Characterising dynamics of new sciences through project collaborations: a project-based scientometrica insight into French bioenergies research

Thomas Tari, Marc M. Barbier, Philippe Breucker

Characterising Dynamics of New Sciences through Project Collaborations:
a Project-Based Scientometrical Insight into French Bioenergies Research

Thomas Tari[1], Marc Barbier[2], Philippe Breucker[2, 3]

[1] Université Paris-Est, LATTS (UMR 8134), IFRIS
[2] INRA SenS (UR 1326), Université Paris-Est & IFRIS
[3] IFRIS CorTexT, Université Paris-Est

Common postal address: Cité Descartes, 5 bd Descartes Champs-sur-Marne 77455 Marne-la-Vallée, France
Contact: thomas.tari@univ-paris-est.fr or barbier@grignon.inra.fr

3rd ENID (European Network of Indicators Designers) Conference
STI Indicators for Policymaking and Strategic Decisions
Conservatoire National des Arts et Métiers, Paris, France, 3rd – 5th March 2010

Paper presented in the 'special session on advanced bibliometrics'

Introduction

This communication proposes to show and discuss some concrete results in visualising and mapping project collaborations within dynamics of new sciences. Projects collaborations are considered as a relevant matter of enquiries, and their visualisation relies on the construction of data mining and database development and on the mobilisation of software and specific computer programming. This empirical and methodological project has been deployed on a specific area of research, namely 'bioenergies'.

We purposefully chose this type of quasi-domain of science and technology, which is politically and epistemologically driven by a quest for more sustainable energy sources. It reflects particularly well the so-called dynamics in 'mode 2' and the idea of emerging fronts of science and technology in relation to technological promise. Our aim here is to ground the study of project collaborations in specific methodological approach of databases, which are accounting for the existence of supports and grants for research and R&D projects. We claim that the study of this layer of information is underestimated in the literature about research evaluation, which is mainly targeting issues of sciences outcomes measurement and focusing on different types of authorship and citations analysis.

Our reflection is grounded in the growing needs, either for decision makers or researchers of the STS and SPS communities, to build the analysis of facts on a convenient visualisation of structural relationships in-between collaborators of projects. The configuration of projects collaborations in dedicated databases is worthy to focus on, as they might reflect also an unexplored part of the dynamics of research and R&D activities.

Our intuition is that the aims, the perimeter, the contents of funded projects represent a relevant account of the under-going technological and scientific dynamics on the one hand, and a relevant account of the mobilisation and choices of scientific communities and science policy "makers" on the other hand. Those configurations rely firmly on spatial-based organisations, mixing European, national and regional scales in formal and informal clusters. Our perspective is thus to enrich the studies of sciences dynamics with the mobilisation of customised databases of research and R&D projects in order to study performative associations of laboratories, scientific teams or R&D firms.

In a first section, we propose some preliminary thoughts on the status of project collaborations to study research dynamics. Reviewing the literature on project collaborations we tried to position the interest of this matter of enquiry for the understanding of sciences and technology dynamics.

In a second section, we shall present quickly our empirical field of study – bioenergies – and present our data sets and the methodology used to build up a database for our research perspective. Therefore, we also present the software needed for the characterisation and visualisation of research collaborations through projects.

Then in a third section we propose our results in two steps. Firstly we deliver a structural characterisation of collaborations in successive research programmes with a classical approach of co-word analysis completed with indications of centrality of main laboratories. Taking into account the politics of research backgrounds in France during the last decade we analyse the dynamics of collaboration as being situated in regional clusters but also muddled by national science policy.

In order to go beyond this type of quite structural interpretation, the second step relies on the hypothesis that project collaborations could be considered independently from the funding programme, meaning that collaborations could be driven by laboratories strategies and interests in research dynamics that would precede or structure research programmes. Our methodology enables to visualise the dynamics of collaborations and we describe complex patterns of retraction, agglutination or emergence of clusters. We refine this visual understanding with a lexical analysis of project abstracts.

In conclusion we discuss the interest of our results for an ecology of research projects. We also propose to go further paying attention to the potential effects of collaboration dynamics: on collaborations themselves, on knowledge production and on technological development.

1. Theoretical backgrounds

1.1. Characterising science dynamics through project collaborations

Without ignoring the existence of a large array of scientific perspectives and information science about the measurement of science productions and science dynamics, we situate our empirical work in the branch of analysis and visualisation of social networks. This field as well as "science of indicators" are central for the evaluation and policy of science (Callon & Law 1986; Law & Callon 1988). At present, the evolution of the analysis of scientific dynamics is largely attached to the question of characterising collaborative and cognitive dynamics of knowledge production (Powell et al. 2005) and to the emergence of multi or trans-disciplinary emerging fields of research (Lucio-Arias & Leydesdorff 2007) or paradigmatic field of research (Chavalarias & Cointet 2008). Nanotechnology e.g. is an important field of enquiry for the study of new sciences and knowledge dynamics (Bozeman et al. 2007). Tracing and mapping knowledge in scientific database or in other electronic sources is representing a huge field of problems for many disciplines dealing with information. More locally, in relation to specific areas of research, mapping heterogeneous networks appears to help the understanding of social dynamic of research activities (Cambrosio et al. 2004; Cambrosio et al. 2006; Bourret et al. 2006; Bonaccorsi 2008).

Whether networks of University-Government-Industry relations (Etzkowitz 2003) constitute an infrastructure or the results of collaboration is still a research question, in order to understand the characterisation of sciences dynamics in relation to innovation processes. There are many institutional facts - but also empirical evidences in many case study researches - showing that innovation processes are situated here in-between. Some empirical works and theoretical considerations within the mode1/mode2 model suggest that multi-disciplinary, multi-institutional collaborations are means in the production of scientific knowledge in society (Gibbons et al. 1994). But this perspective has been challenged (Hessels & van Lente 2008). Bonaccorsi (2008) is for instance quite critical towards the mode 2 description. He rather suggests that new sciences "*are reductionist sciences that address new complex phenomena by breaking the boundary between natural and artificial*" (ibid. p. 296) and that there is "*consistent dynamic properties of the search process that transcend specific disciplines and that can be identified and to a certain extent, operationalised*". He grounds his empirical work in the parsing of new words and their statistical properties within scientific database, one would suggest that project collaborations database represents an other relevant source of information.

Research and R&D projects are seen as episodic financed facilities that support the search and innovation process. Establishing the existence of a kind of "infrastructure" by the means of project collaborations sounds like an invitation to characterise them as a specific layer of information. As already mentioned above, this is also accompanied by the fact that information concerning research programme and science policy is more and more accessible in structured databases.

All STS and SPS researchers rely for their own proof regime on the existence and the reliability of the accountability of scientific productions. Recently, Zitt and Bassecoulard (2008) have indicated three challenges for scientometrics and for the relations between those who study science and technology production and those who create knowledge and techniques: the quality of accountability of knowledge dynamics within various source of data, the necessity to characterise dynamics as well as evaluating positions and the common problem of diversity. Our understanding of these challenges – and of their consequent methodological and ethical requirements – is that there is a need to encourage characterisation of knowledge and technological dynamics for a common understanding, and to do it beyond the necessity of evaluating the performance of normal science and technological creativity. This trend is certainly reinforced by an on-going critical self-evaluation of the reliability of science indicators (Barré 2001; Freeman & Soete 2009). In this trend, characterising and mapping collaboration appears to become a compulsory instrument both for scholars and decision makers, but also for researchers (Heimeriks et al. 2003).

