

A register of public-sector research organizations as a tool for research policy studies and evaluation

Lepori B, forthcoming in Research Evaluation.

Summary

The paper presents the conceptual and methodological design of a register of public-sector organizations, as well as a preliminary delineation of such organizations in Europe. Conceptual and methodological issues are discussed, as well as the potential usage of the register for interlining datasets and analysis. The significance of the register for research policy and evaluation studies is also discussed, as related with changes associated with New Public Management reforms.

Keywords

Public-sector research, organizational registers, higher education institutions, public research organizations, New Public Management, public-sector research evaluation.

1 Introduction

The ecosystem of public research organizations in Europe deeply changed in the last decades, as the outcome of related processes in the governance of public research and in the organization of the sector (Bonaccorsi and Daraio 2007; Crow and Bozeman 1998).

At the policy level, the emergence of New Public Management rationales (Ferlie, Ashburner, Fitzgerald and Pettigrew 1996) transformed the governance of public research from a mode where research organizations were tightly controlled by the state (Capano 2011) to one where Higher Education Institutions (HEIs) and Public Research Organizations (PROs) were given more autonomy (de Boer, Enders and Leisyte 2007) and are increasingly steered through economic incentives (Hicks 2012). Such reforms also pushed public-research organizations to become 'complete organizations' (Brunsson and Sahlin-Andersson 2000), endowed with a clear hierarchy and able to develop coherent organizational strategies and to set priorities in the conduct of research (Fumasoli and Lepori 2011). Empirical studies have shown how research organizations are able to shape the research groups' environment (Heinze, Shapira, Rogers and Senker 2009), as well as the behavior of individual scientists (Laudel 2006), therefore emphasizing the relevance of the organizational level in the study of science.

In the organization of public research, deep changes also took place. On the one hand, the higher education landscape has become more diverse with the emergence of 'second-tier' HEIs such as colleges and Fachhochschulen as a research actor (Kyvik and Lepori 2010) and of private HEIs (Teixeira, Rocha, Biscaia and Cardoso 2014), while many countries attempted to consolidate the system and to reach critical mass through mergers (Pinheiro, Geschwind and Aarrevaara 2016). On the other hand, the PRO landscape increasingly differentiated as the outcome of devolution from the public administration and the emergence of new organizational types such as Research and Technology Organizations (RTOs; Cruz-Castro and Sanz-Menéndez 2018). Finally, blurring between sectors has taken place, such as the establishment of joint laboratories between PROs and HEIs (Jonkers and Cruz-Castro 2010), the association between universities and research hospitals (Davies, Tawfik-Shukor and de Jonge 2010) and the organizational integration of PROs in the university sector. These transformations make more urgent to provide a systematic description of the organizational structure of public research across different national research systems.

This paper reports on a far-reaching effort to map the organizational structure of public-sector research in Europe through the construction of an organizational register (OrgReg; orgreg.joanneum.at), as implemented within the EU-FP7 Research infrastructure project for research and innovation policy studies (RISIS; <https://www.risis2.eu>). OrgReg builds on previous work on Higher Education Institutions in the European Tertiary Education Register (ETER; www.eter-project.com), but extends the coverage to all non-market

oriented organizations with a sizeable research output (OECD 2015), including specifically public research organizations and research hospitals. Further, OrgReg provides a systematic tracking of demographic changes in public research from the year 2000, as well as of linkages between organizations and of multi-level structures such as organizational groups and associations.

In this paper, we first focus on the OrgReg design choices and on how these were driven by the knowledge of the organizational structure of public-sector research. These mainly concerned three issues, i.e. identifying the coverage of public research organizations, dealing with the intricate structure of public-sector research with its pervasive linkages between different (types of) organizations and, finally, tracking organizations consistently over time.

Second, we discuss some directions in which OrgReg might contribute to research policy and evaluation studies. On the one hand, OrgReg allows for the first time for a systematic characterization of the organizational structure of public-research across European countries, therefore enriching comparative studies of research policies (Larédo and Mustar 2001) with the organizational level. This is highly relevant given the increasing role of research organizations in the implementation of public policies, which requires dealing with the impact of policies on the behavior and performance of research organizations as well (Geuna and Martin 2003).

On the other hand, OrgReg functions as an interlinking facility between different datasets and, therefore, is an important step towards open linked data on research organizations (Light, Polley and Börner 2014). Thanks to the use of the register's identifiers, it becomes possible to assign unambiguously inputs and outputs to organizations. This applies to data derived from administrative sources (Lepori, Bonaccorsi, Daraio, et al 2015), from international databases on publications (Waltman, Calero-Medina, Kosten, et al 2012) and patents (De Rassenfosse, Dernis, Guellec, Picci and de la Potterie, Bruno van Pottelsberghe 2013), therefore overcoming a major limitation of many current studies based solely on output data (Abramo and D'Angelo 2016). Further, the availability of a stable list of public research organizations allows matching with data derived from on-line sources, such as institutional websites and Wikipedia, as well as of social media, such as ResearchGate (Lepori, Thelwall and Hoorani 2018) and Google Scholar, therefore significantly expanding the realm of available information on research organizations.

2 Conceptual roots

The establishment of an organizational register as a statistical infrastructure largely builds on New Public Management and its notion that key tasks performed by the state, such as the conduct of public R&D and the delivery of educational services, should be executed by autonomous entities at arm's length from the state (Ferlie, Musselin and Andresani 2008).

Accordingly, policy reforms aimed at constructing – symbolically and practically – public organizations were inspired by a corporate blueprint. According to this model, public organizations should develop a clear identity and boundaries, have an internal hierarchy and rule-based decision-making (Brunsson and Sahlin-Andersson 2000). The extent to which higher education and research organizations conform to this model has been debated in the higher education (Seeber, Lepori, Montauti, et al 2015) and public research organizations literature (Cruz-Castro, Sanz-Menéndez and Martínez 2012). Therefore, despite a common normative blueprint, variation is expected between individual organizations and countries that is also likely to raise methodological issues in the construction of a register.

In this respect, OrgReg represents an effort to provide statistical recognition to the existence of public-sector organizations, moving beyond the current practice of producing statistical data at country and regional level, respectively beyond a longstanding tradition of science studies focused on the individual or group level (Knorr Cetina 1995). Of course, standardization was already undertaken in some areas, such as scientific

publications (Waltman, Calero-Medina, Kosten, et al 2012), but OrgReg pushes it further by providing a unique (source-independent) and certified list of the organizations and by claiming that this list was checked against 'reality'.

Interestingly, this conceptual rooting was never openly advocated in the discussion within RISIS on the design of OrgReg. The underlying model was however visible in the definition of organizations provided in the OrgReg handbook, defined as "a collectivity with a relatively identifiable boundary, a normative order, ranks of authority, communication systems and membership coordinating systems, that exist on a relatively continued basis and is engaged in activities that are usually related with a set of organizational goals". As such, this definition draws on what are considered as the key elements of an organization, namely goals and strategy, boundaries and internal coordination (Pfeffer 1997; Scott 2008); at the same time, this definition is sufficiently vague to provide room for a pragmatic and flexible operationalization.

Of course, science is also composed by scientific disciplines with their social structure and epistemic culture (Knorr Cetina 2009), which largely account for the dynamics and evolution of scientific domains (Clark 1996). The relationship between the organizational level and the disciplinary level has been long debated in science studies (Becher and Trowler 2001), but it is now acknowledged that research organizations, such as universities, are more than simple 'containers' of disciplines (Musselin 2007), but fulfil on their own important functions in the research system, such as managing financial and human resources and shaping the environment of research groups (Carayol and Matt 2004).

