

## RED HAT/GEORGIA TECH OSPI PROJECT

# OPEN SOURCE SOFTWARE POTENTIAL INDEX (OSPI): DEVELOPMENT CONSIDERATIONS

**Douglas S. Noonan, Paul M.A. Baker, and Nathan W. Moon**

**June 30, 2008**

### 1.0 Context

This document provides background and development notes on the Georgia Tech/Red Hat Open Source Software Potential Index project. The model development was grounded in findings from the literature, presented previously in an annotated review of the literature, insights from industry sources, and interviews with country and regional specialists from Red Hat Inc. The first phase of the OSPI development focused on identifying a theoretically relevant and consistent set of component indicators or “criteria” for ultimate inclusion in the index, as informed by input from expert sources. The second phase focused on determining existing direct and indirect measures of the potential indicators identified in the initial phase, and encompassed the examination of a wide variety of public databases maintained by various international, national, and sub-national agencies and other publically available data sources. The third phase compiled the collected data into a multidimensional index, and was tested across a variety of aggregation schemes (weighted averages, nonlinear combinations, etc.), to produce three candidate models discussed in a following section. The end result, of a configurable OSPI is discussed below and parked temporarily online in the Open Source Class t-square website.

### 2.0 Considerations/ Components

#### 2.1 Literature

*General:* While a variety of different approaches exist for the design of an instrument such as the OSPI, increased efficiency as well as generally improved validity flows from a systematic examination of supporting literature. To this end, a review was conducted of the academic and professional literature focusing on opportunities and barriers to open source adoption and use, as well as probing for broader issues of delineating the conceptual boundaries of the project, such as the exact definition/focus of “Open Source,” the functioning at different levels (local, vs. national, versus international), the impact of the various communities of interest, developers/programmers, users, and institutions, and the impact of OSS characteristics such as cost, choice, security, modifiability.

Several themes consistently emerged from the literature:

- Technology adoption at national (country) level
- Public policy issues and adoption within public sector
- Private sector adoption and use
- Developer roles in adoption and use
- Economic issues pertaining to open source software

While technological issues are primary to the adoption of open source, social, cultural and policy issues also impact OSS diffusion and adoption. Weber (2000) identifies three key issues for social scientists to ponder: 1) motivation of individuals who develop open source; 2) coordination of activities in the supposed absence of a hierarchical structure, and 3) growing complexity in open source projects and its management. While we do not directly spend a great deal of time on these specific variables, we did take them to account, and they to some extent embedded in the contextual active/potential variable classification.

Again, with respect to the development of a robust global OSPI, another critical factor is the availability of accurate data sources for as many countries as possible, a fact observed by a number of social

scientists, who have emphasized a need for empirical data in order to substantiate some of the claims made about open source software. Ghosh, in Feller *et al.*, (2007) explains why little empirical evidence exists for explaining why or how the open source model works. Hard data on the monetary value of OSS collaborative development is almost non-existent. Most models for economic evaluation and measurement require the use of money, and noneconomic activity such as the creation and development of free software is hard to measure in any quantifiable sense. Ghosh contends, therefore, that the lack of objective, "census-type" sources means that many indicators, quantitative and qualitative, may require the use of surveys.

*Adoption at national (country) level:* Scholars have examined the adoption of open source by national governments, via policy mechanism/regulatory approaches. By 2001, Peru, Brazil, Argentina, France, and Mexico all had measures pending that would mandate the use of free software on government computers. Other national and local-level efforts were also taken up in such countries as Germany, Spain, Italy, and Vietnam to establish official alternatives to the use of closed, proprietary software by government. For many free software practitioners, it was the seemingly uncontrollable momentum of their movement and the sheer technical strength of free software itself—more than any particular local actions or activities—that were to credit for its global successes. When considering open source adoption at the national level, one key issue is the government's interests in pursuing this option versus those of other stakeholders who stand to benefit from such a decision.

*Public Sector Adoption and Public Policy Issues:* Whereas some governments have begun to *procure* open source software, others have actually channeled public funds to large-scale open source development projects. The distinction here, as made by Lee (2006), is that a nation that "considers" OSS signifies its desire to establish a level playing field within the public sector's IT procurement policies—such policy is not actually pro-OSS policy because it neither constitutes a government preference for OSS or means the government will choose it. However, when policy makers decide to "prefer" OSS over proprietary software, the decision is likely to be criticized as procurement discrimination by proprietary software developers. Other issues germane for policy makers include OSS's impact on e-government initiatives. Berry and Moss (2006) discuss circumstances in which the discourse and practice of non-proprietary software contribute to opening-up and democratizing e-government. OSS can protect and extend transparency and accountability in e-governments and offer scope for technology to be socially shaped by citizens and associations as well as by administrators and private interests. Finally, political issues such as standards settings and open licensing can impact the deployment of open source software.

*Private Sector Adoption and Use:* Even within national contexts, the private sector remains an important factor when considering the opportunities and barriers to the adoption of open source. Notably, Bonaccorsi and Rossi (2006) call attention to the larger issues surrounding the private sectors' decisions whether to embrace open source, including economic (price/license policy independence), social (conforming to values of OS community), and technological (exploiting feedback and contributions from developers, promoting standardization, security issues) motivations.

*Role of Developers in Adoption and Use Decisions:* The motivations of open source developers in the literature have generally been described by a taxonomy that considers two components of motivation—intrinsic (e.g., fun, flow, learning, community) and extrinsic (e.g., financial rewards, improving future job prospects, signaling quality). Krishnamurthy (2006) identifies four factors as important mitigating and moderating factors in the conversation surrounding developer motivation: 1) financial incentives, 2) nature of task, 3) group size, and 4) group structure. Such issues are important because the motivations of open source developers shape socially the adoption of these systems by firms and governmental agencies. Lin (2006) argues that open source development entails a global knowledge network, which consists of: 1) a heterogeneous community of individuals and organizations who do not necessarily have professional backgrounds in computer science but have developed the competency to understand programming and working in a public domain; and 2) corporations, which results in a hybrid form of software development.

*Economic Issues Pertaining to Open Source Software:* Much of the extant literature on OSS adoption involves the work of economists, many of whom are intrigued OSS's distinct mode of technological development, innovation, and distribution. Lerner and Tirole (2005b) for instance, suggests four questions/issues of interest to scholars studying open source software: 1) technological characteristics conducive to smooth open source development, 2) optimal licensing of open source, 3) the coexistence of

open source and proprietary software, and 4) the potential for the open source model to be carried over to other industries (i.e. the portability of the “open source” concept). An especially important issue has been the relationship between open source software and the software market in general. Forge’s (2006) analysis of the packaged software industry suggests the way forward in economic terms for Europe may well be to follow and encourage open source software for reasons of creating a strong software industry and for a counterbalance to current monopolistic trends. The paper’s findings emphasize the need for investment, education and encouragement in open source software, by both the public and private sectors, to build a strong knowledge-based society in Europe.

## 2.2 Conceptual/Construct Issues

The design of the study posed several interesting challenges. One of these was the dynamic tension between the design of an index that captured “activity” (which we refer to as “active,” and conceptually similar to adoption) and “potential” (roughly related to propensity or capacity) of open source software. We address this by developing two different indices one based on the former and one based on the later. In terms of construct, the OSPI model is composed of dimensions, indicators, and variables. The *dimensions* of OSPI are roughly composed of government, firm, institutional, infrastructure, culture and population categories, and are measured by *indicators*, which are generated by a transformation and/or aggregation of the actual underlying *variables* (or data). Each variable in the inventory of data sets is therefore linked to the dimensions, indicators. The variables are also aggregated into a separate database of variable names and other categorizations, linked to either the Active or Potential index, one of the three dimensions in each index, whether it is direct (related to or impacting OSS) and indirect, contextual variables (e.g., GDP, employment by sector, civil liberties, etc.).

A second design consideration related to both transparency and modularity. Given that more than 750 variables were used in the construction of the OSPI, it is useful to have some sort of organizing heuristics. While the proposed OSPI is not specifically designed around the three sector model (government, education and firms), we have coded the variables so that it is possible to easily identify the government, business, and education/community variables. The OSPI model allows for variable selection manipulation along a variety of these dimensions. The candidate variable list was developed based on the theoretical issues from the literature, consideration of insights and observations of industry sources, and work produced by students in the open source software class.

A third design concern relates to coding and the allocation and “weighting” of variables differently depending on the model used. We initially proposed three models (named “Trinity,” Matrix” and “Hybrid Neo”) described below, that allocated variables slightly differently depending on the model. The models do not have a direct equivalency to each other as they were experimental conceptual approaches to capturing different components of OSS adoption and diffusion. The final proposed model, which is essentially, a modification of the Hybrid Neo model, draws on aspects of the other two models, and is also described below.

