

PON RICERCA E COMPETITIVITA' 2007 - 2013

DALL'OPEN DATA ALL'OPEN GOVERNMENT: BUONE PRATICHE DI RIUTILIZZO DEI DATI

L'Autorità di Gestione del PON R&C, interpretando correttamente quanto richiesto dal Regolamento (CE) 1828/2006 (artt. 6-7), fin dall'avvio della Programmazione ha garantito la massima trasparenza e visibilità delle informazioni relative ai progetti finanziati.

Nel periodo 2007-2013 si è passati dalla semplice pubblicazione delle informazioni relative ai beneficiari dei progetti all'implementazione e alla messa online di una banca dati con tutti i progetti finanziati.

La spinta internazionale verso l'*open government*, che va ben oltre le attuali prescrizioni normative a livello nazionale in materia di trasparenza (D.Lgs n. 33 del marzo 2013), è stata il punto di riferimento per l'Amministrazione nella gestione del PON R&C. In questa ottica, nel 2012 l'Autorità di Gestione ha reso disponibili i dati di tutti gli interventi del programma sul sito www.ponrec.it - con informazioni sui progetti finanziati sempre più dettagliate e aggiornate con cadenza bimestrale – secondo i principi dell'opendata. In particolare, i *datasets* sono stati pubblicati nel formato .csv secondo i termini della licenza Creative Commons 3.0 (CC-BY) con la possibilità di utilizzo, redistribuzione e riuso dei dati da parte di chiunque, anche per finalità commerciali.

Sempre sul sito www.ponrec.it è disponibile una sezione dedicata alla trasparenza che consente a chiunque di estrarre facilmente informazioni su progetti, beneficiari, località, ambiti e linee di intervento finanziate grazie ad un applicativo che consente diverse modalità di interrogazione dei dati.

Successivamente al lavoro svolto per rendere trasparente l'azione amministrativa e dare evidenza degli investimenti realizzati con i fondi strutturali, l'Autorità di Gestione si è attivata - in questa ultima fase della programmazione - per favorire il riuso dei dati pubblicati, partecipando a diverse occasioni di studio, confronto, analisi al fine di valorizzare il più possibile l'utilizzo dei dati anche attraverso il coinvolgimento dei cittadini.

In questo modo l'Autorità di Gestione ha raggiunto un duplice obiettivo: promuovere la diffusione del PON R&C attraverso azioni di pubblicità e trasparenza, acquisendo contestualmente maggiore consapevolezza sulla necessità di utilizzare gli Open data come strumento di *policy making*.

In particolare, nel corso dell'ultimo anno il PON R&C:

- ha partecipato all'Open data Day - del 22 febbraio 2014 - con un contributo dell'Autorità di Gestione su "**Open data e Open Gov**", organizzato presso l'Università di Napoli Federico II;

- ha partecipato all'incontro annuale della Comunità italiana sugli OPEN DATA "**Spaghetti Open Data**" che si è tenuto a Bologna dal 28 al 30 marzo. Un'occasione in cui presentare esperienze, conoscere nuovi approcci per descrivere, visualizzare dati e favorire la partecipazione della società civile (c.d. *civil hacking*);
- ha collaborato, fornendo informazioni e creando occasioni di confronto, **con il progetto [Monithon](#)**, progetto di Monitoraggio civico nato per verificare dal basso (da gruppi spontanei di cittadini) la corretta attuazione dei progetti finanziati dall'Unione europea in Italia partendo dal Dataset OpenCoesione. Attraverso questa iniziativa fino ad oggi sono stati realizzati 49 report di monitoraggio di cui 5 nell'ambito del PON R&C (quattro progetti di Innovazione sociale e un progetto di potenziamento di laboratorio Pubblico-Privato).

Inoltre, il PON R&C è stato invitato a partecipare a due interessanti iniziative nelle quali elaborare il proprio dataset:

1. Il Secondo **Masters of Network (MON)**, promosso nell'ambito del progetto europeo [Insite Project](#), **il 9 e il 10 aprile a Roma.**

MON è un workshop che riunisce policy maker, personalità accademiche ed esperti internazionali di Social Network Analysis (SNA) – tecnica matematica che permette di studiare fenomeni che possono essere rappresentati con la teoria delle reti – con il compito di indagare da un nuovo punto di vista alcuni problemi specifici di *public policy*.

