

# Understanding opinion and language dynamics using massive data (OpLaDyn)

## 1 Statement of significance

In the last few years, Human Society has undergone unprecedented changes, driven by the sudden increase in communicating technological devices that surround us, keeping traces of a large amount of our daily activities. This extremely rapid evolution, which is just starting; is associated with a fast adaptive dynamics that induces changes in our everyday practices and has consequences for the nature of the social relations that we develop.

On the one hand, the ubiquity of communication devices, by allowing us to interact with many more individuals, changes the way in which we interact with each other, for example, by allowing interaction among geographically distant people, or by bringing together very different people, who would not have been able to share a discussion or support a cause before, to converge on a particular action. These changes seem to affect the very notion of *social interaction*. On the other hand, an increasing number of our common actions leave digital traces that are collected by different kinds of agents such as governments, scientific societies, commercial firm and NGOs, etc. The fact that the activities of human society can be massively monitored and stored is a new feature in history, and the impact of this fact on our behavior is far from trivial. This rapidly increasing amount of data, called *Big Data*, leads to the urgent question of its storage, organization, retrieval and control. *Data scientists* concentrate on the technical aspects of these problems in order to make massive data ready to be used. This project addresses two different questions: (a) *How can relevant information be obtained from this massive data, useful for the elaboration of explanatory models of different aspects of human social activities?* (b) *What are the possible ethical consequences of the application of Big Data analyses in the study of self-organized human actions?*

We are interested in both, searching for traces of social activity in raw data and studying the impact that this new situation has on social behavior. Our aim is to contribute to the construction of an interdisciplinary view of the relation between informational pattern correlations, available in Big Data, and the dynamics of social actions, bridging the social and the natural sciences. Specifically, we will focus on the study of opinion dynamics and language evolution based on data issued from two very different media, on one hand a historical newspaper (the New York Times) and on the other a new online medium (Twitter).

The involved Principal Investigators are:

**Argentina:** José-Ignacio Alvarez-Hamelin, Facultad de Ingeniería, Universidad de Buenos Aires

**Brazil:** Maria Eunice Quilici Gonzalez, Philosophy Department, Faculty of Philosophy and Sciences, University of São Paulo State - UNESP

**France (coordinator) :** Laura Hernández, Laboratoire de Physique Théorique et Modélisation (LPTM) UMR8089 CNRS-Université de Cergy-Pontoise.

## Table of Contents

|  |    |
|--|----|
| <b>1. Statement of significance</b>                                  | 1  |
| <b>2. Table of contents</b>  | 2  |
| <b>3. Narrative</b>  | 3  |
| 3.1 Description of the subject, objectives, and research questions   | 4  |
| 3.2 State of the art   | 5  |
| 3.3 Description of the data chosen for the project                   | 8  |
| 3.4 Research questions, contributions of different teams             | 9  |
| 3.5 History of the project   | 11 |
| 3.6 Teaching through research  | 12 |
| <b>4. References</b>   | 13 |
| <b>5. Data management plan</b>                                       | 15 |
| <b>6. Project management, dissemination, and communications plan</b> | 17 |
| 6.1 Roles and responsibilities                                       | 17 |
| 6.2 Collaborations   | 18 |
| 6.3 Work plan  | 19 |
| 6.4 Dissemination and communication                                  | 21 |
| <b>Budget</b>  | 22 |
| <b>Resumes</b>   | 42 |
| <b>Letters of Commitment</b>   | 45 |
| <b>Appendices</b>  | 48 |

## 3 Narrative

### 3.1 Description of the subject, objectives, and research questions

New technologies are rapidly changing the informational environment in which people operate, and are taking the interdependence of behavior and decision-making of social agents, to an entirely new level. The speed at which information flows in online networks generates bursts of activity in information exchange, and this translates into new forms of organization. Such an effect has recently been manifested in political protests as well as in the way in which peer-to-peer innovation networks operate. Online technologies are also enlarging personal networks, making the group of reference from which individuals get information less restricted to geographic boundaries or traditional social circles.

Our theoretical understanding of collective human social actions can be greatly improved by the higher-resolution data facilitated by these technologies; however, these data also require new models and methodologies to account for the greater complexity of human interaction, expanded more than ever before across networks with different reach and spatial scope. This increased connectivity, along with the reduction of the time-scale associated to online communication, have important consequences for the dynamics of diffusion and contagion, which provides the methodological approach we propose with not only theoretical but also practical relevance. It is clear that data generated in this way will increase at an extremely high speed, and beyond the urgent technical challenges, related to storage and processing, fundamental questions arise : *What **can** we do with all this massive data? What **should** we do with it?*

In fact, the avalanche of digital traces generated by our activities *is already being used* with the objective of controlling or modifying our behavior, for example, by inducing new habits of consumption. Publicity is certainly not a novelty; what is new is the fact that this personal information, carelessly left available, is used to create new needs for which new products will be later proposed to each of us in a personalized manner. The aim of this process is to improve commercial activity, but this is certainly not the only consequence; it may also modify our behavior. This point implies the need for models of social behavior, such as opinion formation or decision making, which integrate this personalized feedback. Then, global approaches alone are insufficient for such modeling.

In order to study models of social systems based on Big Data, we propose a *Complex Systems* approach, searching for regularities (stylized facts) in social data, in the form of textual, spatial, or temporal patterns, which help us to characterize the collective phenomena studied. This requires Computer Science expertise, in order to develop tools that help to represent and analyze the traces of social behavior that can be extracted from raw data. In parallel, Social Science expertise is needed at all stages of the process: as a guide to detect the data that could be usefully extracted from the databases, to contribute to the choice of the relevant models and for the analysis of the results, including the ethical implications of this procedure.

This project therefore belongs to the emergent interdisciplinary field of *Computational Social Science*, combining expertise from both the formal sciences (here mainly Statistical Physics, Dynamical Systems and Computer Science) on one hand, and the social sciences including Philosophy, Cognitive Sciences, Linguistics, Law, and Sociology, on the other.

The data we produce daily contains information about our behavior, such as typical habits of movement or consumption as well as our preferences and opinions: our favorite books, our inclination to adopt a new word or concept, our opinion on social conflicts or our choices... This

data thus contains quantitative information about what large groups of people actually do, think, like or believe. Based on this one can not only describe or predict what people do, under given external conditions, but also try to get an insight on *why they do so*.

As mentioned, algorithms that perform large scale analyses of this data are already widely used, which implies the existence of data based models of social behavior behind them. However the models involved in these applications aim at suggesting or predicting a given behavior, and do not intend to explain social action in terms of the properties of agents composing the system and their interactions. It is important to understand that the difference between a "predictive" and an "explanatory" model is not only formal. It is at the root of the very definition of the model and conditions the results that one can expect from it.

Our objective is to build data-driven explanatory models. In order to do so, we propose to use two very different data bases on social systems. Based on the information issued from the (possibly) detected regularities, we propose to build multi-level models that take into account not only microscopic interactions (agent-agent interaction), as is usual, but also the interaction at the level of groups in order to account for retro-action loops which has been much less studied so far.

The leading question can therefore be put in this way: can we obtain information about the *most relevant social interactions* that govern the dynamics of different aspects of social behavior, by studying the ICT traces left by social actors?

This proposal is based on a positive answer to the above question. The main hypothesis is that, by analyzing significantly large amount of data, one should be able to detect local patterns of behavior. Based on them, and with the help of the knowledge acquired from the different disciplinary fields involved in each specific problem, it is then possible to build a model of an artificial society and study the emergent regularities to which the chosen dynamics could lead, in order to compare them with the stylized facts observed in the real world. This procedure, which can be called "from data to models and back", will be developed along with some case studies, described in the next sections.

We will focus on two particular aspects of social behavior: opinion dynamics and language evolution. We will address the study of these problems using data obtained from two databases that have very different constraints in terms of style, type of users, and limits of expression: on one hand Twitter data, and on the other a more traditional medium, the New York Times data base.

### 3.2 State of the art

Very recently, two independent articles in the form of a *manifesto*, pleaded for the implementation of what they called *Computational Social Science*. Though very different in their approach and analysis of the problem, both groups of authors agreed on the notion that the large amount of data that we generate in our everyday lives offers an opportunity to obtain new knowledge about our societies. The first sentence of the article, published in Science in 2009 by a group of researchers from different universities in the U.S.A. [1] is : *We live life in the network*. This statement, intentionally provocative is certainly not true. Most of the important, pleasant or difficult features of our life occur off-line. However many of these activities leave traces on-line, often outside the limits that most of us know or can control. The main claim of this work is that Big Data provides the key for a revolution of the Social Sciences which, until now, has relied on self-reported data about the relationships or preferences of relatively small groups of individuals. This huge amount of data is nowadays stored by the different public and private agents. As the ownership of the this data is not yet subject to specific regulations, the respect of privacy is a major issue that cannot

be ignored. The main concern of the authors is that the rules that could be introduced to organize these aspects might eliminate this new Social Science before it is born. They propose instead a self-regulatory regime of procedures and technologies that reduces risk while preserving the research potential. This pragmatic approach fears regulation.

The second article called *Manifesto of computational social science*, published in 2012 by a group of scientists, mainly from European countries [2], is a deep collective reflection about the objectives, challenges and opportunities of this emerging scientific domain. They emphasize the innovative approach that characterizes this way of studying social systems and discuss how this approach constitutes a change of paradigm from the traditional Social Sciences. This article focuses on the epistemological advantages and drawbacks of this new approach to the field.

No matter which of these approaches is selected, it is clear that the possibility of handling very large data bases allows for new approaches in the quantitative social sciences. This is, not only because some results may be obtained with better statistical accuracy, but also and more importantly, because new issues can be addressed. As in biology, where the access to massive data (for example, in genomics) has stimulated the development of areas of research, such as bioinformatics, requiring the use of modeling and computational tools not used before; there is a need in social science for the development of new models and algorithms for dealing with the mass of data coming from ICT systems.

On the modeling side, the search for relevant simple models reflecting stylized facts is a frequent approach in physics, as exemplified by the recent emergence of the new terms, *Econophysics* and *Sociophysics*, coined by H.E. Stanley [3] and Serge Galam [4] respectively, although there are much more ancient contributions from physicists, for example, in quantitative geography. Similar modeling approaches have been developed by social scientists, notably in the 1970s by T. C. Schelling [5] on various topics dealing with collective phenomena, in social and economic contexts. Many questions concerning social systems have been studied using this *theoretical* approach, including opinion and cultural dynamics, language dynamics, the behavior of crowds, spreading phenomena in social networks, and hierarchical structures, amongst others. A review on these topics can be found in [6]. These theoretical models provides us with a better understanding of how a social system might behave when a particular dynamics is considered. They allow the observed social structures, (such as formation of consensus or communities, the segregation of populations with different characteristics, the persistence of cultural heterogeneity in spite of homogenizing rules), to be related to the general rules governing the studied dynamics. However, in most cases, due to the abstract nature of the model, the comparison with real systems remains qualitative.

A different, empirical approach, that has recently seen rapid development, called *phenomenological* in the natural sciences, focuses on the analysis of statistical properties of Social Data in order to identify stylized facts, which are regularities that help to characterize the studied phenomenon. However, dealing with social data is not a minor problem. This is due not only to difficulties generated by data collection and handling, but also to the very definition of the pertinent quantities to be studied. Unlike in the natural sciences, where the variables of interest are clearly defined and the measurement protocol is established, here we face the problem of the choice of those variables. For example, in studying the a diffusion of a culture, how do we define and measure a cultural feature?

However, under some circumstances, the data collected by the ICT devices can be considered as objective data, although the information we can obtain from it may be less clear. In some cases even the "opinion" of the agents may be obtained from the traces left in social networks and ICT

data can be used to infer personal habits, psychological characteristics, social roles (such as being a leader, or a spreader of information) and recently, even the interaction between the on-line and the off-line worlds. For example, a study of Twitter exchanges during the Spanish demonstrations of *Los Indignados* in May 2011 showed the dynamic interplay between on-line contacts and the off line consequences [7].

It is worth noting that this detection of stylized facts cannot be the result of a "brute force" analysis of Big Data; instead an insight into the problem we are studying is necessary in order to target the variables that are the best candidates for revealing regularities present in the sample.

Concerning the theoretical approach to the two specific topics addressed in this project, a large number of models exist.

Opinion dynamics models deal with the way in which a given opinion appears, diffuses and structures into communities of supporters in society. The leading interactions that are commonly considered come from sociological hypotheses, like homophily (like attracts like) and social influence (an agent has a higher probability of being influenced by another that holds a similar opinion). The dynamics studied in different models that integrate these pairwise interactions have taken different forms: imitation processes, local majority rules, global majority rules, influence of mass media, etc.

In network representation, a scalar variable is associated with each vertex in order to represent the agent's opinion on the subject considered, and the links represent the interaction among agents. The early studies on opinion formation considered the extreme situation of binary opinions and different dynamics such as the voter model, majority rule or social impact theory (where each agent is influenced by the majority opinion of its own neighborhood). These models have subsequently been extended to multiple opinion and different network topologies, see [6].

Moreover, there are situations where opinions may actually be of an infinite variety and can be represented by a continuous variable. The *Bounded Confidence model* is of this kind and concerns the situation where the interaction between agents depends on how similar their opinions already are (see Deffuant model in [6]). The intermediate situation has also been studied, where the opinions are continuous but induce discrete actions [8]. Recent works have considered the situation where the opinions of the agents and their social links may evolve in comparable time scales, leading to a coevolving dynamics of links and nodes in the network [9,10]. Given that this project deals with the study of the organizational dynamics of human social action, we believe that a research proposal based on a broad view of systems theory would benefit the project. Furthermore, from the complex systems perspective, there need not be any conflict between the qualitative humanistic study of opinion formation and language evolution and the use of quantitative methods. In discussing, for example, how languages might be considered complex adaptive systems, Beckner et al. state that *An idiolect is emergent from an individuals language use through social interactions with other individuals in the communal language, whereas a communal language is emergent as the result of the interaction of the idiolects (2009, p. 15)* [11].

Language emergence may be studied on different timescales, depending on whether one wants to focus on the evolution of language, traditional processes of language change and diversification, language contact, shorter-term processes such as creolization, or the state of language at a given point in time. For instance, the phylogenetic approach explores how different languages are related making use of an analogy with the phylogeny of natural species, assuming an underlying evolutionary-like mechanism. In the same vein, the distribution of languages has been described as being similar to that of species in ecology [12]. The sociolinguistic approach considers that language changes are due to social interactions and influence, leading for example, to the integration and

diffusion of linguistic variants [14]. Theoretical works have studied the competition with another language [13]. Recently the language dynamics resulting from the use of social media has been studied. In particular, multilayer networks have been introduced to model lexical innovations in order to take explicit account of the coupling of the social network of agents and the *media network* composed of different media that diffuse information to social agents (radio, television, www) [15].

On the other hand the number of phenomenological studies has naturally exploded in recent years, due to the increasing availability of online data bases. Studies of opinion diffusion and formation can easily go beyond the traditional statistical treatment of opinion polls, as is the case of the study of electoral regularities [16–20].

A large number of recent opinion formation studies are based on data issued from online media such as blogs and Twitter, for instance detecting opinion leaders [21], testing the social influence principle [22] or understanding the evolution of the online discussions [23]. Recent work investigated how the transfer of information through the network allows for the emergence of social phenomena at a global scale, in particular showing that a change in the time scale of the information flow indicates the onset of global social phenomena [24].

The same development of empirical studies based on large corpora is also observed in language dynamics. Examples of this development are dialect characterization, using Twitter [25] or the mapping of language geography at different scales, from country level to specific neighbourhoods [26]. A large scale analysis of words in millions of books in seven different languages and covering the last two centuries, using the Google Book data base, showed regularities in the evolution of these languages that could be described in terms of a gas that cools under expansion [27]. Nonetheless, the quality of this database for such a linguistic analysis has been questioned [28]. Grammaticalization, the process by which an item acquires greater grammatical status, has been studied using a modeling approach based on the analysis of a large French corpus [29].

These are a few examples of the many works based on ICT data analysis. In our project we propose to go further by linking both, the theoretical and the phenomenological approaches.

Our team is composed in such a way as to ensure an interplay between the different approaches. The rationale of our work will be as follows: starting with the study of the data, in order to target the important features that can help to characterize the studied phenomenon, we will propose models to represent the data in order to search for patterns (if any) in the relevant features. Then, using the information obtained from the data (inferring information about the interrelation of the social actors, if possible), we will build models of interacting agents with a restricted number of free parameters. Finally we will return to the data in order to check whether the dynamics considered resulted in patterns that are compatible with the observed ones. In parallel, consideration will be given to ethical consequences concerning the criteria of relevance employed in the selection and the analysis of the important social features chosen to build the models.

