



HAL
open science

The many dimensions of interdisciplinarity

Pablo Jensen, Katsyarina Paradzinets

► **To cite this version:**

Pablo Jensen, Katsyarina Paradzinets. The many dimensions of interdisciplinarity. 2013. hal-00785249v1

HAL Id: hal-00785249

<https://hal.science/hal-00785249v1>

Preprint submitted on 5 Feb 2013 (v1), last revised 5 Dec 2013 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The many dimensions of laboratories' interdisciplinarity

Pablo Jensen, Katsiaryna Lutkouskaya

Institut Rhône-Alpin des Systèmes Complexes (IXXI) and
Laboratoire de Physique, UMR CNRS 5672, École Normale Supérieure de
Lyon, 69007 Lyon, FRANCE

Abstract : Interdisciplinarity is as trendy as it is difficult to define. We propose three different operationalizations of a discipline and two levels (article or laboratory) of integration of these disciplines. This leads to six indicators of interdisciplinarity, achieving a rich characterization of laboratories publication practices. Thanks to a statistical analysis of these indicators on 600 CNRS laboratories, we suggest that, besides an average value of interdisciplinarity, different laboratories can be distinguished by the “distance” between the disciplines in which they publish and by the level of interdisciplinary integration (article or laboratory).

Introduction:

CNRS is the largest scientific organization in Europe, covering most fields of science. According to its website, “CNRS encourages collaboration between specialists from different disciplines [...] thus opening up new fields of enquiry to meet social and economic needs. CNRS has developed interdisciplinary programs which bring together various CNRS departments as well as other research institutions and industry.”

Recently, CNRS launched a “Mission for Interdisciplinarity” to “promote, facilitate and coordinate interdisciplinarity at CNRS.” This Mission aims to foster the development of new themes, or new disciplines”. To this end, it commissioned a study on the quantification of interdisciplinarity of the roughly 1000 laboratories affiliated to CNRS, with the help of scientometrics. In this paper, we present the results of this study.

We start by defining a variety of indicators, to capture many facets of interdisciplinarity through publications. Interdisciplinarity means, at the most generic level, some degree of integration of different disciplines (Weingart & Stehr, 2000; Porter & Rafols, 2009; Marcovich & Shinn, 2011; Wagner et al. 2011). To transform this idea into quantitative indicators, we need to answer to three questions:

1. How to define a discipline?
2. At what level the integration is achieved?
3. What is the degree of disciplinary integration achieved?

Let us briefly comment on these three points. There are several ways to define a discipline from a scientometrics' point of view. Since we are dealing with CNRS labs, the most natural would seem to use the disciplinary organization of CNRS in 10 “institutes” and 40 subdisciplinary “sections”¹. A convenient alternative is to adopt the categorization included in the database chosen, as the Journal Subject Category (JSC)

¹ <http://www.cnrs.fr/en/aboutCNRS/institutes.htm>

used by Web of Science (WoS). There are 224 JSCs, from Acoustics to Zoology, leading to a detailed description of the disciplinary fields, compared to that of CNRS. Finally, instead of using institutionally predefined partitions of science, one could use a more “bottom-up” definition of cognitive communities akin to disciplines. To obtain these communities, we use the roughly 300 000 French articles published between 2007 and 2010. We group them into “cognitive communities” using clustering algorithms based on shared references. More details are given below.

We now turn to the second question, the level at which the various disciplines are integrated. Again, there are several possibilities, including an article, a single author, a team and a laboratory. For example, an article may cite articles from different fields, or a laboratory may gather teams or authors from different fields, and try to have them working together on an interdisciplinary topic. Alternatively, a single scientist may have started his career in physics and then changed to history of science or environmental sciences, or collaborate and publish with scientists from different disciplines.

Finally, the hardest question from the point of view of quantification is related to the degree of disciplinary integration achieved. Does the knowledge produced simply juxtapose different disciplines, as in an encyclopedia that gathers entries from different fields? Has the interdisciplinary work achieved a full integration of the different fields to produce a new discipline? A wide range of integrations is possible, from the more superficial (pluridisciplinarity) to the deepest (creation of a new discipline). So far, although we consider it a key issue, we have not been able to distinguish between the different possibilities of integration using scientometrics’ indicators.

In this paper, we will use three definitions of discipline and two integration levels (laboratory, article) to calculate six interdisciplinary indicators. We do not pretend to find the “best” indicator of interdisciplinarity, but to show that interdisciplinarity has many relevant dimensions, to elaborate a rich description of the interdisciplinary practices of CNRS labs.

