Davinia Hernández-Leo, Kostantinos Michos (Eds.)

Research about 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
December 2016

Editors

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

Kostantinos 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/

Preface

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 2016-2017 academic year.

The papers were written as part of an integrative assignment entitled "Research about 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 17 papers were written by the students and presented in the classroom. Assessment included peer-review by students during the presentations, 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 scientific dissemination for the general public, the social impact of science, interdisciplinary in research, open science and PhD process and life.

Table 1. Topics and accepted papers

Topics	Submitted	Accepted to publish
Social Impact of Science	2	1
Gender in science	1	0
Social networks for researchers	1	0
Research integrity	1	0
Interdisciplinarity in research	2	1
Research careers	2	0
Research-Industry collaboration	2	0
Scientific dissemination for the general public	1	1
Open Science	2	1
Science communication	1	0
PhD process and life	2	1
Total	17	5

Table of Contents

PhD Selection: Factors to take into account 4

Natalia Delgado, Tessy Troes, Meghana Sudhindra, and Helena Cuesta

Portrait of a science disseminator 13

Beatriz Cabrero Daniel, Alberto Martínez-Rodríguez, Mari Celi Morales Muñoz, and Carla Ten Ventura

Social impact of Artificial Intelligence: Robotics 19

Kushagra Sharma, Manaswi Mishra, Siddharth Bhardwaj

Is Interdisciplinary research the way to go? 24

Adrian Arroyo Perez, Giorgos Neokleous, Pantelis Stylianides, and Vibhor Bajpai

Analysis of AcousticBrainz, Essentia and Freesound platforms from an Open Science perspective. 30

Pablo Alonso-Jiménez, Gerard Erruz, and German Ruiz-Marcos

PhD Selection: Factors to take into account

Natalia Delgado, Tessy Troes, Meghana Sudhindra, and Helena Cuesta

Master in Sound and Music Computing
Universitat Pompeu Fabra (UPF), Barcelona, Spain
{natalia.delgado01, tessyannevera.troes01, meghana.sudhindra01, helena.cuesta01} @estudiant.upf.edu

Abstract. Taking a PhD is a time demanding career choice and well known for the intensity of the dedication required. Hence, taking the right choices before starting a PhD can be crucial for the upcoming years. In this paper we compile what we believe are the most important issues to be considered when selecting the PhD. We identified the following factors as relevant: the location, both from a personal and professional perspective; the financial resources and funding; the topic selection; and future ventures into academia or industry. We explore these categories and conclude by stating the most important points to be considered when selecting a PhD.

Keywords: PhD, research, funding, University, selection

1 Introduction

University students and graduates get confronted at some point during their studies whether to take their gained skills and enter the professional market, or pursue an academic career. Doing a PhD is the main path to become an academic researcher, being successful on this journey that usually takes more than 4 years however depends on many factors. This paper looks at the various factors that should be taken into account for selecting the best possible PhD. This research stems from a personal motivation of creating an overview answering personal questions, but also breaking with the one-sided, glorified presentation that can be found on official web pages, which pursue a financial interest. Choosing a PhD is a decisive step in the academic future and should hence be treated with care.

Most sources deal with how to write the PhD thesis and leave surrounding factors that inevitably influence the level of research and writing process out. This has mostly to do with these factors not being quantifiable and/or subjective. This paper aims at gathering information from both informal and formal sources in order to create a map of factors that should influence the selection process.

In sought of the research problem, we identified four fields as relevant for the prospective PhD candidate: the location, the topic selection, the funding and future career opportunities. In this paper, we aim at exploring these points and their impact on the PhD selection process.

Firstly, we detail the research methodology. The following result section will provide a description of the different factors, then a conclusion will sum up those factors according to our research.

2 Research Methodology

In this project, we follow a literature search based methodology. In our case, this meant gathering information from different sources and extract conclusions afterwards. This methodology presents some advantages: collecting qualitative data from many different sources, not just focusing on one; do independent research on each of the categories we address; avoid focusing only in one case study (i.e. a single PhD student), which could be very useful to obtain detailed information about some specific processes, not would not provide the general information we are interested in.
Since this is a broad topic that can be approached from different perspectives, the sources used range from students' blogs over formal papers and articles, to conversations with PhD students. Some informal sources have been used (i.e. blogs where PhD students talk about their experiences) because during the research, we noticed that personal factors and experiences from others were also very relevant to answer our research question.
The research has been divided into four main categories, which together provide a great overview of the factors to take into account when selecting a PhD: location (where to study: country, city, university), topic (research field and question), funding (types of funding, institutions, scholarships), and future opportunities (academia, industry).
The types of data that have been used for the research are basically two. First, most of the information gathering has been bibliographic: articles, books, student blogs. Then, some statistics have also been taken into account to obtain quantified data. In Figure 1 the methodology is displayed.

3 Results

In this sections, we present the results of the research, divided into the four categories that have been mentioned.

3.1 Location

Where to do a PhD? This section discusses important aspects when being confronted to this questions, namely: choosing the country and city, university or research lab.