Thus, the nature of collaborative relationships is to be questioned in accordance to the reality of practices of "making science and technology". We would like to raise the contention that the production of knowledge in society has issued a cognitive bias about the understanding of science and technological regime. As a consequence - or perhaps due to - the focus on co-authorship and citation measurement has blurred our understanding of collaborations in science and technology, since publication of articles and patents is certainly an indicator of knowledge dynamic but not necessarily *the only* dimension. Our communication is pulsed by this claim and we would like then to focus on collaborations at the level of research project, not considering those collaborations in a pure communication theory perspective but as collaborations existing within a social world of science and technology.

1.2. Project collaborations: elements for a preliminary review

The importance of the article of Katz and Martin (1997) for our perspective is obvious and the research agenda proposed by Beaver (2001) has also fostered our interest to consider scientific and technological collaborations as a central matter of enquiry. Katz and Martin explicitly grounded their analysis of empirical findings in scientometrics of scientific publications, though their review and definition of motives for collaboration has established the reasons to collaborate in the multiple dimensions of "being a productive researcher". Their approach made explicit the different levels of collaborations (inter and intra organisational forms from individuals to nation), but they rather stay within the stream of measuring collaboration through authorships relation (see ibid, pp 6-14) and suggest keeping very sceptical towards the cost and advantage of a research policy based on collaborations.

Beaver's empirical work on motives of collaboration (Beaver, 2001: 372-375) is based on an account of the past evolving conditions of collaborations in the face of "teamwork" research policy. He acknowledged a series of motives (ibid; p.373) that clearly do not put the publication of articles as unique and central. The heterogeneity of those motives particularly fits well with the conditions in which the practices of Internet and the self-accountability of research works through web pages are representing a growing enhancer of collaborations, far over institutional or geographical proximity. The multiple sources of scientific knowledge accounts (articles, patents, grants database, funding programme) certainly represent a growing field of enquiry for informetrics (Bar-Ilan 2008) but could also issue in a restrictive approach of science dynamics through the measurement of scientific publications.

Within this recent evolution, Bozeman and Rogers (2002) issued a very contending claim towards a culture of counting outputs, when they have called that research evaluation should take into account the social configurations through which new scientific and technical knowledge is produced. Following their anchorage in a tradition focusing on laboratory life or laboratory profile (Latour 1983; Larédo & Mustar 2000) and R&D programme they have claimed that "*many scientists do no conceptualize their work in terms of funding source or the project account and, instead, view project as chiefly a bureaucratic artifice*", ibid, p.771). Their conceptual model (the churn model) reflects their strategic approach of interactions between researchers[1] and thus they distinguish between "*projects that exists in the mind of bureaucrats*" and "*knowledge communities that exist as human interactions with information*", ibid, p.392. At the stage of development of our work on projects collaborations, we do not intend to challenge this model but rather suggest that - even though such a divide between the realm of bureaucrats and Knowledge Value Collective would exist - the strategic use of project collaborations becomes a matter of enquiry. This may particularly be the case if one follows the fact that R&D or Research programme policy is particularly open to researchers and scientific experts (because of ex-ante evaluation with peer-reviewing of proposals) or to lobbying pressure of visible or invisible colleges. All this is more or less related to science dynamics and to researchers at work, not only to bureaucrats, and very often to boundary organisations (Guston 2001). Thus, in order to consider this dimension of boundary work at the frontier of funding agency or national programme committees, it seems relevant to characterise project collaborations as a heuristic.

2. Mapping project collaborations: data demining and methods

The 1973 oil crisis put forward the coupling bio-energy as a solution to the end of an energy paradigm relying on fossil fuels, or at least as a way to improve national autonomy, therefore including substantial economic and geopolitical consequences. Brazil launched in 1975 the Proalcool program, which consists in the large-scale transformation from sugar cane to ethanol for individual cars, while the United States was building pilot plants for corn- and maize-based ethanol production. If ethanol for fuel was known and used since the end of the 19th century, for example in the Ford Model T until the prohibition, the term and therefore the object "biofuel" appears as an energy production medium, in a liquid form, from diverse biomass in the 1970's. The oil prices fall-down in the mid-1980s froze all research programs in the United States, but the "peak oil crisis" led to envisage a profitable biofuel production.

Why focus on R&D programmes and projects? Bioenergies as a new science deals with a lot of issues. If the quest for energy independence always constituted the chief goal of the

[1] This strategic approach is also present in Lee & Bozeman (2005).

ethanol development in Americas, the European promotion of biodiesel mainly depended on agricultural and environmental contexts. The 1992 MacSharry reforms of the Common Agricultural Policy (CAP) incited the production of non food crops on fallows. Since then, agricultural lobbies and producers are fervent supporters of bioenergies. At the turn of the century, climate change and sustainability challenges led the European Union to seeing in biofuels an ideal technological fix, quickly applicable, which could compensate greenhouse gases emissions, thanks to the absorption of carbon dioxide from the air by biomass during the growth of dedicated plants. The incentive EU Directive 2003/30/EC “On the Promotion of the Use of Biofuels or other Renewable Fuels for Transport” proposed a reference target of 5.75 % biofuel blend by 2010; the “EU Strategy for Biofuels” from the European Commission brought this purpose to 10% by 2020. Until 2006, national plans abounded, and a eulogistic media buzz grew up in a regime of techno-economic promise.

However bioenergies also had to face some serious controversies as since 2007 biofuels have been accused to starve the developing world. This global dilemma “Food vs. Fuel” was the main one and highlighted the land use change to non food crops in detriment of the food supply. A large number of sub-controversies erupted and the following list is not exhaustive. Life Cycle Analysis on biofuel production are contradicting each other, biomass absorptions of greenhouse gases may not be sufficient to compensate the emissions according to the Nobel Prize Winner Paul Crutzen because of the N_2O (Crutzen et al. 2007), the energetic efficiency and economic profitability are being discussed, deforestation is condemned, the work conditions of sugar cane hand harvesters in Brazil are denounced... Which plants should one use, with genetically modified organisms (GMO) or not? Which production processes have to be supported: local, international, biorefinery? For which products: ethanol, biodiesel, straight vegetable oil that is prohibited in some countries when used as a private fuel, in which blend? Which spectrum of bio-based products (food, feed, materials, chemicals) and energy (fuels, power, heat) should be targeted? Is there any tangible difference between technological “generations” of biofuel? Are petroleum substitutes the right path, instead of moving out of the car civilisation? So many questions we would like to see how researchers and research managers deal with in France. In such a context, did bioenergies emerge as a new science, what dynamics pulsed its development? On a national scale, classical scientometrics studies based on publications are not numerous enough to be relevant (Tari 2009); as bioenergies is highly trans-disciplinary there are not easy to gather even if some dedicated journal were born, *Bioenergy* e.g.. We then decided to focus on research projects and the collaborations within.

Unlike articles, books, conference proceedings or patents, research projects have not yet been conformed to a standard and collected into a global database. Although most funding institutions – especially national agencies – have created their dedicated structures, the latter are partially opened to public access and relate to their subsidies only. The constitution of mutual research project databases has been promoted by the European Research Area NETwork (ERA-NET) scheme, in order to develop networking and opening of national and regional research programmes. The importance of such a construction in scientific work, as research tool, has been a subject of inquiry for science studies (Hine 2006), especially focusing on the “practices of accessibility” (Hine 2005) that enable collaterally the virtual social sciences to study them. Concerning the bioenergies field, we took great advantage of the BioMatNet project base[2] that covers a scope from ECLAIR, AIR, FAIR programmes to the FP6. Created within the FP6 EPOBIO project, it is, alas, not updated. Once again however, as in Cordis, Framework Programmes’ projects are the only one inventoried. In the United States, a trans-institutional structure, Grants.gov^SM[3] provides an overview on all R&D programmes through grants, still does not scale down to the project unit.

