Firms' involvement in the projects of the OSS community
Some preliminary empirical evidence and a research agenda

RESEARCH TEAM (UP TO NOW)

Dario Lorenzi – DIG, Politecnico di Milano – dario.lorenzi@polimi.it

Cristina Rossi – DIG, Politecnico di Milano – cristina1.rossi@polimi.it

Rationales under the research

It has been acknowledged that **OSS challenges** the traditional way of doing business out of **software solutions** (Bonaccorsi et al., 2006)

- Increasing **economic important** (Ghosh et al., 2006)
- **Open innovation/knowledge paradigm:** knowledge is produced and shared by developers and users (West and Gallagher, 2006)
- Increasing role of **firms' involvement** in the movement
  - **Large incumbents**: IBM, Sun Microsystems, and, even, Microsoft
  - SMEs: entered the market to explore this new business opportunity

On the relationships between FIRMS AND THE OSS
- **Open Source Initiative** as the starting point (1998)
  - Addressing firms' *GPL fear* and showing that it is possible **to profit** from open standards
- The **economic literature**
  - Some **theoretical contributions**
  - **Few empirical studies**, mainly based on qualitative methodologies

Rationales under the research

Research team of OSS of Sant'Anna School

Gathering data on software **firms involved** in the **OSS movement**

☐ Sharp **quantitative approach**

☐ Survey methodology

- ▪ **ELISS I survey**: on 146 Italian firms providing to their customers OSS-based products and services (**Open Source firms**)
- ▪ **ELISS II survey**: on 918 software firms (NACE code 72.2) from 5 European countries (Finland, Germany, Italy, Spain, and Portugal)

MAIN EMPIRICAL FINDINGS FROM THE ELISS II SURVEY

It has emerged that **Open Source**

(vii) Is a **sustainable** business model, even without **appropriability**

▪ 1/3 respondents have chosen an OSS-based business model

(ii) Leading role of **hybrid business models**

- ▪ Mix of **OSS** and **proprietary offerings**
- ▪ POSS, MOSS; and LOSS firms

ELISS findings

Particularly, **OPEN SOURCE FIRMS**

(iii) Do **not** attach **much importance** to traditional IPRs

■ Patents do not foster innovation, in many cases they hamper it!

(vi) Rely on a large community and succeed in having a **wider product/service portfolios**

The **community**

■ Provides **complementary assets**

■ Bears the cost of producing the first copy of the software

■ Makes user-innovation

(xii) Are active in **more market segments**

□ Also in **less mature ones** (Antispam, antivirus,…)

(xv) Participate in **community projects**

The sample

Size distribution of sample firms across countries

COUNTRY	SIZE								Anova F-test p-value
	N	Min.	Max	Mean	Std. Dev.	p50	p75	p95	
Finland	134	1	640	31.03	77.41	7	18	200	
Germany	92	1	1 400	82.72	203.20	21	55	523	
Italy	243	1	140	5.87	12.46	3	6	18	0.000
Portugal	98	1	380	15.82	40.24	7	13	61	
Spain	200	1	1 300	66.80	169.26	17	42	302	
Total	767	1	1 400	36.64	120.40	8	20	160	

☐ The size (number of employees) distribution differs across countries and **reflects that of firms** in the software sector at **the national level**

☐ **German and Spanish firms** are, in general, **larger** than the others: only 28% of them hire less than 10 employees and 8% and 3%, respectively, hire more than 500 employees