I Presentation of six indicators

I – 1 Indicators built on the diversity of publications' JSCs

These indicators are based on Journal Subject Categories (JSC) of Web of Science. We exclude JSC “Multidisciplinary Sciences” from analysis because by definition this JSC mixes articles from many disciplines. If a journal is attributed to n JSCs, it contributes $1/n$ to each of them. From the list of a lab’s publications, we calculate its distribution over the JSC through p_i , the proportion of articles of this laboratory in JSC_i . Two examples of this distribution are given in Figure 1.

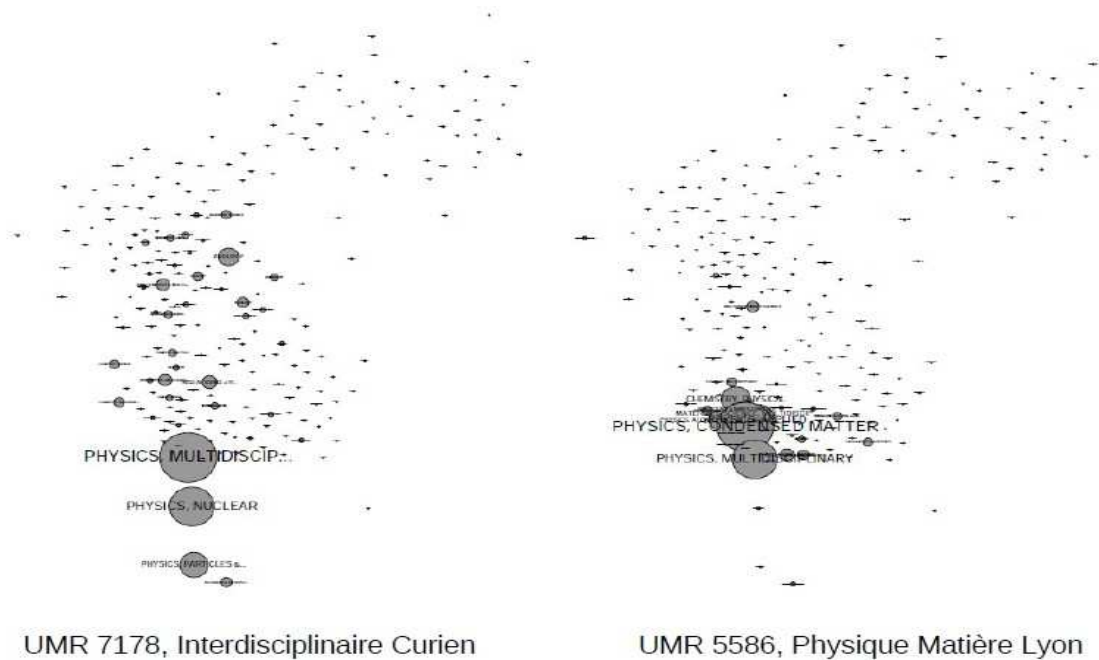


Figure 1 : Each lab can be represented in a JSC map similar to the one proposed by Leydesdorff & Rafols (2008).

I - 1 a - Entropy of SCs of laboratory:

A simple indicator of the spread of the disciplines where a laboratory publishes is given by the entropy:

$$e_{jsc} = - \sum_{i=1}^{N_{jsc}} p_i \ln(p_i) \quad (1)$$

where p_i is the proportion of articles of the laboratory in JSC_i and N_{jsc} is the number of JSCs in which the laboratory has published.

This indicator does not take into account how similar or dissimilar these fields are: publishing 50% of the papers in “Applied Physics” and 50% in “Condensed Matter Physics” leads to the same value of the entropy indicator as publishing 50% of the papers in “Applied Physics” and 50% in “Sociology” which is, however, not equivalent from a cognitive point of view.

I - 1 b- Integration

As we would like to include the idea of “distance” between disciplines, we calculate the integration indicator (Stirling 2007, Porter & Rafols, 2009) which combines both the spread of the disciplines through the p_i and the distance between them through s_{ij} .

$$dist_{jsc} = \sum_{i,j,i>j} p_i p_j \frac{1}{s_{ij}} \quad (2)$$

where s_{ij} is the cosine measure of similarity between SCs i and j . Practically, s_{ij} is measured through the citations from publications in SCs i to publications in SC j (Porter & Rafols, 2009). When discipline i never quotes any reference to articles belonging to discipline j , s_{ij} takes the value 0, while s_{ij} is close to 1 when such references are common. If we compare the two examples mentioned above, the first case leads to $dist_{jsc} = 0.267$, as the JSCs are very similar ($s_{ij} = 0.935$), while the second case leads to a much higher value ($dist_{jsc} = 1250$, since $s_{ij}=0.0002$ between “Applied Physics” and “Sociology”).