[2] http://www.biomatnet.org
[3] http://grants.gov

We question research dynamics through a large scope of sources related to research programs. Within those programs, the R&D project is our unit. The variety of project ecologies and organisational features lead us to conduct systematic data mining, data extraction and to constitute an adequate and robust information structure through the creation of a heterogeneous and relational database. Our methodology, also based on the grounded theory, has a certain similarity to the one described in Hellström et al. (2001), especially dealing with the construction of a relational database thanks to a "snowball effect" mode of data collection. Thus we conducted via the Web a manual survey of more than 200 research projects, or 130 laboratories, 180 institutions including corporations and 260 researchers more or less involved in bioenergy research in France. Project properties, specifications concerning researchers, laboratories and institutions, as well as traditional scientometrical indicators but also geolocation data have been informed as precisely as possible.

We collected material from several French national programmes: the energy related projects in AGRICE (AGRIculture pour la Chimie et l'Energie, 1994-2007), the whole PNRB (Programme National de Recherche sur les Bioénergies, 2005-2007) and BIOENERGIES (2008-). The first one was lead by the funding agency ADEME (Agence de l'Environnement et de la Maîtrise de l'Energie), the third one by the ANR (Agence Nationale de la Recherche), while the second was a mix funded by the ANR and directed by ADEME. We also took into account some regional programmes, mainly oriented through the regional cluster structure of « pôles de compétitivité » and the French participants in European framework programmes through BioMatNet.

First, we completed the programme and project table from funding agencies websites: project's name, acronym, first year, duration, funders, cost, subsidies, abstract, manager, research theme, url… Then, we entered each participant in the researcher table, with his self-declared affiliations, laboratories and institutions, in a data demining perspective (Zitt & Bassecoulard 2008, p.51). Online, we checked the personal, lab and institution pages to fill in diverse fields: id, contact, profession, geolocation and compared them to ISI Web of Knowledge, CAB Abstracts and Thomson Factiva press data for biofuels is a controversial subject[4]. If a researcher belongs to one lab and is paid by one institution, laboratories are often patronised by several institutions. Participation in projects is traced through links in the collaboration table; we wished our data to be accurate enough to join only individuals and projects, but, as represented in figure 1 by dotted arrows, we often had to content ourselves with relationships between laboratories, and in a few cases institutions when the involvement of big corporations is scarcely mentioned.

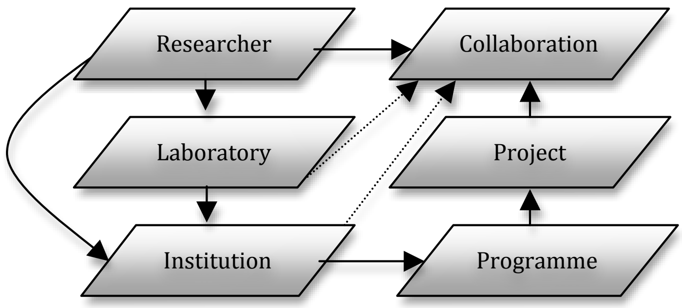


Figure 1 Schematic representation of the heterogeneous biofuel projects' relational database structure

The relations between chosen entities are specified in order to formulate requests based on our research hypothesis. We developed a semi-automatic language translation tool between our database, requests and various existing visualisation software. The first step consists in extracting with traditional relational database requests a list of nodes, edges, their associated

[4] Those who have once dealt with such crossovers would understand how time consuming it is.

weight and other properties later used to create typologies. Our tool can convert this .csv list into .graphml, .xml and .net files, current standards used in network mapping visualisation tools.
The use of specific tools to index database and visualise networks in the context of characterising emerging domain of collaborations within Research and R&D projects represents a step forward to avoid an evaluative perspective with scientometrics. It also enables to open the field of design of methodology and visualising solution to enhance new ways of tacking with relational data. This attitude towards methodological equipments for the visualisation of co-words clusters and mapping shares many ideas of shifting the use of tool from a scientific context to a science policy context (Noyons 2001). We think that this turn should also apply to the technical construction of facts within the scientific stance.

Many tools have been used; we hereby list the only ones providing visualisations in this paper. They are all open source software that have been developed in an academic context; we have interoperated their use within the platform CorTexT[5] of IFRIS:

- Adequate and heterogeneous database construction: MySQL
- Visualisation of social networks: ViSoNe (Brandes 1999; Baur 2008), SoNIA (Bender-deMoll & McFarland 2006; Moody et al. 2005)
- Cartography based on geolocation: AMMAP, OECD Regional Statistics Tool (Jern 2009)
- Semantic analysis based on abstracts: BELUGA (Turenne & Barbier 2004)

Languages for development: FLASH, GRAPHML, NET, PHP, SQL, XML...

3. A Project-based Insight into French Bioenergies Research

When Miettinen (1998) described the "construction of a research object and agenda" in the community exploring cellulose-degrading enzymes (mainly the Finnish one), a main research theme in bioenergies, he centred his analysis on the concept of "application objects" born from the innovation network co-constructed. Quoting Latour (1993), he insisted on the idea that such "an object exists first in the form of a project. If a strong network of actors can be constructed, it turns into an institution". Research project is therefore the first step forward leading to the co-construction of a research object and a scientific community. Moreover, bioenergy, then unnamed, appears to be a pertinent new science to investigate this way.

3.1. Exploring to understand programmes' structure

The below maps (figure 2 and 3) represent the network of laboratories' collaboration through projects, respectively in the national programmes AGRICE and PNRB thanks to the Visone "software for the analysis and visualization of social networks".

The analysis unit is the laboratory. Node size designates the number of projects in which a lab is involved, while the edge width counts project collaborations between two institutions. The node colour depicts a five-classes typology relating to the public or private degree of an institution. Off-white stands for French Public Scientific and Technical Research Establishment (EPST), lighter grey for Industry-Oriented Public Establishment (EPIC) and the medium one (e.g. GIE ARVALIS ONIDOL) for occupational collective structure. The two shades of dark grey and black indicate big corporations and small and medium enterprises.