Questa tecnica che è stata impiegata per approfondire fenomeni sociali come le epidemie, l'obesità può essere applicata anche allo studio delle relazioni che si stabiliscono nell'attuazione di programmi come il PON che è stato oggetto dell'analisi in questa due giorni.

L'esperienza di lavoro è stata quella di un vero e proprio laboratorio, che ha coinvolto studiosi di SNA, esperti di statistica, sviluppatori che grazie a software e competenze hanno elaborato – seguendo le indicazioni del personale MIUR – i dati PON R&C.

Lo studio condotto nel corso dei due giorni è stato finalizzato all'analisi dei dati con particolare riferimento alle relazioni che si stabiliscono tra progetti e soggetti nella assegnazione dei fondi del programma PON (v. ***Allegato 1***). L'analisi ha riguardato gli oltre 2000 progetti e i soggetti - Università, Enti di Ricerca, Amministrazioni, imprese ed individui - finanziati dal programma.

L'**ipotesi iniziale** che ha guidato il lavoro è la seguente: "**Ci sono Reti di soggetti che hanno avuto maggiore successo nell'ottenere finanziamenti da Fondi PON R&C?**". La conferma di questa ipotesi-guida sottendeva un'analisi delle evoluzioni e delle caratteristiche delle Reti costituite.

Una **seconda ipotesi** di studio si è concentrata su una porzione di dati - quelli relativi ai progetti presentati per rispondere al Bando sulla Ricerca industriale - con lo scopo di studiare le **relazioni** che si sono create tra **soggetti ammessi ai finanziamenti** e **i consulenti** che li hanno coadiuvati nella presentazione delle domande.

Il risultato del lavoro ha evidenziato la presenza di **13 network predominanti** che si costituiscono a livello di Programma. Ad ogni modo, considerata la complessità e l'elevata densità delle reti emerse sarà necessario approfondire in modo empirico le relazioni che effettivamente sono presenti – a livello di settore, ambito o territorio di appartenenza – tra i soggetti che costituiscono le predette reti al fine di validare i risultati emersi dall'analisi algoritmica.

Il Report completo del lavoro di analisi è disponibile in allegato (oltre che online al sito <http://www.insiteproject.org/>).

2. Un hackathon di DataJournalism che si è tenuto nell'ambito dell'International Journalism Festival a Perugia il 2 e il 3 maggio.

L'hackathon organizzato dalla società Dataninja, in collaborazione con Datamediahub e il quotidiano La Stampa, prevedeva diverse track, tra cui: **“Ricerca e sviluppo: a chi vanno i fondi strutturali del PON-REC?”**

Nel corso della giornata del 2 maggio è stato svolto il lavoro di analisi, confronto e di visualizzazione dei dati che ha portato all'elaborazione di un articolo giornalistico, presentato pubblicamente il giorno successivo.

Il gruppo che ha lavorato sui dati PON R&C, coordinato dalla giornalista Giulia Annovi, ha prodotto un articolo di taglio divulgativo sul Programma soffermandosi in particolare sull'analisi dei tempi che ha caratterizzato il Programma: dal ritardo nell'avvio alla necessità di procedere in fretta per non perdere risorse.

L'articolo completo, pubblicato anche sul sito di Dataninja, è disponibile in allegato (**Allegato 2**).

Allegato 1 – “Where does Research funding go?”

The context

The context is the National Operative Programme for Research and Competitiveness (PON R&C) which is funded by the European Union for Regional development. This program provides, in the period spanning 2007-2013, almost 4.5 billions euro for research and innovation in the so called “convergence” regions of Italy (Calabria, Campania, Puglia and Sicily) in an attempt to bring these regions closer to the European average in terms of economic development.

The PON R&C has 3 priorities and 12 different interventions - competitive based calls for proposals - some managed by the Italian Ministry of Education, University and Research and others by Italian Ministry of Economic Development, with more than 2.200 projects financed. The subjects eligible for fundings, depending on the type of intervention, are Universities, Research bodies, Small, Medium and Large enterprises, Individuals.

