Reflections on Conducting an International Survey of Software Engineering

Reidar Conradi 1) 2)
Jingyue Li 1)
Odd Petter N. Slyngstad 1)
1) Department of Computer and Information Science
Norwegian University of Science and Technology
NO-7491 Trondheim, Norway
{conradi, jingyue, oslyngst}@idi.ntnu.no

Vigdis By Kampenes 2)
Christian Bunse 3)
2) Simula Research Lab
P.O. Box 134, NO-1325 Lysaker, Norway
3) Fraunhofer IESE
Sauerwiesen 6
D-67661 Kaiserslautern, Germany
bunse@iese.fraunhofer.de
vigdis@simula.no

Maurizio Morisio 4)
Marco Torchiano 4)
4) Dip. Automatica e Informatica
Politecnico di Torino
Corso Duca degli Abruzzi 24
I-10129 Torino, Italy
{morisio, torchiano}@polito.it

Abstract

Component-based software engineering (CBSE) with Commercial-Off-The-Shelf (COTS) or Open Source Software (OSS) components are more and more frequently being used in industrial software development. We therefore need to issue experience-based guidelines for the evaluation, selection and integration of such components.

We have performed a survey on industrial COTS/OSS development in three countries – Norway, Italy and Germany. Concrete survey results, e.g. on risk management policies and process tailoring, are not being described here, but in other papers.

This is a method paper, reporting on the challenges, approaches and experiences gained by conducting the main survey. The main contributions are as follows: At best we can achieve a stratified-random sample of ICT companies, followed by a convenience sample of relevant projects. This is probably the first software engineering survey using census type data, and has revealed that the entire sampling and contact process can be unexpectedly expensive. It is also hard to avoid national variations in the total process, possibly leading to uncontrollable method biases.

Keywords: *Surveys, component-based software engineering, COTS/OSS components, open source software, ICT industry.*

0-7803-9508-5/05/$20.00 ©2005 IEEE

1. Introduction

Component-based software engineering (CBSE) is considered one of the most important technologies to reduce time-to-market and cost and to improve software quality. Although there are many proposed CBSE methods in research papers and some books, few representative industrial studies have been done to study such methods in practice.

Although the survey topic is CBSE, the focus of this paper is to convey some reflections and lessons learned in performing a large-scale, multinational survey of software engineering in the Information and Communication Technology (ICT) industry. For a definition of "component", see [30]. Let us summarize the overall survey process:

1) We scanned the literature to formulate research hypotheses and questions. Sampling was planned.
2) NTNU and the Simula Research Lab in Oslo performed a *pre-study* on COTS (Commercial-Off-The-Shelf) development through on-site, structured interviews of 16 projects in 13 companies in autumn 2003 [18].
3) Based on the pre-study, we started preparing the *main study* in late 2003. This included questions on OSS components, and extended the survey to Germany and Italy. We designed a more *formal questionnaire* for this study in spring 2004, translated it into Norwegian, German, and Italian

and installed it on the SESE web server at the Simula Research Lab [28].

4) The main study was performed in May 2004 – March 2005. We got name lists of ca. 500 ICT companies in each country. For each company, we got hold of a possible respondent over the phone. This person was asked to identify a suitable *OTS-based* project (*Off-The-Shelf, being either COTS or OSS*), and encouraged to answer the questionnaire. Most respondents filled in the web-questionnaire him/herself, but some were filled in by phone interviews. Per 16 August 2005, 133 company responses have been recorded.
5) We are now analyzing and publishing the results [19]. The survey has taken much more effort than anticipated, so we need to *reflect* on the choices, e.g. on sampling, approaching the companies, filling in the questionnaire etc.

The three research partners (NTNU/Simula, IESE, Torino group) have been working on software reuse, CBSE, process improvement, and industrial studies for over 15 years. We have planned and conducted the entire survey with no physical meetings, just email contact and two-weekly phone meetings – even the web server has been accessed remotely.

The main conclusions of this paper are that we at best can obtain a stratified-random sample of ICT companies, and thereafter a convenience sample of relevant projects. The sampling and contact process can be exceedingly expensive, and national process variations are difficult to avoid. This in turn can lead to uncontrollable method biases.

This paper is a *method paper* to reflect on the main study. The remainder of the paper is structured as follows. Section 2 presents some related work. Section 3 discusses definition of population, and Section 4 how to make a sample from this population. Section 5 discusses questionnaire design, and Section 6 discusses the company contact process. Section 7 has lessons learned, and Section 8 concludes the paper.

