1. Publishable summary

Project acronym: INNOS&T
Project title: S&T indicators combining patent data and surveys: Empirical models and policy analyses
Contacts
Scientific Coordinator: Prof. Alfonso Gambardella
Project Manager: Dr Serena Giovannoni
Tel: +39-02 58363591
Fax: +39-02 58363790
E-mail: innost@unibocconi.it, alfonso.gambardella@unibocconi.it
Project website address: www.innost.unibocconi.it

Project objectives
The aim of this project for the first 18 months is to carry out all preparatory work to develop and collect novel, systematic and more adequate science and technology indicators.

The four objectives for this period are the following:
1. To carry out a new survey data collection, PatVal-EU II, which builds on the PatVal-EU survey (hereafter PatVal-EU I). The new survey has a broader European coverage as compared to the first PatVal-EU survey. We interview inventors of patents granted by the EPO located in 20 European countries and Israel. The European countries include: (i) the 15 most active European countries in terms of number of granted patents with priority date in 2003-2005; and (ii) New Member States and other European countries that would not fall into the 15 top patenting countries in Europe, but that increased their patenting activities in the last few years and were granted at least 50 patents in 2003-2005. Specifically, we survey inventors in the following countries: Germany, France, Great Britain, Italy, Netherlands, Switzerland, Sweden, Finland, Belgium, Austria, Denmark, Spain, Norway, Ireland, Greece, Slovenia, Hungary, Czech Republic, Poland, Luxembourg and Israel. The survey will add new relevant indicators (also with respect to PatVal-EU I) that address important policy issues. We add new questions about the inventors' careers, their social setting, the reasons for not using or not licensing patents, the value of patenting vs. the value of the patented inventions, the characteristics of new firms created from patents and of their founders, the importance of local, national and international collaborations in the invention process, etc. The survey aims to cover a more recent period of time compared to PatVal-EU I, i.e. 2003-2005. This is crucial in order to identify changing patenting behaviours during 2000s, which have been characterized by an upsurge in patenting and licensing activities, and by increasing complexity of the patenting process.

2. To carry out two complementary survey data collections in US and JP, i.e. PatVal-JP and PatVal-US. These surveys will allow for a direct comparison of invention processes in Europe, Japan and the US. We expect that our comparisons of European with non-European inventors will yield deep insights into comparative advantages and disadvantages that European researchers and inventors might be subject to in their inventive efforts. Such information is currently lacking in the literature, and our inventor survey methodology is uniquely suited to generate this information. For policy-makers in Europe, these data would prove particularly valuable in addressing impediments on the road towards the realization of the Lisbon Agenda.

3. To develop indicators for industry-science links based on patent citations to science for all OECD countries, over time, across industries, by firms, by universities and by firms to universities and public research institutions. In this project, we propose various ways of using patents to identify interactions between scientific research, technology and innovation. From combining several databases related to patents and publications (USPTO, EPO-REFI, PATSTAT, ISI Web of Science), the project will develop the following 'industry-science' indicators: Science Intensity of technology (i.e. i. Scientific Non Patent References, ii. Patent References to patents held by scientific organizations); Technological orientation of scientific organizations (i.e. iii. Patents held

by universities and public research organisations). These indicators will be constructed for all OECD countries and will cover the 1990-2005 period. Indicators and methods that take adequately into account the linkages between science, technology and industry are currently not available, especially for such a broad coverage of countries and years.

4. To build complementary indicators based on patents (citations, oppositions etc.), publications, individual inventors, companies, public research institutions and universities, technologies, regions, sectors. These indicators will be matched with the survey data. The resulting integrated dataset will take into account simultaneously a mix of key factors that may have important impacts on policies. For example, the use of these indicators in our empirical models will enable us to identify which factors contribute to the production of valuable inventions (i.e. the size of the firm, the R&D investments, the technological area of the patent, the characteristic of the inventor, the local technological environment in which the invention has been developed) and to measure the size of the impact of each relevant factor. This will allow assessing the economic importance of each factor and improving policies fostering the production of valuable inventions. Also, the construction of alternative indicators for the same topic will help to check the validity and reliability of key indicators.