The Managing Authority of PON R&C, through the [Open Data section](#), intended to publish all information on financial management, controls and results of the program, ensuring transparency and data sharing.

The data can be used for any purpose, personal or commercial, and without limitation, in order to better understand the use of the European Funds for projects on scientific research, competitiveness and industrial innovation in the Convergence Regions.

The reuse of these data can be undertaken according to Creative Commons attribution 3.0 (CC-BY), selecting different formats for download (xls, pdf, csv) and surf by filter such as: Region, Sector, intervention description, beneficiary type, amounts, localization, etc.

The goal

The group has explored the network shape and properties of the firms and institutions involved in the national program supporting industrial research and development PON R&C. This seven-year program currently in its last phases of implementation has recently [published the entire database](#) of its projects in Open data format.

In this case the research questions identified were the following:

(a) do certain networks of subjects tend to win the majority of PONREC grants at Program level (PON R&C) and **(b)** at Regional level ?

Then **(c)** what are the linkages between the beneficiaries with third party suppliers (consultants) within a project of the intervention ?

The team

A diverse mix of policy analysts and network scientists included staff of the [Italian Ministry of University and Research](#), and of the Evaluation Unit at the Department of Development policies, staff from the [European Center for Living Technology](#) (Venice, IT), network scientists from [INRIA](#), [Université de Bordeaux](#) (FR) and University of Bologna (IT). The team has worked on this task in the course of the two days.

The experience has been challenging, but in the end was able to provide some first hypotheses and insights into possible future avenues for policy-relevant analysis.

Why call on Social Network Analysis (SNA)?

Standard data analysis usually is able to reveal correlations or patterns between observed variables. Sometimes, however, testing variables for correlations fails at revealing higher level patterns that take place in the data. This is precisely what is expected here from SNA: identify influential actors based on different aspects involved in research programs and funding schemes. Answers to this type of questions can certainly support policy makers in decision-making processes and activity tracking.

The data

The entire database, that consists in more than 5460 records that represent more than 2200 Projects, is divided in three different datasets, which are updated every 2 months:

1. Plans and Beneficiaries
2. Amounts
3. Locations.

In order to ensure a proper interpretation of the data it is necessary to specify that each project is identified by a Project code. Each subject which refer to a project has assigned a “CUP” and a “Local Project Code”. The Local Code Project and the CUP represent two ways to uniquely identify an intervention of public co-financing. (For example: If a beneficiary submit a research and a training project within the same call for proposal, it will receive two CUP codes and two Local Project Codes)

The preliminary difficulties that have been confronted by the group in the first hours of the meeting included the establishment of a minimal common dictionary between the two sub-groups of the policy and the network analysts, the study of the database aimed at excluding the irrelevant information and

controlling possible errors. These complex problems have been dealt within the context of the statement and discussion of possible research questions of interest of the policy debate.

The sense-making loop

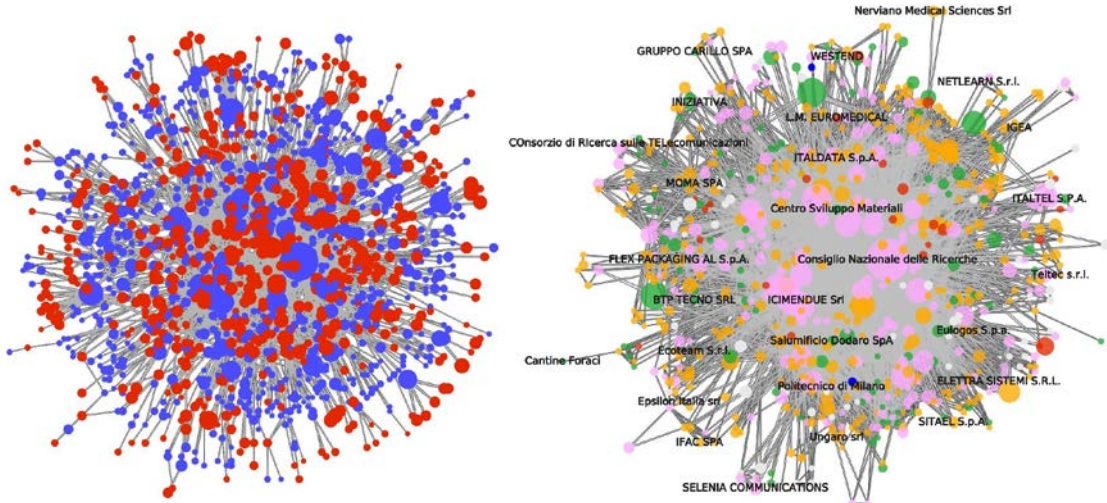
In fact this process is not linear as one would like to present it: the formulation of a research question to guide the analysis in itself represents a gradual and iterative process of refinement of the definition of an area of investigation that has the property of being commonly understood, relevant, and pursuable with the tools at hand. In the case of Group 2 the research question identified gravitated around the concept of persistence of coalitions of firms and research institutions formed in response to the competitive calls. The R&C program is in fact a repeated game in which public funds are awarded competitively to firms and research institutions in subsequent rounds following a set of rules which evolve over time, but that usually require firms to apply together through public-private coalitions.