Country & City
Choosing the country and city in which one will carry out their PhD includes many things to consider: language(s) spoken and culture within the country, the

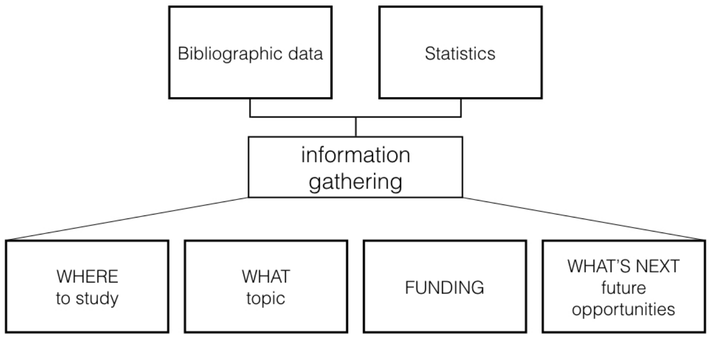

Fig. 1. Methodology chart

lifestyle of the city and extracurricular activities offered, or salary compared to living expenses.
Below, Table 1 compares the average salary of a PhD amongst four countries known for their research. The prospective student then shall research into their expected living expenses, which heavily depend on personal lifestyle and city.

	Spain	Netherlands	Sweden	USA
PhD Salary	17.000 €/ year	52.500 €/ year	29.000 €/ year	28.000 €/ year

Table 1. Average salary of a PhD student. From [1]

University
One way of determining interesting universities is to search for authors of one's research field and determine where they do their research. Some countries, ratings and more objective evaluations for universities and departments can be found, e.g. in the UK exists the Research Excellence Framework [2]. One shall be careful however, as the evaluations in the same university can vary heavily for different departments.
Research Lab
As mentioned in the previous part, tracking down relevant and published authors in one's research field and hence potential supervisors is a crucial aspect of the selection process. It will also help narrow down and clarify possible topics. In

these lines, one has to consider at what career point the potential supervisor is currently at - veterans might be the obvious, but not always the best choice. As stated by Dr Nathalie Mather-L'Huillier in [3] "work with a younger academic at the start of their career can have many advantages. They are likely to have much more time to give to you and will be very pleased to have you on the team".
Furthermore, once the options come down to up to three research labs, it is recommendable to visit them (i.e. at an open day) or schedule a video conference with members of the lab. It provides insight into interpersonal aspects and general environment: working schedules, PhD students lifestyle, how team work is organised. All factors that can only be determined on a personal basis, to know whether one is feeling comfortable in a place that will be frequented for more than three years.
To finish with this section and help you with thee evaluation of the possible research lab when visiting or doing a video conference, a few questions are proposed in [3]:

- If the project is team-based, how do they operate?
- How often do fellow researchers present group seminars?
- Is the supervisor often absent (and does it make a difference when they are)?
- Even if yours is an individual research project (as is common in the Arts, Humanities and Social Sciences) the culture of your department is still important. Are there other PhD students working in your field? Do they participate in any regular shared activities such as internal discussion groups or presentations?

3.2 Funding

The need for funding, thus financial resources, depend on the personal circumstances [4], i.e. age, lifestyle, previous experience and salary. Self-financed PhDs are an option, however most candidates rely on funding to attain financial stability. The application process should be started 6 to 8 months in advance, as it depends on the nationality and offered vacancies [5]. Candidates eligible for funding should provide a decent CV and an excellent track records from previous research [4].
The different types of funding (with examples for Catalonia) are given in the table below [6]: Funding comes with certain conditions and restriction depending

Serial No	Government Funding	Private Organization Funding	University Funding
1	ERC	Santander Grants Program	UPF
2	MAEC - AECID	CatalunyaCaixa	FI
3	AGAUR	La Caixa Scholarships	FPU

Table 2. Different types of PhD funding

on its source, which creates advantages and disadvantages for the candidate that should be taken into account before the selection. They are listed in Table 3:

Advantages	**Disadvantages**
Financial self-sustainability	Unnecessary intrusion on research by funding agencies
Value for research	Conflict of interest may arise due to the private funding agencies which have a hold on the market

Table 3. Advantages and disadvantages of receiving funding

3.3 Topic Selection

There is not an algorithm to choose the right PhD topic. As Davis says in [7], everything depends on one's imagination and energy. According to all the literature that has been read for this research, there are two imprescindible aspects to take into account to choose the topic for a PhD: the candidate must be really interested in and motivated by the topic. However, most authors also agree that what is really motivating nowadays, might not be so in three or four years. According to [8], one might want to start a PhD out of two reasons: either they are passionate about research and gaining new knowledge in general or they have a lot of motivation for a specific topic. Usually, people wanting to start PhD have already been involved in some kind of research project, which facilitates the topic selection.
Clua-Losada points out in [8] that there are three relevant issues to take into account when choosing the topic: the viability (i.e. is it a manageable question?); humility (i.e. one may not always be loyal to one's first ideas, but this is not necessarily an issue); flexibility, to cope with unexpected changes. In [9], Leigh, a former PhD student, says: "if you don't have the slightest idea where to start looking for your topic, now may not be the right time to start your PhD". However, there are some options to look for topics, such as advertised PhD positions on websites (country-dependent information) and looking for existing projects. Another option highlighted in [9] is seeking professional help, i.e. ask academics or potential supervisors for advice. Furthermore, the topic has to be of relevance and novelty; in order for a PhD to be considered good, it has to make a contribution.
Wisker suggests in [10] that performing a good literature search and reviewing the work that has been done in the field will help identifying gaps in knowledge and hence enable the prospective student to find a topic to explore and to research about. The typical path, once some possible locations have been identified, would be, according to Wisker [10], to join a team with an ongoing project in that university and adapt a topic to be in the scope of that project.

