SSH8-CT-2009-225368
INGINEUS
Impact of Networks, Globalisation and their Interaction with EU Strategies

Type of funding scheme: Collaborative Project (large-scale integrating project)
Programme addressed: FP7 Theme 8 "Socio-economic Sciences and the Humanities" (SSH)

Deliverable 2.2
Complete standardised data set containing all the information collected in all countries

Due date of deliverable: Month 27
Actual submission date: Month 29

Start date of project: 1st January 2009 Duration:36 months

Name of lead contractor for this deliverable:
University of Pretoria / Gordon Institute of Business Science (UP / GIBS) Participant no.12)

Revision:

Project co-funded by the European Commission within the Seventh Framework Programme (2007-2013)		
Dissemination Level		
PU	Public	**PU**
PP	Restricted to other programme participants (including Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CP	Confidential, only for members of the consortium (including the Commission Services)	

Table of contents

The INGINEUS survey: methodology report

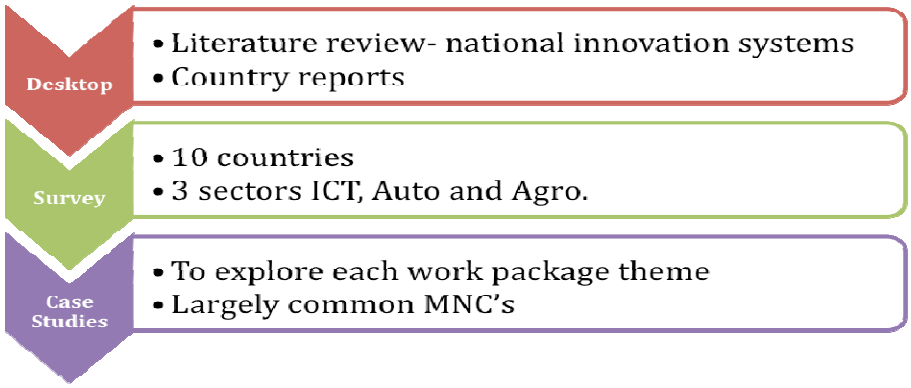
1.1 Introduction

The INGINEUS project focuses on the developed and developing world to determine the extent to which innovation is taking place in globally dispersed networks. The Survey was conducted as part of the broader methodology of the INGINEUS research project. This project aims to capture the dimensions of GINs (global, innovative and networked enterprises) through desktop research, a survey questionnaire and by means of appropriate cases which are examined and researched according to a pre designed set of parameters and constructs (See figure 1).

The survey was conducted across nine countries: Brazil, India, China, South Africa, Norway, Sweden, Germany, Estonia and Denmark. Past experiences of poor response rates in Italy resulted in a shift in strategy for the Italian team. A similar survey was conducted in Italy however the format differed as specific INGINEUS questions were included as part of a larger survey managed by the Turin Chamber of Commerce.

Each country had a dedicated sector of focus in either ICT, Automotive or Agro processing (Sweden had a small number of auto surveys in addition to ICT). Each institute conducting the survey across the nine countries chose a sector which was of economic importance within their national or regional context.

Figure 1: INGINEUS methodology design-triangulation of data sources



Anecdotal evidence suggests that the Southern or emerging nations are increasingly becoming a source for knowledge generation and the inception of ideas, forming a so-called new 'technological frontier' for the development of knowledge.

In order to systematically and rigorously examine this trend, the INGINEUS survey and the project overall, was designed to incorporate data from the Northern developed regions of Norway, Sweden, Germany, Estonia, Italy and Denmark with data which is drawn from Brazil, India, China and South

Africa or the emerging economies. This design affords the study a unique geographic and economic perspective.

The evidence also suggests that the outsourcing of knowledge-intensive activity and the emerging phenomenon of knowledge creation in developing locations outside the EU is spreading from the electronics sector to many other sectors of the economy, that is, from traditional low-tech industries such as agro-processing or medium-tech such as automobile to high-tech such as ICT. A deeper examination of this phenomenon formed the motivation behind the sectoral choices of the ICT, Automotive and Agro-processing sectors for this study. In order to adequately map this trend across the differing tech intensities of each sector the design ensured that each sector was covered by a more developed and less developed country as seen below:

- Agro-processing: **South Africa** and Denmark
- Automotive: **Brazil**, Germany, Italy and a small sample from Sweden
- ICT: **India, China**, Sweden, Norway and Estonia

Table 1: INGINEUS survey results by sector and by country

COUNTRIES	ICT	AUTO	AGRO	TOTAL
Brazil		69 (25.9%)		
China	243 (2.7%)			
India	324 (25.2%)			
South Africa			84 (16.9%)	
TOTAL emerging markets	**567**	**69**	**84**	**720**
Denmark			49 (23.3%)	
Estonia	17 (14%)			
Germany		53 (4.7%)		
Norway	181 (11.9%)			
Sweden	171 (10.3%)	24 (14.3%)		
TOTAL developed countries	**369**	**77**	**49**	**495**
Total	**936**	**146**	**133**	**1215**

In all sectors and across all countries 1215 responses were collected. The combined INGINEUS sample was dominated by ICT responses. This was in part due to the size of the Indian and Chinese market but also due to the nature of the agro processing and Auto industries which tend to be more concentrated. Table 1 above offers a summary of the results and number of responses received from each sector and each country.

1.2 Novelty of INGINEUS data

The paragraph which follows describes what the INGINEUS data delivers which cannot be obtained from existing datasets. It is taken from the INGINEUS DOW:

INGINEUS is of course not the first endeavour to understand GINs. Evidence of emerging GINs have been around for some time and is being regularly reported on in the business press. Scholarly analysis is also paying attention to the problem. Yet unlike the rich body of work on GPNs, GINs are too recent a phenomenon to have attracted exhaustive treatment. In fact, much evidence presented is anecdotal – it refers to instances of individual firms that have undertaken cross-border investments in knowledge-intensive activities without discussing causes and effects of such decisions systematically. It is against this baseline that the INGINEUS project starts. The empirical and theoretical points of departure are the insights summarised in Section B1.2.1, on the state of the art of our knowledge about the extent and the causes of GINs. The difference between INGINEUS and previous work manifests itself in the following dimensions.

Individual instances of firms that constitute GINs: INGINEUS provides a theoretical framework within which to situate case studies systematically so as to construct cumulative evidence of the genesis, performance, and implications of decentralized knowledge-intensive activities (see WPs 4-9).

The scope of GINs: INGINEUS is first at integrating existing databases for the analysis of the extent and depth of GINs throughout the world (see WP3).

The determinants of GINs: Through a customised survey instrument, INGINEUS creates a new data set in representative markets that show what motivates firms in offshoring and outsourcing R&D and other knowledge-intensive activities (see WP2).

The impact of GINs: INGINEUS contributes an over-time analysis to the heretofore primarily static assessments GINs have sparked (see WP8).

According to its research design, INGINEUS will therefore overcome a disparate and rather incomplete understanding of GINs through the investigation of theoretically derived propositions that are examined on the basis of both newly interpreted and entirely new information. Progress of the project will be measured in terms of how systematic the produced evidence is.

Reviews of INGINEUS should focus on how well the project explains the phenomenon whereby innovative activities are increasingly being located in places other than their home country. Within this context, relevant questions include the relative roles of firm decisions versus regional and country characteristics, industry patterns (e.g. whether different knowledge bases lead to different knowledge networks), and so on.

1.3 Background to the survey design

It proved to be quite challenging to develop a questionnaire and conduct a survey that captured global innovation networks in meaningful depth and with adequate specificity, but that was able to work across very different contexts. The survey design was achieved by communicating extensively with all relevant partners and committing to a ‘bottom up’ approach in the design process. Work package 2

provides input into a number of other work packages, notable WP 4, 6, 7, 9 and 10. For this reason, it was important to work very closely with the relevant work package leaders in the process of developing the questionnaire and determining a strategy for execution.

As a starting point each Work Package leader was asked to submit a document which contained the following:

- A proposed a set of questions for the questionnaire.
- An explanation of which deliverables the survey questions contributed to
- An explanation of the theoretical foundation / conceptual underpinnings / mental model informing the survey questions (details of what they hoped to achieve through the specific question)
- A description of how they saw the analysis being conducted
- (e.g. correlate X with Y to see if ....)

Each country team conducting the survey also had to provide a list of the databases to be used, propose their sampling technique (random, stratified random) and sample criteria, e.g. minimum size of firm. The sectoral definition per sample (ISIC/NACE codes) was also established.

The document needed to introduce the delivery method as this was expected to be different across for example India and Brazil as compared with Sweden and Norway due to infrastructural, language and cultural differences across the participating economies. It was also anticipated that in some cases we would have a regional sample and in others a national sample, depending on the sector and country specificities.

1.3.1 Language compatibility

One of the dangers with surveys across multiple language samples is the reliability and consistency of the survey questions across the various countries. In order to reduce the possibility that the survey questions were not distorted by translations all translated surveys were sent to WP 2.

In order to reduce the likelihood of mis-translations WP 2 conducted and completed back translations of the survey from Mandarin, German, Portuguese and Danish. South Africa, India, Sweden, Estonia and Norway conducted the survey in English.

1.3.2 Testing the survey instrument

Survey partners were instructed to **carry out pilots** and by early October a feedback report on the pilot was produced and sent to the Executive from WP 2.

Changes to the survey based on the pilot feedback were then forwarded to the question contributors. This process caused significant delay with partners reluctant to reduce and edit their sections of the survey.

On the 29th July 2009 the finalised survey questionnaire was forwarded to the **Scientific Advisory Committee.** Positive feedback was received from the Committee on the 13 September 2009. Survey partners were then requested to run a test batch of surveys through Survey Monkey in order to test the electronic survey system and get users acquainted with the software.

1.3.3 Defining an industry

The concept of defining an industry for this project was fully explored in the methodology document: *Making methodological decisions in cross-country, cross-industry research: Articulating the assumptions underlying indicators* by Barnard &Ismail to be presented at the 8th International Globelics Conference. An extract from this paper is included below to explain the processes involved in defining the industries for the INGINEUS project in greater detail.

'The INGINEUS project covers three sectors, high research-intensive (Information and Communications Technology, ICT), medium research-intensive (automotive) and low medium-research intensive (agro-processing). Research-intensity is only one dimension of an industry, and the chosen industries are similar and different in a number of other ways. But given the recognised albeit complex link between research-intensity and innovation (for a recent reassessment, see Mairesse & Mohnen, 2005), it was hoped that insights could be gathered across a spectrum of different industries.

Simply defining the boundaries of each industry presented major challenges. ICT is a General Purpose Technology (Breshanan & Trajtenberg, 1995), and therefore ICT firms can be found in a number of other industries, including but not limited to automotive and agro-processing. In turn, both automotive and agro-processing are cross-sectoral, integrative industries. The end product for the customer may be a motor vehicle or a packet of food, but both industries are characterised by the importance of a number of very different inputs needed to obtain that end product.

Because the central goal of the INGINEUS project was to track the globalisation of innovation networks, i.e. how global production networks were (or not) evolving into innovation networks, and what types of innovations were originating where, the project needed to adopt an inclusive rather than exclusive approach. However, for the sake of rigour, the research still needed to be clearly delimited.

The project relied to a substantial extent on industry codes such as the ISIC (International Standard Industrial Classification of All Economic Activities) to provide a "shared language" when we were trying to delimit the research. We did not set out to interrogate how those codes had been defined and derived, and instead used them to ensure that researchers in very different contexts had a shared understanding. The major challenge in using those codes was that not all countries' databases were organised according to the same coding system, not to mention the same version of a given coding system. In a few cases, especially in less developed countries, the available databases were not categorised at all in terms of industrial codes, and using industrial codes served more to signal the type of industries we were looking for. Even in the European countries, research partners used different versions of ISIC, NACE ('Nomenclature générale des activités économiques dans les Communautés Européennes') and local adaptations like EMTAK from Estonia. Extensive communication with the methodology coordinator was needed to ensure the greatest possible comparability.

But while standardised codes provided a "language" for talking about industries, it was still hard to define the boundaries of an industry. The increasing specialisation in global value chains forced partners to think very carefully about the firms we wanted to include and not in the database. For example, in the case of ICT, India has been making a name for its software services, while China has been relatively more active in the manufacture of hardware. Emphasising one over the other would bias the results, and therefore the decision was taken to include the two very different sectors in the questionnaire. To increase cross-country comparability, all the other partner countries were asked to consider both hardware and software sectors, in spite of the fact that these two sectors differ

fundamentally in terms of aspects like capital intensity and the nature of firms' and employees' required capabilities. After struggling with the response rate and realising that it would need to play a more active role in data gathering, one country, Norway, focused only on the sector with the greatest number of firms[1], but generally partners were able to cover both manufacturing and services in ICT.