[5] http://www.cortext.fr

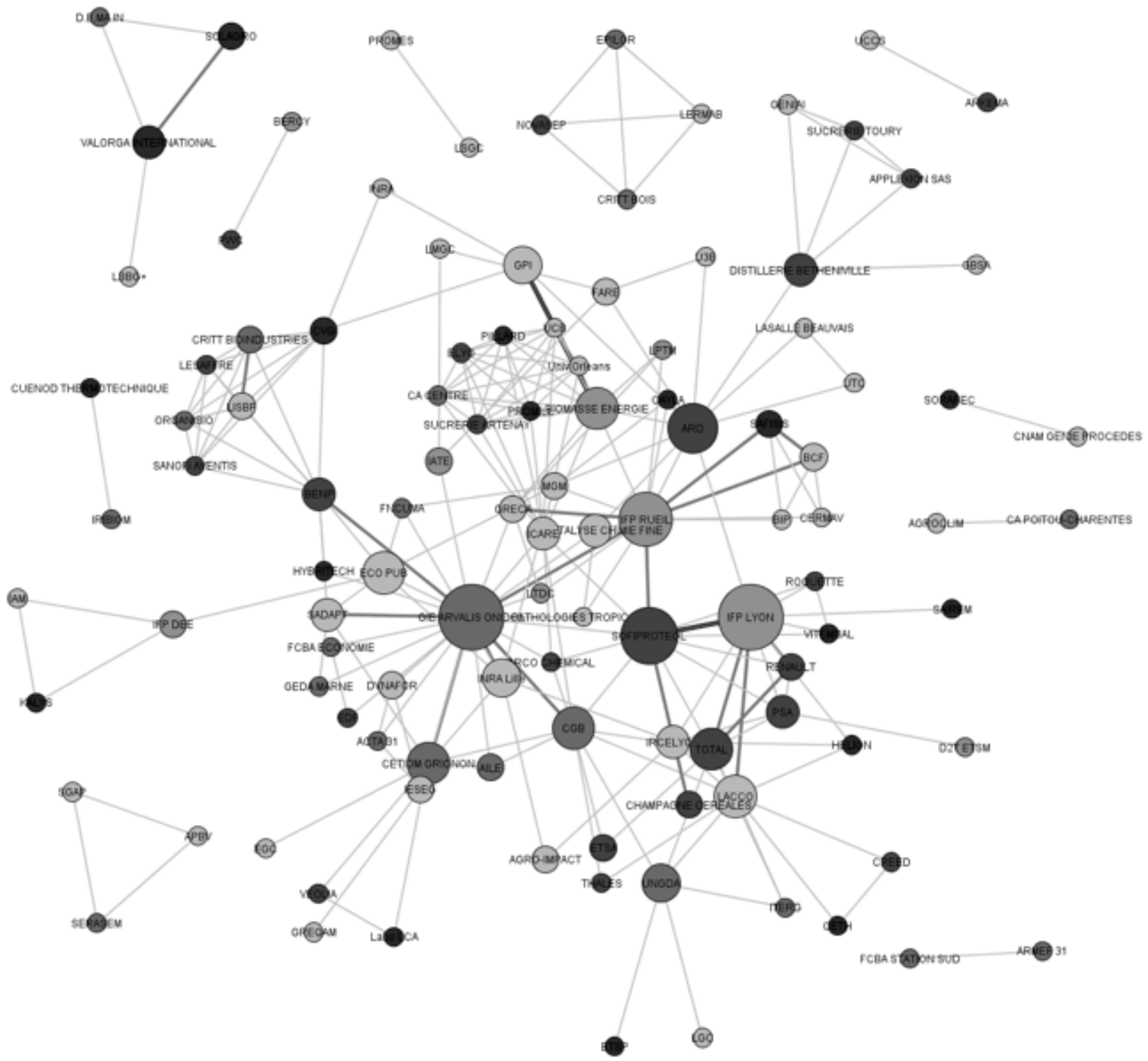


Figure 2 Network of project collaborations between laboratories or institutions within the national programme AGRICE (1994-2007)

AGRICE structure is scattered in several clusters; one major however constituted itself around some agricultural professional organisations like GIE ARVALIS ONIDOL, BENP, CETIOM and SOFIPROTEOL for cereals and oleaginous, CGB for sugar beet: a “first generation” world of ethanol and biodiesel production in close relationship with IFP (the national oil research centre) labs. Close to the latter, the oil company TOTAL FINA ELF, but also car manufacturers collaborated within some projects while the agroindustrial and green chemistry research firm ARD developed a regional cluster (Picardy and Champagne-Ardennes) around it.

We only sketch here a general frame as we are going to study more precisely project collaboration through temporal dynamics. The agricultural world has disappeared as PNRB is built on a grape-like distribution frame, for clusters are linked to a technological and industrial centre through obligatory passage points. The top-western one e.g. is clearly a wood-resource network, which communicates with the bioenergy community through FCBA, a wood technological institute. Industry-oriented public establishments like CEA (LPTM, LITENN, LB3M), IFP, CIRAD (BIOMASSE ENERGIE) stand in a central position: it is typical of a national project management ideology that privileges public-private partnership, in order to guarantee both public good and effectiveness through profitability.

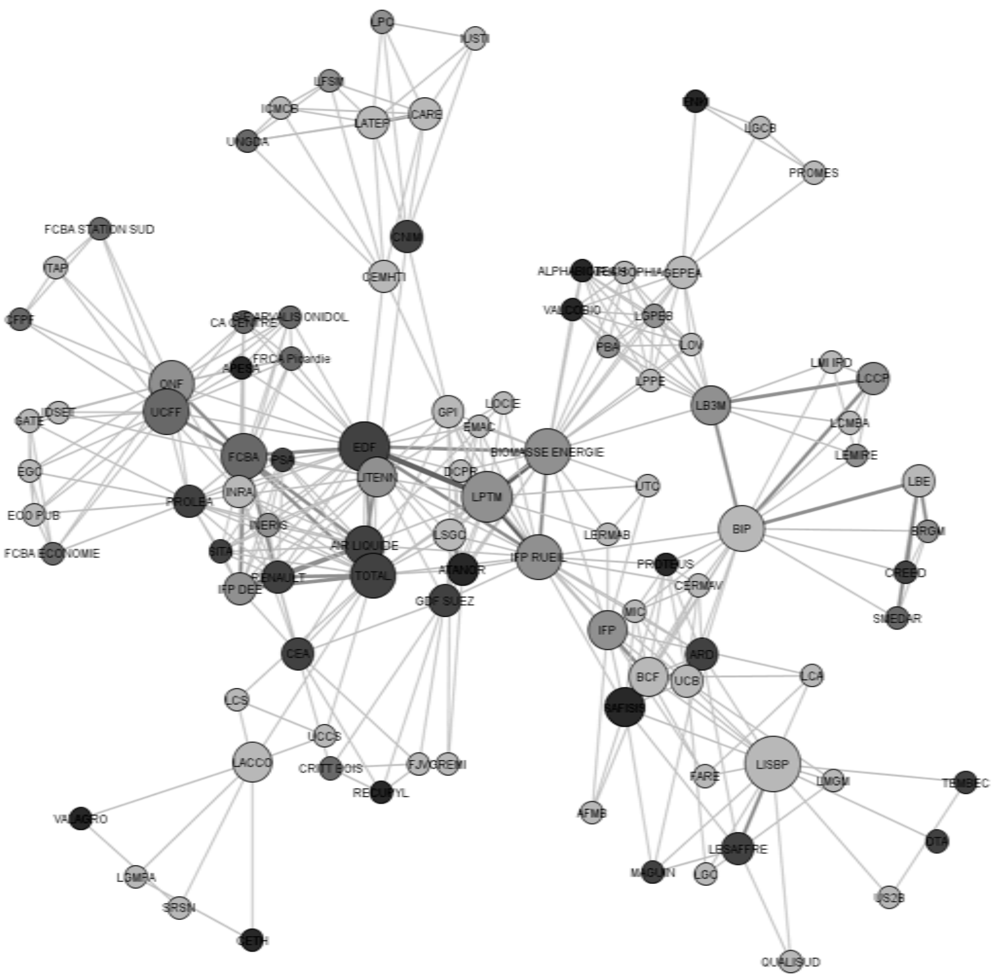
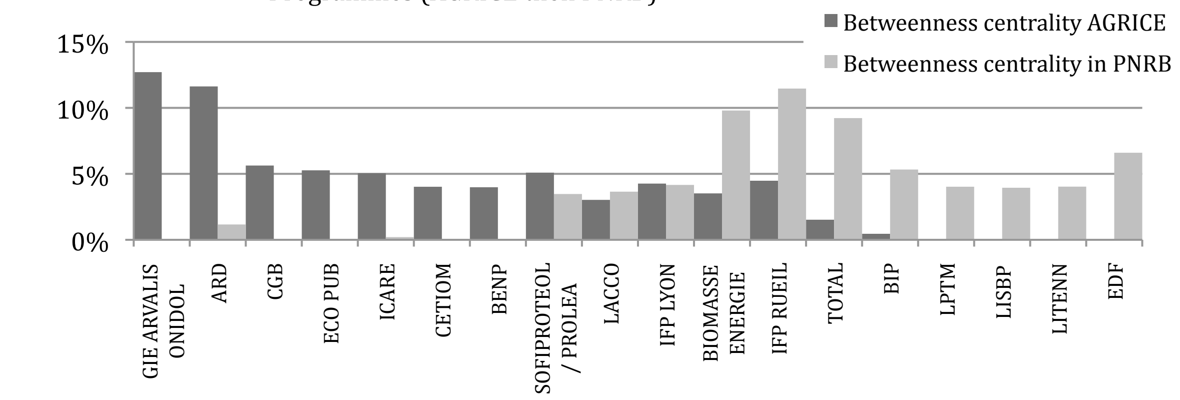


Figure 3 Network of project collaborations between laboratories or institutions within the national programme PNRB (2005-2007)

Analysing betweenness centrality reveals the obligatory passage points best than degree centrality; underneath (figure 4) are to be found the main actors that play a 'broker' role in the previous networks (BC>3,5% in PNRB or AGRICE).