3.4 What's next?

Depending on the subject and the progress made, a PhD can span more than half a decade. Indeed, data from the PhD Completion Project found that in America, less than 50% complete their degree in 6 years or under, no matter if the student was working in Life Science, Engineering or Mathematics / Physical Science [1].
However, having in mind what the desired outcome is, can heavily influence previously mentioned factors in an early stage of the PhD process.
According to statistics published by the Atlantic [11], drawn from National Science Foundation data, the number of completed PhD holders going into the industry grows closer to those going into academic positions. The article reveals another important argument for post-PhD decisions: the average salary in the industry is more than 20% higher than in academia. Also, The Guardian states that 'The number of PhDs vastly outnumber demand for postdocs and permanent contracts.' [12]. Hereby follows a list of possible post-PhD career opportunities:

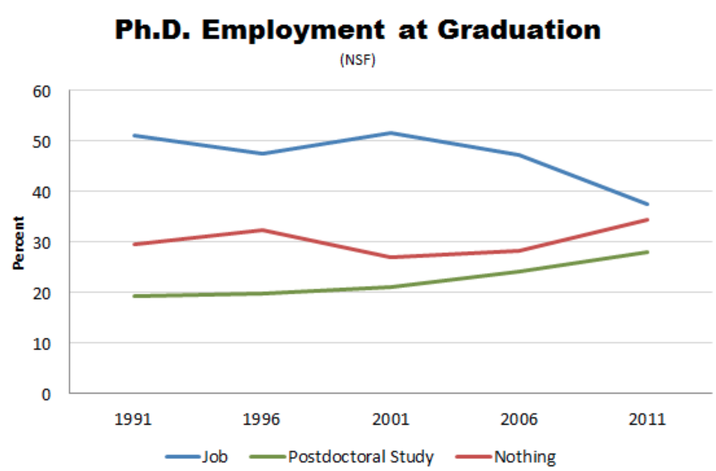

Fig. 2. Evolution of PhD employment at graduation between 1991 - 2011, in the USA

- Moving up the academic ladder: Postdoc, research assistant, lecturer/ professor:
  Postdoc and research assistants continue conducting research after completion of their PhD. This position is often on a temporary contract and low salary [13], however, gives the opportunity to transition to a new place and

[1] http://www.phdcompletion.org/

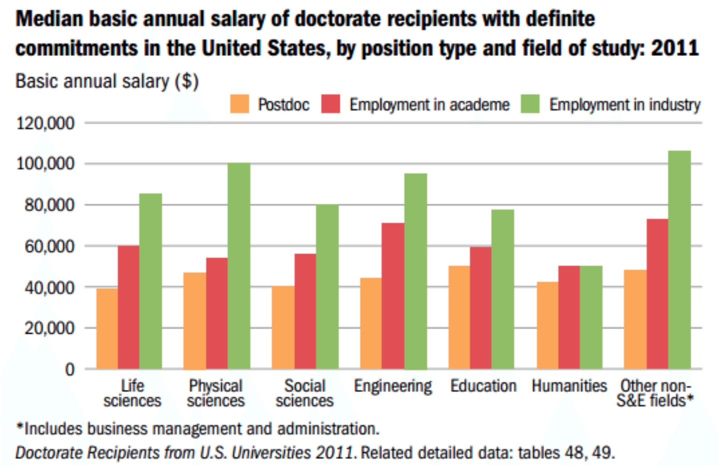

Fig. 3. Average salary in academia and industry in first year after PhD, graduating from different fields

/ or continue researching, teaching and publishing in the lead-up (usually between three months and three years [14]) to getting a permanent academic faculty position.

- Research & Development, Publishing, Consultancy:
  The PhD acquired skills can be put into use in an outside-academia world. These groups work in research, design new products, and also take part in strategy decisions, both in the public or private sector. Dedicated communities exist in order to ease the transition into and guarantee networking among PhDs across industries[2].
- Start-Up: spin-off of research:
  A start-up offers more flexibility, more control, but also more responsibility for an aspiring PhD; and a possibility to take one's research project or associated projects of the research lab to the industry. Within UPF's Music Technology Group, appropriate examples would include Voctro Labs [3] and ReacTable [4] and Musicmuni labs [5].
- Professional:
  Last but not least, a change of career, in no way related to the completed PhD is a possibility. Reasons for that can be broad, however, no research nor meaningful articles could be retrieved by us.

---

[2] http://phdsatwork.com/
[3] http://www.voctro-vocaloid.com/
[4] http://reactable.com/
[5] http://musicmuni.com/

4 Conclusions

Not without a reason, selecting a PhD is called a process: it is a highly personal and thus subjective decision. In this paper, we have attempted to illuminate the relevant factors separately:

In terms of location, a suitable environment is needed, which includes both personal and professional factors; related to the country and city (language, culture, lifestyle, expenses) as well as university and research lab (present research authors and projects, teamwork and working environment).

