Davinia Hernández-Leo, Konstantinos Michos (Eds.)

Meta-Research

Research Methods course

Master in Sound and Music Computing, Master in Intelligent and Interactive Systems, Master in Computational Biomedical Engineering, and Master in Wireless Communications

ICT Department, Universitat Pompeu Fabra,
Barcelona March 2018
Editors

Davinia Hernández-Leo
Associate Professor, Serra Húnter Fellow
Universitat Pompeu Fabra, Barcelona
E-mail: davinia.hernandez-leo@upf.edu

Konstantinos Michos
PhD Student, Teaching Assistant
Universitat Pompeu Fabra, Barcelona
E-mail: kostas.michos@upf.edu

Universitat Pompeu Fabra, Barcelona
e-repository UPF, http://repositori.upf.edu/

Licensed under Creative Commons Attribution-NonCommercial-NoDerivatives 4.0
This document collects a selection of papers written by master’s students in the context of the “Research Methods” course common to the Master’s Programmes in Sound and Music Computing, Intelligent and Interactive Systems, Computational Biomedical Engineering and Wireless Communications, of the Information and Communication Technology Department at Universitat Pompeu Fabra, Barcelona, during the 2017-2018 academic year.

The papers were written as part of an integrative assignment entitled “Meta-Research”, where students were expected to do a small piece of research about a transversal research topic. Students worked in teams and selected a topic, among those suggested in Table 1. A refinement of the topic, the particular research questions to study and the methodology to apply were proposed by the students and discussed with the course educators in tutoring sessions. A total of 15 papers were written by the students and presented in the classroom. Assessment included peer-review by students during the presentations and through a conference management program, assessment by the educators, and self-assessment. The results from the self-assessment were especially considered in the selection of the papers to include in this open document. Selected papers tackle among others interdisciplinarity in research, the influence of gender in science, open science and research-industry collaboration.

<table>
<thead>
<tr>
<th>Topics</th>
<th>Submitted</th>
<th>Accepted to publish</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interdisciplinarity in research</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Gender in science</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Open science</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Research integrity</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Social impact of Science</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Research-Industry collaboration</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Research careers</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>PhD process and life</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Measuring impact of research: impact factors</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>15</strong></td>
<td><strong>13</strong></td>
</tr>
</tbody>
</table>
Contents

Interdisciplinarity in research
Is interdisciplinary research ready to normalize within the conventional research institutions?
Ricard Castellà, Guillem Cañizares, Marc Mayans, and Martí Sánchez

Evaluation of Issues in Interdisciplinary Research
Davide Busacca, Joaquin Jimenez, Tae Hun Kin, and Lorenzo Romanelli

Gender in science
Analysis of sex distribution and glass ceiling effect in academic programs of UPF
Sara Andres Caballero, Silvia Badal Carsi, Mireia Masias Bruns, and Celia Penella Hernández

Analysis of gender ratio in authorship of medical Spanish journals using scopus database.
Mar Altaba i Rosa, Susana Amorós García de Valdecasas, Sarra Arranz Lombana, Irene Cuenca Ortolá, and Marina Echeverría Ferrero

Open science
Implications of various licensing modes in the context of open science
Emir Demirel, Furkan Yesirel, and Marc Kaivon Jones

Exploring relationships between open science and open education: Europe case study
Philip Tovstogan, Bezhad Haki, Martin Disley, and Aniel Rossi

Measuring impact of Open Science
Álvaro Quílez, Fabio Stefanizzi, Keith Kennedy, and Octavi Font

Social impact of science
Social impact of science and technology
Jorge Sevilla Ramos, Jordi Selva Vinagre, Jordi Casals Grifell, and Cristian Izquierdo Morcillo

Social impact of research: n-person games in real life
Marc Alvinyà, Alejandro Hinojosa, Albert Marin, Moscoso, and Lara Solà

Research Industry-Collaboration
A systematic review on common factors for successful, long-term University-Industry collaborations
Domingo de Abreu, Frida Hentschel, José Piñero, and Marielby Soares

Systematic overview of industry-academic collaborations in the field of artificial intelligence
Kami Artik, Enric Cosp, Sara Galindo Martinez, Juan Luis Moral Perez, and Siming Zhang

Phd process and life
The importance of supervisor during PhD-related crisis
Omar Ghatasheh, Bernat Puig, Borja Requena, and Emilia Vulpe

Research careers
Research as a career: common and specialized backgrounds
Ibai Genua, Jake Phillips, Sira Mogas, and Ariadna Mas
Is interdisciplinary research ready to normalize within the conventional research institutions?

Ricard Castellà, Guillem Cañizares, Marc Mayans and Martí Sánchez
Universitat Pompeu Fabra, Barcelona

Master in Wireless Communication
1ricard.castella01@estudiant.upf.edu
2guillem.canizares01@estudiant.upf.edu
Master in Intelligent and Interactive Systems
3marc.mayans01@estudiant.upf.edu
4marti.sanchez01@estudiant.upf.edu

Abstract. The most complex problems of today’s society are those that involve different scientific environments and require diverse research disciplines to work together. Interdisciplinary research has arrived to change the world, since it has become a hot topic in the current panorama. But, is interdisciplinary research ready to be exploited in nowadays research projects? In this article we provide a global picture of how this kind of research should be applied and how is actually used. Generally, we see that neither scientists or research institutions are ready to put in practice interdisciplinary research in a systematic way because of a lack of experience of institutions in this field.

Keywords: Interdisciplinary, Research, Institution Collaboration

1 Introduction

In the global world we live at nowadays, merging and combining different fields to improve the current systems is becoming more and more common. Research, has not been an exception. The word ‘interdisciplinary’ is compounded by the prefix ‘inter’ which means ‘in the midst of’ or ‘together’ and the noun ‘discipline’. So, interdisciplinary research (IR) can be understood the research technique that involves different disciplines to explain or develop a single topic, using the contribution of different points of view. Interdisciplinarity provides an important amount of benefits in research, such as the ability to ask questions and solve problems that have never come up before or to solve complex social problems. It can also address old problems, especially those that have proved unwilling to yield to conventional approaches. Despite these mentioned benefits, there are some disadvantages of employ this kind of research technique. IR could require more time, effort, imagination and capital than the traditional research methodologies does, and could also imply more risks to be unsuccessful. Furthermore, IR requires the researchers to accomplish with a set of skills needed to cooperate between research teams. Figure1 shows a node map provided by the Institute of Scientific Information that connects different scientific fields taking into account the citations similarities. This map provides evidences of the most probable fields to work together in order of their similarities. Of course, there are several fields of knowledge that are more susceptible to be fitted in an interdisciplinary research project than others. For example, Material Science, Chemistry and Environmental Technology are more suitable to work together [1], meanwhile it is more difficult to find similarities with Economics field.

Authors in [2] have studied the impact of interdisciplinary in research world and have done midterm prognostics predicting that, although its situation can gradually improve, interdisciplinary research is not seen with the same status as the traditional one from a publication point of view.
This paper tries to explain with a clear, easy and direct point of view if interdisciplinary research is ready to be normalized within the conventional research institutions, so to whom it concerns are the institutions and people directly related with research. This paper is organized as follows: section 2 introduces the chosen methodology to try to answer the question that this paper addresses. Section 3 shows the results obtained in our research, while conclusions are shown in section 4.

2 Methodology

With the aim to carry out a special contribution to the study field of the interdisciplinary investigation we look over the state of the art and the already existent literature in this field in order to get a new perspective or point of view. In this paper we are using an empirical methodology to develop our investigation over the interdisciplinary research. This methodology can be defined as the research based on the observations or evidences to test an hypothesis [4], it perfectly fits with the goal of this paper because there is not a theory or model to describe it so the hypothesis has to be validated according with the evidences. For this reason other methodologies, such as the scientific, the analytical or engineering method, have been discarded because alternative variations of the hypothesis are not tested, results are not contrasted with results from other observations and the results are not improved after the investigation, respectively. In addition, they would not contribute to answer our hypothesis: "is interdisciplinary research ready to be normalized within the conventional research institutions?"

In spite of its simplicity, the empirical methodology is powerful but sometimes presents limitations, since our knowledge is only based on evidences and observations. To develop the study of the hypothesis it will be used the known ‘historical methods’, where the data of journals or papers which talks about this topic will be saved and related with our hypothesis. With the aim to carry out the proposed empirical investigation it is needed to work using an empirical cycle which has the following steps [4]:

1. Observation: Search for papers, journals and projects related with the actual situation of the interdisciplinary research.
2. Induction: The hypothesis is formulated basing on the relevance of the previous observations.
3. Deduction: Where it is deduced if the created premises can drive us into some conclusions. Proves: The hypothesis is check with new data.
4. Evaluation: Analysis of the results obtained at the previous step.

A crucial step in this methodology is the process to find the correct observations to check our original hypothesis.
3 Results

Today, the interdisciplinary research is a hot topic. Solving important problems in our society requires international multidisciplinary teams of experts that can work together and giving a new point of view, contributing everyone with his own expertise. In this specific kind of projects, several skills are valuable for researchers profile due to the high dynamic range of fields and knowledge. Persons with adaptability and creativity to solve questions and problems are required for IR. Teamwork skills are also important to get on plural squads, as well as to be open minded, to try to understand the given problem by different ways and avoid establish premature bounding on possible solutions. In interdisciplinary research teams, communication is essential. A vital affair in IR is a proper project management. Due to the complexity of these projects and the wide range disciplines that they require, having a general knowledge of the whole project is not enough, it also important to understand all concrete areas, and have the ability to cooperate among different field experts. Interpersonal diplomatic skills and proactive attitude are required to manage IR projects in order have a better understanding of the application areas for the outcomes of the project.

In terms of the stakeholders for interdisciplinary projects, the majority of funding agencies don’t dedicate explicit programs for IR projects, however there are some countries which dedicates more resources specifically to interdisciplinary projects, as research centers or facilities. A practical example that is DFG in Germany, who works to facilitate interdisciplinary work by providing incentives for workshops [2]. The figure below shows the results of [1] study in 2015, they classify countries with major number of IR papers published.

![Figure 2: Publications in world’s top 10% of interdisciplinary papers (%) [1].](image)

This lack of funding may take place because of interdisciplinary research is still in a growing phase. The curve evolution of a project in IR and the production of outcomes requires a long period of time [5], and current institutions have to get used to this methodology yet. Several challenges or barriers that the IR projects have to deal with are creating an effective research team, identifying a framework around which to build the research, working within resource constraints or simply find a common vocabulary for communications across the different fields between others. The combination of diverse research programs also extends conceptual challenges when multiple researchers from different disciplines try to coordinate the logistics of carrying out interdisciplinary fieldwork [6]. Research activity is always limited by quantity of resources as time and money available. Interdisciplinary research may be harmed by the need of more incomes than the necessary for a similar single project.

Taking into account the impediments for IR to be normalized in today’s research, we can say that the current society it is not prepared yet to develop efficiently this type of studies [7]. More dedication has to be invested among institutions to move forward into the IR path.

4 Conclusions

The current situation of research is oriented to face the problems in a singular way, and this is not wrong in certain cases. However, in order to solve the more complex future problems, a more global approach has to be done. This is where interdisciplinary research stands out.

The traditional single-field in research has still a big influence among researchers and institutions, but the future needs will induce to introduce interdisciplinary structures and procedures. Different science field are now starting to collaborate in common points to face new challenges. Still, this methodologies are in growing phase. The transition to upgrade research institutions with
the basic principles of IR will require high effort and time. The new scientists generation will have to reach further challenges, and those will force us to collaborate all together.

References

Evaluation of Issues in Interdisciplinary Research

Davide Busacca  
Joaquin Jimenez Tae  
Hun Kim  
Lorenzo Romanelli

Sound and Music Computing  
davide.busacca01@estudiant.upf.edu  
joaquinfernandez Jimenez01@estudiant.upf.edu  
taehunkim01@estudiant.upf.edu  
lorenzo.romanelli01@estudiant.upf.edu

Abstract. Interdisciplinary Research can overcome the limits of traditional research approaches, but it presents issues that should be considered. Through a literature review, we aim to provide an overview of some of the issues that appear to be frequently found. The focus of this work has been put on fundings, communications, environment and academic impact. These topics turned out to be interconnected at different levels, thus making it difficult to conduct an analysis without considering all the aspects of the problems.

Keywords: Interdisciplinary, Research, Communication, Academic, Impact, Citation.

1 Introduction

Interdisciplinarity is an approach to different academic tasks which, in the recent decades, is gaining impact and accolades [1], especially in the field of research. This approach formally arose in the twentieth century in response to the progressive segmentation, isolation and compartmentalization happening across the plethora of fields of study. The underlying philosophy, however, is based on the concept of the unity-of-knowledge, which dates back to ancient Greece [2]. A more modern interpretation of the concept can be found in the work of Wickson et al.: “the researcher can seek to integrate the different knowledges by looking for coherence, correspondences and ‘ridges’ across the differences, generating knowledge by finding, identifying and communicating patterns across diverse disciplines and discourses.” [3]

Still, interdisciplinary research (IDR) has not a generalized and shared definition [1]. In fact, it can be defined in different ways [2,4,5,6], each one of them highlighting slightly different aspects of a bigger and more complex picture. In this work, we adopted the definition based on [7,8], which proposed that IDR is the mode of research that integrates concepts or theories, tools or techniques, information or data from different areas of knowledge.
One might inquire the advantages of adopting an IDR approach. One benefit relates to the innovative insights and methodologies that researchers from other disciplines can bring to a new field [9]. Another one consists in the oversight that researchers from different disciplines can bring to an ongoing project, by making sure that knowledge coming from their field is up-to-date. In addition, many complex or practical problems can only be understood by pulling together insights and methodologies from a variety of disciplines. [1]

While benefits for doing IDR exist, some drawbacks may arise. In this paper we aim to provide an overview about some of the issues that emerge when dealing with research with the interdisciplinary approach.

2 Research methodology

For this research, we used the literature research methodology. After reading, comparing and discussing information from different academic sources, we tried to draw our own conclusion. The paper that started our discussion is Ten Cheers for Interdisciplinarity [1]. The article explains benefits and possible barriers of IDR. However, it was written around 20 years ago, so the motivation to use it as a reference came from the idea that IDR may have changed since then. We took several approaches such as comparing with other sources to assess its validity or finding the views of other authors and institutions.

After discussing and referencing to similar ideas on other articles, we selected the topics that we believed are the most relevant and interesting. We decided four topics: Communication, Environment, Funding and Academic Impact. These four topics were represented with abundant amount of research whereas some others were not studied well or too philosophical to be discussed in scientific research. Having more than four topics was too complex for the scope of our research.

Because these topics are derived from many parts of disciplines, we decided to review it from many points of view. There are many entities involved in IDR such as institutions, companies, and researchers. Not all the information we found reached conclusions in general terms, as their studies are made in the context of their own experience or disciplines. Some articles perform research in general sense, but it may not apply to all disciplines of research. Interesting studies were found that helped us draw stable conclusions.

3 Results

Fundings

There is a study that shows that the probability of successful funding depends on the level of Interdisciplinarity. Namely, high level of interdisciplinarity shows less probability to be funded [10].
First, interdisciplinary research is hard to evaluate using the current assessment methodologies, because panel or reviewers in charge of grant evaluation are ill-equipped for assessing interdisciplinary projects [11]. For instance, in the case of the Australian research governance system, a formal classification is employed by the government to facilitate centralised forms of evaluation and funding of research, and this tends to disadvantage interdisciplinary project, which need more flexible evaluation systems [12].

Secondly, the amount of resources needed in projects increases for higher level of interdisciplinarity. In this case, a lot of effort is needed to build collaborative relationships, to develop a shared language and to hone a common perspective from disparate viewpoints [13].

Thirdly, following classical evaluation systems for higher interdisciplinarity projects is less effective, and puts interdisciplinarity as high risk research. A study in the UK showed that interdisciplinary research is associated with lower citation impact overall, but higher level of citations in patent applications[14]. More globally, the outputs of interdisciplinary projects may be fewer than those of projects with a narrower disciplinary focus[11,15].

**Communication**

The issue of an effective communication in science is not limited to the scope of this work, and is a very broad topic. However, for interdisciplinarity, many agree that communication in fact constitutes a barrier [16,17,18,19]. The causes are attributed to two factors: disciplinary jargon [9,20] and lack of a shared common language.

It is necessary for every discipline to rely on its own specific terminology [21]. As the philosopher Holbrook puts it, “we could say that disciplines each have – or perhaps are – something like their own language”. The outcome of his analysis is the definition of three answers to the problem of interdisciplinary communication, which we can refer to as (1) integrative consensus, (2) recognition of incommensurability, and (3) reflective invention [4]. In the first point, it is stated that interdisciplinary communication should aim at achieving reciprocal comprehension, shared knowledge, and consensus between actors from different disciplines. This “common understanding” view is in fact the dominant one [6,20].

However, what this hypothesis doesn’t consider is the possibility of incommensurability between disciplines, which arises from the differing basic orientations of each discipline toward the world. In other words, for an effective interdisciplinary communication to take place, one must adopt the point of view of a different discipline. Such perspective suggests that an interdisciplinary approach should be adopted when members from a specific field realize that they are not able to address a problem using only the resources native to that discipline, and thus a paradigm shift in the definition of the goal of interdisciplinary research itself is introduced.

The third possible answer to the issue states that such incommensurability arises only when attempts at communication fail, at which point a new ad-hoc language is necessary to overcome this gap. The lack of a “common language” in science is indeed an open question which has been addressed elsewhere. While some, like Tress [22], argue that creating a common language (if not universally shared across science, at least shared across a research team) is desirable, the general opinion seems to agree that this step may not be necessary [6,20], and in some cases it is even to be avoided, in order to avoid dumbing down what is one of the main advantages of interdisciplinary research, i.e. the possibility to draw from disciplinary knowledge [23].
Communication is an issue that should be addressed as early as possible during the development of an interdisciplinary research, when it is the most difficult to gain a common understanding of the problem. It is a process that takes time [6,9,20,22], and is only possible between individuals who risk their disciplinary identities and sacrifice them to the possibility of co-creating a new, shared genre of discourse [4,19].

Creating good environment

One of the challenges that interdisciplinary researchers face is creating good environment. IDR requires more attention for creating effective and friendly environment. According to our study, many researchers seemed to agree that interdisciplinary environment is affected in multiple layers: individuals, teams, and institutions.

IDR innately has higher barrier for incorporating researchers together. Turner et al. has discovered several types of inevitable "tensions" in interdisciplinary research when scholars from different disciplines attempt to produce one coherent work. [24] In their study, respondents had mentioned that interdisciplinary research environment is chaotic and fuzzy. However, Turner found that strategies to create good environment were not uniform, depending on the interest and structure of institutions. Many studies were conducted to find effective process in IDR. Siedlok et al. demonstrated different types of "interventions" in the level of individual researchers, team level, and institution levels [25]. Through these interventions, researchers could learn and perform better practices in IDR. Beyond the institution, Lyall et al. discussed the importance of the role of funding agencies in creating an effective environment [26]. Szostak's case study concluded that team building process may not be one static action, but can be repeated and reshaped as the research advances and encounter different requirements in a team [27].

Finally, there have been attempts to formulate a framework to facilitate IDR. Roux et al. suggested a framework that incorporates funders, researchers and users together to increase accountability of IDR [28]. König et al. presented different framework for IDR management at institutional level [29].

Academic Impact

In the research field, the fact that some publications have been highly cited confers them a special status as providers of important ideas in their area of speciality [30].
Results indicate that the impact of citation in publications is positively related to variety, but negatively related to balance and disparity. Evidence has been found that relationship between Interdisciplinary Research is not simple [31].

It is believed that being highly cited is associated with innovation and surprise, but this differs from one discipline to another. It is generally claimed that IDR is particularly relevant for addressing social problems and for fostering innovation [32]. The top highest cited papers ever are monodisciplinary and the top 7 belong to Biology area [33]. This result may vary by periods of time.