☐ **Italian firms** are by far **the smallest** ones

The sample

Other firms' structural characteristics

		YEAR OF FOUNDATION						SKILLS[a]			MAIN CUSTOMERS						
COUNTRY	N	Min	Max	p50	p75	p95	Kruskal-Wallis test p-value	Mean	Std. Dev.	Anova F-test	SMEs	Large Firms	Universities	Public sector	End users	Others	Chi-square test p-value
Finland	135	1968	2004	1997	2001	2004		61.02	33.16		49.63	37.04	2.96	8.15	0.74	1.48	
Germany	93	1968	2004	1992	1998	2001		50.19	32.12		45.16	41.94	1.08	7.53	2.15	2.15	
Italy	243	1971	2004	1994	1998	2002	0.000	32.30	37.93	0.000	66.67	16.87	1.65	10.29	2.88	1.65	0.000
Portugal	98	1979	2004	1995	2000	2004		36.67	31.07		69.39	20.41	0.00	7.14	2.04	1.02	
Spain	200	1971	2004	1994	1999	2003		83.03	27.99		63.00	21.00	0.50	9.50	1.50	4.50	
Total	769	1968	2004	1994	1998	2003		53.26	38.86		60.47	24.97	1.30	8.97	1.95	2.34	

[a] *The variable SKILLS refers to the share of graduate personnel, data on skills are available only for a subset of respondents (N=682).*

▢**Respondents** are usually **young**

▢In all the countries but Italy and Portugal firms count **a considerable share** of **graduate** personnel

▢Respondents serve mainly **business customers** whereas very few work for universities or end users

Is OSS a sustainable business model?

□Information on **profitability** tend to be considered **confidential**. Therefore, we had to **tackle the topic** of sustainability **indirectly**

Distribution of the OSS turnover in 2000 and 2003

Share of turnover generated by OSS	Year 2000		Year 2003	
	N	%	N	%
0%	85	33.33	28	10.98
< 10%	77	30.2	72	28.24
10% - 30%	42	16.47	60	23.53
31% - 50%	7	2.75	30	11.76
51% - 70%	10	3.92	19	7.45
71% - 90%	10	3.92	17	6.67
91% and 99%	14	5.49	18	7.06
100%	10	3.92	11	4.31
TOTAL	255	100	255	100

□The **share of OSS turnover** out of the total **has increases** over time indicating **the sustainability** of the **OSS business model**

Firms' assessment of IPRs: patents

□The **importance attached** by firms to **patents** as an instrument for increasing revenues **is generally low**

Firms' assessment on patents

We think that patents…[a]	YES (%)				Fisher exact test p-value
	MOSS	LOSS	NOSS	TOTAL	
	N = 49	N = 142	N = 370	N = 561	
1 Promote innovation	4.08	27.46	37.57	32.09	0.000
2 Hamper innovation	73.47	52.82	34.86	42.78	0.000
3 Do not prevent our potential competitors to enter the market	71.35	75.35	67.35	71.70	0.581
4 Need a too long legal procedure	87.76	69.72	65.95	68.81	0.030
5 Are costly	81.63	76.06	70.00	72.55	0.166
6 Constraint versioning	61.22	50.70	35.41	41.53	0.001
7 Provide information about innovations and product development by other firms	18.37	28.87	25.68	25.85	0.055

[a] *Possible answers: YES, NO; MAYBE. Few firms chose MAY BE and this justifies the use of a Fisher exact test.*

□Patents **increase the cost of innovations** whereas the impact on expected revenues may be dubious