2. Related work on surveys of software technologies

2.1. Empirical study types for CBSE

There are many different types of empirical studies for software engineering [31]. Several guidelines have been published for how to conduct such studies [15] [4]. Aside from action research and case studies, the following two study types seem most relevant for CBSE, but are not discussed in [31]:

- **Structured interviews (qualitative surveys)** with an interview guide, to investigate rather open and qualitative research questions with some generalization potential. The Torino group has investigated the success of 32 reuse projects under the PIE program in EU [23]. In 2002, NTNU and the Torino group did structured interviews of seven small software companies in Norway and Italy. The results were synthesized as six theses [32] that contradicted many claims from the literature. This was followed up with a qualitative *pre-study* [18] in Norway by NTNU and the Simula Research Lab. The aim was to confirm these new findings and perhaps discover more revealing "truths".
- **Quantitative surveys** with a questionnaire containing mostly closed questions. This is suited to investigate formalized hypotheses upon a sample from a larger population, where the results can be generalized (see 2.2). For instance, the above pre-study encouraged us to perform a *main survey* of industrial practice regarding *OTS*- (COTS- and OSS) *based development* [19].

Sometimes there is a sequence, even an iteration, in the studies. That is, we start with local case studies and through interviews end up with a survey, reflecting increased theory building and generalization. As an example, the research partners have previously carried out case studies of CBSE at NASA [22], at Ericsson [20] and in German companies [2].

2.2. How to perform sound surveys in software engineering

There is a massive literature from the social sciences on how to perform valid and reliable surveys - e.g. regarding questionnaire design, sampling, and how to contact the respondents [10]. A challenge in international surveys is to stay methodologically similar across countries to avoid national bias, e.g. with sampling, contacting, and questionnaire fill-in.

Typical samples are based on convenience (reusing previous contacts), snowball (from a few seeds like a pyramid game), randomly selected subsets (from some population list), or stratified subsets (considering e.g. ICT sub-sector, company size, or country). Typical ways to fill in a questionnaire are by paper copy via post or possibly fax, by phone or site interviews, and recently by email or web. The questionnaire can be filled in either by the respondent her/himself or by an interviewer. Interviews, or phone contact in general, may result in a higher response rate, and may also

contribute to quality-assure the returned data. However, they cost more to perform and some answers may be influenced by the interviewer, possibly resulting in dishonest answers (a "mode" problem, see later discussion in Section 5).

When asking about past events, in our case finished software projects, a respondent may have to think a bit, retrieve old documents, or speak to colleagues. This may take time and effort. We must therefore offer a print-out version, in addition to a web-questionnaire - read on the bus, or similar.

There have been published some more specific advice for surveys on software engineering [25], and for how to use web-based questionnaires in surveys [26]. However, even these more ICT-specific papers have not been of much help in our sampling and contact process.

2.3. A review of previous surveys in software engineering

In light of the issues discussed in Sections 2.1 and 2.2, we will briefly present some published software engineering studies. The external validity or generalizability of all these surveys is speculatively or not at all discussed, because the *population and associated sample* are not clearly specified. First some *non-CBSE* surveys, summarized in Table 1, then an annotated list of *CBSE surveys* as shown in Table 2.

Table 1: Some non-CBSE surveys.

Study topic	Study object	Sample selection	Sample size	# responses
1.Software maintenance processes [13]	*Portfolio of maintained software systems in ICT user-organi-zations*	*Random sample of 800 company members of Norwegian IFIP subsidiary*	*500*	*53 valid of 97 returned responses (Mean: 656 for all empl., 11 in IS-dept.)*
2.SPI success criteria [6]	*Software engineering pro-cesses of ICT companies*	*Complete member lists of two Norwegian ICT industrial organizations*	*154*	*120 (Mean: 196 empl., median: 55)*
3. Estimation accuracy [21]	*Finished software projects in ICT companies*	*From 100 largest Norwegian ICT companies, listed in a business magazine*	*37*	*18 companies with 52 projects (Mean: 141 empl., median: 75)*
4. Effort estimation, COTS development [1]	*IncrementalCOTS projects*	*Convenience sample of aerospace and defense contractors*	*100*	*20*
5. Risk management policies [27]	*Finished software projects in IT departments/ companies*	*Random sample of managers in IT companies, from a Finnish IT Assoc.*	*248*	*83*
6. Efficiency of inspection proc's [26]	*State of the practice of inspections*	*Convenience sample of contact persons in German industry and academia*	*865*	*121*

Table 2: Some CBSE studies, mostly surveys.

Study topic	*Study object*	*Sample selection*	*Sample size*	*# responses*
1. Software reuse success factors [11]	*Software organizations*	*Convenience sample of 113 persons in 29 organizations*	*NA*	*113 (Median: 25,000 empl., range 35 to 350,000)*
2. Reuse proc's with Ada [16]	*Software professionals mostly in IT industry*	*Subscriber lists of three Ada magazines*	*NA*	*75*
3. Software reuse success factors [23]	*Finished EU projects on software reuse*	*From official list of 288 PIE projects in EU*	*32 projects*	*24 proj. interviews in 19 comp. (9 small, 9 medium, 6 large)*
4. COTS dev. proc's [32]	*COTS-based projects*	*Convenience sample of 7 small sw companies*	*7*	*7 (Mean and median: 7 empl.)*
5. Company attitudes towards OSS [3]	*Companies using OSS*	*Snowball sample of ICT companies using OSS*	*275*	*146 (Mean: 17 empl., range 1-320)*
6. Company experience and motivation on CBSE [7]	*Companies practicing CBSE*	*Convenience sample*	*1300*	*109 (Median: ca. 100)*
7. COTS/OSS dev. proc's [19] (this study)	*Finished sw projs with OTS dev.*	*From e.g. stratified lists of ICT companies from Census Bureau*	*1083*	*122 (Mean: 2240, median: 34)*