Table 1. Top cited articles ever [33]

<table>
<thead>
<tr>
<th>Rank</th>
<th>Authors</th>
<th>Title</th>
<th>Journal</th>
<th>Subject</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Lowry, O. H., Rosebrough, N. J., Farr, A.</td>
<td>Protein measurement with the folin. J. Biol. Chem.</td>
<td>Biology lab technique</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Laemmli, U. K.</td>
<td>Cleavage of structural proteins during PAGE. Nature</td>
<td>Biology lab technique</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Bradford, M. M.</td>
<td>A rapid and sensitive method for the molecular assay of nucleic acids. Biochem.</td>
<td>Biology lab technique</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Towbin, T., Staehelin, T., &amp; Gordon, J.</td>
<td>Electrophoretic transfer of proteins. Proc. Natl Acad. Sci. USA</td>
<td>Biology lab technique</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>LeC. C., Yang, W., &amp; Farr, R. G.</td>
<td>Development of the Calcofluor. (Proc. Rev. B)</td>
<td>Physical chemistry</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Thompson, J. D., Higgins, D. G. &amp; Gibson, G.</td>
<td>Clustal W: improving the sensitivity ofnucleic. Acid Res.</td>
<td>Bioinformatics</td>
<td></td>
</tr>
</tbody>
</table>

Institutionalization of research in terms of disciplines creates disadvantages when measuring the impact of IDR, because disciplines define 'gold-standards' in a field and suppress or marginalize methods, objects and concepts that don't belong to these standards. Many studies have found that particular perspectives are used to assess IDR [34]. The work of Levitt and Thelwall [35] found that number of citation of multidisciplinary journals was roughly 50% less than monodisciplinary articles.

The work of [31] studied the impact of interdisciplinary by analyzing and correlating the results of 62,408 articles with a minimum of 4 references. Three aspects were measured: 1. Variety (number of distinctive categories), 2. Balance (evenness of the distribution of categories), and 3. Similarity or Disparity (degree to which the categories are different or similar). Part of results of this analysis is presented in the Figure 1. Here we can see that in order to increase citation, the degree of IDR cannot be too low or too high.
4 Conclusions

Interdisciplinary research is a broad topic challenged with multiple issues. Our study has found that there is profound intricacy in the aspects of IDR. One must consider that the situation of IDR is different for each country, society, industry, institution, scientific field, and researcher. Although IDR is not a new concept anymore, opinions about its meaning, focus and value are not agreed upon. Many of the studies we found conclude that IDR brings benefits, but this is hard to leverage, measure and maximize.

For the funding issue, IDR has less probability to be funded than monodisciplinary research. Because it is harder to evaluate with current methods, it requires more resources, and is perceived as more risky. Thus, funding is evaluated in terms of the level of interdisciplinarity. Communication is a critical problem that has to be addressed early in the process, in order to reduce to the minimum field-specific terminology, lay a common understanding of the problem, and avoid misunderstandings. Creating a good environment for IDR depends on many factors such as a size of research, a characteristic of institution, funding agencies, and researchers themselves. Every stakeholder of research must participate in creating a positive and stable environment. For the academic impact, studies we found don’t come to a clear conclusion about evaluation and generalization of such research. Too
much degree of IDR is not positive, and too little may be hard to define, as the boundary between mono and IDR is blurry.

While it is possible to address these issues separately, we observed that there exist links between them that influence one another. From the information presented, it is clear that there is a strong relationship between the above issues, especially between communication and environment. Communication is the link between disciplines, which is essential for creating adequate environment. The relationships between above issues might not be always unilateral. During the IDR process, communication might lead to a shift to the definition of IDR goals. While many funding institutions tend to discourage IDR, they perhaps should participate into creating effective environments for IDR. We couldn’t find evidence of accurate evaluation of academic impact. This might result in negative funding chances when agencies make decisions based on academic impact of IDR.

We analyzed four issues, which already created complex and conflicting relationships regarding IDR. With other issues combined, the complexity will increase. It is important to understand for different parties of IDR that it cannot be judged with the same criteria of single disciplinary research. If we evaluate IDR with a monodisciplinary approach, we will not be able to understand it correctly nor completely. Thus, if we wish to shed more light about the issues related to IDR, more meta-research is required, taking into account the dynamic nature of IDR.

References

13. Haythornthwaite, C., Lunsford, K.J., Bowker, G.C. & Bruce, B. In New Infrastructures for Science
14. Pan,L.,&Katenko,S.AReviewofh5metricsinterdisciplinaryresearchusingacitation-based
15. Laudel, G.Conclave in the Tower of Babel: how peers review interdisciplinary research proposals.
16. Stoddart,D.To claim the high ground: Geographical for the end of the century. Transactions of
18. Bark,R H.,Kragt,M. E.,&Robson,B. J. Evaluating an interdisciplinary research project: Lessons learned for
(2016)
L. Tools for enhancing interdisciplinary communication. Sustainability: Science, Practice, & Policy, 7(1).
(2011)
22. Tress, G., Tress, B., & Fry, G. Analysis of the barriers to integration in landscape research projects. Land
use policy, 24(2), 374–385. (2007)
scholarship: navigating challenges in affect, epistemologies, and structure in environment–society
26. Lyall,C.,Bruce,A.,Marsden,W.,Meagher,L.: The role of funding agencies in creating interdisciplinary
reflection on the accomplishment of transdisciplinary research programs. Environ. Sci. Policy. 13,
733–741 (2010)
30. Garfield, E., Malin, M. V., & Small, H. Citation Data as Science Indicators. In Y. Elkana,
New York: Wiley (1978)
31. Yegros-Yegnas A, Rafols,I, D’EstePdoes Interdisciplinary Research Lead to Higher Citation Impact?
32. Chavarro, D., Tang, P., Rafols, I. Interdisciplinarity and local issue research: evidence from a developing
Publishing Group. (2014)
34. Mallard G, Lamont M, Guettzkow J. Fairness as Appropriateness : Negotiating Epistemological
35. Levitt, Thelwall M. Is multidisciplinary research more highly cited? A macro level study. Journal of
Analysis of sex distribution and glass ceiling effect in academic programs of UPF

Sara Andrés Caballero, Silvia Badal Carsi, Mireia Masias Bruns, Celia Penella Hernández

CBEM
sandrescab@gmail.com, silvia.4994@gmail.com, mmasiasbruns@gmail.com, celiangenell@gmail.com

Abstract. During the second half of the last century, women have been breaking barriers in the access of higher education programs, until a point where they have even become more than half of the students’ community in many countries, with equal or larger attainments than men. Nevertheless, there is still observed a segregation on the field and many authors suggest that the glass ceiling effect prevail in the professional level. With data of Universitat Pompeu Fabra (UPF), we are aimed to determine whether this tendency is also occurring in different levels and fields of study.

Keywords: Breaking Barriers, Higher Education, Gender, Ceiling Glass, Segregation.

1 Introduction

Over the past 40 years, there has been observed a large increase of women enrolling and graduating in higher education programs: while around 1975 undergraduate women represented about a third of the total; by the end of the millennium the percentage of women students was about 50%. Nevertheless, by that time, some authors reported significant differences in the career choices and percentage of degree attainment depending on gender [1][2], with all the implications that this may have entailed and with a probable association to worse conditions in women professional prospects.

Several efforts had been done to identify factors contributing to this difference and some authors suggested a combination of biological and psychological factors to account for this gap. As stated by Eurydice in 2010 [2], it has been observed that girls score better in general knowledge, spelling and writing when compared to boys, in early stages of education; but, from the adolescence boys tend to score better in spatial skills, perception and mathematical assessment. Whichever the reason for this difference is –whether genetic, hormonal or a strong socio-economic influence factor, according to different authors [3]–, they then hypothesized that female students tend to attribute this failure to a lack of ability to such fields and, thus, they tend to avoid
this technical-subjects that would later let them get enrolled into more technical careers, predominantly male-dominated.

McNabb et al. [1] supported the idea that women are less likely to enroll to more technical-like careers and tried to model and identify the causes of attainment differences depending on gender. Nonetheless, they concluded that gender itself did not represent a factor highly influencing the outcomes of individual careers but instead, aptitude, maturity, material status and many other socio-economic factors (such as parents’ professions) influencing individuals could explain the gap.

In the light of these results, huge efforts had been performed to evolve towards a more equally-distributed education model, through policy measures and awareness campaigns; reaching a point in which, nowadays, some countries have even claimed a growing imbalance towards an under-representation and lower-attainment of male students with respect to women [2]. Notwithstanding, even though what would be expected, many studies still report a glass-ceiling for women in their professional careers [5].

Considering the previous evolution in other European countries, in the following pages, the authors of this report will try to determine whether this trend can also be observed in Universitat Pompeu Fabra; and if this glass-ceiling can be observed in professional academic careers.

2 Research methodology

The source of information for this study is the official website of the Universitat Pompeu Fabra (UPF), specifically the section known as UPF in figures [6].

The dataset used covered a period of five years, from 2011 to 2016, and the variables included were the number of students enrolled in the different years divided by bachelor’s degree, master’s degree and doctoral degree (PhD).

The methodology applied consisted of a temporal analysis conducted by sex of the evolution of the number of students in each academic degree structured by field of study (e.g. Experimental vs Health Sciences.).

The dataset required some preprocessing since there was not any structure or aggregation. Both the preprocessing of the information and the temporal analysis were applied with the software Microsoft Excel 2016, which was also used to construct the results tables and present the numerical information.

Glass-ceiling term was first used in 1986 in a Wall-Street Journal to refer to invisible barriers that many women are claimed to be exposed to within organizations, limiting their professional careers progression.
The main strengths of the proposed research methodology are the reliability of the dataset and the reproducibility of the results obtained. The simplicity of the analysis is an advantage since it allows to interpret and reproduce the results easily.

About the limitations of the methodology, the most relevant is the fact that the analysis only included the data from one university in Spain, therefore there is no guarantee that the results are representative. On the other hand, the dataset itself presents some inconsistencies since for each year different degrees and masters studies were included and the PhD data is only available for one year.

3 Results

In this section, the results obtained will be presented.

![Number M/F Degree 2016-17](image1)

![Number of M/F Master 2016-17](image2)

![Number of M/F PhD 2016-17](image3)

**Figure 1. Number of males/females in the different academic degrees divided by field of study in the period 2016-2017.**

In the academic level of degree, the number of females in Experimental and Health Science is much higher than the number of males and the behavior is the opposite for ICT studies. As the academic level increases, specially in the PhD studies, these differences decrease (Figure 1).

![% of M/F 2016-17](image4)

**Figure 2. Percentage of male/female individuals by academic degrees in the period 2016-2017.**

In the analysis of the ratio male/female by academic degrees, the number of female students is larger in the degree and master studies. However, in the PhD level the
number of females decreases and the proportion of both sexes is approximately the same (Figure 2).

Figure 3. Percentage of male/female individuals in the different fields of study in the period 2011-2016.

Studying in more detail the distribution of sexes by field of study, the number of females is always higher in Health Science. On the other hand, in the Engineering studies, there is a greater proportion of males in all cases but Biomedical Engineering, where the ratio is more equilibrated (Figure 3).

Figure 4. Number of males/females in the different fields by bachelor degrees in the period 2011-2017.

Figure 5. Number of males/females in the different fields by master degrees in the period 2011-2017.
The temporal evolution in the degree studies does not show significant variations (Figure 4). In the case of the masters, there is a sudden rise in the number of students because new masters are included in the dataset.

4 Conclusions

As previously reported on the introduction, over the last 40 years the number of women accessing to higher education has increased significantly in all fields of study, representing more than half of the current UPF students. However, taking a closer look to the figures, it can be addressed that at degree level there is still a large segregation in the different areas of study: while the health sciences and humanity studies are dominated by women, the technical careers are still lead by men. This means that there is a large pathway to go to understand the causes of such inequality and the need for a change in the promotion of those fields.

For master’s programs the tendency is the same although it is observed an increase of female students in the ICT field due to the introduction of technical health related programs such as the CBEM. Pure technical studies maintain the male preeminence.

Comparing the evolution of the male/female ratio in PhD studies it seems that the glass ceiling effect is slightly observed in professional academic careers as it has been described for non-academic careers in general. This can be addressed in the health science departments, where the ratio of PhD students is close to 50% while there is a clear female dominance on bachelor and master’s degrees.

The ceiling effect observed on this study can not be proved with the available data and a larger dataset including more sociological features and/or hiring policies would be required to completely verify the hypothesis.

References

Analysis of Gender Ratio in Authorship of Medical Spanish Journals using Scopus Database

Mar Altaba i Rosas, Susana Amorós Gª de Valdecasas, Sara Arranz Lombana, Irene Cuenca Ortolá, Marina Echeverría Ferrero
Computational Biomedical Engineering
mar.altaba01@estudiant.upf.edu, susana.amoros01@estudiant.upf.edu, sara.arranz01@estudiant.upf.edu, irene.cuenca02@estudiant.upf.edu, marina.echeverria01@estudiant.upf.edu

Abstract. Numerous studies in different countries shed light on the unequal representation of men and women in leadership positions and scientific recognition. In this paper, a study was performed to examine the female presence in both authorship and editorial boards of 5 Spanish medical journals, which were selected based on their impact factor in JCR. The authorship of a total of 75 articles was analyzed using Scopus database. Results show that men are more likely to publish in medical field, as well as to appear as a first author. Moreover, women are fewer represented in editorial boards and, generally, present a lower h-index. Taking into account these results, it can be concluded that though the proportion of females is moderately increasing over time, it is still clearly below the level of fair equity.

Keywords: Gender disparities, research, medicine, authorship, inequality, gender ratio, Scopus, Spain.

1 Introduction

The history of women in medicine is marked by challenges and achievements. Although the role of women in the “art of healing” can be traced back through time, traditionally only male figures are highlighted [1].

Over the past four decades, the number of female physicians has increased from less than 10% of the medical students and less than 15% of practicing physicians to being the majority of new graduates in medicine and more than 40% of practicing physicians in Western countries [2]. In Spain, the number of female doctors has raised considerably along the years (see Figure 1.1). Indeed, the proportion of female registered physicians is greater than that of men in most of the working age groups, and there is a 70% female presence.
among the under 35 year-old doctors. In addition, in the present year (2017), 63% of all the candidates attending the MIRexam were women [3].

Despite their increasing numbers in medical training, women are still underrepresented in academic medicine leadership positions. Recent studies have shown significant differences in medical positions held by men and women [5, 6] manifesting the so-called ‘leakypipeline phenomenon’, that is, a disproportionately low number of women achieving leading medical positions. Similarly, along the history of science, women have participated in remarkable medical achievements yet they have not been given the credit they deserved and their discoveries remained forgotten due to their male counterparts [7]. Furthermore, there is evidence of a growing interest in medical investigations involving the topic of gender disparities, since Medline/PubMed\(^1\) database comprises a larger number of publications related to gender in the preceding years, compared to the proportion among older publications, as shown below in Figure 1.2.

---
In light of this, we have conducted a study to examine the female presence in the context of medical research. Numerous studies in different countries shed light on the inequality representation of men and women in leadership positions and scientific recognition. In Spain, this type of studies have been also carried out by one of the most relevant scientific societies, SESPAS\[8\].

For this purpose, we have selected leading Spanish medical journals (with internationally recognized quality standards) and we have focused on the authorship of several publications. The study represents an attempt to fill the gaps in the current understanding of sex differences in the attainment of upper level among academic medicine positions, and was designed to answer the following question: Are women physicians in academic medicine as likely as their male counterparts to write relevant scientific publications in Spain?

The final aim of our study is not only to examine the current status of gender distribution on medical research from a statistical point of view, but also to bring a critical reflection on the subject of gender disparity in science.

2 Research methodology

The selected approach to perform the study is explained in the present section. Journal Citations Reports (JCR) has been used to select the journals in which the analysis is based on. First of all, in order to narrow the search the journals considered are exclusively from Spain (thus, country filtration was needed), and from the Medicine field. The field decision was made because of its relation with our Master’s area (CBEM). Then, the first 5 journals with major impact factor were selected.
After that, Scopus was used to unify the methodology of the analysis and taking advantage of the searching tools of this database, top 5 most cited papers of each journal have been selected. Taking into account the first year of publication of all journals, three different time intervals of 6 years each have been determined to perform the analysis: 1999 - 2004, 2005 - 2010 and 2011 - 2016.

When data was not available in Scopus, an alternative search in Google Scholar was carried out following the same methodology explained above. This was the case of ‘Emergencias’ journal, due to the fact that only information from 2010 to present time was found.

Thereafter, authorship of each paper was analyzed. The complete name of each author was obtained from its Scopus database profile and, with that information, the author’s sex was determined. In the event of finding an ambiguous name, the API known as ‘genderize.io’ was used. This tool allows to obtain the sex probability for a given name and nationality. However, it was not always possible to obtain the complete name of an author, thus that cases were classified as unknown.

The data obtained from the authorship of each paper consisted on:
- Number of male authors
- Number of female authors
- Unknown authors
- Percentage of women among authors, calculated as:

\[
\text{Percentage of women} = \frac{\text{Number of female authors}}{\text{Total number of authors - unknown authors}} \times 100
\]

Furthermore, the sex of the first author of each article and its h-index was also included in the analysis. Then, mean h-index of both male and female authors was calculated, as well as the total number of male and female authors (for all papers of all time-intervals). Finally, the directive committee of each journal was analyzed by determining the sex of the chief editor and associated editors.

A summary of the methodology can be seen in Figure 2.

![Figure 2.1 Stage 1: articles selection](image)

**Data obtained**
- n° male authors
- n° female authors
- % female authors
- 1st author sex
- 1st author h-index
Results have been organized in tables and graphs to clarify its interpretation and ease the understanding.

For this type of analysis, the number of articles studied (a total of 75 papers) is considered significant and representative. However, it would be interesting to expand the research to a greater amount of papers and journals. On the other hand, the first 5 journals with the major impact factor in Spain have been considered, as well as the most cited papers (see annex 1) in different time intervals, so that the obtained results can be a good representation of the presence of women in Medicine journals in Spain over time. Some limitations of the analysis are the fact that some authors’ sex is unknown and the fact that sometimes their sex is based on a probability.

3 Results

A total of 577 authors have been examined classifying them by their sex (male, female or unknown). For each of the 75 papers analysed, the sex (M or F) and the h-index (1:95) of their first author have been extracted, and the proportion of the number of females over the total number of authors has been computed. Table 1 gathers the results obtained (extended table in annex 2).

Figure 2.2. Stage 2: authorship analysis (for each article)
In addition, the Editor in chief and the Associate editors of each of the five journals have been classified by their sex (M or F) as shown in Table 2. A total of 6 Editors in chief and 22 Associate editors have been used for this study.
### Table 2. Editorial board classification by gender for each journal.

<table>
<thead>
<tr>
<th>Journal</th>
<th>Editor-in-chief</th>
<th>Associate editors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>REVISTA ESPAÑOLA DE CARDIOLOGÍA</strong></td>
<td>Ignacio Ferreira</td>
<td>Emad Abu-Assi M</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Miguel Ángel Arias M</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pastora Gallego F</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ángel Sánchez-Recalde M</td>
</tr>
<tr>
<td><strong>AIDS REVIEWS</strong></td>
<td>Vincent Soriano</td>
<td>Anne-Mieke Vandamme F</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Walid Heneine M</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Genoveffa Franchini F</td>
</tr>
<tr>
<td><strong>JOURNAL OF INVESTIGATIONAL ALLERGOLOGY AND CLINICAL IMMUNOLOGY</strong></td>
<td>A.G. Oehling</td>
<td>I. Dávila M</td>
</tr>
<tr>
<td></td>
<td>M.L. Sanz</td>
<td>P.M. Gamboa M</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>R. Lockey M</td>
</tr>
<tr>
<td></td>
<td></td>
<td>J.M. Olaguíbel M</td>
</tr>
<tr>
<td></td>
<td></td>
<td>J. Sastre M</td>
</tr>
<tr>
<td></td>
<td></td>
<td>J.M. Zubeldia M</td>
</tr>
<tr>
<td><strong>EMERGENCIAS</strong></td>
<td>Oscar Miró i Andreu</td>
<td>Guillermo Burillo Puzte. MD, M</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>Agustín Julián Jiménez. MD, M</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Francisco Javier Martín Sánchez M</td>
</tr>
<tr>
<td><strong>ARCHIVOS DE BRONCONEUMOLOGÍA</strong></td>
<td>Esther Barreiro</td>
<td>Victor Bustamante M</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>José Luis López-Campos M</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Xavier Muñoz M</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ana Menezes F</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hugo Neffen M</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Rogelio Pérez-Padilla M</td>
</tr>
</tbody>
</table>

Information gathered in Table 1 is used for creating the following 3 figures. The gender proportion of the most cited authors is shown in Figure 3.1. From the 577 authors, 414 (72%) were male and 148 (26%) female, while 15 (2%) remained unknown. Similarly, the gender proportion of these same authors is shown in Figure 3.2 distributed in the three different time periods used for this study analysis.
Figure 3.1 Authors’ gender proportion in the 75 most cited papers.