### 3.3 Description of the data chosen for the project

In this project we have chosen to concentrate on two different data bases in order to study opinion and language dynamics. These are *Twitter* and the *The New York Times* article (NYT) data base. The motivation for this choice lies mainly in their very different characteristics.

The NYT data base (<http://developer.nytimes.com/>) is one of those proposed in the list of repositories included in this call, so no explicit authorization is needed.

This database is particularly interesting because unlike newer online media, like Twitter, it includes data from 1851 up to the present day, enabling dynamical studies that cover a very long

period. We intend to use the *Article Search API* that contains meta-data (including the title, author, date, abstract, and URL of the article). After having contacted the staff responsible for the data, and after a first inspection of the database, we have found that, with some extra technical work (see tasks), full articles can also be obtained.

We will also use the *Community API* of this database, which gives access to the commentaries of the readers. We intend to build the social network of readers in order to study its statistical and dynamical properties, together with the spreading of opinion in this network.

The NYT data base is the digital version of a traditional medium, a journal, where the codes of communication correspond to those of the written language, with a deep reflection prior to publication.

In contrast, we have collected Twitter data (in Spanish), corresponding to the last electoral period in Argentina. This includes 35 million tweets of around 32 millions of users, related to 8 super-nodes (the most prominent politicians) from September 9, 2015, to March 31, 2016. We are able to study the tweets and re-tweets through the followers of the eight super-nodes, as well as the second order followers.

This database is particularly interesting for opinion formation studies, because due to the extreme polarization of the political field in Argentinian society, serious discussions took place, involving an important part of the society. Unlike the NYT database, here the expression of opinions was spontaneous, using spoken style, as it is often the case in Twitter, where new words and concepts are introduced and adopted at extremely high speed.

The possibility of covering different time-scales for the dynamics involved, together with the different constraints in the communication habits, is one of the research interests of our project.

As required in this challenge, in this project we employ multi-form databases to answer fundamental research questions in the social sciences enabling a novel application of this data, different from its initial usage. These questions aim at enlightening the problem of opinion and language dynamics from different perspectives and are related to the research work that has interested different parts of this team.

### 3.4 Research questions, contributions of different teams

Studies concerning the way in which opinions form and diffuse in society often involve complex network representation. In theoretical models, different network structures are considered ad hoc. and it has been shown that the result of a given dynamics may depend on this choice. However, what are the properties of the actual networks where opinion diffuses? The questions concerning the structure of real networks interest several members of this team who have worked on network characterization in socio-economic systems [30–32], the characteristics networks formed in online Knowledge Bases [36], community detection [33–35] and the evolution of discussions in online social media [23]. In this project we propose to study the way in which the structure of the social network of Twitter users compares to that of the users of the NYT commentaries. However this implies answering a previous fundamental question: which network? From the chosen databases we can build directed networks of emitters and followers, directed networks of answers between users, and undirected networks where the links are based on the common occurrence of a word, hashtag, concept, etc. The choice depends on the phenomenon we wish to target and in this respect, the contribution of the Human Sciences specialists in the group will be of great importance.



Once we have characterized the chosen network, the next question concerns the dynamics followed by the opinion diffusion process. This question interests different members of this team who have studied theoretical models of coevolving networks [10, 37] or epidemic spreading in networks [38]. The data bases chosen will enable us to study the evolution of both the network and the opinion of selected users whose network may vary over time. It is then of interest to investigate if this reconfiguration is related to the predominance of a particular opinion.

These general questions will be applied in a case study of particular interest to the Brazilian team. They aim at searching for traces reflecting the historical evolution of the references made to it in one of the largest journals of the U.S.A. One aspect particularly will be considered : once the statistical properties of the network have been determined, is it possible to detect *anomalies* that do not fit the general correlation patterns, using the traces left by users of the New York Times website data concerning Brazilian political news from 1964 to 2016?

Concerning specifically to language dynamics, we know that language trends exist (Sapirs drift), and many recent studies of language evolution have focused on the constraints bearing on these changes, especially since the work undertaken by Kroch(1989) [39], who showed that there seem to be universal constraints, such as the *s-curve*, which predetermine the speed of language change. Recent models of language change (see [40] for an overview) have tried to explain these constraints on theoretical grounds. In previous work, undertaken by members of the team, it was shown that a simple model could explain the characteristics of the S-curve [41], and that the predictions of the model were borne out in real-life testing of French diachronic corpora [42]. The contribution of the group in this project will be to perform a systematic analysis of language changes in the NYT corpus, and to use it to test their model. This corpus is an ideal complement to existing corpora such as the COCA, COHA and the TIME Magazine corpora (<http://corpus.byu.edu/>), because it offers not only raw text data but also the possibility of linking diachronic changes to real-life events. Specifically, it is proposed to track various types of language changes: borrowings (from other languages), lexicalization (newly coined words for new objects), and grammaticalization (shift from lexical to grammatical uses [43]), in order to see how well they fit the developed version of the S-curve model. Additional features will be studied, such as the latency period, observed in the French corpus, and the dynamics after the maximum of the S curve has been reached which most frequently involves a decrease. The initial model developed will be extended to take into account this last stage.

Moreover our two data bases involve very different time-scales, and language styles, enabling determination of whether the language evolution, observed in the two media could be considered to obey to the same dynamical rules.

Finally, in parallel with the preceding studies a reflection will be carried out concerning ethical issues related to: (i) privacy; (ii) the unauthorized handling, for commercial purposes, of digital traces left by users; (iii) the emergence of an ubiquitous computing panopticon that could generalize mind over mind control, affecting the dynamics of self-organized human social actions, and (iv) the possibility of using data correlations to predict and prevent undesirable human social actions related to security and epidemics.

The research questions described above show how the different members of the group would benefit from the synergies created in the team, conducting research considering the different aspects of a given topic. This will require a continuous inter-exchange of ideas and mutual assistance throughout the entire project (see the detailed examples in the working plan). It is clear that such an ambitious research plan could not be implemented by a single partner. We expect the engagement

in the TAP-DD challenge will result in the establishment and consolidation of a research team that will continue to work together beyond the term of this project.

### 3.5 History of the project

This project is the result of the convergence of interest of a group of researchers on both sides of the Atlantic that have been studying similar problems from different points of view

**Jose Ignacio Alvarez-Hamelin and Mariano G.Beiró** have worked on community detection in social and technological networks with a focus on understanding community formation from a Statistical Physics perspective. Some of their recent works include the proposal of a local algorithm for community detection in large networks [35], the characterization of clustering in networks [33], the study of modularity as a Potts-model hamiltonian and its limitations, and the problem of community detection at multiple resolution levels [34].

**Lidia A. Braunstein's** group has intensively investigated epidemic models in isolated networks because they were focused on understanding the propagation of non-recurrent diseases such as influenza and Ebola . The last few years they concentrated on processes that evolves on top of Networks of Networks, such as mitigation and repairing strategies on interdependent networks under cascade of failure epidemic spreading in multilayer networks and competing dynamical processes on two interacting networks [38, 44, 45]. Her group has ongoing collaborations with H.E. Stanley's and S. Havlin's groups.

**Leonardo Ermann** has worked in complex networks, complex systems, quantum chaos and transport phenomena. He has applied different tools developed in quantum and classical chaos to the study of properties of huge matrices built from real networks [30]. In this novel approach Google matrices of real or modeled networks were treated as Markov processes. Topology, dynamics and centrality indices of real networks were analyzed and modeled for the World Trade Network, Wikipedia, Linux Kernel, University webpages, etc. He has a close collaboration with Shepelyanky's group in Toulouse.

**Maria Eunice Quilici Gonzalez** is mainly interested on the interdisciplinary analysis of philosophical, ethical and political issues related to the influence of big data and ubiquitous computing resources into autonomous decision-making processes. She is also interested in the Semiotic and Epistemological foundations of the concept of Information and its role in the direction of individual and collective actions. Her central contribution to the project will be with the elaboration of an interdisciplinary epistemological view of the relation between information and action bridging the social and the natural sciences.

**Mariana Claudia Broens** has been studying the processes of social habits formation, from the ecological theory of information and complex systems perspectives. Her research concerns the investigation of possible ethical consequences of information technologies related to ubiquitous computing and the Internet of Things, in terms of the patterns of human social behaviour. Her central contribution to the project will be to analyze positive and negative ethical implications of the correlations detected by the models developed.

**Jose Artur Quilici-Gonzalez** main contribution to the project will be the analysis of the impact that Big Data is having on the way that scientific research is undertaken. Traditionally, faced with a problem, the researcher formulates a hypothesis and starts experiments in order to confirm or reject specific hypotheses. With the resources of Big Data, researchers can start the work by seeking correlations among available data, and these correlations may be used as a hypothesis.

He claims that the use of Big Data carries the risk of spurious correlations, but the traditional way of doing research presents the risk of "confirmation bias," where the researcher conducts the work based on prior beliefs. Daniel Martinez vila research interest includes the study of critical theories in technological and journalistic environments. He has studied the application and education of critical theories and methods in media knowledge organization on different critical theories and methodologies in the field of knowledge organization. Combining with experience and interest in technologies, critical theories and ethics in information, he intends to contribute to the analysis of the data, including aspects such as ethical concerns in the documentation and policies of the systems, and the political and ideological features of the commentators.

**Anderson Vinicius Romanini** worked for 20 years as a professional journalist both in Brazil and abroad. He witnessed the emergence of Internet based platforms for journalism as well as the changes brought by social media and mobile devices in processes of political representation and deliberation. As a trained semiotician, he has been working on a semiotic theory of communication based on the logical concept of information and the study of communication in complex systems. His main contribution to the project will be to help with the semiotic investigation of a general logic behind purposeful action, embracing also the esthetical and ethical aspects of social interaction.

During the last two decades, the members of the **Brazilian team** have been working (at UNICAMP, UNESP, UFABC and USP) on the topics of self-organization, information, and complex systems. Ethical aspects of these topics have been the subject of published articles by members of our Brazilian team, dealing with the hard problem of ethics, information ethics, possible ethical consequences of ubiquitous computing in social habits formation, and the influence of the Internet of Things (IoT) on the feeling of social trust, among others [46]. Their more focused contributions to the project concern the development of an Evolutionary Ethics approach to the study of the traces left by users of the New York Times website data on Brazilian political news from 1964 to 2016. A fundamental presupposition of Evolutionary Ethics, considered from the Complex Systems perspective, is that moral habits are emergent properties of social affordances embodied in human social interactions. Social affordances are dispositional collective properties that indicate possibilities of action provided to organisms by other organisms that share co-evolutionary histories. From this perspective, we will examine the main properties of the networks (selected from the Big Data analysis undertaken by the Argentinean and French research teams) of common words and concepts used by the emitters and followers of Brazilian political news.

**Laura Hernández** is interested in theoretical and data based modeling of social and natural systems. She has worked on Cultural and Opinion Dynamics [10,37] models, and in the characterization of the network's structure describing mutualistic (social and natural) systems [31,32,48]. She has studied local elections on many different countries, showing a universal population-size effect, related to the fact that the decision made at a local election (unlike in national elections) only concern the local deciders as a group [20].

**Dimitrios Kotzinos**, is interested in studying the evolution of networks and more specifically online personal networks both in terms of network structure and information diffusion. He is also interesting in identifying inner social circles inside broader communities that can be detected in online social networks, i.e. social circles that might correspond to the actual social behavior of people, for example corresponding to family, work or education ties. He has also worked in extracting characteristics for networks used to describe formal knowledge [36] and in capturing, understanding and visualizing the dynamics of discussions in social media like twitter [23].

**Jean-Pierre Nadal** is working in the field of complex systems in social sciences, with tools

coming from statistical physics. He has been and is working on the analysis and modeling of collective behaviours when individual choice depends on the ones of others. Some specific topics related to the project are opinion dynamics, language change (in collaboration with B Fagard), riot contagion and other social urban dynamics. Through interdisciplinary collaborations, he performs both theoretical and data driven modeling studies.

**Benjamin Fagard and Thierry Poibeau:** The Lattice laboratory has worked on language emergence and evolution for quite some time, with a rich interdisciplinary approach combining mathematical modes (developed by Bernard Victorri), computer-assisted models (Thierry Poibeau and PhD students supervised by him) and diachronic corpus studies (Benjamin Fagard), with multiple collaborations (e.g. with the Institute of complex systems, and the Sony laboratory in Paris). This search for a multi-disciplinary approach to language change explains our collaboration with Jean-Pierre Nadal (LPS laboratory), and our inclusion in this project.

This project is supported by the *Institute of Complex Systems of Paris and Ile de France* (ISCPif), in particular by providing access to their platforms of modeling and data representation (OpenMole, and Gargantex). The latter should prove very useful for the first tasks of this project because it allows to obtain a useful visual representation of a corpus.

It is also supported by the Complex Systems Digital Campus (CS-DC), an UNESCO UniTwin that proposes a collaborative platform for the diffusion of research and teaching in Complex Systems. The CS-DC has already provided a virtual room for our preparatory meetings, and allows us to largely diffuse our results to the Complex Systems community. It is also an excellent tool to organize international seminars which help to easily bring together students and young researchers of the three countries for whom travel expenses are difficult to afford.

### 3.6 Teaching through research

We intend to take profit of the interdisciplinary character of our team to propose two different approaches of student training.

On the one hand, the french team, whose local agency authorizes the recruitment of temporary staff, is not only proposing to hire young researchers (in the form of 2 post-doctoral fellowships) but also devotes part of its budget to the training of six master students through internships in their laboratories. This potentially involves students on different masters, *Theoretical Physics and Applications* and *Master in Communicating and Intelligent Systems* at (U.C.P.), *Physics* (ENS), *Computer linguistics (Plurital, Paris 3, 10 and Inalco)*, *Language sciences (Paris 7)*.

On the other, the technical platforms of CS-DC, to which we have access, allow us to program on line courses (live lectures and not MOOCS) where instructors of the three teams could participate to bring a multidisciplinary training at the level of Master 2 or PhD.

Concerning the training of young researchers, two PhD students (already financed), one at LPTM and another at Lattice, will directly be involved in this project. Moreover a substantial part of the demanded budget is devoted to the financial support to recruit three Post-Docs to work with us (one at Sao-Paulo, one at LPTM and another at Lattice).

## References

- [1] David Lazer, et al., Science, **323**, 721 (2009).
- [2] R. Conte, N.Gilbert, G. Bonelli, C. Cioffi-Revilla, G. Deffuant, J. Kertesz, V. Loreto, S. Moat, J-P. Nadal, A. Sanchez, A. Nowak, A. Flache, M. San Miguel, and D. Hebling. Eur. Phys. J. Special Topics **214**, 325 (2012).
- [3] H.E.Stanley and Rosario N. Mantegna.*An Introduction to Econophysics: Correlations and Complexity in Finance*, Cambridge Universtiy Press, (2000).
- [4] Serge Galam, *Sociophysics : a personal testimony*. Physica A336, 49 (2004)(and references therein).
- [5] Schelling, Thomas C. *Micromotives and Macrobehavior*, Norton & Cy, 1978.
- [6] Claudio Castellano, et al.,Rev. Mod. Phys. **81**, 591 (2009).
- [7] Javier Borge-Holthoefer et al., PLoS ONE 6(8): e23883. doi: 10.1371/journal.pone.0023883 (2011)
- [8] André C. R. Martins. Int. J. of Mod. Phys. C, **19** 617624, (2008).
- [9] P. Holme and M. E. J. Newman,Phys. Rev. E**74** 056108 (2006).
- [10] E. Burgos, L. Hernández, Ho. Ceva, R.P.J. Perazzo, Phys. Rev. E **91**, 032808 (2015).
- [11] Beckner, C.; Blythe, R.; Bybee, J.; Christiansen, M. H.; Croft, W.; Ellis, N. C.; Holland, J.; Ke, J.; Larsen-Freeman, D.; Scheonemann, T. 2009. Language Learning. vol. 59. supplement 1. pp.1-26.
- [12] Sutherland, William, J., Nature **423**, 276279 (2003).
- [13] Abrams, Daniel, and Steven H. Strogatz, Nature **424**, 900 (2003).
- [14] Ke Jinyun, et al.,Communications in Computational Physics 3.4, 935949 (2008).
- [15] International Journal of Modern Physics C **25**,10 (2013).
- [16] S. Fortunato, C. Castellano, Phys. Rev. Lett. **99**, 138701 (2007).
- [17] N. A. M. Araújo, et al. PLoS ONE **5**(9), e12446 (2010).
- [18] C. Borghesi, J. Chiche, J.-P. Nadal, PLoS ONE **7**(7), e39916 (2012).
- [19] M. C. Mantovani, et al. Phys. Rev. E **88**, 024802, (2013).
- [20] C. Borghesi, L. Hernández, R. Louf, F. Caparros, Phys. Rev. E **88**, 062813, (2013).
- [21] Feng Li, Timon C. Du, Decision Support Systems **51** 190197,(2011).
- [22] Sinan Aral, Dylan Walker, Science **337**, Issue 6092, pp. 337-341 (2012)
- [23] Kleisarchaki S., Kotzinos D., Tsamardinos I., Christophides V. (2013), *A Methodological Framework for Statistical Analysis of Social Text Stream*, Springer Series Communications in Computer and Information Science Volume 146, 2013, pp 101-110.
- [24] J. Borge-Holthoefer, et al.,Science Advances Vol. 2, no. 4, e1501158 (2016)
- [25] B. Goncalves, D. Sanchez, PLOS ONE, **9** e112074 (2014).