Offering profile: products

Firms' offering in the 18 products categories

ID	Classes[a]	Product category	NOSS N 514		LOSS N 166		MOSS N 70		POSS N 19		TOTAL N 769	
			N	%	N	%	N	%	N	%	N	%
1	S	Web servers	163	31.71	93	56.02	51	72.86	11	57.89	318	41.35
2	S	Other kinds of servers	153	29.77	73	43.98	47	67.14	11	57.89	284	36.93
3	N	Back up Systems	146	28.40	57	34.34	40	57.14	10	52.63	253	32.90
4	N	Firewall	135	26.26	73	43.98	39	55.71	8	42.11	255	33.16
5	N	Antispam	117	22.76	69	41.57	39	55.71	8	42.11	233	30.30
6	N	Antivirus	130	25.29	68	40.96	37	52.86	7	36.84	242	31.47
7	N	User and Identity Management	123	23.93	53	31.93	35	50.00	10	52.63	221	28.74
8	W	E-mail Client	133	25.88	63	37.95	40	57.14	10	52.63	246	31.99
9	W	Instant Messaging	82	15.95	43	25.90	28	40.00	11	57.89	164	21.33
10	W	Web Browser	37	7.20	38	22.89	22	31.43	3	15.79	100	13.00
11	W	Digital Signature Systems	41	7.98	28	16.87	13	18.57	6	31.58	88	11.44
12	W	Content Management System	127	24.71	73	43.98	45	64.29	12	63.16	257	33.42
13	W	E-commerce solutions	150	29.18	70	42.17	38	54.29	10	52.63	268	34.85
14	W	E-learning Tools	63	12.26	39	23.49	27	38.57	10	52.63	139	18.08
15	O	Management Software	333	64.79	84	50.60	38	54.29	13	68.42	468	60.86
16	O	Data Management Software	274	53.31	82	49.40	44	62.86	16	84.21	416	54.10
17	O	Workflow Systems	130	25.29	42	25.30	25	35.71	10	52.63	207	26.92
18	O	Office Automation Packages	140	27.24	67	40.36	34	48.57	7	36.84	248	32.25

[a] *Products have been grouped as follows. S: server products, N: network infrastructure products; W: Web products; O: Other kind of products*

Offering profile: services

Compared with NOSS, **OSS firms** have a *broader* **product portfolio**

☐ The number of **supplied products increases** with the **degree of openness**

Firms' offering in the 11 service categories

ID	Software related services	NOSS N 514		LOSS N 166		MOSS N 70		POSS N 19		TOTAL N 769	
		N	%	N	%	N	%	N	%	N	%
1	Consultancy	459	89.30	152	91.57	67	95.71	18	94.74	696	90.51
2	Integration	384	74.71	137	82.53	66	94.29	18	94.74	605	78.67
3	Installation	409	79.57	124	74.70	62	88.57	16	84.21	611	79.45
4	Assistance	439	85.41	133	80.12	67	95.71	17	89.47	656	85.31
5	Maintenance	431	83.85	135	81.33	66	94.29	15	78.95	647	84.14
6	System Management	242	47.08	83	50.00	47	67.14	11	57.89	383	49.80
7	Training	368	71.60	110	66.27	54	77.14	10	52.63	542	70.48
8	Application Management	232	45.14	67	40.36	40	57.14	10	52.63	349	45.38
9	Adapting codes written by others to suit customers' needs	219	42.61	99	59.64	54	77.14	17	89.47	389	50.59
10	On order software development from the scratch	356	69.26	129	77.71	62	88.57	14	73.68	561	72.95
11	Generating documentation	285	55.45	95	57.23	51	72.86	17	89.47	448	58.26

☐ This is made possible by the **exploitation** of the **open knowledge base** created by the community of developers

Offering profile: summary

NOSS are **active in mature segments**

☐ Mainly in **Office Automation**, **Management Software** & Databases

☐ MS Office is the leader in Office Automation (over 90% market share), SAP is the leader in Management & Data Management

OSS FIRMS

☐ Have entered **also less mature segments**

- ▪ As **Antivirus**, **Antispam**, **Firewall** & **Security Appliances**, **e-commerce Solutions**, **CMS** (data from IDC 2002, 2004, 2005..)

☐ Seem to be very active in segments in which it is acknowledge the existence of a **wide community of users/producers**