In the case of automotive and agro-processing, the solution was somewhat less elegant because of the large number of technologies used in those industries. There are firms that are classified as automotive companies, but focusing only on those firms may lead to an underrepresentation of innovative activities by firms active in the automotive sector, because the firms that are part of the automotive chain of production are often quite different from the companies which are formally classified as part of the sector. For example, Ingineus would be interested to know about smart fabrics for vehicle seats, but finding that information through a survey would require the inclusion of the textiles industry in the survey. According to the same logic, rubber, glass, and a number of other sectors would need to be included, but the bulk of firms in each of those particular sectors would be unlikely to have links with the automotive industry.

The research team on the German auto industry attempted to use industrial categorization as one of the criteria in defining the automotive industry, but did not use it as the main criterion for inclusion into their sample frame. Instead, the German partners built their sample frame by using a private database, Hoppenstedt, that could be sorted by the sector of the main client. By looking at the frequency with which the automotive sector was mentioned as a client in that sector, it was possible to identify at the 5-digit level six sectors that specifically served the automotive industry. These categories were not only extremely specific – ranging from incandescent lamps to bearings and gears and design[2] – but also represented six separate 2-digit NACE codes. It seemed highly unlikely that matching those specific categories across a range of countries would yield productive insights into the evolution of global innovation networks: Even before gathering any survey data, it seemed as if Germany had a clear advantage in those niche areas, and that innovation globally would be likely to be in complementary areas.

Brazil did not use industrial categorization at all as a starting point for defining the Brazilian automotive industry. Instead, partners attempted to deal with the challenge of multiple industries differently by combining the somewhat out-of-date information of the auto parts union, SINDIPECAS, with the official Annual Registry of Social Information (RAIS). They also supplemented the list by conducting interviews with employees at some of the largest automotive firms and asking them for details about their suppliers. After having identified the specific firms working in automotive, the Brazilian partners were able to link firms to specific (sub-) sectors[3], but those subsectors were not instrumental in defining the boundaries of the industry. Instead, the heuristic that they Brazilian partners used was whether a given firm was, to put it simply, contributing to building a car.

Italy was able to add a number of the questions from the Ingineus questionnaire to a survey conducted annually through the Chamber of Commerce of Turin, where the Italian automotive industry is

[1] The Norwegian partners decided to focus on computer services, NACE J62 with 756 firms in the national business registry, the Brønnøysund Register Centre.

[2] The German automotive firms were in NACE 1 25241, 28408, 29140, 31610, 34300, and 74205

[3] About two-thirds of the Brazilian automotive firms were in NACE 1 34, but the other firms were distributed across a wide range of other industries.

concentrated[4]. Because the contact was made through an institution concerned with facilitating business, the (large) Italian sample therefore consisted of firms that self-selected as belonging to the automotive industry, without reference to their core technological activities. Although the selection process is different to that of the Brazilians, the underlying principle governing the sample frame is the same: Whether firms provide some inputs, whatever they may be, to the building of a car.

The only partner country that used an industrial code classification as the basis for their automotive sample frame was Sweden. Swedes targeted two NACE 2 codes: 29.31 (Manufacture of electrical and electronic equipment for motor vehicles) and 29.32 (Manufacture of other parts and accessories for motor vehicles) to capture the Swedish automotive industry. The Swedish team had sent out questionnaires to both ICT and automotive firms, and the team reported that a small number of firms' self-identified as ICT when the database of Statistics Sweden had identified them as automotive firms, and vice versa. Given the challenges encountered by the German, Italian and Brazilian researchers, this fluidity is hardly surprising.

Indeed, this fluidity in the definition of the core industry usefully highlights that both supply-side and demand-side considerations can be used to delineate an industry, and that in a given industry one approach may be more appropriate than the other. In the case of the automotive industry, the most appropriate way of drawing the boundaries of the industry seemed to be by looking at the demand-side. Automotive firms are firms across a large set of technologically diverse industries that provide inputs resulting in the production of a motor vehicle. In contrast, in ICT, the choice of sectors and sub-sectors was best made with reference to supply-side considerations because ICT, as a General Purpose Technology, is used by virtually every firm in every industry. Using demand-side considerations would likely have resulted in the inclusion of a number of firms that used ICT, but was extremely unlikely to ever innovate in ICT.

The most established and least research-intensive industry of the three, agro-processing, seemed to present an even greater challenge. There was firstly a demand-side criterion, that the end product be consumable by humans. This criterion led to the exclusion of a wide range of industries that processed agricultural products, but not for human consumption, like the manufacture of ethanol from corn, cotton from cotton plants, or paper and furniture from pine and other plantations.

But there was also a supply-side criterion, in the sense that there was almost by virtue of the definition an expectation of some form of industrial processing. In a low research-intensive industry, it is a given that innovation is not necessarily tied to narrowly defined technological advances (Tunzelmann & Acha, 2005). The researchers therefore wanted to draw the boundaries of the industry using an inclusive definition of innovation and processing, so that agro-processing could also refer to, for example, a packing plant in South African that met the quality and traceability requirements of the European Union well enough to allow organically grown apples to be boxed and exported to Europe.

[4] The automotive industry has been extensively researched, and many automotive firms suffer from "survey fatigue". The Chamber of Commerce of Turin surveys the automotive industry annually, and because the institution administering the questionnaire is seen as a mouthpiece for business and the company running the survey has accumulated experience in running the survey, the survey has a relatively high response rate and number of responses (about 800). Moreover, because the survey has been run on a relatively stable sample for some time now, some comparisons over time are possible. It was therefore very attractive to send out the INGINEUS questions as part of the annual Turin Chamber of Commerce auto survey, but it was not possible to include all the questions from the INGINEUS project. In short, the decision of the Italian researchers was to sacrifice some depth for breadth.

However, our expectation was interpreted differently by many of the firms we had approached. Butchers and bakers would claim that they were not really using "technology", reflecting a view that "technology" could not consist of practices and processes, but only of machinery – and new machinery at that. In turn, organic farmers would insist that they should not be included in the agro-processing sector because their products were precisely not processed, even though considerable organisational and logistical innovations are typically needed to farm organically and deliver produce to markets in urban areas.

As a result, the sample frame reflected a more "conservative" definition of this industry than the research team had originally intended. Within the context of a research question on globalising innovation networks, it is probably of little consequence that a local chain of bakeries does not consider itself a part of the agro-processing industry. But if our central goal was to understand the performance of and interface between the agro-processing industries of the developing world and Europe, pre-existing views about what technology and agro-processing entail may well have hamstrung the study.

Two key insights emerged from our process of determining a sample frame for the three industries covered in the study. The first is that the increasing complexity of global supply chains has complicated the definition of industries substantially. Although industrial categories are a useful tool at a more atomic level, it is often necessary to aggregate different categories to obtain a more coherent definition of a given industry. But it is often hard to find the common elements among the very different firms that are contributors to an industry, and researchers make decisions and trade-offs in how they decide to operationalise a given industry through their construction of a sample frame, and selection of firms.

Our second insight is that it is perhaps not useful (and probably not even possible) to strive for a definitive definition of a given industry. Instead, it seems more appropriate to operationalise an industry relative to a given research question. Because our concern was with the globalisation of innovation, we regarded it as important to obtain an overview of the different activities taking place across the global value chain, even when it meant more "loosely" defining an industry. However, for a comparative study of how, for example, national institutions affect how firms are structured, it would be important to find as close as possible a match of firms in an industry. In sum, we expect that the issue of industry definition may have to get more attention in future, for example by the development of a vocabulary for better explaining on what basis aggregation took place – or not.

1.3.4 Firm size

A large number of smaller firms populated the Swedish and Norwegian ICT databases. Excluding these firms would have compromised the size of these datasets and resulted in the possible oversight of GINs as smallness may bear no relation to the ability of the firm to be innovative, global or networked. For this reason it was decided that the minimum size of a firm for the survey would be five employees. There was no upper ceiling set for the size of the firm.

1.4 Survey method

Each survey country was required to set up an online survey tool through '**Survey Monkey'**. The Survey Monkey site 'Username' and 'Password' was sent through to UP-GIBS. In this way the

survey progress could be monitored from a central site. The purpose of using a common survey tool which looked exactly the same across all survey countries was intended to assist with consistency across the project, streamline the research processes and to simplify the assimilation of the data at the end.

1.4.1 Method of delivery

The use of 'Survey Monkey', an online survey tool, made it possible to track the progress of different partner countries while responses were coming in. However, the developing countries (bar South Africa) all opted to conduct the questionnaires either in person or telephonically in order to increase response rates, and only later entered the data into the shared survey software. This meant that it was not possible to map the progress of the partner institutions with the least experience participating in EU projects.

The survey could be delivered electronically by mail or link, by face-to-face interviews, through telephonic interviews or by written mail. This choice was left up to the delivering institute based on their past experience of survey dissemination and their historical knowledge of the best methods utilised for high response rates.

The following countries used electronic channels: South Africa, Norway, Sweden, Germany, Denmark and India.

The following countries used face- to –face interviews: India, Brazil

The following countries used telephonic channels: China, Brazil

The countries using electronic channels were required to first make contact with the enterprise and get a telephonic consent indicating that the survey link could be mailed through to the enterprise. If the company could not be contacted or refused permission then the survey was not mailed to that enterprise on the database. This pre-survey consent process resulted in a significant reduction in the number of surveys being sent out as compared with the full database size. (For detail on the original database name and size please refer to the individual country reports in Section 2)

In the case of India, past experience of poor electronic response rates motivated the team to visit companies and complete the questionnaire face-to face. In order to bolster their results the Indian team at CDS also sent out mail surveys to their database. This channel of survey dissemination garnered 31 additional responses.

1.4.2 Data processing

Each survey partner was required to enter all survey responses into Survey Monkey even if their surveys were not conducted online. All partners then cleaned and checked responses prior to downloading the data. The downloaded files were then mailed to WP 2 in a common Excel format for analysis. The WP 2 statisticians assimilated all the responses into a single spreadsheet. Following this a finalised dataset and codebook were assembled along with a set of reports and pivot tables based on the various levels of analysis.

1.4.3 National versus regional sample

The Swedish, South African, Norwegian, Danish and German surveys were national. Due to the size and geographical spread of the population in Brazil, China and India, these surveys were regional.

One of the assumptions made in the survey methodology was that the survey would be carried out on a nationally representative database. In the developed countries of Sweden, Norway and Denmark, databases were comprehensive and regularly updated. These countries are also characterised by high IT penetration, excellent national IT infrastructure and a culture of survey completion, all of which contribute to good survey response rates online.

In Brazil, the survey partner was required to create a composite database built from 3 separate databases (RAIS, SINDIPECAS and Supplier case list) in order to create a representative sample. Further complicating matters was the historically poor response rate to mailed and electronic surveys. This meant that interviews would have to be conducted face-to- face. The survey was thus confined to the region of Minas Gerais. The Brazilian automotive industry is however concentrated in the region of Minas Gerais therefore the majority of the relevant auto firms were represented in the databases.

In South Africa a composite agro-processing database was also created from 4 separate databases as the databases acquired were not updated and had a large number of invalid contacts.

In India, historically, a difficulty with low electronic response rate was experienced requiring that a face-to-face interview strategy be implemented. Since it was not viable to conduct this nationally, a regional profiling of the NASSCOM database was undertaken. Cities with IT dense clusters were chosen as targets for the survey. These cities represented 93% of all the firms in the database. These cities included Bangalore, Delhi, Mumbai, Pune, Trivandrum, Hyderabad and Kochi.

In China the vastness of the geography and the sheer number of ICT firms nationally made it very difficult to access and approach firms with the survey. In China, face-to-face interviews or telephone interviews were found to offer the highest response rate. These challenges necessitated a regional approach therefore two regional databases were used, one focussing on Beijing and the other on Shenzen.

The Italian survey is carried out in collaboration with the Turin Chamber of Commerce. This ‘piggy back’ strategy was adopted as it greatly enhances response rates. This meant however that the format of the questions differed slightly from the rest of the survey partner countries. The Italian data will therefore be compared against the assimilated and integrated data set from the rest of the partners.