Related to funding, it can be provided by either government bodies, private institution or come directly from the university: the prospective student shall gain an expansive knowledge of funds available, as they might be linked to conditions and restrictions, and be aware of applying in the right time window.

In regards to finding a topic, three aspects are decisive in the candidate's attitude: before the PhD, they need to have interest and motivation; in addition to that have a clear research domain and question; during the process, flexibility and adaptive capacities are needed.

We deduced that future opportunities for completed PhDs are not of utmost priority at the start, but become more important along the way and might influence the actual course of the PhD.

In a real-life situation, all listed factors should be taken into account, however, seeing how intertwined they are and individual needs and expectations, the path to selecting a PhD is a winding one.

References

1. GlassDoor: PhD Student Salaries. https://www.glassdoor.com/Salaries/stockholm-phd-student-salary-SRCH_IL.0,9_IM1136_KO10,21.html
2. HEFCE, SFC, H.D.: Research Assessment Exercise. http://www.rae.ac.uk (2008)
3. Mather-L'Huillier, N.: Deciding Where to Study for Your PhD. https://www.findaphd.com/advice/finding/where-to-study-your-phd.aspx"
4. Pitchforth, J. *et al.*: Factors affecting timely completion of a PhD: a complex systems approach. Journal of the Scholarship of Teaching and Learning **12** (2012)
5. Council of Graduate Schools: Financial Support. http://www.phdcompletion.org/promising/financial.asp
6. UPF: Universitat Pompeu Fabra Webpage for PhD. https://www.upf.edu/web/doctorats#altres
7. Davis, D.: PhD Thesis: Where do I Start?.
8. Clua-Losada, M.: Choosing and Developing Your PhD Research Project. https://www.findaphd.com/advice/finding/choosing-phd-research-project.aspx
9. Leigh, L.: How To Select a PhD Research Topic. https://100thousandwords.wordpress.com/2012/08/31/how-to-select-a-phd-research-topic/ (2012)
10. Wisker, G.: The Postgraduate Research Handbook. 2nd edn. Palgrave Macmillan (2007)
11. Weissman, J.: The Ph.D Bust: America's Awful Market for Young Scientists - in 7 Charts. The Atlantic (2013)
12. Anyangwe, E.: QA best bits: Life after a PhD. https://www.theguardian.com/higher-education-network/blog/2011/apr/27/career-options-after-a-phd
13. Fazackerley, A.: Why are many academics on short-term contracts for years? https://www.theguardian.com/education/2013/feb/04/academic-casual-contracts-higher-education (2013)
14. Claeys-Jackson, D.: Your PhD, what next? https://www.prospects.ac.uk/postgraduate-study/phd-study/your-phd-what-next#what-do-other-phd-graduates-do (2016)

Portrait of a science disseminator

Beatriz Cabrero Daniel[1], Alberto Martínez-Rodríguez[2], Mari Celi Morales Muñoz[3], and Carla Ten Ventura[4]

Master in Intelligent Interactive Systems
[1]beatriz.cabrero01@estudiant.upf.edu
[2]alberto.martinez02@estudiant.upf.edu
[3]araceli.morales01@estudiant.upf.edu
[4]carla.ten01@estudiant.upf.edu

Abstract. In the last few years, human communications have deeply changed both in form and content. Social networks and broadcasting technologies, as well as mobile devices, are the root of this evolution. And so has changed scientific diffusion. The monopoly of traditional publishing methods is slowly disappearing and the popularity of Web 2.0 tools is on the rise. Researchers can exchange ideas, data, and knowledge with its peers much faster than before. There is also a tendency to address a wider audience by creating and broadcasting educational content. In this context, we analyse the differences between the profiles of traditional science communicators and the Web 2.0 educators and discuss the importance of role-models in science.

Keywords: demographics, open science, science dissemination, web 2.0

1 Introduction

Even if classical means of scientific communication (journals, conferences, printed books, etc.) are not likely to disappear in the near future, the growth and potential of Web 2.0 tools (Social Network sites (SNS), on-line archives, blogs, etc.) are triggering some important changes in publication of scientific results [1][2]. This evolution draws academia towards a more open framework - a context in which researchers share results, ideas, and knowledge much earlier and to a wider audience.

Traditionally, communication with the broader public has been done in conferences, press releases and by having reporters and writers publish articles in specialized magazines such as National Geographic. Nowadays, this knowledge dissemination can be done through the novel ICT tools: iTunes; SNS such as Facebook; YouTube or Vimeo; microblogging sites e.g. Twitter; on-line publishing platforms like WordPress, Blogger, or Research Blogging; etc [3].

One of the outcomes of opening science will be a gradual rapprochement between academia and the general public. Table 1 lists a sample of tools that

Table 1: Tools used by researchers and science educators based on the public they are addressing.

	Science 1.0 media	Web 2.0 communication tools	
Cooperation among researchers	Printed journals and books, conferences	Science-specific SNS (publishing of unfinished and/or negative results)	On-line repositories, specialized blogs, science and technology themed educational content
Diffusion for the general public	Science-themed magazines	Science-themed on-line magazines	

science communicators use based on their motivation and the public they are addressing [4].