- [26] Delia Mocanu et al., PLoS ONE 8(4): e61981. doi:10.1371/journal.pone.0061981, (2013).
- [27] Alexander M. Petersen<sup>1</sup>, et al. Sci. Rep. DOI: 10.1038/srep00943 (2012).
- [28] E A Pechenick, et al. PLoS ONE 10(10): e0137041 (2015)
- [29] Q. Feltgen, B. Fagard, J.-P. Nadal , Chapter 3 in La Mantia F., Licata I., Perconti P. (Eds.), *Language in Complexity: The Emerging Meaning*, Lect. Notes in Morphogenesis, Springer, 2016
- [30] L. Ermann, K.M. Frahm, D.L. Shepelyansky, Rev. Mod. Phys. **87** (4), 1261 (2015); L. Ermann, D.L. Shepelyansky, Eur. Phys. J. B **75** (3) 299-304 (2010).
- [31] E.Burgos, H.Ceva, L.Hernández, R.P.J.Perazzo, M.Devoto, D.Medan, Phys.Rev. E **78**, 046113, (2008)
- [32] E.Burgos, H.Ceva, L. Hernndez, R.P.J. Perazzo. Comp. Phys. Comm. 180, p532 (2009).
- [33] P.Colomer-de-Simón, M.A. Serrano, M.G.Beiró, J.I.Alvarez-Hamelin, M.Boguñá, Sci. Rep. **3**, (2013).
- [34] Beiró, M. G., Grynberg, S. P. and Alvarez-Hamelin, J-I., EPJ Data Science 4:12, (2015).
- [35] M.G Beiró, J.R. Busch, S.P. Grynberg, J.I. Alvarez-Hamelin, Physica A **392**, 2278 (2013).
- [36] Theoharis, Y., Tzitzikas Y., Christophides V, and Kotzinos D. (2008) *On Graph Features of Semantic Web Schemas*, IEEE Transactions on Knowledge and Data Engineering, 07 Dec 2007. IEEE Computer Society Digital Library. IEEE Computer Society, 20 January 2008.
- [37] C.Gracia-Lázaro, F.Quijandría, L.Hernández, L.M.Floría, Y.Moreno, Phys. Rev. E **84**, 067101 (2011).
- [38] L.G.A. Zuzek, H.E. Stanley, and L.A. Braunstein, Nature Scientific Reports 5, 12151 (2015).
- [39] Kroch, A. S. 1, 199244,(1989).
- [40] Vogt, P., and De Boer, B. Adaptive Behav.: Anim., Animats, Soft. Ag., R., Ad. Syst.,**18**, 511 (2010).
- [41] Feltgen, Q., B. Fagard and J.-P. Nadal. Traitement automatique des langues **55**(3),(2014).
- [42] Feltgen, Q., B. Fagard and J.-P. Nadal. In F. La Mantia et al. (eds), *Language in Complexity*. New York: Springer. (to appear in 2016).
- [43] Hopper, P. and E. Traugott. *Grammaticalization*. Cambridge: Cambridge University Press, (2003).
- [44] C. Buono, L. Alvarez-Zuzek, P. A. Macri and L. A. Braunstein, PLoSone **9**, e92200, 2014.
- [45] C. Buono and L.A. Braunstein, Europhys. Lett. (EPL) 109, 26001 (2015).
- [46] Quilici-Gonzalez, J. A. , Broens, M. C., Gonzalez, M. E. Q., Kobayashi, G., Scientiae Studia (USP), v. 12, p. 161-179, 2014.
- [47] S. González-Bailón, et al., Scientific Reports**1**, 197 (2011).
- [48] R.P.J.Perazzo, L.Hernández, H.Ceva, E.Burgos, J.I.Alvarez-Hamelin, Physica A **394** p124135 (2014).

## 5 Data Management Plan

### 5.1 Big Data Processing

The data processing pipeline will be centered around Data Analysis languages and tools (e.g. Python) and parallel-processing cloud based architectures (like e.g. Spark or Hadoop), using one or more of the cloud based data stores either relational or NoSQL or a mixture of those (like e.g. the HDF5 data store, MongoDB or PostgreSQL databases). This processing will be done in high-performance servers available in one of the countries, and accessible through SSH for all the researchers. This will provide a secure access to data and allow cooperative work at the same time.

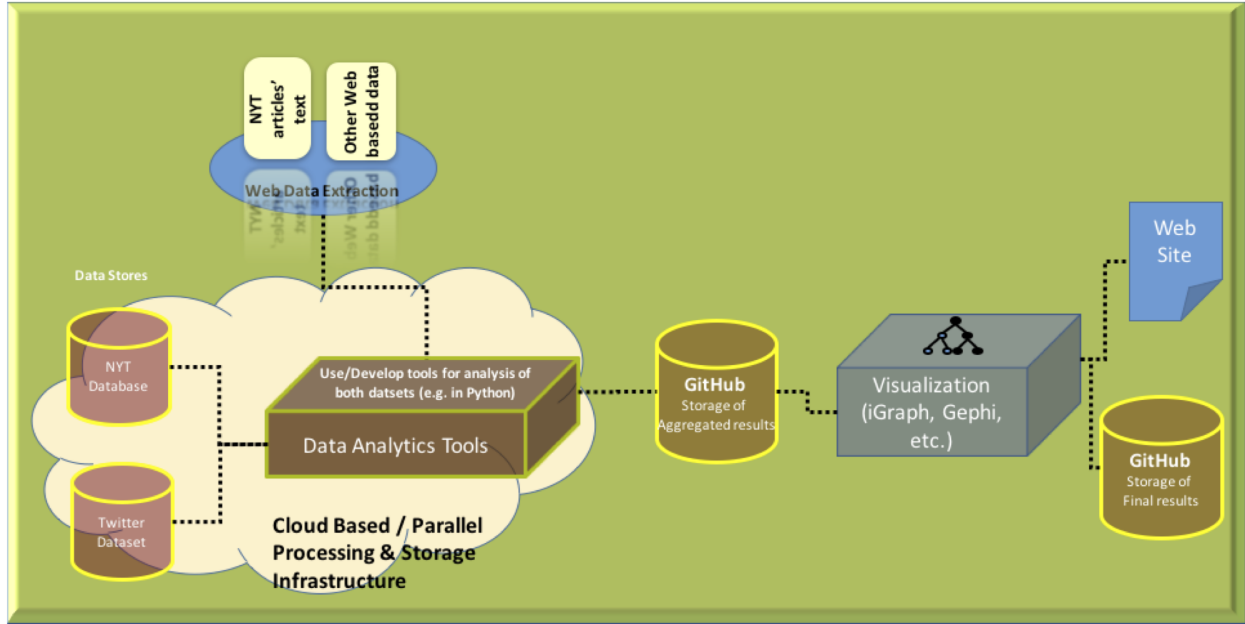


Figure 1: The project's big data processing pipeline.

For network analysis and visualization more specific software will be used, including the igraph library and Gephi. We need to visualize data in order to understand better the online communities that might be formed and also the flow of information among them or the individuals participating in them. Besides these we plan to exploit various plot generation techniques to visualize the results of the calculation of various properties of the networks.

The documentation of the data workflow will be assisted by the use of the UML modeling language with which we plan to model all the data processing activities including the storage of any intermediate data. This practice will allow the communication and discussion through the data workflow among the specialists from the different areas and will make the architecture and the code easily reusable in the future.

### 5.2 Big Data Storage

All the original data will be stored as needed internally and we will also try to exploit the replication capabilities offered in the high end cloud based servers that will be used. Besides that we will

establish (to the extend necessary) a system for storage preservation to account for hard disk and other technical failures, so that we can make sure that we do not need to retrieve and process the same data again. This will include all intermediate (aggregated) data sets and networks that will be obtained from the raw data processing; this data will be tracked and shared through GitHub among all the researchers since they do not carry the same publication restrictions as the original ones.

Finally, and in order to better organize the information infrastructure we plan to provide links to the available data sets through the project's website, where we plan to centralize all the produced information. This website will point at the GitHub platform and will be pointed by the CS-DC.

### 5.3 Data Licensing

As the original data sets used in this project (obtained through NYT API and Twitter API) do not allow for redistribution, the original data will not be made open publicly. However, the results of our project in terms of software, algorithms and publications will be diffused under one of the Open Source Licenses (for the software) and one of the Creative Commons Licenses (for documentation and publications) in the website of the CS-DC UNESCO UniTwin, Wikiverstiy portail, where members of our team have constituted an e-Laboratory. (<https://en.wikiversity.org/wiki/Portal:ComplexsystemsDigitalCampus/E-Laboratoryonsocialself-organization>). In addition, to simplify their access and share the produced software we will use a specialized collaborative platform like GitHub, which might also serve as a versioning system.



## 6 Project Management, Dissemination, and Communications Plan

The management of this project is guaranteed by the fact that several members of different laboratories have already worked together in different projects, as stated in the "Narrative", and also because the International Coordinator (L.H.), who is the one having personal contacts with all the teams, may act as a bridge between those that have not collaborated yet.

Each represented country has a Principal Investigator that coordinates the activities of the corresponding group of researchers. As each participant country finances its own laboratories, with different rules, the laboratories of each country have decided together of their method to manage the demanded budget. However, in order to maintain fluid exchanges among the members we have planned that, besides the virtual meetings that the team (or parts of it) may regularly hold, using the virtual rooms of the CS-DC UniTwin, at least, one meeting of the whole team per year, is necessary to discuss, evaluate, correct the the project all along its evolution. This means one meeting in each one of the countries. During these meetings, mid-term evaluation and key decisions on the orientation and advancement of the project will be discussed. The related travel expenses are taken into account in the corresponding budgets. Moreover the final conference organized by the funders at the end of the grant period, has been included in the local budgets.

### 6.1 Roles and responsibilities

#### Argentina

- **José Ignacio Alvarez-Hamelin** Role : PI Affiliation: Facultad de Ingeniería, Universidad de Buenos Aires. Grade: Professor. Contact: ihameli@cnet.fi.uba.ar ; dedication : 30% (10 months/ period).
- **Mariano G.Beiró** Role : participant Affiliation: Facultad de Ingeniería, Universidad de Buenos Aires. Grade : Assistant Professor. Contact: mbeiro@cnet.fi.uba.ar; dedication : 20% (6,5 months/ period).
- **Leonardo Ermann** Role : participant Affiliation: Theoretical Physics Dept. TANDAR, CNEA. Grade: Associate. Researcher CNEA and CONICET. Contact: ermann@tandar.cnea.gov.ar; dedication: 15% (5 months/ period).
- **Lidia A. Braunstein** Role : participant. Affiliation: IFIMAR , Universidad de Mar del Plata, CONICET. Grade : Researcher, IFIMAR. Contact: lbrauns@mdp.edu.ar; dedication : 16.6% (6 months/ period).

#### Brazil

- **Maria Eunice Quilici Gonzalez** Role : PI. Affiliation: Faculty of Philosophy and Sciences - University of Sao Paulo State - UNESP. Grade: Associate Professor. Contact:gonzalezquilici@gmail.com; dedication : 20% (6,5 / period).
- **Mariana C. Broens** Role : Participant. Affiliation: Faculty of Philosophy and Sciences - University of Sao Paulo State - UNESP. Grade : Associate Professor. Contact: mariana.broens@gmail.com; dedication : 15% (5 months/ period).

- **Daniel Martinez Ávila** Role : Participant. Affiliation: Faculty of Philosophy and Sciences - University of Sao Paulo State - UNESP. Grade: Assistant Professor. Contact: dmartinezavila@gmail; dedication : 10% (3 months / period).
- **Anderson Vinicius Rimanini** Role : Participant. Affiliation: School of Communication and Arts - University of Sao Paulo - USP. Grade: Assistant Professor. Contact: viniroma@gmail.com; dedication : 10% (3 months / period).
- **Jose Artur Quilici Gonzalez** Role : Participant. Affiliation: Center of Mathematics, Computing and Cognition - Federal University of ABC - UFABC. Grade: Assistant Professor. Contact: jose.gonzalez@ufabc.edu.br; dedication : 10% (3 months / period).

### France

- **Laura Hernández** Role : PI, Coordinator. Affiliation: Laboratoire de Physique Thorique et Modlisation (LPTM) UMR8089 CNRS-Universit de Cergy-Pontoise. Grade : Associate Professor (MCF-HDR). Contact: Laura.Hernandez@u-cergy.fr, dedication : 30% (10 months / period).
- **Dimitrios Kotzinos** Role : Participant. Affiliation: Laboratoire Equipes Traitement de l'Information et Systmes (ETIS) UMR 8051, CNRS- ENSEA, Universit de Cergy-Pontoise. Grade : Professor. Contact: Dimitris.Kotzinos@u-cergy.fr; dedication : 10% (4 months / period).
- **Jean-Pierre Nadal** Role : Participant. Affiliation: Laboratoire de Physique Statistique (LPS, UMR CNRS 8550), Ecole Normale Supérieure, PSL Research University, Paris. Position: CNRS (Senior Research – DR1) and EHESS (Director of studies). Contact: nadal@lps.ens.fr; dedication : 5% (1.5 months / period).
- **Thierry Poibeau** Role : Participant. Affiliation: Laboratoire Langues, Textes, Traitements Informatiques et Cognition (LATTICE, CNRS, Ecole Normale Supérieure, and University of Paris 3, PSL Research University, USPC). Grade : Senior Researcher (DR2). Contact: thierry.poibeau@ens.fr; dedication : 5% (1.5 months / period).
- **Benjamin Fagard** Role : Participant. Affiliation: Laboratoire Langues, Textes, Traitements Informatiques et Cognition (LATTICE, CNRS, Ecole Normale Supérieure, and University of Paris 3, PSL Research University, USPC). Grade : Researcher (CR1). Contact: benjamin.fagard@ens.fr; dedication : 10% (3 months / period).

## 6.2 Collaborations

This project comes out as a natural aggregation of colleagues on both sides of the Atlantic, who have already worked together in different projects, with the project coordinator (L.H.) situated at the crossroad of the three teams.

Some members of the present project are co-authors of scientific publications (M.Beiró and J-I, L.H. and J-I. A-H, M.Broens and M-E. Q-G., B.F., T.P. and J-P.N.) in the cited bibliography, members of the project are underlined. More broadly there is a long history of interactions between the Latin-American and French participants. For instance L Ermann coorganized in 2012 an interdisciplinary workshop in Buenos-Aires with JP Nadal among the invited speakers.

Laura Hernández, Maria-Eunice Quilici-Gonzalez et Mariana C. Broens, have participated together to the creation and development of the Complex Systems Digital Campus (CS-DC), an UNESCO UniTwin that proposes a platform for the diffusion of research and teaching in Complex Systems. Thus the present project will benefit from the different tools provided by the CS-DC: broadcasting and registering visio-conferences, up to 150 participants, useful for organization meetings, research discussions and remote teaching purposes. These tools have been used for the preparation of the present project.

Jean-Pierre Nadal was coordinator, and Laura Hernández participant, of the French component of the European Flagship project FuturICT. They have together coordinated a French research network, HumanICT, devoted to the modeling and simulation of large techno-social systems supported by the *Réseau National des Systèmes Complexes of France*. They were both members of a research project KynDyMo (Kinetic and dynamic modeling of markets and opinion) supported by the CNRS (two "PEPS" projects); they have co-organized an international workshop *Complexity in Social Systems: from Data to Models*, held at Cergy-Pontoise University in June 2013, and they are both members of the Steering Committee of the *Complex Systems Institute of Paris and Ile de France Region* (ISCPif, <http://iscpif.fr/about-isc-pif/steering-committee/>).