The Germans experienced a low response rate of 8 percent and requested an extension on order to target a second group of smaller firms. Despite the German sample being small and having a poor response rate the German sample was national.

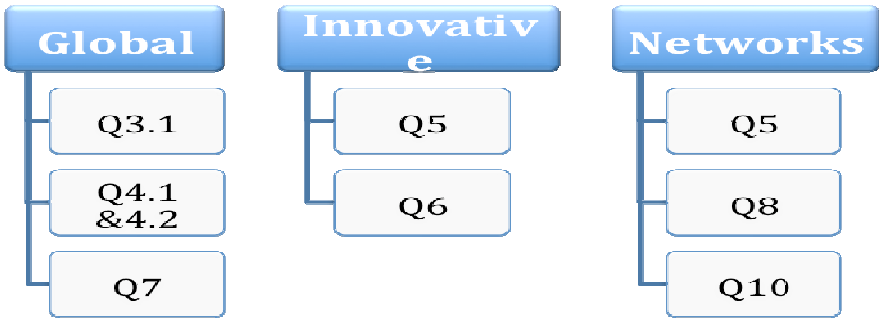
1.5 Structure of the questionnaire

The survey instrument consisted of 14 questions. Many of the questions had multiple sub questions. The questionnaire was structured to elicit information on firm behaviour around a clearly defined set of theories.

- **-Question 1** asked the respondent to briefly describe the enterprise's main product (goods or services)**,** the respondent was then directed to a menu which allowed them to select the option which best described the firm's main area of focus (the NACE code and description were tabulated).
- **-Question 2 to 4** elicited background information about firm size, market, sales information and R&D activity.
- **-Question 5 and 6** were innovation based questions.
- **-Questions 7 and 8** probed the firms geographic network and collaborations with customers, suppliers, Universities, research institutions, government etc.
- **-Question 9 and 10** were detailed questions around offshoring and regional attractiveness.
- **-Questions 11, 12 and 13** were policy based questions
- **-Question 14** examined the impact of the global economic crisis on innovative activity.

The overarching goal of the survey was to establish the presence of GINS therefore how global, how innovative and how networked the sample was. These three themes could be mapped to specific questions as illustrated in Figure 2 below.

Figure 2: Graphic representation of INGINEUS themes and matching survey questions



1.5.1 Levels of data analysis

The survey was designed in order to allow for various levels of analysis, these being:

- Country analysis
- Sector analysis

- Developed/developing analysis
- Assimilation across all countries

1.5.2 Survey timelines

- Zebula South Africa March 2009: Methodology workshop in Johannesburg hosted and run by GIBS to begin the bottom up process of methodology development. Thirty participants.
- July 2009 Finalisation of survey questionnaire
- July 2009 Set up of online survey tool, Survey Monkey
- August Survey pilot launch
- April 2010, all surveys are closed
- June 2010 data from surveys is collected
- September 2010 the first assembled and assimilated data analysis is conducted
- October 2010 the final cleaned and checked dataset is disseminated for online downloading off the password protected section of the INGINEUS site.

The survey was launched in Sweden, Norway and Denmark before October 2009 as these countries databases were updated and ready to use. The survey was launched in late January 2010 in Estonia, South Africa, India, China and Brazil as the database management was far more complex and demanded greater preparation.

1.6 Specific challenges

The definition of "region" proved surprisingly complex – in the smaller countries, a region would often be associated with a single city and surrounds, while large countries like India, China and Germany had a very different understanding of how big a region would be. This concept was fully explored in the Methodology paper: Making methodological decisions in cross-country, cross-industry research: Articulating the assumptions underlying indicators by Barnard &Ismail to be presented at the 8th International Globelics Conference. An extract from this paper is included below to explain the concept of 'region' as understood in this project in greater detail.

'Scholars have increasingly been concerned that the concept of a national innovation system operates at a very high a level of analysis, and have started to examine so-called regional systems of innovation (Cooke, 2001, Padilla, Vang-Lauridsen, & Chaminade, 2009). A core differentiating characteristic of regional systems of innovation is embeddedness – it is possible for firms to develop close enough relationships with each other and with enabling institutions such as universities and funding organisations for there to be the potential for the exchange of tacit knowledge. Most of the studies of such regions have focused on a specific industry in a specific geographic region, and comparability was not such a big concern. In a study of the Netherlands, Beugelsdijk and Cornet (2002) challenge the importance of proximity (except in the case of universities) without defining "far" and "near".

Almeida and Kogut (1999) in the US use a very rigorous approach when they determine regions by matching county level data with economic activity evidence. However, it is extremely hard to find a shared, single yardstick like Almeida and Kogut did when a study examines firms across not only national boundaries, but also from countries at very different levels of development. Both the political and the economic points of reference in the different locations are very different. The Ingineus project was going to use self-reported data, but it was still necessary to define 'region' in some way. In particular, it was necessary to find some easily understood word or phrase that would be meaningful to executives, and adequately capture the fairly complex concept of a location with a number of related firms and supporting institutions fairly close to each other.

While exploring this issue, a number of non-theoretical observations were made. Some questions were on how big a region typically was, and how close 'close' needed to be. The commonly used heuristic of "commuting distance" was mentioned, but the Danish and Estonian sense of distance still seemed quite different to the sense of distance in larger countries like Brazil or Germany. It seemed that when people identified a region, they considered the presence of certain resources, e.g. universities, not in absolute terms but by their relative presence, by considering how close the alternatives were. It also seemed as if the presence of abundant other entities led to people defining smaller regions than when those entities – whether firms or supporting institutions – where scarce.

Thus in Sweden with its better developed and denser institutional infrastructure, a relatively small city like Lund with its immediate surrounds was considered a region, whereas in both India and China, the region seemed to be best represented by a large city and the area around it – around Pune and Bangalore in India, and Beijing and Shenzhen in Guangdong province. In contrast, in less densely populated developing countries like South Africa and Brazil, the region was often equivalent to a province.

The group of research partners had not expected 'region' to prove such an elusive concept. A question that came up in the discussions around the definition of region is whether and how much nuances and more substantial differences in different contexts matter. Does it have a material impact – and to whom – if different partner countries define regions differently? In the context of the Ingineus project, we answered the question by referring back to the research question: Given that our focus was global innovation networks, we did not want to become side-tracked by the definition of sub-national regions. We therefore followed the example set by previous researchers, and defined 'region' somewhat loosely.

Ultimately, we decided to use the term "region" in the questionnaire, and to explain that region was a "sub-national area". We also signalled through the layout of the question – first "your region", then "your country" and then a list of larger territories like "North America" that we wanted information about a unit that was operating at a more intimate level than the national level. In truth, while we have gathered useful information about the functioning of innovation at a "sub-national" level, we cannot provide information on the boundaries of a region. But because research builds on prior work, we have further entrenched an inherited loose definition.'

1.7 Analysis of global innovation networks

Before detailed information is provided, it may be useful to provide an overview of the process. For each of the three concepts (Global, Innovative and Networked), relevant questions in the survey were chosen and then weighted according to their importance. A scoring system was devised, and a formula specified which gave each instance in the dataset a continuous value greater than or equal to 0. This value was divided by the maximum value in the dataset, so that each instance had a continuous score between 0 and 1, with the instance with score 1 being that which most epitomised the concept in question. This resulted in each instance being scored relative to the other instances in the dataset.

These scores were displayed on a scatter plot, and a combination of cluster analysis and inspection of the scatter plot used to identify the cut-off point between categories, e.g. highly global, somewhat global and not at all global. Alternative scoring systems were explored to test the robustness of the original scoring. Once the scoring was determined, each instance in the dataset was classified as one of the types of GINs.

1.7.1 Globalness

The purpose of this measure is to establish globalness (rather than innovativeness or networkedness), and it was therefore deemed important to not give greater weight to more "complex" activities (like innovation) than to "simpler" tasks like exports – what matters is global reach. We therefore considered all questions that asked respondents about the locational spread of their activities, regardless of what those activities were.

Table 2: Questions used for calculation of globalness

Question from survey	**Possible answers**	**Points**
3.2.1. Please provide the percentage (%) of total sales derived from export.	Continuous value *p* between 0 and 100	*p*/100
4.2. Indicate the three most important destinations in terms of sales: 4.2.1. North America 4.2.2. South America 4.2.3. Western Europe 4.2.4. Central & Eastern Europe 4.2.5. Africa 4.2.6. Japan & Australasia 4.2.7. Rest of Asia 4.2.8. Rest of the world (developing)	Yes Yes* No *If the answer reflects a regional focus, e.g. Africa for a South African firm, then it was coded as * and given only half the weight. An extra-regional response, e.g. North America was given full weight.	1 0.5 0 To a maximum of 3 points (for three destinations)

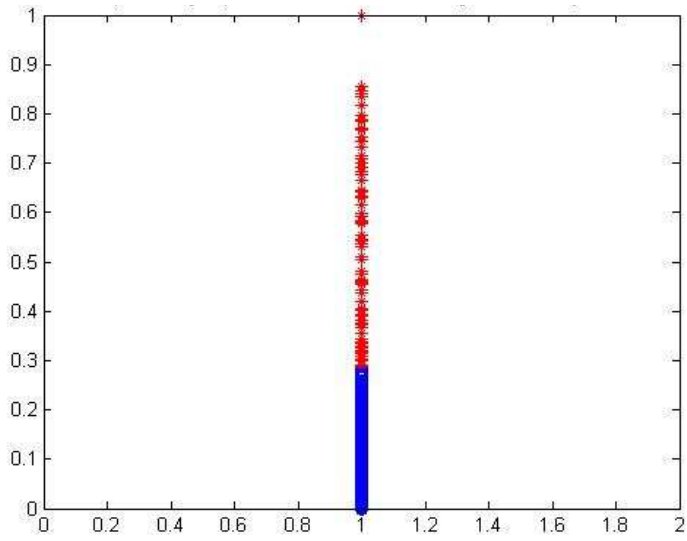
7. Regarding the development of the most important innovations of your firm in the last three years: who did you actively collaborate with and in which geographical location? 7.1. Clients 7.2. Suppliers 7.3. Competitors 7.4. Consultancy companies 7.5. Government 7.6. Local universities / research institutions / labs *(by definition local and therefore excluded from measure)* 7.7. Foreign universities / research institutions / labs 7.8. Other (please specify). The regions are: 1. Your region *(by definition local and therefore excluded from measure)* 2. Your country *(by definition local and therefore excluded from measure)* 3. North America 4. South America 5. Western Europe 6. Central & Eastern Europe 7. Africa 8. Japan & Australasia 9. Rest of Asia	Yes Yes* No *If the answer reflects a regional focus, e.g. Africa for a South African firm or Europe for a Swedish firm, then it was coded as * and given only half the weight. In contrast, an extra-regional response, e.g. North America for either of these examples, was given full weight.	1 0.5 0 To a maximum of 42 points, (6x7), since all regions can be selected for all sub-questions, excluding sub-questions 7.6. and 7.8. (other)
9.1. Regarding internationalisation, does your firm offshore production or any R&D activities?	Yes No	1 0
10. Please indicate how the following functions are performed by your enterprise, including different subsidiaries of the same firmStrategic Management 10.2. Product Development 10.3. Marketing, sales and account management 10.4. Operations 10.5. Procurement, logistics, distribution 10.6. Corporate governance 10.7. Human Resource Management 10.8. Technology and process development 10.9. Firm infrastructure 10.10. Customers and after sales service. The regions are:	Yes No For sub-questions 10.1-to 10.10, options 1 and 4 are left out because they deal with local locations. Options 2 and 3 are not clear; those locations can include subsidiaries at home. Therefore only options 5 and 6 are used.	1 0 To a maximum of 20 points, since both regions 5 and 6 may be selected for each of the ten sub-questions

1. By your unit in your location 2. At subsidiaries of firm in a developed location 3. At subsidiaries of firm in a developing location 4. Outsourced to a partner in your country 5. Outsourced to a partner outside your country in a developed location 6. Outsourced to a partner outside your country in a developing location.		

After transforming each value so that they all had a score between 0 and 1, all five categories listed in the table were used to calculate an average. For the robustness test an average was calculated where questions 4.2 (regarding sales) and 7 (regarding innovation) were given greater weight. Those questions are more fine-grained and force the respondent to state precisely which regions are involved.