Many have already addressed the situation of Science and Web 2.0 focusing on the knowledge sharing among researchers. Though many educators are reluctant to use Web 2.0 tools, opening science through technology is considered the key to the second scientific revolution by many others [2][5][6].

The publication of ideas, preliminary and negative results may cause an improvement of academia's performance e.g. preventing repeating experiments. In 2012, Nicole Forster was able to ask her peers for help with her research after none of her colleagues at the General Hospital Cancer Center in Boston, Massachusetts, were able to help her [7]. Another example is the contribution of Grigory Perelman, a Russian mathematician that published the solution to a famous unsolved mathematical problem at the open on-line repository ArXiv.org [8].

This paper focuses, instead, on the profiles of those researchers that try to reach broader audiences. By comparing the traditional science communicators, as opposed to scientists using features enabled by Web 2.0, we intend to gain a deeper understanding of the demographics of supporters of open science [7]. On the one hand, the novel publication methods of Science 2.0 provide multiple benefits for academics such as:

- removing biases due to publication fees and economic pressures,
- greater social and professional recognition (new sources for metrics),
- increase impact and visibility of research papers,
- decreasing the time needed to publish results,
- being able to share (and read) even negative or preliminary results,
- agile cooperation among researchers and peer review,
- exchange of ideas, views, experiences, and knowledge, etc.

On the other hand, the cultural impact of technologies like on-line encyclopedias and science and technology themed educational Facebook pages and YouTube channels, show the need for greater efforts to bring science to the general public (even though science related topics represent only 6% of the most visited Wikipedia pages [9]).

So, there is a demand for science and technology themed content and being a creator of such resources is both beneficial in terms of visibility and useful for work [7]. But, who is in charge of disseminating scientific knowledge? Is the profile of the Science 2.0 disseminator different from the traditional media communicator? What do they have in common? Have the role-models of Science, Technology, Engineering and Mathematics (STEM) professionals evolved?

We try to answer these questions by conducting a demographic study of science educators that address the general public. In the Research Methodology Section we describe how the data was gathered and how it was interpreted. The results of the study can be found in the Results Section as well as a discussion on other academics' results compared to the findings presented here. The last Section, Conclusions, we summarize the key ideas presented and propose possible lines of research that build on the results of this paper.

2 Research methodology

To compare the profiles of science communicators, we gathered data and studied it from different perspectives [3]. Authors from the National Geographic magazine represent traditional science communication and some of the most popular Social Media contributors represent the scientists 2.0. The features contained in the data are: the name of communicator, born date, the date of his/her work publication or creation, the age the author had when he or she published or created his/her work, owner's gender, nationality, current occupation, education, and popularity (only for TED conferences, and YouTube and iTunes content).

On the one hand, contributors of traditional media were chosen among National Geographic authors with an article in volumes 10 to 27 of the Spanish edition (corresponding to the years 2008, 2009 and 2010). Also, we gathered information about the scientists in the most popular TED conferences (according to YouTube's rank).

On the other hand, the platforms included for communication in 2.0 are Facebook, arXiv, iTunes, YouTube, and a number of blogs [3]. Unfortunately, often when studying Science 2.0 communicators one can only get a glimpse of who is really behind the content. The barrier between the editing team, the writer of the script, and the host is often blurry.

A number of academics have also gathered data about contributors of an specific site to conduct their research. In the Results Section, our statistics are compared with previously published studies [2][5]. Accessing the studies by these researchers, and therefore being able to compare our conclusions to theirs, was an easy task since those academics that conduct research about open science tend to be more likely to publish their findings in open archives.

3 Results

In Table 2 we show the resulting statistics extracted from the gathered data. We can observe that science communicators are more often male both in traditional and in novel communication media. Also, while in traditional science dissemination almost all of the communicators are more than 35 years old, the users of Web 2.0 tools are younger. In fact, the older *youtuber* we considered opened his YouTube channel being a 43 year-old while the average age for a National Geographic reporter is almost 50 years. This could result from the fact that younger

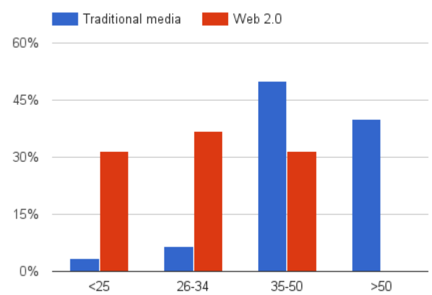

Fig. 1: Age groups based on the tools used for disseminating knowledge.

academics are more used to the Web 2.0 and therefore, are more prone to use general-application tools such as non-specialized SNS [7]. Meanwhile, older scientists benefit from the traditional media setting, where experience and a record of achievements are given great importance.

Table 2: Percentage of communicators by gender, media and age group.

Age	**Gender**	**Media 1.0**	**Science 2.0**
≤ 25	Male	0.00%	12.24%
	Female	2.04%	2.04%
26-34	Male	4.08%	12.24%
	Female	0.00%	2.04%
35-50	Male	16.33%	12.24%
	Female	14.29%	0.00%
>50	Male	24.49%	0.00%
	Female	0.00%	0.00%

Regarding the education level of communicators, there are not differences in the educational level between 2.0 and traditional media, nor between genders. Most of them have a bachelor's degree and pursued further studies (16% of the studied communicators have not studied science at university). Many of the traditional media communicators that we chose are also writers of a number of books but this is due to their age and experience. Also, while they complement their scientific knowledge with editing studies and journalism expertise, many of Web 2.0 educators are video producers.