Figure 3.2 Authors’ gender proportion in the 75 most cited papers distributed in time periods: (a) 1999 – 2004, (b) 2005 – 2010 and (c) 2011 – 2016.

Likewise, 75 h-index of the first author of each paper have been classified by gender and sorted in descending order; the average of this h-index has been computed for male and females as shown in Figure 3.3.
Figure 3.3. First authors’ h-index classified by gender (males h-index and females h-index). Dotted lines stand for each gender average values.

Additionally, pulling together the information in Table 2, the editorial board gender proportion is presented in Figure 3.4, distinguishing between the Editors in chief and the Associate editors positions.

Figure 3.4. Editorial board gender classification: (a) Editors in-chief and (b) Associate editors.
4 Conclusions

The results of this study show that male authors notably outnumber female authors in the medical research field. Although women and men begin their scientific careers in fairly equal numbers, the proportion of female medical researchers to male in Spain is far from being balanced. This proportion is moderately increasing over time, having reached a 28% between years 2011 and 2016. However, it is still clearly below the level of fair equity.

The fact of finding a higher presence of male authors in the analysed papers might have an effect on the h-index of the first authors, where a clear dissimilarity between both sexes is revealed. The mean of h-index of males is approximately 27 while the females' h-index means barely reach 20. Assuming a relation between author's sex and their acknowledgement might be too hazardous but it is still a possibility that women have less credibility due to their shortage in the field.

Gender disparities in career opportunities and contract conditions could be partly related to age or family issues. Many women participate more actively in research in the early stages of their career, but lower their professional ambitions in order to raise children. Some studies state that female researchers are more likely to hold temporary contracts or scholarships, which often means lower job satisfaction and higher levels of stress [9]. This is also reflected in our study, where the leading positions in the editorial board of the most relevant journals also bring to light a clear predominance of men over women.

Achieving gender equality remains an important challenge for the scientific community at large. Improvements, in terms of policies and staff management, are necessary.

Some initial steps have already been taken in Europe to strengthen the contribution of women to research development and decision-making. First, a mentoring program has been designed to fit the needs and expectations of female scientists. Second, specific criteria are regularly applied both at early and advanced career stages to receive training and mobility funds. Third, a public event was organized at the Parco Tecnológico Padano in June 2007 [9] to highlight the 'gaps' in opportunities that still hamper the progress of women in science to members of the scientific community.

These are all appropriate starting points to make gender not only a matter for individual careers but also a structural concern for research institutions and teams in general.
References


67. Llorens, P., Escoda, R., Miró, Ò., Herrem-Puente, P., Martín-Sánchez, F.J., Jacob, J., Garrido, J.M., Pérez-Durá, M.J., Gil, C., Fuentes, M., Alonso, H., Muller, C., Mebazaa, A. (2015). Characteristics and clinical course of patients with acute heart failure and the therapeutic measures applied in spanish emergency departments: Based on the EAHFE registry (Epidemiology of Acute Heart Failure in Emergency Departments) [Características clínicas,
terapéuticas y evolutivas de los pacientes con insuficiencia cardiaca aguda atendidos en servicios de urgencias españoles: Registro EHAF (Epidemiology of Acute Heart Failure in Spanish Emergency Departments). Emergencias, 27(1), 11-22.


## Annex I

### REVISTA ESPAÑOLA DE CARDIOLOGÍA

- 1999-2004: References 10-14
- 2005-2010: References 15-19
- 2011-2016: References 20-24

### AIDS REVIEWS

- 2005 – 2010: References 30-34
- 2011-2016: References 35-39

### JOURNAL OF INVESTIGATIONAL ALLERGOLOGY AND CLINICAL IMMUNOLOGY

- 1999 – 2004: References 40-44
- 2005 – 2010: References 45-49
- 2011-2016: References 50-55

### EMERGENCIAS

- 1999 – 2004: References 56-60
- 2005 – 2010: References 61-65
- 2011-2016: References 66-70

### ARCHIVOS DE BRONCONEUMOLOGÍA

- 1999 – 2004: References 71-75
- 2005 – 2010: References 76-80
- 2011-2016: References 81-85
<table>
<thead>
<tr>
<th>PERIOD TIME</th>
<th>AUTHORSHIP</th>
<th>FIRST AUTHOR</th>
<th>INDEX</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MALE</td>
<td>FEMALE</td>
<td>UNKNOWN</td>
</tr>
<tr>
<td></td>
<td>59</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>REVISTA ESPAÑOLA DE CARDIOLOGÍA 1999-2004</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>2005-2010</td>
<td>1</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>36</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>2011-2016</td>
<td>1</td>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>AIDS REVIEWS 1999-2004</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2005-2010</td>
<td>1</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2011-2016</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>JOURNAL OF INVESTIGATIONAL ALLERGOLOGY AND CLINICAL IMMUNOLOGY 1999-2004</td>
<td>1</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>2005-2010</td>
<td>2011-2016</td>
<td>Total</td>
</tr>
<tr>
<td>-------</td>
<td>-----------</td>
<td>-----------</td>
<td>-------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>13</td>
<td>6</td>
<td>19</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>35%</td>
<td>14%</td>
<td>41%</td>
</tr>
<tr>
<td></td>
<td>35%</td>
<td>81%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>40%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>25%</td>
<td>78%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>35%</td>
<td>14%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>25%</td>
<td>20%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20%</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>33%</td>
<td>33%</td>
<td></td>
</tr>
<tr>
<td>Year</td>
<td>Male (N)</td>
<td>Female (N)</td>
<td>Males</td>
</tr>
<tr>
<td>------</td>
<td>----------</td>
<td>------------</td>
<td>-------</td>
</tr>
<tr>
<td>2011-2016</td>
<td>4</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>2011-2016</td>
<td>1</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>16</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>22</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>135</strong></td>
<td><strong>24</strong></td>
<td><strong>1</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Implications of Various Licensing Modes in the Context of Open Science

Emir Demirel, Furkan Yesiler, Marc Kaivon Jones
MS Sound and Music Computing · Universitat Pompeu Fabra
emir.demirel@upf.edu, myesiler@ku.edu.tr, kaivon@virginia.edu

Abstract. There are many licensing models available to scientists and researchers whose work is primarily in the open-source domain. Each model has its own distinct set of regulations, leading to a variety of consequences. Their intricate variations complicate collaborative work, access to information, as well as public understanding of their implications. Examining copyright alongside alternative licensing models reveals their intercorrelational legal, economic, and social impact on both creators of work and users alike.

Keywords: open science, licensing, copyright, copyleft, public domain.

1. Introduction

Today we have three main modes of intellectual property (IP) licensure: private licenses, open licenses, and unlicensed. In products with private licenses, "the creator has the absolute right to decide individually about every utilization of their work."[3] In products with open licenses, depending on the license type, grant people the freedom to copy the product and use it in a way that they choose, with some restrictions. On the other hand, products with no licenses are under no conditions for people to use them without any conditions. Since the advent of copyright litigation, the complexity of IP licensure has grown exponentially alongside the proliferation of the internet. The same technologies that facilitated web connectivity and subsequently magnified the interpersonal channels by which we exchange information, can also be attributed to enhancing access of IP. Whether obtained through openly available information found in the public domain, or through traversing insecure barriers, the use cases of diversely licensed IP have ramifications that often reach scopes outside the knowhow of most IP creators and users alike. Problems tend to arise in collaboration of cross-licensed technologies. The limitations copyright created, prevented a worldwide community of researchers from building atop the state-of-the-art. In the face of this problem, open science movements have worked to lower the barriers of collaboration and access to IP. Here we seek to better understand the legal, economic, and social implications of the various modes of licensing in the context of open science.
2. Research methodology

This research aims to successfully identify and categorize the implications of varying models of IP licensing. We use a historical method to investigate the implications of various licensing models leveraging qualitative data obtained from previous studies near or related to this topic. Moreover, the goal of our method applied in this research is to clearly determine the trends and outcomes of utilizing different common open source licensing models. Before delving into an analysis on open source licenses and their ramifications, we must first define “open source.” According to the Open Source Initiative [7], the list of criteria that the distribution terms of an open-source software must comply with can be seen in Table 1:

<table>
<thead>
<tr>
<th>Table 1: Open Source Initiative Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Free distribution of software</td>
</tr>
<tr>
<td>- Free access to the source code (just reproduction costs are covered)</td>
</tr>
<tr>
<td>- Authorization of modifications and the distribution of derived works</td>
</tr>
<tr>
<td>- No discrimination between people and fields of endeavor</td>
</tr>
<tr>
<td>- No restriction on other software use</td>
</tr>
<tr>
<td>- Technological neutrality (independence from a specific product)</td>
</tr>
</tbody>
</table>

While the idea behind copyright is fairly straightforward, the differentiation between private, public licensing and open licensing lead to murkier waters. From Apple’s Public Source License to Sam Hocevar’s “Do What The Fuck You Want Public License”, there are a wide variety of open source licenses, all with varying degrees of restrictiveness. Even though their use cases often overlap, productions that are unlicensed and those classified under public domain or not synonymous. The main idea of open licenses is that they allow users to copy the document, change it, distribute the changed versions but those versions have to be under the same license as the original document. Where open licensing works to ensure the access to the work cannot be privatized in the future, unlicensed work (aka “no known copyright”) allows for commercial redistribution.

Considering the large number of open source licenses in existence, we provide a brief description for the most common models for open-source licensing in this section. These different models can be distinguished according to several criteria.
Probably the most obvious one of these criteria is the origin of the license, being either Free Software Foundation (FSF) or Open Source Initiative [6].

2.1 Open Source License Models

2.1.1 GNU General Public License

As seen from Table 1, an open source software is allowed to be used, modified and distributed freely with the obligation of authorization from the creator. In the case of completely free softwares (codes, scripts that can be found online without any licenses or references), authorization is not necessary. Probably the most obvious and most commonly used example is the GNU General Public License, which is completely free to use, distribute and modify. In general, softwares with such licensing is called 'freesoftware', instead of 'opensource software'.

2.1.2 Creative Commons

Creative Commons (CC) is a non-profit corporation founded in 2001 by Stanford University with the intention of facilitating the sharing of and building upon the work of others. CC offers practical methods of protection of the right of subsequent access to the work. Moreover, CC makes possible for a user to generate a specific licensing model according to his/her personal needs. These different terms of conditions are listed in Table 2:

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Allows others to copy, distribute, derive works based upon the software</th>
</tr>
</thead>
<tbody>
<tr>
<td>ShareAlike</td>
<td>Distribution of derivative works is allowed only under a license identical to the license that governs the first creator's work.</td>
</tr>
<tr>
<td>Non-Commercial</td>
<td>Others are allowed to copy, distribute, display and perform the work for noncommercial purposes only.</td>
</tr>
<tr>
<td>No Derivative Works</td>
<td>Others are allowed to copy, distribute, display and perform only verbatim copies of the creator's work, but no derivatives based upon it</td>
</tr>
</tbody>
</table>

2.1.3 GPL and LGPL—"Copyleft"

Another way to distinguish different types of open source license models is by considering the Copyleft principle and the extend where the specific type of licensing
allows the licensee to stick to this principle. The license models GPL and LGPL could be given as examples which have Copyleft principle. An explicit discussion on Copyleft is provided in the Results part.

General Public License (GPL) is one of the most dominating license models in the software world. This type of licensing overs act of copying, distribution, modification of licensed work. When performing these actions, the user is obliged to refer to GPL and the owner. Charging a license fee is not permitted with GPL. Furthermore, the owner of the license has to make the source code directly available. Lesser General Public License is designed for software libraries in particular. This open source license model covers situations where a combination of GPL software and non-free software is required to solve a problem. With LGPL this combination is allowed.

2.1.4 Other Types of Open Source Licensing

The European Union Public License (EUPL) was developed upon the initiative of the IDABC project (Interoperable Delivery of European eGovernment Services to Public Administrations, Business and Citizens). This model is based on other Open Source Licensing Models, which is adapted to the specific characteristics of the laws of the countries within the European Union. EUPL allows to the licensee to copy, modify, distribute the original software. The licensor must grant the licensee royalty-free, non-exclusive usage of rights to any patents of the licensor. Furthermore, there exists many other types of open source licensing that were created for more specific users. These licenses may have national/geographical specifications for an already existing open source model. For instance, CC-NL = Creative Commons for Netherlands, DPPL = Germany, LA = France are country specific licenses that adopted the principles of CC to the law of that country. Also there are license models that were developed for specific tasks. These licenses are basically modifications of already existing models. ‘Sampling Licenses’, ‘Music Sharing License’ are in this category.

2.2 Public Knowledge on Open Source Licensing Models

In this analysis we must also take into account public knowledge ability of the topic, as the vast community of creators and users are sure to be affected by the minute details differentiating licensing modes. A study conducted by Universidad de Extremadura found that this topic is not emphasized enough [5]. The authors asked the students their perspectives of copyright, finding that 28% of the students did not know what types of use rights are included under copyright, how long a copyright lasts (58% responded 50 years after the author’s death while 23% responded 60 with years). Further, under the specified lens of open licenses in the context of use and economic implications: 80% of the students responded with the assumption that they could not use contents accessed on the internet, 27% responded “the author renounces all economic rights, but not the moral rights”, while 17% responded “the author
renounces all her/his rights.” Lastly 89% of the students do not know how to use freely distributed (often referenced as “copyleft”) work.

2.3 Framework: A Copyright Comparison

In order to observe and classify the implications of open source licensing, we have decided upon a comprehensive comparison between aforementioned licensing types (private, open, and unlicensed). In this study, we are introducing a framework for such analysis. In our proposed framework, explicit comparisons are made between the legal, economical and social implications of the different licensing types.

<table>
<thead>
<tr>
<th></th>
<th>Private Copyright</th>
<th>Open Source Licenses</th>
<th>Unlicensed</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Legal Implications</strong></td>
<td>• All rights reserved for the creator</td>
<td>• Free access to the source code</td>
<td>• Users access creators work with no warranty</td>
</tr>
<tr>
<td></td>
<td>• Not possible to modify or distribute</td>
<td>• Free distribution of the software</td>
<td>• Can be stolen by or credited to a third party</td>
</tr>
<tr>
<td></td>
<td>• Copyright lasts 70 years after the creator’s death.</td>
<td>• Authorization of modifications and the distribution of derived works</td>
<td></td>
</tr>
<tr>
<td><strong>Economic Implications</strong></td>
<td>• Capital potential reserved for the creator</td>
<td>• Depending on the license chosen, the creator can force derivative work to also be free, thus affecting the third party</td>
<td>• Creator has no right to claim any economic outcome</td>
</tr>
<tr>
<td></td>
<td>• Allows creator to change price of product with changes in market</td>
<td>• Reproducibility, ease of distribution</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Bigger community of contributors</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Modifications on terms may reduce reusability</td>
<td></td>
</tr>
<tr>
<td><strong>Social Implications</strong></td>
<td>• Less access to information</td>
<td>• Product could be malicious</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• High barrier potential to collaborative work</td>
<td>• Reproducibility, ease of distribution</td>
<td></td>
</tr>
</tbody>
</table>
3. Results

3.1 Legal Implications

3.1.1 Private Licenses

Under copyright, the creators “have the right to prohibit any use that does not suit them and to demand any kind of remuneration.” [3] It makes impossible for users to copy the content and build derivations upon it. This rule has exceptions depending on the publication year and whether the work was for hire or not.

3.1.2 Open Licenses

The main idea of open licenses is that they allow users to copy the document, change it, distribute the changed versions but those versions have to be under the same license as the original document. Where open licensing works to ensure the access to the work cannot be privatized in the future, unlicensed work (aka “no known copyright”) allows for commercial redistribution.

When we investigate the legal implications of open licenses, we see that it gives more freedom to the users. It allows free redistribution without royalty fees. Also for the modification of the original product or any kind of derivations are allowed but they have to be under the same license as the original product. An exception of this case can be found in some open license types and they may require the modified product to have a different name or a version number. Another feature open licenses have is that they cannot restrict any other software to be used.

3.1.3 Unlicensed

In the case of unlicensed software, the users can access the software without any warranty. There is no mechanism to protect the software from piracy.

3.2 Economic Implications

3.2.1 Private Licenses

The products under copyright mostly give their users no right to copy or modify the original product. As a basic result of this, the creator has the right to claim all the possible revenue. Copyright also allows the creators to change the price of a product.

---

[3] However as a general rule, for the works published after 1977, the copyright lasts 70 years after the author’s death if the work has not been published (with exceptions depending on the publication year and the work was for hire or not).
according to the competition in the market. Although it seems like it is the perfect scenario for the creator, it also can have negative implications to the creator.

To understand the economic implications of copyright, a distinction between short-term and long-term impacts can be made. In the short-run, because of the absolute rights the creator has, it clearly has more potential for the creator to have greater income. The creators' rights are above the user's rights and this situation benefits the creator more. An increase in price of a product affects the users in a negative way, and this is a trade-off that the creator must consider the effects of. The users will be exposed to costs, administrative costs etc. There seems to be no solution for the creator and the users can benefit at the same time.

This situation may be different in the long-run. The unauthorized copying of a product clearly damages the economical situation of the creator. Since supplying new works has a cost for the creator, this economical challenge may affect the business. Thus, in the long-run, a system where the creators and the users both benefit the situation is more possible.

### 3.2.2 Open Licenses

The number of products that are under an open license is increasing every year. Aforementioned advantages such as ease of collaboration, freedom to use and modify the product etc. are important factors for this trend. Another crucial factor affecting this trend is the economic implications of open licenses. It has been observed that in many cases choosing an open license or changing the open license to a private one for the product can have significant ramifications for developers personally and for the products.

When we look at this issue from the perspective of products, we can say that there are many cases in which changing the license of a product from a private license to an open license or even changing type of open licenses to another type can have serious results.

"Changing the license of a free and open source software (FOSS) system might result in users no longer being able to reuse the software"[1]. An example of this was about a firewall package called IPFilter. The author has added an extra sentence to the licensing statement and as a result, the users stopped preferring this software because of a conflicting situation in licenses of IPFilter and its environment. As a solution, some developers have created another software to replace it.

"Changing the license of a system can promote and ease the distribution and reuse of a software system"[1]. This situation was observed with Java and its inclusion in Linux. Before changing their license, Java could not be included in Linux distribution and after the change, as a natural consequence, every Linux user had Java included in their system.

"A change to a more permissive license (and in particular, allowing commercial derivative works) may increase the size of the community of contributors to a FOSS system"[1]. A framework called Mono is a good example of this. Although it had already an open license, because of the license type, its derivative work had to be under the same license as well. After changing the license to another type of open
license that allow people to use any license they want for the derivative work, their community has witnessed a drastic increase in numbers.

"Changing the license of FOSS system towards a more permissive might cause the abandonment of a competing system" [1]. As an example, a widget library called QT can be studied. After the first version, they have created an open license for their own needs and this license caused conflicts with a project called KDE which QT formed a basis. Resulting from this conflicts, another project was started, a QT-free KDE project, and this move caused serious problems for QT.

In a system where the software developers maintain the continuity, we should also talk about the economic interests of the developers. When the product of a developer is reaching a bigger audience, one of the positive results for the developer is that she/he attracts software firms as a potential employee. Lerner and Tirole mentions this as the signaling mechanism. [4] "One may add that supplementary markets for applications and services, such as support and training for open source software, are also emerging." In examples of Red Hat and Novell, this business model provide the original product to the public and individual developers may have income by providing support for the original product.

3.2.3 Unlicensed

The implications for copying or building derivations on products that have no license are minimal. Users access creators work with no warranty, further the work can be stolen by or credited to a third party.

4. Conclusions

The complexities of the various forms of licensing have lead to a select few to becoming commonplace standards. Traditional private licenses will continue to serve their purpose for the firms who maintain an economic interest above all. Meanwhile Creative Commons, GNU, GPL, and MIT are all vying to become the de facto license model. For as long as academic and private parties continue to create and build a top work that is licensed with any form of copyright, we must ensure that their implications are well understood. The clear lack of understanding by members of academia (which make a large part of the research community) show us that we have to spend more time on educating the public and the students regarding the capacities and rules of various licensing types.
References

Abstract. There are many institutional and regulatory initiatives to promote (and even mandate) open science for publicly funded research activities. On the other hand, with the rise of online educational platforms, open education has also gained more popularity. Institutions involved in open science may also be involved in open education; however, this is not necessarily the case. Different indicators for involvement in open science and involvement in open education can be used to study this correlation. This paper presents both a discussion of current states of the fields, the tasks of measurement and also the results of a study of correlation between the number of open repositories and the number of MOOCs offered in different European countries.