We use k-means cluster analysis with two groups and the squared Euclidean distance as the distance measure between points. The silhouette plot for the analysis where greater geographical distance has greater value is shown below. The red markers indicate Cluster 1 and the blue markers indicate Cluster 2. The mean of Cluster 1 is 0.5178 and the mean of Cluster 2 is 0.0552. Looking at the scatter plot, the value 0.283 is a natural break point and we classify all instances >0.283 as G, all instances greater than 0 and up to 0.283 as g, and all instances of zero as *, with G denoting truly global, and g denoting somewhat global.

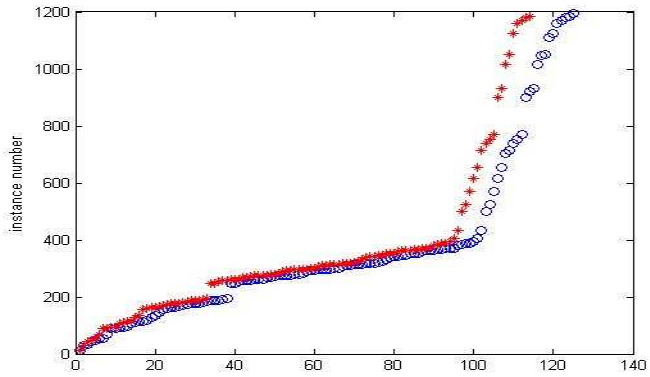
Figure 2: Distribution of values for globalness using equal weights



A similar process for the model where all instances of globalness are given equal values results in a cut-off point for >0.27 as G, and for all instances greater than 0 and up to 0.27 as g.

Comparing the two models, we observe that these two formulae (based on different questions) give similar groupings. Numerically, 99.09% of all 1215 instances in the dataset have the same value under each of the models. This implies that the scoring system for globalness is robust.

Figure 3: Robustness check of two models for globalness



1.7.2 Innovativeness

With regards to innovation, respondents were asked to indicate if they have innovated in 2006 to 2008 in any of five categories:

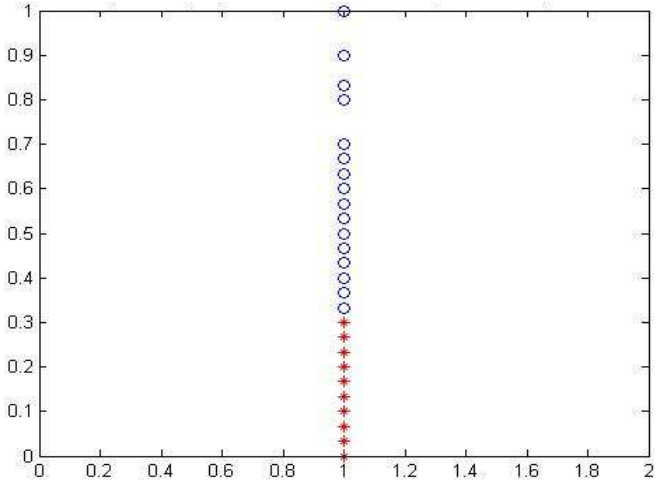
- New products
- New services
- New or significantly improved methods of manufacturing or producing
- New or significantly improved logistics, distribution or delivery methods for your inputs, goods and services
- New or significantly improved supporting activities for your processes (e.g. purchasing, accounting, maintenance systems, etc.).

For each of the options selected, the respondent was asked to indicate if the innovation was new to the world (which was given a value of 3), new to the industry (with a value of 2) or new to the firm (with a value of 1). This yielded a maximum score of 15. However, that scoring system implies that there is a linear progression from new-to-the-firm to new-to-the-industry to new-to-the-world innovations, whereas it may be significantly more complex to generate more novel innovations. To test for robustness, all scores for "new to the world" are multipled by 3 (to a maximum of 9), and all scores for "new to the industry" by 2 (to a maximum of 4). This approach provides greater weighting by degree of innovativeness.

We first do a cluster analysis using the linear scale. The red markers indicate in Figure 4 indicate Cluster 1 and the blue markers indicate Cluster 2. There seems to be a break at around 0.7. However, this is a very strict cut-off point, as less than 2% of the values fall above this point. Therefore, we choose the next most obvious cut-off point (by inspection), which is just below 0.6. The values get much denser below this point, and increasingly sparser above this point. We classify

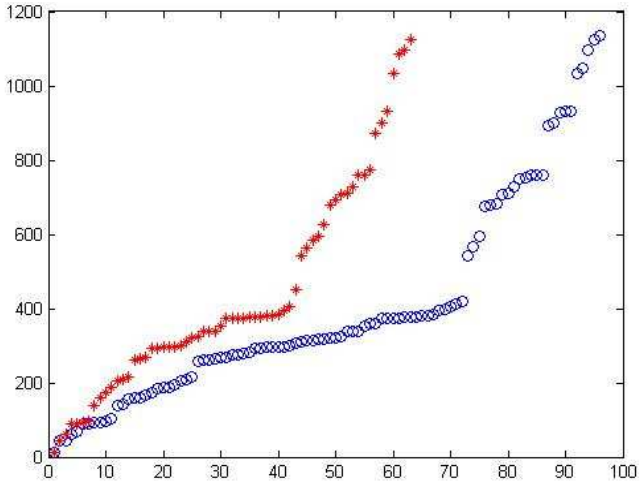
all instances >=0.6 as I, all instances between 0 and 0.6 as i, and zero as *, with I denoting "Innovative" and i denoting "somewhat innovative".

Figure 4: Distribution of values for innovativeness using linear scale



To test the robustness of the cluster analysis, we do a similar analysis, but one where innovations are given much greater weightings for greater novelty. The graphical representation of the comparison indicates that although the two sets of markers are not one on top of the other, they follow the same general trends. Since the y-axis denotes the instance number, it is clear that many of the same instances occur for the two formulae, although the ordering may be slightly different (as each formula has a slightly different number of instances classified as "I"). In other words, since the markers for both scores appear on the same *horizontal gridlines*, the two scoring systems must classify most of the instances in the same way. Doing a logical check, we find that 95.72% of the values for the two scores are identical. This suggests that the scoring system for innovativeness is robust.

Figure 5: Robustness check of two models of innovativeness



1.7.3 Networkedness

In operationalising the concept of networkedness, we considered debates about the indicators of a "strong" network. Formal linkages may be seen as especially strong, as they provide the benefit of legal protection (Zhou & Poppo, 2010). However, there is also an argument that trust may be reduced by formal control mechanisms (Das & Teng, 1998; Malhotra & Murnighan, 2002) and that informal linkages may signal especially strong relationships. Similarly, although it is plausible that the strongest network would be within the firm – where people share an organisational culture and goal – it is also possible that a firm may be less inclined to take for granted and therefore take more care to nurture important external networks.

We therefore incorporate two measures of connectedness, span and depth. An enterprise is highly networked firstly if it has connections or relationships with many other people, enterprises or institutions. The more connections which an enterprise has with people or bodies outside of the enterprise itself (e.g. clients, suppliers, competitors, universities, etc.), the larger is the span of the network. Secondly, an enterprise is highly networked if those connections or relationships are deep. A deep connection is one which is meaningful or even crucial to the running, development or success of the enterprise.

In developing the measures, we considered both internal/external (to capture span) and formal/informal linkages (to capture depth). We calculate three scores for networkedness, one where all scores are given equal weighting, one where external linkages are given greater weight than internal linkages, and one where formal linkages are given greater weight than informal linkages.

Table 3: Questions used for calculation of networkedness

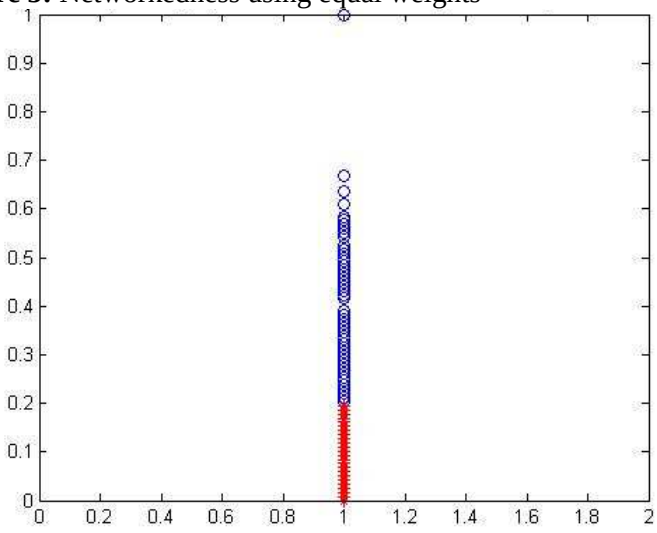
Question from survey	Possible answers	Points
7. Regarding the development of the most important innovations of your firm in the last three years: who did you actively collaborate with and in which geographical location? 7.1. Clients 7.2. Suppliers 7.3. Competitors 7.4. Consultancy Companies 7.5. Government 7.6. Local universities / research institutions / labs 7.7. Foreign universities / research institutions / labs 7.8. Other (please specify). The regions are: 1. Your region 2. Your country 3. North America 4. South America 5. Western Europe 6. Central & Eastern Europe 7. Africa 8. Japan & Australasia 9. Rest of Asia	Yes No The regions themselves do not matter. Each region specified indicates a new connection with people in that region. For this reason, each region will be awarded a point for networkedness	1 0
8. Has your enterprise developed formal/informal linkages (e.g. research relationships) with the following kinds of foreign organizations? 8.1. Clients 8.2. Suppliers 8.3. Competitors 8.4. Consultancy companies 8.5. Government 8.6. Foreign universities / research institutions / labs 8.7. Other (please specify) The options are:	Formal Informal Both Neither	1 / 2 1 / 1 2 / 3 0 / 0 Both a formal and an informal linkage indicate the presence of a connection, and they are initially equally weighted. For the robustness test, formal linkages are more heavily weighted

• Yes, formal • Yes, informal • No		
10 Please indicate how the following functions are performed by your enterprise, including different subsidiaries of the same firm: 10.1 Strategic Management 10.2 Product Development 10.3 Marketing, sales and account management 10.4 Operations 10.5 Procurement, logistics, distribution 10.6 Corporate governance 10.7 Human Resource Management 10.8 Technology and process development 10.9 Firm infrastructure 10.10 Customers and after sales service The regions are: 1. By your unit in your location 2. At subsidiaries of firm in a developed location 3. At subsidiaries of firm in a developing location 4. Outsourced to a partner in your country 5. Outsourced to a partner outside your country in a developed location 6. Outsourced to a partner outside your country in a developing location	For option 1 (own unit) no points were given For options 2 and 3 (internal network): Yes No For options 4 to 6 (external network): Yes No	1 / 1 0 / 0 1 / 2 0 / 0 Both an internal and an external linkage indicate the presence of a connection, and they are initially equally weighted, and for the robustness test, exxternal linkages are more heavily weighted To capture the internal/external dimension, sub-questions 7.8 and 8.7 are also used. The entity indicated is considered, e.g. a subsidiary would be regarded as an internal but a broker as an external connection.

Figure 5 maps values for networkedness with an equal weight for all indicators. The red markers represent Cluster 1 and the blue markers represent Cluster 2. Although the figure indicates that the two clusters are separated around 0.2, the data points at the break for clusters 1 and 2 are very close together – almost on top of each other. At the same time, the above plot shows a slight gap in data values around 0.4. Looking at the scatter plot, this seems to be closer to where the natural break occurs. Therefore we reject this cluster analysis, and rely instead on inspection of the scatter plot in order to decide on a natural break in the data values. Taking into account that the percentage of

values greater than 0.2 = 15.3909%, greater than 0.28 = 8.9712%, greater than 0.32 = 6.6667% and greater than 0.4 = 3.7860%, we consider a natural break at 0.32 (by inspection). Following this model, we classify all instances >0.32 as N, all instances with a value greater than 0 and as high as 0.32 as n, and all instances of zero as *, with N denoting “truly networked” and n denoting “somewhat networked”.