The questions and tasks

Policy makers would like to know many things about these coalitions: are they stable over time? How do they evolve? Are they formed only in response to the monetary incentive, or do they pre-exist the opportunity offered? Do they survive after the policy measure stops being active, or if the group is not awarded the funds?

After discussing these intertwined policy issues, the analysis has proceeded in two directions.

The **first line of inquiry** into the data guided by the Bordeaux team has used the broader database of projects funded by the PON R&C through the different rounds of competitive calls. With the objective of excluding projects irrelevant for the question defined, the analysis has required the filtering of the actors (nodes) involved in collaborative research projects, which graphically appeared to belong to a single biggest sub-network. This network component appeared to be separate from many small isolated nodes, which have been excluded in that they represented firms funded individually.



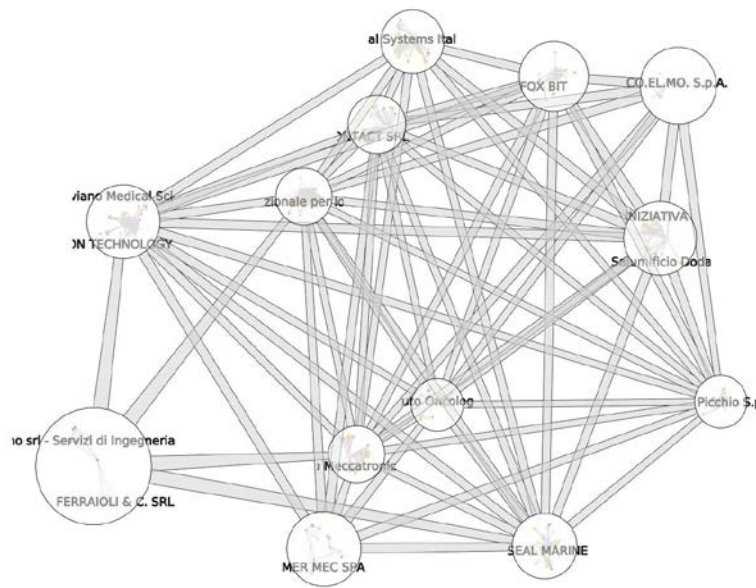
Two-mode versus one-mode

The data that was provided actually only *indirectly* link firms and research institutions (left image: firms and institutions are in blue; projects are red). Indeed, it is because they jointly participate to a project that they do interact. As a consequence, it was necessary to compute a different network directly linking firms and institutions based on their co-participation to projects, turning the original two-mode network into this one-mode network only involving actors.

The resulting network was, as expected, quite dense. Indeed partners of a project form a maximally linked subnetwork since they are all pairwise connected, by construction (right image: nodes are colored according to regions; node size is mapped on total funds for a firm or institution). Various techniques can be used to filter out some of the links and preserve the backbone structure of the network. However, the case we were faced with resisted all filtering techniques we tried. The network is, by nature, tightly connected.