Progress of work and main results achieved

During the first 18 months of the project we have carried out the following activities and achieved the following results.

We pursued all preparatory activities (included in WP1) necessary for carrying out the survey of inventors of patents granted by the EPO with priority date 2003 and 2005 and located in one of the following 20 European countries (Germany, France, Great Britain, Italy, Netherlands, Switzerland, Sweden, Finland, Belgium, Austria, Denmark, Spain, Norway, Ireland, Greece, Slovenia, Hungary, Czech Republic, Poland, and Luxembourg) and Israel. We also conducted several complementary activities necessary for carrying out the surveys of US and Japanese inventors.
In particular:
1) We produced the questionnaire. The final version of the questionnaire was translated into 11 languages (CZ, DE, ES, FR, EN, HU, IT, NL, PL, SI, JP) before running the pilot test, while we used English questionnaires for the Nordic countries (SE, DK, FI, NO) as well as for IL and GR.
2) We prepared the sample of patents to be surveyed. Our targeted number of patents is about 20% of the population of granted patents of each country and at least 50 patents for the countries with a small number of patents.
3) We defined the exact procedures for preparing the inventors' addresses for each country and prepared the dataset of inventors' addresses.
4) We defined the survey methods. LMU sub-contracted to TNS-Infratest, a market research company operating world-wide, the task of contacting inventors and collecting their responses. In particular, Infratest contacts inventors by mail and ask them to fill out an online questionnaire on a website that they can access through an ID and password, generated for the specific inventor by the market research company. Inventors in some countries with low response rates will also be provided with a hard copy of the questionnaire.
5) We defined the procedures for running a pilot test and ran the test. The pre-test was conducted in all countries, except for Japan (postponed). On June 22, 2009, we sent out 1,375 letters inviting the inventors to fill out the online questionnaires. The overall response rate amounted to 8.9% (corrected response rate: 10.1%).

The preparatory work illustrated above has been completed in order to proceed with the full scale surveys by the beginning of the second 18 months of the project (WPs 2 and 3).

Regarding the development of Industry-Science Link Indicators using patents (carried out in WP4), we carried out the following activities:
1) We identified patent assignees by organizational type (company, university, public research organization, individuals etc.). This activity was based on a methodology, developed in 2005 at

K.U. Leuven by INCENTIM/SOOI in collaboration with Eurostat/PATSTAT, which has been extended and refined in 2009 within this project.
In particular, we constructed indicators that relate to: patents held by universities and public research organisations and patent references to patents held by scientific organizations. The sector allocation and name harmonizing steps that were required for extracting these indicators, were accompanied by validation and quality control of the results obtained.
2) All non-patent references (NPR's) from the PATSTAT database were extracted. Because not all NPR's can be considered as scientific (Callaert et al., 2006), two complementary methodologies were developed to evaluate the scientific nature of non-patent references. On the one hand, a set of key terms was developed to create classification rules, based on exact matching procedures. On the other hand, a supervised machine learning environment was set up for identifying references as scientific or non-scientific.
The exact matching approach revealed that almost 59% of the allocated NPR's are scientific (i.e. they refer to the serial journal and proceedings literature). 31% were identified as non-scientific and the remaining 10% referred to databases or abstract services that represent a mixture of source material.
3) Validation of the NPR identification was performed separately for each approach. At the same time, the exact matching and the machine learning approach were cross-validated to check their consistency in identifying NPR's. Whereas the exact matching provides a very transparent and near 100% accurate way of identifying the specific nature of the reference, it allocates 'only' half of the NPR's. The machine learning approach from its side identifies 99% of the references, but is somewhat less accurate and much less specific in identifying the exact nature of the references. Complementing both approaches allows to harvest the best of both worlds, resulting in a 99% coverage with 95% accuracy in identifying scientific references; and an additional 100% accurate and detailed specification of about 50% of all non patent references.
4) A dataset of indicators for industry-science links was developed. It includes patent references to patents held by scientific organizations, and patents held by universities and public research organisations. The outcomes of the NPR identification are being used to build indicators of scientific NPR's and to map the occurrence of science technology relatedness for the broader patent universe in all OECD countries (time period 1990 -2005).