This remark points to the complementarity between studies at the organizational and disciplinary or individual level. Integration between the two levels could be achieved in different ways. For studies focusing on the organizational level, disciplinary heterogeneity can be controlled through the repartition of students by fields within HEIs (Lepori, Baschung and Probst 2010) or by breaking down outputs by scientific domains (Waltman, Calero-Medina, Kosten, et al 2012). In turn, for studies at the disciplinary or individual level, it becomes now possible to control for organizational characteristics, as a possible explanation of observed differences, for example through the design of multi-level models (Lepori, Wise, Ingenhoff and Buhmann 2016).

3 Identifying public-research and higher education organizations

To provide a consistent register of public-sector research organizations, a number of methodological issues needed to be resolved.

First, according to the OrgReg definition organizations should be distinct and identifiable. Since not all entities in the public sector have a legal personality, we eventually resorted to a combination of criteria such as visibility (for example on the Web), having their own budget, managing staff, etc. Consistently with previous analyses on PROs (Cruz-Castro, Sanz-Menéndez and Martínez 2012) and universities (Musselin 2013), we found out that most public research organizations can indeed be identified as distinct from the public administration.

This applied also for research centers that used to be part of the public administration, but received in the last decades an independent status. Nevertheless, we also identified a number of entities that conduct R&D, but are still integrated within the public administration, the most frequent instances being research units of meteorological services, statistical offices and metrology services. Since these cases might be of substantive interest, they have been included in a specific subcategory of 'public administration' (PA) entities, under the condition that they have identifiable research units and visible research output. Ministerial research conducted for the sole purpose of supporting policy was excluded.

A further issue was related to quasi-organizations created by national research programs, which are in fact groupings of institutes within universities or PROs without their own staff, such as the Virtual Vehicle Centre

in Austria. As a general rule, they were excluded from the register since the focus of OrgReg is on lasting organizational forms rather than on transient forms established to foster research collaboration. Admittedly, the distinction between 'real' and 'virtual' organizations is not always easy to draw and might evolve over time, so that inclusion of some of these entities might be considered in the future.

Second, the notion of 'public' is complex and partially depends on the country considered. Especially after NPM reforms, legal status is a weak indicator as organizations within the public sector may have a private legal status, while some purely market oriented companies are within the public sector. 'Publicness' as the extent to which an organization is influenced by the political authorities (Bozeman and Bretschneider 1994) is a more relevant definition that is however difficult to operationalize – public influence might be exerted through varied forms such as hierarchical control, public regulation or resource dependencies.

Eventually, we adopted the criterion that entities in OrgReg should not be mainly market-oriented, i.e. selling R&D services to third parties. This criterion is easier to operationalize, for example by looking to the revenue composition, and is consistent with the sectoral delimitation in the Frascati Manual (FM; OECD 2015). The OrgReg coverage therefore corresponds to the higher education, government and private-non-profit sectors in the FM.

In practice, this excluded consultancies, technical centres providing on-pay services to companies (without a prevalent public mission) and companies' research centers, while it includes private non-profit centers whose research is mostly public domain, even if they are managed by charities and mostly funded through private sponsors, such as the French Curie Institute. Research centers of public companies were excluded, as they serve market-oriented companies.

The application of this criterion was rather straightforward. The main ambiguous cases were Research and Technology Organizations (RTOs), for which it is was not always fully clear whether the public mandate of promoting technology transfer is more important than selling services to companies; some of them are real 'hybrids' as they receive public support, but should also acquire funding by selling services (Gulbrandsen 2011). Eventually, it was decided to include such centers in OrgReg at the condition they have a public mission, some public entities as partners and perform some general-purpose R&D (even if it cannot be ascertained whether this is their main activity). In this respect, we decided to prioritize the consistent inclusion of a group of entities over a strict application of the market orientation criterion.

Third, organizations in OrgReg should be involved in R&D and/or in higher education. This criterion was operationalized by thresholds in terms of the number of publications (at least 10 per year in in the Web of Science) and of participations in European Framework programs (at least 10 EU-FP projects), as well as a minimum number of tertiary education students enrolled for HEIs (100 students or at least 30 Full Time Equivalent of staff). This criterion maintained the number of entities in OrgReg within a manageable level, as there are many entities with low level of activities.

The following entities were generally excluded: policy think thanks, science museums and organizations in charge of the dialogue between science and society, except when they have an explicit research mission, and public-private cooperation entities with no R&D mission, such as pure technology transfer centres, technology parks, innovation support centres, etc. Again, the most ambiguous cases were (public) technological centres serving the needs of the industry, where the borderline between R&D and services was not always easy to draw.

An important addition to OrgReg are Research Funding Organizations (RFOs), i.e. organizations that distribute public project funds, such as research councils (Braun 1998). They are included given their role in public research policies and the fact that in some countries, notably the UK, RFOs also manage national research facilities.

By summing up, when considering the core of public-sector research, i.e. Higher Education Institutions and Public Research Organizations, there were few ambiguous cases and OrgReg is likely to provide a reasonably robust and comparable coverage. Two areas of ‘hybridity’ deserve closer investigation, i.e., first, the boundary between public administration services and PROs for activities where research and public service are closely associated and, second, public-private centres at the crossroads between public applied R&D and industrial support. Both point to changing organizational forms over time that are of substantive interest to analyse the relationships of public-sector research with the state on the one hand, with the private sector on the other hand.

4 Multi-level structures and linkages

The concept of OrgReg is to provide a list of units with a clear and distinct perimeter. Yet, it was quickly noticed that public-sector research includes nested structures that are connected by some kind of linkages. The model of organizations as self-standing entities did not adequately deal with large ‘umbrella’ organizations, such as CNRS, comprising largely autonomous laboratories, as well as with mixed units between Public Research Organizations (PROs) and universities. This remark pointed to the varied forms of research organizations in European countries and to the fact that a register, while achieving some level of standardization, should not be blind to these differences. It called for introducing some more complexity in the register’s design.

In the case of firms, the organizational structure can be highly complex, with many level of subsidiaries controlled by a parent company, leading to a highly complex and demanding identification of linkages. Yet, their legal structure provides for a clear criterion to identify linkages, i.e. control through the majority of shareholders’ voting power (Eurostat 2003, chapter 7). Unfortunately, this criterion does not fit the reality of public organizations, where forms of control are more varied and difficult to track.

To address this issue, OrgReg distinguishes between *levels of entities* on the one hand and provides for a separate table characterizing *linkages between entities* on the other hand. Both features required substantial methodological effort.

First, similarly to the distinction between groups and enterprises in the business units registers, OrgReg distinguishes between three entity levels, i.e. groups of organizations, organizations and components, meaning units that are part of a larger organization.

Consistently with the focus on the organizational level, components should be singled out only in exceptional cases, since OrgReg does not aim to delve into the intra-organizational structure of public research. As a matter of fact, OrgReg includes only 323 components out of 6,003 entities. Most of them are centers that have their own visibility, such as the Swiss Supercomputing Center (a component of the ETH Zurich), some Max Planck centers in Germany and Research Councils’ facilities in the UK, such as The Rutherford Appleton Laboratory. Some components are university hospitals that are legally part of the university, in order to treat them on a similar footing as hospitals that are legally independent, but closely associated with the university. We remark that the distinction between components and affiliated entities is sometimes difficult, particularly concerning institutes affiliated to universities. In many countries, efforts to consolidate the public research sector implied that previously independent institutes were associated with universities, but the exact form and how this should be coded is not always straightforward – these institutes might for example keep their legal personality, but delegate decisions concerning appointments to the university board. Admittedly, the notion of control is difficult to assess concerning public-sector organizations, in which the extent of autonomy might differ between administrative and academic decisions, while resource dependency may be more relevant than formal control.