Inferring higher-level structure

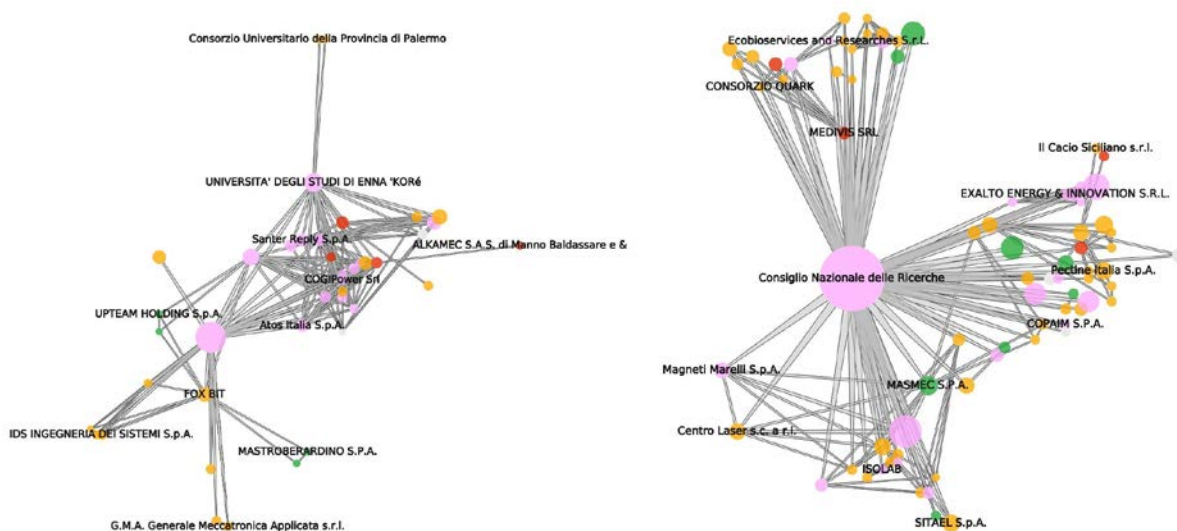
This was the network subjected to the analysis, that a procedure the [Louvain algorithm](#) was able to further partition into 13 sub-networks corresponding to agglomerations of firms and research institutions entertaining dense relations with each others, and weaker with the outside, in the course of the program implementation.



It should be noted that on the above graph, the size of circle for subgroups does not mean anything, also not the size of links between them.

Interpreting the results

Large part of the time in the second day of work was devoted to understanding better the properties and the characteristics of these sub-networks. Are they defined by territorial proximity, by sectoral affinity, by some other common characteristic, or by a mix of those? This exploration was done first by examining the identity of the main players within those groups and their placement within the network. In response to a curiosity which has been raised regarding the amount of program resources that each player commanded and its correlation with location within each group, the graphic representation was modified by increasing the size of the sphere representing each actor in proportion to the amount of money they received through the program funding. Other visual tricks were attempted to emphasize the industrial sector prevalence of each subgroup.



(Detailed image of subnetworks identified through the application of the Louvain algorithm. Node size maps to total funds for a firm/institution. Node color maps to regions (blue - Calabria; green - Campania; orange - Puglia; pink - other; gray - unknown).

These first attempts did not support the conclusion that the persistent coalitions are industry or territorial based. However, quality of the data was not sufficient to answer these questions conclusively in the time available.

In general, the first graphs provide a preliminary snapshot of sub-networks which could possibly correspond to persistent aggregates whose internal relations may be a by-product of the program intervention. To the extent that these images correspond to real world informal networks of relations and trust, these social resources could be employed in different, non policy-induced contexts, to the benefit of research capacity and system competitiveness.

The second attempt was guided by Matteo Fortini who has used the dataset of firms and research institutions awarded funds per effect of one single most relevant intervention of the program ([Industrial research](#)). The analysis has been able to highlight different roles of the program beneficiaries: the recipients of the funds and their suppliers of services, indirectly benefiting from the project funds. The dataset used for this analysis has not yet published on the opendata section of www.ponrec.it.

The data was converted to a directed graph in which the nodes were the proponents and the third party suppliers, and the edges represented the relation “subject X was a third-party supplier for subject Y in project Z”.