As with these studies, most published surveys in our field fail to report on how the population is defined, what ICT sub-sectors are covered, what sampling procedure has been applied, how representative the final sample is, what the mean and median company sizes are etc. Even most controlled

experiments in our field, where the demand for rigor is high, have similar problems. Furthermore, generalizability raises more fundamental concerns than valid statistical sampling [17].

3. Defining the relevant population

This has two parts: how to define the ICT sector (Section 3.1), and how to define our study objects inside this (3.2). We will use the terminology at (www.socialresearchmethods.net/kb/sampterm.htm):

- *Theoretical population:* all OTS-based software projects globally,
- *Study or target population:* all OTS-based projects in the entire ICT industry in the three countries.
- *Sampling frame:* list of OTS-based projects in stratified subset of above (by NACE codes and company size).
- *Sample:* complete or random selection from the above sample frame.
- *Subsample:* final or responding subset of sample.

3.1. Defining the population of ICT companies

We generally have a problem in defining our *study population* of ICT companies, or the commercial ICT sector. In our study, this should be defined for three countries, and is expected to be representative for the ICT sector world-wide, the *theoretical population*. A quick glance at public statistics, however – using Standard Industry Codes from the US Census Bureau or their EU/OECD adaptation as four digit NACE codes [24] – gives little clue of what is ICT or not. For instance, the software business is bafflingly categorized under sector K for "Real estate, renting and business activities". Thus, OECD, SSB (the Norwegian Census Bureau), and similar national agencies have jointly defined the following four sub-sectors to constitute the ICT sector in a country:

- 3x.xx ICT *industry* (Motorola, Nokia, etc.)
- 5x.xx ICT *trade* (Computerland etc.)
- 64.2x ICT *telecom* (AT&T, BT, Telenor etc.)
- 72.xx ICT *software*, both vendors (Oracle etc.) and consultants (Accenture, Cap Gemini etc.)

We will use *ICT sector* as the general name. In Norway, this sector comprises 4-5% of the economy, excluding public services, i.e. it is the biggest industrial sector after petroleum. Of course, the trade sub-section (5x.xx) is not very R&D-intensive, and hence outside the scope of our study.

In addition comes the *ICT contents sector* (database and web services, media industry etc.), expected to be finally classified by OECD et al. in 2007. There is also internal software development *outside* the ICT sector – as part of finance, engineering, public services etc. This is considered by the EU to be larger than the ICT sector itself. We will not deal with the contents sector or the non-ICT sector (including the public one) here, other than including some large non-ICT companies in our survey.

As a start, let us look at the structure of the ICT sector in Norway [29]. By *employee* is meant both owners and hired staff. By *company* (the Norwegian SSB uses the term *enterprise*) is meant a "legal" economic unit (typically a public shareholding company), not just an internal development department or a branch office. Some comments are:

- For software, ICT as well as the total, over 2/3 of the companies are *small* (1-4 employees) and with a median of 1 employee. The company average is 4.5 employees for software, 7 for all ICT, and 5 for the total. Only 0.5% (46) of the software companies, 1.5% of the ICT companies, and 0.5% of all companies have 100 or more employees.
- 8,228 (about 2/3) of the 12,861 ICT companies are software companies (NACE code 72.xx), having over 40% of the employees. Of the 7,171 software companies with 0-4 employees, 4,113 have one employee and 1610 have no employees!
- Ca. 30% of both software and ICT employees work in companies with less than 20 employees, 30% in companies with 20-99 employees, while 40% work in companies with 100 or more employees.
- There are regrettably some deviations in the totals, since they are drawn from different SSB sources.

The ICT company distribution in Germany [8] and Italy [14] in 2002 show that, as in Norway, most ICT companies are small. In Germany, the average software company has 7 employees and the ICT average is 13. 1% of German software companies have 100 employees or more. In Italy, the average software company has 4.5 employees, the median is 1 employee, and the average ICT company has 7 employees.

The software company distribution in the US in 1997 can be found in [9]. The US pattern is similar to that described in Europe, with many small companies and a median of 1 (!) employee. However, the average company size is 150 employees in the US (4.5 in Norway, 7 in Germany, 7 in Italy), and 3% of the software companies had 100 or more employees.