The ENS participants (from LPS and Lattice) have been collaborating for a few years on data driven modeling of language change, with funding for a PhD student (Quentin Feltgen). This project has already lead to one publication (Feltgen et al. 2016, to appear).

### 6.3 Work plan

Roughly speaking, one can say that the aim of this project is to study communication dynamics in both NYT and Twitter data bases. By *communication* we mean both the language as the vector of communication and the opinion that is transmitted through it.

This project may be structured into the following tasks :

#### 1. Experimental set-up

- (a) Choosing the pertinent data to be retrieved from our data sources. It involves the whole group, as it requires on one hand, information from disciplinary specialists and modelers, and on the other, the expertise of Data Scientists. This task includes a bibliographic work that will help to choose our particular case studies. It will determine the type(s) of networks that will be used to represent our system. For example, should we target some special language item (words, groups of words, concepts), a special period or person? This may lead to directed networks (who is talking to who), to bipartite networks, (who is talking about what), to co-occurrence networks (who are talking about the same subject). We plan to measure its success by: the number of possible networks to be used in an interdisciplinary way, covering the skills and interests of the group. The main difficulty lies on the interdisciplinarity and the associated communication problems. But this is part of the work and we have already some training in doing so, due to previous collaborations. (ALL)
- (b) Set-up of a project-specific NYT query interface. Programming and testing of specific algorithms to retrieve data from NYT article, community and semantic APIs, in particular to search articles from the metadata information. The success will be measured by the amount of gathered data in relation to the networks defined in task 1a and finishing

it in time. The major risk is to take more time than it is planned. Strategy: start as soon as some queries are defined, this will help to evaluate the difficulty of the massive data retrieval (in particular for the articles) (D.K., J-I.A-H., M.B.).

- (c) Data Collection for statistical analysis (sampling over short time periods) (D.K., J-I.A-H., M.B.).
- (d) Data collection for dynamical analysis (sampling over long time period) (D.K., J-I.A-H., M.B.).

Tasks 1.c and 1.d depend on 1.b, so they suffer from the same risks. To minimize them we start 1.b as soon as possible

## 2. Data analysis

- (a) Statistical analysis, large scale network structure: This involves the build up of networks from recovered data, and their characterization. Success means that we have characterized the networks. Risk: we do not find interesting stylized facts. Solution: this is typical of Complex Systems, and means that the chosen variables were not the pertinent ones. Strategy: change the network construction (D.K., J-I.A-H., M.B., L.H., L.E.)
- (b) Community detection: This involves to choose (or develop) appropriate community detection algorithms. Here, risks are related to the richness of the real data, where finding communities is a hard task. Its success will be characterized by the isolation of certain phenomena instead of a complete description. (D.K., J-I.A-H., M.B., L.H., L.E.)
- (c) Study networks' dynamics: this involves the analysis of both networks' dynamics and dynamical processes on networks. We plan to study the evolution of networks corresponding to different time slots and the time correlation patterns between network evolution and opinion evolution (by the observation of the opinion evolution on some chosen subject). Risks: associated to task 1d. Given the nature of the data sets the risk of not being able to detect the dynamics is minimal, however data sets can be augmented (e.g. if needed, we are prepared to download the full text for the NYT articles to better capture language dynamics).(D.K., J-I.A-H., M.B., L.H., L.E., J-P.N., B.F.,T.P.)
- (d) Analysis of semantic networks. Correlated with tasks 1a and 1b, the main risk is related to the way to handle the size of the networks and the scope the analysis. Success measured by the degree of novelty introduced here, which is quite probable since we are planning to explore new network definitions based on data. (M-E.Q-G, M.C.B., A.V.R., D.M.A., L.E., J.A.Q.G.)
- (e) Interpretation and comparison of results issued from both data bases for both problems. Correlated to tasks 2b, 2c and 2d. The risk here would be to have a *negative* result. Given the differences in the data bases, if task 2 gave interesting results, there is low probability for this. (ALL)

## 3. Modeling dynamics of opinion and language evolution

- (a) Choice of the models, programming and tests. Risk: As usual dealing with data based models, there is a compromise between realism and the number of parameters that make a model explanatory. This is part of the research work, and the modelers of the team are used to it. (ALL)

- (b) Calculations and interpretation of results Risk: the size scaling of the models that may increase computing time. Strategy: thanks to the ISCpif, we have access to the Grid. (L.H., L.E., L.B. JPN, B.F.)
4. **Reflections on ethical aspects of the problem** The choice of criteria of relevance in the selection of targeted social features, and its consequences on social behavior. (developed along all the project) (M-E.Q-G., M.C.B, A.V.R., D.M.A, J.A.Q.G.)
5. **Diffusion of results** Starting from the end of task 2a until the end. There are no risks associated by this task since the consortium members have a track record that proves that they can successfully disseminate scientific results and make them available to the greater scientific community. (ALL)

## 6.4 Dissemination and communication

A special attention will be paid to the communication of our results. The specific software concerning data retrieval and analysis that will be developed, will be published as open source under one of the main open source licenses. As mentioned before, we will use the facilities of the Unitwin CS-DC to diffuse our results to the Complex Systems community.

Moreover we will publish our scientific results mainly in open access journals in order to facilitate the broadest possible diffusion (part of the demanded budget is devoted to this). We also plan to participate to relevant high-end conferences in the field so that we can disseminate the results faster to the broader scientific community in the area. Given the multidisciplinary nature of the project and the team we envision that we can reach multiple scientific communities, extending the reach of our results.

As the project is being developed, we intend to organize a cycle of online seminars in order to enlarge the discussion group to other researchers working in similar or related programs. The seminars will also be registered and stored in the CS-DC system.

Finally, we will create a website for our project which will centralize all the information around it, will contain the related publications or links to them and will have links to the generated data sets as described earlier. This website will point at the GitHub platform for the possible Open Source software production, and will be also pointed by the CS-DC.

|     |         |     |     |     |     |     |     |     |     |     |    |
|-----|---------|-----|-----|-----|-----|-----|-----|-----|-----|-----|----|
| 3m  | 3m      | 3m  | 3m  | 3m  | 3m  | 3m  | 3m  | 3m  | 3m  | 3m  | 3m |
| 1.a | 1.a/1.b | 1.b | 1.c |     |     |     |     |     |     |     |    |
|     | 4       |     |     |     |     |     |     |     |     |     |    |
|     |         |     |     | 2.a | 2.b |     |     |     |     |     |    |
|     |         |     |     |     | 5   |     |     |     |     |     |    |
|     |         |     |     |     | 2.d |     |     |     |     |     |    |
|     |         |     |     |     | 1.d |     |     |     |     |     |    |
|     |         |     |     |     | 2.e |     |     |     |     |     |    |
|     |         |     |     |     |     | 2.c |     |     |     |     |    |
|     |         |     |     |     |     |     | 3.a | 3.a | 3.b | 3.b |    |

Figure 2: Working plan time-line. Each color corresponds to one of the tasks listed in the text.

Budget Summary

| Country   | Total Budget (in €) |
|-----------|---------------------|
| Argentina | 39942               |
| Brazil    | 141002              |
| France    | 179712              |
| Total     | 360656              |



**Applicant Institution:** Universidad de Buenos Aires (UBATEC)  
**Project Director:** Laura Hernández (Coordinator), J.I.Alvarez-Hamelin (PI-Argentina)  
**Project Grant Period:** 2017-2019

|  | Computational<br>Details/Notes                 | (notes) | Year 1      | (notes) | Year 2      | (notes) | Year 3      | Project Total |
|--|--|---------|-------------|---------|-------------|---------|-------------|---------------|
| <b>1. Travel</b>   |  |         |             |         |             |         |             |               |
|  | 3 (flight + 13 per-diem)                       | euros   | <b>9390</b> | euros   | <b>9390</b> | euros   | <b>9390</b> | <b>28170</b>  |
| <b>2. Supplies &amp; Materials</b>   |  |         |             |         |             |         |             |               |
|  | Computers for processing and storage for data. | euros   | 9870        |         |             |         |             | <b>9870</b>   |
| <b>3. Services</b>   |  |         |             |         |             |         |             |               |
|  |  |         |             |         |             |         |             | <b>0</b>      |
| <b>4. Total Indirect Costs</b>   | <b>Per Year</b>                                | euros   | 963         | euros   | 469.5       | euros   | 469.5       | <b>1902</b>   |
| <b>Indirect Cost Calculation: just for cost administration</b>                         | 5.00%  |         |             |         |             |         |             |               |
| <b>5. Total Project Costs Requested (Direct and Indirect costs for entire project)</b> |  |         |             |         |             |         |             | <b>39942</b>  |
|  |  |         |             |         |             |         |             |               |
|  |  |         |             |         |             |         |             |               |

|  | Computational<br>Details/Notes | (notes) | Year 1 | (notes) | Year 2 | (notes) | Year 3 | Project Total |
|--|--------------------------------|---------|--------|---------|--------|---------|--------|---------------|
|--|--------------------------------|---------|--------|---------|--------|---------|--------|---------------|

Currency in euros

No subcontracting cost will be covered

We propose UBATEC as UVT



## RESEARCH PROPOSAL FORM

PROTOCOL

**Trans-Atlantic Platform (T-AP) Digging into Data / 2016  
REGULAR RESEARCH AWARD****1. PRINCIPAL INVESTIGATOR - FAPESP (do not omit or abbreviate names)**

NAME: Maria Eunice Quilici Gonzalez

**2. HOST INSTITUTION (SÃO PAULO)**

INSTITUTION (University): São Paulo State University - UNESP

UNIT (College, Center, Institute): Faculty of Philosophy and Sciences

DEPARTMENT: Philosophy

**3. PRINCIPAL INVESTIGATOR – (do not omit or abbreviate names)**

NAME: Laura Hernandez

INSTITUTION (University): Université de Cergy-Pontoise

UNIT (College, Center, Institute): Laboratoire de Physique Théorique et Modélisation (LPTM)

DEPARTMENT: Physics

AGENCY: ANR - Agence Nationale de la Recherche

**4. PRINCIPAL INVESTIGATOR – (do not omit or abbreviate names)**

NAME: José Ignacio Alvarez-Hamelin

INSTITUTION (University): Universidad de Buenos Aires - UBA

UNIT (College, Center, Institute): Faculty of Engineering

DEPARTMENT: Department of Electronics

AGENCY: Ministry of Science, Technology and Productive Innovation - MINCYT

**5. PRINCIPAL INVESTIGATOR – (do not omit or abbreviate names)**

NAME:

INSTITUTION (University):

UNIT (College, Center, Institute):

DEPARTMENT:

AGENCY:

**6. PRINCIPAL INVESTIGATOR – (do not omit or abbreviate names)**

|                                    |
|------------------------------------|
| NAME:                              |
| INSTITUTION (University):          |
| UNIT (College, Center, Institute): |
| DEPARTMENT:                        |
| AGENCY:                            |

**7. PRINCIPAL INVESTIGATOR – (do not omit or abbreviate names)**

|                                    |
|------------------------------------|
| NAME:                              |
| INSTITUTION (University):          |
| UNIT (College, Center, Institute): |
| DEPARTMENT:                        |
| AGENCY:                            |

**8. PROJECT TITLE (do not abbreviate)**

|  |
|--|
| Compreendendo a dinâmica da opinião e da linguagem utilizando Big Data |
| Understanding opinion and language dynamics using massive data         |

**9. PROJECT TYPE (according to FAPESP's list) [www.fapesp.br/2000](http://www.fapesp.br/2000)****PROJECT DURATION**

|  |   |   |   |   |   |   |                              |
|--|---|---|---|---|---|---|------------------------------|
| AREA OF EXPERTISE: <b>Philosophy</b>   |   |   |   |   |   |   |                              |
| PROPOSED START DATE: 01/06/2017  |   |   |   |   |   |   |                              |
| CODE: <table border="1" style="display: inline-table; vertical-align: middle;"> <tr> <td>7</td><td>0</td><td>1</td><td>0</td><td>5</td> </tr> </table> - 00 - <table border="1" style="display: inline-table; vertical-align: middle;"> <tr> <td>6</td> </tr> </table> | 7 | 0 | 1 | 0 | 5 | 6 | DURATION (MONTHS): <b>36</b> |
| 7  | 0 | 1 | 0 | 5 |   |   |                              |
| 6  |   |   |   |   |   |   |                              |

**10. RESEARCH PROJECT SUMMARY**

A sociedade humana está passando por mudanças sem precedentes decorrentes do rápido desenvolvimento de tecnologias de comunicação, as quais registram traços de um significativo número de atividades cotidianas de seus usuários. Tais mudanças, extremamente rápidas, propiciam uma dinâmica adaptativa que altera atividades do dia-a-dia e afeta a natureza das relações sociais humanas. O crescente número de dados gerou o que é conhecido como Big Data, suscitando questões sobre seu armazenamento, organização, recuperação e controle. O interesse nesses dados tem até agora se concentrado em aspectos técnicos que possibilitem sua rápida utilização. Em uma perspectiva diferente, o presente projeto tem como objetivo investigar as seguintes questões: Como selecionar informação relevante, a partir da análise de Big Data, para elaborar modelos da ação humana coletiva? Quais são as possíveis consequências éticas do uso da análise de Big Data na ação humana auto-organizada? Investigaremos traços da conduta social humana registrados em bases de dados, questionando possíveis impactos da manipulação de Big Data na ação social. Explicações causais da ação humana, individual/coletiva, são alvo de críticas de pesquisadores que consideram seus aspectos sociais e éticos. A hipótese a ser investigada é a de que o estudo de conexões entre informação disponível nos Big Data, que indica uma quantidade massiva de correlações entre os dados coletados, e as explanações causais da ação podem propiciar novidades promissoras na compreensão da dinâmica da ação social humana na perspectiva dos sistemas complexos. A pesquisa proposta focalizará o estudo da dinâmica da formação de opinião, bem como da evolução da linguagem que pode afetar tal dinâmica, a partir da análise dos Big Data resultantes de duas diferentes mídias: The New York Times e o Twitter.

Human society has undergone unprecedented changes, driven by the sudden increase in communicating technological devices that surround us, keeping traces of a large amount of our daily activities. This extremely rapid evolution is associated with a fast adaptive dynamics that induces changes in our everyday practices and has consequences for the nature of human social relations. The rapidly increasing amount of data has generated what is known as "Big Data", leading to urgent questions concerning the storage, organization, retrieval, and control of this information. Efforts have been concentrated on the technical aspects of these questions, in order to make Big Data ready to be used. From a different perspective, the present project addresses the following questions: How can sensible information be obtained from Big Data and used to help in elaborating explanatory models of human social actions in different specific circumstances? What are the possible ethical consequences of the application of Big Data analyses in the study of self-organized human actions? We are interested in both searching for traces of social behaviour in rough data, and studying the impact that manipulation of Big Data might have on social action. Causal explanations of human individual/collective actions have been the object of criticisms by researchers concerned with the social and ethical aspects. It is our hypothesis that the connection between Big Data, which provides massive amounts of data correlations, and causal explanations might bring about promising novelty in the study of human social actions from the Complex Systems perspective. Our research will focus on the study of opinion dynamics and the language evolution process that might affect this dynamics, based, as a study case, on Big Data issued from two different media: The New York Times and Twitter.