- ▪ In the absence of a large incumbent
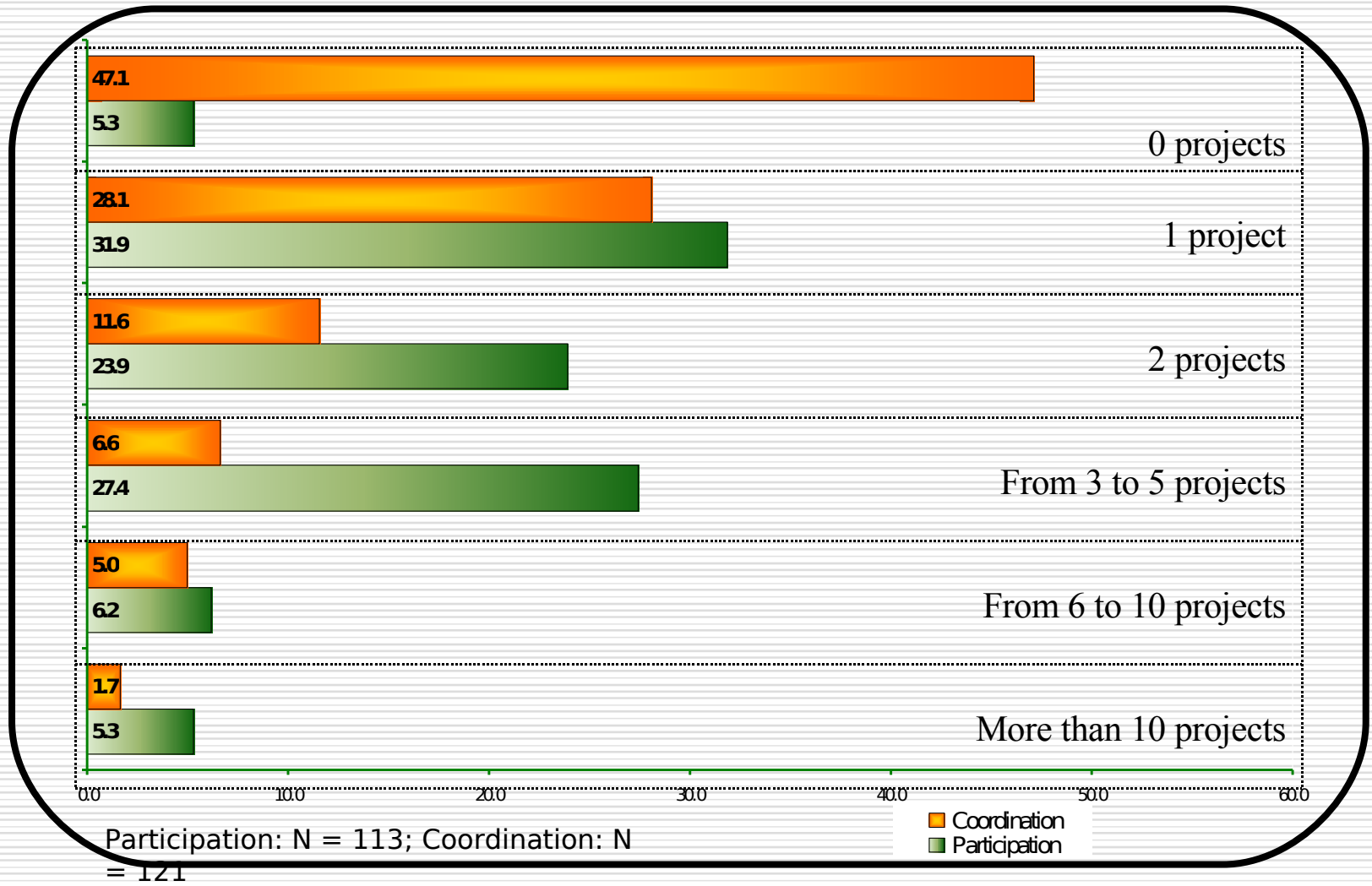- ▪ The **Apache Web server**

Starting point

From the ELISS II survey: **OSS firms contribute to the projects of the community**

☐ By **fixing bugs**, providing **mailing list assistance**, and, even, **writing code**

This **belies** the idea that they simply **exploit** the code of the OS developers

	PARTICIPATED		COORDINATED	
	From the start of OS	2003	From the start of OS	2003
	119	113	115	121
Min	1	0	0	0
Max	150	50	30	15
Mean	7.2	4.0	2.1	1.4
Standard deviation	17.7	8.1	4.2	2.6