Note that we use $1/s_{ij}$ to account for a distance between disciplines that increases with decreasing s_{ij} , contrary to Porter & Rafols (2009) or Roessner et al (2013) who use a quantity close to $(1 - s_{ij})$. Clearly, using the inverse leads to a much higher contrast between close and distant disciplines, which is what we are after here in order to detect different styles of interdisciplinarity.

I - 2 Indicators built on the interdisciplinarity of single publications

A possible bias of both the entropy and integration indicators is that they may show high values for a laboratory that gathers teams of different disciplines (and that, therefore, publish in different SCs), even when there is no significant interaction between the different teams. To further characterize the level of interdisciplinarity, one can characterize the interdisciplinarity of single articles, to test if the integration of the disciplines is also effective at this cognitive level.

I - 2 a Single article interdisciplinarity

The integration of a single article is calculated as for labs (equation 2) by integrating the JSC of the *references* listed in the article. This supposes that one can attribute each reference to a JSC, which is not always the case. We have obtained the JSC of a reference from the name of journal where it appears. Journal names of references are given by WoS in the form of a 29-character abbreviation. We could identify 6,960,940 references out of 9,626,563 (after data cleaning described in <http://www.sebastian-grauwin.com/>) leading to a recovery rate of 72.3%, to be compared to the recovery rate of 80% achieved by (Porter& Rafols, 2009).

Specifically, $art_integration$ is calculated as:

$$art_{integration} = \sum_{i,j,i>j} p_{a_i} p_{a_j} \frac{1}{s_{ij}}$$

where pa_i is the proportion of references of article in JSC_i . Integration is therefore equal to 0 if all the references of an article belong to a single JSC.

In practice, we need to use a corrected value for *art_integration* because WoS contains errors in cited references. A cited reference attributed to a distant field (for example, to Sport instead of Physics) results in an erroneously high value of integration. To overcome this problem, we calculate the *corrected value of integration*:

$$art_integration_corr = min(art_integration_1)$$

where *art_integration_1* is the “integration without one reference”. It is calculated as follows: one reference is dropped from the list of references of the article and the value of integration is calculated. Then, the next reference is dropped (the first reference is restored) and the second value of integration is calculated, and so on. The minimal integration from this set of integrations is selected as *art_integration_1*. In this way, we avoid not only false references but also artificially high values of the interdisciplinarity of a single paper which would result from a single reference. In the following, we use the corrected value for *art_integration*.

We have also calculated an alternative indicator of the interdisciplinarity of single papers through the integration of the JSC of their *connected* papers. To do this we connect a given paper to those which it shares at least two references, and compute the integration formula above using the JSCs of the connected papers. This indicates the range of disciplines to which a paper links through papers sharing references with it. We will not show the results obtained with this indicator as we found it to be strongly correlated with *art_integration_corr*.

I - 2 b Laboratory level indicator

One can aggregate this article indicator at the laboratory level by averaging over all the articles published by that laboratory :

$$art_{jsc=} = \frac{\sum_i art_integration_corr}{nbpap} \quad (3)$$

where *nbpap* is the number of articles for which at least one reference was identified.

I - 2 c- Integration of article without distance:

We can also use a calculation of integration similar to the integration of *art_jsc_integration* but without similarity matrix s_{ij} :

$$art_integration_{nodist=} = \sum_{i,j,i>j} pa_i pa_j \quad (2a)$$

This indicator is expected to be more robust when looking at dynamics of interdisciplinarity over large periods of time, especially over 10 and more years, since the relationships between the disciplines, and hence the measure of similarity s_{ij} may change in time.

I - 3 Indicators built on CNRS' disciplinary « Institutes »

CNRS is divided in 10 disciplinary institutes: Institute of Biological Sciences (INSB), Institute of Chemistry (INC), Institute of Ecology and Environment (INEE), Institute for Humanities and Social Sciences (INSHS), Institute for Information Sciences and Technologies (INS2I), Institute for Engineering and Systems Sciences (INSIS), National Institute for Mathematical Sciences (INSMI), Institute of Physics (INP), National Institute of Nuclear and Particle Physics (IN2P3) and National Institute for Earth Sciences and Astronomy (INSU). Each laboratory is assigned to an Institute which is supposed to reflect the majority of the research fields studied by the lab. We have built two indicators based on the spread of the lab's academic production over the disciplinary Institutes.