**11. KEY WORDS (up to six)**

|                  |                    |
|------------------|--------------------|
| Big data         | Opinion formation  |
| Complex networks | Self-organization  |
| Social action    | Information Ethics |

## 12. SUMMARY OF FUNDING REQUIRED FOR PROJECT – FAPESP

2  
8

|   | EXPENSES IN R\$<br>(separate decimals with a comma) | EXPENSES IN US\$<br>(separate decimals with a comma) |
|---|---|--|
| EQUIPMENT (MPN   MPI)   | R\$9.073,37   | US\$2.583,82   |
| MATERIALS AND SUPPLIES (MCN   MCI)  | R\$6.273,00   | US\$1.792,28   |
| SERVICES (STB   STI)  | R\$14.840,00  | US\$4.240,00   |
| TRAVEL ALLOWANCE (PER DIEM) (DIP   DIE)   | R\$70.000,00  | US\$20.000,00  |
| TRAVEL EXPENSES (TRA)   | R\$42.000,00  | US\$12.000,00  |
| <b>OVERHEAD FUNDS</b> <a href="http://www.fapesp.br/rt">www.fapesp.br/rt</a> (Please, copy the following required values from your filled budget sheet - Excel) |   |  |
| Project Infrastructure<br>(RTPCIDP - 15% of the total budget)   | R\$10.827,95  |  |
| Institutional Infrastructure<br>(RTPCIIP - 10% of the total budget requested)   | R\$14.218,64  |  |
| Complementary Benefits<br>(BC - R\$ 8.000,00 per project/year)  | R\$24.000,00  |  |
| Provision for Importation<br>(15% of the total budget requested in US\$)  | R\$0,00   |  |
| <b>FELLOWSHIPS</b>  |   |  |
| Technical Training<br><a href="http://www.fapesp.br/capacitacaotecnica">http://www.fapesp.br/capacitacaotecnica</a><br>(see item 11 below)                      | R\$40.910,40  |  |
| Post-doctoral (PD) fellowships<br><a href="http://www.fapesp.br/5389">http://www.fapesp.br/5389</a> (exception for APR in this call)                            | R\$245.494,80                                       |  |
| Fellowships overheads<br>(Reserva Técnica - Post-docs)  | R\$36.824,22  |  |
| OTHER COSTS (please specify)  |   |  |
| <b>SUB-TOTALS</b>   | <b>R\$514.462,38</b>                                | <b>US\$38.796,10</b>                                 |
| <b>TOTAL</b>  | <b>R\$514.462,38</b>                                |  |

## 13. SUMMARY OF FUNDING REQUIRED FOR PROJECT – partner 1

|                              | AMOUNT REQUESTED (US\$) |
|------------------------------|-------------------------|
| STAFF                        | US\$ 2.511,00           |
| EQUIPMENT                    | US\$ 5.896,00           |
| MATERIALS AND SUPPLIES       |                         |
| SERVICES                     | US\$ 18.426,00          |
| TRAVEL ALLOWANCE (PER DIEM)  | US\$ 18.426,00          |
| TRAVEL EXPENSES              | US\$ 18.426,00          |
| FELLOWSHIPS                  | US\$ 148.644,00         |
| OTHER COSTS (please specify) |                         |
| <b>TOTAL</b>                 | <b>US\$ 212.329,00</b>  |

**14. SUMMARY OF FUNDING REQUIRED FOR PROJECT – partner 2**

|   | AMOUNT REQUESTED (US\$) |
|---|-------------------------|
| STAFF   |                         |
| EQUIPMENT   | US\$ 10.925,00          |
| MATERIALS AND SUPPLIES  |                         |
| SERVICES  |                         |
| TRAVEL ALLOWANCE (PER DIEM)   | US\$ 13.919,00          |
| TRAVEL EXPENSES   | US\$ 13.919,00          |
| FELLOWSHIPS   |                         |
| OTHER COSTS (please specify) FINAL<br>CONFERENCE AND INDERECT COSTS | US\$ 5.569,00           |
| <b>TOTAL</b>  | US\$ 44.332,00          |

**15. SUMMARY OF FUNDING REQUIRED FOR PROJECT – partner 3**

|                              | AMOUNT REQUESTED (US\$) |
|------------------------------|-------------------------|
| STAFF                        |                         |
| EQUIPMENT                    |                         |
| MATERIALS AND SUPPLIES       |                         |
| SERVICES                     |                         |
| TRAVEL ALLOWANCE (PER DIEM)  |                         |
| TRAVEL EXPENSES              |                         |
| FELLOWSHIPS                  |                         |
| OTHER COSTS (please specify) |                         |
| <b>TOTAL</b>                 |                         |

**16. SUMMARY OF FUNDING REQUIRED FOR PROJECT – partner 4**

|                              | AMOUNT REQUESTED (US\$) |
|------------------------------|-------------------------|
| STAFF                        |                         |
| EQUIPMENT                    |                         |
| MATERIALS AND SUPPLIES       |                         |
| SERVICES                     |                         |
| TRAVEL ALLOWANCE (PER DIEM)  |                         |
| TRAVEL EXPENSES              |                         |
| FELLOWSHIPS                  |                         |
| OTHER COSTS (please specify) |                         |
| <b>TOTAL</b>                 |                         |

## 17. SUMMARY OF FUNDING REQUIRED FOR PROJECT – partner 5

|                              | AMOUNT REQUESTED (US\$) |
|------------------------------|-------------------------|
| STAFF                        |                         |
| EQUIPMENT                    |                         |
| MATERIALS AND SUPPLIES       |                         |
| SERVICES                     |                         |
| TRAVEL ALLOWANCE (PER DIEM)  |                         |
| TRAVEL EXPENSES              |                         |
| FELLOWSHIPS                  |                         |
| OTHER COSTS (please specify) |                         |
| <b>TOTAL</b>                 |                         |

## 18) BOLSAS ACADÊMICAS

|                                      | QUANTIDADE | DURAÇÃO (meses) | VALOR TOTAL |
|--------------------------------------|------------|-----------------|-------------|
| <a href="#"><u>PÓS-DOCTORADO</u></a> | 1          | 36              | 245.494,00  |
|                                      |            |                 |             |

As bolsas solicitadas, caso concedidas, devem ser cadastradas no **SAGe em bolsas concedidas como itens orçamentários em auxílios.**

## 19. TECHNICAL TRAINING FELLOWSHIPS – FAPESP

ARE YOU REQUESTING FAPESP TECHNICAL TRAINING FELLOWSHIPS? ☒ YES ☐ NO  
 IF YOU SELECTED "YES", PLEASE SPECIFY THE LEVEL AND TOTAL NUMBER OF FELLOWSHIPS REQUESTED

| TECHNICAL TRAINING - TT                     | NUMBER | HOURS DEDICATED PER WEEK | DURATION (months) | TOTAL REQUESTED (R\$) |
|---|--------|--------------------------|-------------------|-----------------------|
| <b>TOTAL TT</b>                             |        |                          |                   |                       |
| <input type="checkbox"/> LEVEL 1            |        |                          |                   |                       |
| <input type="checkbox"/> LEVEL 2            |        |                          |                   |                       |
| <input checked="" type="checkbox"/> LEVEL 3 | 1      | 40                       | 36                | 40.910,40             |
| <input type="checkbox"/> LEVEL 4            |        |                          |                   |                       |
| <input type="checkbox"/> LEVEL 4A           |        |                          |                   |                       |
| <input type="checkbox"/> LEVEL 5            |        |                          |                   |                       |

**20. GRANTS REQUESTED TO OR AWARDED BY OTHER SOURCES RELATED TO THE PRESENT PROPOSAL (please indicate currency)**

3

| SOURCE AND GRANT DESCRIPTION | TOTAL REQUESTED<br>or AWARDED |
|------------------------------|-------------------------------|
|                              |                               |
|                              |                               |
|                              |                               |
|                              |                               |
|                              |                               |

**21) FAPESP SCHOLARSHIPS AND/OR GRANTS RELATED TO THIS PROPOSAL**

Are there any FAPESP grants related to this proposal? If so, please inform below their corresponding process numbers.

| GRANT ID             |                      |                      |                      |   |                      |                      |                      |                      |                      |   |                      | Comments |
|----------------------|----------------------|----------------------|----------------------|---|----------------------|----------------------|----------------------|----------------------|----------------------|---|----------------------|----------|
| <input type="text"/> | <input type="text"/> | <input type="text"/> | <input type="text"/> | / | <input type="text"/> | <input type="text"/> | <input type="text"/> | <input type="text"/> | <input type="text"/> | - | <input type="text"/> |          |
| <input type="text"/> | <input type="text"/> | <input type="text"/> | <input type="text"/> | / | <input type="text"/> | <input type="text"/> | <input type="text"/> | <input type="text"/> | <input type="text"/> | - | <input type="text"/> |          |
| <input type="text"/> | <input type="text"/> | <input type="text"/> | <input type="text"/> | / | <input type="text"/> | <input type="text"/> | <input type="text"/> | <input type="text"/> | <input type="text"/> | - | <input type="text"/> |          |
| <input type="text"/> | <input type="text"/> | <input type="text"/> | <input type="text"/> | / | <input type="text"/> | <input type="text"/> | <input type="text"/> | <input type="text"/> | <input type="text"/> | - | <input type="text"/> |          |

**22. STATEMENT FROM THE DIRECTOR OF THE UNIT WHERE THE PROJECT WILL BE DEVELOPED**

|  |
|--|
| <p>1- I declare to be aware of the project's infrastructural needs and requirements and that the part of the Technical Reserve allocated to Research Infrastructural Expenses is destined towards the support of collective research infrastructure that benefits research projects supported by FAPESP in the Unit, according to the Employment Plan to be approved by the Unit's collegiate body and submitted to FAPESP. The guidelines for the application of the part of the Technical Reserve allocated to Research Infrastructural Expenses can be found on <a href="http://www.fapesp.br/rt">http://www.fapesp.br/rt</a> (in Portuguese).</p> <p>2- Therefore, I declare that in case this Project is approved and for the duration of its contract, the researcher and group of researchers taking part in the project will have all the institutional support necessary for its development, as previously agreed with the principle investigator. Specially, the researcher and research group taking part in the Project will be granted physical space for the adequate installation and operation of the requested equipment, permission to use all facilities (laboratories, computer networks, library, databases etc.) and access to all services (lab technicians, administrative support etc.) available at the institution and that are relevant for its fulfillment. If the development of the project is hindered or made impracticable due to the non-fulfillment of this clause and without previous consent from FAPESP, the Institution commits itself to reimburse FAPESP's expenditure towards it.</p> <p>3- I am also aware that the Principal Investigator and the Host Institution are held totally responsible for requesting, obtaining, holding and showing whenever requested by FAPESP all applicable and legal authorizations required for the good execution of the Project. Such authorizations are to be issued by the competent official Agencies with whose policies the conduction of the research must comply as required by Law. (The currently required authorizations are described on <a href="http://www.fapesp.br/4476">http://www.fapesp.br/4476</a>, in Portuguese). Because such authorizations can change according to the ongoing applicable legal requirements, it is strongly recommended that the Host Institution sector responsible for the pertinent orientation be consulted.</p> <p>4- I am aware that the failure to fulfill the terms of this statement may compromise the course of future requests applied to FAPESP by researchers from this Unit.</p> |
| NAME: <b>José Carlos Miguel</b>  |
| POSITION/TITLE: <b>Director of the Faculty of Philosophy and Science</b>   |
| PLACE, DATE AND SIGNATURE: <b>23/06/2016</b>   |

**23. SÃO PAULO APPLICANT STATEMENT**


|   |
|---|
| <p>1- I declare to be aware of the procedures adopted by FAPESP for the analysis of requests in this Call for Proposal. I authorize this request to be analysed according to these procedures and specifically to submit it to the analysis of reviewers chosen by FAPESP, whose identities will not be disclosed.</p> <p>2- The pieces of information herein and those contained in my CV accompanying this proposal application were revised by myself and are correct and updated.</p> <p>3- I am also aware that the Principal Investigator and the Host Institution are held totally responsible for requesting, obtaining, holding and showing whenever requested by FAPESP all applicable and legal authorizations required for the good execution of the Project. Such authorizations are to be issued by the competent official Agencies with whose policies the conduction of the research must comply as required by Law. (The currently required authorizations are described on <a href="http://www.fapesp.br/4476">http://www.fapesp.br/4476</a>, in Portuguese). Because such authorizations can change according to the ongoing applicable legal requirements, it is strongly recommended that the Host Institution sector responsible for the pertinent orientation be consulted.</p> <p>4- I am aware that the failure to fulfill the terms of this statement may compromise the analysis and the eventual granting of this proposal.</p> |
| PLACE, DATE AND APPLICANT'S SIGNATURE:<br><b>23/06/2016</b>   |



## 24. DOCUMENTS TO BE ATTACHED - (see detailed instructions in the Call for Proposal)

| <b>REGULAR RESEARCH AWARDS</b> - SUBMITT ONLY ONE SET OF ALL LISTED DOCUMENTS<br><b>COMPLEMENTARY DOCUMENTS OR OTHER MEDIA (MAPS, CDs, FOLDERS, ETC.) MUST BE SUBMITTED IN THREE SETS</b>  | Check-list               |                          |
|--|--------------------------|--------------------------|
| (MANDATORY DOCUMENTS REQUIRED FOR ANALYSIS) (PLEASE DO NOT BIND)   | Applicant                | FAPESP                   |
| Research Proposal Form (this document)   | <input type="checkbox"/> | <input type="checkbox"/> |
| Joint Research project   | <input type="checkbox"/> | <input type="checkbox"/> |
| <a href="#">Researcher's Registration Form (in Portuguese)</a>   | <input type="checkbox"/> | <input type="checkbox"/> |
| Summary of the Principal Investigators' CVs and of each researcher involved on both sides (in English - guidelines at <a href="http://www.fapesp.br/en/6351">www.fapesp.br/en/6351</a> )   | <input type="checkbox"/> | <input type="checkbox"/> |
| <a href="#">Worksheet with members of the team</a>   | <input type="checkbox"/> | <input type="checkbox"/> |
| When applicable, the corresponding <b>activity plans</b> for each one of the requested Technical Training <b>Fellowships</b> (in English)  | <input type="checkbox"/> | <input type="checkbox"/> |
| When applicable, the corresponding <b>activity plans</b> for each one of the requested Post-doc <b>Fellowships</b> (in English)  | <input type="checkbox"/> | <input type="checkbox"/> |
| Budget Worksheets ( <b>please, enclose detailed budget according to FAPESP standards</b> ) (In English)  | <input type="checkbox"/> | <input type="checkbox"/> |
| <a href="#">Description of the scientific equipment base</a> of the Institution(s) hosting the Project. (In English - guidelines available at: <a href="http://www.fapesp.br/137#4785">www.fapesp.br/137#4785</a> )  | <input type="checkbox"/> | <input type="checkbox"/> |
| <b>ANNEX II</b> - Approved information disclosed by the Host Institution(s) as to the institutional infrastructure needed for the project conduction. <b>This document should be presented at the moment of the Grant Contract Signature, if the proposal is granted</b> (in Portuguese).  | <input type="checkbox"/> | <input type="checkbox"/> |
| A Letter of Agreement between the Higher Education or Research Institutions to which the national team PIs are affiliated, establishing how Intellectual Property rights, confidentiality, and publications will be treated jointly, in observance of the policies of each funding Party. This document is not mandatory for the submission of proposals, but the presentation of a copy of the signed Agreement is mandatory for the grant contract, in case of approval. | <input type="checkbox"/> | <input type="checkbox"/> |
| <b>ATTENTION: ALL REQUESTS THAT ARE NOT ACCOMPANIED BY THE MANDATORY DOCUMENTS FOR ANALYSIS WILL BE RETURNED.</b>  |                          |                          |

FAPESP, JUNE 2016

|  |                                      |  |
|--|--------------------------------------|--|
|  | Programme ANR T-AP Digging into data | Réservé à l'organisme gestionnaire du programme<br>N° de dossier : |
|  | Document financier de soumission     | Date de révision :<br><b>Edition 2016</b>                          |

Fiche d'identité du projet

|                                |   |
|--------------------------------|---|
| Acronyme                       | OPLADYN   |
| Titre du projet<br>en français | Etude de la dynamique d'opinion et l'évolution du langage basée sur du "Big Data" |
| Titre du projet<br>en anglais  | Understanding opinion and language dynamics using massive data                    |
| Durée du projet                | 36 mois (24 à 36)   |

Récapitulatif : partenariat, budget et main d'œuvre

|                             | Sigle du partenaire | Coût Complet (€) | Aide demandée (€) | Personnel permanent (personne.mois) | Personnel non permanent (personne.mois) |
|-----------------------------|---------------------|------------------|-------------------|-------------------------------------|---|
| Partenaire 1 (coordinateur) | UCP - LPTM          | 339 208          | 124 072           | 16,00                               | 42,00                                   |
| Partenaire 2                | ENS - LATTICE       | 192 036          | 55 640            | 8,50                                | 18,00                                   |
| Partenaire 3                | 0                   | -                | -                 | -                                   | -                                       |
| Partenaire 4                | 0                   | -                | -                 | -                                   | -                                       |
| Partenaire 5                | 0                   | -                | -                 | -                                   | -                                       |
| Partenaire 6                | 0                   | -                | -                 | -                                   | -                                       |
| Partenaire 7                | 0                   | -                | -                 | -                                   | -                                       |
| Partenaire 8                | 0                   | -                | -                 | -                                   | -                                       |
| Partenaire 9                | 0                   | -                | -                 | -                                   | -                                       |
| Partenaire 10               | 0                   | -                | -                 | -                                   | -                                       |
| TOTAUX                      |                     | 531 244          | 179 712           | 24,50                               | 60,00                                   |

**Résumé (non confidentiel) du projet en français** (1500 caractères maximum)

Les nouvelles technologies, et les données immenses (ou big data) qu'elles permettent de récolter de manière automatique, sont en train de changer très rapidement le cadre dans lequel l'information est partagée : cette récolte systématique de données entraîne effectivement des changements dans les comportements et la prise de décision des acteurs sociaux . Nous proposons d'utiliser les données massives générées par la participation directe ou indirecte aux médias sociaux, pour mieux saisir l'émergence de nouvelles structures de l'activité humaine – plus spécifiquement, dans le domaine de la diffusion de l'opinion et du changement linguistique. Notre objectif est ainsi de construire des modèles explicatifs basés sur les données et de questionner les conséquences éthiques de l'utilisation de données massives.