## 2.3 Data Limitations

*Countries for inclusion:* In a perfect world the OSPI would be able to draw on a wide variety of data sets that were populated with identical global data. In practice, there is a trade-off between the number of countries directly modeled and the range of variables included such that as the total number of countries becomes more complete the number of completely represented data variables decreases. Conversely, the larger the number of variables included in the OSPI the smaller the number of countries for which complete and up to date data exists. There are of course several ways in which to deal with this. Due to time and resource limitations our development of the OSPI was based on available (secondary) data sets which were variable in inclusiveness both cross-sectionally (number of countries) and longitudinally (over time). Future efforts and wish lists, might be oriented to capturing additional variables, particularly in some of the more complex variables (for example, related to specific OSS characteristics in countries). This would be primary data collection, and could be done by surveys on a country by country basis. Another approach for generating absent data was by achieved through the use of proxy or imputation (e.g., multiple imputation of chained equations, alternative models using inclusive vs. robust approaches) where absent data can be estimated based on relationship or similarity to associated variables.

The following countries were included in the robust (“short”) list, that is countries for which the richest number of variables were available, and for which there was sufficient level of OSS related activity, and for which the literature, and expert sources, indicated would be good candidates for selection.

Argentina	Australia	Belgium
Brazil	Canada	China
Czech Republic	Finland	France
Germany	Hong Kong	India
Ireland	Italy	Japan
Korea	Malaysia	Mexico
Netherlands	Singapore	Spain
Sri Lanka	Sweden	Switzerland
Taiwan	United Kingdom	United States

The master list of variables considered for the OSPI included more than 750 base variables which were tested in the model development. The actual variables included in the OSPI are detailed elsewhere.

### 3.0 OSPI Model

We have included a brief discussion of the three preliminary approaches (below) to help demonstrate the logic behind the development of the OSPI. We experimented with several different approaches to variable selection and interaction, and the final OSPI is based on a variation of the Hybrid Neo conceptual model, and uses two sub-indices (Active and Potential), representing aggregations of indicators along several dimensions (government; firms or commercial enterprises; and community and educational system) of variables.

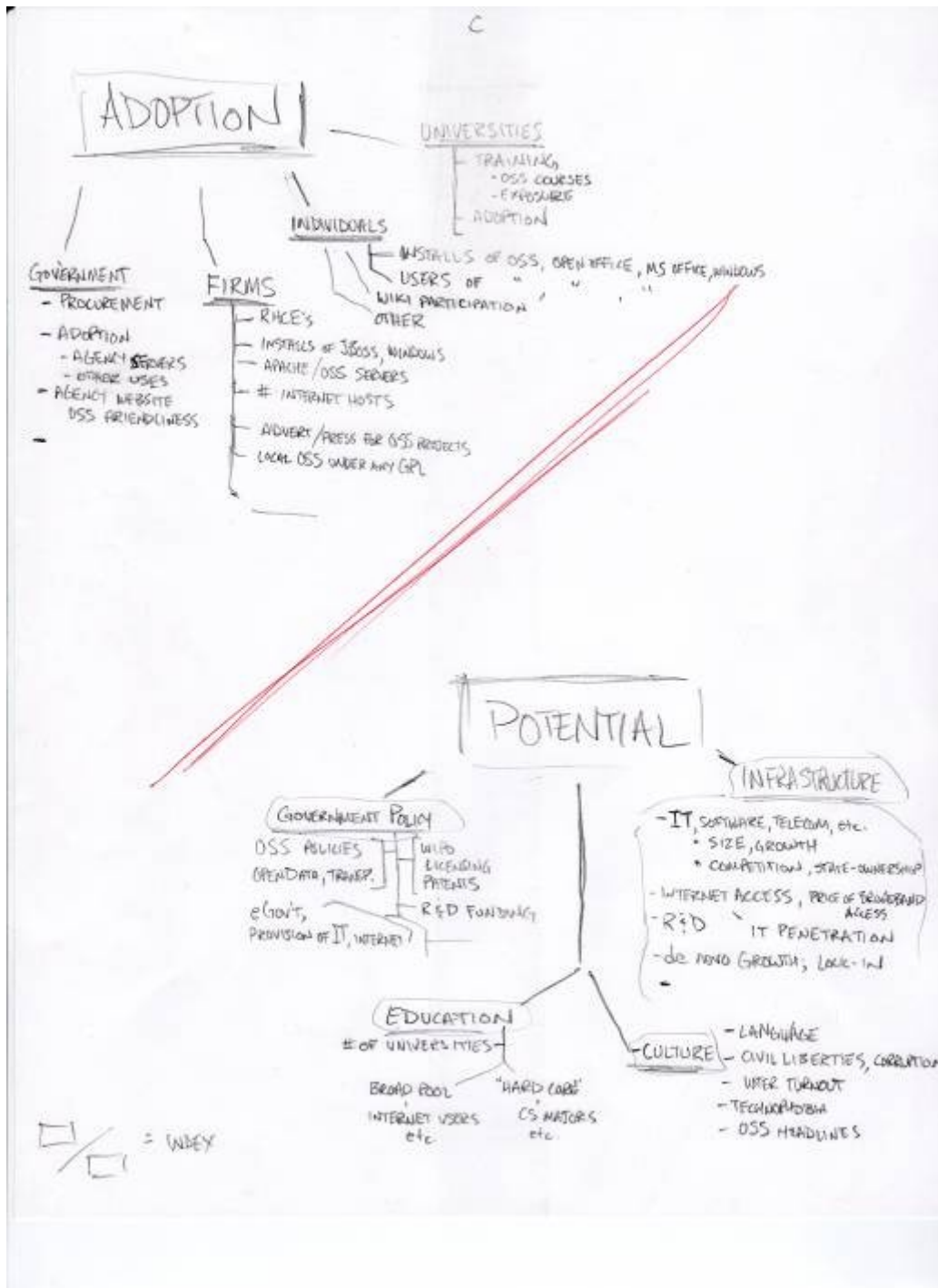
3.1 *Preliminary models.* We initially proposed three different OSPI models, each with a “long” (reduced number of indicators but drawing in a larger number of countries) and “short” (that is a larger array of variables, but reducing the number of countries for which data exists) version. Each of the models captured different aspects of the underlying phenomena and consequently have different advantages and limitations. The original conceptual framework diagrams appear in Appendix III. We have assigned working “names” to each of the models for internal purposes. These are:

*The Trinity Index* (Index A in Appendix III ) is an index with three dimensions (sub-indices) based on RedHat’s three-legged stool metaphor. The sub-indices (dimensions) are roughly related to business, education, and government clusters and composed of weighted averages of proxy measures that correspond to indicators drawn from an extensive review of the literature in conjunction with input from Red Hat and Industry sources. While an intuitively appealing model, one of the concerns of the model is that the a priori assumptions of the importance, or impact of the variables composing the dimensions may obscure or limit phenomenon the index seeks to capture..

*The Matrix Index* (Index B in Appendix III) is a data-intensive approach that divides proposed variables into two types, “Direct and Indirect Indicators: those *directly* related to OSS (e.g., Firefox downloads, articulated government policies, availability of internet access, and RHCEs, among others) and *indirect*, contextual variables (e.g., GDP, employment by sector, civil liberties, etc.). Once divided, we constructed the index as a weighted average of the direct variables. The “direct index” is regressed on each of the indirect variables, which is in turn used to predict the Matrix Index. While robust methodologically, this model reduces the ability of the analyst to apply expert knowledge, or to experiment with their own weighting preferences.

The Hybrid Neo Index (Index C in Appendix III), and the basis of the proposed OSPI is a hybrid approach to the index, composed of two component sub-indices: Activity (i.e. adoption/diffusion related variables) and Potential (or the aggregate of variables that suggests potential for future adoption). The Hybrid Neo Index is not, a priori, a restricted weighted average of Activity and Potential; it allows for the testing of different aggregations of variables (Potential  $\times$  Activity), (Potential/Activity), or other types of algorithmic manipulation, intentionally left open-ended. This algorithm was used to explore relationships between indicators and composite sub-indices.

### Conceptual Framework of Final OSPI Model [Derived from "Hybrid Neo" [Index C]]



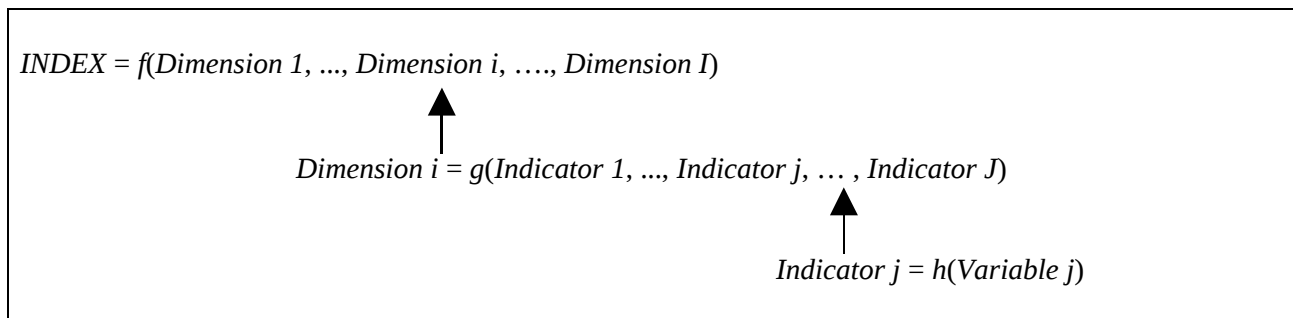
3.2 *Final OSPI Model*. The final proposed model, which is essentially, a modification of the Hybrid Neo model, draws on aspects of the other two models. The model code performs data transformations and imputes values for some countries that are missing values (where possible). It then pools these variables together into indicators, roughly along the lines of the equations that appear below. The elements inside the parentheses in those equations represent placeholders for reasonable "indicators" and the dimensions they feed into. The code aggregates the variables into dimensions (roughly analogous to "sub-indices") using a variety of different aggregation algorithms (e.g., arithmetic means, geometric means, minimum values). The model aggregates data into the dimensions/sub-indices, allows for the comparison of the rankings of countries in all of the various indices. The OSPI is contained in three files (attached) (1) OSI.txt (which contains the final index values), (2) transformed.txt (which has all of the "transformed" variables (z-scores, etc.) and other massaged variables, including the weights that were used in the "matrix" approach), and (3) OSdata.txt (with raw data that feeds into #2, which, in turn, feeds in #1). A separate file is attached (VarList.xls) containing the component variables.