Figure 4 Betweenness Centrality of Project Participants to two almost successive Programmes (AGRICE then PNRB)



Three schemes appear: six out of seven key actors in AGRICE that are not any longer central in PNRB are resource-oriented, persisting lead institutions act in the oil and combustion world (six out of six) while the five emerging in their new inbetween role labs are specialised in bioprocesses (BIP: Bioénergétique et Ingénierie des Protéines and LISBP: Laboratoire d'Ingénierie des systèmes biologiques et des procédés) or thermochemistry and electricity nuclear generation: LPTM and LITENN are Commission for Atomic Energy laboratories, EDF is France main electricity provider.

Project collaborations between laboratories enabled us to sketch inherent logics in programmes' structures, logics that are strong. This is not stunning since the call for a so-named bioenergies research action, focusing on second generation biofuels, was performed into a 2005 programme (PNRB) in the newly born research-funding French leading agency ANR. Other structural logics can be unveiled. In project ecologies, local and regional networks play a major role, but if Grabher (2004) proposes to think knowledge spaces topologically, it is specifically in order to overcome the strong/weak-tie dichotomy that ascripts strong ties and social coherence to the local level and sparse networks to the non-local realm. The following map (figure 5) is an attempt to mingle the scales as it counts the national projects in which at least one lab is involved per region, and at a given time compares the number of still running or newly selected one. We entered our geolocated data in the OECD regional statistics display (http://stats.oecd.org/OECDregionalstatistics) since it provides a both spatial and time dynamic representation.

We also represented the geolocated networks of project collaborations thanks to AMMAP (maps are not shown here) and could interpret the embeddedness of research places and relations in territories. AGRICE network was the archetypal star-network around Paris, although some links do not trace an effective partnership within Ile de France region but denote the numerous headquarters of big corporations in the area. It draws an interesting triangle of regions: Ile de France – Nord Pas de Calais – Champagne-Ardenne enclosing Picardy, which matches the traditional first generation biofuels territory since 1986 and the first R&D platforms construction resulting from a strong commitment of agricultural cooperatives, and which gathers some biotechnology-specialised universities that developed vivid links with local industry (Cassier 1995). This sub-network progressively disappeared from national funding scope, on the one hand thanks to regional programmes development in which local universities have been involved and on the other hand *via* the trans-regional Industries and Agroresources (IAR) cluster creation, a structure dedicated to bioenergies and green chemistry and which became his own projects funder.

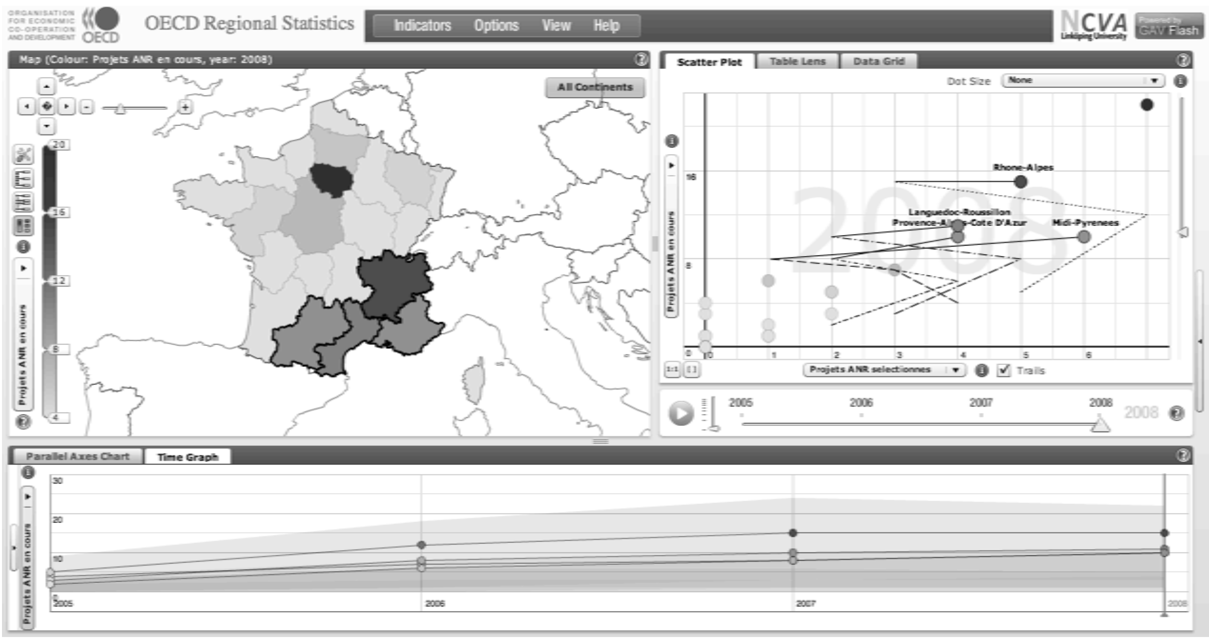


Figure 5 Repartition and evolution of the number of running and newly selected ANR bioenergies projects per region

On the top right window we can see that Southeast regions (selected on the left geographical map) follow a similar pattern over the period 2005-2008, as major new actors in bioenergies research. Paradoxically, this regional trend does not reflect strong regional logics in the field. Like IAR, the new "pôles de compétitivité" clusters prevail and their creation explains the shift. Unlike IAR, those one are not dedicated to the biofuel world; therefore researchers still have to seek support through national calls for projects even if they get a regional cluster label that will favour their acceptance[6]. This label helped us to identify the involved clusters and their associated research themes that are: Mer (sea) and Capenergies (GHG saving energies) in Provence-Alpes-Côtes d'Azur, Tenerrdis (technologies, new and renewable energies) and Axelera (chemistry and environment) in Rhône-Alpes, Agrimip Innovation (agroindustry) in Midi-Pyrénées and Xylofutur (wood) in Aquitaine. The mapping of collaboration delivered the evidence that research localisation follows resources production, as it is planned to produce second-generation biofuels from lignocellulose out of wood and dedicated plants, and third generation ones from microalgae or waste.

It is deceitful to ignore the opacity of such cluster structures as *pôles de compétitivité*, which have been created as a spatial planning policy tool but shifted from an equity logic to an effectiveness one (Duranton et al. 2008) and "*pôles*" became not only competitive but also in competition one with another. Even if some projects are half-publicly-funded, we cannot easily find data. This is when some *demining work* has to be done: how can one measure such opacities?

Programme *per year*	Number of projects involving at least one EC member	Total number of projects selected
PNRB	**10**	**32**
2005	*5*	*10*
2006	*4*	*13*
2007	*1*	*9*
BIOENERGIES	**0**	**13**

Figure 6 Involvement of ANR Evaluation Committee members in selected projects

Thanks to our database requests, we can e.g. easily compare the composition of national programmes' scientific committees and their involvement in selected projects. In the Programme National de Recherche sur les Bioénergies (PNRB), eight out of seventeen members of the Evaluation Committee (EC) have been funded through a research project they evaluated. If there was no public information on AGRICE I and II scientific committees, in 2005, as the new programme has been launched the same year the ANR was created, the EC appeared to be a self-funding structure (cf. figure 6), i.e. an award for the successive co-construction of a 'hype object', the biofuel, and a related research agenda, as described in Miettinen (1998). The flow of collaborations in project exemplifies this coproduction. Progressively, the programme opened to others scientists and communities. How can we depict these research communities?