The documented structure of the ICT sector in all these countries has clear consequences for defining our population and later samples of this (Section 4). We have four main sources for concrete *names* of ICT companies in our *study population*, called P1-P4:

- P1) A *convenience* subset of company contacts among the research partners, as in the pre-study. Relevant contact persons are often available.
- P2) The 100 largest ICT companies and similar, often summarized in *annual lists in business magazines.* No contact person is usually given.
- P3) *Members* of one or several *industrial organizations*. Norway has three such organizations, Germany two, and Italy two. However, we do not know how comprehensive or accurate ("ICT-related") such member bodies might be. In Norway, one of these organizations covers 5% of the total (600 of 13,000 companies). The listed contact is often an administrative person.
- P4) *Complete set of registered ICT companies* defined by public sources, e.g. by NACE-based lists from census bureaus or as sector lists in the yellow pages in phone books. Such sources should give a trustworthy coverage of the ICT sector.

In all such lists, there will be a substantial share of outdated information: ceased companies, mergers or split-ups, or companies with e.g. new names, contact persons, locations, or phone/email addresses. All this reflects a very dynamic marketplace.

3.2. Defining the population of finished OTS-based projects

For many ICT surveys, including this one, we do not only have a problem in defining the population of all ICT companies, and later samples with respondents of these. Our real problem is that the actual *study object* (experimental unit) – a special kind of project – is not identical to the *primary object* (accessible sampling unit), being a company or a respondent.

To be precise, our *study object* is "finished software project using OTS components". The corresponding *study topic* is "OTS-based development processes". In other studies, the study object could be "maintained software system" (to study software evolution processes), "finished software project" (to study estimation accuracy), or "applications of inspection methods" (to study related cost-efficiency).

Thus, we should first ask "every" ICT company (see 3.1) in the relevant countries to comprehensively document all recent instances of OTS-based projects, e.g. by an *explorative pre-study*. We might even conduct such a pre-study in two steps: first identify "typical" software projects, and then identify the "OTS" subset of this. But which project attributes should be collected to serve as later sampling criteria, and how to validate such criteria? We could consider company size, company profile, project size and duration, application domain, lifecycle process, developer skills, development technology, actual OTS-components being used etc. etc.

The above circumstances mean that many companies may have to give us project data, that later will be unused ("wasted"). Due to resource constraints and logistical problems, it seems totally unrealistic to perform such an explorative pre-study of a potential population of relevant study objects, both from the research and industry side.

Furthermore, only 1/3 of the sampled ICT companies in our survey had OTS-based development. This share would be close to 100%, if the study topic was effort estimation or software maintenance – i.e. activities that all companies routinely perform.

In surveys in medicine and social sciences, the population is often well-defined and coming from reliable public sources, although use of volunteers leads to an open-ended population. There, however, a responding person often matches the actual study object.

4. Sampling from the population

A *valid sample* is a *representative subset of the study population.* Sections 3.1 and 3.2 have indicated some problems in defining the study population in software engineering surveys in the ICT sector. Then, how to pick a valid sample from a poorly or even un-documented population of study objects? Even if we had a complete list of all OTS-based projects, what would be the selection method – complete, stratified, random, or combinations?

At best, we can establish a hopefully representative population of primary objects (ICT companies) from which we can draw a valid sample. The study population of ICT companies may be defined as described in 3.1. We could then draw either a *complete* or a *random-subset* sample of primary objects as defined by P1-P3, or a *stratified-random-subset* sample from P4, as mostly done here.

The final sample will be a *convenience sample* of study objects (OTS-based projects), that the actual contact person (respondent) in the selected ICT company is willing to provide. We can expect one or two such projects per company, but mostly none.

In our case, we needed about 50 study objects from each country to achieve statistical significance, given our research questions and related questionnaire questions. We therefore planned to ask about 250 companies in each country. Assuming an overall response rate of 1/3 or around 80 answers in each country, that would give us a satisfactory yield with some margin. (In reality, we needed 400-500 companies to get 50 responses per country). However, we should not draw a random sample from "all" ICT companies. Based on the documented structure of the ICT sector, our *sample frame* should ideally be stratified according to NACE code and company size.

For the **Norwegian sample,** our sample frame consists of three groups of ICT companies: large (all ICT sub-sectors including some non-ICT ones), medium (only software), and small (also software). Each group has roughly the same amount of software developers. The sample frame is then as follows:

1) All large ICT companies (with 100 or more employees) with any ICT-relevant NACE code; totally 114 companies and with average size of 330 employees. We sampled the topmost 100 of these 114 ICT companies, and 26 turned out to be software companies.
2) Additionally, simply the 3 largest companies in 5 non-IT sectors (finance, engineering, energy, food processing, public/other). We sampled all these 15 companies.
3) All medium-sized software companies, NACE codes 72.xx (20-99 employees), totally 228 companies with average size of 40 employees. From this we randomly sampled 100 companies (increased to 200, due to poor response rate).
4) All small software companies, NACE codes 72.xx (5-19 employees), totally 783 companies and with average size of 9 employees. From these 783 companies we randomly sampled 100. We removed those with 0-4 employees, as they represent 87% (7171) of the companies, but only 19% (7514) of the employees. They are not expected to have very structured work processes, and are assumed irrelevant for our study.