Keywords: open science, open data, open access, open methodology, open peer review, open education, mooc

1 Introduction

1.1 Open Science

There is not a universal definition for open science (OS). However, in most contexts, open science refers to the practice of making parts or the entirety of a research publicly available. There are six commonly known pillars of open science: open data, open access, open methodology, open source, open peer review and open education [23].

Open Data refers to the process of sharing raw and processed data for the purpose of public analysis. There are no infrastructural limitations with regards to sharing large volumes of data. Open data allows for transparent and well documented peer reviews. Moreover, openly sharing data has shown to increase the citation rate [20].

Open Access refers to methods of paper publication that allow easier access. There are four commonly known open access standards [16]:

Exploring relationship between open science and open education: Europe case study

Philip Tovstogan, Behzad Haki, Martin Disley, and Aniel Rossi

Sound and Music Computing

{philip.tovstogan01, behzad.haki01, martinjake.disley01, aniel.rossi01}@estudiant.upf.edu
1. **For-profit gold open access**: articles are freely available at the time of publication. The costs are covered by the funders or researchers.

2. **Non-profit gold open access**: articles are freely available and there are no costs to be covered.

3. **Hybrid open access**: some articles are freely available while others are for purchase. To access open access articles in hybrid open access journals, the publisher mandates paid subscription.

4. **Green open access**: pre-print or peer-review versions of articles are freely available at personal or institutional webpages or repositories, and there are no costs to be covered.

**Open Methodology** refers to providing clear descriptions of a research methodology so as to enable other researchers to repeat the research procedure. Sharing research methodologies or experimental procedures increases the level of trust in the “quality”, “integrity”, “claims and conclusions” of a research. Open methodology increases the level of transparency of a research activity [13].

**Open Source** refers to computer software for which the source code is openly and freely available. Software used as part of the research methodology should be well documented and also be available to read, compile and run. Open science research rely heavily on open source software. Open source software also promotes transparency and improves the peer-review process [3].

**Open Peer Review** refers to non-anonymous peer reviews which facilitate the author-reviewer collaboration. Peer review is very critical to the validation process of a research activity. Traditional peer review is time consuming, costly, highly subjective, biased, and commonly fails to detect fraud. Open peer review is an alternative method to overcome the shortcomings of traditional peer-review [21].

**Open Education** refers to free and openly accessible educational resources. It should be noted that open education does not mean that the students can not be charged for the education, rather, the resources are freely available. Open education has given rise to many platforms commonly known as massively open online courses (MOOCs).

**Open research** values and promotes “accessibility, sharing, transparency and inclusivity” [13]. As shown in Fig. 1, Lyon has proposed a 3D-model for the continuum of openness. In this model, ’Access’ dimension refers to the level of (free) accessibility of an article, the 'Participation’ dimension refers to the degree of collaboration involved in a research, and the 'Transparency’ dimension refers to international availability of data and documentation used in a research activity [13].
1.2 MOOCs

In this paper, we will focus on one particular aspect of the Open Education (OE) world, due to its incredible growth and the positive feedback it has received from the scientific and wider community. A massive open online course (MOOC) is an online course aimed at unlimited participation and open access via the web. In addition to traditional course materials such as filmed lectures, readings, and problem sets, many MOOCs provide interactive user forums to support community interactions among students, professors, and teaching assistants (TAs) [17]. MOOCs have the following characteristics:

1. **Massive**: MOOCs are intended to be run at scale, with hundreds or even thousands of participants and without any limit to student numbers being imposed.
2. **Open**: MOOCs are intended to be open. With the word “open” used to imply that access to MOOC is both free of charge and also that access to MOOCs is unrestricted. MOOCs have no entry requirements and are open to learners of all educational background, age, and location.
3. **Online**: MOOCs are delivered completely online and involve no face-to-face contact. They are delivered through Internet technologies and so make it easy for students to communicate with each other while learning and for students to access resources that are available elsewhere on the web.
4. **Course**: One of the key attributes that differentiates MOOCs from an open educational resource is that they have the characteristics of a traditional course they are run during a specific time period, based upon prescribed content, and instruction is provided to the student during that period of time. As with traditional courses, there is also usually an element of assessment included in MOOC, and this may include some form of accreditation (which is a thorny topic and one to which we will return later on) [18].
When one of the first massive open online courses appeared at Stanford University, 160,000 students enrolled. It was 2011, and fewer than 10 MOOCs existed worldwide. It has been four years since then, and according to a new report, the cumulative number of MOOCs has reached nearly 4,000 [25]. In these 6 years a surprisingly number of MOOCs and providers were born online, suggesting that their importance was received by people worldwide.

![Growth of MOOCs](image)

**Fig. 2.** Growth of MOOCs from November 2011 to February 2016 [25]

The top three providers by number of courses in 2015 are Coursera, edX, and Canvas Network [2]. Other common providers are FutureLearn, Udacity and Khan Academy, that is not explicitly cited in the graph. Nowadays the percentages showed are very similar (the graph refers to the 2015).

![Course Distribution by Providers](image)

**Fig. 3.** MOOCs Providers in 2015 [2]
1.3 The trouble with measuring open science

A 2017 report by the European Commission (EC) [24] presents the work of an Expert Group (an interdisciplinary research group), whose remit was to review the “strengths, weaknesses, and future possibilities of next-generation metrics”.

Before discussing the metrics themselves it would be worthy conduct some analysis of the application of metrics in science and the metrics themselves.

Firstly, metrics, in the abstract, are not free from criticism, some view that metrics can facilitate a culture of bureaucratic scrutiny in universities, focusing primarily on that which can be measured easily and objectively and pushing researchers to work on lower-risk incremental research aimed at high impact journals [4]. With some even suggesting that academia is at risk of becoming a “self-referential system where quality is measured mostly in bibliometric parameters and where societal relevance is undervalued”[6].

However, the EC Report argues that if a proper assessment of the metrics themselves is conducted, and a more holistic goal-orientated understanding is come to, then the application of such can certainly be of benefit. Monitoring the development of the scientific system’s “openness” will provide accountability institutions and other actors involved, and provide reward-based incentives for improving ways of working at the group and individual level.

The first task of a meta-analysis would be to ask, for whom or what is this analysis for, and, in what ways does the act of analysis further our goal. The EC Report suggests that relying on a single metric through which to measure open science fails to acknowledge its two-sided goals. To understand this they consider the political economy of science; suggesting that open science is in itself is the reconciliation of current ‘supply-side efforts with the social aim of a more ‘demand-side approach. In other words, within the remit of open science lies the task of stimulating the ‘demand for scientific output by increasing the access the ‘consumer has to the output. The act of analysis must be for the benefit of a dualistic approach. They, therefore, conclude that a single metric will not be sufficiently flexible.

The findings of the EC Report suggest that the shift towards web-based networked research requires a suite of novel indicators, which they group together as “altmetrics. Before presenting these, however, they give a summary of the most widely employed metrics used currently.

Bibliometrics and usage-based metrics Conventional metrics attempt to measure research output, the two basic and most commonly considered are, the number of publications, and the number of citations the publication receives. The data for these measurements is sourced from bibliometric databases such as Web of Science\(^1\), Scopus\(^2\) and Google Scholar\(^3\). From these sources, higher-order indicators have been created, such as the Journal Impact Factor (JIF) [9]. the

---

\(^1\) https://webofknowledge.com
\(^2\) https://www.scopus.com
\(^3\) https://scholar.google.com
h-index [12], field normalized citation indicators [22], Eigenfactor [1] and SJR [10].

Criticism of the Journal Impact Factor has in recent years been rising [5] largely due misuse of the JIF as an attribution of the articles “impact”. Moed [14] and Hicks [11] also criticize the h-index for its bias towards and emphasis on researcher seniority.

Altmetrics The concept of Altmetrics was first introduced by Neylon and Wu [15] as “article-level metrics” and by Priem and Hemminger[19] as “Scientometrics 2.0”. Data for Altmetrics is sourced mainly from social media platforms, blogs, Twitter and more professional networked sources such as Research Gate and Mendeley.

The EC Report groups the strengths of Altmetrics into the three following categories:

1. Formats of relevance: altmetrics can identify new formats of scholarly products to measure, which have not been considered in research assessments before, e.g., research data and software;
2. Forms of impact: these refer to the new audiences captured, who interact with or react to scholarly products and scenarios related with that, e.g., policy makers and policy documents;
3. Targets and uses: these reflect the purposes for which altmetrics can be used, e.g., budget allocation or self-assessment and career development.

In other words, altmetrics present a broader approach, for they are capable of measuring not only scholarly influence, but reach in broader audiences too. Furthermore, altmetrics can measure a greater variety of research objects, including data, software tools and applications, whilst offering a multi-faceted approach by measuring multiple signals simultaneously [24].

2 Research methodology

2.1 Problem statement

As mentioned in Section 1.1, there are six well-known pillars of open science: open data, open access, open methodology, open source, open peer review and open education. However, in most contexts, open education is considered to be separate from open science.

A concern about open science can be that the active involvement of institutions/countries in open research does not necessarily imply active involvement in open education. The goal of this paper is to find possible correlations between countries’ involvement in open research and their involvement in open education. This paper specifically focuses on European countries.
2.2 Method

Within the European Commission, there are many initiatives regarding both open science and open education. The goal of this research is to use available historical data to investigate whether involvement in open science is correlated to involvement in open education.

Different indicators can be used to evaluate the involvement of an institution/country in open science. Fig. 4 illustrates the indicators proposed by European Commission for open science.

![Fig. 4. Indicators for Open Science Proposed by European Commission](image)

These indicators are used to assess different characteristics of open science: open research data, open access to publications and open scholarly communication. Some of these indicators deal with qualitative assessment of open science characteristics, while others can provide a quantitative assessment. Some quantitative indicators that can be used to assess openness of science are: number of research data repositories, number of open peer reviews, number of corrections.
and retractions, amount of open science funding, number of reprints and number of open access publications.

Involvement in open education can be assessed using indicators such as number of open courses and educational resources provided by an institution on their own specific platforms, number of open courses (MOOCs) and educational resources provided on dedicated educational platforms, amount of funding dedicated to open education, and policies towards open education.

2.3 Data sources

The European Commission website provides statistical data about the number of data repositories per country [8]. Fig. 5, obtained from the EC website, shows the descending ranking of European countries based on the number of open data repositories\(^4\).

The European Commission initiative regarding open education is called OpenEducationEuropa. Their website provides data on the involvement of European

\(^4\)the provided data were obtained from re3data.org
countries in MOOC platforms [7]. Fig. 6, obtained from [7], illustrates the number of MOOCs offered by each European country up to June 2015.

![Distribution of MOOCs per country](image)

**Fig. 6.** Involvement of European Countries in MOOCs [7]

In this paper the number of data repositories per country (Fig. 5) has been used as an indicator for involvement in open science and the MOOC involvement data (Fig. 6) has been used as an indicator for involvement in open education. All data is referencing year 2015 to make comparison valid.

The goal is to investigate correlation between data available on open science and open education. The simplest metrics will be used, particularly correlation coefficients like Pearson and Spearman.

### 3 Results

#### 3.1 General analysis

Using the data available from [7] and [8] you can see the relation between number of open data repositories and number of MOOCs in Fig. 7. The right figure is the zoom on the bottom left corner of the left figure.

From Fig. 7 we can see the general trend and that those quantities are positively correlated. We can see UK in the top right corner what means that it is involved at a great amount in both open science and education. Germany, on the other hand, has more open data repositories than UK, but number of MOOCs provided by it is less than half of UK. So Germany is a country that does not really follow the trend. Another one is Spain with much more involvement in open education than in open science. And France confirms our hypothesis. Austria, while it has much lower numbers than Germany, shows similar bias towards open science.

There are a lot of countries in bottom left corner with low number of both repositories and MOOCs, and that data is actually more important to our correlation analysis. The reason is that countries with higher GDP (like UK, Spain,
Fig. 7. Scatter plot of open science vs open education

Germany) obviously have more resources to spend on both education and science.

3.2 Correlation

To analyze the correlation of our data we will be using Pearson and Spearman correlation coefficients, but because our data is not normally distributed, Spearman’s might be more useful.

We separate our data into 2 groups: countries that have large numbers for both open science and open education: Germany, United Kingdom, France, Spain, Austria and Netherlands; and other countries

<table>
<thead>
<tr>
<th>Measure</th>
<th>Top 6</th>
<th>Other</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson</td>
<td>0.545</td>
<td>0.557</td>
<td>0.545</td>
</tr>
<tr>
<td>P-value</td>
<td>0.035</td>
<td>0.043</td>
<td>0.030</td>
</tr>
<tr>
<td>Spearman</td>
<td>0.768</td>
<td>0.816</td>
<td>0.777</td>
</tr>
<tr>
<td>P-value</td>
<td>0.0008</td>
<td>0.0002</td>
<td>0.0010</td>
</tr>
</tbody>
</table>

Table 1. Correlation coefficients

From data presented in Tab. 1 we can indeed see that p-values for Pearson coefficient are two orders of magnitude larger than for Spearman. Moreover the Spearman coefficients are larger than 0.75 which is a very good indicator that correlation is indeed present what confirms our hypothesis. To rule out the possible hidden variables we looked at rankings of European countries by GDP and GDP per capita, and generally the coefficients of correlation of either of our data with those were not far from 0.
Another important observation is that the correlation is typically much more evident for countries outside of our top 6 groups. That shows that if we exclude countries with much more power and freedom to decide where do they spend their budget on, open science and open education are still naturally related.

4 Conclusion

We have presented a discussion on both open science and open education, with emphasis on open data and MOOCs. We have presented the trends in growth of MOOCs and discussed the difficulties encountered in measuring the growth in open science. The European Commission Report[24] outlines a working method- ology that seeks promote the use of a suite of metrics that can be used in tandem to facilitate this. Given this we also present the results of some preliminary re- search conducted using data from European Commission initiatives[7] regarding the number of open data repositories per EU State and number of MOOCs per EU State. We have considered open education and open science to be different fields as this is consistent with the working methodologies of the literature we have reviewed. However, it was our initial assumption that there was indeed a correlation between the two fields, our results go some way to prove that this is indeed true.

References

9. Garfield, E., Others: Citation analysis as a tool in journal evaluation. American Association for the Advancement of Science (1972)
Measuring Impact of Open Science

Álvaro Quílez, Fabio Stefanizzi, Keith Kennedy and Octavi Font

Computational Biomedical Engineering keithtopher@gmail.com
octavifont01@estudiant.upf.edu álvaro.fquilez@gmail.com
fabio.stefanizzi01@estudiant.upf.edu

Abstract. Open Science is a movement that aims to transform scientific research into a more accessible practice. Nowadays, transformation is already happening, and it is expected that scientific research will continue to become more reliable, lucrative, and principled. The aim of this study is to determine the impact of open science practices in various fields of research. This is measured qualitatively through number of citations and qualitatively through progression within the specific fields. The infrastructure, incentives, and regulation of open science practices in various fields of research are discussed. In studying the effect of open availability of articles on the average number of citations within specific journals, statistical analysis revealed a positive correlation for three out of twelve journals tested. However, there appeared to be no effect of open availability when comparing among journals.

Keywords: Open science; data sharing; collaboration; infrastructures; archive; influence; impact; citations, accessibility; research

1 Introduction

Open science is an implicit construct in research that can be difficult to define absolutely. It includes many components, which have varying levels of importance to different types of research. At the most general level, open science refers firstly to “technological advancement, particularly in networks and data management,” and secondly to “social adaptation among researchers, particularly in collaboration and transparency.” [1] Regardless of the type of research being conducted, open science deals with the dissemination of information, which inevitably involves cooperation among different groups.

Open science practices have arisen in various fields throughout the years as the result of internal motivations. Environmental and economical sciences have a long history of large-scale, large-data projects that rely on international, multi-disciplinary teams who work together to combine data into a global picture [1]. Mathematics has always been inherently open, although its manner of work is more individualized. There is no data to share as it is not an experimental science, but nonetheless the research culture is open and readily adapts new technologies [1].

Some fields, however, are not as willing to accept all the practices of open science. For example, in microeconomics, sometimes it is not lucrative from a business...
standpoint to share data. An organization may have a novel method for building a product or analyzing data and do not want competitors to steal their ideas [1]. Also, field ecologists are leery to share data as there is often much time in between data collection and analysis. They do not want other groups to analyze their work before they have had a chance to analyze it themselves.

Local groups have arisen over the years with the aim to improve the use of open science in their respective fields. In the 1980s, a water utility company in the UK worked with researchers to study water quality, resource management, and consumer behavior. The team has since developed models, provided accessible data sources, and extended their work overseas. This was done prior to the internet, with the data stored on floppy disks and mailed from location to location [1]. Programs such as these paved the way for larger, more encompassing organizations to implement and regulate aspects of open science.

In 2000, the European Research Area (ERA) was born, and its purpose was to increase the competitiveness of European research institutions by bringing them together [2]. The way to reach this aim is through infrastructures which allow for an optimal reuse of data on the condition that the data is FAIR (Findable, Accessible, Interoperable and Reusable) [3],[4].

Having infrastructures means having money to create them, and for this reason an EU platform was born called Horizon 2020. From 2014 to 2020, this programme will finance 80 billion euros to create an open access European database and thus reach the goal of the ERA. When this project is complete, it will contain mechanisms for open/collaborate/public peer review. Thus, its primary goal is to offer a publishing service that is fast, cost efficient, and fit to the 21st century [5].

The wider range of open research benefits can vary between specific areas. Nevertheless, open science practices have been associated with the potential for collective intelligence, like the one shown in the development of the Internet [6]. Open science leads towards a greater openness in research design, project management, data release, reproducibility and publication practices. Those, at the same time, make new century challenges more feasible transforming unaffordable research into something economically realistic [1].

The outcome of open science practices can not only be measured in economical impact but also in terms of time, collaboration and, generally speaking, available resources for the researcher. Open science also facilitates the formation of multi-disciplinary teams, breaking the barriers between different fields and thus, making easier the emergence of new approaches for a given research.

As it has been introduced, the benefits of open science can be measured from different points of view. The study being presented focuses on showing the relationship between open science practices and the impact of the articles in the scientific community. The hypothesis is that articles that are openly available tend to receive more citations when compared to articles that require payment or access.
2 Research methodology

A dataset of relevant journals from a diverse group of disciplines was selected so the trends of open vs non-open access articles could be tracked to see if there was a significant correlation between this and the impact of the paper.

First, a group of 6 disciplines was selected. The disciplines were selected due to their overall wide range of subject matter in order to capture the maximum variation regarding the approaches to open science.

Table 1. Selected journals by discipline

<table>
<thead>
<tr>
<th>Discipline</th>
<th>Journal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social Science</td>
<td>Journal of Sex Research, International Social Science Journal</td>
</tr>
<tr>
<td>Psychology</td>
<td>Journal of Clinical Psychology, Journal of Experimental Social Psychology</td>
</tr>
<tr>
<td>Computer Science</td>
<td>Neural Networks and Learning Systems, IEEE WIRELESS COMMUNICATIONS</td>
</tr>
<tr>
<td>Economics</td>
<td>Journal of Economic Literature, Quarterly Journal of Economics</td>
</tr>
</tbody>
</table>

The selection of the journals was done by picking 2 of the Top 5 JCR Journal Citation Reports for each category. The 2 picked have been selected by balancing impact factor with overall number of citations (based on 2016 stats).

Once the journals and disciplines were selected, the top 400 papers were pulled for each publication in the period 2012-2016. Google Scholar was used to come up with the results, filtering by date and publication. A modified version of this script https://github.com/ckreibich/scholar.py was employed to automatically pull the paper references and parse the results. This tool returns a list with the article, number of citations, journal link to the publication and, if available, the PDF link. Still, this information was not complete, since many publications offer a pdf view to the paper, although this link is not directly available from Google Scholar. For this reason, we performed an extra step, where every journal link was assessed to find any direct link to the pdf. If the journal page had a pdf link, the article was considered to have open access. That last preprocessing step was automated as well with some code developed for that purpose.

The subsequent analysis of the dataset was performed in Python. First, some basic graphical analysis was completed. In each journal, the number of citations for each article was plotted against the year in which the article was published. Also, the mean number of citations per journal was plotted with respect to the fraction of articles within the journal that have open availability.