I - 3 a- Proportion of publications outside the Institute's mainstream JSC

We start by creating a list of the most common JSCs for each institute. For this, we take all the publications of the laboratories belonging to an institute and classify them by JSC. For example, for Life Sciences, the list is given in Table INSB. Then, we choose a threshold to define the most common JSCs for each institute. We have studied thresholds of 80, 90 and 99% and found that there were no significant variations in the results obtained. We therefore choose a threshold value of 90% (Table INSB).

#	JSC	percentage, %	cumulative sum, %
1	BIOCHEMISTRY & MOLECULAR BIOLOGY	16.65	16.65
2	NEUROSCIENCES	9.13	25.78
3	CELL BIOLOGY	6.50	32.28
4	ENDOCRINOLOGY & METABOLISM	3.87	36.15
5	MICROBIOLOGY	3.83	39.97
6	GENETICS & HEREDITY	3.76	43.73
7	PLANT SCIENCES	3.34	47.07
8	MULTIDISCIPLINARY SCIENCES	3.32	50.39
9	IMMUNOLOGY	3.14	53.54
10	BIOPHYSICS	2.46	55.99
11	ONCOLOGY	2.45	58.44
12	PHARMACOLOGY & PHARMACY	2.40	60.84
13	DEVELOPMENTAL BIOLOGY	2.30	63.14
14	HEMATOLOGY	2.08	65.22
15	CLINICAL NEUROLOGY	1.99	67.20
16	VIROLOGY	1.69	68.89
17	PHYSIOLOGY	1.53	70.42
18	BIOTECHNOLOGY & APPLIED MICROBIOLOGY	1.53	71.95
19	CHEMISTRY, MULTIDISCIPLINARY	1.43	73.38
20	BIOCHEMICAL RESEARCH METHODS	1.42	74.80
21	MEDICINE, RESEARCH & EXPERIMENTAL	1.04	75.84

22	BEHAVIORAL SCIENCES	0.98	76.82
23	BIOLOGY	0.96	77.78
24	CHEMISTRY, MEDICINAL	0.89	78.66
25	CHEMISTRY, ORGANIC	0.83	79.49
26	GASTROENTEROLOGY & HEPATOLOGY	0.82	80.31
27	DERMATOLOGY	0.80	81.11
28	INFECTIOUS DISEASES	0.77	81.88
29	PSYCHIATRY	0.76	82.64
30	CHEMISTRY, PHYSICAL	0.75	83.39
31	PATHOLOGY	0.65	84.04
32	ZOOLOGY	0.59	84.63
33	REPRODUCTIVE BIOLOGY	0.58	85.21
34	CHEMISTRY, ANALYTICAL	0.57	85.78
35	TOXICOLOGY	0.56	86.34
36	RADIOLOGY, NUCLEAR MEDICINE & MEDICAL	0.55	86.88
37	IMAGING	0.54	87.42
38	PERIPHERAL VASCULAR DISEASE	0.51	87.93
39	RHEUMATOLOGY	0.40	88.33
40	CARDIAC & CARDIOVASCULAR SYSTEMS	0.40	88.74
41	FOOD SCIENCE & TECHNOLOGY	0.39	89.12
42	PARASITOLOGY	0.37	89.49
43	MEDICINE, GENERAL & INTERNAL VETERINARY SCIENCES	0.36	89.86

Table INSB : List of the most common JSCs for CNRS's Biology Institute (INSB) for the threshold of 90%

Then, for each laboratory, we count the percentage of articles outside this 90% list and normalize the "marginal JSC" count by dividing by the expected value, i.e. the average value 0.1.

$$j_{sc90} = 1 - \frac{\sum_i p_{i, JSC \cap JSCs_{INST}}}{0.1} \quad (4)$$

where $p_{i, JSC \cap JSCs_{INST}}$ are the probabilities of the JSCs that belong to the institute's JSC list.

I - 3 b- Proportion of co-publications with CNRS laboratories belonging to different Institutes

Another way of measuring interdisciplinary collaborations is to compute the proportion of a lab's publications that involve authors from laboratories belonging to a different Institute. We therefore compute:

$$copubli_{inst=} = \frac{\sum_i article_{i, nb_{INST} \geq 2}}{N_{articles}} \quad (5)$$

where the sum counts the number of articles that involves authors from at least two institutes and $N_{articles}$ is the number of articles published by the laboratory.