We got company names for year 2002 for category 1, 3, and 4 of ICT-companies. Randomization for category 3 and 4 was performed beforehand, and the data sent to us as spreadsheets via email. The 15 large non-ICT companies in category 2 are from a business magazine for year 2002.

Some observations: The Norwegian SSB struggles with incomplete or erroneous company lists. SSB has one part-time person on ICT and twelve on agricultural statistics! SSB has cooperated with one of the Norwegian industrial ICT organizations, IKT-Norge, since 1998 to help establish a reliable database of Norwegian ICT companies. IKT-Norge has recently set up the Infosector subsidiary [33] to sell services around this database. This subsidiary turned out to be the only practical way to get a stratified sample of the official database on electronic form, based on NACE category and company size. We hence got the requested company list very fast at a low price (1250 euro) – on the condition that we reported back anomalies in the data.

For the **German sample,** we got a list of company names for the years 2001 and 2002 from the German census bureau (Statistisches Bundesamt) for the four sub-sectors of the ICT industry, enhanced by major non-ICT companies whose revenues are largely based on software (e.g., Siemens, DaimlerChrysler, Allianz). These lists were then validated using our customer database (e.g., removing falsely classified organizations or companies which clearly do not develop software, such as hardware stores). From the resulting sample frame of 430 companies, we selected a random sample for each category. Each sample was then cross-checked with the IESE customer database in order to obtain a contact point.

For the **Italian sample,** there is no national directory of ICT companies available to us. We therefore started with a study population of 43,580 assumed software companies from the yellow pages in 2001 (the same database as in the ELISS project on OSS [3]). From this population, we made a sample frame of 6,000 possible company contacts. From this sample frame, we randomly picked a sample of 747 companies, which we cross-checked (by web etc.) that their addresses were still valid and that they really did software development. We were left with 248 software companies, whereof 125 were using OTS components. 118 of these accepted to fill in the web-questionnaire. However, many potential respondents didn't know what an OTS component was.

Tables 3 and 4 show the response rates in the three countries. In "Likely" we have pruned duplicates or irrelevant items such as holding companies. "Contactable" means the company coordinates were still valid.

Reflections on sampling and response rates:

- For all countries, there were problems in classifying large and perhaps international ICT corporations, with many daughter companies and divisions. That is, what is the relevant *company unit*? A company (legal economic unit) may have 10,000s of

employees, but a local business unit (e.g. R&D department) less than 100. E.g., we twice "adjusted" a too high company size in Norway.

- The relatively high mean size for companies in Germany and especially in Italy is caused by a few large "outlier" companies..
- The local business unit size is considered the most representative measure. This does not vary much between the three countries, with the larger business units in Italy and Germany, as expected.
- In the pre-study in Norway, 17 of 34 (50%) of the ICT companies in our convenience sample had COTS-based development, and most of the remaining said they also used internal components. 13 of these 17 companies (almost 80%) agreed to participate and all responded.
- In the main study in Norway, 94 of 334 (28%) of the "contactable" companies had OTS-based development, and only 41 of the 94 (44%) eventually responded with a total of 47 projects. That gives an overall response rate of 12% (41 of 334).
- In Norway, 50 of 380 (13%) of the companies were no longer "contactable", as the company lists were two years old in 2004. This is not surprising, and represents a very dynamic marketplace, as mentioned earlier.
- In Norway, we found that 70 of 285 (24%) software companies (NACE code 72.xx) claimed to have *no software development at all*. Many of these companies were sales subsidiaries of international ICT corporations, perhaps belonging to NACE code 51.84? However, no "penalty" for a company to be registered with a wrong NACE code exists!
- The response rate in Norway increases with decreasing company size. The response rate was especially low in large companies, cf. "gatekeeper" problem in Section 6.

Table 3: Response rate for Norwegian companies.

	Init. sample	Likely (pruned s.)	Contactable	Do SW dev.	Use OTS comp.	Decl. willing	Final sample N
Large ICT comp. (where-of SW)	100 (26)	82 (23)	80 (23)	51 (21)	28 (13)	8 (6)	3 (2)
Large *non*-ICT comp.	15	15	15	9	1	0	0
Medium SW comp.	200	183	152	110	41	30	22
Small SW comp.	100	100	87	61	24	18	16
Sum main study	415	380	334	231	94	56	41
Sum pre-study	34	34	34	34	17	13	13

Table 4: Company responses in the three countries.

	Initial sample	Likely (pruned s.)	Contact-able	Do SW dev.	Use OTS comp.	Declared willing	Final sample "N"	Mean/median in final sample :company	Mean/median in final sample: business unit
Norway	415	380	334	231	94	56	41	279 / 30	20 / 11
Germany	430	430	430	314	206	60	48	1933 / 200	74 / 40
Italy	747	747	323	248	125	118	38	3091/ 25	76 / 9
Total	1592	1557	1087	793	425	234	122	3371 / 44	56 / 20

5. Questionnaire design

The (english) interview guide for the pre-study included standard meta-questions to test understandability, was held in two runs in fall 2003 on ten industrial colleagues in Norway, and lead to 10% of the questions being revised.