Next, the data were examined more closely within their respective journals. Articles in each
journal were first split into two subgroups corresponding to whether the article has an available pdf. The means and standard deviations of the number of citations in the two subgroups were calculated for each journal. Statistical analysis was performed on the difference between the means of the two subgroups using a paired t-test with a significance value of \( \alpha = 0.05 \). The null hypothesis for each test was that there is no difference in the two means for each journal. The alternative hypothesis is that there is indeed a significant difference between the means. In the context of this study, it is assumed that open science, in the form of pdf availability, has no effect on the number of citations for a particular article.

Since the sample sizes and standard deviations differed, the following corrections were used for the standard deviations and degrees of freedom:

\[
\text{Std. Dev.} = \sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}},
\]

\[
\text{DoF} = \frac{(s_x^2/n_x + s_y^2/n_y)^2}{\frac{(s_x^2/n_x)^2}{n_x-1} + \frac{(s_y^2/n_y)^2}{n_y-1}}
\]

where \( s_x \) and \( s_y \) are sample standard deviations, and \( n_x \) and \( n_y \) are sample sizes.

Next, the data were examined among the different journals. Statistical analysis was again performed using paired t-testing. For each journal, the mean number of citations was compared separately to the means of the remaining journals. In total, 66 paired t-tests were conducted, and the significance level \( \alpha = 0.05 \) was changed accordingly to \( \alpha = 0.05 / 66 \approx 0.001 \). The hypotheses of this test were similar to those of the first test, with the null stating that there is no difference between the means of citations of journals, and the alternative stating that the means are indeed different.

3 Results

The graphical analysis of the data did not reveal any trends in relation to the impact of open science practices. In Figure 1a below, the only clear trend that can be seen was that the number of citations decreases as the year increases. This information is trivial, as the older articles have more citations most likely because they have had more time to receive citations. Also, there appears to be no correlation between the average number of citations and the fraction of articles with open availability. The results are plotted below in Figure 1b.
Figure: 1a. This plot shows the relationship between the number of citations and the year the article was published. Only the Journal of Sex Research is shown, however, the remaining eleven journals have similar graphs. 1b. This graph shows the relationship between the average number of citations from the articles and the fraction of articles with available pdf's in a particular journal.

The statistical analysis produced some significant results. In the testing between the mean number of citations of openly and non-openly available articles within a specific journal, five of the twelve showed significance. Although there were indeed differences between the means, there was not a clear outcome on whether open availability leads to more citations. For International Social Science, New England Medicine, and Nature Cancer Reviews, the mean number of citations was higher for articles with open access, while Quarterly Economics and Sex Research had the opposite relation. For the other journals, the availability of the articles appears to not affect the average number of citations.

In the tests between the mean number of citations for different journals, 61 of the 66 pairs were significant. This shows that there is a difference in the mean number of citations. However, as there appears to be no relation between the number of citations and the fraction of openly available articles, these differences in means likely come from other sources.

4 Conclusions

Based purely on the statistics applied to citations of academic articles studied, it can be said that, nowadays, freely available articles do not have such a profound impact on the scientific community. Within some individual journals, the number of citations was greater for open articles, while others showed the opposite or no relation. In summary, no conclusions can be made on the significance of open access to articles one way or the other. This result could lead one down two different currents of thought regarding the relationship between freely available publications and traditional journals that require payment to access.
On the one hand, perhaps trivially, behind the deep economic interests of private publications, there is also very strict peer review. This gives more credit to the publishers and therefore reinforces the belief that the worth of publication increases with the cost. On the other hand, the data that was studied is extremely current and therefore not validated and supported by years of experience as is the case for private publishing. The reputation of an established, private publication may have a significant impact on how widely it is referenced in other publications.

Looking beyond the scope of this study, the results do not negate the impact that open science as a whole has had in the scientific community. What makes this topic difficult to examine is its implicit root in research. There is no definitive line between groups who practice “open science” versus “closed science.” Additionally, many of the supposed benefits are difficult to describe quantitatively and equally difficult to isolate in terms of their true impact. In order to more effectively study the impact of open science practice, clearer definitions of the topic must be established, as well as more precise manners of analyzing the data.

Considering the future direction of open science practice, there is hope that the current EU policy reforms, which allow for a substantial liquid dilation in research, will allow openly accessible publications to achieve the same credibility standards as seen in private competitors. This would likely lead to exponential growth and ‘well controlled’ open diffusion of knowledge. Therefore, it can be concluded that open science today is comparable to a child, full of potential and uncertainty. This child can and must learn much from two fathers; the EU, the father giving rise to possibility, and publishing, the wise father who gives critical meaning, which will help the child grow on his own.

References

4. G7 2017 Programme (point 19, 20), http://www.g7italy.it/sites/default/files/documents/G7%20Science%20Communiqu%C3%A9%202017.pdf
Social impact of science and technology

Jorge Sevilla Ramos, Jordi Selva Vinagre, Jordi Casals Grifell and Cristian Izquierdo Morcillo

Master in Intelligent Interactive Systems UPF

Abstract. This paper presents a study of 67 papers about which specific areas of society do science and technology have an impact on, especially artificial intelligence and robotics. In the last years, these fields have quickly grown and improved, because there is a lot of research focusing on them. For this reason, society has got to adapt to these changes and take advantage of this new development. In this research, we study the papers’ impact on Economics, Politics, Community, Ethics and Other areas.

Keywords: Artificial intelligence, robotics, science, technology, society, impact, research about research, social impact

1. Introduction

Paper production has been increasing steadily for years, with about 2.5 million of them published in 2014 in journals from around the world [1]. Their authors reach a series of conclusions that the public can also interpret. Sometimes the paper’s impact on society does not match between what the author thought it would be and what people actually do with that research.

There are a lot of parameters that can be considered, such as implementation cost, national laws, or state of the art of the needed technology. In this study, we have considered six divisions of society that the authors of the papers have believed that their work may have an impact on.

We have selected three different journals –one from the United States of America, another from China and the last from the European Union–, all of them from 2017, for a total amount of 67. They are listed on the Annex. The questions presented in them are usually solved by artificial intelligence, autonomous systems or robotics. Some of them does not focus on a single field, but present new technologies.

The rest of this paper is structured as follows: in Section 2 we briefly discuss the different factors and characteristics that our society has and their relation to the issues we have studied. In Section 3 we introduce the followed methodology to extract the
needed data from the papers. In Section 4 we present the conclusions that we have arrived after our research

2. Fields of social impact of science

Society is a complex organisation; it cannot be defined in a general description. There are multiple factors that must be identified and studied separately to extract the proper conclusions about the state of society nowadays. The level of different concepts has been increasing during the past centuries considering the addition of new features that can explain how the humans and their psychology, and therefore the collective, work. In this paper, we considered the study of six social divisions of our society [2, 3, 4] that can give a wide explanation of the most remarkable aspects affected by the social impact of science: Economy; Politics; Community and Individuals; Culture, Ethics and Values; and Other Institutions.

2.1 Economy

In a world regulated by the rules of capitalism, the money, business, and enterprises have become one of the most important mainstays to understand the performance of the society and its governments [3].

Humanity is at a crossroads [5]. Present social science theory and public policy are no longer adequate to meet the multidimensional challenges posed by rising social aspirations, unemployment and inequality, wasteful patterns of production and consumption, globalization of markets, technological advances, demographic changes, and ecological constraints. Incremental changes in public policy based on the present conceptual framework will only aggravate problems that are already acute. At the same time, the resources and capabilities of global society have never been greater than they are today. Scientific knowledge, technological developments, infrastructure and productive capabilities, educated and skilled human resources, entrepreneurial skills, commercial organisation and a rapidly expanding global social network offer unprecedented opportunities for rapid social progress [4, 6].

Our problems arise from a mismatch between resources and opportunities: social science theory that is divorced from real world functioning, financial markets that siphon off funds from investment in the real economy, capital and technology-intensive manufacturing strategies that eliminate labour resulting in higher levels of unemployment and falling consumer purchasing power, educational systems that fail to impart the knowledge and skills required by the labour market, energy production
technologies that threaten the ecosystem, mispricing of natural resources leading to overexploitation and waste, national level institutions incapable and unwilling of coping with global level problems

To summarize, the impact of science on the economy can be detailed in three aspects: employment, consumer purchasing and overexploitation of resources (which is also an area of impact in terms of the environment).

According to IDC’s FutureScape [7]: Worldwide Robotics 2017 Predictions report, in 2019, 30% or more of the world’s leading companies will chief robotics officers and we will have to regulate and build a new legislation surrounding it. Low-skilled people, according to the report, will reduce their job opportunities in a 78%. As an apocalyptic final statement, “For every job created by robotic automation, several more will be eliminated entirely. At scale, this disruption will have a devastating impact on our workforce. And we seem not to be ready for that”.

2.2 Politics

Fundings. Since politics are in charge of regulating almost everything about the social organisation, it is plausible to admit that science has a huge direct impact on it. Science, politics, and technology have a complex relationship between politicians and scientist. After all, science is the pursuit of knowledge, knowledge is power and power is politics [8].

One of the most controversial aspects of this dispute is related to fundings addressed to science. Most of the researchers in the world do their investigations for public institutions like universities. These organisations (almost all of them) depend directly from public money, and the decisions about funding allocation are made by politicians, so deciding what science we find valuable is not a scientific decision, it’s a political one. Most governments choose to spend more money on defense instead of cancer research or space travel. So science affects the roots of politicians that now find themselves in the debate about dedicating it part of the budget.

2.3 Population and individuals

In terms of community as a social society and its individuals, its features can be related to the rest of the points mentioned before and below, but there is one specially dedicated to a community impact: coexistence. Technology is giving us some tools that are useful in many contexts. Super-computers for advanced mathematics, the Internet and other networks to be interconnected with each other around the world in real-time. all these brand new social networks are rewriting the definition of friendship, love, and fellowship. The Internet of things is giving us today even more
Robots present a scenario with multiple interpretations. Who are they? Are they real? Can I have a human-like relationship with them? Are they conscious to live, to feel the way I do? For sociology represent a terrible shock for their standards of humanity. And they are supposed to help us out, but what if they take my job?

It is difficult to theorize about these scenarios, and the only concern of scientists may be only the scientific improvement leaving behind the consequences of this new technologies [9].

2.4 Culture, ethics and values

Society usually respects scientists over other professionals such as politicians, doctors or artists. Although people have faith in them in order to solve our current problems, the number of controversial issues involving science is increasing. At the same time, people demand a say in science, or at least a better access to information. There are public debates about whether science leads to progress or to danger. On the other hand, scientists have the responsibility to address the social implications of their research [4]. Several decades ago investigators only focused on conducting their research, and expected the rest of society to deal with the consequences of the knowledge they created. They wanted to be as objective as they could. But science is a part of society, and scientists cannot detach themselves from it [2].

Science is not value-free. Values determine the importance of the actions that lead to a desired goal. These goals can be epistemic (truth, knowledge, rigor) or not (moral, social, economic), but all of them have some influence in some aspects of the research. For example, a company may decide to fund a project which targets a lot of population because it will increase its profit. But scientists also make value judgements when they accept or reject a hypothesis because the amount of evidence needed to accept one depends on the consequences of accepting it.

Social responsibility may present difficult ethical conundrums to scientists, their obligation to society. These dilemmas usually arise in three different moments of the research: problem selection, data sharing, and public engagement. A worthwhile research may lead to social dangers (discrimination, animal and human test subjects), and the results of some experiments may land in the wrong hands (terrorism, war). Public engagement can lead people to question scientist objectivity, but it can be protected by distinguishing between the data presented and the researcher’s opinion about them, and encouraging discussion. Researchers should learn to present the results of their experiments without hiding its limits and uncertainties [10].
2.5 Other Institutions

Social institutions are associations with a common purpose that organise relatively stable patterns of human activities [11]. Education, Health, Space Agencies, etc. are all institutions. We are going to address all of them which do not fall into the previous categories we have defined above.

3 Research Methodology

To perform this research, we gathered data from three scientific publications: Science Robotics, Journal of Robotics, and Robotics and Autonomous Systems. The first one is a new American journal that has run monthly for only a year (2017); we used 38 of its papers. The second one is a Chinese-based free access journal, with one issue per year; we selected 15 of its articles from 2017. The last journal is from the EU, and we have chosen a Special Issue on New Research Frontiers for Intelligent Autonomous Systems published in April 2017; the papers taken into account have been 14 in this case. Thus, we have 67 papers from three different world regions.

We have read their abstracts, introductions, and conclusions to understand the main purpose of each one of them. What we found is that some publications did not have a unique field-dedicated purpose but they presented new technologies of robotic designs that can be implemented in many fields like transport, medicine or space travel. So potentially, robotics (science field we focused on in this report) is a transversal investigation and one single improvement in their performance could mean a huge impact in many different fields.

For the purpose of this paper, we have divided the possible categories into the ones defined in Section 2. Papers labelled under Economy dealt with industrial and market organisation. Those involving regulations and civil or military protocols were nested under Politics. Anything that affected the way of life of the general population, from prosthetics to robot-person communication was classified with the Community tag. Technology ethics and mind introspection applied as Culture, ethics and values. Other papers affecting areas such as medicine, aeronautics, space exploration, etc., were labelled as Other Institutions. A few of them could not be fitted into any category, such as editor’s summaries of their journal issues.

4 Results

The results we obtained in our analysis are presented below (Figure 1). The graphic describes the areas of impact in the EU, US and China regarding the papers we covered in our study. We can see that for Europe and the United States the papers are focused on Other Institutions in about the 45% of them, with Community being the
label with more frequency (with a difference of a 10% between US and EU). This category is the most issued in China.

![Figure 1. Bar chart showing our papers classification in five different areas of society, divided by the country of origin of the journals.](image)

It is important to mention that, although Economics and Politics are not usually specifically addressed in these scientific papers, both are transversal domains that will be affected by the new knowledge presented in them. Ethics and Culture label walks down the same road, even considering that the robotics field often raises moral conflicts and controversies, as described in Section 2.4.

5 Conclusions

The future is uncertain. The arrival of the robots, the upcoming industrial revolution of the XXI century is among us. Science and technology are nowadays increasing and improving exponentially, accomplishing goals that seemed to be unreachable not long ago. Society can’t hide from this progress; humanity must grow up at the same time. The impact of science is huge and has an important effect in every aspect of our society, and a future even more technologic will force us to adapt ourselves to the new ways of work and relate to each other.

In terms of the graphics obtained, it seems that scientists are focusing their research on different ways regarding the geographic area. In Europe and the US, they address to specialized institutions like health or space, while in the other region they take much more into account the community impact in their papers.
The first conclusion we understand from this report is that science does not grow up just on its own, it echoes among other domains, and many other features of society follow the same path. Science is in charge of revealing the secrets of our world and faith, but many other areas must follow so we can deal with some future events that will put us in a situation like never before. Society will have to understand, the sooner the better, how the relationships with machines and conscious robots will change the way we see each other. And that is because scientists are improving robotics and producing an enormous amount of knowledge in this field. Politics and regulatory institutions will have to regulate how robots will face the commitment of the law. The Economy will have to build up a new structure able to hold the new employment situation for many people that will be replaced by robots in their actual roles.

The second conclusion is the necessity of funding and encouraging scientists to keep on doing research. If the impact that science makes is on the community, it has to be properly provided with the resources needed. The only way to improve is to go beyond our knowledge and reaching new goals. Disease investigation, space travel, particle physics, among other domains, will give us the future explanations to fix current issues.

The third conclusion, considering the more addressed labels, is that medicine is one of the main fields where science has an impact, which we have classified either in Community or in Other institutions depending on if the paper results were more practical or theoretical. Diseases and therefore death is one of the greatest concerns of humans and seems quite reasonable to dedicate efforts to improve and to extend human life.

The final conclusion is that robotics, AI and science related to this field has an undoubtable purpose: human (or human parts) replacement in many aspects, This will have a huge impact, transversally, in economics and politics, although its investigation is addressed to improve other concrete areas.

References


Annex

Papers from Science Robotics


Papers from Journal of Robotics


Papers from Robotics and Autonomous Systems

Social impact of research: \textit{n-person} games in real life

Marc Alvinyà\textsuperscript{1}, Alejandro Hinojosa\textsuperscript{1} Albert Marin\textsuperscript{1}, and Lara Solà\textsuperscript{1}

Universitat Pompeu Fabra, Spain, \{marc.alvinya01, alejandro.hinojosa01, albert.marin01, lara.sola01\} @estudiant.upf.edu

Abstract. Research can be a broad concept and is the base of science. We can find research in practically all fields known to humanity: from arts to mathematics passing through biology or physics. However, some of the research is focused in opening more research questions instead of focusing in non-academic life applications. In this paper, we have evaluated whether a paper on mathematics (root paper) has had any impact at a social level, by checking the papers that have cited the root. We found out that there are fields in which it was prominently used but some others lacked direct applicability. We suggest that refinement in the measurement of the social impact and the underlying concept is crucial to conduct further studies on this topic.

Keywords: Research Methodology, Science, Social Impact, Nash Equilibrium.

1 Introduction

Analysing, communicating and improving social impact is one of the most pressing demands to all scientific fields. Indeed, science and technology dominate every aspect of our lives. Our day-to-day activities would be unrecognisable without the constant forward march of technological and scientific advancements we have had as a society. The scientific and technical advances translate into new products of mass consumption such as radios, computers, TVs, cell phones and services such as the web, social networking, etc. It also provides us with new and better medical inventions that have helped increase the average life span and the quality of life. However, among all the amazing contributions that science and technology have equipped humanity with, not all are scented with roses. Indeed, one may think of what technology has supplied to war: mass destruction instruments. World hunger, food and water resources, population growth, air quality and atmosphere, and disease. Those are just some of the non desired contributions of science and technology, which we are ironically trying to fight back now with the same tools. Quoting professor Jay Van Bavel from New York University, Science is a human activity and is therefore prone to the same biases that infect almost every sphere of human decision-making.
Our research has thus been focused on how scientific and technological advances impact on society. We have chosen the field of mathematics, for which its contribution to society is not always clear. For this, we have decided to investigate how a particular research paper [1] has had direct impact on society. We have chosen this particular publication because it introduced a simple and yet very powerful idea. As we shall see, its impact spreads over very different fields of science and has had numerous applications.

In [1], Nash introduced the notion of equilibria for non-cooperative games. A game is said to be non-cooperative if each individual plays selfishly and logically (the player does not play random moves) to maximise its own reward without any communication or cooperation between himself and the rest of the players. Nash was not the first person to work on non-cooperative games. Indeed, a very complete work on 2-person zero-sum games had been presented by mathematician John von Neumann and economist Oskar Morgenstern previously in 1928. Nash extended the idea to n-person games.

In order to understand what an equilibrium point is, imagine you are playing a game against 3 more players. You come up with a strategy to obtain the best result you can, and so do the other 3 players. An equilibrium is reached when, even by knowing the strategies of the others, you do not feel the need to change your own strategy (and so do the others), since you are already obtaining the highest reward even though the others strategies. That is, if you went to change your strategy, you would be decreasing you reward. Everyone is satisfied with their choices.

While there are broadly accepted indicators for scientific impact of research (e.g. the H-index) that are currently being used, there is lack for those on social impact. Social impact plays however a major role when selecting what lines of research are worth following.

The remainder of this article is organised as follows. In Section 2 we present an explanation of the research methodology we have used and alternatives we have thought of. As we will see, we have considered a small sampling to evaluate the social impact of our chosen article due to its difficulty. In Section 3 we present the key results of our research. Finally, in Section 4 we discuss the conclusions we have drawn.

2 Research methodology

In order to check the social impact of the research in the field of game theory we propose the following method, depicted in figure 1. The process started picking up the field of mathematics we wanted to focus on. Given that we are at different levels of understanding mathematics, the field that balanced the interest that each of us had and the inherent complexity of the topic was selected. After narrowing the search, we limited the process to selecting an article that met the following conditions:

1. It has been around during the last century (XX).
2. Published by an author that had a h-index considered high for us ($h \geq 10$).
3. The publishing journal has an impact factor in that field such that makes it position in, at least, the quartile Q2.

