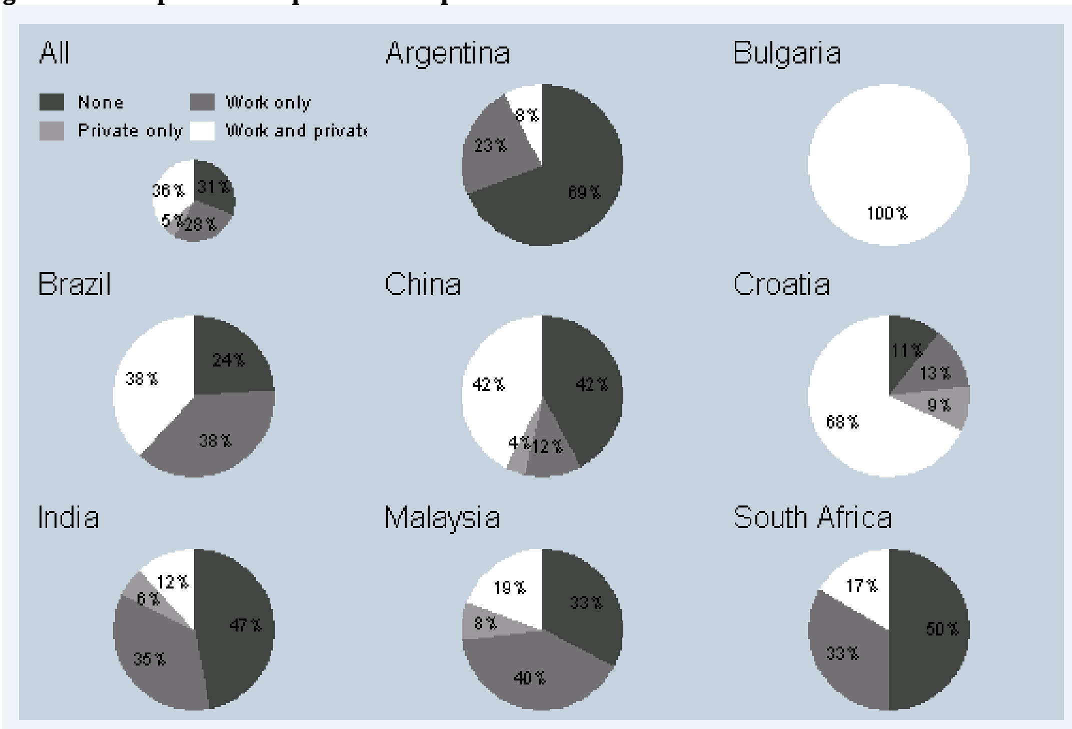
Based on 403 observations. Mean=0.45. Graph shows mean across universities. A Rao-Scott F test rejects the null hypothesis that responses are the same across countries (p-value=0.00). The statistical appendix explains the details of this test.

Figure 14. Policy for licensing out software



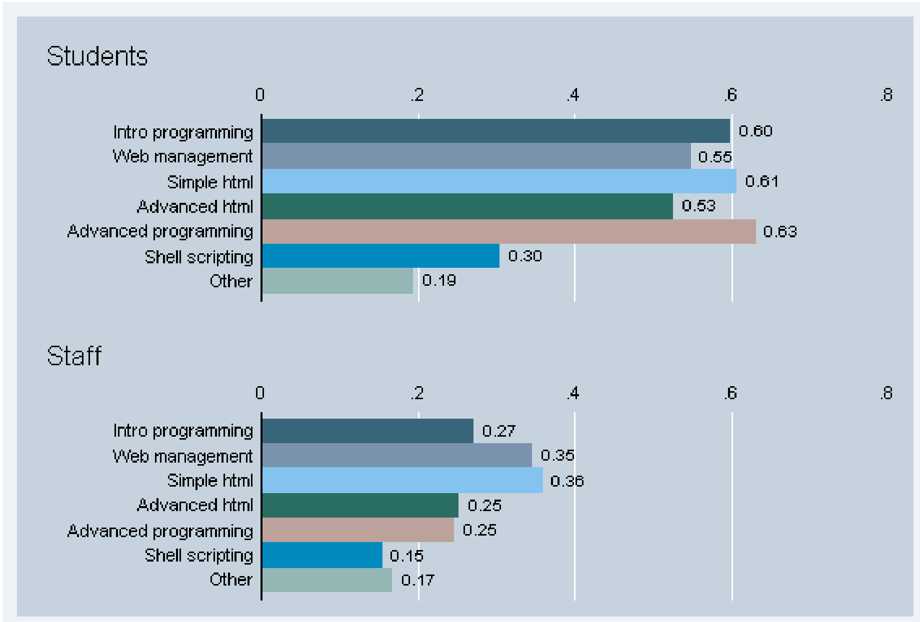
Panel A based on 446 observations; Panels B and C based on 32 observations (including only those who answer "yes" in Panel A); Panel D based on 30 observations.