The main study questionnaire was designed in cooperation with the partners in Dec. 2003 – April 2004, relying on email correspondence and phone meetings. Most of the questions from the pre-study were reused, and we used much effort to reliably map the research questions and hypotheses into the questionnaire. The design went through 16 major revisions, and this lead to changes in hypotheses, measure scale, and answer categories. Also, we created an analysis plan and conducted a questionnaire pre-test to increase survey quality.

The final questionnaire had an introductory letter, a set of concept definitions (component, OTS etc.) and some control questions to the reader. Thereafter 10 pages with 80 questions divided in 6 parts. Questions on personal or company matters were put at the end, emphasizing the confidentiality of the responses.

The pre-test had 18 respondents (industrial and academic) in the three countries. 15% of the questions were altered based on this. All work was done in English using an ms-word (.doc) file.

The questionnaire was finally translated into the native language of the three partners to ensure equal comprehension across cultures, and this revealed only few unclarities. Filling in took ca. 20-30 minutes.

Reflections on the questionnaire:

- The questionnaire showed robust, with very few misunderstandings. Only four incompletely filled-in forms (of 47) came in for Norway.
- In Germany the fill-in procedure was improvised to be phone interviewing of respondents. This introduced a *mode problem,* that could affect "sensitive" answers, the response rate, or cause misunderstandings [12]. However, most questions are factual and technical, and the few national differences in German answer are not likely caused by mode problems, yet this mode difference represents an unplanned method bias.
- In Germany, we also had to offer extra anonymity for respondents and companies, in spite of an explicit confidentiality clause.

6. Contacting the companies

Through a list of potential companies, we have to find a contact person in each company. The initial contact person (not the switchboard operator!) in a survey is called a *gatekeeper* [5]. Ideally, this person should connect us with a *respondent* for the actual study object. Note again, that the respondent is not our study object, neither is the company itself.

For convenience samples, there are few problems in getting hold of a potential respondent, as we often have personal and professional ties from earlier, joint activities. But for a large corporation, from a census or member list and with thousands of employees, it can be very frustrating to try to locate a quality manager, development/project manager, or just an "IT responsible". The final struggle is to convince a possible respondent over the phone to use his/her time and effort to fill in a questionnaire, sent by email.

Phone or site-interviews were discussed thoroughly in March 2004, but, despite recommendation [21] by researchers at Simula Research Lab, rejected, as we at that time estimated having 80-100 respondents per country and therefore did not have the capacity to do so (especially based on German arguments!).

Problems (low response rate etc.) were anticipated, so written procedural guidelines were made early:

1) Establish a representative sample of companies (Section 4). Generally, this meant no pre-known gatekeeper, respondent, or phone number.
2) Rewards for survey participation were the final report, invitation to an industrial sum-up seminar (www.idi.ntnu.no/grupper/su/cbse-survey/), and participation in a holiday lottery worth 1250 euro.
3) Call up the company, ask to be connected to a probable gatekeeper (first try the "IT manager"), and eventually to a prospective respondent. Explain the survey context including reward incentives, and ask if the company has software development and also OTS-based development. A positive return should yield an operative email address and a direct phone number.
4) The respondent is then emailed a "paper copy" of the questionnaire, telling her/him to please fill in the questionnaire online by the web (preferred option), on paper, or as an email attachment.
5) A few days later, re-phone and ask if the respondent is willing to participate in the survey. If yes, issue a userid/password to the web server.
6) Then wait for a final response, re-phone if necessary etc. This process could take months, even if the requested work took only half an hour.
7) In case of an incompletely or wrongly filled in questionnaire, we would try to call back.

Though this may sound smoothly, we have some **reflections on the company contact process:**

- All this turned out to be very costly and with a rather low response rate. Given a person-year of total effort just in Norway, the cost of a filled-in questionnaire is ca. one person-week (2000 euro).
- As mentioned, the German fill-in procedure was performed through interviews, i.e. different from that used in Norway and Italy.
- Italy lost their research assistant (a summer student) in Oct. 2004, not replaced before March 2005. Likewise, IESE had 5 months delay (May-Sept. 2004) in PhD student upstart. Also, Norway hired an extra research assistant (upstarting PhD-student Slyngstad in March-June 2004). All this lead to much improvisation.
- There was marginal interest for any reward system - maybe we were not inventive enough?

7. Lessons learned

7.1. Method Aspects

- A **pre-study** often clarifies both research questions and concrete questions in a later questionnaire. For

instance, we "discovered" several variants of OTS-based development processes in our pre-study [18].