The model algorithm consists of two primary indices (Active & Potential) based on five different aggregation rules and two different sets of indicators depending on the inclusion of the robust ("short") or inclusive ("long") lists of countries, and three different dimensions (government related, firms, community/cultural) for each of these. This generates, essentially, some sixty different possible indices, which are subjected to a sensitivity analysis.

*Index Construction*. Two indices are developed here, following a large literature on the development of indices to measure (unobserved) environmental conditions (e.g., index of sustainability, human development index). The first index captures the OSS-related *activity* in a country. The second index captures the OSS-related *potential* of a country. The conceptual theory behind the construction of the index is overviewed briefly here.

The conceptual layout of the indices follows this form:  
 $INDEX = f(Dimension\ 1, Dimension\ 2, \dots, Dimension\ I)$ , where  $f$  is a function that aggregates  $I$  different dimensions (i.e., sub-indices). For the  $i$ th dimension of the index, it is based on several indicators.  
 $Dimension\ i = g(Indicator\ 1, Indicator\ 2, \dots, Indicator\ J)$ , where  $g$  is another aggregating function for the  $J$  different indicators in the  $i$ th dimension. In practice, each indicator is derived from a variable that measures or proxies for that indicator. The variables typically require some transformation and normalization, as they are generally measured in wildly different units. Let  $Indicator\ j = h(Variable\ j)$ , for some transformation function  $h$  that might include some imputation, transformation, and normalization of  $Variable\ j$ .

To summarize the general form:



This generalized conception of an index is applicable to a wide variety of indices. Regardless, the arbitrary choices of functions  $f$ ,  $g$ , and  $h$  can greatly affect the ultimate rank-orderings from the index aside from their obvious impact on the ultimately index values. Ebert and Welsch (2004) theoretically

demonstrate the arbitrariness that follows from certain combinations of variable types and aggregation rules – the result being dubious indices that violate some fairly standard ordering rules.

In the present context, there are two different indices – each with the same general structure of multiple dimensions, several indicators for those dimensions, and variables measuring those indicators. After collecting hundreds of variables from preexisting datasets, the variables are assigned to corresponding indicators and the functions  $f$ ,  $g$ , and  $h$  are decided. Sensitivity analyses follow to identify the robustness of the indices to the designers' arbitrary choices.

The two indices are for activity (A) and potential (P). Both of the indices generated here follow the same dimensional structure: government (G), firms or commercial enterprises (F), and community and education system (C). Although the dimensions are the same in the A and the P index, the indicators for each of the dimensions are different between the two.

G = government;

F = firms or commercial enterprises;

C = community and educational system

Active =  $f(G_A, F_A, C_A)$

Potential =  $f(G_P, F_P, C_P)$

$G_A = g(\text{procurement, policy, use})$

$G_P = g(\text{software policy, corruption and liberties, e-government, IP law})$

$F_A = g(\text{RHCEs \& other developers, firms' installs/users, firms developing/supporting OSS})$

$F_P = g(\text{ICT industry size/competition, ICT growth, R\&D, internet access, de novo growth})$

$C_A = g(\text{household installs/users, OSS courses or adoption by educators, discussion in media, language supported})$

$C_P = g(\text{culture, education, CS majors, internet users})$

Because of the nature of existing international data, most variables cover only a limited number of countries and years. Missing values are prevalent in the dataset and, unfortunately, require difficult choices and compromises in order to produce an index. To facilitate this, variables are roughly classified according to whether they cover a “short” ( $N < 100$ ) or a “long” ( $N > 120$ ) list of countries. As is usual, “short” (S) variables tend to be of higher quality or more directly related to important indicators, whereas “long” (L) variables are more general and indirectly relate. The index construction recognizes this balance and attempts to separately create a “short” and a “long” index – where the latter sacrifices some variable quality in order to obtain greater coverage of countries.

All variables used in constructing the index relate either “directly” or “indirectly” to an indicator. This classification merely captures how closely the variable pertains to the indicator. For instance, the “Firefox users” variable relates directly and “PCs per capita” variable relates indirectly to the *household installs* indicator (in  $C_A$ ). Direct measures are preferred to indirect measures, when available. When both are available, the distinction between direct and indirect variables is relevant to the use of regression techniques to derive dimension weights (as discussed below).

Each numeric variable is either a ratio-scale or interval-scale measure. The former (e.g., population, expenditures) has the important attribute of a natural zero in its measurement allowing for meaningful ratios. The latter (e.g., civil liberties index) has no such advantage. As Ebert and Welsch (2004) show, geometric means of ratio variables can preserve orderings independent of the choice of transformations  $h$ . Thus, variables are categories by whether or not they are ratio-scale.

In summary, any *Variable j* used to create an *Indicator i* will be classified as “best” (B), “ratio” (R), “direct” (D), or any combination of those. Each *Variable j* is also either “short” (S) or “long” (L) depending on how many missing values it has. Because there up to 3 different indicators (B, R, D) and 2 different length, for each indicator, for each dimension, for each index ... a multitude of indicators must be created. Many of these are the same, however. The table below shows this. First a table shows the naming

conventions for each indicator in both indices. The second table shows the variables actually selected for the corresponding indicators (note that the top variable of each pair is the “long” variable).

### INDICATOR NAMES

Index	Dim	Indicator	best	direct	ratio	
A	G	procurement	AGprocBL AGprocBS	AGprocDL AGprocDS	AGprocRL AGprocRS	
		policy	AGossBL AGossBS	AGossDL AGossDS	AGossRL AGossRS	
		use	AGuseBL AGuseBS	AGuseDL AGuseDS	AGuseRL AGuseRS	
	F	RHCEs & other developers	AFdevelBL AFdevelBS	AFdevelDL AFdevelDS	AFdevelRL AFdevelRS	
		firms’ installs/users	AFuserBL AFuserBS	AFuserDL AFuserDS	AFuserRL AFuserRS	
		firms developing/supporting OSS	AFossBL AFossBS	AFossDL AFossDS	AFossRL AFossRS	
	C	household installs/users, Wiki participants	ACuserBL ACuserBS	ACuserDL ACuserDS	ACuserRL ACuserRS	
		OSS courses, adoption by educators	ACadoptBL ACadoptBS	ACadoptDL ACadoptDS	ACadoptRL ACadoptRS	
		discussion in media	ACdiscussBL ACdiscussBS	ACdiscussDL ACdiscussDS	ACdiscussRL ACdiscussRS	
		language supported	AClangBL AClangBS	AClangDL AClangDS	AClangRL AClangRS	
	P	G	software policy	PGsoftBL PGsoftBS	PGsoftDL PGsoftDS	PGsoftRL PGsoftRS
			corruption and liberties	PGlibertBL PGlibertBS	PGlibertDL PGlibertDS	PGlibertRL PGlibertRS
e-government			PGegovBL PGegovBS	PGegovDL PGegovDS	PGegovRL PGegovRS	
IP law			PGiplawBL PGiplawBS	PGiplawDL PGiplawDS	PGiplawRL PGiplawRS	
F		IT industry size/competition	PFiteconBL PFiteconBS	PFiteconDL PFiteconDS	PFiteconRL PFiteconRS	
		IT growth	PFitgroBL PFitgroBS	PFitgroDL PFitgroDS	PFitgroRL PFitgroRS	
		R&D	PFrndBL PFrndBS	PFrndDL PFrndDS	PFrndRL PFrndRS	
		internet access	PFaccessBL PFaccessBS	PFaccessDL PFaccessDS	PFaccessRL PFaccessRS	
		de novo growth	PFdenovoBL PFdenovoBS	PFdenovoDL PFdenovoDS	PFdenovoRL PFdenovoRS	
C		culture	PCcultBL PCcultBS	PCcultDL PCcultDS	PCcultRL PCcultRS	
		education	PFeduBL PFeduBS	PFeduDL PFeduDS	PFeduRL PFeduRS	
		CS majors	PFcsmajBL PFcsmajBS	PFcsmajDL PFcsmajDS	PFcsmajRL PFcsmajRS	
		internet users	PCinterBL PCinterBS	PCinterDL PCinterDS	PCinterRL PCinterRS	