Static network structures do not allow us to question the cohesiveness of a bioenergies scientific community. Grabher (2004) suggested a project-related typology through social and communicative logics: "*whereas* communality *signifies lasting and intense ties,* sociality *signifies intense and yet ephemeral relations and* connectivity *indicates transient and weak networks*". These stylised features, also analysed with 'substance' (respectively narration or knowledge or information) and 'governance' (trust or swift trust or peer recognition) categories, are primarily defined by time dynamics. We therefore need to rely on the embedded structure of projects collaborations, their evolution, this focusing on to the flows of collaborations.

[6] This analysis is not only extracted from the lecture of mapping but reflects what have been found in the 20 interviews done with the actors.

3.2. The project collaboration as analysis unit of dynamics in new sciences

The following hypothesis has to be considered as a relevant heuristic to tackle with the complexity of motives and conditions of project collaborations: research entities (such as laboratory, research unit, department or even board of research institute or university) are developing strategic collaborations through projects in order to frame a loosely coupled organisational form (with social, cognitive and technical interests) with the view to foster their own development, independently of the policy of programmes.

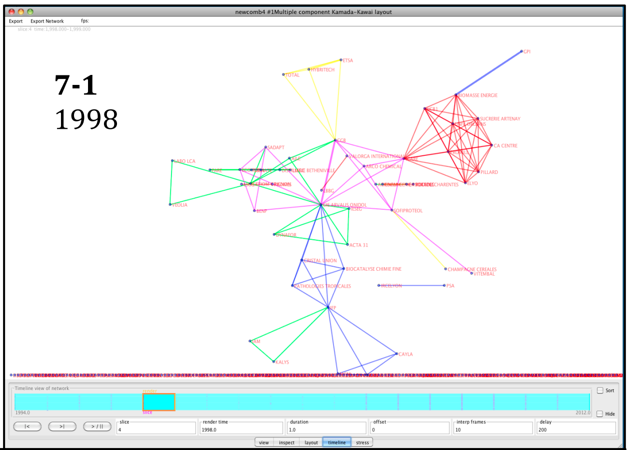
Taking a set of projects as a population of reciprocal collaborations evolving through time offers then a way to look at science dynamics from this point of view. Moreover, when these dynamics correspond to the affirmation of the emergence of a new science or to the existence of technological promises, the characterisation of project collaborations in this context – and possibly their effects on science production – is becoming a good matter of enquiry.

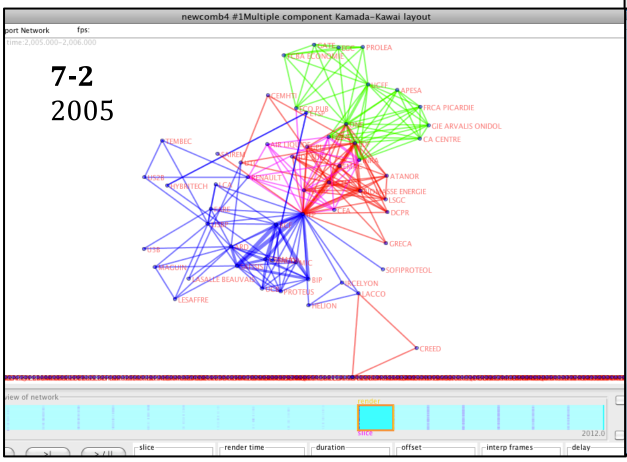
We used SoNIA, the Social Network Image Animator[7] (Bender-deMoll & McFarland 2006; Moody et al. 2005) to visualise our data and study the patterns in project collaborations, all programmes aggregated, through years. Colours stand for thematic axes, according to an ANR typology evolving continuously that we stabilised and adapted for a coherent trans-programmes investigation.

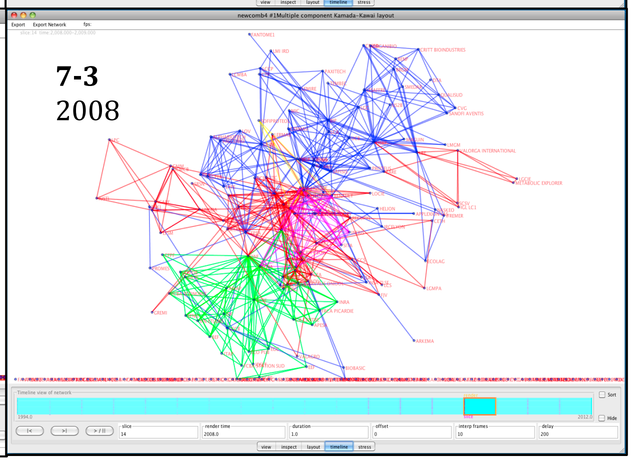
Figure 7 Networks of projects collaborations in the bioenergies field in France in 1998 (right), 2005 and 2008 (below). Visualisation from SoNIA software.

Networks become dynamic in a video that can be seen at:

http://sens.inra-ifris.org/Project-collaborations-dynamics.html







After a start with separated resource-based projects (green) and engine and fuel combustion research (yellow) the AGRICE programme has evolved towards other directions. Firstly a socio-economical and -environmental project (pink) emerged in relation to one of the resource projects, the cereal institute ARVALIS being central. Then in 1996, new comers and

[7] http://www.stanford.edu/group/sonia/index.html

new thematic axes surfaced, with a big cluster of thermochemistry (red) research that showed no link at all with the previous project while new bio-resource projects have aggregated with previous ones. Schematically thermochemistry is one way of producing oil *via* BtL, Biomass to Liquid process through combustion and gasification followed by a liquefaction (Fischer-Tropsch) while another, biochemistry, deals with enzymes that separate fermentable cellulose from lignin, or microalgae that under stress conditions produce oil.

1997 displayed the appearance of ICARE, a combustion research lab that, being at the same time in the pink cluster and in the red one, played thus a role in joining technological research on thermochemistry and an evaluation oriented research gathering all the agroresources: CGB, SOFIPROTEOL, VALORGA Int. etc. In 1998 (see figure 7-1) the dynamic of the first phase of AGRICE is clearly revealed, (cf. previously the betweenness centrality in figure 4): GIE ARVALIS is still major in centrality; but IFP is landing through a set of biochemistry projects (blue). Exogenous sources of information – French participants in framework programmes censed in our database and interviews – allow us to precise that ICARE and ARVALIS are having a key position in the European landscape and possibly are also central aggregates for this reason. Until 2001, new comers do not reshape the landscape and research themes are staying apart; some actors only interfacing; the economical-sociological-environmental theme is creating an evaluative stance of the effect projects and major institutional actors implicated produce. A new wave clearly arose in 2002 around the oil national research institute IFP and biochemistry oriented projects. Our interpretation is that for the 1st generation of biofuels – direct fermentation of sugars or oil pressurisation from traditional agroresources – research had sufficiently achieved good outcomes in terms of industrial research (adaptation of engine and results about combustion), then it seems that the promise emerging in this year is to figure out a 2nd generation biofuel initiative. One notices a real structuring of the biochemistry theme (blue) and thermochemistry one (red) in a kind of butterfly wings feature that is centered on IFP.