Figure 15. Respondents' personal experiences with FLOSS.



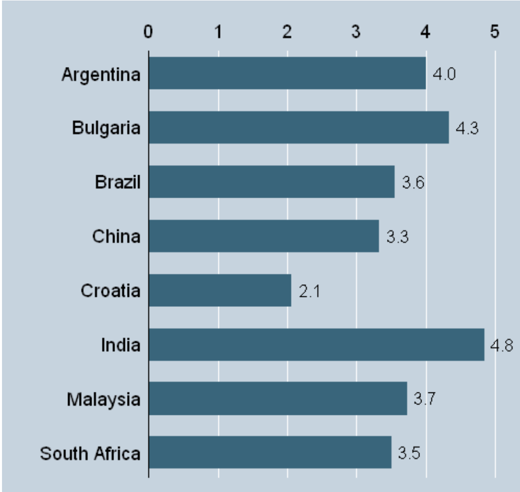
Graphs based on 240 observations. (This question appeared in the IT survey only.) Graphs show mean across individuals. A Rao-Scott F test rejects the null hypothesis that responses are the same across countries with p-value below 0.000. The statistical appendix explains the details of this test.

Figure 16. Availability of computer programming courses to students and staff at institution



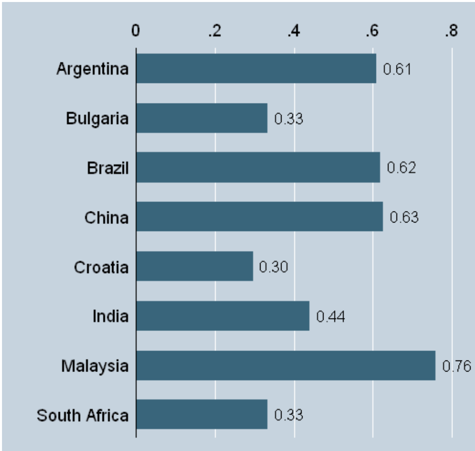
Based on 228 observations. Graphs show mean across universities. Questions appeared in IT survey only.

Figure 17. Mean number of courses available to students, by country



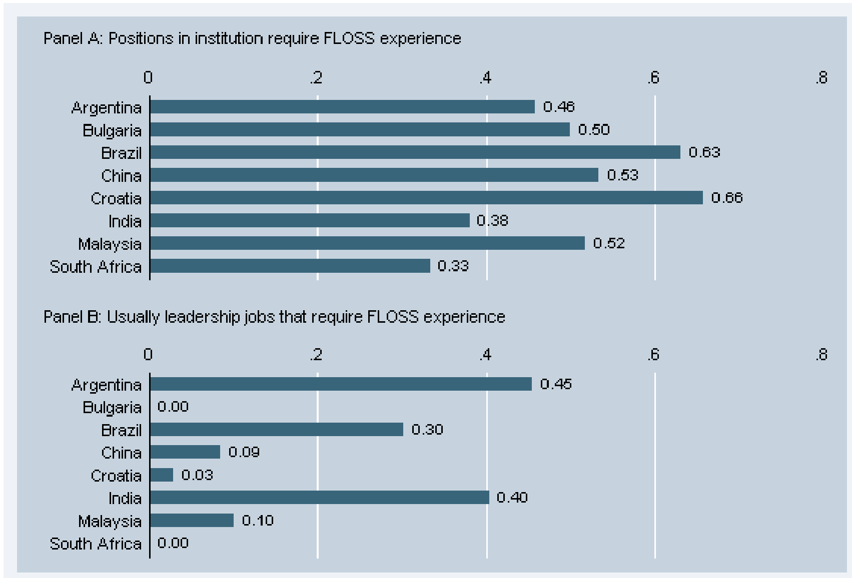
Based on 227 observations. Graphs represent means across universities. Each bar represents the total number of student courses listed in Figure 16 that a respondent institution offers. A Rao-Scott F test rejects the null hypothesis that responses are the same in each country (p-value=0.00). The statistical appendix explains the details of this test.