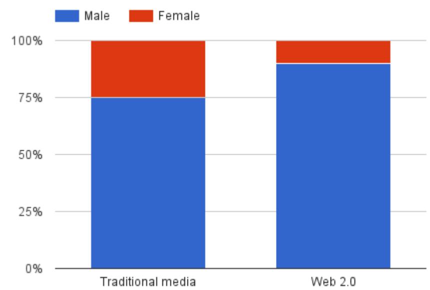

Fig. 2: Gender percentage based on the media to which they contribute.

About 20% of the educators we considered are women. When studying the female representation in Science 2.0 media, the percentage is even lower: 10%. The results shown in Figure 2 are consistent with the findings of Sajjapanroj et al., stating that Wikibookians are 97% male [10], and the results of a survey by the Wikimedia Foundation (13% of female contributors in 2011) [11][12]. As their male counterparts, communicators using Web 2.0 tools tend to be younger.

A 2013 survey by the International Women's Media Foundation stated that 66% of traditional media communicators are men. The under-representation of women in science communication is due to a number of cultural and environmental factors. According to the UNESCO Institute for Statistics, 30% of women chose to follow studies in science and 28% of the world's researchers are women [13]. Moreover, the trend shows a shrinking percentage of women in STEM [14].

4 Conclusions

The use of new communication and information technologies, both among researchers and as a bridge between academia and the general public, are the key to the second scientific revolution. As discussed by many academics and educators, there is a will to make science more accessible to a wider audience [15][6]: many regard public dissemination of science a researcher's duty and the demand for quality educational content is a growing trend.

In this paper we focus on the profile of content creators and science communicators that try to address a wider audience using some of the most popular Web 2.0 tools for general-use applications. We try to understand the difference between them and the traditional Science 1.0 communicators.

What we found is that even if science educators 2.0 are younger than the traditional media disseminators, there are still disproportionately male. This could be due to a combination of factors:

- the already smaller percentage rate of women in research,
- the lower percentage rate of women in STEM, and
- the lower female representation in broadcasting media.

We believe that gender imbalances in STEM affect creation of knowledge. Therefore, society and academia may need to address the opening science issue while providing female role-models. Future research might attempt to understand the reasons for this gap and try to find strategies for eliminating it.

References

1. Chin-Lung Hsu and Judy Chuan-Chuan Lin. Acceptance of blog usage: The roles of technology acceptance, social influence and knowledge sharing motivation. *Information & Management*, 45(1):65 – 74, 2008.
2. Michael Nentwich and René König. Cyberscience 2.0. `http://www.press.uchicago.edu/ucp/books/book/distributed/C/bo13136505.html`, June 2012. Accessed: 2016-11-23.
3. Jason Priem and Bradely Hemminger. Scientometrics 2.0: New metrics of scholarly impact on the social web. *First Monday*, 15(7), 2010.
4. Sara Kjellberg. I am a blogging researcher: Motivations for blogging in a scholarly context. *First Monday*, 15(8), 2010.
5. Caroline Lego Muñoz and Terri Towner. Back to the "wall": How to use facebook in the college classroom. *First Monday*, 16(12), 1 December 2011.
6. David M. Eagleman. Why public dissemination of science matters: A manifesto. *Journal of Neuroscience*, 33(30):12147–12149, 2013.
7. Sönke Bartling and Sascha Friesike. *Towards Another Scientific Revolution*. Springer International Publishing, Cham, 2014.
8. Vladimir B. Teif. Science 3.0: Corrections to the science 2.0 paradigm. *CoRR*, abs/1301.2522, 2013.
9. Anselm Spoerri. What is popular on wikipedia and why? *First Monday*, 12(4), 2007.
10. Suthiporn Sajjapanroj, Curtis J. Bonk, Mimi Miyoung Lee, and Meng-Fen Grace Lin. A window on wikibookians: surveying their statuses, successes, satisfactions, and sociocultural experiences. *Journal of Interactive Online Learning*, 7(1), Spring 2008.
11. Shyong (Tony) K. Lam, Anuradha Uduwage, Zhenhua Dong, Shilad Sen, David R. Musicant, Loren Terveen, and John Riedl. Wp:clubhouse?: An exploration of wikipedia's gender imbalance. In *Proceedings of the 7th International Symposium on Wikis and Open Collaboration*, WikiSym '11, pages 1–10, New York, NY, USA, 2011. ACM.
12. Noam Cohen. Define gender gap? look up wikipedia's contributor list. *The New York Times*, January 2011. Accessed: 2016-11-19.
13. UNESCO Institute for Statistics. Women in science. `http://www.uis.unesco.org/ScienceTechnology/Pages/women-in-science-leaky-pipeline-data-viz.aspx`, 11 2015. (Accessed on 11/24/2016).
14. National Center for Education Statistics (NCES). Digest of education statistics-advance release of selected 2013 digest tables. `http://nces.ed.gov/programs/digest/2013menu_tables.asp`. (Accessed on 11/24/2016).
15. Vladimir B. Teif. Science 3.0: The future of science in the internet.