OSS projects the firm has participated to/coordinated in 2003

Percentages

Category	Coordination	Participation
0 projects	47.1	5.3
1 project	28.1	31.9
2 projects	11.6	23.9
From 3 to 5 projects	6.6	27.4
From 6 to 10 projects	5.0	6.2
More than 10 projects	1.7	5.3

Participation: N = 113; Coordination: N = 121

■ Coordination
■ Participation

Question 25b

Contribution to OSS community

Percentages

N = 139

Contribution	Percentage
Fixing bugs	74.8
Contributing code that was accepted in projects' official versions	69.1
Writing documentation	61.6
Providing user assistance within the projects' mailing lists	56.6
Contributing code that was NOT accepted in projects' official versions	51.5

30 35 40 45 50 55 60 65 70 75 80
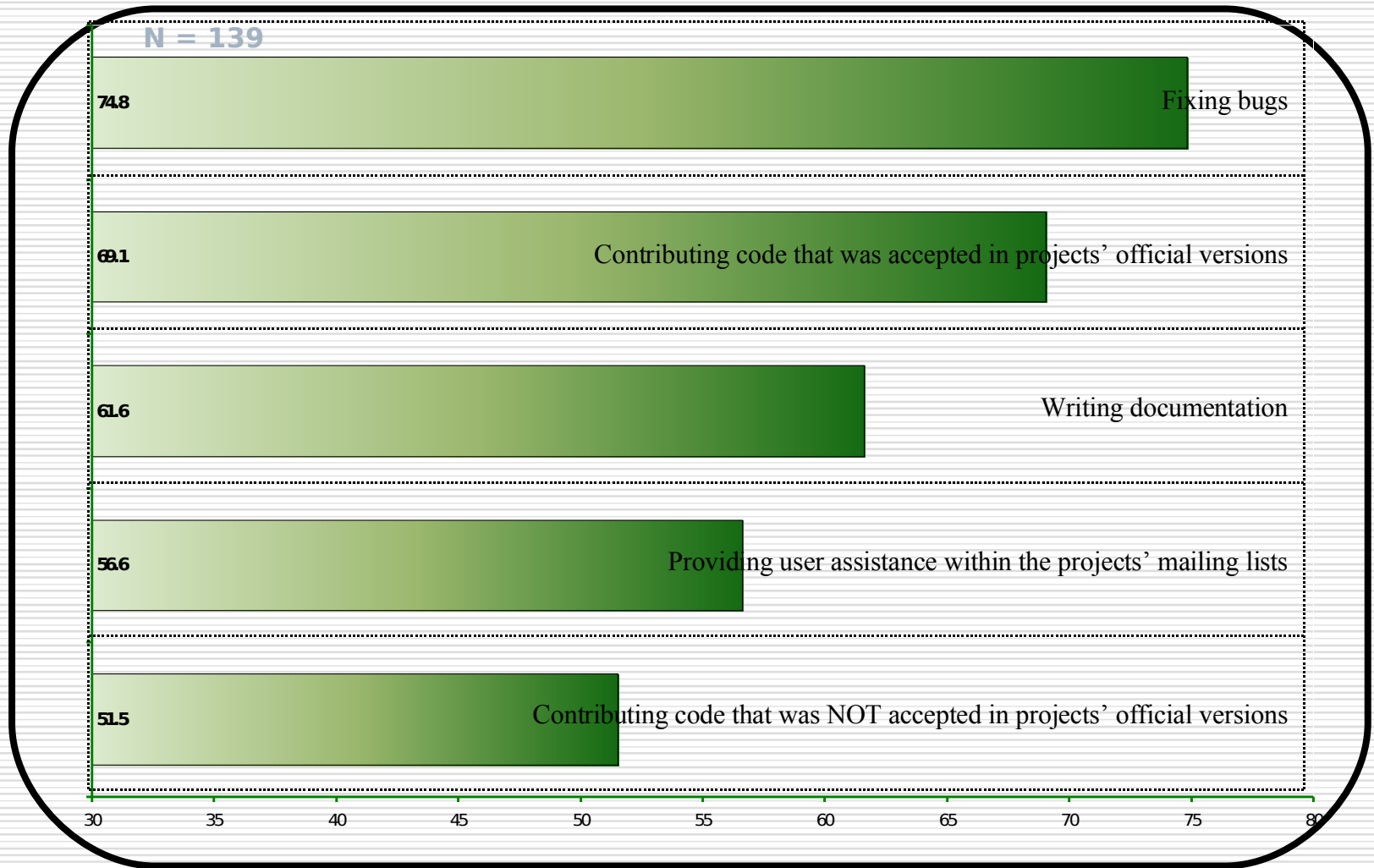
Challenging issue