In 2005 joined in a whole new bunch of projects; we cannot totally forget that this year is the starting year of the PNRB programme. There is one and only obligatory point of passage: IFP (see figure 7-2). A paradox is emerging: biochemistry, centred on enzymatic hydrolysis in relation to ligno-cellulosis, reveals no link with the resource cluster. It gives the impression that there is no research front that would associate bioresource and biochemistry. This is surprising since resource-oriented researches now only concern the wood sector, which tends to indicate the driven force of the 2nd generation promise. We also note that major companies and industrial actors are staying close to thermochemistry researches when they are involved in the evaluation theme. We think boards of managers or directorates of big organisations are involved in such projects, exactly for this central position, in-between technological clusters. It remains to be studied whether or not this position is transferred to the inner organisation of those big institutions (Total, CEA, Air Liquide…) in terms of transdisciplinary research in the inside. New comers (cf. figure 7-3) through new collaboration networks are then very much concerned with so-called forthcoming 3rd generation projects on microalgae (CEA LB3M is central), waste valorisation (INRA LBE) and still bioprocesses (CNRS BIP, LISBP at INSA Toulouse). This new combination of laboratories not much linked to agro-resource production represents a new potential in favour of researches focused on bioprocess linked to thermochemistry. It also introduces a clear difference compared to the coalition of major fossil oil companies and car industries that has been largely driven, until recently, by a substitution and no innovation strategy.

We do not endorse a generation based analysis as close attention to the field and definitions of these co-called generations vary extremely with actors – a quali-quantitative approach mixed with interviews is essential, but we indentified several search regimes (as defined by Bonaccorsi (2008)) based on diverse actors, centralities and themes. In fact the ‘generation’ speech is here seen as a performative involvement, therefore real, without taking into account the actual development of biofuels and market configuration.

This research axes typology can be refined with a more precise, yet dynamic, discourse analysis focused on research projects abstracts. We purposefully used the BELUGA tool (Turenne & Barbier 2004) that enabled a textual analysis of motives and their evolution through time, cf. figure 8 below.

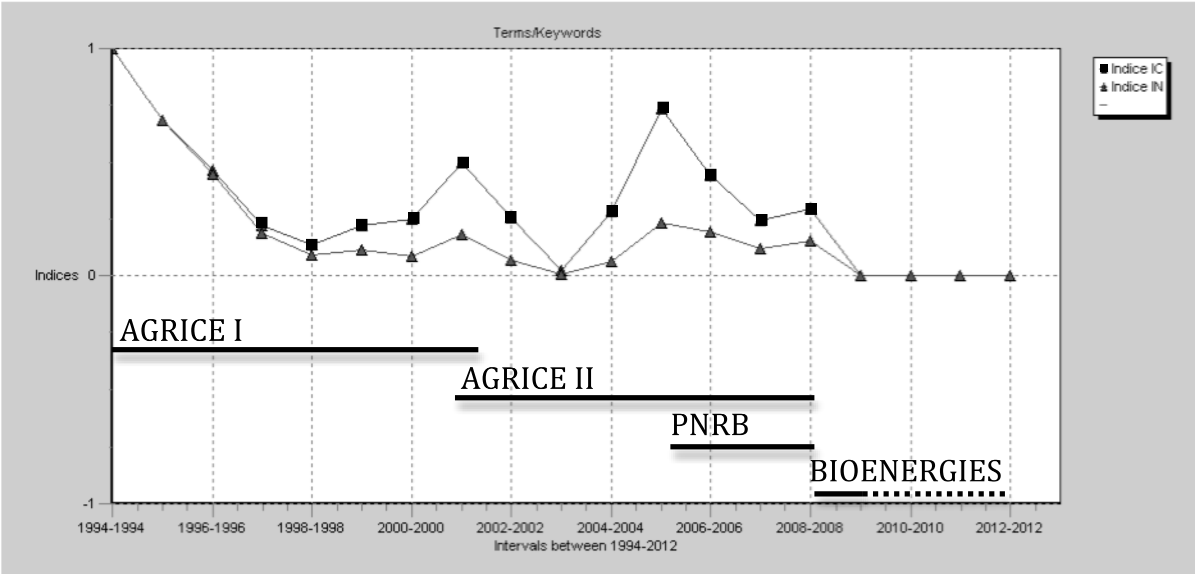


Figure 8 A visualisation of discourse changes in project abstracts through time, BELUGA tool.

BELUGA first performs a terminological extraction with lemmatised forms, then through class extraction algorithms, is able to produce indices of knowledge production dynamics. The square dots stand for the Change Index (IC) and triangle ones for the Novelty Index (IN). If N_t characterises the number of new descriptors between t-1 and t, T_t the global number of descriptors at time t, IC_t and IN_t are then defined such as: $IC_t = \frac{N_t}{T_t} \quad and \quad IN_t = \frac{N_t}{\sum Ni}$.

IC and IN are indicators for the intensity of patterns change, here discourses, IC in regard to the previous period and IN relatively to the whole.

Of course discourse motives are all new in 1994 and it is only from 1997 that we really can analyse the indices trend. We entered projects until 2008 so the period 2008-2012 neither is to be studied as it only indicates still running projects. New research themes or at least new ways of presenting them emerged in 2000-2001, 2004-2005 and 2008, confirming our previous description in terms of social network of collaborations within projects.

Abstracts in R&D projects are interesting media as they are a performative discourse that convinced funding agencies and resulted in research collaborations. It is not surprising that new narrative infrastructures (Deuten & Rip 2000) precede new programmes: the invention of a speech in terms of generations, whether it actually deals with researches and results or not, is a major tool to structure new programmes and therefore new sciences, before thinking of new patents or new journals for dedicated publications. Such a project collaborations analysis is then essential to understand how researchers collectively construct their own science, not only through science policy media; this is a research question we intend to investigate further on.

Conclusion

In this paper we have only addressed the matters of characterising the dynamics of collaborations and we have tried to open a perspective enriching the 'ecology of projects' perspective developed by Grabher (2004). When one enriched the follow up of project collaborations with a preliminary fastidious work of de-mining the information on principal collaborators (PC) dataset, we have shown that many social dynamics become accessible for the characterisation of research dynamics. We have proposed a characterisation of the implication of PC in the programme scientific board. We have also shown with a geolocation tool that project collaborations could be mapped and we have advocated for changes in the localisation of regional clusters in relation to a 3rd generation of bioenergies. We also have tried to push a hypothesis, which might be taken as counter-intuitive: we assess that project collaborations could be explored as if there were independent from the attachment of collaborators to programme for financial resources-based strategies only. This is a heuristic hypothesis that we want to keep in mind for further empirical work.

We have developed the idea that project collaborations can be analysed thanks to the type of ties that are at work according to the duration, the profile and the intensity of those relations. In this communication we have basically paid attention to the visualisation of the structure and the dynamic of those ties, with a quick look of the centrality of laboratories involved in project collaborations. The visualisation of project collaborations enables to shed light on the dynamics of the technological promises of 2nd and then 3rd generation of bioenergies. We have established the localisation and the enforcement of new coalition of principal collaborators linking technological research on bioprocess and on thermochemistry. The structure of ties and their thematic composition shows a potential innovative turn in the coalitions between major companies of fossil oil and car industries, which have been largely driven until recently by a substitution strategy.

Further development will be needed to characterise the communality, the sociality or the connectivity of projects collaborations profiles of bioenergies in France. This type of empirical work requests to go further with the characterisation of projects collaboration paying attention to the potential effects of collaboration dynamics. Firstly we have to measure the effect of collaborations in a period on collaborations in another period. As we noticed there is a promising methodology of characterising the trend of association of specific item sets (authors, keywords or indexed textual features) with ratios (see also the work realised with *Key-Words Lab* on the nanotechnology domain). Secondly there is a need to measure the effect of collaborations through projects on publication. Of course, basic or advanced statistics could be requested to measure the effect of project collaborations on ranking or co-citation indexes or also on patents. But we would be very curious to keep on paying attention to qualitative effect such as changes in the profile of ties in between periods, or such as changes in lexical composition of the repertoire of laboratories and principal investigators. Because we think that, there, in the performative narrative infrastructure of research projects and through R&D collaborations, is a good place to look at new sciences emerging and dynamics shifting, for it is there that actors construct them.