The raw data was filtered to keep only projects which were eventually approved in the tender. The resulting graph had 2281 nodes - 771 of which were beneficiaries, 1248 third-party suppliers and 131 were

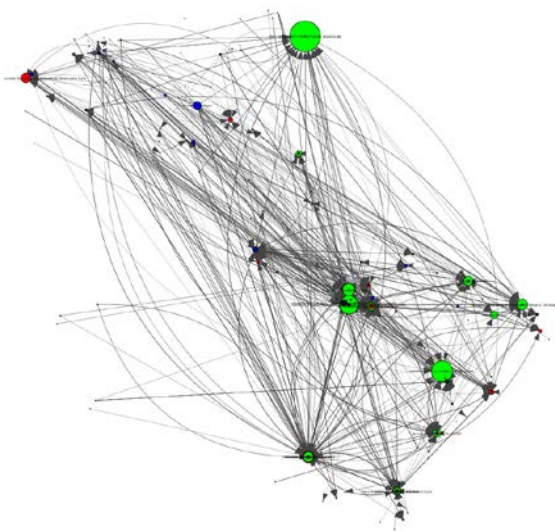
both - and 782 edges, representing collaborations on 142 distinct projects.

On average, on this intervention, there were 3 third-party suppliers every 2 proponents, and each project involved 5 to 6 collaborations.

Degree distribution

The degree distribution of the nodes is close to that of a scale-free network, suggesting that interactions between proponents and 3rd-party suppliers mimic a “small world” in which there are a few very connected nodes (hubs) and many loosely connected ones.

Below a graph of all the interactions, georeferencing proponents (in red), 3rd-parties (in blue) and subjects acting both as proponents and 3rd-parties (in green). The size of a node is proportional to its degree (sum of in+out degree).

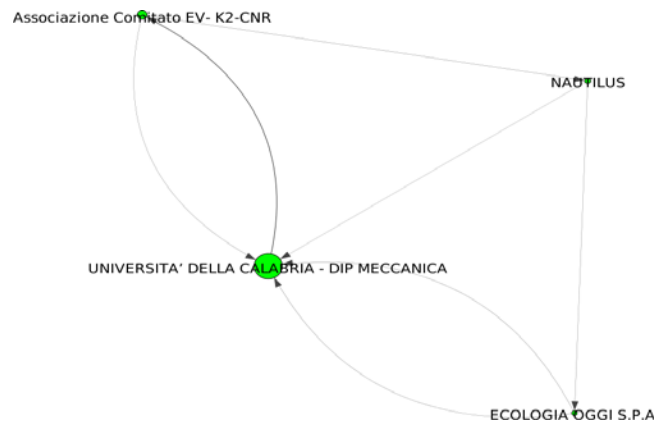


Communities

A first analysis was done on simple connectivity. There is one giant component with 421 nodes and 606 edges which represents the main structure of the interactions and 1696 other tiny components with less than 9 nodes, which are small groups of companies participating in one or two projects.

We also applied both the infomap multilevel community detection and the Louvain algorithm. The Louvain algorithm finds smaller communities, which could help in finding groups of parties which tend to collaborate more closely. An example of one of the communities found by the Louvain algorithm is this:

Below, for example, it is represented a ring with 4 nodes of subjects who submitted 3 different projects.



The data has been polished and represented as a graph. This helped to understand more clearly the interactions between proponents and 3rd-party suppliers. In the future, it would be very interesting to insert into the graph how the total value of the project was split among the different parties, to understand better how the money flows (for now, only the total costs were given): the size of a node could be proportional to the amounts it has managed, rather than the simple number of in/out connections.

Conclusion

These intense two days of collaborative work did convince the group of the potentialities of the SNA. One by-product clearly was to refine the questions policy makers originally had, in light of what the data was able to uncover. As is often the case, much time was devoted to data curation and organization before the analysis and sense-making loop could be triggered.

Members of the group plan to pursue their collaborative work and pursue this line of inquiry. The following steps required from this analysis involve:

- A deeper understanding and sharper definition of the sub-networks identified:
 - are they really 13? would their number and membership change under different specifications of the model?
 - how these subgroups are related to particular sectors or territories?
- An improved description of the territorial mapping of these network entities. The attempt at associating the region of “residence” of the firms returned too many errors to produce any significant outcome
- A better characterization of the predominance of specific industrial sectors within each subgroup.