Social impact of Artificial Intelligence: Robotics

Kushagra Sharma, Manaswi Mishra, Siddharth Bhardwaj

Music Technology Group, Universitat Pompeu Fabra
kushagra.sharma01@estudiant.upf.edu, manaswi.mishra@upf.edu,
siddharth.bhardwaj01@estudiant.upf.edu

Abstract. We conduct our research on International Journal of Robotics Research 2015 with the aim of studying the social and moral implications of their research. We study 76 research papers to know how many researchers keep in mind the social impact of their work. We divide them in terms of their impact categories and state the percentage of papers that mentions any kind of social or ethical concerns. With this research, we show that even though robotics research can directly or indirectly influence so many aspects of society, majority of the researchers do not explicitly state the positive/negative impact that their research can have.

Keywords: Artificial Intelligence, Assistive Robots, Human Computer Interaction, Social

1 Introduction

The recent developments and ongoing research in science and technology of Artificial Intelligence (AI) coupled with the advent of Machine Learning has brought to us smart artifacts which will have profound ethical, psychological, environmental, and economic consequences on the human society. The field of science and technology has been ever evolving, but now we have been able to build machines which perceive, decide and act on their own. This is a radical shift from our earlier image of technology. Apart from making knowledge widely available, these machines can take unanticipated actions beyond our control. Here we try to skim the surface of some of these consequences.

One of the major topics in AI is the development of assistive robots and human-computer interfaces. Research and development in this field will bring together a wide variety of expert systems: “for aiding medical diagnosis and prescription, for helping scientists, lawyers, welfare advisers and providing people with information and suggestions for solving problems” [1]. Influential research and development has already started in areas like of Automation (the advent of self-driven vehicles) [2] , SLAM (Simultaneous Localization and Mapping) [3], Meta Cognition (provides a sense of self – analysis to the machine by which it can change its behaviour and adapt to the environment) [4] and Human Computer Interaction (focuses on interfaces between people and computers) [5] etc. The 3rd Generation Partnership Project (3GPP) is working on the standardization of the so-called Internet of Things and increase internet speeds and capacity for seamless exchange of data [6]. Therefore, we can easily foresee a future society where all these technologies and machines integrate and work in tandem with each other which is going to impact humans and how they interact with each other and machines, tremendously.

Robots are “an automatically controlled, reprogrammable, multipurpose manipulator programmable in three or more axes, which may be either fixed in place or mobile, for use in industrial automation applications” [7]. While these emerging technologies improve the speed, quality, efficiency and cost of goods and services, they dramatically increase the scope of replacing human labor since they reduce the need the need for human intervention in automated processes, leaving majority of population jobless. This trend will surely lead to “change the nature of the work across industries and occupations” in the future, but it is still debatable as “workers will reallocate to tasks that are non-susceptible to computerization” [8].

Some academics have questioned the use of robots for military combat, especially when such robots are given some degree of autonomous functions [13]. Researchers argue about the impact and consequences of these “smart” machines becoming more intelligent than humans and entering an era of singularity [9] or superintelligence [10]. Shane Legg [11] proved that a machine (such as a human brain) cannot predict and control a machine of greater algorithmic complexity. “It could be designed with motivations that initially appeared safe (and easy to program) to its designers, but that turn out to be best fulfilled (given sufficient power) by reallocating resources from sustaining human life to other projects” [12]. This raises the issue of safety and transparency which can be a threat to human dignity. Hence, it becomes very important for researchers to keep in mind this issue of “ethicality” and “morality” of the machines they propose. In similar terms, Issac Asimov presented the following three laws [14]:

- A robot may not harm a human being, or, through inaction, allow a human being to come to harm.
- A robot must obey the orders given to it by human beings except where such orders would conflict with the First Law.
- A robot must protect its own existence, as long as such protection does not conflict with the First or Second Laws.

The remaining part of the paper is organized as follows. In Section 2, we present the Research Methodology that we based to study the research papers of International Journal of Robotics Research, 2015. Section 3 describes our findings as the percentage of papers belonging to a category. Finally, we conclude the paper by enumerating the limitations of research and future work in this domain.

2 Research methodology

An empirical study was performed on the International Journal of Robotics Research, 2015. The journal was chosen for its high impact factor of 2.489 and ranking of 4 out of 25 in robotics as reported by the 2016 Release of Journal Citation Reports and 2015 Web of science data[15]. We analyzed the Abstract, Introductions and Conclusions of the 76 papers published in the 2015 edition of the journal and made a qualitative study of the social impacts communicated, categorizing them into the following aspects of society – 1) Improves Lifestyle, 2) Industrial Manufacturing advancement, 3) Medical advancement, 4) Disaster (search and rescue operations), 5) Weaponization and 6) No direct or indirect impact on society.

Table 1: Categories of social impact and their representation in the journal.

Social Impact Category	Percentage Representation (%)
Improves Lifestyle	55.26
Industrial Manufacturing	40.79
Medical advancement	21.05
Disaster (search and rescue ops)	9.21
Weaponization	6.57
No direct or indirect impact	5.26

The medical advancements include both therapeutic as well as surgical advancements. Robotics research was seen to have a wide range of impacts within its scope. Qualitative assessment of ethical impacts, positive and negative, were also analyzed.