During the first 18 months we also started to create complementary indicators (WP5) that will be integrated with the datasets resulting from the surveys of inventors.
In particular, we produced a dataset of patent indicators (including backward and forward citations classified by category, oppositions, claims, filing date, date of receipt at EPO, IPC information, priority date, main technological class used for classification of areas, number of EP equivalents, family size/total number of equivalents) and carried out several preparatory tasks for creating datasets of indicators at the company, regional and technological level.

Some initial activities for the validation of industry-science indicators (WP6) have been carried out at this stage of the project. The completion of validation activities will be carried out in the second 18 months of the project

Expected Final Results
The indicators and the empirical methods and models of this project will represent an important step forward with respect to the current state of the art of indicators and methods. In particular the expected results of the work are the following.
We will produce innovative indicators with broad coverage (countries and years), by employing advanced and rigorous methodologies and addressing relevant issues for which adequate indicators do not exist. Indeed, this work will fill an important gap by creating indicators (currently unavailable elsewhere) on the following issues: the motivations for the actual use or non-use of the patents, the extent of strategic patenting; the actual decision of cross-licensing a patent; the characteristics of the founders, the owners, the activities or the geographical location of new firms created from patented inventions; the social setting of inventors (family, children), their role and position in the organization in which they are employed; the monetary value of patenting, the links between industry and science.

An important result of this project will be the development of appropriate and effective empirical methods that enable to assess simultaneously the impact of a variety of factors on key economic phenomena (i.e. the value of inventions, the use of patents). This will be possible because we will develop survey-based, industry-science link and complementary indicators at different levels of analysis. All these indicators will be selectively included in empirical models that will employ advanced econometric methods, which are appropriate for survey-based indicators and for the use of multiple indicators at different levels of analysis. By drawing on these empirical models we will build tools for assessing the expected impact of policies within a range of possible outcomes. For instance we may estimate the expected impact of a change in the share of inventors with a PhD on the production of valuable inventions. We will obtain a range of values as a result of the variation of a factor, or from combined changes in more than one factor.
A final but still important achievement of our project will be the training of young researchers working in our own institutions and in institutions of other countries that will be involved in various stages of this project.

Impact and use
Indicators from the surveys, complementary indicators and S&T indicators will be used to develop empirical models that can contribute to improve policies on the following topics:
1) The economic use of the patents: (i) unused patents and strategic patenting, (ii) licensing, and (iii) creation of new firms from patents. Understanding these issues will help to define policies aimed at removing barriers to licensing, such as the reduction of transaction costs, the creation of intermediaries in the market for technologies, the support to small and new technology-based firms.
2) Science-industry linkages and innovation performance: under this topic we will develop empirical models and policy oriented analyses on different levels: countries (innovation systems), organizations and inventions stemming from research organizations. These analyses could provide an important contribution on understanding the relevance of policies aimed at fostering entrepreneurial behaviour of universities and/or interaction between industry and academia.
3) Gender, education and mobility of inventors: we are interested in factors like age, gender, family setting (single/couple, children), level of education, field of university degree and PhD, professional background, career, national and international mobility of European inventors. Information on individual inventors/scientist is often available at an aggregate level. New indicators may contribute to improve the impact of national and European policies supporting S&E education, national and international mobility, the participation of women in S&T. They also contribute to better understanding the impact of age, family and career on the productivity of women vs. men in S&T, for which indicators are very scarce. The collection of these types of information has been strongly encouraged by the European Commission in the STI report (2003), by claiming that the sex disaggregation is an EU priority in data collection activities on Science and Technology.
4) The economic value of patents: Understanding the characteristics of high-value patents and their producers can contribute to improve policies supporting the production of valuable patents.