Further, OrgReg currently includes 183 groups, which are heavily concentrated in France (44) and in the UK(102). In both countries, the main occurrence are hospital groups, i.e. the National Health Service Trust in the UK and the Regional/Hospital Centers in France. The second important case are the French Communities of Universities (COMUEs): their status is complex as some enroll students or PhD students, others are simple coordination structures. Moreover, COMUEs are subject to continuous change, such as the entry of additional partners. The list of groups includes also few private university groups and some ‘umbrella’ PROs, such as the Helmholtz Society in Germany and the Czech Academy of Sciences. Additionally, OrgReg includes few (15) consortia of national relevance that have official recognition at the country level, but, unlike groups, do not control their members. The most important case is the Leibnitz Association in Germany. Admittedly, the distinction between groups and consortia is not always clear-cut.

A major issue for OrgReg was how to cover large ‘umbrella’ organizations that are composed by multiple laboratories, such as CNRS in France and the Academy of Sciences in Eastern Europe. Consistent with a focus on organizations, OrgReg decided to treat these entities depending on whether the laboratories are autonomous organizations or not. Therefore, CNRS and Max Planck, in which laboratories are within a common hierarchical structure, have been coded as organizations, while individual laboratories are not included in the register. On the contrary, Helmholtz laboratories are formally independent and have their own visibility – a prime example being the German Aerospace Center DLR – and are therefore included individually, while Helmholtz is coded as a group.

In practice, the distinction was not always easy to draw, particularly for some of the Academy of Sciences in Eastern Europe that are in the transition from hierarchical structures to more loosely coupled groupings of laboratories. A useful criterion was provided by patents since these are usually filed by legal entities. Indeed, data showed that, for CNRS, almost all patents are under the name of the organization, while for Helmholtz most patents are filed by individual centers.

Second, OrgReg includes a table with linkages between organizations, defined as stable and structural relationships that are lasting over time. Distinguishing between structural linkages and weaker ties such as strategic alliances was a clear concern here, also for practical purposes. Currently, OrgReg, distinguishes between three linkages types: *membership*, where a component (for example a research institute) is directly controlled by a parent unit; *affiliation*, meaning a strong structural relationships between two independent units which share some essential elements; *association*, a structural but less strong relationships than affiliation.

For example, in the case of university hospitals, affiliated hospitals share a common educational and research mandate and/or have strong integration within the medical faculty – for example hospital staff is also affiliated with the university -, while associated hospitals still cooperate in education, but there is little integration in research and staff. This distinction is based on previous work at the University of Leiden on the identification of university hospitals¹ and allows for a fine-grained understanding of the types of relationships between universities and hospitals. Within OrgReg, the coverage of research hospitals has been further extended, since the extent of medical research performed within hospitals without a connection with universities is substantial in countries such as Italy and the UK.

The current number of linkages is 1,844, a figure that is certainly underestimated, but shows how important such relationships are in the public sector. Most of them are in France (508) and in the UK (268). While in the UK this is essentially accounted by the hospital sector, in France this high number reflects the structure of national research system that is strongly based on linkages and joint units between research performers (Mustar and Larédo 2002).

¹ <https://www.cwts.nl/blog?article=n-q2w264>

Most of the tracked linkages are membership linkages (1,143), which allow identifying the relationships between groups and member organizations, respectively between organizations and their components. The remaining linkages are either affiliation (355) or association (317), most of them being between university and university hospitals.

An important type of linkages which is not yet covered by OrgReg are 'joint' or mixed units between two organizations. Such structures are pervasive in France, where most of the CNRS laboratories are joint with universities (Mustar and Larédo 2002), but some cases were found also in other countries. Clearly, the scope of a register can only be to identify such situations, without entering into the complex issue of how to attribute inputs and outputs. Two variants can be envisaged, one tracking the linkage at the organizational level, one including also information on the specific units. In the first option, most French universities would be linked to CNRS in an unspecific way, while the second would imply tracking more than 1,000 units only in France. The first option is therefore uninformative, while the second would be very complex to implement – also taking into account the longitudinal nature of the register.

5 Tracking organizations over time

Organizations change over time, new organizations are created, some are closed and some are merged together. While these issues are well known in the private sector, where demographic turnover is a major issue for registers, the more stable structure of public research implied that demography was largely overlooked. Yet, after 2000, many countries undertook reforms of public-sector research through extensive mergers and consolidation processes to achieve critical mass and international visibility (Pinheiro, Geschwind and Aarrevaara 2016), hence the need to document more systematically these processes. Dealing with demography in a proper way is also important to construct panels for longitudinal analysis.

Demographic change in registers can be observed either in a discontinuous or in a continuous way: in the first approach, the one followed by the US Census Bureau for the firms' register, the register is composed by distinct time strata, but organizations that are present in more than one strata are connected through unique identifiers and, in case of mergers or take-overs, organizations are linked through parent-child linkages. In the discontinuous approach, strata can be more than one year apart and, hence, changes between two strata are not observed, reducing the burden for data collection particularly in case of rapid change as observed for firms.

In the continuous approach, the register is composed of a list of entities, while demographic events are recorded in a separate table - for example, in the case of a merger the merged organizations is attributed a closure year corresponding to the merger, while a new ID is attributed to the new organization created with the merger with foundation year corresponding to the merger. A yearly panel can be reconstructed by combining the entities and demographic events table. This approach is preferred when demographic events are of interest as such, while their number is reasonably limited.

This approach requires a clear-cut distinction between demographic events that involve the creation of a new entity and the change in existing entities, including the change of name, legal seat, legal status, which does not involve a change in the identifiers.

Quantitative data has been complemented with descriptive information on the complex history of many European research organizations, which constitutes an incredibly rich material for the analysis of the evolution of the system over time.