References

Bar-Ilan, J., 2008. Informetrics at the beginning of the 21st century—A review. *Journal of Informetrics*, 2(1), 52, 1.

Barré, R., 2001. Sense and nonsense of ST productivity indicators. *Science and Public Policy*, 28, 259-266.

Baur, M., 2008. *Visone Software for the Analysis and Visualization of Social Networks*. Ph.D. Thesis in Computer Science. Karlsruhe: Universität Fridericiana.

Beaver, D., 2001. Reflections on Scientific Collaboration (and its study): Past, Present, and Future. *Scientometrics*, 52(3), 377, 365.

Bender-deMoll, S. & McFarland, D.A., 2006. The Art and Science of Dynamic Network Visualization. *Journal of Social Structure*, 7(2).

Bonaccorsi, A., 2008. Search Regimes and the Industrial Dynamics of Science. *Minerva*, 46(3), 285-315.

Bourret, P. et al., 2006. A New Clinical Collective for French Cancer Genetics: A Heterogeneous Mapping Analysis. *Science Technology Human Values*, 31(4), 431-464.

Bozeman, B., Larédo, P. & Mangematin, V., 2007. Understanding the emergence and deployment of "nano" S&T. *Research Policy*, 36(6), 807-812.

Bozeman, B. & Rogers, J.D., 2002. A churn model of scientific knowledge value: Internet researchers as a knowledge value collective. *Research Policy*, 31(5), 769-794.

Brandes, U., 1999. *Layout of Graph Visualizations*. Ph.D. Thesis in Computer Science. Universität Konstanz.

Callon, M. & Law, J., 1986. *Mapping the Dynamics of Science and Technology: Sociology of Science in the Real World*, Sheridan House Inc.

Cambrosio, A. et al., 2006. Mapping the emergence and development of translational cancer research. *European Journal of Cancer*, 42(18), 3140-3148.

Cambrosio, A., Keating, P. & Mogoutov, A., 2004. Mapping Collaborative Work and Innovation in Biomedicine: A Computer-Assisted Analysis of Antibody Reagent Workshops. *Social Studies of Science*, 34(3), 325-364.

Cassier, M., 1995. *Les Contrats de recherche entre l'université et l'industrie: l'émergence d'une nouvelle forme d'organisation industrielle*. Thèse en sociologie. Grenoble: A.N.R.T Université Pierre Mendès France.

Chavalarias, D. & Cointet, J., 2008. Bottom-up scientific field detection for dynamical and hierarchical science mapping, methodology and case study. *Scientometrics*, 75(1), 37-50.

Crutzen, P. et al., 2007. N2O release from agro-biofuel production negates global warming reduction by replacing fossil fuels. *Atmos. Chem. Phys. Discuss.*, 7, 11191-11205.

Deuten, J.J. & Rip, A., 2000. Narrative Infrastructure in Product Creation Processes. *Organization*, 7(1), 69-93.

Duranton, G. et al., 2008. *Les pôles de compétitivité : Que peut-on en attendre ?*, Rue d'Ulm.

Etzkowitz, H., 2003. Innovation in Innovation: The Triple Helix of University-Industry-Government Relations. *Social Science Information*, 42(3), 293-337.

Freeman, C. & Soete, L., 2009. Developing science, technology and innovation indicators: What we can learn from the past. *Research Policy*, 38(4), 583-589.

Gibbons, M. et al., 1994. *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies*, London: SAGE Publications Ltd.

Grabher, G., 2004. Learning in Projects, Remembering in Networks? Communality, Sociality, and Connectivity in Project Ecologies. *European Urban and Regional Studies*, 11(2), 103-123.

Guston, D.H., 2001. Boundary Organizations in Environmental Policy and Science: An Introduction. *Science, Technology & Human Values*, 26(4), 399-408.

Heimeriks, G., Hörlesberger, M. & Van Den Besselaar, P., 2003. Mapping communication and collaboration in heterogeneous research networks. *Scientometrics*, 58(2), 391-413.

Hellström, T., Eckerstein, J. & Helm, A., 2001. RD management through network mapping: using the Internet to identify strategic network actors in cooperative research networks. *R&D Management*, 31, 257-263.

Hessels, L.K. & van Lente, H., 2008. Re-thinking new knowledge production: A literature review and a research agenda. *Research Policy*, 37(4), 740-760.

Hine, C., 2006. Databases as Scientific Instruments and Their Role in the Ordering of Scientific Work. *Social Studies of Science*, 36(2), 269-298.

Hine, C., 2005. The politics and practice of accessibility in systematics. Dans *Past, Present and Future of Research in the Information Society*. World Summit on the Information Society (WSIS). Tunis, Tunisia.

Jern, M., 2009. Collaborative Web-Enabled GeoAnalytics Applied to OECD Regional Data. Dans *Cooperative Design, Visualization, and Engineering*. p. 32-43. Available at: http://dx.doi.org/10.1007/978-3-642-04265-2_5 [Accédé Février 23, 2010].

Katz, S. & Martin, B., 1997. What is research collaboration? *Research Policy*, 26(1), 18, 1.

Larédo, P. & Mustar, P., 2000. Laboratory Activity Profiles: An Exploratory Approach. *Scientometrics*, 47(3), 515-539.

Latour, B., 1993. Ethnography of a "High-Tech" Case, About Aramis. Dans *Technological choices*. Pierre Lemonnier. Routledge, p. 372-398.

Latour, B., 1983. Give me a laboratory and I will raise the world. Dans *Science Observed: perspectives on the social study of science*. Sage, p. 169, 141. Available at: http://www.bruno-latour.fr/articles/article/12-GIVE%20ME%20A%20LAB.pdf [Accédé Février 27, 2010].

Law, J. & Callon, M., 1988. Engineering and Sociology in a Military Aircraft Project: A Network Analysis of Technological Change. *Social Problems*, 35, 284.

Lee, S. & Bozeman, B., 2005. The Impact of Research Collaboration on Scientific Productivity. *Social Studies of Science*, 35(5), 673-702.

Lucio-Arias, D. & Leydesdorff, L., 2007. Knowledge emergence in scientific communication: from “fullerenes” to “nanotubes”. *Scientometrics*, 70(3), 603-632.

Miettinen, R., 1998. Object Construction and Networks in Research Work: The Case of Research on Cellulose-Degrading Enzymes. *Social Studies of Science*, 28(3), 423-463.

Moody, J., McFarland, D.A. & Bender-deMoll, S., 2005. Visualizing Network Dynamics. *American Journal of Sociology*.

Noyons, E., 2001. Bibliometric mapping of science in a policy context. *Scientometrics*, 50(1), 83-98.

Powell, W.W. et al., 2005. Network Dynamics and Field Evolution: The Growth of Interorganizational Collaboration in the Life Sciences. *American Journal of Sociology*, 110(4), 1205, 1132.

Tari, T., 2009. *A la recherche du biocarburant, entre promesses et controverses. Etude de la forme projet dans les dynamiques de recherche sur les bioénergies en France*. Master Thesis. Paris: Ecole des Hautes Etudes en Sciences Sociales (EHESS).

Turenne, N. & Barbier, M., 2004. BELUGA : un outil pour l'analyse dynamique des connaissances de la littérature scientifique d'un domaine. Première application au cas des maladies à prions. Dans EGC Conference. Clermont-Ferrand.

Zitt, M. & Bassecoulard, E., 2008. Challenges for scientometric indicators: data demining, knowledge-flow measurements and diversity issues. *Ethics in Science and Environmental Politics*, 8, 60, 49.