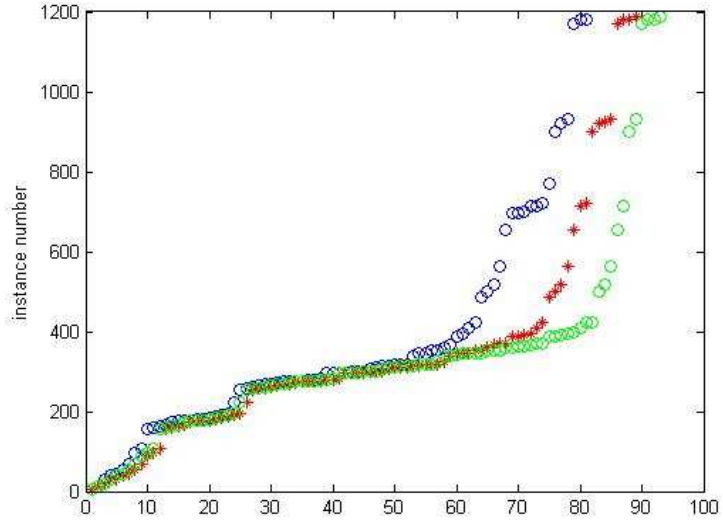
Figure 5: Networkedness using equal weights



The second and third models emphasize the relative formality of linkages, and external scope of linkages respectively. Each time, using the same process described before, it is calculated which respondents can be considered truly networked, somewhat networked, and not at all networked.

Figure 6 compares the three models. In most instances, the same value is obtained regardless of which model us used. Numerically, 97.2% of all 1215 instances in the dataset have the same scoring for networkedness. This implies that the scoring system is robust.

Figure 6: Robustness of scoring for networkedness.



1.7.4 Results: Global Innovation Networks

The fact that the indicators for globalness, innovativeness and networkedness proved to be robust to different operationalisations of each construct suggests that they tap into robust constructs.

Using the calculated scores, we classify each firm within one of the types of global innovation networks. We use a capital letter to indicate that the firm is highly global (G), highly innovative (I) or highly networked (N), and small letters if the firm has been classified as somewhat global (g), somewhat innovative (i) or somewhat networked (n). Finally, we use an asterisk (*) in cases where a firm is not at all global, innovative or networked. Mathematically, twenty-seven (3x3x3) permutations are possible, but to the extent that firms are engaging in some form of GIN not on a random basis, but because of an underlying logic, we expect that only some combinations will be seen.

The results indicate that there is an underlying logic for firms' behaviour. Certain combinations are not found – it is extremely rare to find a firm scoring highly on one dimension, and not at all on another dimension.[5] In fact, only twelve of the possible 27 categories account for more than 97% of the dataset, and it is possible to combine those twelve categories into six main types. The types are presented in Table 4.

In addition, there are indeed some strong-form GINs. They represent only 15 firms (just more than 1%) in the sample, but given the emergent nature of the phenomenon, this is to be expected. The strong-form GINs are discussed in more detail later.

[5] In terms of how we designate types, it virtually never happens that a firm would be described with both an asterisk and a capital letter.

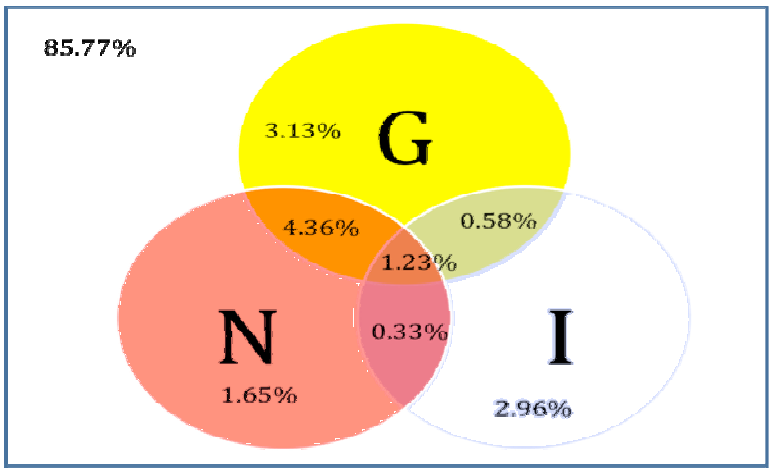
Table 4: Types of GINs

Type of GIN	Description	Values
Balanced GINs	All the elements are in alignment	GIN (1.23% of sample) gin (40.41%) *** (12.18%)
Global asset exploiters	Global reach is greater than the extent of innovation or networkedness	Gin (2.96%) g** (1.65%)
Innovators	Firms are relatively more innovative than their global reach or the extent of their networks would suggest	gIn (2.63%) *i* (1.89%)
Networkers	Strength of networks is greater than global reach or innovativeness	giN (1.48%) **n (5.76%)
Global networkers	Innovation is not as high as both the globalness and the networkedness. This is the only common combination of two stronger dimensions	GiN (4.36%) g*n (3.79%)
Domestics	Firms that have no supra-national footprint at all, but are innovative and networked enough to (presumably) survive domestically or locally – this category accounts for the second largest group of firms.	*in (18.93%)

The greatest proportion of firms, 40% of the sample, consists of firms that are somewhat global, somewhat innovative and somewhat networked, and the third most commonly found category (12% of the sample) of firms that are not at all global, innovative or networked. These firms are all "balanced", in that their globalness, innovativeness and networkedness are at an equal level of development.

Almost a fifth of the dataset (the second-largest category overall) consists of firms that have no supra-national connection at all, but are still somewhat innovative and networked. These firms are clearly focused on a domestic market. But for the categories of global asset exploiters, innovators and global networkers, the firms that are *somewhat* global, innovative and globally networked are outnumbered by those with *high* scores on those dimensions. It seems that there could be some kind of momentum or logic by which it is easier for firms to have intensive than somewhat global, innovative and/or networked behaviour when they participate in a global innovation network, even when it is not yet a stronger form GIN.

Figure 7: Distribution of stronger form GINs



Mapping the entire dataset is useful in order to quantify the relative importance of GINs, and it provides evidence that about 15% of the firms in the dataset are truly global, innovative and/or networked. These firms belong to various stronger forms of GINs, and it is worth investigating the characteristics of the stronger form GINs.

Table 5 provides evidence of some core characteristics. The global asset exploiters and global networkers have a similar distribution in terms of both size (large firms) and firm type – mainly the subsidiaries and headquarters of MNCs. Among the global asset exploiters, the European locations are relatively well represented. These firms seem to follow a fairly traditional model of market-seeking expansion. In contrast, the global networkers is the single category where developing country firms are most prevalent – with almost 7% of all developing country firms in the dataset represented in this category. Networkers are also large firms, also predominantly subsidiaries and headquarters of MNCs, but firms from developing countries are not as readily found as among the global networker category.

The comparison between networkers and global networkers is useful because the main dimension of difference is the scope of the network. It is telling that the developing country firms are so much more global, and that high levels of globalness and networkedness co-occur, but not innovativeness. This pattern is consistent with previous evidence about the relatively lower innovativeness of developing country firms. We suggest that the less munificent institutional context of entities in less developed countries is an important explanatory factor in their strong drive for global networking.

In contrast, innovators are more often from Europe than any other category, more often small (less than 50 employees) standalone firms, and more likely to generate new to the world product and/or service innovations than any other category. It seems that these players are most able to draw on an appropriate regional institutional infrastructure, and that these players are organisationally small enough for them to focus on customer-focused innovations. If growth derives primarily from innovation in new to the market offerings, these are critical firms in an economy. However, there is limited evidence of value capturing, as innovators have a very low proportion of exports and international clients. This raises the question of whether there is a ceiling to their economic value.

Of the fifteen firms that are highly global, highly innovative and highly networked (which we will call strong-form GINs), two are in the agro-processing industry, and the other thirteen all in ICT. This to some effect reflects the dominance of ICT in the dataset, although this result is also quite coherent with the literature that has long argued that globalization is more likely to occur in some industries than in others, due to the different nature of their knowledge bases (Pavitt, K. 1984; Asheim, B. and Gertler, M. 2005). The fact that no automotive firms (with their strong tiered supplier structure) are strong-form GINs is also consistent with that evidence.

As regards the size distribution of the strong-form GINs, one very small firm is found, and the others range in size from 50 to more than 1000 employees. This is smaller than would be the case for most traditional industries (e.g. much of manufacturing), and suggests that there may be a current optimal point in terms of number of employees in terms of the complexity of managing a GIN. Those firms with a global footprint (global asset exploiters and global networkers) that are only somewhat innovative are generally large firms with 1000 plus employees, and those firms that are innovative but with a limited global footprint tend to be very small (around 50 employees). In contrast, the strong-form GINs have a considerable footprint, although they have clearly not internalised all activities. This could also be related to the fact that the majority of firms are in ICT, which has a stronger skills than labour component and often fewer in-house employees.

The location of the strong-form GINs is somewhat surprising. One strong-form GIN is found in China, two in South Africa, and eleven in India. Five of them are the subsidiaries of advanced (and in fact, US) MNCs in India, as is the single Chinese strong-form GIN. But an additional five of the strong-form GINs are subsidiaries or headquarters of emerging MNCs, and four more are stand-alone firms. Apart from the Norwegian firm, the only European participation in this list of strong-form GINs is through two of the emerging MNCs whose subsidiaries are represented have dual headquarters, both in their country of origin and in a European country.

The evidence suggests that it would be wrong to regard strong-form GINs as the domain primarily of the most advanced MNCs of the developed world. Strong-form GINs seem to have two origins: Some are advanced MNCs evolving into GINs, who are able to manage the complexity of a global network and achieve substantial innovation. The other strand is of developing country firms that have long had the global networks, but are also achieving true innovation.

In terms of industry, the auto industry has a strong showing in two categories – innovators and global networkers, but it does not have any strong-form GINs. The fact that firms are either capable of strong innovation, or of global networking, suggests that there may be some trade-off between managing advanced innovation, and managing extensive global networks. In addition, it seems that there are “integrator firms” in the industry that are tasked with global sourcing and integration of innovations that come from specialist innovative suppliers, and this most likely links to different positions in the value chain

Table 5: Characteristics of firms participating in stronger form GINs

Global asset exploiters, 38 cases, 3.13% of dataset

Industry	#	% of all firms in that industry	Country	#	% of all firms in that country	Size	#	% of all firms of that size	Firm type	#	% of all firms of that firm type	Location of responding unit	#	% of all firms of that type in that location
Auto	3	2.05%	China	2	0.82%	<10	2	1.53%	Standalone	16	2.31%	Developing	10	2.98%
Agro	2	1.50%	India	26	8.02%	<50	7	1.94%				Europe	6	1.83%
ICT	33	3.53%	South Africa	1	1.19%	<250	13	4.39%	Subsidiary	16	6.50%	Developing	15	7.61%
			Developing	**29**	**4.03%**	<1000	8	4.71%				Europe	1	1.79%
			Denmark	1	2.04%	>1000	7	6.80%	MNC HQ	6	4.44%	Developing	4	3.39%
			Germany	2	3.77%	No info	1	0.65%				Europe	2	11.11%
			Norway	3	1.66%									
			Sweden	3	1.54%									
			Europe	**9**	**1.82%**									

Innovators, 36 cases, 2.96% of dataset

Industry	#	% of all firms in that industry	Country	#	% of all firms in that country	Size	#	% of all firms of that size	Firm type	#	% of all firms of that firm type	Location of responding unit	#	% of all firms of that type in that location
Auto	5	0.53%	Brazil	4	5.80%	<10	1	0.76%	Standalone	20	2.89%	Developing	12	3.57%
Agro	2	1.50%	China	3	1.23%	<50	13	3.60%				Europe	8	2.44%

ICT	29	3.10%	India	17	5.25%	<250	13	4.39%	Subsidiary	10	4.07%	Developing	8	4.06%
			South Africa	1	1.19%	<1000	5	2.94%				Europe	2	3.57%
			Developing	**25**	**3.47%**	>1000	4	3.88%	MNC HQ	5	3.70%	Developing	4	3.39%
			Denmark	1	2.04%							Europe	1	5.56%
			Norway	5	2.76%				No info	1				
			Sweden	5	2.56%									
			Europe	**11**	**2.22%**									

Networkers, 20 cases, 1.65% of dataset

Industry	**#**	**% of all firms in that industry**	**Country**	**#**	**% of all firms in that country**	**Size**	**#**	**% of all firms of that size**	**Firm type**	**#**	**% of all firms of that firm type**	**Location of responding unit**	**#**	**% of all firms of that type in that location**
Auto	2	1.37%	India	10	3.09%	<10	1	0.76%	Standalone	7	1.01%	Developing	2	0.60%
Agro	3	2.26%	South Africa	3	3.57%	<50	4	1.11%				Europe	5	1.52%
ICT	15	1.60%	**Developing**	**13**	**2.29%**	<250	5	1.69%	Subsidiary	8	3.25%	Developing	7	3.55%
			Germany	2	3.77%	<1000	3	1.76%				Europe	1	1.79%
			Sweden	5	2.56%	>1000	5	4.85%	MNC HQ	4	2.96%	Developing	4	3.39%
			Europe	**7**	**1.41%**	No info	2	1.30%				Europe	0	0.00%
									No info	1				