**Variables selected:**

Index	Dim	indicator	best	direct	ratio	
A	G	procurement	OSSpolNatman GovExppGDP	OSSpolNatman OSSpolNatman	GovExppGDP GovExppGDP	
		policy	OSSpolNatRD OSSfunding	OSSfunding OSSfunding	OSSpolNatRD OSSfunding	
		use				
	F	RHCEs & other developers	RHCEpc RHCEpc	RHCEpc RHCEpc	RHCEpc RHCEpc	
		firms' installs/users	LinuxUserspc LinuxUserspc	LinuxUserspc LinuxUserspc	LinuxUserspc LinuxUserspc	
		firms developing/supporting OSS				
	C	household installs/users, Wiki participants	GoogleApp GoogleApp	GoogleApp GoogleApp	GoogleApp GoogleApp	
		OSS courses, adoption by educators	SchoolNet		SchoolNet	
		discussion in media	rOSSnews rOSSnews	rOSSnews rOSSnews	rOSSnews rOSSnews	
		language supported	LinuxLang LinuxLang	LinuxLang LinuxLang		
	P	G	software policy	nPiracy OOXML	OOXML OOXML	nPiracy nPiracy
			corruption and liberties	nCivLib nCivLib	nCivLib nCivLib	Turnout Turnout
			e-government	eGov eGov	eGov eGov	eGov eGov
IP law			nTRIPS nIPRI	nTRIPS nIPRI		
F		IT industry size/competition	ICTtop250pGDP ICTtop250pGDP	ICTtop250pGDP ICTtop250pGDP	ICTtop250pGDP ICTtop250pGDP	
		IT growth	newCellGro TelcomInvestpc	TelcomInvestpc TelcomInvestpc	newCellGro TelcomInvestpc	
		R&D	SciArticlespc RnDeploypc	RnDeploypc RnDeploypc	SciArticlespc RnDeploypc	
		internet access	nNetPrice nNetPrice	Broadband Broadband	nNetPrice nNetPrice	
		de novo growth	inewGrowth inewGrowth	inewGrowth inewGrowth	inewGrowth inewGrowth	
C		culture	TVpc TVpc	. nTechphob	TVpc TVpc	
		education	College GradEngpgrad	College GradEngpgrad	College GradEngpgrad	
		CS majors	PCspc PCspc		PCspc PCspc	
		internet users	InternetUserspc InternetUserspc	InternetUserspc InternetUserspc	InternetUserspc InternetUserspc	

Notice the grey-shaded cells, where only 11 out of 57 cells did not have a suitable and available variable (obtainable during the project period). Only 2 out of the 23 total indicators have no variables available, and none for the Potential (P) index. Note also a few other conventions: variables with an 'n' prefix enter the index as a negative value, the 'pc' suffix indicates a "per capita" scaling, the 'pGDP' suffix indicates "per GDP" likewise, and throughout the prefix 'z' indicates that the variable has been Z-score transformed. Most variable definitions are available in the "Variable definitions" worksheet of the VarList.xls file. Many other variables (beyond those listed in the table above) were available for at least some of the countries, and all were thoroughly considered before settling on the above selections.

Alternate variables can easily replace the chosen ones, given the structure above, for calculating new versions of the index. All variables are normalized (i.e., transformed to a Z-score) before entering the index.

After all of the requisite scaling and normalizations (and a few imputations) have been completed, the next step is to settle on the  $g$  functions that aggregate the multiple indicators into single dimension (i.e., "sub-index") values. For each dimension in each index, several  $g$  functions are computed. This includes an arithmetic mean ('am'), a geometric mean ('gm'), a maximum value ('ma'), and a minimum value ('mi'). For the minimum, maximum, and arithmetic mean aggregations, missing values for the constituent indicators are simply ignored and the operation is applied to the remaining indicators for which a value exists (unless, of course, all indicators are missing, in which case the dimension value is also missing). For the geometric mean, a similar approach is taken, except that the dimension value is assigned a "missing" value if all or all but one constituent indicators have missing values.

The final step in initially constructing the indices involves deciding on the aggregation function  $f$  to compile the three dimensions into a single index value. Common choices for aggregating the dimensions include arithmetic means, minimum values, and maximum values. Because the dimensions themselves are aggregates of indicators, these  $f$  functions selected are referred to as arithmetic means ('am', which is the arithmetic mean of dimensions that are themselves arithmetic means of indicators), maximins ('mx', which is maximum value of the dimensions that are themselves minimum values of indicators), and minimeans ('mi', which is the minimum value of dimensions that are themselves arithmetic means of indicators). The first is straightforward to interpret, as an average of indicator averages. The second is the "best" dimension, where dimensions are measured by their "worst" contributor. The third is the "weakest" dimension, where dimensions are themselves averages. Of course, other aggregations are possible as well (e.g., a minimin function would be the minimum dimension where dimensions themselves took the minimums of their constituent indicators – a sort of "weakest link" function).

Two other novel aggregations offered here are worth emphasizing. First, the geometric mean ('gm') aggregation bears some distinction as having the being the most robust, in theory, to arbitrary scaling effects for ratio-scale variables (see Ebert and Welsch, 2004, and others). The advantage of the geometric mean indices arises when ratio-scale variables are used. Thus, the 'gm' indices should be ones that are also 'R' indicators to take full advantage. (A trade-off arises here because several components of the indices, at least in theory, are largely immune to being measured in ratio-scale, such as measures of "liberty" or "language" or "IP law" that are typically only found in interval-scale.)

Second, the "Matrix" index approach from our preliminary work has been extended to provide a weighted average where the weight derives (endogenously) from a seemingly unrelated regression (SUR) approach. Hence, the prefix 'su' indicates this "Matrix" approach to letting the data suggest how to weight the various dimensions. In short, this approach uses only dimensions derived from direct (D) indicators, estimates a model that predicts those D dimension values using a wide array of indirect (not D) variables, and then creates an index value that is the weighted average of the predicted dimension value (thereby reducing the influence of outliers and using 'expected' rather than actual dimension values for each country) where the weights are proportional to the goodness-of-fit measures ( $R^2$ 's) from the predictive models – therefore weighting the dimension in direct proportion to our ability to explain its variance using the theorized variables at hand. In practice, with this data, the weights are not generally wildly different. Moreover, the SUR approach proved unnecessary as the independence of the equations cannot be rejected and separate regressions can suffice.

Armed with these many candidate indices (and sub-indices), thorough analyses were performed to test for robustness to difference in aggregation rules, sample size, and measure type. Special attention was paid to the correlations in rank-orderings derived from each index, more than just correlations in raw values. Overall, a good deal of stability is found across aggregation rules, although the geometric mean rankings were the least correlated with the other common types. The ‘su’ (endogenous weights) and the ‘am’ (equal weights) models were regularly quite similar.

Based on this analysis, and in the interest of paring down the resulting indices into more manageable results, we would recommend three types for the A and the P indices: the arithmetic mean, the geometric mean, and the weighted mean {amABL, gmARL, suamADL, amABS, gmARS, suamADS}. All three are closely and very significantly correlated. The ‘am’ model is correlated with the other two at over 0.79 for both the A and the P index. Moreover, its rank-ordering is correlated with the ‘gm’ P index at 0.87 and the ‘gm’ A index at 0.67. All in all, these are closely related index values but the rank orders do still differ noticeably. The ‘am’ indices have the advantage of spanning most countries on the globe (N>132), while the theoretical rigor of the ‘gm’ index is counterbalanced by its only covering half as many countries (N = 51). The ‘suam’ weighted index serves as a compromise in coverage. For the ‘am’ and ‘gm’ indices, the L indices have about twice as many observations as the S indices.

*Final Comments.* Finally, yet another index is proposed to accomplish two intriguing tasks. The “efficiency” index – based on a stochastic frontier analysis (inspired by the Phoenix Center’s “broadband efficiency index”) – serves a first purpose in using a country’s background attributes or “endowments” to explain its OSPI or OSAI score. This offers us a first inductive glimpse into the kinds of things that appear to determine a country’s index score. (This is a strange advantage, given that we designed the index from scratch and know full well all of its determinants. Nonetheless, by explaining the variation in the OSPI values using only variables not also included in the index itself, we can see what other attributes tend to correlate with high OSPI values. This permits hypothesis testing about various relationships between high OSPIs and, say, income or trade policy or education levels.) Initial results do point to some fascinating relationships, but detailed discussion goes beyond the scope of this project. The second task that this final analysis accomplishes is that it produces a measure of “technical efficiency” for a country. Given its background conditions or endowments, some countries do better or worse at achieving their OSPI. The efficiency index measures this efficiency at converting endowments into high index values. Future research can readily investigate why some countries are more efficient than others (perhaps due to policy? culture? economics?). The efficiency index is available for any index value imaginable. Sensitivity checks suggest that the ‘gm’ indices are the most stable at this preliminary point. Hence, the ‘gmARLeff’ and ‘gmPRLeff’ indices can usefully be added to the collection of indices.