6 OrgReg in practice

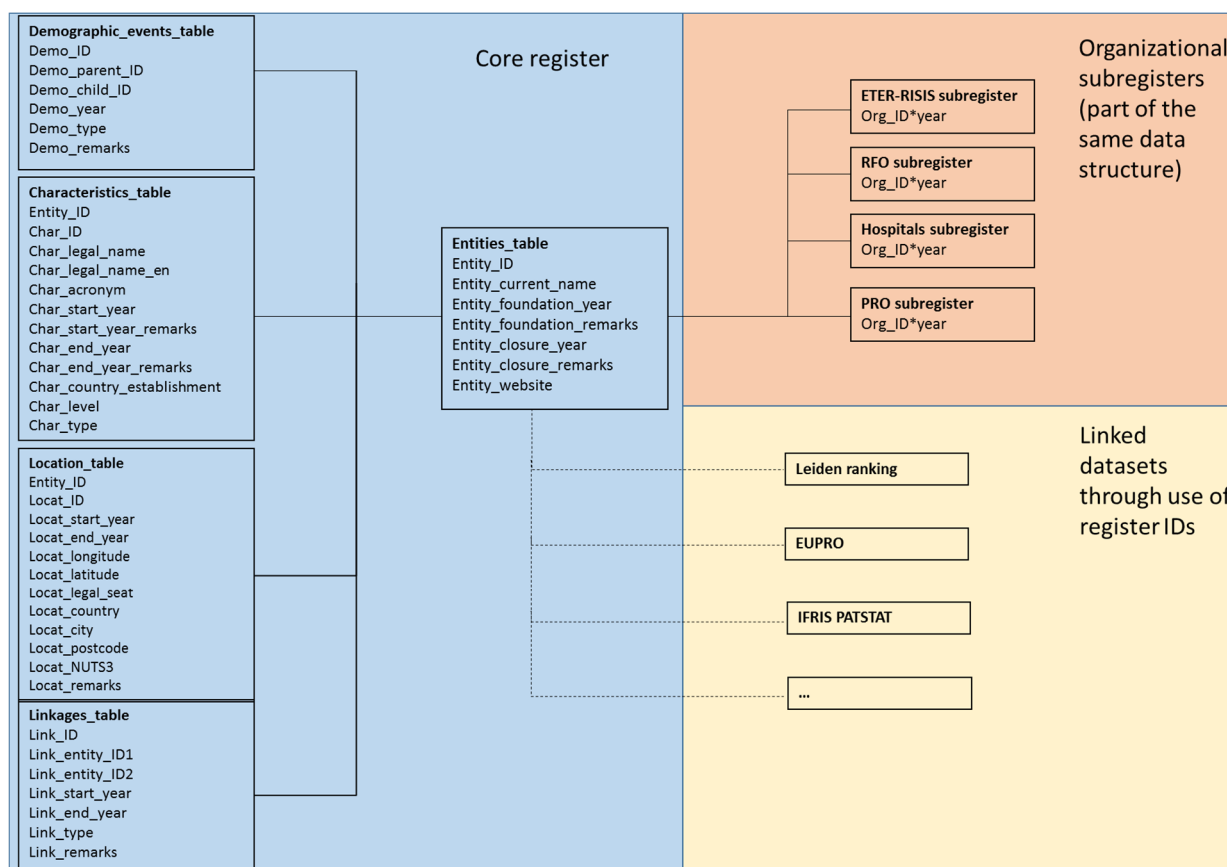
Two main choices grounded the design of OrgReg. First, while achieving some level of standardization, the design should leave flexibility to allow for future changes and extensions. Second, a register is not made to collect statistical data, but only some key variables for the identification of organizational structures.

Accordingly, the core element of OrgReg is a table listing all the entities and attributing them a unique ID of the form (country code + 4 digits) (Figure 1). This table includes only those elements that are by definition unique for an entity, specifically foundation and closure year, as well as the current website and name for identification purposes, while all other attributes are stored in separate table and, by construction, can be non-unique, i.e. entities can have multiple sets of characteristics during non-overlapping periods of time.

Beside name changes, this applies to entities that became part of a group, such as the telecommunication institutes in France or, even more complex, the Greek Institute of Geology and Mineral Exploration that was first taken-over by another entity and then became independent again. The OrgReg structure allows reclassifying entities as of their type or level without reassigning IDs, while new classifications can be added. This design is essential to deal with the methodological issues highlighted in the previous sections.

A similar approach was adopted for the location table. Entities can have different locations over time and during the same time period, the only constraint being that the main seat is unique at a time. Therefore, OrgReg was constructed from the beginning in order to track multi-campus HEIs and large national PROs such as the CNRS. To some extent, multi-location is meant to replace the identification of individual laboratories. The corresponding information has been mostly worked out from institutional websites, but an alternative approach that will be tested is clustering the addresses of scientific publications. In the current version of OrgReg, about 20% of the entities have multiple locations (in a different NUTS-3 region than the one of the main seat).

FIGURE 1. ORGREG STRUCTURE AND MAIN LINKAGES



Consistently with its function of interlinking datasets, OrgReg is a public resource that is accessible upon registration from a dedicated website (orgreg.joanneum.at) that provides a search interface and a tool to download the data. Data can be downloaded in two formats, either as a list of all active entities for a period of time or as panel data, showing by each year the active entities with their respective characteristics.

The infrastructure can also be interfaced through Advanced Programming Interfaces (API), a feature that will allow for automated interlinking of datasets.

6.1 Constructing OrgReg

Unlike the Business Units Register, in which the main unit, i.e. the ‘enterprise’, is a purely statistical construct, OrgReg claims to identify ‘organizations’ as they exist in reality, therefore assuming a realist conception of statistical units (Desrosières 2016). Of course, this claim may be debated since, as discussed in the previous section, there is always an element of conventionality in the identification of (public) organizations.

But, methodologically, this understanding of OrgReg was translated into an approach where the existence and characteristics of the OrgReg entities was checked against sources that are supposed to provide information on the organizations themselves, including institutional websites, Wikipedia entries and the knowledge from national experts. Of course, some of these practices are also adopted by database owners to standardize their organizations, but it is fair to say that the combination of different sources and the systematic checking with national experts are a unique feature of OrgReg.

Lists of candidate entities were extracted from different sources: the European Tertiary Education Register (ETER; Lepori, Bonaccorsi, Daraio, et al 2015) for HEIs, the list of institutions classified in public-research in the WoS database at CWTS in Leiden (Waltman, Calero-Medina, Kosten, et al 2012), as well as in the EUPRO database of European projects (Roediger-Schluga and Barber 2008). The combined list was then manually verified by national experts. In this process, many organizations were excluded, as the list also included a number of firms, duplicates and university institutes. The primary list of university hospitals was derived from the CWTS Web of Science database, but was enriched by national experts, particularly in France and Italy where most of clinical research is performed by hospitals not directly associated with universities. Similarly, the basic list of RFOs was derived from the Public Research Funding contract from the EU (Lepori, Reale and Spinello 2018), but was then revised and harmonized with OrgReg definitions.

These sets of entities were then integrated in a common list for each country, plus a specific list of international organizations, such as the EU Joint Research Center or CERN. In a further step, national experts added the linkages and recollected descriptors; demographic changes were then reconstructed by checking manually all entities founded after the year 2000. National experts reviewing the lists were left the possibility of adding organizations based on their importance in the system, which happened especially for some Central and Eastern European Countries.

The process was manageable since the number of entities is reasonably small in most European countries. Only France and UK proved to be very time-consuming because of the large number of entities (more than 500 each), but also of the presence of complex structures and linkages between different types of organizations.

This process was highly interactive, involving a continuous exchange between the experts dealing with individual countries and the core team of people involved in the design of OrgReg; unclear and ambiguous cases were discussed collectively in order to come to a shared solution by taking also into account their general implications. In a number of cases, adaptations in the OrgReg handbook were also needed. A further round of standardization was performed centrally, by checking systematically organizational websites and Wikipedia entries, while also correcting different ways of writing cities or imprecise geographical coordinates.

Such quality control processes are central for the robustness of the register, but also for its legitimacy in terms of a trusted representation of organizational reality (Desrosières 2016).

In this respect, there is ample potential for further automation in the retrieval of information and in data quality. Currently, a simple set of consistency checks is implemented to detect mistakes, such as organizations being merged, but without a correct foundation year for the new organization. However, there is a potential for retrieving automatically information on foundation years and geographical information, as well as for early detection of changes – the experience shows for example that websites are frequent markers of demographic events. Automated procedures will also be required to maintain consistency between OrgReg and other databases and have a potential for pre-processing and for reducing the burden for maintenance, but final decisions will be made by experts, particularly on the most complex cases.

Relatedly, there is a need for a more systematic documentation of why certain choices have been made and of tracking changes over time, for example in the classification of entities and for what concerns their inclusion and exclusion. Such a documentation will be important for many purposes: to avoid repeating the same processes or changing back decisions without a clear rationale; to allow methodological consolidation in the long term through ex-post analysis of cases; finally, as a source for ethnographic analysis of the underlying principles and assumptions of the register. A more systematic set of remarks (at the entity level) and of metadata (at the country or variable level) is therefore a central future extension of the register.