I - 4 Indicator built on cognitive disciplines

Finally, we use an alternative definition of discipline, not institutional as before (WoS or CNRS), but emerging from the articles themselves. To build these “cognitive disciplines”, we use bibliographic coupling (Kessler, 1963) between the 320 000 papers sampled from WoS. Links between pairs of articles are calculated through their common references (at least two references are needed to create a link, to avoid non significant links based on general references). Links are weighted, i.e. we define a similarity between two articles i and j as the cosine distance:

$$\omega_{ij} = \frac{|\mathcal{R}_i \cap \mathcal{R}_j|}{\sqrt{|\mathcal{R}_i| |\mathcal{R}_j|}}$$

where R_i is the set of references of article i . By definition, $w_{ij} \in [0,1]$ and is equal to zero when i and j do not share any reference and to 1 when their sets of references are identical. In comparison to a co-citation link (which is the usual measure of articles’ similarity), bibliographic coupling (BC) offers two advantages: it allows to map recent papers (which have not yet been cited) and it deals with all published papers (whether cited or not).

The reason why weighted links are used is that they reinforce the dense (in terms of links per article) regions of the BC networks. This reinforcement facilitates the partition of the network into meaningful groups of cohesive articles, or communities. A widely used criterion to measure the quality of a partition is the modularity function (Fortunato & Barthélémy 2007), which is roughly is the number of edges inside communities (as opposed to crossing between communities), minus the expected number of such edges if the partition were randomly produced. We compute the graph partition using the efficient heuristic algorithm presented in (Blondel et al 2008). The whole method is described in (Grauwin & Jensen, 2011).

Applying this algorithm yields in a partition of French papers into roughly 250 communities containing more than 100 papers each. Simple frequency analysis then allows to characterise each community through its more frequent items (keywords, authors, etc...). Once these cognitive disciplines are created, we calculate the proportion of each lab’s publications in each community (each paper belongs to a single community) and call this p_i . The distance between communities is set as the inverse of the mean bibliographic coupling weight between articles in communities i and j :

$$\langle w \rangle_{IJ}^{-1} = \langle w_{ij} \rangle_{i \in I, j \in J}^{-1} = (\Omega_{IJ} / N_I N_J)^{-1}$$

where Ω_{ij} is the total weight of the links between communities I and J , N_i and N_j the number of articles in communities I and J respectively. By analogy with the integration formula based on JSC (formula 2), we define the integration across “cognitive communities” as:

$$art_{cogn} = \sum_{i,j,i>j}^i p_j \frac{1}{s_{ij}} \quad (6)$$

where $p_{i,j}$ is the proportion of papers published by the lab in the community i .

II Results

II 1 Orders of magnitude: how interdisciplinary are average papers?

Before analyzing in detail the results obtained for the laboratories, it is useful to compute some order of magnitudes on the range of disciplines used, on average, by papers and laboratories.

It turns out that articles citing more than 10 references (2/3 of the French articles, with an average number of references equal to 32.5) refer to papers from almost 10 different disciplines (9.8 JSC). However, when considering those JSC that are used in more than 10% of the references list, this average drops to 2.7. This means that, on average, an article spreads its references on 3 main JSCs and 7 additional which benefit from roughly a single reference.

At the laboratory level, counting the 680 that have published more than 50 papers over 2007-10, they publish in journals belonging to 34 JSC, which reduces significantly to 2.5 distinct JSC when imposing the same threshold of 10%.

Finally, it is interesting to note that, contrary to what is sometimes asserted (but in agreement to other quantitative studies, Porter & Rafols, 2009; Larivière & Gingras 2013), there is only a small increase in the interdisciplinarity of articles, as shown in Figure history. This slow increase may reflect a progressive drift of scientific practices away from fixed JSC, without any real crossing of cognitive frontiers.

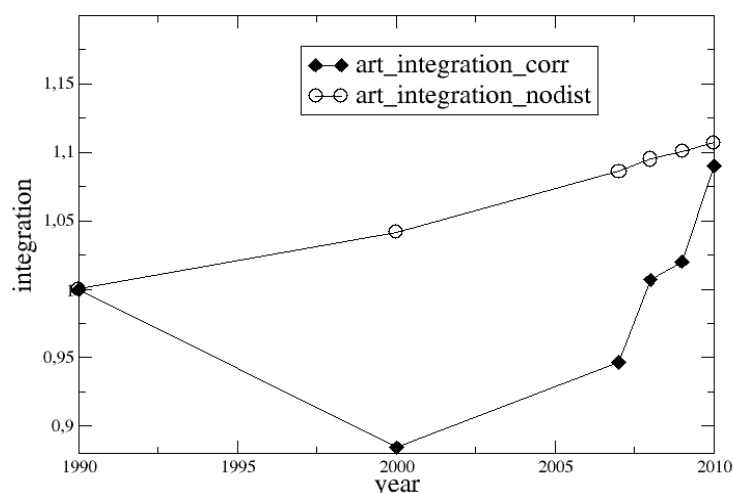


Figure history: Evolution of the interdisciplinary integration of articles over the last 20 years, normalized to 1 in 1990. We plot both the integration with (*art_integration_corr*) and without (*art_integration_nodist*) distance between disciplines (formula 2a).