Global networkers, 53 cases, 4.36% of dataset

Industry	#	% of all firms in that industry
Auto	3	2.05%
Agro	4	3.01%
ICT	46	4.91%

Country	#	% of all firms in that country
Brazil	2	2.90%
India	45	13.89%
South Africa	3	3.57%
Developing	**50**	**6.94%**
Germany	2	3.77%
Sweden	1	0.51%
Europe	**3**	**0.61%**

Size	#	% of all firms of that size
<10	1	0.76%
<50	3	0.83%
<250	14	4.73%
<1000	18	10.59%
>1000	17	16.50%

Firm type	#	% of all firms of that firm type
Standalone	11	1.59%
Subsidiary	24	9.76%
MNC HQ	18	13.33%

Location of responding unit	#	% of all firms of that type in that location
Developing	9	2.68%
Europe	2	0.61%
Developing	24	12.18%
Europe	0	0.00%
Developing	17	14.41%
Europe	1	5.56%

Strong-form GINs, 15 cases, 1.23% of dataset

Industry	#	% of all firms in that industry
Auto	0	0
Agro	2	1.50%
ICT	13	1.39%

Country	#	% of all firms in that country
China	1	0.41%
India	11	3.40%
South Africa	2	2.38%
Developing	**14**	**1.94%**
Norway	1	0.55%
Europe	**1**	**0.20%**

Size	#	% of all firms of that size
<10	1	0.76%
<50	0	0.00%
<250	4	1.35%
<1000	7	4.12%
>1000	3	2.91%

Firm type	#	% of all firms of that firm type
Standalone	4	0.58%
Subsidiary	10	4.07%
MNC HQ	1	0.74%

Location of responding unit	#	% of all firms of that type in that location
Developing	3	0.89%
Europe	1	0.30%
Developing	10	10.71%
Europe	0	0
Developing	1	0.84%
Europe	0	0

As can be seen from Table 6, GINs also seem to a certain extent to be an "India" phenomenon with a third of the Indian dataset showing up as a stronger form G, I and/or N. Part of the reason may be that the India survey was conducted in the ICT sector with its emphasis on connectedness, and the virtual (and therefore easily globalised) nature of many of its offerings. However, countries like China and Norway also conducted the survey in ICT, and do not seem to have so many GINs. This suggests that firm strategy matters: India is English-speaking, it is a popular outsourcing destination for established MNCs, and domestic Indian firms often target the global market first. In contrast, China and Norway experience not only language barriers, but there is also a stronger domestic focus among IT firms.

Table 6: Participation in some stronger-form GIN

Respondents participating in a stronger form GIN	**#**	**% of all respondents from that country**
Brazil	6	8.70%
China	6	2.47%
India	109	33.64%
South Africa	5	5.95%
Total developing countries	**126**	22.22%
Denmark	2	4.08%
Estonia	0	0.00%
Germany	6	11.32%
Norway	9	4.97%
Sweden	14	7.18%
Total Europe	**31**	6.26%
TOTAL	**157**	12.92%

1.7.5 Methodological limitations

It is important to note that although the paper theorises global innovation networks, what is polled is not the network, but a single node of the network. The evidence can at best be described as an "ego network", and it suffers from the typical shortcomings of ego networks. The evidence is self-reported, and respondents are likely to provide more accurate information on local matters (e.g. the number of people employed at that unit) than on more distant matters (e.g. the size of the organisation overall).

Another issue of concern is ownership and control. First, although the data provides the location of the unit, which is adequate for standalone firms, it provides inadequate information about the location of the parent of subsidiaries. Specifically related to the strong representation of firms from developing countries, the evidence does not allow us to adequately distinguish between a subsidiary which is part of a strong-form GIN because it is part of the complex network of an advanced MNC and a subsidiary that uses a strong-form GIN to compensate for not only a weaker institutional context, but also the absence of the advanced MNC's rich network. Stated differently, if participation in a stronger form GIN can be regarded as a form of created asset seeking, it is not

possible to establish whether the motive lies with the unit in the responding location or with the parent.

Although the considerations related to ego networks affect respondents from Europe and the developing countries equally, it may also be the case that the two groups have a different reference point on certain matters. For example, when assessing the novelty of a given innovation, an entity in the developing world may judge it relative to other innovations in its less developed context, and judge it as more innovative than an entity in Europe would, since new-to-the-world innovations are more common there.

This shortcoming relates to the substantial challenges of conducting and interpreting a standardised survey across very different countries and industries. In spite of considerable efforts to ensure concordance between different countries and different industries, there are considerable differences in the types of databases used and response rates between countries. At a conceptual level, it must be asked to what extent even "objective" measures like the number of people working in a firm in two contexts as different as, for example, Denmark and India, can be regarded as comparable.

This is especially consequential because the analysis relies on *relative* measures for the construction of groups. The highly globalised, innovative and/or networked respondents are so relative to the other responses in the dataset, not according to some objective external measure. A relative measure is useful in the case of an emerging phenomenon such as GINs, as it allows us to capture the patterns that already exist. However, it also makes the conclusions vulnerable to the specifics of a dataset. The size and the breadth of the dataset may mitigate that limitation in this case.

Finally, it is important to remember that especially the final list of strong-form GINs is a short one, and that the limited data allow only tentative conclusions. For example, a more balanced dataset may or may not reveal fewer GINs in the ICT sector. The current era is dominated by the emergence of ICT, and advances in ICT have been described as a "carrier branch" in the overall economy (Cantwell, 2001). It may be that ICT firms lend themselves to operating in a global innovation network. However the relatively strong showing of agro-processing firms (2 out of a total of 133 agro-processing responses compared to the 13 out of 936 ICT responses) suggests that GINs may actually function across a range of industries. Further research is needed to clarify the link between the nature of the industry and GIN participation.

1.8 References

Asheim, B. and M. Gertler (2005). The geography of innovation: Regional innovation systems. The Oxford Handbook Of Innovation. J. Fagerberg, D. Mowery and R. Nelson. Oxford, OUP**:** 291-317.

Cantwell, J. 2001. Innovation and information technology in the multinational enterprise. In A. Rugman & T. L. Brewer (Eds.), The Oxford Handbook of International Business: 431-456. New York: Oxford University Press.

Das, T. K. & Teng, B.-S. 1998. Between Trust and Control: Developing Confidence in Partner Cooperation in Alliances. Academy of Management Review, 23(3): 491-512.

Malhotra, D. & Murnighan, K. J. 2002. The Effects of Contracts on Interpersonal Trust. Administrative Science Quarterly, 47(3): 534-559.

Pavitt, K. (1984). "Sectoral patterns of technical change: Towards a taxonomy and a theory." Research Policy **13**(6): 343-373.

Zhou, K. Z. & Poppo, L. 2010. Exchange hazards, relational reliability, and contracts in China: The contingent role of legal enforceability. Journal of International Business Studies, 41: 861-881

Country survey reports

1.9 Sweden: ICT and automotive

1.9.1 Database

The database was provided by Statistics Sweden.

1.9.2 Establishing survey sample: sector and company size

The dataset contained all the Swedish companies that according to Statistics Sweden operate in the ICT and Automotive sector in the following NACE 2 codes:

- 26.30 Manufacture of communication equipment
- 62.01 Computer programming activities
- 62.02 Computer consultancy activities
- 62.03 Computer facilities management activities
- 62.09 Other information technology and computer service activities
- 29.31 Manufacture of electrical and electronic equipment for motor vehicles
- 29.32 Manufacture of other parts and accessories for motor vehicles).

The organizations considered all have more than 5 employees.

1.9.3 Contacting the survey sample

In the original dataset 2181 companies were listed.

Of these, 585 companies provided general e-mail contact details. The database was expanded by searching for email contacts for each company. When possible personal email contacts were obtained, if this proved unavailable general email contacts were collected.

Sweden were able to find and add 1596 contacts, 1195 were general e‑mail contacts while the remaining 509 were direct e‑mail contacts.

In total Sweden had at least one e‑mail contact for 1981 companies from the total of 2181.

The next step was to check whether there was more than one company related to any email address and remove any duplicates leaving us with a list of 1868 companies.

1.9.4 Response rates

Table 4.1: Sweden: response summary

Number of companies contacted	E-mail	1830
	Total	**1830**
Responses		632

No Response		1198
Valid Responses		206
Response Rate		35.00%

Survey Monkey couldn't send 14 e-mail contacts to the INGINEUS Swedish questionnaire since those contacts had opted out of Survey Monkey.

The survey was initially sent to 1854 companies. During the survey other 24 companies were removed from the company list after they contacted us by email to inform us that they had been included in the survey by mistake and were not in the ICT or Automotive sector.

The final number of companies we were able to contact by mail totaled 1830 (1662 ICT; 168 Automotive).

To increase the response rate when we sent the second reminder we decided to create a second collector and to contact the companies using a second or a third email contact where we had this information.

The final number of completed responses in the Swedish survey was 195 (not including 11 completed responses that Lund had to cancel as explained below): 171 in ICT and 24 in Automotive; of which 5 didn't answer to Q.1 therefore Lund proceeded to answer and classify them using the sector classification provided by Statistic Sweden data (4 ICT and 1 Automotive).

There were 426 partial respondents (326 in ICT, 51 in Automotive, 13 did not answer Q.1).

We noticed that in some cases the companies classify themselves differently to Statistics Sweden.

In the first collection we noted those contacts that opted out or where delivery had failed. 133 contacts opted out and Survey Monkey noted 90 failed deliveries. We however received 150 failed delivery messages.

1.9.5 Post-survey data processing

Once the survey was concluded, we checked the cases where we had more than one completed answer per company and assessed which one to keep.

The criteria we used to decide were:

- degree of completion;
- position held in the company by the person who answered the survey;
- contact details fulfilled.

We had 9 cases of double answer and 1 case of triple answer to analyze and select.

Lund first had to reorganize the data in a format compatible to the original INGINEUS Survey since the Swedish one was partially different; mainly the questions were ordered in a different way and we had some additional questions.

1.9.6 Challenges

Lund had 5 completed responses that didn't answer Q.1. therefore Lund opted to answer and classify them using the sector classification provided by Statistics Sweden data.

In the case of Q.3.2 and Q.3.2.1 in the INGINEUS survey:

"Does your enterprise have a significant share of sales activity abroad?; 3.2.1 If you answered 'Yes' to the question above then please provide the percentage (%) of total sales derived from export."

We only had a single question asking to provide the percentage (%) of total sales derived from export. What we did to be able to give the answer to Q.3.2 of the Ingineus survey is to consider the answer was YES if the answer to Q.3.2.1was >=30%. If there was no answer to Q.3.2.1 we left empty the answer to Q.3.2.

In the case of Q.6 in the INGINEUS survey:

"Please indicate if you experienced innovation in the past 3 years (2006 - 2008) in any of the following. You may tick more than one option: new to the world, new to the industry, new to the firm and none"

In the Swedish survey there were 5 options: new to the world, new to the industry, new to the domestic market, new to the firm and none. When a firm answered in the Swedish survey "new to the domestic market", we decided to consider that equivalent at least to "new to the firm" in the INGINEUS survey.

As a control, to test that companies which did not respond to the survey were not different to those which did respond to the survey, we looked at existing information contained in the initial database from Statistics Sweden. We compared the group of companies, not respondents and respondents, by size of firm and by activity code (NACE 2).

1.10 Germany: automotive

1.10.1 Database

DIE bought data from Hoppenstedt, a private business information service. Hoppenstedt had to process raw data according to our specifications because the automotive industry is a cross-sectoral industry.

1.10.2 Establishing survey sample: sector and company size

In the Hoppenstedt database, firms are able to indicate the sector of their main clients (such as "car", "automotive", "motor vehicle"). Therefore, the frequency of these indications per sector was considered a good indicator for the final choice of sectors to be addressed in the survey.

Only sectors with a minimum of 19 addresses were considered (NACE 1 25241, 28408, 29140, 31610, 34300, and 74205).