Nous avons choisi à cet effet deux bases de données ayant des caractéristiques très différentes mais complémentaires, l'une correspondant à un journal traditionnel, le New York Times et l'autre à un média très récent, Twitter. Notre équipe inclut des chercheurs en Informatique, Physique, Linguistique, Philosophie et Droit. Nous souhaitons contribuer à la construction d'une vision interdisciplinaire qui tire profit des corrélations observées dans le Big Data pour améliorer la compréhension de la dynamique des actions sociales, par un travail conjoint des sciences « dures » et des sciences sociales.

**Résumé (non confidentiel) du projet en anglais** (1500 caractères maximum)

Big Data technologies are changing the informational environment in which people act, keeping traces of activities, merging behavior and decision-making processes of social actors. Based on traces left by social media, we will study emerging patterns in social actions, focusing on opinion diffusion and language evolution. The following questions will guide our research: How can relevant information be obtained from Big Data and used to elaborate explanatory models of social actions? What are the ethical consequences of the application of Big Data in the study of human actions? Two databases are selected: The New York Times collection, a traditional journal, and a collection of political tweets from Twitter, a newer online medium. Our team, with expertise in Data Science, Physics, Linguistics, Philosophy and Law aims to develop an interdisciplinary view of the relation between information patterns in Big Data and the dynamics of social actions, bridging the Social and Natural Sciences.

**NOTA : Tous les champs de cet onglet "Fiche Identité" doivent être renseignés.**

**Remarque** : toutes les informations figurant ci-dessus à l'exception de celles relatives aux trois derniers champs ont vocation à être publiées si le projet est retenu pour financement (sous réserve d'une mise à jour si besoin). En déposant un dossier, les partenaires acceptent la publication de toutes ces informations.

Fiche Partenaire 1 : Identification et budget

Responsable scientifique et technique

Coordinateur du projet (France) : oui

Genre

F

Prénom

Laura

Nom

HERNANDEZ

Titre

Tél.

Tél. portable

E-mail

[laura.hernandez@u-cergy.fr](mailto:laura.hernandez@u-cergy.fr)

Identification du partenaire

Nom complet du partenaire  
(laboratoire / entreprise / ...)

UNIVERSITE CERGY-PONTOISE

Sigle du partenaire  
(laboratoire / entreprise / ...)

UCP - LPTM

Catégorie de partenaire

Organisme public de recherche

Base de calcul pour l'assiette de l'aide

Coût marginal

Etes vous un EPIC dans un projet partenarial public-privé ?

Pour un laboratoire d'organisme public de recherche :

Type d'unité : UMR

Numéro d'unité : 8089

Tutelle gestionnaire du financement  
si le projet est retenu :

UNIVERSITE CERGY PONTOISE

Autre(s) tutelle(s) :

CNRS (DR05)

Pour une entreprise :

N° Siret : 199 517 939 00013

Effectif (si PME ) :

Adresse de réalisation des travaux

numéro de rue : 33

adresse : Boulevard du Port

adresse (suite) :

code postal : 95011

ville : CERGY PONTOISE CEDEX

cedex :

pays : France

Demande financière (montant HT en €, incluant la TVA non récupérable)

| Tâches  | Investissement<br>t<br>(€) | Personnels        |          |  |          |   |          | Prestations de<br>service<br>externe (€) | Missions<br>(€) | Autres<br>dépenses de<br>charges<br>externes<br>(€) | Dépenses sur<br>facturation<br>interne (€) | Totaux<br>(€) |
|---------|----------------------------|-------------------|----------|--|----------|---|----------|--|-----------------|---|--|---------------|
|         |                            | permanents        |          | NON permanents avec<br>financement ANR demandé |          | NON permanents <b>sans</b><br>financement ANR demandé |          |  |                 |   |  |               |
|         |                            | personne.<br>mois | Coût (€) | personne. mois                                 | Coût (€) | personne. mois  | Coût (€) |  |                 |   |  |               |
| Tâche 0 |                            | 3,00              | 16 785   |  |          |   |          |  | 6 300           | 3 000   |  | 26 085        |
| Tâche 1 |                            | 2,50              | 13 988   |  |          |   |          |  | 5 500           | 3 000   |  | 22 488        |
| Tâche 2 |                            | 2,50              | 13 988   | 21,00  | 33 750   |   |          |  | 5 500           | 3 000   |  | 56 238        |
| Tâche 3 |                            | 3,00              | 16 785   | 21,00  | 33 750   |   |          |  | 5 500           | 3 000   |  | 59 035        |
| Tâche 4 |                            | 2,50              | 13 988   |  |          |   |          |  | 5 500           | 3 000   |  | 22 488        |
| Tâche 5 |                            | 2,50              | 13 988   |  |          |   |          |  | 5 500           | 3 000   |  | 22 488        |
|         |                            |                   |          |  |          |   |          |  |                 |   |  | -             |
|         |                            |                   |          |  |          |   |          |  |                 |   |  | -             |
|         |                            |                   |          |  |          |   |          |  |                 |   |  | -             |
|         |                            |                   |          |  |          |   |          |  |                 |   |  | -             |
| Totaux  | -                          | 16                | 89 520   | 42   | 67 500   | -   | -        | -  | 33 800          | 18 000  | -  | 208 820       |

Pour information : montant maxi des frais de gestion/frais de structure pris en compte par l'ANR = 4 772 €

Frais de gestion / frais de structure (€)

4 772 €

80%

Frais d'environnement (€)

125 616

Coût complet (€)

339 208

Coût éligible pour le calcul de l'aide : Assiette (€)

124 072

Taux d'aide demandé

100%

Aide demandée (€)

124 072

Engagement du partenaire (Les signatures sont à apposer uniquement sur le document papier)

Après avoir pris connaissance de l'ensemble du dossier de soumission ([http://diggingintodata.org/#quicktabs-quick\\_nav\\_interior\\_about=2](http://diggingintodata.org/#quicktabs-quick_nav_interior_about=2)) et du règlement relatif aux modalités d'attribution des aides de l'ANR, je donne mon accord pour la participation de mon laboratoire/organisme/entreprise au projet, dans les conditions décrites de répartition des tâches et de financement demandé, et garantis les informations données. Je m'engage à envoyer une copie de ce dossier à chacun des organismes de tutelle de mon laboratoire (pour les laboratoires d'organisme public de recherche uniquement, hors EPIC)

Responsable scientifique et technique

Directeur de laboratoire ou de l'unité d'accueil

Représentant légal

Prénom : Laura

Nom : HERNANDEZ

Prénom : Philippe

Nom : LECHEMINANT

Prénom : François

Nom : GERMINET

Signature

Signature

Signature

Les informations personnelles transmises dans ces documents sont obligatoires et seront conservées en fichiers par l'ANR ou par la structure support mandatée par elle pour assurer la conduite opérationnelle de l'évaluation et l'administration des dossiers.

Conformément à la loi n° 78-17 du 6 janvier 1978 modifiée, relative à l'Informatique, aux Fichiers et aux Libertés, les personnes concernées disposent d'un droit d'accès et de rectification des données personnelles les concernant. Les personnes concernées peuvent exercer ce droit en s'adressant à la structure support, ou l'ANR quand il s'agit d'un programme géré directement par elle (voir coordonnées dans le texte de l'appel à projets).

Fiche Partenaire 2 : Identification et budget

Responsable scientifique et technique

Coordinateur du projet (France) : non

Genre

H

Prénom

Benjamin

Nom

FAGARD

Titre

Tél.

Tél. portable

E-mail

benjamin.fagard@ens.fr

Identification du partenaire

Nom complet du partenaire  
(laboratoire / entreprise / ...)

ECOLE NORMALE SUPERIEURE

Sigle du partenaire  
(laboratoire / entreprise / ...)

ENS - LATTICE

Catégorie de partenaire

Organisme public de recherche

Base de calcul pour l'assiette de l'aide

Coût marginal

Etes vous un EPIC dans un projet partenarial public-privé ?

Pour un laboratoire d'organisme public de recherche :

Type d'unité :

UMR

Numéro d'unité :

8094

Tutelle gestionnaire du financement

ENS

si le projet est retenu :

Autre(s) tutelle(s) :

CNRS

PARIS 3

Pour une entreprise :

N° Siret :

Effectif (si PME) :

Adresse de réalisation des travaux

numéro de rue :

45

adresse :

Rue d'Ulm

adresse (suite) :

code postal :

75005

ville :

PARIS

cedex :

pays :

France

Demande financière (montant HT en €, incluant la TVA non récupérable)

| Tâches  | investissements<br>(€) | Personnels permanents |          | Personnels NON permanents avec financement ANR demandé |          | Personnels NON permanents sans financement ANR demandé |          | Prestations de service externe (€) | Missions (€) | Autres dépenses de charges externes (€) | Dépenses sur facturation interne (€) | Totaux (€) |
|---------|------------------------|-----------------------|----------|--|----------|--|----------|------------------------------------|--------------|---|--------------------------------------|------------|
|         |                        | personne. mois        | Coût (€) | personne. mois   | Coût (€) | personne. mois   | Coût (€) |                                    |              |   |                                      |            |
| Tâche 0 |                        | 1,50                  | 9 503    |  |          |  |          |                                    |              |   |                                      | 9 503      |
| Tâche 1 |                        | 1,00                  | 6 705    |  |          |  |          |                                    |              |   |                                      | 6 705      |
| Tâche 2 |                        | 1,00                  | 6 705    |  |          |  |          |                                    |              |   |                                      | 6 705      |
| Tâche 3 |                        | 2,00                  | 11 190   | 9,00   | 26 750   |  |          |                                    |              |   |                                      | 37 940     |
| Tâche 4 |                        | 2,00                  | 11 190   | 9,00   | 26 750   |  |          |                                    |              |   |                                      | 37 940     |
| Tâche 5 |                        | 1,00                  | 6 705    |  |          |  |          |                                    |              |   |                                      | 6 705      |
| Tâche 6 |                        |                       |          |  |          |  |          |                                    |              |   |                                      | -          |
| Tâche 7 |                        |                       |          |  |          |  |          |                                    |              |   |                                      | -          |
| Tâche 8 |                        |                       |          |  |          |  |          |                                    |              |   |                                      | -          |
| Tâche 9 |                        |                       |          |  |          |  |          |                                    |              |   |                                      | -          |
| Totaux  | -                      | 9                     | 51 998   | 18   | 53 500   | -  | -        | -                                  | -            | -                                       | -                                    | 105 498    |

Pour information : montant maxi des frais de gestion/frais de structure pris en compte par l'ANR = 2 140 €

Frais de gestion / frais de structure (€)

2 140 €

80%

Frais d'environnement (€)

84 398

Coût complet (€)

192 036

Coût éligible pour le calcul de l'aide : Assiette (€)

55 640

Taux d'aide demandé

100%

Aide demandée (€)

55 640

Engagement du partenaire (Les signatures sont à apposer uniquement sur le document papier)

Après avoir pris connaissance de l'ensemble du dossier de soumission (http://diggingintodata.org/#quicktabs-quick\_nav\_interior\_about=2) et du règlement relatif aux modalités d'attribution des aides de l'ANR, je donne mon accord pour la participation de mon laboratoire/organisme/entreprise au projet, dans les conditions décrites de répartition des tâches et de financement demandé, et garantis les informations données. Je m'engage à envoyer une copie de ce dossier à chacun des organismes de tutelle de mon laboratoire (pour les laboratoires d'organisme public de recherche uniquement, hors EPIC)

Responsable scientifique et technique

Directeur de laboratoire ou de l'unité d'accueil

Représentant légal

Prénom :

Benjamin

Nom :

FAGARD

Prénom :

Thierry

Nom :

POIBEAU

Prénom :

Marc

Nom :

MEZARD

Signature

Signature

Signature

Les informations personnelles transmises dans ces documents sont obligatoires et seront conservées en fichiers par l'ANR ou par la structure support mandatée par elle pour assurer la conduite opérationnelle de l'évaluation et l'administration des dossiers.

Conformément à la loi n° 78-17 du 6 janvier 1978 modifiée, relative à l'Informatique, aux Fichiers et aux Libertés, les personnes concernées disposent d'un droit d'accès et de rectification des données personnelles les concernant. Les personnes concernées peuvent exercer ce droit en s'adressant à la structure support, ou l'ANR quand il s'agit d'un programme géré directement par elle (voir coordonnées dans le texte de l'appel à projets).

Tableaux récapitulatifs

Récapitulatif des dénominations des partenaires

|               | Sigle du partenaire | Nom complet du partenaire |
|---------------|---------------------|---------------------------|
| Partenaire 1  | UCP - LPTM          | UNIVERSITE CERGY-PONTOISE |
| Partenaire 2  | ENS - LATTICE       | ECOLE NORMALE SUPERIEURE  |
| Partenaire 3  | 0                   | -                         |
| Partenaire 4  | 0                   | -                         |
| Partenaire 5  | 0                   | -                         |
| Partenaire 6  | 0                   | -                         |
| Partenaire 7  | 0                   | -                         |
| Partenaire 8  | 0                   | -                         |
| Partenaire 9  | 0                   | -                         |
| Partenaire 10 | 0                   | -                         |

Récapitulatif financier du projet (montant HT incluant, le cas échéant, la TVA non récupérable)

|               | Equipements (€) | Personnels     |          |   |          |   |          | Prestations de service externe (€) | Missions (€) | Autres dépenses de charges externes (€) | Dépenses justifiées sur facturation interne (€) | Sous-totaux (hors frais) (€) | Frais de gestion / frais de structure (€) | Frais d'environnement (€) | Coût complet (€) | Assiette de l'aide (€) | Taux d'aide demandé | Aide demandée (€) |
|---------------|-----------------|----------------|----------|---|----------|---|----------|------------------------------------|--------------|---|---|------------------------------|---|---------------------------|------------------|------------------------|---------------------|-------------------|
|               |                 | permanents     |          | NON permanents avec financement ANR demandé |          | NON permanents sans financement ANR demandé |          |                                    |              |   |   |                              |   |                           |                  |                        |                     |                   |
|               |                 | personne. mois | Coût (€) | personne. mois                              | Coût (€) | personne. mois                              | Coût (€) |                                    |              |   |   |                              |   |                           |                  |                        |                     |                   |
| Partenaire 1  | -               | 16,00          | 89 520   | 42  | 67 500   | -   | -        | -                                  | 33 800       | 18 000                                  | -   | 208 820                      | 4 772                                     | 125 616                   | 339 208          | 124 072                | 100%                | 124 072           |
| Partenaire 2  | -               | 8,50           | 51 998   | 18  | 53 500   | -   | -        | -                                  | -            | -                                       | -   | 105 498                      | 2 140                                     | 84 398                    | 192 036          | 55 640                 | 100%                | 55 640            |
| Partenaire 3  | -               | -              | -        | -   | -        | -   | -        | -                                  | -            | -                                       | -   | -                            | -   | -                         | -                | -                      | 100%                | -                 |
| Partenaire 4  | -               | -              | -        | -   | -        | -   | -        | -                                  | -            | -                                       | -   | -                            | -   | -                         | -                | -                      | 100%                | -                 |
| Partenaire 5  | -               | -              | -        | -   | -        | -   | -        | -                                  | -            | -                                       | -   | -                            | -   | -                         | -                | -                      | 0%                  | -                 |
| Partenaire 6  | -               | -              | -        | -   | -        | -   | -        | -                                  | -            | -                                       | -   | -                            | -   | -                         | -                | -                      | 0%                  | -                 |
| Partenaire 7  | -               | -              | -        | -   | -        | -   | -        | -                                  | -            | -                                       | -   | -                            | -   | -                         | -                | -                      | 0%                  | -                 |
| Partenaire 8  | -               | -              | -        | -   | -        | -   | -        | -                                  | -            | -                                       | -   | -                            | -   | -                         | -                | -                      | 0%                  | -                 |
| Partenaire 9  | -               | -              | -        | -   | -        | -   | -        | -                                  | -            | -                                       | -   | -                            | -   | -                         | -                | -                      | 0%                  | -                 |
| Partenaire 10 | -               | -              | -        | -   | -        | -   | -        | -                                  | -            | -                                       | -   | -                            | -   | -                         | -                | -                      | 0%                  | -                 |
| Totaux        | -               | 24,50          | 141 518  | 60,00                                       | 121 000  | -   | -        | -                                  | 33 800       | 18 000                                  | -   | 314 318                      | 6 912                                     | 210 014                   | 531 244          | 179 712                | 100%                | 179 712           |