Other interesting points on which to query the data for future analysis are the following:

- What is the relation between the above sub-networks and the results that they produce?

- Can we see this relation at program, intervention, sector and project level. (ex: To achieve better results, is it more profitable to participate with different kind of subjects? Is there a dimension of the partnership, in terms of numerosity, which allow to reach better results in any given sector?)

One result from this exercise is that it is necessary to polish the actual datasets in order to ensure a better reuse and analysis of the same and above all a greater transparency. It is therefore necessary to improve it with additional data add related for example to the result indicators and the accountability of expenditure and also the publication of all the consultants recruited by the beneficiaries of the projects.

Allegato 2 - Ricerca e sviluppo: a chi vanno i fondi strutturali del PON-REC?

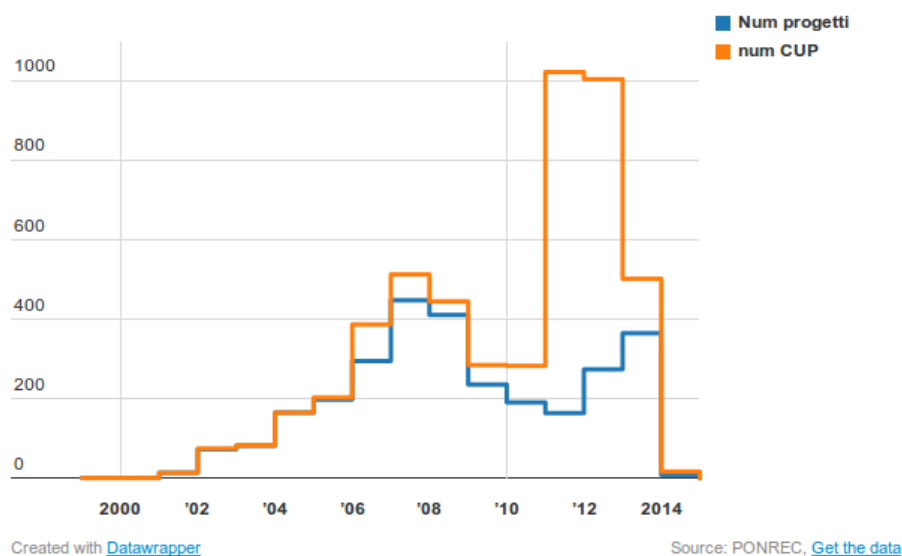
L'obiettivo è il raggiungimento dei livelli medi di sviluppo europeo, e la cifra messa sul piatto è superiore ai 4 miliardi di euro destinati a ricerca e innovazione con il [Programma PONREC](#). A stanziare una tal somma di denaro in favore di Campania, Puglia, Calabria e Sicilia è il FESR, il Fondo Europeo di Sviluppo Regionale. Perché solo queste regioni? Sono state scelte perché hanno un grado di sviluppo inferiore rispetto alla media europea.

I progetti finanziati per questo scopo ad oggi sono più di 3000, e finora sono stati sborsati 2 miliardi di euro. Tanti? No, perché il Programma si è chiuso nel 2013 e i soldi da spendere sono il doppio. I progetti selezionati saranno in grado di assorbire queste risorse? Per rispondere analizziamo lo scenario.

Il numero di progetti finanziati comincia lontano nel tempo: anche se il ciclo di finanziamenti è partito nel 2007, alcuni dei partecipanti avevano già iniziato a lavorare nel 1999. Nel momento in cui si avvia un nuovo ciclo di finanziamenti, il MIUR può scegliere alcuni progetti già avviati purché il loro contenuto sia compatibile con il nuovo percorso. Questi progetti coprono il picco del 2007. I nuovi progetti che si appoggiano a questi fondi sembrano invece partire un po' in ritardo. Nel 2011 erano ancora pochi, ma hanno assorbito un quarto del totale dei finanziamenti (1 miliardo di euro). La ripresa corrisponde all'anno 2012 e il punto più alto si raggiunge nel 2013. Le lentezze dovute all'uscita a scaglioni dei bandi o le complicazioni in fase di approvazione dei progetti, hanno ritardato la messa in opera di questi investimenti. Il tempo stringe: entro maggio 2015 i soldi vanno spesi, pena restituzione alla Commissione europea.