II 2 Laboratory indicators

We have computed the six indicators on the 680 laboratories which have published more than 50 papers over 2007-2010. To allow comparisons and statistical analysis, since the absolute values have no intrinsic meaning, we have scaled all the figures to achieve an average value of 0 and a variance of 1. We then carried out a Principal Component Analysis of the (680 × 6) matrix using the free software R (www.r-project.org/).

II 2 a PCA analysis

Table I shows the coordinates of the first four axis of the PCA. Their cumulative variance is : 0.38; 0.57; 0.72 and 0.83.

Indicator	PCA1	PCA2	PCA3	PCA4
<i>e_jsc</i>	0.40	-0.43	0.23	-0.62
<i>copubli_inst</i>	0.37	-0.35	0.49	0.70
<i>jsc90</i>	0.49	-0.24	-0.18	-0.17
<i>art_cogn</i>	0.31	0.64	0.45	-0.12
<i>art_jsc</i>	0.40	0.01	-0.68	0.26
<i>dist_jsc</i>	0.45	0.47	-0.11	0.02

Table 2 : coordinates of the first four axis of the PCA analysis

We now turn to the interpretation of the main four axis.

II 2 b Some representative examples

Let us now discuss the principal dimensions that emerge from this analysis, together with some representative labs.

II 2 b 1 PCA1: Average interdisciplinarity axis

The main axis represents an average of the various interdisciplinarity indicators. A representative example of the laboratory with low interdisciplinarity is UMR7652, an Organic Synthesis laboratory from the Chemistry Institute located at Ecole Polytechnique (*pca1* = -3.2). All its six interdisciplinarity indicators are well below average values. The lab publications are highly centered on the Chemistry JSCs, with only one publication over 79 published in a journal not entirely related to Chemistry (JSC = "Pharmacology"). Similarly, 50 of the 51 publications of the

“Observatoire Aquitain” (UMS2567) are within a single JSC “Astronomy and Astrophysics” ($pca1=-3.2$).

On the other side of the spectrum, Laboratoire Interdisciplinaire Hubert Curien (UMR7178) has the strongest $pca1$ (+8.4). It has also the strongest int_coeur and $dist_jsc$ because it publishes papers on very distant JSCs: 143 papers on “Physics, Nuclear”, 50 on “Zoology”, 27 on “Chemistry, Multidisciplinary”... over a total of 929 papers. However, its papers are less interdisciplinary than average ($art_jsc = -0.24$), pointing to a low degree of integration of its different teams from different disciplines, as will be pointed out by $pca3$ (see below). Another very interdisciplinary lab is TIMC-IMAG (Techniques for biomedical engineering and complexity management, UMR 5525, $pca1=+6.9$). Out of its 304 papers, 24 were in published in JSC=“Neurosciences”, 8 in “Genetics & Heredity”, 7 in “Sports Sciences”, 4 in “Mathematics”...

II 2 b 2 PCA2: Short and long distance axis

This axis distinguishes those labs that connect distant or nearby disciplines. For example, the Center for the study of divided matter, UMR 6619, is quite interdisciplinary ($pca1=+1.4$) but on rather short distances ($pca2=-2.1$), as its main JSC indicate: Chemistry, Physical (19%), Materials Science (14%), Physics, Atomic (7%)... On the contrary, the “Center of research on the mathematics of decision”, UMR 7534, with a similar overall interdisciplinarity ($pca1=+2.1$) reaches longer distances ($pca2=+3.1$), as its JSC include “Astrophysics” and the very distant “Biology”, or “Automation”. Note that the calculation of the integration including as “distance” the inverse cosine measure ($1/s_{ij}$) strongly increases the contribution of distant fields, even if their representation in the publications’ list of a lab is small. In this last case, “Astrophysics”, “Biology” or “Automation” represent each 1% of the lab’s publications, but contribute hugely to its total integration, as they are very distant (huge $1/s_{ij}$).