DIE chose a minimum firm size of 50 employees for two reasons. Innovation activities are generally rather low in small firms in Germany, and innovation activities are rather concentrated on the larger supplier firms in the hierarchical automotive production system.

DIE had an initial database of 689 addresses. At the end of March 2010, we were asked to enlarge the database by including small firms (with more than 20 employees). We had to negotiate with Hoppenstedt and finally bought 434 more addresses and extended the survey to these firms in May 2010. We closed this additional survey on 10 June 2010.

1.10.3 Contacting the survey sample

Firms have different ways of organising production diversity:

- by different legal firms (but only one responsible; we finally contacted this person only)
- by a decentralised in-house organisation (we had to look for one or several of these)
- by a holding structure (we contacted all firms belonging to the holding, not the mother firm).

This clearing process reduced the database by 110 addresses. Clearing the added database of small firms reduced the number of addresses by a further 50.

We found that many addresses were not correct or incomplete, so we first would call to establish correct and complete addresses. An e-mail invitation was then sent to participate in Survey Monkey. Small companies (below 50 employees) were contacted directly by email.

38 firms noted that they preferred to get the invitation by postal mail (none of these firms responded to the survey).

When possible (phone numbers available) incomplete questionnaires were completed by further phone calls.

1.10.4 Response rates

Table 4.1. Germany: response summary

Number of companies contacted	E-mail	925
	Post	38
	Total	**963**
Responses		54
No Response		909
Valid Responses		33
Response Rate		5.60%

From the initial database of 689, 50 addresses were either unavailable, refused the inquiry or were dissolved because of mergers or insolvency. A further 4 cases were deleted because we realized that they are not automotive suppliers.

Thus the cleared database contained: 579 large companies (50 employees and more), of whom 541 were addressed by email and 38 by postal mail (according to their wishes), and 384 small companies (below 50 employees), in sum: 963.

1.10.5 Challenges

The response rate from the small companies (below 50 employees) proved to be extremely low (1.6 %; 6 responses from 384 addresses). This can be partly explained by the fact that these companies often do not specialize for the automotive sector.

1.11 Norway: ICT

1.11.1 Database

The main dataset was obtained from the public central business register in Norway, "The Brønnøysund Register Centre". This register is one of many sources that commercial enterprises use to build up databases for publishing business statistics and analysis and is also used by Statistics Central of Norway. This is the register that almost all official business statistics are based on.

The specific dataset used was extracted from a commercial register (Proff Forvalt - Eniro), as this is the solution subscribed to by BI Norwegian School of Management, and therefore readily available for researchers at this institution. The data are national and since the original source is the national register centre, the selections of firms that are included in datasets are mostly independent of the provider.

1.11.2 Establishing survey sample: sector and company size

NIFU-STEP identified units operating within the three selected industries (C10+11, C26.3 and J62). The organizations considered all had more than 5 employees.

1.11.3 Contacting the survey sample

NIFU-STEP noted 2477 potential contacts in the initial dataset.

However some were units of a single company with different outlets, some were published without e-mail addresses. After manually working through the list we were left with 1522 respondents with address information.

1.11.4 Response rates

Table 4.1. Norway: response summary

Number of companies contacted	E-mail	1522
	Total	**1522**
Responses		182
No Response		1440
Complete Responses		127
Response Rate		11.96%

NIFU conducted a pilot survey which targeted five selected firms, these provided feedback on the questionnaire. This feedback was communicated to the project management. Thereafter, an electronic questionnaire was sent to all 1522 respondents on which we had address information. The response rate was abysmal, with only 38 partial or completed responses. We thereafter decided to 1) focus on one industry (J62 with 756 firms) and 2) to use a commercial polling bureau to contact all firms and ask for an agreement in advance to respond to the survey.

Finally 519 firms had agreed to be contacted. The contacts were mostly managing directors in the organizations. The email-address of these firms were then fed into Survey Monkey, and when the survey was completed there were 182 partial responses and 127 completed surveys.

1.12 Estonia: ICT

1.12.1 Database

The database utilised was the Estonian Business Registry. The business registry database contains information on all companies registered nationally.

1.12.2 Establishing survey sample: sector and company size

A sub-set of the database was obtained, which contains information on all companies with EMTAK codes 26301, 62011, 62021, 62031, 62091.[6]

The database contains the information on the annual sales, exports, revenues, and number of employees. Given this, for the current survey in Estonia, a modified version of INGINEUS survey questionnaire was used. The original questionnaire was simplified by removing the questions on the industry sector of the respondent (ICT, automotive or agro-processing). Also, the questions on the number of employees and export volume were removed, as all this information was already known from the survey sample database. Also, a limited number of extra questions were added, e.g., open text questions clarifying the nature of innovative activities the respondents have actually undertaken.

In summer-autumn 2009, the NACE codes of individual companies contained in this database was cross checked and cleaned of any errors by the Institute of Baltic Studies in collaboration with the PRAXIS Centre for Policy Studies and the Estonian Association of Information of Information Technology and Telecommunications. The resulting database contained 1156 enterprises in total.

	NACE code	**Number of enterprises**
Manufacturing of telecommunications equipment	26301	16
Computer programming	62011	522
Computer consultancy	62021	240
Computer facilities management	62031	93
Other ICT activities	62091	285
Total		**1156**

After filtering the above database 121 companies with five or more employees remained in the survey sample.

[6] EMTAK is the Estonian equivalent of NACE code system. See: http://www.rik.ee/orb.aw /class=file/action=preview/id=33320/EMTAK%2B2008%2Bstruktuur-inglise%2Bkeelne.pdf

1.12.3 Contacting the rurvey rample

The contact details contained in the Estonian Business Registry are, typically either general e-mail addresses or the director's / owner's e-mail addresses.

All companies in the survey sample were invited to participate in the survey through the Survey Monkey web based survey system. The first invitation was sent out on 3 March 2010. Subsequently, to the contacts who had not yet responded, 2 reminders were sent.

1.12.4 Response Rates

Table 4.1. Estonia: response summary

Number of companies contacted	E-mail	121
	Total	**121**
Responses		17
No Response		4
Complete Responses		17
Response Rate		14.04%

During the survey process, 7 addresses bounced back, and 5 addressees opted out of the survey. In total, 17 responses were collected. The results do not allow, unfortunately, for any statistically valid analysis of the internationalisation of innovative activities in Estonian ICT enterprises at subsector level. The responses collected are still valid for both analysing the internationalisation of innovation in the Estonian ICT sector in general, and being included to the total INGINEUS survey data set.

1.12.5 Post-survey data processing

As part of the post-survey data processing, all responses were complemented with the already existing information on the ICT subsector, number of employees, exports, etc., and the full information was put into the original INGINEUS questionnaire format and once more to the Survey Monkey database.

1.12.6 Challenges

The most notable reasons for a relatively low response rate is that majority of the ICT companies in Estonia do not undertake formal R&D nor do they have significant international activities. It was therefore perceived that the current survey on internationalisation of R&D and innovation is not particularly relevant to their daily business activities. Also, an earlier ICT export survey, which was addressed to the same set of companies less than half a year ago, is likely to have reduced the response ratio of this survey.

The number of companies with five or more employees is in most of the ICT sub-sectors in Estonia very low. It is also our understanding of the survey data set that the allocation of individual companies to one of the specific computer programming or services subsectors is somewhat random, as the programming companies provide also consulting services and vice versa.

1.13 Denmark: agro-processing

1.13.1 Database

We utilised Orbis, a company database offered by Bureau Van Dijk, The Netherlands, which lists 241.000 Danish companies. According to the official Danish Statistics from 2007, there are 305.319 companies in Denmark.

1.13.2 Establishing survey sample: sector and company size

All companies in NACE rev. 2 codes:
- 10 manufacture of food products and
- 11 manufacture of beverages.

Companies with a minimum of 5 employees were selected.

This resulted in a total of 474 companies in the initial database.

Thereafter a number of companies were taken out of the database as these were not in really the manufacturing part of the sector, e.g. local meat shops (66 instances) or local bakeries (113 instances).

Companies which had closed down since the updating of the database were also cleaned (37 companies).

1.13.3 Contacting the survey sample

Companies without e-mail addresses were excluded (39 companies).The final cleaned database consisted of 219 companies.

Companies were contacted using Survey Monkey, following these criteria:
- For companies with between 5 and 30 employees CBS used the general company e-mail account.
- 31-250 employees we sent the link to the company manager.
- 251+ employees we send it to the Research/innovation manager.

1.13.4 Response rates

Table 4.1. Denmark: Response Summary

Number of companies contacted	E-mail or link	219
	Total	**219**
Responses		48
No Response		171
Valid Responses		48
Response Rate		21.91%

In total, 48 responses were received and there were 9 e-mails which were not delivered.

1.13.5 Post-Survey Data Processing

CBS then checked the results with the official Denmark's Statistics R&D and innovation survey. Unfortunately this does not allow for cross references; however the Statistics include all Danish companies (compulsory in Denmark) and is therefore are very reliable.

Feedback from some of the non-respondents indicated that the survey was irrelevant to them as farmers (maybe organic). After consulting the official Statistics Denmark, our results seem to be fine.

1.14 China: ICT

1.14.1 Database

Two databases were used for the INGINEUS project: one is focused in the region of Beijing and the other is focused on Shenzhen (Guangdong province). The first database ("Beijing database") is owned by Sinotrust, a market research company located in Beijing, and consisting mainly of a firms list published by the Beijing Administration for Industry and Commerce, as well as the Beijing Taxation Bureau. The databank is renewed every three months.

The second database ("Shenzhen database") is owned by CVISC, a similar research company located in Shenzhen, and consisting of a firms mainly from several science and technology entrepreneurship service centres in Shenzhen, as well as Shenzhen small and medium enterprises service centre. Thus this survey is regionally focused on Beijing and Shenzhen.

The information contained in the two databases include: company name, address, zip code, telephone number, fax number, main business, annual sales volume, number of employees, industry, ownership type, and corporate representative.

1.14.2 Establishing survey sample: sector and company size

GUCAS targeted the ICT sector for this survey, with the following sub-sectors

- Manufacture of communication equipment
- Computer programming activities
- Computer consultancy activities
- Computer facilities management activities
- Other information technology and computer service activities

The sample was equally distributed among small, medium and large companies: small-size companies (less than 49 employees) account for 37%, medium-size companies (50 to 249 employees) account for 35%, and large companies (250 or more employees) account for 38%.

1.14.3 Contacting the survey sample

For the "Beijing database", the only mode of contact GUCAS used were phone interviews; while for the "Shenzhen database", three modes of contact were used: face to face visits, face to face interviews on public activities and email.

For phone interviews, GUCAS held a training course for all telephone interviewers on each question of the questionnaire.

Initially a pilot sample of 20 companies were surveyed to see if any procedure modifications were needed.

The sample drawing method for phone call interviews was:

- GUCAS used a program to conduct random sampling every three companies on the list, if a company could not be contacted, the computer skipped to the next one automatically.

- If the number of companies did not reach expectations during the first-round of phone interviews, the team then conducted the second-round of phone interviews using the same method as above; and so forth.

For face to face interviews, GUCAS selected companies at random, to either have face to face visits with or to invite participants to public activities, such as the Shenzhen Indigenous Innovation Forum, the Product Innovation Forum and Innovation Salon, to fill out the questionnaire.

For small and medium enterprises, the interviewee in most cases was the owner-manager or top-level manager, while in large firms the interviewee was usually the R&D Head or his/her deputy.

1.14.4 Response rates

Table 4.1: China "Beijing Database": Response summary

Total number of companies contacted	**Telephone Interviews**	**8692**
Responses		208
No Response		8484
Response Rate		2.39%

Table 4.2: China "Shenzen Database": Response summary

Total number of companies contacted	**E-mail, face-to-face and invitations**	**427**
Responses		148
No Response		279
Valid Responses		35
Response Rate		34.7%

Of the responses 'Manufacturers of communication equipment', accounted for 33%, Computer programming activities 32% and computer consultancy activities accounted for 9%.

The sample was distributed throughout the five most developed provinces in China. Results from Beijing accounted for 60% of the total questionnaires; Guangdong province, accounted for 21%; Shanghai, accounted for 14%, Zhejiang province which accounted for 4%, and Shandong province, accounted for 1% of the total questionnaire.

1.14.5 Post-survey data processing

Since our survey responses came from two databases, we conducted T-test in selected variables for the responses from the "Beijing database" and "Shenzhen database". There was no significant difference in the tested variables (including company size and the nature of the company) between the mean of the "Beijing database" and the "Shenzhen database". It was noted that companies from the "Shenzhen database" have more significant R&D compared with the "Beijing database".