DO FIRMS TAKE PART IN COLLECTIVE ACTION PROCESSES?

❑**Few investigations** on **whether and how** OSS firms directly feed open code basin by contributing their own developments back to the community

METHODOLOGICAL CONCERNS

❑Focus on **a single project** or on few firms (Henkel, 2006)

❑**Qualitative** methodologies (Dahlander and Magnusson, 2005; Lin, 2006)

❑**Shortcomings** of survey data (Bonaccorsi and Rossi, 2003)

   ❑No information on the projects to which respondents take part

   ❑Few information about firms' activities within the projects

   ❑The very concept of project **participation** is **surrounded** by **confusion**

      ❑ Over/Under/Estimation

      ❑ Small samples

Research questions

Within this framework, **four basic research questions** come to the limelight

1. Do firms act not only as *takers* but also as *givers* by directly contributing to OSS projects?

☐ **Methodology**: data collection from **SourceForge** and its linked Web sites

2. If yes, what do **firms do** within the projects?

☐ Do they only carry on ancillary works, or do they also provide code and undertake coordination activities?

3. Does the presence of firms shape **the evolution of the projects**?

☐ Are there significant differences between projects participated by firms and the others? Or after and before firms' entrance?

4. Do firms joining in OSS projects **innovate more** or in a **different manner?**

Basing on preliminary **empirical findings**, **a research agenda** for the future is proposed

Data and methodology

Data are collected using **SourceForge**: the largest OSS repository on the Internet

□Plenty of information about the hosted projects

□No risk of subjective interpretation of the information

Project selection: on the basis of the **level of activity**

□Ranking of projects' activity determined by the repository

□Selection of the 300 most active projects (Klincewicz, 2005)

Data collected from

□**SourceForge repository**

■Number of developers and administrators

■Date of registration on the repository

■Type of licence under which the code is released…

□**Web sites and other instruments** outside SourceForge and linked to it

■Information on companies' participation (Mainly through projects' mailing

Main results: projects' characteristics

A. PROJECT DIMENSIONS. The developing team is fairly narrow

- In line with other researches on the topic (Ghosh et al., 2002a, 2002b)
- The median number of programmers is 7, 15% of one man projects

B. LICENSES. The most widespread licence is GPL (Lerner and Tirole, 2005)

- GNU GPL: 57.91%
- LGPL: 12.84%
- BSD licence: 7.76%

C. TECHNICAL ASPECTS

- Widely use of the instruments put at the disposal of developers
- Java as the most widespread programming language
- Compatibility with the Windows operating systems
  - Evolution of the OS movement from its strong ideological origins
- Projects mainly targeted on advanced users or developers
  - High skills required

Main results: firms' participation

97 PROJECTS (32.33%) SHOW THE INVOLVEMENT OF FIRMS

A. COORDINATION: the **most frequent** form of participation (60 cases)
- Good relationships between firms and the OS community: the leadership emerges from the bottom up (O'Mahony, 2003)
    - **Foundation of the project**, provision of valuable code, bright solutions to critical technical problems (Bonaccorsi et al., 2006)
- Ways in which firms succeeded in achieving the leadership
    - Creating the project
    - Entering an existing project and replacing the coordinator
    - The firm was settled up by the members of the project coordinating group

B. COLLABORATION to development in different ways (bug fixing, testing or offering services, 37 cases)

C. PROVISION OF CODE (7 cases)

Main results:
Peculiarities of the projects participated by firms

Several **disparities** emerge between projects participated by firms and the others

□Firms' presence shapes project evolution?