- A **representative sample** of national ICT companies assumes access to public data sources, like NACE-based company lists from census registers, but even such data may contain many errors. Close contact persons in such organs are needed to obtain stratified samples and expedite delivery of data in electronic form.
- Better **criteria for selecting projects** are needed. Since the study population (of finished OTS-based projects) was undefined, we had to do a **two-phase** approach: first pick a representative sample of companies, then a convenience sample of projects.
- For a variety of reasons it seems impossible to perform the same random sampling and data collection in different countries. The mentioned **national differences** now give an **unknown mode effect** in result interpretation between the countries. However, the observed national differences for Germany are unlikely to be linked to mode effects.
- **Random** sampling and the ensuing contact process appears to be **much more expensive** (5 times?) than **convenience** sampling. In our case, this is because we got stuck with no previous contact person, and scant pre-knowledge and "goodwill" from such companies, especially the large ones.
- If the final sample size is under **100 respondents**, use **phone interviews**. This applies particularly when phone contact is anyhow needed, e.g. due to a non-convenience sample. A written procedure for the contact process is also needed.
- A baseline can be useful for comparison purposes. In our case, we did not ask the companies about non-OTS projects, so we missed this.
- Interviewing at this level requires solid background knowledge of the material; Company phone contact assumes at least PhD student qualifications.

7.2. Practical experiences

The industrial ICT environment seems extremely busy in recent years. It is very hard to ask for any extra commitment of effort. Also, of an audience of 25 at our final seminar, only 5 were respondents.

The SESE web server in Oslo has generally worked fine. One of the researchers had earlier experience as a webmaster, and spent two days in Oslo to learn about the SESE tool. The tool has already been used by over 20 studies, mostly controlled experiments, including a shared one by NTNU and the Simula Research Laboratory on software maintenance.

7.3. Research cooperation

The researcher cooperation has gone very well, due to previous contacts, shared research interests, and good person chemistry. No physical meetings were held, and all data are sharable among the partners.

Extra resources through an external project would have helped to minimize national process differences, and having a social scientist on the team is strongly advised, as we found it highly useful.

Finally, the senior researchers must be thoroughly acquainted with details around sampling etc., i.e. getting their hands "dirty" by serving as research assistants.

8. Conclusion

This study is the first survey on CBSE and indeed software engineering, which uses stratified-random sampling of ICT companies in several countries. All previous surveys have had problems in defining the study population and in selecting a valid sample of this, especially since the study object is a special project, not a company.

We had few real problems with questionnaire design and implementation, and some problems with randomized company sampling. On all these points we anticipated problems, and put in a lot of effort. However, unexpected much work went into contacting the companies and having the questionnaires filled in. National differences in sampling procedures and questionnaire fill-in may have introduced method biases, although not serious it seems.

We have learned a lot in running a representative, multinational survey, and also on OTS-based development. We plan to follow up the survey in other countries (China and US), and to supplement with more qualitative studies.

9. Acknowledgements

This study was funded in part by INCO (Incremental COmponent based development, 2001-2004,http://www.ifi.uio.no/~isu/INCO) and SPIKE (SPI by better Knowledge and Experience), 2003-2005, http://www.idi.ntnu.no/grupper/su/spike.html). We thank our colleagues in these projects, social scientist Ola Listhaug at NTNU, and all participants in the survey.

10. References

[1] C. Abts, B.W. Boehm, and E.B. Clark, "COCOTS: A COTS Software Integration Lifecycle Cost Model – Model Overview and Preliminary Data Collection Findings", Proc. of the ESCOM-SCOPE 2000, 18-20 April 2000, Munich, Germany, Shaker Publ., pp. 325-333.

[2] C. Atkinson, C. Bunse, and J. Wüst, "Driving Component-Based Software Development through Quality Modelling", In A. Cechich, M. Piattini, and A. Vallecillo (Eds.), "Component-Based Software Quality: Methods and Techniques, State-of-the-Art Survey", Springer Verlag LNCS 2693, 2003, pp. 207-224.

[3] Andrea Bonaccorsi and Cristina Rossi, "Why Open Source software may succeed", Research Policy, 32(7):1243–1258, 2003.

[4] Reidar Conradi and Alf Inge Wang (Eds.), *Empirical Methods and Studies in Software Engineering -- Experiences from ESERNET*, Springer Verlag LNCS 2765, ISBN 3-540-40672-7, Aug. 2003, 278 pages.

[5] M. Corra and D. Willer, "The Gatekeeper", Sociological Theory, 20(2):180-207, July 2002.

[6] Tore Dybå, "Factors of SPI Success in Small and Large Organizations: An Empirical Study in the Scandinavian Context", Proc. Joint ESEC'03/FSE-11 conf., Helsinki, Finland, 1-5 Sept., ACM Press, pp.148-157.

[7] Marisa Escalante and Pedro Gutierrez, "CBSE: State of the Practice", Proc. 16th Int'l Conf. on Software & Systems Engineering and their Applications (ICSSEA'2003), 2-4 Dec. 2003, Paris, Vol. 2, Session 9: Reuse & Components, ISSN 1637-5033, 16 p.

[8] EuroStat office, EU, Luxembourg, http://europa.eu.int.comm/eurostat, 2002.