6.2 OrgReg as an interlinking tool

OrgReg is not meant to be an isolated facility, but a tool to interconnect different data sources. Three levels have been identified (see Figure 1).

First, the creation of organizational subregisters that store statistical data on the OrgReg entities and share identifiers and basic information. This is currently the case with ETER that shares the same identifiers as OrgReg and provides administrative data such as finances, staff and students on Higher Education Institutions, but similar datasets could be developed in the future for PROs, research hospitals and research funding organizations.

Second, a number of core datasets within the RISIS project, including those on scientific publications, European projects and patents, are connected to OrgReg by sharing the organizational identifiers. A key feature of these datasets is the standardization of organizations that is required for analytical purposes, but also to produce robust indicators. The construction of OrgReg however displayed a number of inconsistencies in the identification of public-sector organizations, with some companies being included in the public sector or the incomplete treatment of demographic events. In this respect, OrgReg has the potential to improve organizational standardization within each dataset, but, more importantly, across datasets. Such a process will have to be maintained over time by tracking and standardizing new organizations that appear in the interlinked datasets and integrating them into OrgReg.

This matching of OrgReg generates the possibility of providing within OrgReg aggregated indicators on key research outputs for all public research organizations in Europe. While such indicators could also be retrieved from the parent datasets, similarly to the scientific publication indicators provided by the Leiden Ranking, providing for all research organizations in Europe a core set of output data, covering both science and technology, will represent a major advance in terms of data availability to end users.

Third, having a closed list of organizations is a valuable tool in order to search for information from the Internet and social media, as it makes disambiguating names much easier. Besides the organizational name (in English and national language), localization information and the website, OrgReg also includes organizational acronyms as experience showed that they are frequently used when referring to the organization. The experience made in matching OrgReg with institutional profiles in ResearchGate (Lepori,

Thelwall and Hoorani 2018) showed that this information is sufficient for a precise matching, without having to store in the register name variants, with the exception of (major) changes in the official name.

7 European public-sector research as seen from OrgReg

Table 1 provides an overview of the current status of OrgReg². The figures are expected to increase slightly in the future. On the one hand, OrgReg is continuously updated based on feedback received from the project partners, for example concerning new demographic events taking place. On the other hand, the project team is still completing the coverage of Research Funding Organizations and of large ‘umbrella’ PROs, as well as of a few countries where coverage is not fully complete, notably in Eastern Europe. However, we have ample evidence that coverage is now very extensive for what concerns public entities with a sizeable output in terms of tertiary education students, scientific publications and patents. Student coverage in tertiary education, as compared with EUROSTAT, is well above 90% for most countries, while very few entities not included in OrgReg could be identified in publications and patents databases.

By design, the coverage of OrgReg is by and large limited to ‘classical’ knowledge production manifested in outputs such as graduations, publications and patents, as well as tertiary education. actors involved in social innovation, such as citizens’ associations and grassroots movements (Edwards-Schachter and Wallace 2017), are not covered as these cannot be easily identified in the data sources used.

A first relevant outcome of the establishment of OrgReg is that the number of public research and higher education institutions in Europe does not exceed a few thousands entities; this is a low figure as compared, for example with firms, where preliminary work on a parallel firms’ register (FirmReg) showed that the number of innovative firms in patent databases might be around 100,000. The figure for the public sector would be even lower when considering research only, as among the 4,500 HEIs covered only about 1,000 have a sizeable publication and/or patent output (Lepori, Geuna and Veglio 2017). The availability of data on scientific and technological output will allow qualifying more precisely this statement and identifying the core of European public research,

TABLE 1 ORGREG IN A NUTSHELL

DATA REFER TO THE VERSION OF ORGREG DOWNLOADED 20.11.2019.

Total entities included (all years)	6,002 entities 1,573 entities founded from the year 2000, 854 closed between 2000 and 2018. 706 In Germany, 700 in France, 599 in Poland, 544 in the UK, 388 in Italy, 251 in Spain.
Covered countries	39 (plus international organizations). EU-28, Iceland, Liechtenstein, Norway, Switzerland, Israel, North Macedonia, Montenegro, Kosovo, Serbia, Bosnia-Herzegovina, Albania, Turkey.
Entities included by type (2016)	4,332 3,437 HEIs, 876 PRO, 676 research hospitals, 87 Public Administration, 86 PNP, 21 RFOs, 155 mixed..
Entities included by level (2016)	4,853 organizations, 178 groups, 287 components, 12 consortia.
Demographic events by type (2000-2016)	468 unique events 195 merger, 41 spin-out, 10 split, 222 take-over.
Linkages by type (2000-2016)	1844 linkages 1143 membership, 356 affiliated, 317 associated, 26 join units. 508 linkages in France, 268 in the UK, 203 in Germany, 104 in Bulgaria.
Access	OrgReg is a public resource that can be accessed and downloaded on-line upon pre-registration at OrgReg.joanneum.at

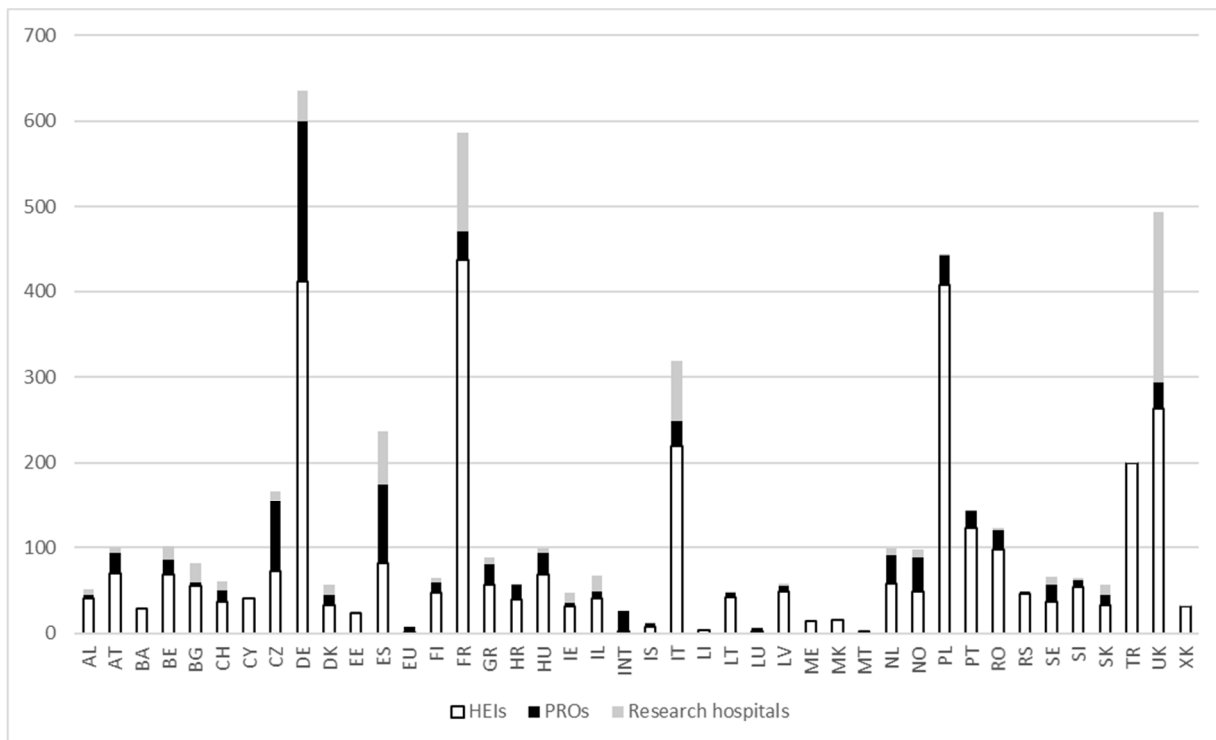
² All data refer to the 2019 edition of OrgReg, as downloaded 20.11.2019.