Once we had the paper selected, we read the paper and understood the main gist of it. The next step was to find all the citations of this article. All of this work has been done using *ISI Web of Knowledge* and *Google Scholar*. After we had the list of all the papers that have quoted ours, we randomly subsampled the references to limit ourselves the evaluation of 4 references. This references were analysed to see the social implications that were drawn from them and classified whether they were giving a result that society would use straight away or if it was more a purely theoretical research. This gave us the direct applicability measure, which we have defined by:

\[ app = \frac{a}{n} \]  

where \( a \) is the number of articles that had direct applications and \( n \) is the total number of analysed articles. A direct application was considered to be something that society would benefit of right after its publication. For instance, in the economy field, knowing the best strategy to make contracts with a client could be considered as an application (if I start a business tomorrow, I can use it) whereas providing a new framework for the brain research is not (given that only researchers would benefit from this).

We repeated the process with the cites of the previous articles but now only with two referenced articles. This allowed us to obtain a general equation for the applicability: the direct applicability measure at a given depth, \( d \), is defined by:

\[ app_d = \frac{a_d}{n_d} \]  

Formally, we can define the methodology as a tree graph with the following parameters:

1. \( d = 0 \) is the root node. It has a branching factor of 4.
2. For $0 < d < 3$, they are the child nodes with a branching factor of 2.
3. For $d > 3$, there are no nodes defined. Therefore, a total of 29 articles were analyzed.

In the discussion for the method, we have thought about other ways to measure the social impact of game theory. For instance, another approach that we have considered was to obtain 29 articles (the total number of articles is constant along all the methods) selected at random from the main field, with the conditions stated above, and to read and classify them depending on whether or not they could be used as a direct impact for social implication. The problem we found with this approach is that, given that not all of us have the same mathematical knowledge, it would have been very hard to do regular meetings to share what each one of us understood of each article and it would have meant a greater load of work that was totally impractical.

Another one that we have discarded was to use another topic, for instance medical engineering. For a similar reason, the group is very heterogeneous and none have the same level of knowledge in each of the possible fields so we considered that the best idea was to pick a topic and a field that we were very interested in: this resulted to be game theory.

Our method has a few key points, which are:

1. The requirements of the papers.
   - Deciding papers that had been on the literature for the past century allowed us a timeframe with higher chances that the article has been cited already.
   - Similarly, by limiting the papers to authors that are well-known, we can ensure that the probability that this research has made it to a real life application is higher than if the author was a fresh PhD.
   - On the same line of reasoning, by picking a journal with high relative scoring in the field (i.e. high quartile), we cover papers that have been important in the moment of the publication.
2. The way of retrieving the citations. By randomly subsampling the citations from the root paper we eliminate bias from the results.
3. Interest on the topic was defined as a key point because that made us be motivated to learn and read the papers even though it was not our main strength.

However, we can also find some limitations:

1. Having more real knowledge of the subject would have made the task easier. However, by having such heterogeneous team, this weakness has been significantly reduced.
2. Mathematical research has not generally a fast way to get implemented directly into the society. At the same time, we cannot trace back research to an ancient time with good reliability.
3. Scalability issues: given that we are dealing with a non-cyclic directed graph [2], we have the following problem. If $b$ is the number of child nodes from a
parent node, the branching factor, and \( d \) is the depth of the tree, if we had to look 4 references for each node and we did 3 generations, we would have ended up with \( 1 + 4^1 + 4^2 + 4^3 = 85 \) articles to analyze in total. Therefore, we have a scalability problem, that we solved by setting:

\[
d_i = \begin{cases} 
4 & i = 0 \\
2 & \forall i \in \{1, 2\}
\end{cases}
\]

Where \( i \) is the branching factor from the node \( i \).

3 Results

After analyzing 28 papers randomly chosen, we have found that the basis in game theory from the root article has been extended through several and diverse fields. Each of the four articles that cite the initial one revolve around a different topic. That has allowed us to find the influence of Nash equilibrium theory in Economics, Biomedicine, Artificial Intelligence and Politics.

In Economics, it has been found that the notions from Nash equilibrium theory can be applied to a duopoly of companies in a specific market and the strategy that should be followed in order to obtain better profits. From this paper, two more were obtained, one talking about cost strategies in media marketing and another one about patent licensing, both of which are also applied in the society. The last level of the tree contains four articles about free and pay digital content strategies, the phenomena of social media, entrepreneurship and patents. Half of them were more theoretical, while the other two were more explicit about their direct application [3–9]. This means that from the Economics branch we obtained five out of seven articles whose knowledge has an impact in society and are directly or indirectly related to the topic of the original paper.

In Healthcare, the branch started with a theoretical article about free-energy principles and their relation to the brain. Nash equilibrium theory is mentioned in terms of the concept of cost minimization for biological processes. Two of the papers selected from all of those that cite the previous one revolve around the neuroanatomy of speech production and Brain-Machine Interfaces. The first of them is has a theoretical, while the second explains how BMI are useful for neurological recovery and neural control of machines. Lastly, in the third level two of the articles have current applications, one in data analysis with deep learning algorithms and the other with the treatment of speech disabilities, while the other two papers are related to research for the improvement of existing knowledge [10–16].

The third branch has been encompassed under the label Automation, but it also involves some content related to Engineering, Electronics and Mathematics. The first article is an extension of the root paper, also written by Nash. It is theoretical, but also gives an example of application of concepts in the game of Poker. From this one, two articles with real examples sprung out, one about an algorithm for Artificial Intelligence applied to non-cooperative games
and another one about power control of wireless data networks. In the last four articles we found three real applications about congestion management of liberalized electricity markets, maximization of the comfort of the inhabitants of smart homes with minimum energy consumption and reallocation of resources in wireless networks to maximize user experience. A theoretical article about the prospects of 5G was also obtained, but nowadays this technology has not arrived to the public [17–23].

Lastly, in the Politics branch we have found that much of the research done in this area is more theoretical than practical or applied in the society. The first paper selected revolves around the outcome of strategy-proof voting procedures of three or more alternatives. In the next layer we obtained theoretical articles about Nash equilibrium applied to welfare and the deliberative nature of democracy according to social choice theory. The only applied example, found in the third level of this branch of the tree, is an alternative to the Human Development Index used for policy evaluation, international comparisons, growth assessment, and inequality measurement. The other three articles provide theoretical knowledge about social choice rules, citizen participation in democracy and the changes in the instruments or processes used for governance[24–30].

Overall, a total of 14 studies of the 28 branched out from the root article have knowledge with a direct impact on society. However, the distribution has been found to be different depending on the field of study. A complete breakdown of the applicability of the papers analysed can be found in Table 1.

Table 1. Results of the research: on the totals of the left we can see the totals of applicability per depth index \(d\). On the bottom, the totals per sub-field.

<table>
<thead>
<tr>
<th>Economy</th>
<th>Healthcare</th>
<th>Automation</th>
<th>Politics</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>(d_1)</td>
<td>1.00 (1/1)</td>
<td>0.00 (0/1)</td>
<td>0.00 (0/1)</td>
<td>0.25 (1/4)</td>
</tr>
<tr>
<td>(d_2)</td>
<td>1.00 (2/2)</td>
<td>0.50 (1/2)</td>
<td>1.00 (2/2)</td>
<td>0.63 (5/8)</td>
</tr>
<tr>
<td>(d_3)</td>
<td>0.50 (2/4)</td>
<td>0.50 (2/4)</td>
<td>0.75 (3/4)</td>
<td>0.25 (1/4)</td>
</tr>
<tr>
<td>Total</td>
<td>0.71 (5/7)</td>
<td>0.43 (3/7)</td>
<td>0.71 (5/7)</td>
<td>0.14 (1/7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.50 (14/28)</td>
</tr>
</tbody>
</table>

4 Conclusion

The study of mathematics is a particular field of science for which is very difficult to see its social ramifications, one of the main reasons is because it usually takes a great amount of time to find an application for the particular research topic. With this research we wanted to obtain a feeling of the possible social outcomes along the time of a particular paper that focuses on the game theory subfield of mathematics. After analysing some of the papers that cite our Equilibrium points in n-person games, we have been surprised by the variety of fields it has had an impact in, especially for finding ramifications in politics theory.

We have found that only 50% of the reviewed papers have a direct application which was something to be expected given the nature of the chosen topic.
From these 14 articles, the fields that had a higher social impact are economics and automation, with each one representing 35.71%. Following is the area of healthcare with 21% and finally a minimal impact in politics with 7%. These percentages are not surprising, we can extract a correlation between the areas that use mathematics as the base for their research and the actual application of Nash's Equilibrium definition.

Although we have been able to find several papers with direct applications, we are aware that this is also a result of the paper chosen as the base of our analysis. For another mathematical research topic we might have had more trouble finding its social impact.

Future indications for a deeper research could involve increasing the branching factor, the depth of the tree (as depicted in section 2) and picking as root different papers from different subfields of mathematics. Tuning the measurement of the applicability with the previously stated indications would help to answer the more general question of Social impact of mathematics.

References

A systematic review on common factors for successful, long-term University-Industry Collaborations

Domingo De Abreu, Frida Hentschel, José Piñero and Marielby Soares

Master in Intelligent Interactive Systems
domingo.deabreudias01@estudiant.upf.edu
frida.hentschel01@estudiant.upf.edu
josegregorio.pinero01@estudiant.upf.edu
marielbymercedes.soaresafonso01@estudiant.upf.edu

Abstract. The current technological and economically competitive global markets encourage the collaboration between academia and industry to develop new, efficient and viable products and services. In some cases, the collaboration efforts are not effective or do not lead to long-term cooperative partnerships. This paper evaluates four empirical long-term collaboration cases between universities and industry to identify management factors that have been proven to work correctly, and the relation of these factors to current literature. The outcome was a collection of key, common factors perceived as successful for long-term university–industry collaborations.

Keywords: University-Industry collaboration, Collaboration success factors, Long-term collaboration, Common management practices, Systematic review

1 Introduction

University-Industry collaboration is commonly referred as the interaction between a higher educational system - as a whole or any of its faculties- and industry aiming mainly to encourage knowledge and technology exchange [1]. These collaborations have had a long history, as means of building on knowledge stock to improve, optimize and/or create competitive products and services. Although collaborations between universities and industry are known to bring different benefits to both parties, some can be overshadowed by the need of considerable management efforts in order ensure a successful one [2]. Nevertheless, there has been an increase in these collaborations (e.g. Bodas Freitas et al. [3]; D’este et al. [4]) due to the rapid technological change, competitive global economies, environmental pressures, and budgetary needs - this last one applies specifically to universities [5]. Consequently, the need for more and better relations between universities and industry has led to specific literature reviews and case studies' analysis (e.g. Bruneel et al., 2010[6]; Ankrah & AL-Tabbaa, 2015[7]) on the subject. Nevertheless, there hasn’t been a study that correlates the common factors in long-term collaborations projects with the current literature.

Therefore, the purpose of this paper is to find common factors that have been proven as successful in collaboration projects, by bringing together the empirical evidence.
provided by four separate long-term case studies and the results of a complete review of the current literature in the field of academia-industry collaboration. The selected cases provided an opportunity for the empirical study of successful management practices within university-industry collaborations in different countries. These cases include Procter & Gamble collaboration with University of Cincinnati; Audi Hungary Motor and University Széchenyi István; Siemens and the University of Tennessee; Dairy Crest and Harper Adams University. Since individual circumstances apply to each project, specific factors cannot be generalized or categorized as being always successful. Nevertheless, when key factors are shown to have had an impact in several projects, this could indicate common successful and/or problem areas in collaborative efforts. The results of this analysis aim to be an indicator of such factors.

In the following sections of this paper we will present the process undertaken to reach our conclusions. In section 2, the research methodology is presented. Both, for the case studies' selection and the literature available on the subject. In section 3, we present the common factors encountered as successful management practices for long-term collaborations in the study cases and throughout the literature review. Finally, in section 4, we present the overall conclusions of our analysis. Moreover, a table synthesizing the selected case studies can be found in the Appendix of the paper.

2 Research methodology

For this research piece, a systematic literature review method was used, that is, a study that seeks to answer a clearly formulated question by finding, describing and evaluating evidence from all published research on topic(s) related to that question within a specific set of boundaries [8]. It differs from traditional narrative reviews by adopting a reliable and rigorous process that reduces subjective bias and lowers the risk of overlooking relevant literature [9]. Nevertheless, there is still a potential risk to exclude relevant studies and the limitation of understanding different processes to interpret the results from the studies.

To select the cases of study, we searched for successful ones presented in the web pages of the University-Business cooperation initiative of the European Commission and in the University-Industry Demonstration Partnership organization [10, 11]. We selected two sources to include examples from different continents and analyzed by different groups. The criteria used to select studies were the relevant and long-term partnership between the industry and universities; and the presence of a research center/facility jointly created to conduct projects in different industry sectors. The year of publication was also taken into account for our case studies' selection, as we wanted to use recent information to draw conclusions applicable to current and future years. We decided to exclude the studies that contained little detail about the collaboration between industry and university. Considered four as a good number of studies to analyze, we expected to find in these repeatable patterns of success factors in the scope of this research piece. After the cases were selected, the strategy followed to extract relevant information was a recurrent reading and understanding of each case study, as
well as comparing and contrasting the factors encountered in each one, presented in a table.

To choose the literature to compare, we selected all databases included in Web of Science as our search sources. We defined a specific combination of terms to search: The title should include one word from each of the following groups: (i) 'university' or 'academia'; (ii) 'business', 'industry' or 'firm'; (iii) 'collaboration'. Additionally, we made sure that one of these concepts was also integrated in the topic of the paper: 'success factors', 'barriers', 'obstacles' or 'lessons learned'. This process yielded more than 70 results as our initial sample; then, following our inclusion criteria, we identified every peer-reviewed article from 2010 to 2016 published in the English language. There was no restriction on country of origin or source sector. After applying this criteria, we had 23 results.

One additional step was followed to determine the final sample of 8 articles: reading the titles and abstracts to identify if the study focused mainly on university-industry collaboration and if it included factors that facilitate or help to prevent failure in the collaboration. An initial evaluation of quality was performed in this process and then, we filtered out every article with no citations or published in a journal with an impact factor in Q3 or Q4.

A secondary search was carried out afterwards, based on references contained in primary sources. This process resulted in 2 additional relevant papers. The final process of this phase was to extract relevant information from the articles. This involved repeated readings of the articles, summarizing key concepts and conclusions in a document. After that, the strategy was to compare the findings between the articles and between the cases of study.

3 Results

In this section, we show the key factors that are shared across the different case studies and compare those findings with the current literature. By doing so, we can identify key aspects that would help model a long-term collaboration between the industry and the universities.

3.1 Universal Factors: Trust, Commitment, Common Goals

Throughout the different case studies there were multiple mentions of trust between both partners as a key factor. Trust is what allows both parties to understand each other and have the confidence that the opposite party will cooperate in a fair and helpful way. This will enable both parties to act in a responsible manner upon the problems that may arise during the course of the collaboration [6]. In fact, trust has been identified several times before as a key issue affecting the success and effectiveness of collaborations [12, 13]. Furthermore, according to Santoro and Gopalakrishnan [14], the trust level between partners represents the capacity of both parties to work together and demonstrates a willingness to understand and align themselves with the needs and
expectations of their partners, thus leading the industry to be more inclined to invest across different areas that eventually could lead to a long-term relationship.

A common feature across all our case studies is that, throughout the process, there was always a firm understanding commitment to the scope and the expectations set for the collaboration project. Both parties were aware of their role and responsibilities in their collaboration effort since beginning. This led to a trust improvement amongst the partners, and derived in further collaboration projects. Furthermore, if both parties were able to establish expectations on the scope, output and risks of their research, they would be more willing to assist and support their counterpart to achieve the collaboration common objective [15], and is also widely documented by Barnes et al. [16].

3.2 Mutual Benefit

An important factor for the continuity of the collaboration effort in each of the cases was the common understanding that the research project would be mutually beneficial, and realizing the strategic importance of the goals that both parties were trying to achieve. This is something that Barnes et al. [16] covered by stating that the mutual objectives are "based on established common areas of interest, mutual strategic importance and mutual benefit". Moreno, Bruneel et al. [6] also mention that one of the barriers for successful collaborations is the concern that both parties have regarding their own benefits of the collaborative research projects. Thus, it is important that each party involved establishes beforehand their expected outcomes and projected benefits from the collaboration. This is true, especially for the industry, whose ultimate goal is to gain some level of proprietary gain that can be used as competitive advantages within its market and area of influence.

3.3 Clear Communication

Another factor mentioned on different literatures [6, 16, 17] and the case studies was that effective communications are a key aspect when dealing with collaboration projects. Moreno, in each of the cases the strategies for effective communication were mentioned as key factors throughout the negotiations, discussions and planning of the research projects. From project management teams that developed a clear communication strategy that involved frequency of the meetings and the use of written communications, to the establishment of lead researchers from both parties that would resolve any issues and keep regular contact, makes us think that the success of a collaboration venture is highly influenced by how fluent is the communication between both actors, and is also documented by Barnes et al. [16].
3.4 Clarity in Intellectual Property Rights

Although universities have been constantly increasing their attempts to capture intellectual property (IP) rights for the last years [18] for financial gain, most research efforts focus on only acquiring new knowledge, whereas the industry R&D ventures try to obtain knowledge for the sake of gaining a competitive advantage [19]. Therefore, in both the literature [6, 7] and our case studies the intellectual property rights are amongst the highlighted factors when dealing with university-industry collaboration, where all partners should be able to achieve a reasonable level of proprietary gain from the collaboration [16]. In particular, our case studies reflected that when dealing with big companies (e.g. Audi, P&G) the negotiation of the IP rights was deemed as a barrier in the early stages of the venture and that clear guidelines for IP ownership are paramount for moving forward in their shared efforts.

4 Conclusions

This research has examined and discussed the main success factors that emerge in both, the published literature and evidence drawn from four case studies, each examples of long-term university-industry collaborations. Amongst all mentioned issues, some common themes persistently emerged, indicating that standardized good practice factors can be applied to augment the probabilities of successful collaboration efforts between academia and industry.

The good practice factors presented in the analysis of this research is based on 4 key areas, representing the major common themes in the selected case studies and in our literature review:

- Trust, commitment and common goals in both parties are universal collaboration factors that are always present in the successful collaboration efforts reviewed.
- Large degree of clarity when it comes to mutual benefits that the collaboration will bring to each party.
- Clear and continuous communication channels should always be open to ensure information exchange.
- Precise definition of intellectual property rights in the beginning.