□**Endogeneity** concerns

In general, projects participated by commercial firms

□Are **MORE ACTIVE**

- ▪**Are larger**: more developers and more coordinators
- ▪More intense **bug reporting** activity
- ▪Wider use of the **mailing lists**

□**MANAGE IPRS IN A DIFFERENT WAY**

- ▪Less GPLed projects (45.36% vs. 73.89 %)

□**PROVIDE DIFFERENT TYPOLOGY OF SOFTWARE**

- ▪More products targeted to companies
- ▪The average user seems to have higher computer science skills

□**SHOW SOME TECHNICAL DIFFERENCES**

- ▪**For instance**: use of different programming languages, with a wider presence of the Java language

Conclusions

Our empirical results reveal as, at present, the **OS movement** differs considerably from its origins

A. Increasing role of for profit firms

☐ **Almost 1/3** of the 300 sampled projects: some form of firms' involvement

B. Different types of links between these companies and the OS community

☐ Coordination, code offering, or provision of other kinds of contributions

C. Firms have an impact on the evolution of the projects

☐ Differences between the projects participated by firms and the others

- Are **more active** and **larger:** more developers, more coordinators, more debugging

- Make **less use** of the **GPL** licenses

- Show several **technical peculiarities**

- Solutions targeted mainly on **companies** and **high skill-users**

Research agenda

Our findings do not provide **definite conclusions** → **Research agenda**

1. A wider survey of literature on firms' participation in collective action

☐To disentangle the main aspects of the topic

☐**Interdisciplinary** approach

- ▪ Sociology and psychology have widely contributed to the understanding of PPCG

2. Clear research hypotheses should drive the empirical analyses

H1. Are projects in which firms involved more successful than the others?

Methodological problems

☐The concept of success is hard to define: integration of different metrics

☐Endogeneity concerns: is it firms' involvement to shape the success of a project or, viceversa?

☐Inferential procedures are not suitable to address these issues

- ▪ Definition of empirical models and application of appropriate econometric techniques

Research agenda

H2. What are the characteristics of the firms involved in OS projects?

☐ **Case studies** focusing on the **relationships** between **companies** and **projects**

☐ Data should be gathered on **structural characteristics** of these firms

- Age, size, competence
- Do they differ from those following the traditional model?
- How these difference are related to project participation?

H3. How project participation shapes the innovation activity?

☐ Are firms participating in OS projects more innovative than the others?

☐ *Open innovation model* (Chesbrough et al., 2006): firms can achieve a greater return on their innovation by using a broad range of knowledge sources

☐ OS is a clear example of open innovation (West and Gallangher, 2006)

- The OS community is a large knowledge basin
- OS licenses are designed to foster instead of forbidden the access to the information

Research agenda

METODOLOGY

c. **Enlarging** our sample from 300 to 1,000 projects

B. Data gathered from **different sources**: **underestimation** the phenomenon

☐ Firm making few contribution to the code: no cited in project Website
☐ Collecting information directly from the code posted on SourceForge
☐ **Software scanning applications** (CODD, Ghosh and Prakash, 2000)

C. A more rigorous methodology is needed

Firms' involvement in the projects of the OSS community
Some preliminary empirical evidence and a research agenda

RESEARCH TEAM (UP TO NOW)

Dario Lorenzi – DIG, Politecnico di Milano – dario.lorenzi@polimi.it

Cristina Rossi – DIG, Politecnico di Milano – cristina1.rossi@polimi.it