In terms of organizational types, the largest component in OrgReg are Higher Education Institutions (see Figure 2). While these are indeed core actors in public research, their inclusion also emphasizes how human capital formation is a core task of public-sector research and is broader than research and technological development, for example in terms of regional coverage (Lepori, Guerini, Scherngell and Larédo 2019).

As suggested by previous descriptive studies, the world of Public Research Organizations is highly heterogeneous (Cruz-Castro and Sanz-Menéndez 2018). Its core is composed by a dozen large ‘umbrella’ organizations, either engaged in basic sciences such as CNRS, Max Planck and the Academy of Sciences in Eastern Europe, or in technological development, such as the Fraunhofer Society in Germany. Additionally, the PRO landscape comprises large national facilities (Rutherford in the UK, Swiss Supercomputing Centre), policy-oriented and service organizations to the public sector (Danish Institute for Local and Regional Government Research; Scottish Crop Research Institute) and a sizeable number of Research and Technology Organizations (VITO - Flemish Institute for Technological Research), bridging the public and private sector. In that respect, the novelty of OrgReg is to provide the basis for a systematic comparative analysis of the PRO landscape, complemented with measures of their contribution to scientific and technological output (Sanz-Menendez et al., this issue).

FIGURE 2. ENTITIES IN ORGREG BY TYPE AND COUNTRY

ONLY THE THREE MOST FREQUENT TYPES ARE DISPLAYED. ORGANIZATIONS EXISTING IN THE YEAR 2016, N= 4980.



The third main type of organizations in OrgReg are research hospitals, a frequently overlooked component of public research given their close linkages with universities. OrgReg currently includes 676 hospitals and this number is likely to further increase since current coverage in Eastern Europe is only partial.

A preliminary analysis of the information on linkages displays some interesting findings. First, only 69 hospitals in OrgReg do not have a form of association with a university, most of them being specialized hospitals in domains such as in cancer or pediatric care. These cases are concentrated in Italy (30 hospitals) and France (15). Second, in most European countries, including Austria, Germany, Switzerland, Italy, Netherlands and Norway, the affiliated hospital, which is an independent legal entity, but structurally connected to a university, is by large the prevalent model. Major exceptions to this pattern are France, with the University Hospital Centers, and UK, with the NHS foundation trusts; in both cases, public healthcare is

organized in consortia managing multiple hospitals and having association with different universities within the same region. The most notable case is the Paris Hospital Trust that comprises 44 hospitals, of which 12 are recorded in OrgReg since they have a sizeable publication output; most of them closely cooperate with the universities in the Paris area.

Particularly when complemented on data on co-publications between hospitals and universities, OrgReg has therefore an important potential to analyze the forms of organization of healthcare research and its relationships with universities.

Finally, Orgreg includes only 87 entities within the public administration and 86 Private Non Profit entities. In both cases, the coverage might not fully complete, as these entities are not straightforward to identify and, for instance, to distinguish from private companies for PNP.

The public administration entities are largely concentrated in Italy and France, i.e. countries with a traditionally centralized state (Bleiklie and Michelsen 2013) and in a few domains where the state has a direct regulatory role or public interest, namely defense (France, UK), meteorology (Belgium, Switzerland, Finland), agriculture and veterinary (Finland, Cyprus, Denmark, Germany), health and food safety (France, Austria, Italy) and official statistics (Netherlands, Switzerland).

Finally, public non-profit organizations are strongly concentrated in health, the most notable cases being the Ludwig Institute for Cancer Research and the French Pasteur Institute.

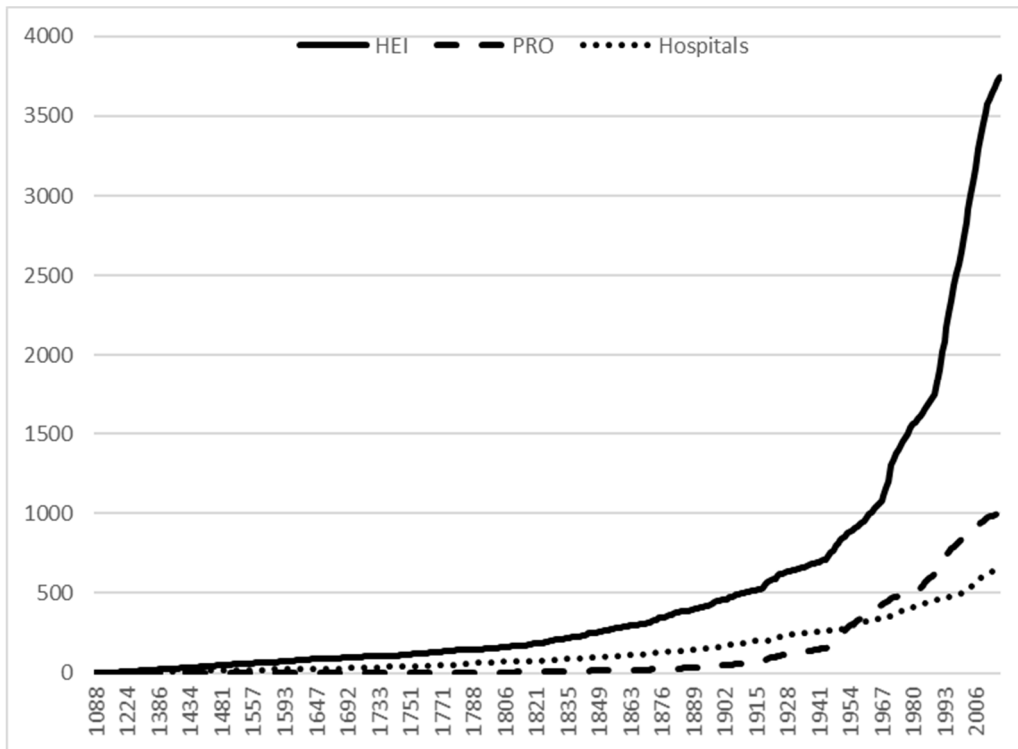
7.1 Organizational dynamics and geographic distribution of public research

For the first time, OrgReg provides systematic information about two relevant aspects of public-sector research, i.e. the organizational dynamics over time and the spatial distribution of (different types of) public research organizations.

On the first aspect, Figure 3 shows that public research organizations tend to be rather young – the median age in OrgReg is only about 30 years, even if we need to take into account that some might have older ancestors. We also observe different historical dynamics depending on the type of organization considered. While there is a sizeable number of universities, which have been founded before the XIX century, most Public Research Organizations date to the XIX and XX century, displaying their close link with national states. Among the 225 organizations in OrgReg founded before the year 1800, 154 are HEIs, 63 research hospitals and only 2 PROs, i.e. the Bavarian Academy of Sciences dating back to 1759 and the Czech Academy of Sciences 1784). While most of the top-ranked international universities, including Oxford, Cambridge, Heidelberg, Paris, date back to the Middle Ages, the most representative PROs date back to the first half of the XX century, such as CNR (1923), CNRS (1939) and Max Planck (1948).

At the same time, the higher education sector witnessed an incredible expansion in the last few decades, pushed by the increase in enrolments and by the differentiation of higher education, with the integration of professional education (Lepori 2020) – half of the HEIs in OrgReg have been founded after 1990. In turn, some research hospitals in France, UK and Italy date back to the Middle Ages, but the increase has been rather slow, hinting to the fact that the research function remained limited to a core of hospitals.