If anyone is interested in constructing their own index, using weights of their own devising, a simple aggregating formula ( $wmABL = w_G \times amAGBL + w_F \times amAFBL + w_C \times amACBL$ ) can accomplish this readily in any spreadsheet program.

#### 4.0 Future Directions for the OSPI

The OSPI should be considered a work in progress, and was devised as a useful tool to explore potential impacts of open source software and approaches at a country level. However the true value of the tool lies in the use to which it can be put, and by extension in crafting policy and strategies for the advancement of open source interests, more broadly. With reference to open source software, a number of gaps appear along several dimensions, in the literature, in communication between different actors, as well as in the general awareness of what the OSS model represents. This lack of awareness, was articulated recently by an industry observer (Asay, 2008c) who noted that the assumption that “everybody knows about this stuff” could be a key barrier to the development of the open source approach, where passing familiarity with a concept/product leads to underestimation of its value, and the consequential free-riding could ultimately undermine the community and the desire to continue to develop in an open source approach. Another gap, evident in the literature, appears between the proscriptive, business-oriented/computer-oriented literature on open source software and the scholarly literature on the same subject. In developing the OSPI, we relied on a survey of the scholarly literature, as well as the input from Red hat and other open source experts, but a word

is in order about the business and computer literature not used. Trade publications and even mainstream periodicals such as *The Economist* have dedicated a great deal of attention to open source software, but such coverage typically based on anecdotal or tightly focused, short-term data to make broad assessments about open source. More often than not, such observations are used to speculate on the direction of open source, either positive or critical, depending on the viewpoint of the publication in question.

The empirical scholarly literature tended to use more rigorous methods to gather very specific data. One such example is Bonaccorsi, Giannangeli, and Rossi's (2006) survey of 146 Italian software firms to better understand OSS business models. While such methods tell the researchers a great deal about a specific topic, such findings are difficult to generalize from. Conversely, as mentioned, the generalized observations of the business and computer literature often lack needed data. What sets the GT/Red Hat OSPI apart is its model that uses comprehensive global data to provide some broad observations. Still, there is a need for better data documenting open source software, especially at the national level.

Turning to policy considerations, government commissions and agencies have proposed, and in some cases implemented, a variety of measures to encourage open source developers. For example, in the United States, the President's Information Technology Advisory Committee (2000) recommended direct federal subsidies for open source projects to advance high-end computing, and a report from the European Commission (2001) also discussed support for open developers and standards. Many European governments have policies to encourage the use and purchase of open source software for government use. And, as is well known, governments can sponsor the development of localized open source projects. Economists have sought to understand the consequences of a vibrant open source sector for social welfare. Perhaps not surprisingly, definitive or sweeping answers have been difficult to come by. But if a tentative conclusion can be made, most analyses have concluded, based on limited data, that government support for open source projects is likely to have an ambiguous effect on social welfare.

The GT/Red Hat OSPI uses indicators that allow one to make some broad observations about a given nation's potential to adopt open source software. In many cases we relied on general economic and socio-demographic indicators, but where possible, we drew upon variables and associated data based on software and computing central to the focus of the OSPI. While we believe strongly that, as an index, the OSPI provides a good "snapshot" of a country's open source potential it is worth noting that with better data collection efforts—beyond the scope of the current project—the index could be improved in subsequent iterations.

But this should not be the end product of research in this area, we believe that there is more to be found out that would reorient (or inform in a different direction), our findings about public policy and OSS. In other words, the assumptions about OSS's liberating nature and positive implications for social welfare (made often by governments themselves) have not necessarily been observable when the (admitted theoretical) research is done. We suggest that it is not necessary to accept such a nuanced and ambiguous view but propose that empirical research be supported that yields objective, generalizable observations.

## WORKING REFERENCES/BACKGROUND

Note this is a shortened version that augments the previous annotated bibliography.

Anderson, Paul. 2006. "Crossing the Chasm: open source software comes of age," A report from the OSS Watch conference, Open Source and Sustainability, held at the Said Business School, Oxford, 10-12th April 2006. [<http://www.oss-watch.ac.uk/resources/sustainability06.xml>]

Asay, Matthew. 2008a. "Open source doesn't make Gartner's "top-ten" list," *CNET: The Open Road Blog*, May 31, 2008. [[http://news.cnet.com/8301-13505\\_3-9956573-16.html](http://news.cnet.com/8301-13505_3-9956573-16.html)]

Asay, Matthew. 2008b. "Should governments legislate open source?," *CNET: The Open Road Blog*, June 6, 2008. [[http://news.cnet.com/8301-13505\\_3-9961828-16.html](http://news.cnet.com/8301-13505_3-9961828-16.html)]

Asay, Matthew. 2008c. "Is ignorance open source's biggest enemy?," *CNET: The Open Road Blog*, June 15, 2008. [[http://news.cnet.com/8301-13505\\_3-9968932-16.html](http://news.cnet.com/8301-13505_3-9968932-16.html)]

Asay, Matthew N. 2006. "Open Source and the Commodity Urge: Disruptive Models for a Disruptive Development Process." In DiBona, Chris, et al., eds. *Open Source 2.0: The Continuing Evolution*. (Sebastapol, CA: O'Reilly Media Inc.) pp.103-120.

Bergquist, Magnus and Jan Ljungberg. 2001. "The power of gifts: organizing social relationships in open source communities." *Info Systems Journal* Volume 11, pp305-320.

Bonaccorsi, Andrea and Cristina Rossi. 2006. "Comparing Motivations of Individual Programmers and Firms to Take Part in the Open Source Movement: From Community to Business," *Knowledge, Technology & Policy* 18, no. 4 (Winter 2006): 40-64.

Boulanger, Alan. Open-source versus proprietary software: is one more reliable and secure than the other? *IBM Systems Journal* July 2005.

Camara, Gilberto and Fonseca, Frederico. 2007. "Information Policies and Open Source Software in Developing Countries," *Journal of the American Society for Information Science and Technology*. 58, no.1: 121-132.

Chae, Bongsug (Kevin) and McHaney, Roger. 2006. "Asian Trio's Adoption of Linux-Based Open Source Development," *Communications of the ACM* 49, no. 9 (September): 95-99.

Comino, Stefano and Fabio M. Manenti. 2005. "Government Policies Supporting Open Source Software for the Mass Market," *Review of Industrial Organization* 26:217-240

Dahlander, Linus and McKelvey, Maureen. 2005. "Who Is Not Developing Open Source Software? Non-Users, Users, and Developers," *Economics of Innovation and New Technology* 14, no. 7, pp. 617-635.

Dalle, Jean-Michel and David, Paul. 2007. Allocation of Software Development Resources in Open Source Production Mode." In Feller, Joseph, et al., eds. *Perspectives on Free and Open Source Software*. (Cambridge, MIT Press: 2007) pp.297-328.

Dedrick, Jason and West, Joel. 2003. "Adoption of Open Source Platforms: An Exploratory Study". HBS - MIT Sloan Free/Open Source Software Conference: New Models of Software Development.

DiBona, Chris, Danese Cooper, and Mark Stone, (eds). 2006. *Open Source 2.0: The Continuing Evolution*. Sebastapol, CA: O'Reilly Media Inc.

Feller, Joseph, Fitzgerald, Brian, Hissam, Scott A., and Karim R. Lakhani (eds). 2007. *Perspectives on Free and Open Source Software*. Cambridge, MIT Press.  
Available online at: [ <http://mitpress.mit.edu/books/chapters/0262562278>].

Fitzgerald, Brian. 2006. "The Transformation of Open Source Software." *MIS Quarterly*, Vol. 30, No. 3

Fitzgerald, Brian 2007. "Has Open Source Software a Future." In Feller, Joseph, et al., (eds). *Perspectives on Free and Open Source Software*. (Cambridge: MIT Press) pp.93-106.

Glass, Robert L. 2007. " "Standing in Front of the Open Source Steamroller." In Feller, Joseph, et al., (eds.) *Perspectives on Free and Open Source Software*. (Cambridge: MIT Press) pp.81-92.

Gallego, Dolores M., Luna, P., and Bueno, S. 2007. "Designing a forecasting analysis to understand the diffusion of open source software in the year 2010." *Technological Forecasting and Social Change*.

Goode, Sigi. 2005. "Something for nothing: management rejection of open source software in Australia's top firms." *Information & Management* (42) 669-681.

Hawkings, Richard E. 2004. "The economics of open source software for a competitive firm, Why give it away for free?" *Netnomics* 6:103-117.

Holck, Jesper et. al. 2004. "Identifying Business Barriers and Enablers for the Adoption of Open Source Software." Working paper, Copenhagen Business School, Department of Informatics.

"Innovation and Its Enemies," *The Economist*, 12 January 2006.

Krishnamurthy, Sandeep. 2002 "Cave or Community? An Empirical Examination of 100 Mature Open Source Projects," May 2002, [<http://opensource.mit.edu/papers/krishnamurthy.pdf>]

Krishnamurthy, Sandeep. 2007 "An Analysis of Open Source Business Models," in Feller, Joseph, et al., (eds.) *Perspectives on Free and Open Source Software*. (Cambridge, MIT Press).