1.14.6 Challenges

Reasons for failed responses to phone interviews

Since phone call interviews were the dominant mode of contact in this survey, GUCAS analysed reasons for the limited responses to this medium. There were eight main reasons leading to the failing responses in the phone call interviews: 28.8% was for the wrong number or unobtainable number of the company; 25.2% was that company denied the interview; and 25.1% was because the call was not answered. The number of failed responses due to the above three reasons accounts for 79.1% of the non responses.

1.15 India: ICT

1.15.1 Database

CBS purchased access to the NASSCOM Directory of IT firms. The NASSCOM Directory is released every year and covers all areas of software production and related industries such as IT Enabled Services. The 2009-10 Directory provided the information on 1380 firms in different areas of the IT industry. CBS downloaded and created a database of all the firms regionally.

1.15.2 Establishing survey sample: sector and company size

The INGINEUS survey was conducted in India covering the Information Technology Sector. The choice of the industry was predetermined by the INGINEUS objectives reported in the research proposal.

The survey was designed to be implemented to cover all the IT firms in India. The following subsectors were examined in the survey:

- Manufacture of communication equipment
- Computer programming activities that include software development
- Computer consultancy services
- Computer facilities management activities and
- Other information technology and computer service activities

1.15.3 Contacting the survey sample

The first step involved a pilot survey in and around Trivandrum with a view to test the questionnaire. The feedback received was given to WP leaders so as to revise the questionnaire.

After getting the revised questionnaire the next step consisted involved collecting the physical and email addresses of all IT firms in the country. The initial idea was to conduct an email survey of all the firms using the NASSCOM Directory. The survey questionnaire link from Survey Monkey was emailed to all firms with a covering letter. However, face to face interviews were later conducted. All the firms in the selected locations were contacted and face to face interviews were conducted where possible.

The survey team consisted of three members in large cities like Bangalore, Delhi, Mumbai and Chennai. For all other smaller regions, a single person was appointed. A total of 16 members were appointed across all regions. Two supervisors were also appointed for the task of manual data feeding. CDS were in constant touch with the field staff, helping them to develop contacts with firms and solving other logistical problems. All the team members were trained by the CDS team and were asked to complete the questionnaire themselves in order to acquaint them with it.

1.15.4 Response rates

Table 4.1. India: Response Summary

Number of companies contacted	E-mail	1380
	Face to Face	307
	Total	**1687**
Responses		338
No Response		1349
Valid Responses		318
Response Rate		20.00%

Based on our earlier experience we expected a response rate of about 30 to 40 percent.

E-mail responses

Of the 1380 firms that the e-mail was sent to, only 21 responded, of which only 9 of them completed the questionnaire. About 80 of the mails either bounced or responded that they were out on vacation. After a month's gap a reminder mail was send to all the firms that did not respond or did not complete the survey. This time CDS offered incentives such as promising them to provide the respondents with the survey report so that they could compare their firm's performance levels with that of the industry average. The yield was still not encouraging. Only 12 more respondents replied, of which only 2 completed the surveys. Therefore the online survey achieved only 31 responses, of which only 11 were complete. So, having clearly failed in achieving a decent sample size we now turned to conducting personalized face to face interviews.

Face to face interviews

For the face to face interviews, it was not viable to cover the entire country as it would be very expensive and time consuming. Instead the team chose cities/ IT clusters that together represented nearly 93 percent of all firms in the NASSCOM directory. The regional profile is presented in Table 3.2. Over the eight weeks from March 1st to April 30th CDS were able to collect a total of 307 surveys completed with a fairly favourable response rate of 24 percent. In addition CDS also received 11 completed schedules through online survey making the total sample size 318.

Table 4.2. India manual (Face to Face) Survey Data

Cities chosen for survey	**Number of Firms as per NASSCOM 2009-10**	**Number of firms surveyed Manually**	**percentage of firms surveyed manually**	**Regional distribution of Firms in NASSCOM (%)**	**Regional distribution of Firms in survey(%)**
Bangalore	281	50	17.79	21.8	16.3
Delhi/Noida/Gurgaon	256	75	29.3	19.9	24.4
Mumbai	185	68	36.76	14.4	22.1

Pune	72	20	27.78	5.6	6.5
Chennai	147	39	26.53	11.4	12.7
Trivandrum	184	20	10.87	14.3	6.5
Hyderabad	107	25	23.36	8.3	8.1
Kochi	55	10	18.18	4.3	3.3
Total	1287	307	23.85	100	100

1.15.5 Post-survey data processing

In order to test the consistency of the data, the CDS team needed to check whether online and face to face data was comparable. They established that although there were some differences between the two surveys they were largely comparable and poolable.

1.16 Brazil: automotive sector

1.16.1 Database

The survey was carried out on a sample of firms, which was created based on three distinct sources:

(a) The Annual Registry of Social Information (RAIS)

RAIS is a registry of social and balance sheet information collected by the Labour and Employment Ministry and it is mandatory for all firms formally registered in the country. The dataset is made available by the Ministry annually with a two-year lag.

(b) Auto-parts Union Contact List (SINDIPECAS)

The Union keeps a database of all affiliates companies. The downside of this source is its maintenance, given that the companies' details are only updated sporadically. Regardless, we took on the information available and updated the details using information available online.

(c) Other known suppliers

Further to the Union List, we had already gathered data from interviews with employees of a few key companies in the automotive sector. These interviews provided a number of contacts of local suppliers. These contacts were added to the sample.

1.16.2 Establishing survey sample: sector and company size

In summary, 107 firms were chosen from RAIS, 66 from the SINDIPECAS and 88 from previous research projects, in a total of 266, which account for 100% of companies directly classified as or pertaining to the automotive sector in the state. Data on size such as number of employees is only readily available on the RAIS database.

The raw dataset was then reduced to 241, after cleaning the sample of companies that closed between the last year of availability of the date sources and the present day, companies without complete contact information (name, address, sector, phone number and potential interviewee).

(a) The Annual Registry of Social Information (RAIS)

From the dataset all manufacturing firms classified as pertaining to the automotive sector, defined by the company's highest source of revenue, from the state of Minas Gerais were extracted, provided the firm declared over 30 employees.

The total number of firms classified in the automotive sector in Brazil is 2,625. Out of these, 233 companies are located in the state of Minas Gerais. Of these companies, 107 (46%) employed, in 2008, 30 workers or more.

(b) Auto-parts Union Contact List (SINDIPECAS)

From experience, CEDEPLAR established that the number of firms which are part of the automotive chain of production tend to differ from the companies which are formally classified as part of the sector. This problem is particularly important for companies producing a large number of goods, making the formal classification less meaningful. For this reason the Brazilian team anticipated that their case study (GIN of local FIAT) identify companies that are part of the chain of production. One of the possible sources is the Auto-parts Union.

1.16.3 Contacting the survey sample

Before sending the questionnaires by e-mail to all 241 companies, a first round of phone calls took place to check the correct details of the interviewees. Preference was given to job titles such as general manager or production supervisor. From the 241 companies, 134 provided the name and phone number of a specific person to answer the questionnaires. For the 107 remaining companies the e-mail was sent to a more general sector of the company.

After the first round of phone calls, e-mails were sent to all companies. The e-mail included the standard letter presenting the research project, a link to the online questionnaire and a phone contact in case the preferred media was a hard copy or fax. Having sent the initial round of e-mails, we waited for replies for a full week, when a new round of e-mails was sent, together with phone calls to the existing contacts. This strategy was kept for the following three weeks.

1.16.4 Response rates

Table 4.1: Brazil: response summary

Number of companies contacted	E-mail	241
	Telephone follow-ups	
	Total	**241**
Responses		55
No Response		186
Complete Responses		30
Response Rate		22.8%

By the end of April the effort rendered 30 fully answered questionnaires and 25 partially completed questionnaires, in a total of 55 questionnaires.

1.16.5 Post-Survey data processing

It is important to note that contrary to most of the INGINEUS partners, CEDEPLAR chose, to host the online survey locally using an own server and tools. This decision was taken based on two perceptions related to the local culture. Firstly, in the teams opinion, companies could see the alternative, i.e., the “survey monkey” website, as informal whereas the university domain name is instantly recognised by the public. Secondly, the word monkey, both in Portuguese and English, is often used with a racially derogatory context and we judged that we should avoid any further complexity to the already difficult task of surveying the industry. Hence, only after completing the effort to obtain answered questionnaires the answers were entered in the common tool to allow for an easier handling of all partners data.

1.16.6 Challenges

Only two companies justified not taking part in the survey. The remaining 158 companies that did not answer did not give any reason for doing so even after further e-mail and phone calls.

1.17 South Africa: agro-processing

1.17.1 Database

The first database procured was the Experian database, a global information services company, which provides data and analytical tools to clients in more than 90 countries. This database consisted of 1096 firms. The Experian database is skewed towards larger firm sizes when compared against the population of agro-processing firms. Absolute values of responses were deemed too low therefore a database was constructed using several other data sources. These were:

- Go Organic Online Directory
- South Africa's premier organic website, directory and marketing company.

The directory was refined by the category of all farmers, producers and wholesalers, in all areas in South Africa, with all types of products selected. (This last option was selected as some companies were dual agro processing/other industries and would not have shown up in the search if only agro products were selected.) Repetitions and non agro processing companies were omitted from contact.

- Tradepage Online Trade and Business Directory South Africa

Tradepage is a dedicated Internet Service Provider (ISP) offering a range of internet services relating to Business on the Internet, enabling businesses throughout South Africa to benefit from internet access.

- Search ZA Directory

SearchZa is the most comprehensive search engine for .ZA domains.

- The Food World

A very broad based database of all food importers and exporters

In the constructed database all repeats from the previous database were excluded. The decision was also taken to eliminate all resellers of agro-processed products.

1.17.2 Establishing survey sample: sector and company size

The survey was conducted in the agro-processing sector and included the following subsectors:

- Processing and preserving of meat and production of meat products
- Processing and preserving of fish, crustaceans and molluscs
- Processing and preserving of fruit and vegetables
- Manufacture of vegetable and animal oils and fats
- Manufacture of dairy products
- Manufacture of grain mill products, starches and starch products
- Manufacture of bakery and farinaceous products
- Manufacture of other food products
- Manufacture of prepared animal feeds

The Experian database distribution in terms of the sizes of the firms is contained in figure 2 below. The figure also shows this distribution as compared with the SARS database which is representative of the population of firms in the agro sector.

Table 2.1. Experian database size of firms by employee number

Number of Employees	Experian Employees	Experian % of Companies (out of 1096)	SARS Employees	SARS % of Companies (out of 8506)
0	14	1.28	3842	45.17
Fewer than 10	162	14.78	2 538	29.84
10-49	490	44.71	1 543	18.14
50-249	309	28.19	443	5.21
250-999	94	8.58	89	1.05
1000-2999	14	1.28	29	0.34
3000-5999	6	0.55	9	0.11
More than 6000	7	0.64	13	0.15
Total Companies	**1 096**	**100**	**8506**	**100**

1.17.3 Contacting the survey sample

An online survey tool was set up with an e-mail link facility. Thereafter, each contact (minus repetitions) on the database was called, given a description of the survey and its relevance and asked to participate. The persons contacted who agreed to participate were then sent the survey link electronically. Those contacts who agreed to participate but who failed to submit their survey responses were contacted again two weeks later. If they failed to respond to this reminder a final reminder was sent again 2 weeks later.

1.17.4 Response rates

Table 4.1. South Africa: Response Summary

Number of companies contacted	E-mail	497
	Total	**497**
Responses		83
No Response		414
Valid Responses		83
Response Rate		17.00%

The Experian database had not been updated therefore many of the contacts were redundant. From the 1096 firms listed only 325 were contactable and agreed to have the survey mailed to them, from this database, 59 responses were received.

In order to raise the sample size several other sources were used to construct an additional database in order to raise the number of respondents.

172 additional firms were called and contacts made which gave permission for the survey to be sent to them. This garnered 24 additional responses bringing the total number of responses for the phased approach to 83 from 497 surveys sent out.

Table 4.2. South Africa: database summary

Database	**Sent**	**Responded**
Experian	325	59
W. Cape	63	7
Tradepage	6	2
Search ZA Directory	17	5
Go Organics	25	8
Foodworld Directory.com	61	2
Total	497	83