Figure 18 Do you ask job applicants about their FLOSS experience?



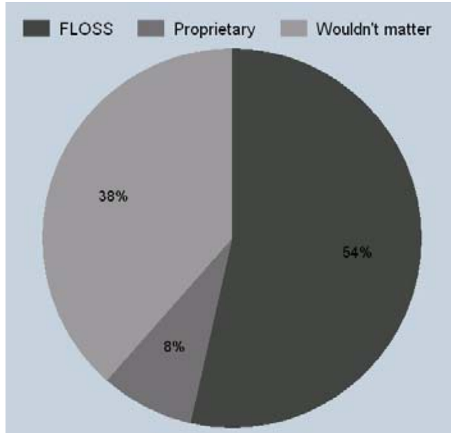
Based on 227 observations. Mean=0.52. Graph shows means across universities. A Rao-Scott F test rejects the hypothesis that responses are the same in each country (p-value = 0.00). The statistical appendix explains the details of this test.

Figure 19 FLOSS jobs in institution



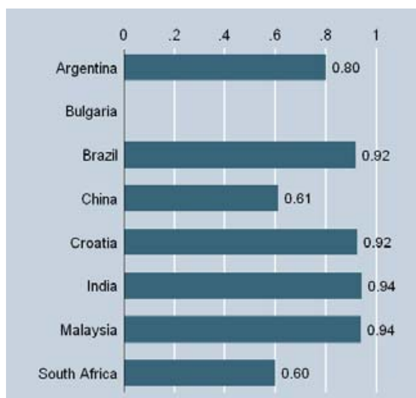
Panel A based on 240 observations. A Rao-Scott F test fails to reject the null hypothesis that response are the same in each country (p-value=0.27). Mean across universities is 0.56. Panel B based on 136 observations and includes only individuals who responded "yes" to the question underpinning first graph. A Rao-Scott rejects the null hypothesis that responses are the same in each country (p-value=0.01). Mean across universities is 0.17. The statistical appendix explains the details of this test.

Figure 20. What type of experience do you prefer in hiring staff?



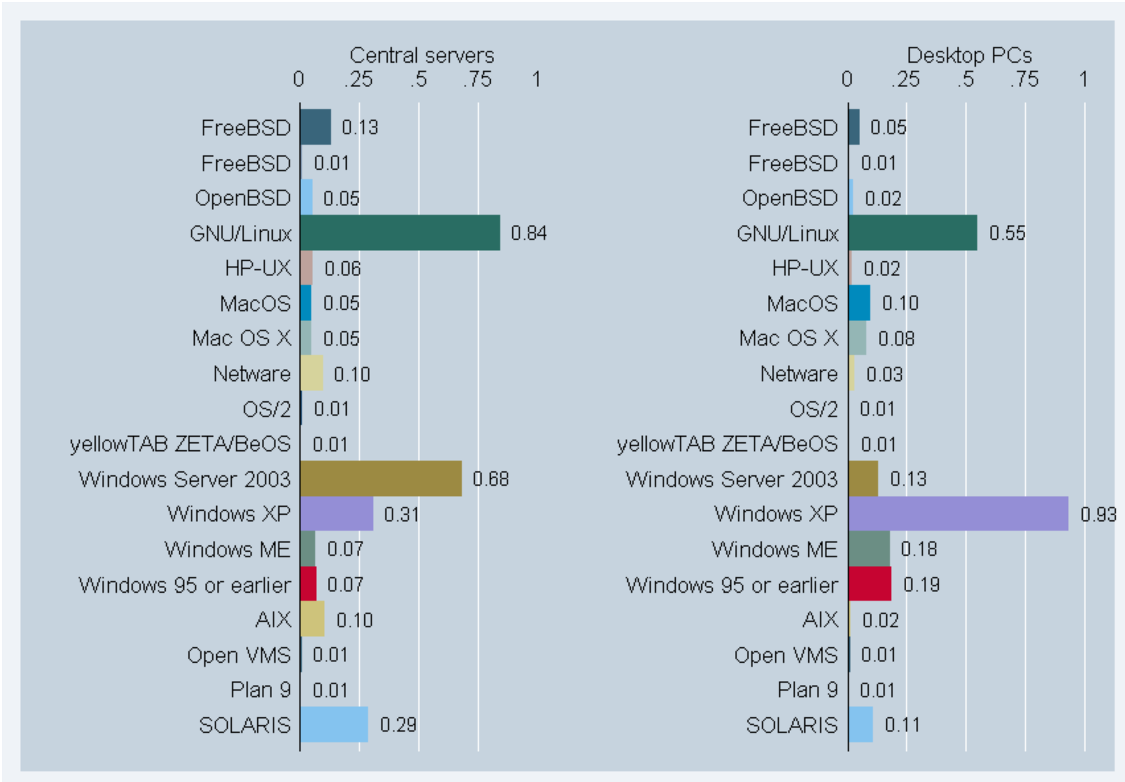
See main text for explanation of exact source question. Based on 226 observations. Graph shows mean across individuals. Question only included in IT survey.