Lakhani, Karim R. and Robert G. Wolf. 2005. "Why Do Hackers Do What They Do?: Understanding Motivation and Effort in Free/Open Source Software Projects, in Feller, Joseph, et al., (eds.) *Perspectives on Free and Open Source Software*. (Cambridge, MIT Press).

Lakhani, Karim R. and Hippel, Eric Von. 2003. "How Open Source Software works: free user-to-user assistance." *Journal of Research Policy*. 32(6). pp. 923-943.

Lerner, Josh and Jean Tirole. 2007. Economic Perspectives on Open Space." In Feller, Joseph, et al., (eds.) *Perspectives on Free and Open Source Software*. (Cambridge, MIT Press) pp.47-78.

Madey, Greg, Vincent Freeh, and Renee Tynan. 2002. "Understanding OSS as a Self-Organizing Process" [<http://www.nd.edu/~oss/Papers/MadeyFreehTynan.pdf>]

May, Christopher 2006. "Escaping the TRIPs Trap: The Political Economy of Free and Open Source Software in Africa." *Political Studies* 54, pp. 123-146.

Open Source Projects (website) <http://osprojects.info/233/open-source-projects>

- Perrin, Chad. 2006. "Security through visibility: The secrets of open source security". *TechRepublic* [<http://articles.techrepublic.com.com/5100-10877-6064734.html>.]
- Rajani, Niranjana and Rekola, Juha and Mielonen, Timo. 2003. "Free as in Education: Significance of the Free/Libre and Open Source Software for Developing Countries". *World Summit on the Information Society*. [[http://www.itu.int/wsis/docs/background/themes/access/free\\_as\\_in\\_education\\_niranjana.pdf](http://www.itu.int/wsis/docs/background/themes/access/free_as_in_education_niranjana.pdf)]
- Raymond, Eric S. 1999. *The Cathedral & the Bazaar*. O'Reilly. ISBN 1-56592-724-9.
- Rosen, Lawrence. 2005. *Open Source Licensing: Software Freedom and Intellectual Property Law*. Upper Saddle River, NJ: Prentice Hall.
- Schiff, Aaron. 2002. "The Economic of Open Source Software: A Survey of the Early Literature." *Review of Network Economics* 1, no. 1, pp. 66-74.
- Schmidbauer, Harald et al. 2007. "Free/Open Source Software Adoption, Public Policies and Development Indicators: An International Comparison", Draft paper, submitted to "The Third Int'l Conference on Open Source Systems" [<http://cs.bilgi.edu.tr/~mgencer/pub/OSS2007.pdf>]
- Vaisman, Daria. 2007. "Coding a Revolution." *Foreign Policy* 159, p. 93.
- Valimäki, Mikko. "Dual Licensing in Open Source Software Industry". *Systemes d'Information et Management* January. 2003. Helsinki Institute for Information Technology.
- Van Rooij, Shahrone. 2007. "Perceptions of Open Source Versus Commercial Software: Is Higher Education Still on the Fence?" *Journal of Research of Technology in Education* 39, no. 4 (Summer): 433-453.
- von Hippel, Eric 2007. "Open Source Software Projects as User Innovation Networks." in Feller, Joseph, et al., (eds.) *Perspectives on Free and Open Source Software*. (Cambridge: MIT Press).
- Walli, Stephen R. 2006. "Under the Hood: Open Source and Open Standards Business Models in Context." In DiBona, Chris, et al., eds. *Open Source 2.0: The Continuing Evolution*. Sebastopol, CA: O'Reilly Media pp.121-136.
- Waring, Teresa and Maddocks, PhWeber, Steven. 2000. "The Political Economy of Open Source Software" BRIE Working Paper 140, E-economy Project Working Paper 15, June 2000.
- West, J. 2006. "The Economic Realities of Open Standards: Black, White and Many Shades of Gray." In Greenstein, S. and Stango, V.(eds.) *Standards and Public Policy*, Cambridge Univ. Press: Cambridge.
- West, Joel and Jason Dedrick, "Scope and Timing of Deployment: Moderators of Organizational Adoption of the Linux Server Platform," *International Journal of IT Standards and Standardization Research* 4, no. 2 (July-December): 1-21.
- Xiaobai, Shen 2005. "Developing Country Perspectives of Software: Intellectual Property and Open Source - A Case Study of Microsoft and Linux in China." *International Journal of IT Standards & Standardization Research* 3, no. 1, pp. 21-43.

## APPENDIX I

### Coding Sheet: References Providing Rationales for Indicators

- 1) Alonso, Andoni, Luis Casas, Carlos Castro, and Fernando Solis, "Research, Development, and Innovation in Extremadura: A GNU/Linux Case Study," *Philosophy Today* 48 (2004): 16-22.
- 2) Berry, David and Giles Moss, "Free and Open-Source Software: Opening and Democratizing E-Government's Black Box," *Information Polity: The International Journal of Government and Democracy in the Information Age* 11, no. 1 (2006): 21-34.
- 3) Bonaccorsi, Andrea and Cristina Rossi, "Comparing Motivations of Individual Programmers and Firms to Take Part in the Open Source Movement: From Community to Business," *Knowledge, Technology & Policy* 18, no. 4 (Winter 2006): 40-64.
- 4) Bonaccorsi, Andrea, Silvia Giannangeli, and Cristina Rossi, "Entry Strategies Under Competing Strategies: Hybrid Business Models in the Open Source Industry," *Management Science* 52, no. 7 (July 2006): 1085-1093.
- 5) Chae, Bongsug (Kevin) and Roger McHaney, "Asian Trio's Adoption of Linux-Based Open Source Development," *Communications of the ACM* 49, no. 9 (September 2006): 95-99.
- 6) Chan, Anita. "Coding Free Software, Coding Free States: Free Software Legislation and the Politics of Code in Peru," *Anthropological Quarterly* 77, no. 3 (Summer 2004):
- 7) Chumney, Wade M., and Zehai Zhou, "Legal and Business Perspectives of Open Source Education Software," *Journal of American Academy of Business* 13, no. 1 (2008): 208-214.
- 8) Coleman, Gabriella. "The Political Agnosticism of Free and Open Source Software and the Inadvertent Politics of Contrast." *Anthropological Quarterly* 77, no. 3 (Summer 2004): 507-519.
- 9) Comino, Stefano, and Fabio M. Manenti, "Government Policies Supporting Open Source Software for the Mass Market," *Review of Industrial Organization* 26, no. 2 (March 2005): 217-240.
- 10) Dahlander, Linus, and Maureen McKelvey, "Who Is Not Developing Open Source Software? Non-users, Users, and Developers," *Economics of Innovation and New Technology* 14, no. 7 (October 2005): 617-635.
- 11) Duenas, Jose *et al.* "Apache and Eclipse: Comparing Open Source Project Incubators," *IEEE Software* 24, no. 6 (November/December 2007): 90-99.
- 12) Ebert, Christof, "Open Source Drives Innovation," *IEEE Software* 27, no. 3 (May/June 2007): 105-109.
- 13) Forge, Simon, "The Rain Forest and the Rock Garden: The Economic Impacts of Open Source Software," *Info: The Journal of Policy, Regulation and Strategy for Telecommunications, Information, and Media* 8, no. 3 (2006): 12-31.
- 14) Gallaway, Terrel, and Douglas Kinnear, "Open Source Software, The Wrongs of Copyright, and the Rise of Technology," *Journal of Economic Issues*, 38, no. 2 (June 2004): 467-74.
- 15) Gaudeul, Alex. "Do Open Source Developers Respond to Competition? The (LA)TEX Case Study," *Review of Network Economics* 6, no. 2 (June 2007): 239-263.