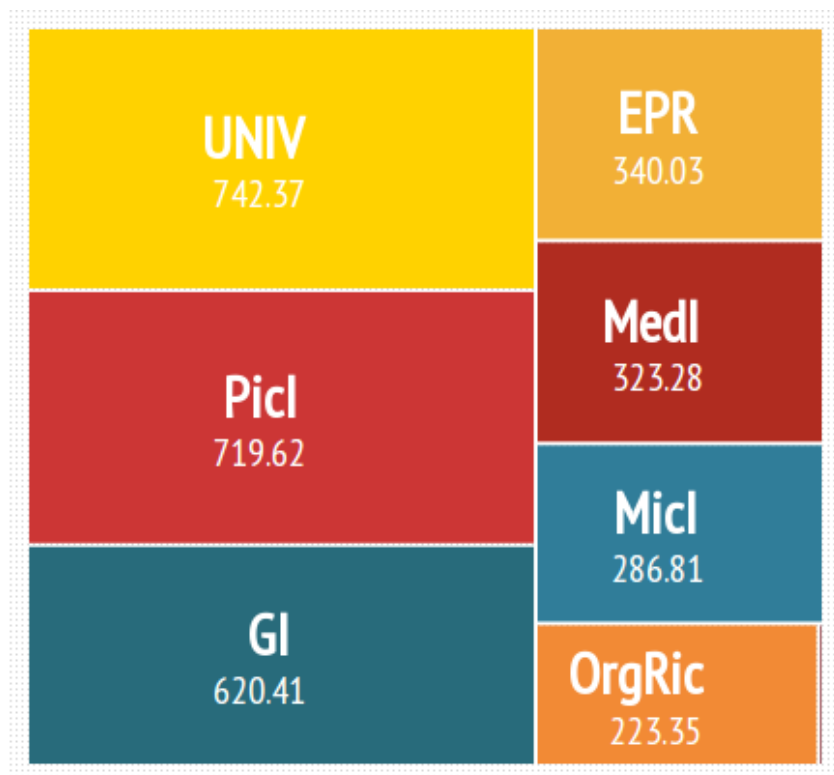
Nel momento in cui i nuovi fondi si sono avviati, c'è stato un punto di svolta: se all'inizio sono stati finanziati molti progetti in mano ad un singolo soggetto, dal 2011 il MIUR ha preferito usare una politica più inclusiva. In ogni progetto sono coinvolti più soggetti: la collaborazione e la formazione di gruppi compositi porta alla realizzazione di progetti di ricerca di maggior impatto. Inoltre i progetti più grandi sono meglio strutturati e vanno a coprire un ventaglio di competenze più ampio, favorendo una minor dispersione di fondi.

Progetti finanziati e soggetti coinvolti (CUP)



Da dove provengono i migliori progetti di ricerca e innovazione del Sud? L'università (23%) e le piccole imprese (22%) sono in testa. Segue la grande impresa nella corsa per beneficiare dei finanziamenti PONREC.

Finanziamenti ottenuti sul PONREC per tipologia di soggetto (in milioni di euro)



In rapporto al numero di interventi, la regione che ha ottenuto più fondi è la Sicilia, a cui seguono Puglia, Calabria e Campania. Ad oggi però, più le regioni hanno ottenuto, meno hanno speso. Il tempo adesso è una variabile importante: se tutti i beneficiari si devono impegnare a spendere al meglio i soldi a loro disposizione, le amministrazioni dal canto loro stanno cercando di eliminare i ritardi burocratici legati all'avanzamento del finanziamento dei progetti. Questi soldi vanno investiti e non restituiti.

Capacità di spesa media per Regione

Importi medi impegnati (disponibili) vs pagati (richiesti ed erogati) per intervento.

■ Importo impegnato medio per sede (migliaia di €) ■ Importo pagato PON medio per sede (migliaia di €)