[9] M. E. Fayad, M. Laitinen, R. P. Ward, "Software Engineering in the Small", CACM, 43(3):115-118, March 2000.

[10] Floyd D. Fowler, *Survey Research Methods*, 3rd edition, Thousand Oaks, Sage, 1988.

[11] W. B. Frakes and C. J. Fox, "Sixteen Questions About Software Reuse", CACM, 38(6):75-87, June 1995.

[12] Robert M. Groves et al., *Survey Methodology*, Hoboken, NJ, John Wiley, 2004.

[13] Knut K. Holgeid, John Krogstie, Dag I. K. Sjøberg, "A study of development and maintenance in Norway: assessing the efficiency of information systems support using functional maintenance", Information & Software Technology, 42(10):687-700, 2000.

[14] Italian Census Bureau, www.istat.it, 2002.

[15] Barbara A. Kitchenham et al., "Preliminary guidelines for empirical research in software engineering", IEEE Trans. on Software Engr., 28(8):721-734, Aug. 2002.

[16] N. Y. Lee and C. R. Litecky, "An Empirical Study of Software Reuse with Special Attention to Ada", IEEE Trans. on Sw. Engineering, 23(9):537-549, Sept. 1997.

[17] A. S. Lee and R. L. Baskerville, "Generalizing Generalizability in Information Systems Research", Information Systems Research, 14(3):221-243, Sept. 2003.

[18] J. Li, F. O. Bjørnson, R. Conradi, and V. B. Kampenes, "An Empirical Study of Variations in COTS-based Software Development Processes in Norwegian IT Industry", Proc. 10th IEEE Int'l Metrics Symposium (Metrics'04), Sept. 14-16 2004, Chicago,p.72-83 (*pre-study*).

[19] J. Li, R. Conradi, O. P. N. Slyngstad, C. Bunse, M. Torchiano, and M. Morisio, "Preliminary Results of a State-of-the-Practice Survey on Motivations of Using Off-The-Shelf Components", in Proc. 6th International Conference on Product Focused Software Process Improvement (PROFES'2005), 13-16 June, 2005, Oulu, Finland, 17 p. (on *main study*).

[20] P. Mohagheghi et al., "An Empirical Study of Software Reuse vs. Defect Density and Stability", Proc. 26th Int'l Conf. on Software Engineering (ICSE'2004), 23-28 May 2004, Edinburgh, Scotland, IEEE-CS Press, pp. 282-292.

[21] K. J. Moløkken-Østvold et al., "A Survey on Software Estimation in the Norwegian Industry", Proc. 10th IEEE Int'l Metrics Symposium (Metrics'04), Chicago, USA, 14-16 Sept. 2004, p. 208-219.

[22] M. Morisio et al., "Investigating and Improving a COTS-Based Software Development Process", Proc. 22nd Int'l Conference on Sw. Engineering (ICSE'00), Limerick, Ireland, IEEE-CS Press, 4-11 June 2000, pp.31-40.

[23] M. Morisio, M. Ezran, and C. Tully, "Success and Failure Factors in Software Reuse", IEEE Trans. on Software Engineering, 28(4):340-357, April 2002.

[24] Nomenclature Generale des Activites Economiques dans l`Union Europeenne, European Union, Brussels,2005,http://europa.eu.int/comm/competition/mergers/cases/index/nace_all.html.

[25] S. L. Pfleeger and B. A. Kitchenham, "Principles of Survey Research", 6 parts in ACM SIGSOFT Engineering Notes, Nov. 2001 – March 2003.

[26] T. Punter, M. Ciolkowski, B. Freimut, and I. John, "Conducting On-line Surveys in Software Engineering", Proc. 2nd Symp. of Emp. Sw. Eng. (ISESE'03), Rome, 30 Sept.–1 Oct. 2003, IEEE-CS Press, pp. 80-88.

[27] J. Ropponen and K. Lyytinen, "Components of Software Development Risk: How to Address", IEEE Trans. on Software Engineering, 26(2):98-111, Feb. 2000.

[28] Dag I. K. Sjøberg et al., "Conducting Realistic Experiments in Software Engineering", Proc. 1st Int'l Symp. on Empirical Sw. Eng. (ISESE'2002), 3-4 Oct. 2002, Nara, Japan, IEEE-CS Press, pp. 17-26

[29] Norwegian Census Bureau (SSB), Oslo, ICT company data 2002, http://www.ssb.no/emner/10 /03/ikt/.

[30] J. Li et al.: "An Empirical Study of COTS Component Selection Processes in Norwegian IT companies", MPEC'04 at ICSE '04, May 25, 2004.

[31] M. V. Zelkowitz and D. R. Wallace, "Experimental Models for Validating Technology", IEEE Computer, 31(5):23-31, May 1998.

[32] M. Torchiano, M. Morisio, "Overlooked facts on COTS-based Development", IEEE Software, 21(2):88-93, March/April 2004.

[33] Infosector company, www.infosector.net