II 2 b 3 PCA3: article or laboratory interdisciplinarity

This axis distinguishes labs that publish rather mono-disciplinary articles (low art_jsc) but in journals from different (and distant) JSC (high $dist_jsc$) from labs that publish interdisciplinary articles (high art_jsc) in a few (and related) JSC (low $dist_jsc$). The two extremes are given by Laboratoire Interdisciplinaire Hubert Curien (UMR7178). As we have seen previously, it has the strongest $pca1$ (+8.4) but also the strongest $pca3$ (+7), arising from its low art_jsc (-0.24). This means that its papers are mostly mono-disciplinary, pointing to a lab constituted of a juxtaposition of rather mono-disciplinary teams, each publishing in its own field. Overall, the lab seems interdisciplinary, but this may well be an artifact of our indicators that, except art_jsc , consider these teams as a single entity, the laboratory, as if it had managed to integrate across the disciplinary teams, an unwarranted assumption... On the opposite side, the lab “Natural history of the prehistoric man”, UMR 7194, is also interdisciplinary on average ($pca1=+1.9$) but has a strongly negative $pca3$ (-4.6), arising from

a small `int_coeur` (-0.5) and a huge `art_jsc` (+6.7). UMR 7194 publishes in journals linked to its disciplinary heart (“Evolutionary Biology”) but its papers gather references not only to this field, but also to the analytical tools used to analyze objects such as teeth, the tools coming from other disciplines such as Chemistry or Physics, which are quite distant from “Evolutionary Biology” or “Paleontology”. This gives rise to highly interdisciplinary papers as quantified by `art_jsc`.

II 2 b 4 PCA4: Publishing in different JSC or co-publishing with laboratories from other Institutes

This axis distinguishes labs that publish in journals from different JSC (high `E_JSC`) from labs that co-publish with labs from different CNRS Institutes (high `copubli_inst`). Consider for example UMR7194 (“Natural history of the prehistory man”) and UMR8151 (“Genetic and Chemical Pharmacology and Imagery”): they show approximately the same average interdisciplinarity (PCA1 equal to (+1.9 and +2.0 respectively) but they appear as opposed along the PAC4 axis. The interpretation is that the first lab, as we have seen (II 2 b 3), publishes in a few JSCs (low `E_JSC`) but co-publishes with many labs from other Institutes, while the Genetics lab does the reverse: it publishes in many JSCs (more than 2 papers on 19 different JSCs) but collaborates with only a few labs outside its Institute.

III Discussion & Perspectives

What do these indicators allow to say about the interdisciplinarity of CNRS labs? First, let us note that using the first four PCA axis gives an overall view about the interdisciplinarity practices of each lab. This view has been compared to the expert knowledge coming from scientists of the labs or scientific advisors from CNRS for about 20 different labs from all the disciplines. There is a strong conviction that the view provided by the indicators reflects correctly the reality as seen from the field. In this respect, our study is successful, as it allows to obtain a reliable characterization of the interdisciplinarity aspects of a lab from its publications.

However, scientometrics does not allow to distinguish real *interdisciplinary* collaborations, giving rise to new concepts, or to a coherent new scientific field, from simple *pluridisciplinary* practices that simply juxtapose different disciplines, as when historians use characterizing tools from physics. It is difficult to learn much about the cognitive dimensions of interdisciplinarity from an automatic analysis of metadata of the papers.

There are open questions that could be addressed by further studies:

- what can actors (in this case, scientists) from the field learn from this analysis? It is reasonable to assume that they are already aware of the degree of interdisciplinarity of their own practices. Maybe they can learn from the comparison with the results obtained by other labs, from the same discipline or not.

- How do CNRS policy officers use or learn from this data? How do they update their policy?

Finally, on a more technical level, it would be interesting to investigate whether the interdisciplinarity of articles obtained through its references correlates well with another indicator, given by calculating the integration not of its references but of its *citations by subsequent papers* (what Roessner et al 2013 call Diffusion as opposed to Integration). The point is that, while references are chosen by the author, citations are not so easy to control and may reflect a truer multidisciplinary impact. While it is quite easy to automate the calculation of the interdisciplinarity based on a paper's references, it is very long to search for its citations and then calculate a similar indicator. Therefore, as a first step, we have performed the calculation on a selection of hundred papers. Figure refcit shows that papers that use references from a few disciplines (low `art_integration_corr`) are likely to receive citations also from a handful of disciplines (low `ID_cit`). However, papers that appear as very interdisciplinary by looking at their reference list can either be cited by very different disciplines (upper - right side) or cited by a few and related disciplines (lower - right side). Therefore, while `art_integration_corr` seems to be reliable to detect low interdisciplinarity papers, however, it is not necessarily the best indicator of high interdisciplinarity papers, as `art_integration_corr` may simply reveal that the author pretends to be relevant for several disciplines, relevance that the corresponding disciplines may not acknowledge, for lack of interest or awareness. In this paper, we have mainly used `art_integration_corr` to detect labs that appear interdisciplinary by their publication fields but not by their articles, and this use seems to be valid since a low `art_integration_corr` seems to be a good indication of low interdisciplinarity. We note that Larivière & Gingras (2013) found that both indicators give essentially the same result.