FIGURE 3. ORGREG ORGANIZATIONS BY THEIR FOUNDATION YEAR
ONLY THE THREE MOST FREQUENT TYPES ARE DISPLAYED. N= 5383.



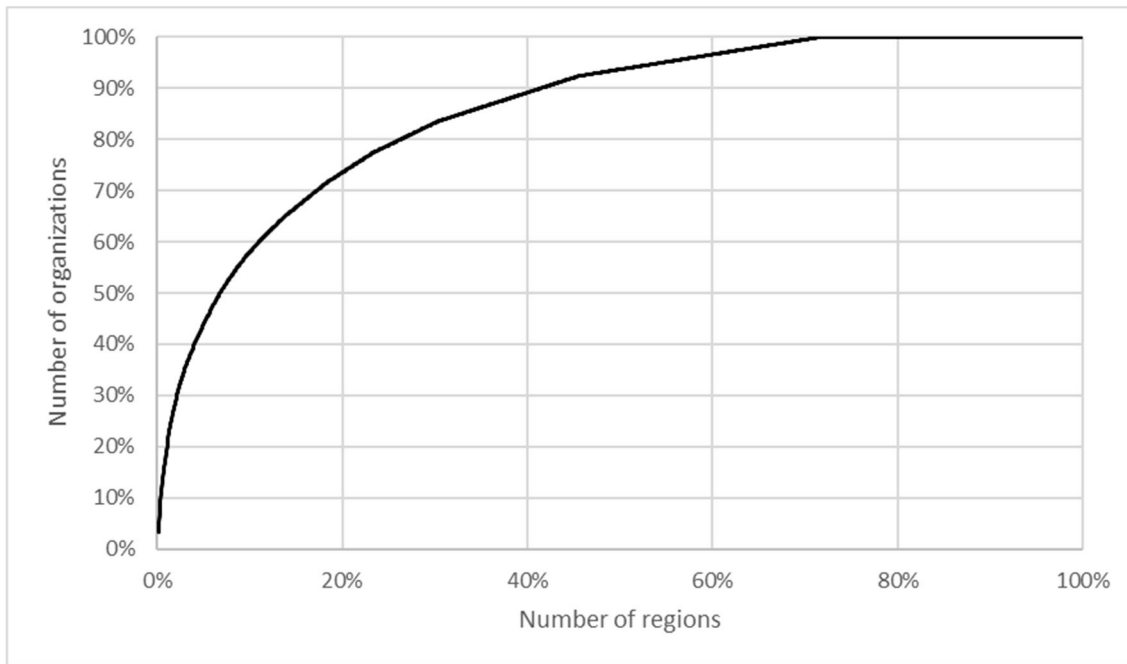
Beyond this long-term information, OrgReg allows for a systematic observation of demographic and name changes over the period 2000-2019. The extent of observed changes is fairly surprising as compared with an overall idea of stability of the public sector. OrgReg includes 468 demographic events, of which 195 mergers, involving about one-quarter of all organizations in the register, showing how demographic turnover was substantial on a two decades timeframe (see Heller-Schuh and Lepori in this issue). Additionally, OrgReg also includes more than 1,000 name changes. Most name changes are either colleges which included ‘university’ in their name, the most frequent occurrence being Fachhochschulen or Polytechnics relabeled in English as ‘universities’ and private colleges renamed as universities. This information is important for identifying organizations over time, but is also informative of changes in the status and self-presentation of organizations.

Concerning geography, OrgReg provides for a systematic localization of the seat of research organizations at the city level; this information is being extended to include also satellite campuses of HEIs, as well as multiple locations of large PROs such as CNRS.

Data display a broad regional coverage, but at the same time strong concentration in metropolitan regions. More than 1,100 NUTS3 level regions out of 1,500 hosted at least a location of a public-research organization. This geographical spread is essentially due to HEIs, which were present in more than 1,000 regions, against only about 300 regions each for PROs and research hospitals. The research function is therefore more spatially concentrated than higher education.

At the same time, half of the locations were situated in 150 regions only, including most capital cities such as Warsaw (120 locations) and Paris (117 locations), as well as some large non-capital cities like Milan or Barcelona. Overall, 20% of the regions hosted more than 70% of the organizations in OrgReg (Figure 4). Particularly when complemented with size indicators of the considered organizations, these data have a potential for in-depth analyses of the structure of European public research overlying organizational structure to spatial information.

FIGURE 4. DISTRIBUTION OF ORGREG ENTITIES OVER NUTS3 REGIONS
ORGANIZATIONS EXISTING IN 2016, N= 5332.



8 Conclusion

OrgReg represents a significant advance in the construction of a data infrastructure for European public-sector research and higher education. For the first time, users can dispose of a reference list of organizations that follows clear methodological principles and has been systematically checked against different sources. OrgReg also documents to an unprecedented level inter-organizational linkages and demographic changes in European public research, allowing therefore a deeper understanding of the organizational diversity of public systems in Europe. Finally, OrgReg represents an important tool for database interlinking and for supporting data collection from different sources, including the Internet.

As highlighted in this paper, the success of OrgReg is also due to choices that limit its scope in terms of what entities are covered, but also of how much the complexity of organizational settings is represented in the register. A key parameter in dealing with methodological issues cases was the trade-off between taking into account the complexity of systems on the one hand and the goal of providing a manageable helicopter view of European public research on the other hand.

In this paper, we have also provided some illustrative examples of the analyses that could be performed by using OrgReg, particularly when combined with interlinked data on organizations. In this respect, the planned integration within OrgReg of data on the research and educational output of organizations will represent a major leap forward.

9 References

- Abramo, G. & D'Angelo, C. A. (2016). A farewell to the MNCS and like size-independent indicators. *Journal of Informetrics*, 10(2), 646-651.
- Becher, T. & Trowler, P. R. (2001). *Academic Tribes and Territories. Intellectual enquiry and the culture of disciplines* Ballmoor, Buckingham / Philadelphia, PA: The Society for Research into Higher Education and Open University Press.
- Bleiklie, I. & Michelsen, S. (2013). Comparing HE policies in Europe. *Higher Education*, 65(1), 113-133.