- 16) Gehring, Robert, "The Institutionalization of Open Source," *Poiesis & Praxis* 4, no. 1 (March 2006): 54-73.
- 17) Gosain, Sanjay, "Looking Through a Window on Open Source Culture: Lessons for Community Infrastructure Design," *Systemes d'Information et Management* 8, no. 1 (2003): 11-42.
- 18) Gruen, Nicholas, "Geeks Bearing Gifts: Open Source Software and Its Enemies," *Policy* 21, no. 2 (Winter 2005): 39-44.
- 19) Hawkins, Richard E. "The Economics of Open Source for a Competitive Firm: Why Give It Away for Free?" *Netnomics: Economic Research and Electronic Networking* 6, no. 2 (August 2004): 103-117.
- 20) Krishnamurthy, Sandeep, "On the Intrinsic and Extrinsic Motivation of Free/Libre/Open Source (FLOSS) Developers," *Knowledge, Technology & Policy* 18, no. 4 (Winter 2006): 17-39.
- 21) Lee, Jyh-An, "Government Policy toward Open Source Software: The Puzzles of Neutrality and Competition," *Knowledge, Technology & Policy* 18, no. 4 (Winter 2006): 113-141.
- 22) Lerner, Josh, and Jean Tirole, "The Scope of Open Source Licensing," *Journal of Law Economics & Organization* 21, no. 1 (April 2005): 20-56
- 23) Lerner, Josh, and Jean Tirole, "The Economics of Technology Sharing: Open Source and Beyond," *Journal of Economic Perspectives* 19, no. 2 (Spring 2005): 99-120.
- 24) Lin, Yuwei, "Hybrid Innovation: The Dynamics of Collaboration between the FLOSS Community and Corporations," *Knowledge, Technology & Policy* 18, no. 4 (Winter 2006): 86-100.
- 25) May, Christopher, "Escaping the TRIPs' Trap: The Political Economy of Free and Open Source Software in Africa," *Political Studies* 54, no. 1 (March 2006): 123-146.
- 26) McGhee, Douglas, "Free and Open Source Licenses: Benefits, Risks, and Steps toward Ensuring Compliance," *Intellectual Property & Technology Law Journal* 19, no. 11 (November 2007): 5-10.
- 27) McGowan, Matthew K., Paul Stephens, and Dexter Gruber, "An Exploration of the Ideologies of Software Intellectual Property: The Impact on Ethical Decision Making," *Journal of Business Ethics* 73, no. 4 (July 2007): 409-424.
- 28) Michler, Carla, "The Procurement Decision – 'Open' or 'Closed' Source Software?" *Deakin Law Review* 10, no. 1 (2005): 261-269.
- 29) Miralles, Francesc, Sandra Sieber, and Josep Valor, "An Exploratory Framework for Assessing Open Source Software Adoption," *Systemes d'Information et Management* 11, no. 1 (March 2006): 85-105.
- 30) Samuelson, Paula, "IBM's Pragmatic Embrace of Open Source," *Communications of the ACM* 49, no. 10 (October 2006): 21-25.
- 31) Seiferth, C. Justin, "Open Source and These United States," *Knowledge, Technology & Policy* 12, no. 3 (Fall 1999): 50-79.
- 32) Simon, K. D. "The Value of Open Standards and Open-Source Software in Government Environments," *IBM Systems Journal* 44, no. 2 (2005): 227-238.

- 33) Vaisman, Daria, "Coding a Revolution," *Foreign Policy* 159 (March/April 2007): 93.
- 34) Van Rooij, Shahron. "Perceptions of Open Source Versus Commercial Software: Is Higher Education Still on the Fence?" *Journal of Research of Technology in Education* 39, no. 4 (Summer 2007): 433-453.
- 35) Van Wendel de Joode, Ruben, Yuwei Lin, and Shay David, "Rethinking Free, Libre, and Open Source Software," *Knowledge, Technology & Policy* 18, no. 4 (Winter 2006): 5-16.
- 36) Waters, John, "Opening a New Door," *T H E Journal* 34, no. 8 (August 2007): 30-35.
- 37) West, Joel, and Jason Dedrick, "Open Source Standardization: The Rise of Linux in the Network Era," *Knowledge, Technology & Policy* 14, no. 2 (Summer 2001): 88-112.
- 38) West, Joel and Jason Dedrick, "Scope and Timing of Deployment: Moderators of Organizational Adoption of the Linux Server Platform," *International Journal of IT Standards and Standardization Research* 4, no. 2 (July-December 2006): 1-21.
- 39) Xiaobai, Shen, "Developing Country Perspectives of Software: Intellectual Property and Open Source—A Case Study of Microsoft and Linux in China," *International Journal of IT Standards & Standardization Research* 3, no. 1 (January-June 2005): 21-43.
- 40) Asay, Matthew N. 2006. "Open Source and the Commodity Urge: Disruptive Models for a Disruptive Development Process." In DiBona, Chris, et al., eds. *Open Source 2.0: The Continuing Evolution*. (Sebastapol, CA: O'Reilly Media Inc.) pp.103-120.
- 41) Bergquist, Magnus and Jan Ljungberg, 2001. "The power of gifts: organizing social relationships in open source communities." *Info Systems J*(2001), Volume 11, pp305-320
- 42) Boulanger, Alan. Open-source versus proprietary software: is one more reliable and secure than the other?. *IBM Systems Journal* July 2005.
- 43) Camara, Gilberto and Fonseca, Frederico 2007. "Information Policies and Open Source Software in Developing Countries," *Journal of the American Society for Information Science and Technolgy*. 58, no.1: 121-132.
- 44) Comino, S., F.M. Manenti and M.L. Parisi (2005), "From Planning to Mature: on the Determinants of Open Source Take Off," Discussion paper, No. 2005/17, Università degli Studi di Trento. [http://www-econo.economia.unitn.it/new/pubblicazioni/papers/17\\_05\\_comino.pdf](http://www-econo.economia.unitn.it/new/pubblicazioni/papers/17_05_comino.pdf)
- 45) Dalle, Jean-Michel and David, Paul. 2007. "Allocation of Software Development Resources in Open Source Production Mode." In Feller, Joseph, et al., eds. *Perspectives on Free and Open Source Software*. (Cambridge, MIT Press: 2007) pp.297-328.
- 46) DiBona, Chris, Danese Cooper, and Mark Stone, eds. 2006. *Open Source 2.0: The Continuing Evolution*. (Sebastapol, CA: O'Reilly Media Inc.)
- 47) Feller, Joseph, et al., eds. *Perspectives on Free and Open Source Software*. (Cambridge, MIT Press: 2007) available at: <http://mitpress.mit.edu/books/chapters/0262562278>

- 48) Fitzgerald, Brian, 2006. "The Transformation of Open Source Software." *MIS Quarterly*, Vol. 30, No. 3, 2006
- 49) Fitzgerald, Brian 2007. "Has Open Source Software a Future." In Feller, Joseph, et al., eds. *Perspectives on Free and Open Source Software*. (Cambridge, MIT Press: 2007) pp.93-106.
- 50) Glass, Robert L. 2007. "Standing in Front of the Open Source Steamroller." In Feller, Joseph, et al., eds. *Perspectives on Free and Open Source Software*. (Cambridge, MIT Press: 2007) pp.81-92.
- 51) Gallego, Dolores M., Luna, P., and Bueno, S. 2007. "Designing a forecasting analysis to understand the diffusion of open source software in the year 2010." *Technological Forecasting and Social Change* article in print.
- 52) Goode, Sigi. 2005. "Something for nothing: management rejection of open source software in Australia's top firms." *Information & Management* 42 (2005) 669-681.
- 53) Hawkings, Richard E. 2004. "The economics of open source software for a competitive firm, Why give it away for free?." *Netnomics* 6:103-117, 2004.
- 54) Holck, Jesper et. al. 2004. "Identifying Business Barriers and Enablers for the Adoption of Open Source Software." Working paper, Copenhagen Business School, Department of Informatics.
- 55) Krishnamurthy, Sandeep. 2002 "Cave or Community? An Empirical Examination of 100 Mature Open Source Projects," May 2002, available online at: <http://opensource.mit.edu/papers/krishnamurthy.pdf>
- 56) Madey, Greg, Vincent Freeh, and Renee Tynan. 2002. "Understanding OSS as a Self-Organizing Process," available online at: <http://www.nd.edu/~oss/Papers/MadeyFreehTynan.pdf>
- 57) Rajani, Niranjan and Rekola, Juha and Mielonen, Timo (2003). "Free as in Education: Significance of the Free/Libre and Open Source Software for Developing Countries". World Summit on the Information Society. [http://www.itu.int/wsis/docs/background/themes/access/free\\_as\\_in\\_education\\_niranjan.pdf](http://www.itu.int/wsis/docs/background/themes/access/free_as_in_education_niranjan.pdf)
- 58) Raymond, Eric S. (1999). "The Cathedral & the Bazaar". O'Reilly. ISBN 1-56592-724-9.
- 59) Rosen, Lawrence 2005. *Open Source Licensing: Software Freedom and Intellectual Property Law*. Upper Saddle River River, NJ: Prentice Hall.
- 60) Schiff, Aaron 2002. "The Economic of Open Source Software: A Survey of the Early Literature." *Review of Network Economics* 1, no. 1, pp. 66-74.
- 61) Schmidbauer, Harald et al.2007. "Free/Open Source Software Adoption, Public Policies and Development Indicators: An International Comparison",Draft paper, intended for submission to "The Third International Conference on Open Source Systems"  
<http://cs.bilgi.edu.tr/~mgencer/pub/OSS2007.pdf>
- 62) Spinellis, Diomidis. 2008. "A Tale of Four Kernels." In Wilhem Schäfer, Matthew B. Dwyer, and Volker Gruhn, editors, *ICSE '08: Proceedings of the 30th International Conference on Software Engineering*. New York: Association for Computing Machinery, pp.381–390.
- 63) Waring, Teresa and Maddocks, Philip 2005. "Open Source Implementation in the UK Public Sector: Evidence from the Field and Implication for the future." *International Journal of Information Management* 25, pp. 411-428.