Figure 21. Preference for FLOSS candidate, by country.



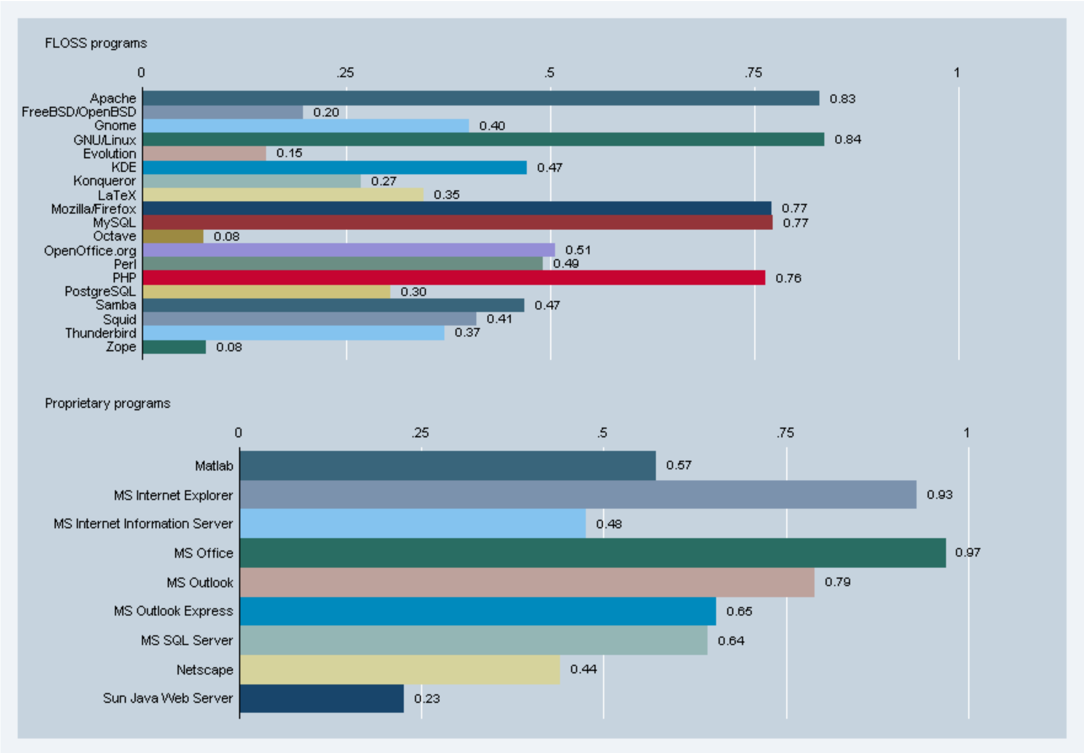
See main text for explanation of exact source question. Based on 226 observations. Graph shows mean within country,a cross individuals. A Pearson χ^2 test fails to reject the null hypothesis that responses are the same in every country (p-value=0.20). Graph includes only respondents who expressed a preference for the FLOSS or proprietary candidate. The statistical appendix explains the details of this test.

Figure 22. Basis of IT system for central servers and desktop PCs.



Based on 236 observations (Central servers); 240 observations (Desktop PCs). Graphs show mean across universities.

Figure 23. Software systems used in institution



Based on 238 observations (first 19 programs) and 244 observations (remaining 9 programs). Graph shows mean across universities.