## Justification of Resources:

|   |  |               |
|---|--|---------------|
| a | 2 Post-Doctoral contracts                | 100000        |
| b | 6 Master internship allowances           | 21000         |
| c | Missions and meetings of the whole group | 20000         |
| d | Dissemination of results                 | 10000         |
| e | Services                                 | 5000          |
| f | Final conference                         | 7500          |
| g | missions/material for post docs          | 3000          |
| h | Local workshop                           | 6300          |
| i | Managment fees                           | 6912          |
|   | <b>Total</b>                             | <b>179712</b> |

Table 1: French JoR. Currency Euro. (a) Post-Docs will integrate tasks 2,3 and 4. (b) Each master student, will integrate task 2 or 3, (c) This includes one working meeting per country, which means two 'long courier' for the five French members, (d) Participation to fees for Open Journals to guarantee wide diffusion of results (typical fee 1500-3000 €) and creation of our website; (e) Participation to mutualised computing platforms (h) Participation to the organization of an open workshop. This includes invited speakers and traveling help for students. Extra contribution will be demanded to other funding agencies. (i) Administrative fees required by local institutions for the project management

|   |  |              |
|---|--|--------------|
| a | Missions and meetings of the whole group | 25150        |
| b | Equipment                                | 9870         |
| c | Final conference                         | 3130         |
| d | Indirect costs (5%)                      | 1902         |
|   | <b>Total</b>                             | <b>39942</b> |

Table 2: Argentinian JoR. Currency Euro. (a) 6 Missions for Argentina with at least 13 nights at destination. (b) This includes computer for processing data (with GPU) and some hard disk to storage the whole dataset. (c) Participation to the organization of an open workshop. (d) It is the cost for the administration of funds by UBA UVT.



|   |                                |                |
|---|--------------------------------|----------------|
| a | Equipment                      | 2.334          |
| b | Materials and supplies         | 1.619          |
| c | Services                       | 3.830          |
| d | Travel allowance               | 18.068         |
| e | Travel expenses                | 10.841         |
| f | Post-doctoral fellowship       | 72.872         |
| g | Technical training scholarship | 10.560         |
| h | Overhead funds                 | 20.878         |
|   | <b>Total</b>                   | <b>141.002</b> |

Table 3: Brazilian JoR. Currency Euro. (a) We are requesting 1 laptop, 1 webcam, 1 digital camera, 4 external storage HDs, and 1 recorder. These devices will be used to perform tasks fundamental to the work, mainly enabling remote communication among the researchers of the Brazilian, French, and Argentinian groups, the production of papers and reports, and data storage, amongst other uses. (b) This item refers to basic supplies (paper, toner, and USB drives). (c) Services related to text editing, translation and transcription of recorded lectures and conferences for teaching purposes (Task 5). (g) The travel allowance is related to eight trips of the PI, co-Pi, and AI during the 3 years of the project, in order to participate in general meetings, workshops, and conferences. (h) Ten international tickets for the PI, co-Pi, and AI during the 3 years of the project, in order to participate in general meetings, workshops, and conferences. (i) The post-doctoral research project sets out the argument that natural languages possess most of the characteristics of complex adaptive systems. As a case study, the evolution of the English language is going to be investigated using data obtained from the New York Times database (Task 3). (j) The technical training scholarship will enable an undergraduate student of Philosophy or Information Science to participate in the research activities, mainly assisting with the data ethical analysis (Task 4).

## LAURA HERNANDEZ-Summary CV

Associate Professor (MCF-HDR) Laboratoire de Physique Théorique et Modélisation (LPTM) UMR 8089CNRS-Université de Cergy-Pontoise France. [Laura.Hernandez@u-cergy.fr](mailto:Laura.Hernandez@u-cergy.fr)

### ACADEMIC BACKGROUND

- 16/2/2015** Research Conducting Qualification Diploma, “Habilitation à Diriger des Recherches” (HDR).  
**1990-1993** Ph.D. in Physics from INSA, Toulouse, France,  
**1988-1990** Ph.D. Student, TANDAR Lab, CNEA, Buenos Aires, Argentina, (*left to continue in France*).  
**1981-1987** « Licenciada en Física » Universidad de Buenos Aires (Master level, Bac+6)

### RESEARCH ACTIVITY

**Key words:** Critical Phenomena, Phase Transitions, Disordered Magnetic Systems, Monte Carlo Methods, Complex Systems, Complex Networks

**10 Related Publications** (full list and pdfs in [https://www.researchgate.net/profile/Laura\\_Hernandez13](https://www.researchgate.net/profile/Laura_Hernandez13) )

- *Entropic determination of the phase transition in a co-evolving opinion-formation model.* E.Burgos, [Laura Hernández](#), H.Ceva, R.P.J. Perazzo. Phys. Rev. E **91**, 032808 (2015).
- *Study of the influence of the phylogenetic distance on the interaction network of mutualistic ecosystems.* R.P.J. Perazzo, [Laura Hernández](#), H.Ceva, E.Burgos, José Ignacio Alvarez-Hamelin. Physica A **394** p124–135 (2014).
- *Une loi universelle pour la participation aux élections municipales.* [Laura Hernández](#), François Savatier, *Pour la Science*, (2014) [http://www.pourlascience.fr/ewb\\_pages/a/actu-une-loi-universelle-pour-la-participation-aux-elections-municipales-32939.php](http://www.pourlascience.fr/ewb_pages/a/actu-une-loi-universelle-pour-la-participation-aux-elections-municipales-32939.php)
- *Universal size effects for populations in group-outcome decision-making problems”* Christian Borghesi, [Laura Hernández](#), Rémi Louf, Fabrice Caparros, Physical Review E **88**, 062813 (2013) (article cited by CNRS journal and general public journal). (\*)
- *Coevolutionary network approach to cultural dynamics controlled by intolerance.* Carlos Gracia Lázaro, Fernando Quijandria, [Laura Hernández](#), Luis Mario Floría, and Yamir Moreno. Rev. E **84** 067101 (2011).
- *Understanding and characterizing nestedness.”* E.Burgos, H.Ceva, [Laura Hernández](#), R.P.J. Perazzo. Comp.Phys.Comm. Comp. Phys. Comm. **180**, p532 (2009).
- *On the application of the Critical Minimum Energy Sub-space method to highly disordered systems.* [Laura Hernández](#) and Horacio Ceva. Physica A Physica A **338**, 2389 (2009).
- *Two classe of bipartite networks: nested biological and social systems.”* E.Burgos, H.Ceva, [Laura Hernández](#), R.P.J. Perazzo, Mariano Devoto, Diego Medan. Phys. Rev. E **78** 046113 (2008).
- *Wang-Landau study of the critical behaviour of the bimodal 3D-Random Field Ising Model.* [Laura Hernández](#) and Horacio Ceva. Physica A, **387**, 2793 (2008).
- *Existence of a tricritical point at finite field in the three dimensionnal random field Ising model,* [Laura Hernández](#) and H. T. Diep. Phys.Rev. B **55** p14080 (1997).

#### Participation to Research Diffusion:

- Regular (at least once a year) participation to international conferences (2 invited). 8 invited seminars.
- Organization of 6 international conferences.
- Development of a vulgarization tool based on (\*) <http://iscpif.fr/projects/municipales/>

### TEACHING AND RESEACH ADMINISTRATION

- Co-director of the Master *Theoretical Physics and Applications* (TPA), Director of *Complex Systems* path.
- Member of the Steering Committee of the Institute of Complex Systems of Paris Ile de France.
- Councilor of the Complex System Digital Campus (Unitwin UNESCO) and of Complex Systems Society
- Co-organizer of HumanICT network of the RNSC (Complex Systems National Network)
- Member of the hiring expert committees (Theoretical and Condensed Matter Physics sections) of the UCP (1993-2006), and Computer Science, 2016.
- Referee for Physica A, Physics Letters A, Journ. Statist. Mech., and several international conferences
- Member of the hiring expert committee (Theoretical and Condensed Matter Physics sections) of the UCP (1993-2006) and of Computer Science committee 2016.

### RESEARCH SUPERVISION

Master:4; PhD: 1 co-direction + 1 ongoing; Post-Doc 1 (supported by DIM Ile de France)

## José Ignacio Alvarez-Hamelin-- Sumary C.V.

Associate Professor, Universidad de Buenos Aires, Argentina  
 Professor, Instituto Tecnológico de Buenos Aires, Argentina  
 Researcher, CONICET, Argentina  
[ihameli@cnet.fi.uba.ar](mailto:ihameli@cnet.fi.uba.ar) (<http://cnet.fi.uba.ar>)

### Academic Background

**Postdoctoral position:** Laboratoire de Physique Théorique, Université Paris-Sud, Orsay, France (2003-2005). Advisor: Alessandro Vespignani.

**PhD on Informatics:** Université Paris-Sud, France (2002).

**Ingeniero Electrónico:** Universidad Nacional de Buenos Aires, 1997 (UBA) Argentina.

### Research Activity

**Keywords:** Complex Networks, Internet, Ad-hoc networks, routing protocols, network measurements.

**10 Related Publications** (<http://cnet.fi.uba.ar/publicaciones/>)

- J. Ignacio Alvarez-Hamelin, Luca Dall'Asta, Alain Barrat, and Alessandro Vespignani. Large scale networks fingerprinting and visualization using the k-core decomposition. In Y. Weiss, B. Schölkopf, and J. Platt, editors, Advances in Neural Information Processing Systems 18, pages 41-50, Cambridge, MA, 2006. MIT Press.
- J. Ignacio Alvarez-Hamelin, Luca Dall'Asta, Alain Barrat, and Alessandro Vespignani. K-core decomposition of Internet graphs: hierarchies, self-similarity and measurement biases. Networks and Heterogeneous Media, 3(2):371-293, 2008.
- José I. Orlicki, J. Ignacio Alvarez-Hamelin, and Pablo I. Fierens. Scalable Faceted Ranking in Tagging Systems, volume 45 of Lecture Notes in Business Information Processing. Springer Berlin Heidelberg, 2010. 10.1007/978-3-642-12436-5\_21.
- Pol Colomer de Simon, M. Angeles Serrano, Mariano G. Beiró, J. Ignacio Alvarez-Hamelin, and Marian Boguñá. Deciphering the global organization of clustering in real complex networks. Sci. Rep., 3:2517, Aug 2013.
- Mariano G. Beiró, Jorge R. Busch, Sebastian P. Grynberg, and J. Ignacio Alvarez-Hamelin. Obtaining communities with a fitness growth process. Physica A: Statistical Mechanics and its Applications, 392(9):2278 - 2293, 2013.
- Roberto P.J. Perazzo, Laura Hernández, Horacio Ceva, Enrique Burgos, and José Ignacio Alvarez-Hamelin. Study of the influence of the phylogenetic distance on the interaction network of mutualistic ecosystems. Physica A: Statistical Mechanics and its Applications, 394(0):124 - 135, 2014.
- Yannick Leo, Eric Fleury, Carlos Sarraute, José Ignacio Alvarez-Hamelin, and Márton Karsai. Socioeconomic correlations in communication networks. In Fourth conference on the Analysis of Mobile Phone Datasets (NetMob 2015), <hal-01134295, MIT Media Lab, United States, April 2015.
- J. Ignacio Alvarez-Hamelin, Mariano G. Beiró, and Jorge R. Busch. Understanding Edge Connectivity in the Internet through Core Decomposition. Internet Mathematics, 7(1):45-66, 2011.
- Eduardo Mucelli Rezende Oliveira, Aline Carneiro Viana, Carlos Sarraute, Jorge Brea, Ignacio Alvarez-Hamelin, On the regularity of human mobility, Pervasive and Mobile Computing, Available online 27 May 2016, ISSN 1574-1192
- Mariano G. Beiró, J. Ignacio Alvarez-Hamelin, and Jorge R. Busch. A low complexity visualization tool that helps to perform complex systems analysis. New J. Phys, 10(12):125003, 2008.

### Teaching and Research Administration

Postdegree: Internet Topology (ITBA), Open Design Master Program (Humboldt-Universität and Universidad de Buenos Aires)

Degree: Introducción a la Ingeniería Electrónica (UBA), Comunicación de Datos (UBA)

Organization of: Workshop about Dynamics Networks 2013 (<http://cnet.fi.uba.ar/wdn13/>), Workshop about Dynamics Networks 2012 (<http://cnet.fi.uba.ar/wdn/>), Latin-American Workshop on Dynamic Networks (<http://cnet.fi.uba.ar/lawdn/>).

Referee of Physica A, since 2011; Computer Networks, since 2009; European Physical Journal B, since 2004; Computer Communications, since 2010.

### Research Supervision

2 PhD (CONICET, ITBA) + 1 ongoing (CONICET), 2 Master (UBA) + ongoing (UBA)

## SUMMARY CV

## Maria Eunice Quilici Gonzalez

Full Time Associate Professor at the Philosophy Department of the University of São Paulo State (UNESP), Brazil

**1) Academic background**

1977 – Bachelor's degree in Physics – University of São Paulo State (UNESP), Brazil.

1984 – Master's thesis in Logic and Philosophy of Science – University of Campinas (UNICAMP), Brazil.

1989 – PhD thesis in Cognitive Science, Language, and Linguistics – University of Essex, UK.

2004 – Postdoctoral research – Nijmegen Institute for Cognition and Information.

**2) Ten most relevant publications**

a. GONZALEZ, M. E. Q.; BROENS, M. C. On Denis Fiset's "Franz Brentano on higher-order of consciousness: A view from the complex system perspective. *Argumentos: Revista de Filosofia* (Online), v. 13, p. 85-94, 2015.

b. KOBAYASHI, G.; BROENS, M. C.; GONZALEZ, M. E. Q.; GONZALEZ, J. A. Q. The Internet of Things and its Impact on Social Relationships Involving Mutual Trust. In: *Proceedings, IEEE International Symposium on Technology in Society (ISTAS)*. Dublin: IEEE, v. 1. p. 1-6, 2015.

c. QUILICI-GONZALEZ, J. A.; BROENS, M. C.; GONZALEZ, M. E. Q.; KOYABASHI, G. Complexity and information technologies: an ethical inquiry into human autonomous action. *Scientiae Studia* (USP), v. 12, 2014.

d. BRESCIANI FILHO, E.; D'OTTAVIANO, I. M. L.; GONZALEZ, M. E. Q.; PELEGRINI, A. M.; ANDRADE, R. C. (Orgs.) *Auto-organização: estudos interdisciplinares*. 1ª ed. Campinas: Coleção CLE, 2014. v. 66.

e. GONZALEZ, M. E. Q.; MORAES, J. A. Complexidade e privacidade informacional: um estudo na perspectiva sistêmica. In: BRESCIANI FILHO, E.; D'OTTAVIANO, I. M. L.; GONZALEZ, M. E. Q.; PELLEGRINI, A. M.; ANDRADE, R. S. C. (Orgs.) *Auto-organização : estudos interdisciplinares*. 1ª ed. Campinas: Coleção CLE, 2014, v. 66, p. 155-174.

f. SILVEIRA, L. F. B.; GONZALEZ, M. E. Q. Instinct and Abduction in the Peircean Informational Perspective: Contributions to Biosemiotics. In: Vinicius Romanini, Eliseo Fernández (Eds.), *Peirce and Biosemiotics*, v. 11, p. 151-169, 2014.

g. QUILICI-GONZALEZ, J. A.; KOYABASHI, G.; BROENS, M. C.; GONZALEZ, M. E. Q. Ubiquitous Computing: Any Ethical Implications? *International Journal of Technoethics*, v. 1, p. 11-23, 2010.

h. GONZALEZ, M. E. Q.; BROENS, M. C.; D'OTTAVIANO, F. L. Abductive Reasoning, Information and Mechanical Systems. *Studies in Computational Intelligence (SCI)*, v. 64, p. 91-101, 2007.

i. GONZALEZ, M. E. Q.; HASELAGER, W. F. G. Creativity: surprise and abductive reasoning. *Semiótica*, v. 153, n. 4, p. 325-341, 2005.

j. GONZALEZ, M. E. Q. Information and mechanical models of intelligence: What can we learn from cognitive science? *Pragmatics and cognition*, v. 13, n. 3, p. 565-582, 2005.