- Bonaccorsi, A. & Daraio, C. (2007). *Universities and Strategic Knowledge Creation. Specialization and Performance in Europe* Cheltenham: Edward Elgar.
- Bozeman, B. & Bretschneider, S. (1994). The “publicness puzzle” in organization theory: A test of alternative explanations of differences between public and private organizations. *Journal of Public Administration Research and Theory*, 4(2), 197-224.
- Braun, D. (1998). The role of funding agencies in the cognitive development of science. *Research Policy*, 27 (1998), 807-821.
- Brunsson, N. & Sahlin-Andersson, K. (2000). Constructing organizations: The example of the Public Sector Reform. *Organization Studies*, 21(4), 721-746.
- Capano, G. (2011). Government continues to do its job. A comparative study of governance shifts in the Higher Education Sector. *Public Administration*, 89(4), 1622-1642.
- Carayol, N. & Matt, M. (2004). Does research organization influence academic production? *Research Policy*, 33(8), 1102.
- Clark (1996). Diversification of higher education: Viability and change. In V. L. Meek, L. Goedegebuure, O. Kivinen & R. Rinne(Eds.) *The Mockers and the Mocked: Comparative Perspectives on Differentiation, Convergence and Diversity in Higher Education* (pp. 16-25). Pergamon Press.
- Crow, M. & Bozeman, B. (1998). *Limited by design: R&D laboratories in the US national innovation system* Columbia University Press.
- Cruz-Castro, L. & Sanz-Menéndez, L. (2018). Autonomy and Authority in Public Research Organisations: Structure and Funding Factors. *Minerva*, 56(2), 135-160.
- Cruz-Castro, L., Sanz-Menéndez, L. & Martínez, C. (2012). Research centers in transition: patterns of convergence and diversity. *The Journal of Technology Transfer*, 37(1), 18-42.
- Davies, S. M., Tawfik-Shukor, A. & de Jonge, B. (2010). Structure, governance, and organizational dynamics of university medical centers in the Netherlands. *Academic Medicine*, 85(6), 1091-1097.
- de Boer, H., Enders, J. & Leisyte, L. (2007). Public Sector Reform in Dutch Higher Education: the Organizational Transformation of the University. *Public Administration*, 85(1), 27-46.
- De Rassenfosse, G., Dernis, H., Guellec, D., Picci, L. & de la Potterie, Bruno van Pottelsberghe (2013). The worldwide count of priority patents: A new indicator of inventive activity. *Research Policy*, 42(3), 720-737.
- Desrosières, A. (2016). *La politique des grands nombres: histoire de la raison statistique* La découverte.
- Edwards-Schachter, M. & Wallace, M. L. (2017). ‘Shaken, but not stirred’: Sixty years of defining social innovation. *Technological Forecasting and Social Change*, 119, 64-79.
- Eurostat (2003). Recommendation manual on business registers.
- Ferlie, E., Ashburner, L., Fitzgerald, L. & Pettigrew, A. (1996). *The New Public Management in Action* Oxford: Oxford University Press.

- Ferlie, E., Musselin, C. & Andresani, G. (2008). The steering of higher education systems: A public management perspective. *Higher Education*, 56(3), 325.
- Fumasoli, T. & Lepori, B. (2011). Patterns of strategies in Swiss Higher Education Institutions. *Higher Education*, 61(2), 157-178.
- Geuna, A. & Martin, B. R. (2003). University research evaluation and funding: An international comparison. *Minerva*, 41(4), 277-304.
- Gulbrandsen, M. (2011). Research institutes as hybrid organizations: central challenges to their legitimacy. *Policy Sciences*, 44(3), 215-230.
- Heinze, T., Shapira, P., Rogers, J. D. & Senker, J. M. (2009). Organizational and institutional influences on creativity in scientific research. *Research Policy*, 38(4), 610-623.
- Hicks, D. (2012). Performance-based university research funding systems. *Research Policy*, 41(2), 251-261.
- Jonkers, K. & Cruz-Castro, L. (2010). The internationalisation of public sector research through international joint laboratories. *Science and Public Policy*, 37(8), 559-570.
- Knorr Cetina, K. (2009). *Epistemic cultures: How the sciences make knowledge* Harvard University Press.
- Knorr Cetina, K. (1995). Laboratory studies: The cultural approach to the study of science. *Handbook of Science and Technology Studies*, , 140-167.
- Kyvik, S. & Lepori, B. (2010). *The Research Mission of Higher Education Institutions Outside the University Sector. Striving for Differentiation* Dordrecht: Springer.
- Larédo, P. & Mustar (2001). Research and Innovation policies in the new global economy. An international comparative analysis.
- Laudel, G. (2006). The art of getting funded: how scientists adapt to their funding conditions. *Science and Public Policy*, 33(7), 489-504.
- Lepori (2020). Non-university education and professional institutions. In J. C. Shin & P. Teixeira(Eds.) *Encyclopedia of International Higher Education Systems and Institutions* Dordrecht: Springer.
- Lepori, B., Baschung, L. & Probst, C. (2010). Patterns of subject mix in higher education institutions: A first empirical analysis using the AQUAMETH database. *Minerva*, 48(1), 73-99.
- Lepori, B., Bonaccorsi, A., Daraio, A., Daraio, C., Gunnes, H., Hovdhaugen, E., Ploder, M., Scannapieco, M. & Wagner-Schuster, D. (2015). *Establishing a European Tertiary Education Register. Final Report* Brussels: European Commission.
- Lepori, B., Geuna, A. & Veglio, V. (2017). *A typology of European Universities. Differentiation and Resource Distribution* Brighton: University of Sussex, SPRU Working Paper Series.
- Lepori, B., Guerini, Scherngell & Larédo (2019). The spatial distribution of knowledge production in Europe. Evidence from KET and SGC. *Proceedings of the 17th International Conference on Scientometrics & Informetrics*, , 685-690.

- Lepori, B., Reale, E. & Spinello, A. O. (2018). Conceptualizing and Measuring Performance Orientation of Research Funding Systems. *Research Evaluation*, 27(3), 171-183.
- Lepori, B., Thelwall, M. & Hoorani, B. H. (2018). Which US and European Higher Education Institutions are visible in ResearchGate and what affects their RG score? *Journal of Informetrics*, 12(3), 806-818.
- Lepori, B., Wise, M., Ingenhoff, D. & Buhmann, A. (2016). The dynamics of university units as a multi-level process. Credibility cycles and resource dependencies. *Scientometrics*, 109(3), 2279-2301.
- Light, R. P., Polley, D. E. & Börner, K. (2014). Open data and open code for big science of science studies. *Scientometrics*, 101(2), 1535-1551.
- Musselin, C. (2013). *Long March of French Universities* Routledge.
- Musselin (2007). Are universities specific organisations? In G. Krücken, A. Kosmützky & M. Torke (Eds.) *Towards a Multiversity? Universities between Global Trends and National Traditions* (pp. 63-84). Bielefeld: transcript.
- Mustar, P. & Larédo, P. (2002). Innovation and research policy in France (1980-2000) or the disappearance of the Colbertist state. *Research Policy*, 31 (2002), pp. 55-72.
- OECD (2015). *Frascati Manual 2015. Guidelines for Collecting and Reporting Data on Research and Experimental Development* Paris: OECD.
- Pfeffer, J. (1997). *New Directions for Organization Theory* New York: Oxford University Press.
- Pinheiro, R., Geschwind, L. & Aarrevaara, T. (2016). Mergers in higher education. *European Journal of Higher Education*, 6(1), 2-6.
- Roediger-Schluga, T. & Barber, M. (2008). R&D collaboration networks in the European Framework Programmes: data processing, network construction and selected results. *International Journal of Foresight and Innovation Policy*, 4(3-4), 321-347.
- Scott, W. R. (2008). *Institutions and organizations. Ideas and interests* Thousands Oaks: Sage.
- Seeber, M., Lepori, B., Montauti, M., Enders, J., De Boer, H., Weyer, E., Bleiklie, I., Hope, K., Michelsen, S. & Mathisen, G. N. (2015). European universities as complete organizations? Understanding identity, hierarchy and rationality in public organizations. *Public Management Review*, 17(10), 1444-1474.
- Teixeira, P., Rocha, V., Biscaia, R. & Cardoso, M. F. (2014). Public and private higher education in Europe: competition, complementarity or worlds apart? *Knowledge, Diversity and Performance in European Higher Education: A Changing Landscape, UK and Northampton, MA, USA: Edward Elgar*, 84-105.
- Waltman, L., Calero-Medina, C., Kosten, J., Noyons, E. C., Tijssen, R. J., van Eck, N. J., van Leeuwen, T. N., van Raan, A. F., Visser, M. S. & Wouters, P. (2012). The Leiden Ranking 2011/2012: Data collection, indicators, and interpretation. *Journal of the American Society for Information Science and Technology*, 63(12), 2419-2432.