Many of the factors discussed in this paper can be considered generic, being also applicable to any other collaboration sorts (e.g. industry-industry, within academia, government-industry, etc.). While this case study research has encountered key success factors which should prove good practices, our research was limited to a small number of successful collaboration case studies. Future research should concentrate on further validation of these findings through additional international cases. Additionally, further research should broaden the research to find evaluation factors for partner selection, management of the outcomes and common traits in unsuccessful collaborations. Such work would enable further testing and shaping of a good practice model, including these success factors, that could be applied to maximize the benefits of university-industry collaborations.
References


### Appendix: Comparative Table of the Study Cases

<table>
<thead>
<tr>
<th>Reference</th>
<th>Descriptions</th>
<th>Activities / Outcomes</th>
<th>Motivation</th>
<th>Barriers</th>
<th>Key Success Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dairy Crest Innovation Centre at Harper Adams University: A comprehensive long-term agri-food collaborative relationship[3]</td>
<td>Dates: Since 2014 Location: United Kingdom Parties involved: Dairy Crest (DC) U: Harper Adams University (HAU) Description: Long-term agri-food collaborative partnership Importance: - Industry with traditionally low R&amp;D - Most Innovative Contribution to Business-University Collaboration award in Times Higher Education Awards 2016</td>
<td>- Set up the Dairy Crest Innovation Centre on the campus of HAU (Shared facility) - Joint research and development of R&amp;D projects - Curriculum development providing additional business-relevant education (design modules and courses) - Annual placement of students and joint final year projects for students - New lecturership in Animal Science and Bioinformatics to expand research</td>
<td>Both: - Support economic growth - Increase interaction and knowledge flow between academia and the food industry for commercial and societal benefits - Encourage career options in the food industry for young people (Agri-food sector is short of high-quality young graduates) HAU: Opportunities for students to apply their knowledge in an industrial setting DC: - New product development with scientific research and technology - Quality access to research and academics - Consolidating the majority of their technical and research expertise, and equipment in one place</td>
<td>At the beginning of the partnership, when the teams did not know each other very well, the negotiations about the partnership's features went slower so the dialogue had to be promoted Securing external funding for joint research activities</td>
<td>- Common vision - Genuine interest in the field - Mutual trust - Throughout the whole process, it has been essential that both partners understand that it is a win-win solution for everyone - High level of commitment - Sufficient resources - Good project management - Negotiations, discussions and planning must be done well ahead of time - The leadership and ownership of the results is agreed upon at the beginning of each joint research project - Relationship grows over time</td>
</tr>
</tbody>
</table>
| **Procter & Gamble / University of Cincinnati Simulation Center [4]** | **Dates:** Since 2008  
**Location:** Cincinnati, Ohio  
**Parties involved:**  
I: Procter & Gamble (P&G)  
U: University of Cincinnati (UC)  
**Description:**  
Strategic academic partnership to develop modeling and simulation capabilities for advancing product and process development  
- Set up the UC Simulation Center  
- Faculty and students contributing in real projects in R&D and Product/Supply for P&G  
- Numerous publications, conference presentations, and patents  
**UC:**  
- Opportunities for students to have insight or collaborations on industrial challenges  
- More funding for research  
**P&G:**  
- Get talent to develop and apply modeling and simulation capabilities across the company  
- Train talented researchers for future recruitment  
**Negotiation of Intellectual Property Rights:**  
Bureaucracy issues in complex initiatives which affects relationship building, especially in the early stage  
- Clear guidelines for IP ownership  
- Highly committed leaders on-site from both P&G and UC (single point of contact to avoid confusion)  
- Focus on the core win-win outcomes  
- Allow students, faculty and researchers to focus on individual projects and areas of expertise  
- Follow UIDP Guiding Principles for university-industry collaboration  
- P&G project teams willing to fund students, train them, and mentor them over a period  
- Locality: Encourage colocation. P&G personnel can spend part or all of a work day at the Center without impact to the work and work/life balance.  
- Relationship grows over time  
- Good relationships between the partners |

| **Audi Hungaria and SZE: The Audi Faculty making Győr the most significant automotive, economic and cultural center of Hungary [3]** | **Dates:** Since 1999  
**Location:** Hungary  
**Parties involved:**  
I: Audi Hungaria Motor Kft. (AHM)  
U: Széchenyi István University in Győr (SZE)  
**Description:**  
- Set up the Audi Faculty at the university.  
- Cooperative development of research, curricula and courses in management, vehicle engineering and manufacturing  
- Provision of practical experience for the university’s student  
Both:  
- Provide high quality education – achieving the highest-level practice-oriented vehicle engineering education in the world  
- Provide practice-oriented engineer training  
- Facilitate and grow cooperation between the research sector and the automobile industry  
- Make Győr one of the most significant automotive, economic and cultural centres of both Hungary and Europe. This includes achieving desired quality thresholds in terms of international competitiveness for both the  
- Working with such an influencing partner needs careful balancing with the needs and requirements of other regional and national stakeholders and partners  
- Location: R&D activities of Audi are carried out at their headquarters in Germany. This leaves only small room for the university to engage in the company’s research activities  
- Firm commitment from collaboration partners to investment and long-term working  
- All the partners share the goal of fostering regional competitiveness and economy  
- Relationship grows over time  
- Good relationship between the partners |
Key strategic partnership for educational, professional and scientific collaboration.

**Importance:**
- Subsidiary of a multinational company
- University with a lot of collaboration with industry (research centers and partnerships)
- Partnership also benefited from the local and regional ecosystem
- Joint training programmes
- Support industry-relevant scientific research work
- Bachelor's and Master's courses in the Automotive Engineering field

- University and the region (Audi is a benchmark in this regard)
- Increases the number of German-speaking engineers in the region

**SZE:**
- Engineering students have access to state-of-the-art technical and technological knowledge, which helps them meet industrial requirements and are better prepared for the world of work [3]

**AHM:**
- Attract highly trained engineering students to work within the region and specifically at Audi’s engine and vehicle manufacturing plant in Győr
- Gain access to highly skilled trained graduates as future workforce

<table>
<thead>
<tr>
<th>Dates: Since 2005</th>
<th>Part 1: Set up Scintillation Materials Research Center (SMRC)</th>
<th>Part 2: Lead the way in the research and development, impacting the lives of thousands of people each day (cancer detection, homeland security)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location: Tennessee, United States</td>
<td>- Joint projects to discover and develop new material technology</td>
<td>- Provide a competitive business advantage for Siemens</td>
</tr>
<tr>
<td>Parties involved:</td>
<td>- Over 40 journal articles have been published, numerous conference presentations have been made, and several patent applications have been filed</td>
<td>- Provide research topics for students</td>
</tr>
<tr>
<td>I: Siemens Medical Solutions</td>
<td>- Discover breakthroughs in medical imaging technology for improved patient outcomes</td>
<td>- Difficulties working out the contractual legal details</td>
</tr>
<tr>
<td>U: University of Tennessee</td>
<td></td>
<td>- Frequent communication (face-to-face weekly and monthly meetings with written reports)</td>
</tr>
</tbody>
</table>

**Description:**
Partnership that seeks to develop new scintillators at the University of Tennessee (UT) that will enhance the performance of medical imaging devices that are manufactured by Siemens Medical Solutions.

**Difficulties:**
- Frequent communication (face-to-face weekly and monthly meetings with written reports)
- Relationship grows over time
- Patience and flexibility by both partners for legal issues
Systematic overview of industry-academic collaborations in the field of artificial intelligence

Kami Artik\(^1\), Enric Cosp Arque\(^2\), Sara Galindo Martinez\(^3\), Juan Luis Moral Perez\(^4\), SiMing Zhang\(^5\)

Masters in Intelligent Interactive Systems
\(^1\)kaminurhan.artik01@estudiant.upf.edu, \(^2\)enric.cosp01@estudiant.upf.edu, \(^3\)sara.galindo01@estudiant.upf.edu, \(^4\)juanluis.moral01@estudiant.upf.edu, \(^5\)siming.zhang01@estudiant.upf.edu

Abstract. The field of Artificial Intelligence (AI) is rapidly accelerating, and experiencing a renewed renaissance. It has undeniable applications across every discipline in science and our day-to-day lives. While the possibilities of Artificial Intelligence are undeniably rewarding, there exist numerous challenges to successfully pursue ambitious applications. Symbiotic engagement between academia and industry is crucial to the success of growing the field of Artificial Intelligence. We conducted an exhaustive literature review to investigate the current climate in these collaborations, their goals and the methods they use. We focused on three such collaborations as case studies. Specifically, we looked at their mission statement, how they are facilitating their vision and the demographics of their employees. We found that in general, the goal of these collaborations is to advance scientific achievements, propel economic growth, and improve humanity’s general well-being. They attract world-class top-talent and are heavily funded by both the public and private sectors.

Keywords: Artificial Intelligence, Machine Learning, Research, Academia, Industry, Collaboration, Economy

1. Introduction

Academia and industry need a symbiotic relationship for both to proper in the field of Artificial Intelligence. Many graduates from academia are quickly absorbed by the industry. Moreover, research work in universities are often used by the industry and turned into products and services, which can benefit society. Therefore, it is natural and necessary for academia and industry to collaborate. Companies in the industry often will request academic institutions to tailor their
courses and research on their main focuses. In exchange, the industry will fund the university’s research, by setting up laboratories, designating industry chairs and providing guest faculty and placement opportunities. Often new research topics, and products arise out of these successful collaborations. [3]

Artificial Intelligence (AI) is a cross-disciplinary field. Therefore, it entails all these facilitators and more in order to group various professional and academic profiles together. This is difficult to happen in universities, as each department is often too remote. Companies are more prepared for these kinds of purposes. However, they can lack the specific expertise in some matters and often the time investment for long term research.

The raise of AI and its applications in recent decades has pushed important change in our society. Many companies and governments are vastly investing in computer power, big data and machine learning. In particular, it is estimated that it employs nearly 6% of the US workforce[1]. Companies such as Facebook are quickly rising in market capitalization lists, where the top three companies are IT companies: Apple, Google and Microsoft.

AI is both a trendy field and a new one, where a lot of research still needs to be done. We present a discussion on the possible benefits of industry-academia relationships in this field. In order to do successfully advance AI, specific expertise, computational resources and a multi-disciplinary workforce is needed. Therefore, we draw and explore the following questions within this report with three case studies:

1. Why is there a need for academic-industry collaborations?
2. How are these relationships being built?
3. How beneficial are they?
4. How can they be improved?

Thus far, there has been no specific literature for academia-industry relationships in the field of AI. This paper aims to provide a systematic assessment and report on the climate of these collaborations, while specifically looking into three world-leading collaborations as case studies.

We begin by enlightening the reader on the need and importance of healthy academia-industry relationships. Moreover, we discuss the main issues in building these links and some proposals to overcome them. Next, we provide an overview on general relationships. Furthermore, we will bring specific data on its goals and forms in AI. Last, we conduct three case studies, each of them focusing in a particular relationship: Vector institute in Toronto, Canada, DeepMind in London, UK, and
MIT-IBM Watson AI Lab in Cambridge, USA; followed by conclusions and future work prospects.

2. Research methodology

We conducted an exhaustive literature review in order to construct the academia-industry collaboration landscape. We followed an observational model to gather our data using resources found on the Internet. In our case studies, we ensured that, we not only gathered information from the official website, but also news articles and fact-based critical reviews to avoid bias. Furthermore, we used LinkedIn as a data-mining tool to gather information about employees of the institutes mentioned in our case studies to construct demographic profiles.

Given the scope and context of the topic we are investigating, it was not possible conduct any experimental research. In our LinkedIn demographics study, we randomly selected 20 employees from each company using a random number generator to avoid bias. The sample size of 20 was selected because it is sufficiently statistically significant, and it would be difficult to gather more data due to the manual collection methods used. Also, the companies mentioned in our case studies are quite new, so some may not have more than 20 researchers yet.

3. Results

While both academia and industry can benefit from collaborating with each other, universities are more prepared to do deep and long term research in specific topic, and companies are best suited to multi-disciplinary projects, as well as having more straightforward funding structures.

Nevertheless, there are some issues regarding these relationships. Researchers at universities are often focussed on early stages or more ‘blue skies’ and futuristic research than the kinds of marketable, application-based research that industry supports. According to Prof. Shunji Murai, Founder, Asian Association of Remote Sensing, “The leader should be a person who has ability of management experience and marketing sense with innovative motivation and positive thinking, which academic scientists sometimes have not enough. Industry should be the CEO while academia should be the CTO.”[2] Moreover, financial matters may be sticky. Prof. Michael Goodchild, Emeritus Professor at University of California states, “Student internships and short-term residence of industry representatives in universities could be a win-win situation for both parties as the costs are minimal.
comparing to the benefits”. However, this kind of collaborations, are insufficient to achieve major long-term goals. Other challenges that surfaced include a lack of trust over issues such as intellectual property, uncertainty about the potential benefits of working together, and the difficulty on both sides of finding the time for initial exploratory conversations. Finally, there is also well known disparity between universities and businesses in the kinds of outputs that would make such collaborations seem worthwhile. While businesses may be seeking shippable products, academics prize excellent research results and publications.

**Keys to success in science and technology**

Numerous sources have cited the following keys for a successful relationship: [5]

- Build a shared vision that identifies the goals and provides a clear framework for all the parties involved.
- Find leaders, capable of crossing the the bridge between business and academia to promote ties.
- Create a shared platform in order to facilitate communication between entities.
- Establish a well-defined agreement for the use of resultant intellectual property in advance in a transparent way.
- Invest in long term relationships, which allow parties to share risk and accountability without overburdening a single entity.

We will now focus on AI collaborations. When looking at interactions within this field, and according to a survey carried out by the Computation Community Consortium [1], there are three main goals:

1. Ideas with actionable intellectual property, such as algorithms, designs/architectures, open source software, or new research directions
2. Resources, data or service moving between academia and industry, when the product of the collaboration takes the form of software or hardware artifacts
3. The transfer of people, research and students, with specific skills to industry, or for the creation of an ecosystem.

Once the goals are defined, we can look at ways of collaboration that are suitable for achieving those objectives. According to the Computation Community Consortium [1], the most common collaborations are:

1. Contracts from companies to academic institutions with a specific statement of work.
2. Fellowships/Internships/Grants/Gifts from industry to academic institutions or units.
3. Direct skill transfer. Instead of contracting a person to do the work, the company hires somebody to teach the skills and knowledge needed to do the work.
4. Shared entities, like on-site labs where personnel from different institutions are embedded.

3.1. Case Studies

IBM - MIT Watson AI Lab

As everyone knows, IBM is the largest information technology company in the world and is a pioneer in Artificial Intelligence. IBM has explored the application of AI across many areas and industries for years. To accomplish this, IBM researchers invented and built a cloud-based AI platform especially for businesses, developers and universities, which is called Watson. The projects carried out in this platform were diverse but all of them denote a significant change in our society. Some of them were focused on fighting cancer, minimizing pollution or enhancing agriculture [8].

Collaboration method and capacity

Their objective is to continue working on fundamental advances in AI, for this reason the 6th of September of 2017 they invested $240 Million over a 10-year period for a joint research with the Massachusetts Institute of Technology (MIT) called the MIT-IBM Watson IA Lab [6]. This partnership would mean a major technological breakthrough for Artificial Intelligence as both have unmatched talent.

The new lab will be one of the largest long-term university-industry AI collaborations to date, mobilizing the talent of more than 100 AI scientists, professors and students to pursue joint research at IBM’s Research Lab in Cambridge, Massachusetts and on the neighbouring MIT campus.

Research fields

IBM and MIT plan to issue a call for proposals to MIT researchers and IBM scientists to submit their ideas to push the boundaries in AI science and technology in several areas.
The Lab is focused on advancing four research pillars: AI Algorithms, the Physics of AI, the Application of AI to industries and Advancing shared prosperity through AI [7]. The objective of each pillar is the following:

- **AI algorithms**: develop advanced algorithms in Machine Learning and Reasoning in order to be able to face more complex problems.
- **Physics of AI**: find out more information about new AI hardware materials, devices and architectures to optimize and speed up the algorithms developed.
- **Application of AI to industries**: develop new applications of AI for professional use like health care and cybersecurity.
- **Advancing shared prosperity through AI**: explore how AI can deliver economic and societal benefits to a broader range of people, nations and enterprises as well as study the economic implications of AI and improve prosperity and help individuals achieve more in their lives.

Another distinct objective of the new lab is to encourage MIT faculty and students to launch companies with the purpose of commercializing AI inventions developer there. In addition, the lab’s scientists will also publish their work, contribute to the release of open source material and, last but not least, foster an adherence to the ethical application of AI [9]. The later, will be really important for our daily life if their projects succeed in the near future.

Prior collaborations

The relationship between MIT and IBM is not surprising at all as in 2016 they announced a multi year collaboration focused on the advanced scientific field of Machine Vision which is a core aspect of AI. Furthermore, IBM and the Broad Institute of MIT and Harvard have established a 5-year $50 million research collaboration on AI and genomics.

**Demographics data for employees (researchers/scientists only)**

![Demographics data for employees](image)

**Fig. 1.** Highest level of education obtained by employees (researchers/scientists only) of IBM-MIT Lab
DeepMind Technologies is a British artificial intelligence company, acquired by Google in 2014, the company has created a neural network that learns how to play video games in a fashion similar to that of humans. Their mission is to push the boundaries of AI, developing programs that can learn to solve any complex problem without needing to be taught how. DeepMind’s AlphaGo program beat a human professional Go player for the first time in 2015 [10]. DeepMind started engaging with the academic communities in order to attract artificial intelligent talents from universities and research institutions. As a result of this collaborations, scientific publications are getting actively published by DeepMind every month.

Collaboration with McGill University

In July 2017, DeepMind opened the first ever international AI research laboratory in Edmonton, Canada in close collaboration with McGill University. Canada is globally recognized as a leader in artificial intelligence research [11].
DeepMind engineers see open collaboration between company research labs and academia as central to the future of AI.

Training of university students

Along with the collaborations with researchers, DeepMind engineers are giving lectures in the universities such as Oxford and University College London (UCL). ‘Advanced Topics in Machine Learning’ class in UCL will be overseen by Thore Graepel, who is a research lead at DeepMind and a computer science professor at UCL. It will focus on areas like deep learning, reinforcement learning, and natural language understanding. UCL is aiming benefit from the training provided by the DeepMind staff. Besides, DeepMind started to teach Deep Learning for Natural Language Processing advanced course at the University of Oxford’s Department of Computer Science. This applied course, focusing on recent advances in analysing and generating speech and text using recurrent neural networks [12]. Furthermore, three Engineering Science researchers from University of Oxford were hired by Google DeepMind to work on image recognition and natural language understanding [13].

Funding of labs and students

Moreover, DeepMind is providing sponsorship for several research labs and their PhD students to pursue their own research priorities in whichever way they choose, including the University of Alberta, University of Montreal, University of Amsterdam, Gatsby Unit at UCL, NYU and Oxford. This year, company provided full funding for 15 doctoral students in Oxford.[14]

Demographics data for employees (researchers/scientists only)

Fig. 4. Highest level of education obtained by employees (researchers/scientists only) of DeepMind
Vector Institute

The Vector Institute was created in an effort for Canada to dominate the space of Artificial Intelligence. Some of its founding members and directors and world-leading scientists in this field, include the father of deep-learning, Geoffrey Hinton.

Mission

“The Vector Institute will drive excellence and leadership in Canada’s knowledge, creation, and use of artificial intelligence (AI) to foster economic growth and improve the lives of Canadians.”[15]
It aims to attract the best international talent in the field of artificial intelligence and machine learning, and grow AI-based innovations by utilizing the powerful potentials of deep learning and machine learning. They will collaborate with Canadian companies and institutions in order to achieve world-class results, while also assisting the growth of AI-based business (i.e. start-ups) within Canada.

Funding and Capacity

The Vector Institute employs only world-class Research Scientists, and Postdoctoral Fellows. While the Vector is not a degree-granting institute, students in postgraduate study from an affiliated school can work with the institute’s researchers. Currently there are two affiliated universities: University of Toronto, and University of Guelph.

The Vector Institute has received $125 million in funding from the Government of Canada to pursue its endeavours. On October 18, 2017 the they also received a $30 million investment from the Government of Ontario to grow the number of Applied Master’s students in the AI-field within five years. Additionally, more than 30 companies have contributed a total of over $80 million to support the Vector Institute.

Methods [17]

• Implement proposals and curriculums for affiliated schools
• Facilitate all aspects industry experience for researchers (e.g. consulting, business development)
• Provide extensive computational infrastructure
• Access to a broad range of datasets
• A team of experienced software engineers who will assist in improving the software development process, and ensure that deliverables meet the standards for commercial use

Researcher requirements [18]

• PhD in Computer Science or related field
• Strong publication record in top-tier journals and conferences (mostly first author)
• Strong expertise in core areas of machine learning, or applied machine learning
• Interest in starting academic research groups, exploring opportunities in the field to start a new business, work closely with real-world data and problems
Demographics data for employees (researchers/scientists only)

Fig. 7. Highest level of education obtained by employees (researchers/scientists only) of Vector Institute

Fig. 8. Field of study (at highest level) by employees (researchers/scientists only) of Vector Institute
4. Conclusions

We have summarized the findings in our case study in the table below.

Table 1. Summary of academic-industry collaboration case studies

<table>
<thead>
<tr>
<th></th>
<th>IBM-MIT</th>
<th>Vector Institute</th>
<th>DeepMind</th>
</tr>
</thead>
<tbody>
<tr>
<td>University</td>
<td>MIT</td>
<td>University of Toronto</td>
<td>Montreal Institute for</td>
</tr>
<tr>
<td></td>
<td></td>
<td>University of Guelph</td>
<td>Learning Algorithms (MILA)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>McGill University</td>
</tr>
<tr>
<td>Company</td>
<td>IBM</td>
<td>Large companies (Google),</td>
<td>Acquired by Google in 2014</td>
</tr>
<tr>
<td></td>
<td></td>
<td>startups, incubators and</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>government</td>
<td></td>
</tr>
<tr>
<td>Main Field</td>
<td>AI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Objective</td>
<td>Working on fundamental</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>advances in AI, both</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>academic and commercial</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>purpose</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contributors</td>
<td>100 AI scientists,</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td></td>
<td>professors and students</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Where is the</td>
<td>IBM’s Research Lab in</td>
<td>Toronto, Ontario</td>
<td>London, UK</td>
</tr>
<tr>
<td>company?</td>
<td>Cambridge, Massachusetts and on MIT campus</td>
<td></td>
<td>Edmonton and Montreal, Canada</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mountain View, California, USA</td>
</tr>
<tr>
<td>Studied sub-field</td>
<td>Computer vision, Health</td>
<td>Neural networks, Probabilistic models, Statistical theory</td>
<td>Health care Solving Game</td>
</tr>
<tr>
<td></td>
<td>care, Cybersecurity,</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Natural language</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>processing, Machine</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>learning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Investments</td>
<td>Private sector (IBM)</td>
<td>Private sector (Google)</td>
<td>Private sector (Google)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Public Sector</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Government of Canada and</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Province of Ontario</td>
<td></td>
</tr>
<tr>
<td>Outcomes</td>
<td>SDK and Papers</td>
<td>-----</td>
<td>Publications</td>
</tr>
</tbody>
</table>
The AI community has made several steps in putting together the efforts of both academia and industry with success. There are remarkable examples, such as the ones presented in our case studies: IBM-MIT Watson, DeepMind and Vector Institute. All these and more have found different approaches to maximizing benefits and avoiding typical problems in collaborations, such as intellectual property ownership issues. Specifically, AI collaborations are not only seeking scientific achievement, but also impulsion economic growth, and enhancing the well-being of mankind. To do so, they attract world-class top-talent in several disciplines to work together in both public and private sectors.