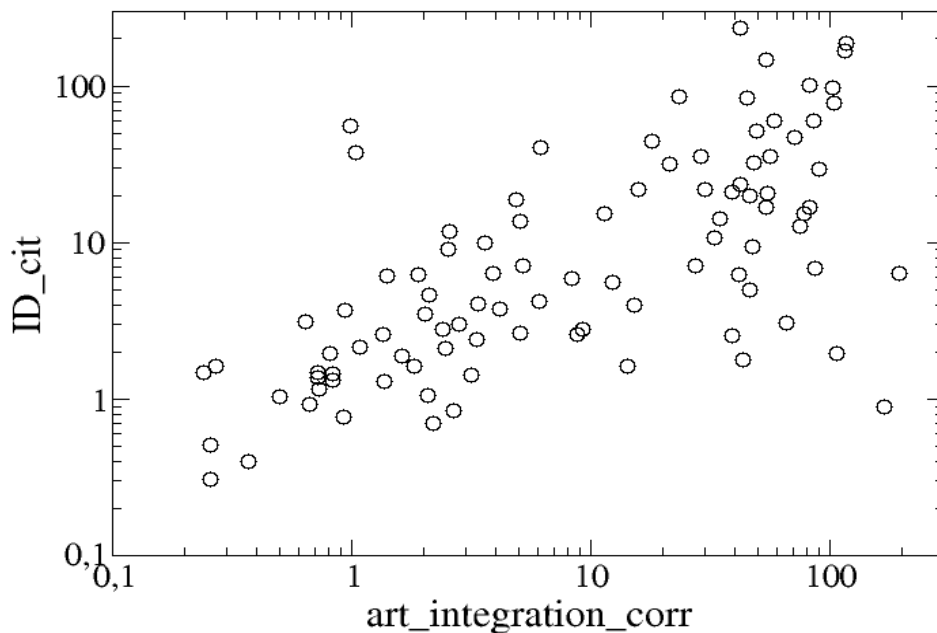


Fig refcit: interdisciplinarity of a paper measured from its citations (ID_cit) compared to that measured from its references (art_integration_corr).

Acknowledgement: We thank the “Service d'appui à la politique et à la prospective scientifiques” of the CNRS which provided publication data related to CNRS laboratories for our research work. We also thank CNRS Interdisciplinary Mission for financial support.

References:

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics*, P10008

Fortunato, S & Barthélemy M (2007). Resolution limit in community detection. *Proceedings of the National Academy of Sciences USA*, 104, 36-41.

Grauwin S & Jensen P (2011) Mapping scientific institutions *Scientometrics* **89** 943-954

Kessler MM (1963). Bibliographic coupling between scientific articles. *American Documentation* **24** 123-131.

Larivière V & Gingras Y (2013) Measuring Interdisciplinarity over the Course of a Century, to be published in B. Cronin & C. Sugimoto (Eds.),

Next Generation Metrics: Harnessing Multidimensional Indicators of Scholarly Performance. Cambridge MA: MIT Press.

LEYDESDORFF, L., RAFOLS, I. (2008) A global map of science based on the ISI subject categories. *Journal of the American Society for Information Science and Technology*.

Marcovich A & Shinn T (2011) Where is disciplinarity going? Meeting on the borderland, *Social Science Information* **50** 1-25

Porter AL and Rafols I (2009) Is science becoming more interdisciplinary? Measuring and mapping six research fields over time *Scientometrics*, *81*, 719-745

Roessner D, AL Porter, NJ Nersessian and S Carley (2013) Validating indicators of interdisciplinarity: linking bibliometric measures to studies of engineering research labs *Scientometrics* **94** 439-468

STIRLING, A. (2007), A general framework for analysing diversity in science, technology and society. *Journal of The Royal Society Interface*, **4** (15) 707-719.

Wagner, C. S., Roessner, J. D., Bobb, K., Klein, J. T., Boyack, K. W., Keyton, J., et al. (2011). Approaches to understanding and measuring interdisciplinary scientific research (IDR): a review of the literature *Journal of Informetrics*, *5*(1), 14-26.

Weingart, P., & Stehr, N. (2000). *Practicing interdisciplinarity*. Toronto: University of Toronto Press