**3) Ongoing research grants**

- Scholarship in Research Productivity 1D (CNPq), from 2016 to 2019. Research project: *Information, mechanismism and autonomous action in complex systems*.

Co-Principal investigator of the FAPESP project on *Systemics, self-organization and information* (Number 10/52627-9), coordinated by Itala Maria Loffredo D'Ottaviano

**4) Ongoing postdoctoral supervision (with scholarship)**

2016 - Edna Alves de Souza. *Information and Perception/Action: an analysis from the information realism perspective*. Postdoctoral research at the University of São Paulo State (UNESP), Brazil (PNPD/CAPES scholarship).

**5) Quantitative data**

a) Books organized: 14; b) Articles: 34; c) Book chapters: 35; d) Master's supervisions: 30; e) PhD supervisions: 4; f) Postdoctoral research: 3; g) Citations according to Google Scholar: 189 (since 2011); all citations: 510.

Link for My Citations (Google Scholar): ([https://scholar.google.com.br/citations?hl=pt-BR&user=JCYyVqIAAAAJ&view\\_op=list\\_works&gmla=AJsN-F4hmRT42s93ua81C07NIQJ5FvuyTVqf8sXXy3UtT3hOHmAdZDvK9Is4SvNxxwQpEhEQiLM3bX99cx8kayt3kAhJjiDb6ZD\\_0VSGdIcSIKZaAN5UHN4bsIeukBbS9OoqSXYX\\_Jj34&undo=untrash\\_citations.It3QlvBKHbMC.1AS7WB7zg6gC.4\\_yl7nwqy4oC.qwy9JoKyICEC.6DzZfY3qH5gC.AYInfyleIOsC&cstart=0&pagesize=20](https://scholar.google.com.br/citations?hl=pt-BR&user=JCYyVqIAAAAJ&view_op=list_works&gmla=AJsN-F4hmRT42s93ua81C07NIQJ5FvuyTVqf8sXXy3UtT3hOHmAdZDvK9Is4SvNxxwQpEhEQiLM3bX99cx8kayt3kAhJjiDb6ZD_0VSGdIcSIKZaAN5UHN4bsIeukBbS9OoqSXYX_Jj34&undo=untrash_citations.It3QlvBKHbMC.1AS7WB7zg6gC.4_yl7nwqy4oC.qwy9JoKyICEC.6DzZfY3qH5gC.AYInfyleIOsC&cstart=0&pagesize=20))

**6) Other relevant information**

- Latin-American representative at the Council of the Complex Systems Digital Campus - CS-DC - UNITWIN/UNESCO (International inter-university cooperation and networking), since 2014.

- Leader of the Research Group on Cognitive studies – GAEC (Ref. UNESP.0175) – CNPq, since 1989.

- Member of the Research Group on Self-Organization at University of Campinas (UNICAMP) Brazil.

- Member of the Brazilian Society on Cognitive Science, the Ecological Psychology Society, Complex System Society, and the International Society for Philosophy and Technology.

Complex Systems Digital Campus  
218 rue du Faubourg Saint Martin  
75010 - Paris

Paris, June 27<sup>th</sup> 2016

The CS-DC UNESCO UniTwin is devoted to worldwide launch the research and education on theoretical and experimental studies on complex systems. It is developing e-departments for its roadmap chapters on the theoretical transversal questions and e-departments for its roadmap chapters on experimental studies in the great classes of complex systems.

This project about opinion diffusion in emergent networks is *exemplar because it combine the theoretical study of emergent behaviour and the experimental study of social cognition*. The CS-DC will give to this project the opportunity to create an e-team attached both to the e-department of emergent behaviour and to the e-department of cognitive systems. *The e-team will thus have a permanent collaborative e-studio for its meetings, its international e-seminar and e-workshops. All these e-events can be recorded and be replayed both for research or education purpose.*

The permanent e-studio will thus assure that the project will create during its duration deep and permanent interactions between the involved students, the young researchers and seniors in three countries. The e-studio will also contribute to largely disseminate worldwide the main theoretical and experimental topics of the project. It can even allow this very important project to recruit other researchers in the world especially when the end of the project is arriving, for its continuation. In particular, an e-team is the good permanent place for postdocs to continue to work together after the end of the project even if they are changing of geographical positions.

There is no doubt that this excellent project and its CS-DC e-team will play an important role in the complex systems community not only during the duration of the project but also after the end of the project.



Prof. Paul Bourguine,  
President of the CS-DC UNESCO-Unitwin  
Honorary Director of the RNSC (National Network of Complex Systems)  
[paul.bourgine@polytechnique.edu](mailto:paul.bourgine@polytechnique.edu)

**ISC-PIF**  
113 rue Nationale,  
75013 Paris



David Chavalarias – Directeur  
[david.chavalarias@iscpif.fr](mailto:david.chavalarias@iscpif.fr)  
113 rue Nationale Paris 75013  
+33 6 22 94 64 60

Paris, June 15 2016

The Institute of Complex Systems of Paris Ile the France (<http://iscpif.fr>), is a unit of the National Center for Scientific Research (UPS3611), one of the largest french research organisation, and is in partnership with more than a dozen of Universities and other french research organisations, as well as with the city of Paris and Ile-de-France Region.

Interface between disciplines, but also between research organizations and higher education, the Complex Systems Institute of Paris Ile-de-France (ISC-PIF) is a place dedicated to the development of innovative and interdisciplinary research on complex systems at the crossroad between high performance computing and big data. It catalyzes, since 2005, the emergence of common and interdisciplinary practices, facilitating access to skills, work areas and pooled research resources.

As the director of ISC-PIF, I am pleased to attest that we supports the application of the project entitled: Understanding opinion and language dynamics using massive data (OpLaDyn) to the call “TransAtlantic Platform Digging into Data Challenge 2016”.

The aims of this project correspond to the research line promoted by the ISC-PIF, according to its Roadmap, and some of the French members are active collaborators to the activities of the ISC-PIF.

If selected, the OpLaDyn will take advantage of the technologies developped at ISC-PIF and will have access to the complex systems grid (part of the european complex systems V.O) for large scale simulations and the OpenMOLE middleware, as well as to Gargantext, an open plateforme for text-mining analysis. The institute may also host the working meetings and conferences organized by the group.

David Chavalarias  
Director of the ISC-PIF

A handwritten signature in black ink, appearing to be 'D. Chavalarias', written in a cursive style.

Répondre

Répondre à tous

Transférer

Supprimer

Déplacer

4  
7**Fwd: Digging - TAP - Letter of eligibility**

27 juin 2016 16:27

Maria Eunice Quilici Gonzalez &gt; Laura Hernandez, Mariana C. Broens

----- Forwarded message -----

From: **chamada\_digging** <[chamada\\_digging@fapesp.br](mailto:chamada_digging@fapesp.br)>

Date: 2016-05-30 9:53 GMT-03:00

Subject: Digging - TAP - Letter of eligibility

To: Maria Eunice Quilici Gonzalez <[gonzalezquilici@gmail.com](mailto:gonzalezquilici@gmail.com)>**Letter of eligibility**

Dear Prof. Dr. Maria Eunice Quilici Gonzalez

Thank you for submitting the consultation regarding your eligibility to the Trans-Atlantic Platform (T-AP) Digging into Data Request for Proposals (RFP) – 2016.

I am pleased to inform you that you have been considered ELIGIBLE for submitting a proposal at FAPESP under the scope of this Call.

Important: the Letter of Eligibility aims only to certify the eligibility of the PI (minimum qualification).

It does not ensure the eligibility of the full proposal. Before merit review of the proposal, there is a step for assessing the adherence of the proposal to the terms of the call and the rules for the requested grant modality.

The present letter is one of the mandatory documents to submit your final proposal at FAPESP. Please, note that proposals lacking this document will not be considered for review.

We look forward to receiving your application.

Scientific Directorate

FAPESP



## **WORK PLAN FOR POST-DOCTORAL GRANT**

**Title:** Big data, social dynamics and natural language

**Supervisor:** Maria Eunice Quilici Gonzalez

In this post-doctoral project, the grantee will participate in the Brazilian team for the research project “Understanding opinion dynamics and language evolution using massive data” (T-AP project). Specifically, the grantee will work with the other members of the Brazilian team on the part of the T-AP project that involves research on language evolution by means of “big data” analyses. The grantee will bring a knowledge of linguistics and semiotics to the investigation, will assist in the development of computational models of language dynamics, and will investigate the following question: Can big data be used to bring relevant evidence to bear in support of the hypothesis that natural languages have most of the characteristics of complex adaptive systems?

### **2. Objectives**

The main objectives of this research project are (1) to investigate, as part of the Brazilian team of the T-AP project, the development of models in computational social science, specifically models of language dynamics, by means of big data tools and in the context of the theory of complex systems, and (2) to investigate how and if evidence from Big Data analyses support the hypothesis that natural languages have the characteristics of complex semiotic systems.

### **3. Work Plan, Methodology and Chronogram of Expected Results**

The activities of the postdoctoral researcher as presented in this proposal will consist initially of bibliographical research and the study of the fundamental concepts of big data analysis, computational social science and the dynamics of language from the point of view of the theory of complex systems and semiotics. It should be noted that the literature on big data includes many non-academic sources, because, as Gandomi and Haider have observed, “The rapid evolution and adoption of big data by industry has leapfrogged the discourse to popular outlets, forcing the academic press to catch up.” (2015, p. 137).

The grantee should participate regularly in the activities of the Brazilian team of the T-AP project and, where relevant, in the activities of other members of the project based in Brazil (for example in the PUC research group of Pragmatism) and other countries. In addition, the grantee must participate in regular personal meetings with the grant supervisor.

The research activity will be divided into three phases in accordance with project objectives. For each phase of the project, a detailed work program shall be presented to the Brazilian team of the T-AP project, according to the schedule below. This work program will be jointly specified by the researcher and the grant supervisor.

As part of the interdisciplinary Brazilian team, the grantee will contribute his or her experience in the areas of linguistics, semiotics and the theory of complex systems. Thus it is expected that the grantee, in addition to academic training in linguistics, have some training in, or knowledge of, the theory of complex systems, semiotics and/or related fields.



| Activity/<br>Semester | 1. Big data analysis | 2. Computational social<br>science | 3. Dynamics of language from<br>the point of view of the theory<br>of Complex Systems and<br>Semiotics. | 4. Writing of the<br>results and of the final<br>report |
|-----------------------|----------------------|------------------------------------|---|---|
| 1st. semester<br>2017 | X                    |                                    | X   |   |
| 2nd. semester<br>2017 | X                    | X                                  | X   |   |
| 1st. semester<br>2018 |                      | X                                  | x   | X   |
| 2nd. semester<br>2018 |                      | X                                  | x   | X   |
| 1st. semester<br>2019 |                      |                                    | x   | x   |
| 2nd. semester<br>2019 |                      |                                    |   | x   |

## 5. References

BECKNER, C.; BLYTHE, R.; BYBEE, J.; CHRISTIANSEN, M. H.; CROFT, W.; ELLIS, N. C.; HOLLAND, J.; KE, J.; LARSEN-FREEMAN, D.; SCHOENEMANN, T. Language is a complex adaptive system: position paper. *Language Learning*, v. 59, s.1, pp. 1-26, dec. 2009.

GANDOMI, A.; HAIDER, M. Beyond the hype: big data concepts, methods, and analytics. *International Journal of Information Management*, v. 35, n. 2, pp. 137-144, 2015.

KRETZSCHMAR, W. A. *Language and Complex Systems*. Cambridge: Cambridge University, 2015

LARSEN-FREEMAN, D.; CAMERON, L. *Complex Systems and Applied Linguistics*. Oxford: Oxford University, 2008.

MICHEL, J. B.; SHEN, Y. K.; AIDEN, A. P.; VERES, A.; GRAY, M. K.; THE GOOGLE BOOKS TEAM; PICKETT, J. P.; HOIBERG, D.; CLANCY, D.; NORVIG, P.; ORWANT, J.; PINKER, S. Quantitative analysis of culture using millions of digitized books. *Science*, v. 331(6014), pp. 176-182, jan. 2011.

PEIRCE, C. S. The collected papers of Charles Sanders Peirce. Burks, A. W. Charlottesville: Intelix Corporation; Cambridge: Harvard University, 1931-1935. vols. I-VI, VII-VII.

RACZASZEK-LEONARDI, J., 2014. Multiple systems and multiple time scales of language dynamics: coping with complexity. *Cybernetics & Human Knowing*, v. 21, n. 1-2, pp. 37-52, 2014.

**Extract of the Code of Federal Regulations**  
**TITLE 45**  
**PUBLIC WELFARE**  
**DEPARTMENT OF HEALTH AND HUMAN SERVICES**  
**PART 46**  
**PROTECTION OF HUMAN SUBJECTS**

\*\*\*

**Revised January 15, 2009**  
**Effective July 14, 2009**

**Subpart A - Basic HHS Policy for Protection of Human Research Subjects**

**Authority:** 5 U.S.C. 301; 42 U.S.C. 289(a); 42 U.S.C. 300v-1(b).

**Source:** [56 FR 28012, 28022](#), June 18, 1991, unless otherwise noted.

**§46.101 To what does this policy apply?**

(a) Except as provided in paragraph (b) of this section, this policy applies to all research involving human subjects conducted, supported or otherwise subject to regulation by any federal department or agency which takes appropriate administrative action to make the policy applicable to such research. This includes research conducted by federal civilian employees or military personnel, except that each department or agency head may adopt such procedural modifications as may be appropriate from an administrative standpoint. It also includes research conducted, supported, or otherwise subject to regulation by the federal government outside the United States.

(1) Research that is conducted or supported by a federal department or agency, whether or not it is regulated as defined in [§46.102](#), must comply with all sections of this policy.

(2) Research that is neither conducted nor supported by a federal department or agency but is subject to regulation as defined in [§46.102\(e\)](#) must be reviewed and approved, in compliance with [§46.101](#), [§46.102](#), and [§46.107](#) through [§46.117](#) of this policy, by an institutional review board (IRB) that operates in accordance with the pertinent requirements of this policy.

(b) Unless otherwise required by department or agency heads, research activities in which the only involvement of human subjects will be in one or more of the following categories are exempt from this policy:

(1) Research conducted in established or commonly accepted educational settings, involving normal educational practices, such as (i) research on regular and special education instructional strategies, or (ii) research on the effectiveness of or the comparison among instructional techniques, curricula, or classroom management methods.

(2) Research involving the use of educational tests (cognitive, diagnostic, aptitude,

achievement), survey procedures, interview procedures or observation of public behavior, unless:

(i) information obtained is recorded in such a manner that human subjects can be identified, directly or through identifiers linked to the subjects; and (ii) any disclosure of the human subjects' responses outside the research could reasonably place the subjects at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, or reputation.

(3) Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures, or observation of public behavior that is not exempt under paragraph (b)(2) of this section, if:

(i) the human subjects are elected or appointed public officials or candidates for public office; or (ii) federal statute(s) require(s) without exception that the confidentiality of the personally identifiable information will be maintained throughout the research and thereafter.

(4) Research involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available or if the information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects.

(5) Research and demonstration projects which are conducted by or subject to the approval of department or agency heads, and which are designed to study, evaluate, or otherwise examine:

(i) Public benefit or service programs; (ii) procedures for obtaining benefits or services under those programs; (iii) possible changes in or alternatives to those programs or procedures; or (iv) possible changes in methods or levels of payment for benefits or services under those programs.

(6) Taste and food quality evaluation and consumer acceptance studies, (i) if wholesome foods without additives are consumed or (ii) if a food is consumed that contains a food ingredient at or below the level and for a use found to be safe, or agricultural chemical or environmental contaminant at or below the level found to be safe, by the Food and Drug Administration or approved by the Environmental Protection Agency or the Food Safety and Inspection Service of the U.S. Department of Agriculture.