At this moment, such academic-industry collaborations are an emerging field, so there are few examples. To improve this work in the future, we can generalize the investigation to more collaborations, and expand our LinkedIn search to all the listed employees. To do this, we will need to automate the data-mining process.

References


The Importance of the Supervisor During PhD-related Crisis

Omar Ghatasheh
Bernat Puig Borja
Requena Emilia Vulpe

MWC, MHS, & CBME
omarmohammadabdehamed@estudiant.upf.edu
bernat.puig01@estudiant.upf.edu
borja.requena01@estudiant.upf.edu
emilieodorova.vulpe01@estudiant.upf.edu

Abstract. The development of a PhD thesis is a non-trivial task that can, at some point, become an overwhelming burden for the student. A great part of PhD students suffer from PhD-related crisis throughout the development of their thesis. When not handled properly, this can lead to mental illnesses that put at risk the health of the student and the success of their thesis. Here we study the role of the supervisor during PhD-related crisis and how they can help prevent bad outcomes. Independent literature review on the issue was performed and a survey with current and former PhD students was conducted. The study shows that the supervisor can play a key role to overcome, mitigate and, in best case scenario, even prevent PhD-related crisis from happening.

Keywords: PhD life, crisis, supervisor, PhD difficulties, mental health, academia

1 Introduction

Developing a PhD thesis is an arduous task that requires strict hard working from the student. This constant pressure can build up to the point that it can become an overwhelming burden to handle and can lead to a crisis [1-4]. In the context of this paper, crisis has a broad definition, which is related to a difficult period for the PhD student and can have various causes. This could refer to a feeling of being overwhelmed, being unable to cope, fear of lacking the necessary skills and time or doubting the successful completion of the research. In extreme cases crisis could also be mental instability or even depression. Regardless of the severity, a situation of crisis requires effective management in order to prevent more serious problems that can have a strong impact on the mental health of the student. The supervisor plays a key role in the PhD life of their students, being them in source of motivation and guidance [5]. However, despite having excellent academic skills, supervisors may lack the expertise to provide the required support without being aware of it [6].

There is recognized need for a standardized framework for skills and qualities that supervisors must develop in addition to the academic capabilities they possess [7]. The goal of our review is to assess the need for this framework to include skills which will help supervisors to manage situations of crisis, demotivation and even depression through which their students might be passing. The outcome can provide a better understanding whether the PhD experience can be enhanced to improve its quality. In the end, both students and advisors can be aware of the most common issues and prevent them from happening in the future.

The remainder of the paper is organized as follows. In Section 2, we describe the research methodology used in this study. Section 3 presents and discusses the results. Finally, Section 4 contains the conclusions drawn.

2 Research methodology

For this research, we hypothesize that the supervisor can have an impact during difficulties in the PhD life of a student. A combination of research methodologies is used. Firstly, we perform a deep literature review on the difficulties PhD life can bring and what is the role of the supervisor. This historical method allows the collection of qualitative data from different independent sources on the topic in an inexpensive manner. However, despite
multiple sources being included, the results may be biased. Thus, to further assess the impact of the supervisor during PhD-related crisis their student may be going through, an observational method was used in the form of an anonymized survey. It was completed by current and graduated PhD students. This method provides quantitative data on the issue to complement the results from the literature review and reduce the bias from it. The questions of the survey can be found in the Appendix.

The aim is to identify if difficulty crises are common among PhD students and if/how the supervisor could help to avoid or mitigate such hard times.

3 Results

3.1 PhD Life Difficulties

The results of the literature review suggest that the complexity and the extensiveness of writing one's own research can sometimes overwhelm a PhD student [1-4], to an extent where they doubt their suitability for the task and their capability to manage it [8]. In addition, they can be the cause for serious anxiety for the researcher and even lead to a lack of their confidence and questioning the academic skill set and achievements which brought them to the programme.

Another difficulty for PhD students arises when the expectations for the time and effort involved become a reality [9-10]. Based on the experience of formal PhD students, the time they had to dedicate for their studies and experiments, deprived them from having a healthy social life [11]. This can result in loneliness and isolation, thus leading to an unstable mental state.

However, the main problem in academia is that mental health issues are a serious concern due to the research conditions in universities [12,13]. The literature review further suggests that stress is widespread and increasing [14], and it is observed predominantly in younger academics [15].

The mental health of a PhD student evidently affects the outcome and the quality of the PhD, in addition to putting at risk the student themselves [16]. A crisis the researcher may be going through can directly affect the research team they are part of, and in particular their supervisor. Furthermore, it may result in the PhD student quitting their studies and the research industry altogether [17]. In fact, according to literature, 30 to 50 percent of PhD students don't finish the programme [18]. This poses a serious threat to the research industry, potentially making it less attractive for new candidates [19].

3.2 The Role of the Supervisor

Supervisors are heavily important for the success of the PhD system and they should be able to dedicate the time to ensure the PhD students are on track to produce a high-standard thesis. Thus, it is necessary for the good progress and the success of the PhD student that there exists a good relationship between them and their supervisor [5,20-22]. Furthermore, the supervisor's mentoring could affect the candidate's confidence and effectiveness [23,24]. According to Sheeban [25] good supervision involves providing support, guidance, constructive appraisal and encouragement to develop individual thinking.

Despite different suggestions from literature on the qualities a good supervisor should possess, unfortunately, there is no common framework to what skills are required for best support hence Pearson and Kayrooz [7] point out the need for developing such a framework. Supervisors may end up having too many students, thus not having enough time for a particular individual, having personality clashes with their student, or even lacking the required expertise in the supervision field [26]. Alternatively, it is possible that they simply don't have the right skill set, and hence, which is also a very important part of supervision as pointed out by Pearson and Kayrooz [7] and Seagram et al. [27]. Despite the numerous categories in which a supervisor may fail, supervisors often don't have awareness that their supervision is not good enough [6]. However, these problems are felt by the students and can be very demotivating, result in depression or even failure to complete the PhD [28].
3.3 PhD Students Survey

3.3.1 Survey Results

The results of the survey are presented in the figures below. Figure 1a shows how the participants are distributed across different stages of their PhD. Figure 1b shows the percentage of students that have experienced a PhD-related crisis during their PhD life. Figure 2 displays a distribution of people from different stages of their PhD with relation to suffering a PhD-related crisis. Figure 3a presents the percentage of people who consider the supervisor is playing an important role during time of crisis versus the ones who think oppositely. Figure 3b shows the distribution of different ways the supervisor can help during hard times of the student.

![Figure 1a](image1.png) ![Figure 1b](image2.png)

Figure 1. a) Percentage of people in each PhD stage. b) Percentage of people that have been through PhD-related crisis.

![Figure 2](image3.png)

Figure 2. PhD-related crisis suffered by PhD stage.

![Figure 3a](image4.png) ![Figure 3b](image5.png)

Figure 3. a) Percentage of people who avoided PhD crisis due to their supervisor versus those who avoided it for other reason. b) Roles of the supervisor that could have helped people who suffered from a PhD-related crisis.
3.3.2 Survey Results Discussion

The survey was completed by 45 PhD students mostly from UPF PhD programmes to whom the survey was distributed via e-mail. The students were classified in 4 different categories according to their research stage, as it can be seen in Figure 1a. From them, 60% have experienced or are currently experiencing a PhD-related crisis (Figure 1b). However, from Figure 2, it is observed that the answers are enomogeneous along the different PhD stages. The percentage of crisis increases along the stages, having 36% of crisis rate for beginners, 55% for middle-stagers and 91% for those who are finishing. Hence, a big portion of PhD students suffer from at least one crisis along the process.

Furthermore, Figure 3 analyses the importance of the supervisor in the time of PhD-related crisis. From Figure 3a, we observe that 67% of those who have not been through a crisis state that it is thanks to the support of the supervisor. This makes it evident that the supervisor plays a predominant role in this difficult period and can help the student to handle the situation. Besides, from Figures 2 and 3b there can be observed that only 3 out of 27 people who have suffered a crisis could not be helped by the supervisor. The different ways in which supervisors can be helpful to resolve the crisis could be attributed to the individual student’s situation and reason for the crisis. The most popular responses, however, indicate the need for more implication in the thesis, better emotional support or helping to keep up the motivation.

In short, the majority of the participants have been through a PhD-related crisis, especially those in middle or advanced stages. In most cases (24/27) this could have been avoided with the help of their supervisors. In addition, 67% of those who haven’t experienced a crisis stated that it was due to the support of their supervisors. Hence, the supervisor has a great influence in the PhD life quality, mental stability of the student and the success of the PhD thesis. Unfortunately, most of the supervisors, despite being respected academics, did not manage to provide the necessary guidance.

4 Conclusions

PhD life can be too challenging at times and can result in students experiencing a PhD-related crisis at some stage of their research, which in turn may affect negatively the completion of the research or the student’s mental health. Supervisors play an important role for the success of the PhD thesis. However, the need for them to provide more support during challenging times is recognised.

Our findings reconfirm an important issue which needs to be resolved in order to ensure the good health, experience and success of PhD students. Therefore, supervisors should be more conscious of their responsibility and the role they play in the academic life of their students, and seek more opportunities to provide help and support, especially during times of difficulties and stress. Parallel to that, students must not be afraid to ask for help when truly needed. The key to good progress and results is the good relationship between the supervisor and the student.

This survey has been carried out with a small part of the PhD student community but large enough to give evidence for a latent problem. A potential improvement would be to bring the study one step further and make it more extensive and conduct it in larger scale in order to determine which actions could be taken to improve PhD life conditions. One possibility could be to evaluate the performance of the PhD supervisors in order to implement some measures such as the regulation of the maximum amount of students that they can have depending on their available time, provide necessary supervisor training or encourage student-supervisor communication.
References


Appendix

This is an anonymous questionnaire. Data is collected as part of an assignment for the Research Methods course. Ethical implications are handled in the context of the course. Data will be deleted by the students collecting it once the data is analysed in an aggregated fashion, only anonymous and aggregated data will be presented in the assignment.

What is your current PhD stage?
- Beginning
- Middle stage
- Finishing
- Finished

Have you been through some kind of PhD related crisis?
- Yes
- No

If so, the role of the supervisor could have helped you... (multiple choice).
- No, the supervisor could not help me.
- By showing my interest in my work.
- By keeping up the motivation
- By providing emotional support
- By implicating more on my thesis

If not, the supervisor had a key role in avoiding such crisis
- Yes
- No
Research as a Career:

Common and Specialized Backgrounds

Genua, Ibai; Phillips, Jake; Mogas, Sira; Mas, Ariadna
Computational Biomedical Engineering

a904045@alumni.tecnun.es
jphillmatic@gmail.com
sira.mogas01@estudiant.upf.edu
ariadna.mas20@gmail.com

Abstract. Being a researcher implies the acquisition of certain skills and knowledges, which vary depending on the specific field you would like to work. In this paper, a study is carried in order to determine the specific requirements and the main organizations and companies for certain job positions which were previously defined, i.e. astronauts, marine biologists and biomedical engineers. To do so, an extensive job searching has been taken, looking for different job listing pages to have a general idea of which are the main asked aptitudes in each field. Moreover, a sample of 20 Linkedin profiles of each research area has been analysed to obtain possible significant differences between the education of each scientific professional.

Keywords: requirements, research career, education

1. Introduction

In the last few years, a significant increase in the number of programs for undergraduate students to start or improve their research careers has been noticed. Being a researcher implies many things, such as having communication, problem solving and critical thinking skills, originality, creativity, curiosity and teamwork, among others [1].

In practice, there are many obstacles to face when developing a research career. There is not a defined single path to undertake an academic or research position [2]. This diversity is positive, as there would be people who would undertake a PhD candidature and endeavour to build a research profile, whereas others who will prefer the industry and would look for a tech company, as well as those pursuing an academic position [3].

In this paper, we are going to study which are the requirements to perform research in different fields. We have chosen three research careers which we believe come from different backgrounds with different requirements. The three chosen careers are astronauts, marine biologists and biomedical engineers.

Astronauts are people trained to command, pilot and perform research operations as crew members of a spacecraft. As we will find with many other research careers, becoming an astronaut requires a considerable amount of dedication, original thinking, and field specific training. In the past, space programs have largely been developed by government organizations such as the U.S. National Aeronautics and Space organization (NASA), the Russian Federal Space Agency (RFSA), and the China National Space Administration (CNSA). In recent years, private corporations such as SpaceX, Blue Origin, and others have set their sights on operating their own space programs. While the motivation behind each program is different, they are all similar in their ultimate goal of expanding human knowledge through meticulous research.
Marine biologists study the oceans and investigate the behaviour and physiology of marine species, together with their diseases, the environmental conditions and the impact of human activities, e.g. commercial shipping, plastics, chemicals, trying to minimize sea life damaging.

Their research activities involve testing [4], taking ocean samples, analyse the behaviour and evolution of some organisms, measure properties of the environment, etc.

Biomedical engineers apply engineering methods for solving medical problems. The biomedical engineering research field can be very wide: from working in the laboratory with biological samples to work with computers, manipulating biostatistical data. The objective of the research can be different between researching fields. Some research studies aim to obtain an innovating pharmaceutical drug; other studies want to create a robotic device; other studies look for algorithms that can solve medical problems. Despite having different objectives, all of them aim to diagnose, treat, prevent or cure patients.

Taken into consideration that large variety between scientific careers which was previously reported, in this study an aggregation is trying to be made in order to highlight specific scientific qualities and help future undergraduate students to orientate themselves and improve their scientific profiles, making them more competitive for their future work. Moreover, to avoid biases, the methodology was carefully reported following an accurate criteria to select the studied candidates and their profiles.

2. Research methodology

As previously mentioned in the introduction, three different research careers were chosen, i.e. astronauts, marine biologists and biomedical engineers.

To understand the similarities and differences between these careers, we first identified industry leaders in all three research field to have an initial reference. From there, we then searched in different databases for job listings positions and we identified the requirements for each of the research areas.

We also tabulated the education background obtained from a sample of LinkedIn and agency profiles (n = 20) for each particular research position. It consisted on looking the course degree and the possible complementation with a master or a PhD.

Using the data gathered from the requirements listed in job listings we made comparisons on multiple levels.
Firstly, we compared the different degrees course within the research career types, i.e. the background, and fixed the main requirements for each of the job positions.
Secondly, we compared the degree level distribution between research career type. In this case, we compared the different level of education, i.e. degree, master or PhD, between the three research groups.
Finally we explored similarities and differences between the various research types.

These comparisons were done by generating histograms which categorized education based on data gathered in a thorough search of profiles in each respective career type.

A workflow of the steps followed is shown below.
3. Results

After a deep study and after searching for these three research subtypes, we have obtained a general idea of which are the requirements and skills needed to work as a researcher for each field.

3.1. Astronauts

Figure 2. Astronaut’s Field of Study for Highest Degree. This figure represents the different studies of astronaut sample.

Becoming an astronaut is a childhood dream for many driven scientists, and as a result, the selection and training processes are quite competitive. Even in some of the earliest groups, when space flight was incredibly dangerous, only seven out of 500 applicants were selected to become NASA astronauts. To be an astronaut, there are two aspects to be taken into consideration. Astronauts must be physically and mentally fit. Physically there are requirements on their stature, eyesight, and blood pressure. Mentally they must pass trials to prove they are able to withstand the psychological pressure of being surrounded in the great emptiness of space. Secondly, an average of
1000 hours of flying experience are required or if not, to have a PhD in biology, chemistry, engineering, technology, mathematics or physics. The limitations in this respect are loose for astronauts. Flight skills can be trained, but expertise in a specialized field takes years to develop. Figure 2 demonstrates that even in a sample of 20 NASA astronauts there are many different backgrounds. Each astronaut is expected to contribute a unique research focus to their squad [5].

3.2. Marine Biologist

It can be concluded that to get a marine biologist job [6] you need to have, apart from the bachelor’s degree, advanced knowledge from a Master or PhD in marine biology, marine ecology, conservation or a related science which include knowledge about biology, maths, zoology, English and physics. On the other hand, the main requirements include personal skills as motivation, problem-solving and especially being good at communication and writing due to the relevance of taking extensive notes about their observations. Also physical requirements as having a good hand-eye-coordination and fitness to do activities such as diving.

In most of job applications, have practical experience [7] is always a good advantage, so in the majority of the universities you have the opportunity of doing an internship or a fellowship. This last aspect is very relevant because then the person is more qualified to lead a group of people and take decisions. To complete this study, we have searched 20 samples of profiles in LinkedIn (see Fig.2) to see if, in general, they comply with the requirements established.

![Figure 3. Marine biologist’s Field of Study for Highest Degree. This figure represents the different studies of the marine biologist sample.](image-url)
3.3. Biomedical Engineering

There are variations in the requirements asked for the researches even within this subgroup. Everybody was asked to have Bachelor title and some knowledge about the Biological aspects of the job (acquired in the Bachelor or in the post-graduated studies). Apart from that, the requirements asked differ considerably. In those jobs related to laboratory research, some laboratory experience is required but this doesn't happen if applying to biocomputational research jobs, at least not in all of them [8].

Figure 4. Biomedical Engineer’s Field of Study for Highest Degree. This figure represents the different studies of the biomedical engineer sample.

As we have explained previously, an important requirement for being a researcher in any field, is to have at least a Master or PhD. In the following graphic (Fig. 5) we can observe the results of the highest degree of the 20 people studied for each career.

Figure 5. Highest Degree of every people studied in LinkedIn. This figure represents the highest studies of the researchers in each field.
4. Conclusions

After analysing the results, some clear conclusions can be extracted. First, academic formation, i.e., bachelor degree or higher education, is crucial to devote a scientific career in a specific field. While academically, when dropout in academic formation within groups is minimum. For marine biologists, the focus is set on biology, zoology and oceans. For astronauts, the main topics are physics and engineering and aeronautic concepts. Finally, for biomedical engineers, the education is focused on computational strategies and robotics, all applied to the biological world.

Although the diversity among scientific careers, it has been seen a common tendency in terms of education, as the majority of the candidates have undertaken a masters programme after the bachelor to have a higher and more specific formation and just few of them have completed a PhD candidature.

After seeing how determinant is education, the required skills for each of the scientific careers have been analysed. Here, despite the differences of each specific job, more coincidences can be observed. At the end, being a researcher implies curiosity, perseverance, a positive attitude and be always looking forward to progress and success. Focusing on those specific skills, astronauts are asked to have strong physical and psychological conditions to face different extreme situations. Marine biologists requirements are more related to academic bases, which are fundamental for their work, together physical skills such as diving. Good communication abilities are also a plus when applying for this career. For biomedical engineers, one of the most relevant skills is computation and programming, together with a good command of technologies. The biological background can be obtained by reading about the research topic you are working at, but the technological skills are harder to acquire and need to be trained.

Having seen the results, although different profiles have been considered, it is difficult to analyse and represent the whole scientific community, as many backgrounds should be considered. For this reason, this study could be improved by adding more scientific careers and thus, having more data for further comparisons and conclusions. Moreover, it could also be improved by incorporating more Linkedin profiles.

Apart for those weaknesses, this study has fulfilled the main objective as it shows significant differences between research careers, but also common patterns within the scientific field. Thus, it gives a general guide or trend of different research careers and the steps which have been followed.

References