## APPENDIX II

### Coding Notes

Final OSPI design parameters

- use endogenous weighting approach (“Matrix” model)
- allow for sub-indices and data aggregation along dimension subcategories
- use active/potential dynamic
- test run, if possible, stochastic frontier analysis {‘frontier’ and ‘xtfrontier’ in Stata} as used in Broadband Index report

Dimensions: G = government; F = firms or commercial enterprises; C = community and educational system

Indices: Active =  $f(G_A, F_A, C_A)$  / Potential =  $f(G_P, F_P, C_P)$

$G_A = f(\text{procurement, policy, use})$

$G_P = f(\text{OSS policy, corruption and liberties, e-government, IP law})$

$F_A = f(\text{RHCE \& other developers, firms' installs/users, firms developing/supporting OSS})$

$F_P = f(\text{IT industry age, IT growth, R\&D, internet access, de novo growth})$

$C_A = f(\text{household installs/users, Wiki participants, OSS courses, adoption by educators, discussion in media})$

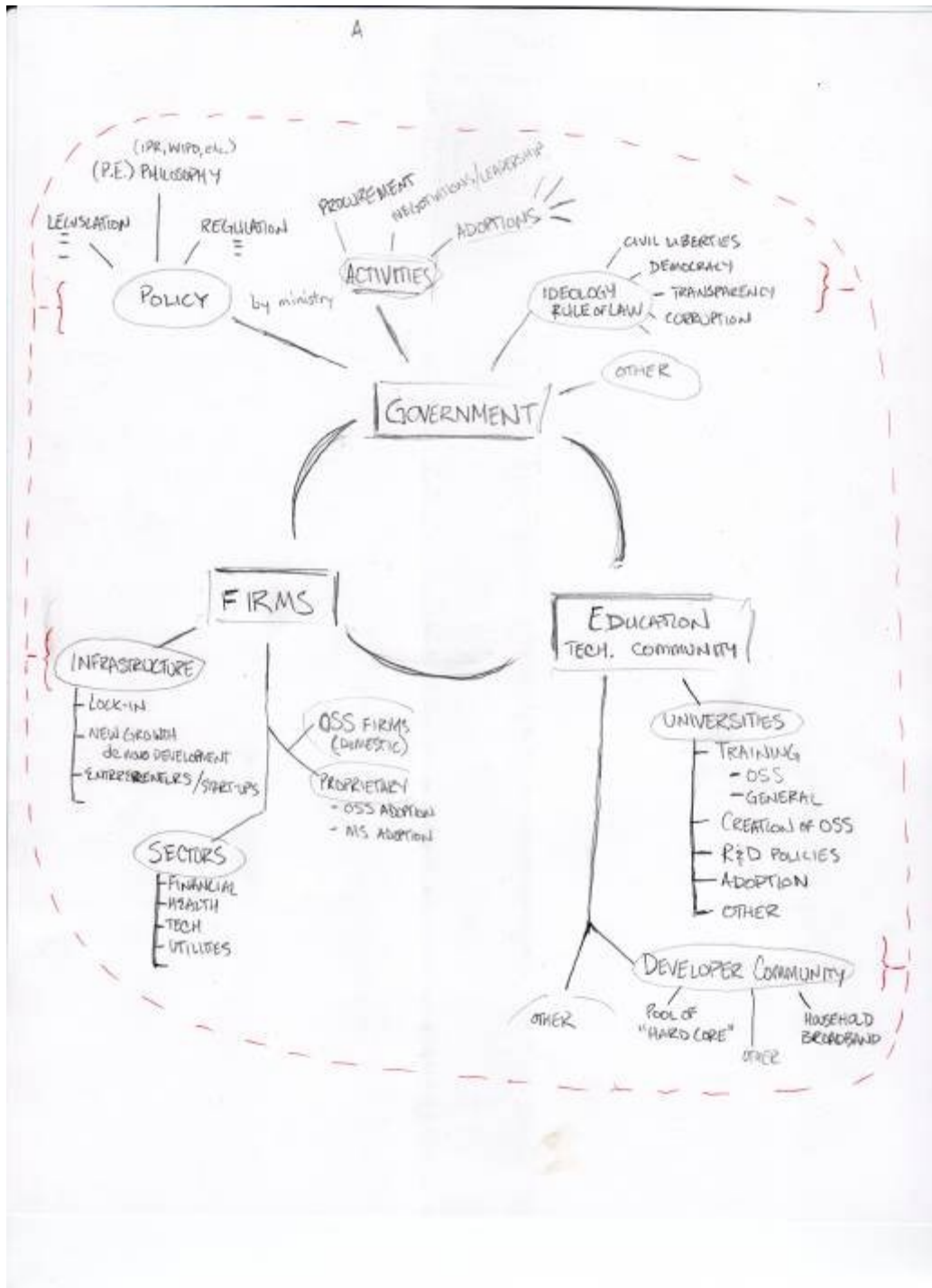
$C_P = f(\text{culture, education, CS majors, internet users})$

Outline of “do file” for Stata:

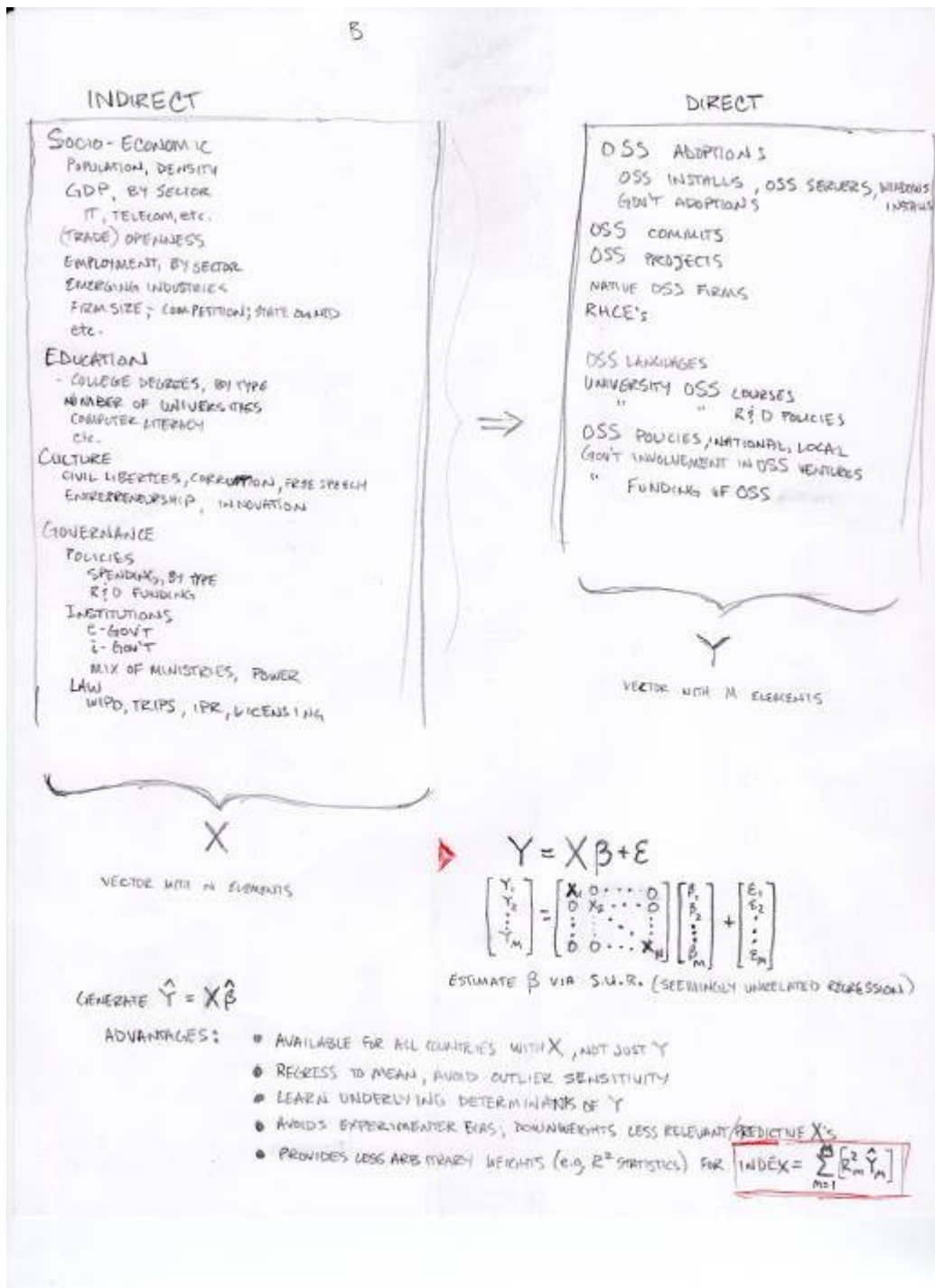
1. clean data
2. Loop1: Active, Potential
  - i. Loop2: Government, Firms, Community
    - a. Pick variables
    - b. Pick indicators
      - I. Transformation
      - II. imputation
      - III. aggregationaggregation methods:
      1. sum
      2. maximin
      3. min
      4. geometric mean
      5. SURE (Matrix Index)
    - c. end Loop2
3. end Loop1
4. comparison of:
  - i. long/short
  - ii. active/potential
  - iii. dimensions within and across active/potential
  - iv. dimension rankings by aggregation methods
5. run a ‘frontier’ analysis, using Active as the frontier and Potential as predictors. Rank countries based on their shortcomings.

**APPENDIX III**  
**Preliminary OSPI Models**

**“Trinity” [Index A] Concept Model**



# "Matrix" [Index B] Concept Model



# "Hybrid Neo" [Index C] Concept Model