3 Results

We say that none of the papers has discussed a direct social impact. From our own understanding, we have analyzed papers and their possible interdisciplinary applications and what they might lead to. We make the following observations from our empirical study.

1) In the field of robotics research though only 42.1% of the papers talk directly about social impacts, 94.7% of the topics discussed had an obvious impact on society broken down into the reported categories.

2) The top two categories of social impact made by robotics research is a) Improving Lifestyle (55.26%) and b) Industrial Manufacturing (40.79%). This seems obvious as the sources of funding come from the industry and the market for service robots is at an all time high. [16]

3) Robotics research also has a very high impact on the medical field, both through therapeutics and assisted surgical advancements. (21.05%)

4) Only 6.57% of the papers have any mention of the ethical concerns of their research despite the 55.26% of Lifestyle improving research that aim to be a part of regular household activities, The 6.5% papers with scope of weaponization must also include an ethical discussion.

None of the papers say anything about an ethical impact. We feel its important for machines working in tandem as a system to have transparency in it. Its also important for these machines to be "moral".

4 Conclusions and Future work

We started with the aim of studying the social impact of robotics research on society in general. We studied 76 papers from a very reputed journal(IJRR) for the year 2015 to get an idea whether the researchers in this area are taking into account the implications that their work can have on the society. We divided all these papers into five main, sometimes overlapping categories of social impact out of which majority are aimed at improving the lifestyle (55.26%) or industrial manufacturing advancement (40.79%). The other significant categories were found to be medical advancement, disaster management and weaponization. It was found that 94.7% of the papers had a direct/indirect impact on society but only 42.1% of the papers talk about this which should be improved. Only 6.57% of the papers have any mention of the ethical concerns of their research.

The limitations of our research are that we just studied the papers from one journal which can limit the type of research and their impact categories. If the research can be extended to more journals, it would give us a more concrete idea of the situation. Also, we have not studied the actual impact that this research has on the society, rather we have just predicted the categories that they can have a significant impact on. In future, we can study research papers from several years and take into account how they have actually influenced the society. Also, we defined the papers into broad categories which can be further subdivided into categories like SLAM, autonomous driving etc. and then the impact in each of these categories can be studied separately.

Finally, we conclude by saying that the places where the research is being conducted should teach the researchers about the ethical, psychological, environmental, and economic consequences of their work. There is a need for paradigms which need to make these machines somehow "moral".

References

1. R. Kurzeweil. The Age of intelligent Machines. 1990. Association of American Publishers
2. R. Savitha, S.Suresh, H.J. Kim, A Meta-Cognitive Learning Algorithm for an Extreme Learning Machine Classifier, Cognitive Computation, June 2014, Volume 6, Issue 2, p 253-263
3. T. Bailey, H. Durrant-Whyte, Simultaneous Localisation and Mapping (SLAM): Part II State of the Art
4. M.D. Schmill, and T. Oates, M. L. Anderson, D. Josyula , The Role of Metacognition in Robust AI Systems, Association for the Advancement of Artificial Intelligence, 2008.
5. F. Karray, M. Alemzadeh, J.A. Saleh and M.N. Arab, Human-Computer Interaction: Overview on State of the Art, International Journal On Smart Sensing and Intelligent Systems, March 2008, Vol. 1, Issue 1
6. “LTE”, http://www.3gpp.org/technologies/keywords-acronyms/98-lte, 3rd Generation Partnership Project, Dec 2013 as on 11/2016
7. ISO 8373:2012, Robots and Robotic Vehicles
8. C.B. Frey and M. A. Osborne, “The future of employment: How susceptible are jobs to computerization?”, Oxford University Programme on the Impacts of Future Technology, September 17, 2013
9. D.J. Chalmers, The Singularity: A Philosophical Analysis, Journal of Consciousness Studies 17:7-65, 2010
10. N. Bostrom, "How Long Before Superintelligence?" Linguistic and Philosophical Investigations, 2006, Vol. 5, No. 1, pp. 11-30
11. S. Legg, Is there an Elegant Universal Theory of Prediction? Technical Report No. IDSIA-12-06, 2006
12. Yudkowsky , The Power of Intelligence, http://www.yudkowsky.net/singularity/power/ , 2007
13. D. Axe, Robot three-way portends autonomous future, https://www.wired.com/2009/08/robot-three-way-portends-autonomous-future/ as on 11/2016
14. I. Asimov, (1950). I, Robot. Doubleday, Garden City, NY.
15. http://ijr.sagepub.com/ as on 11/2016
16. http://www.reportlinker.com/p04224692-summary/Data-Visualisation-report-on-Global-Service-Robotics-Market.html as on 11/2016

Is Interdisciplinary research the way to go?

Adrian Arroyo Perez[1], Giorgos Neokleous[2] Pantelis Stylianides[3], and Vibhor Bajpai[4]

[1] Masters in Intelligent interactive systems
adrian.arroyo01@estudiant.upf.edu
[2] Masters in Intelligent interactive systems
adrian.arroyo01@estudiant.upf.edu
[3] Masters in Intelligent interactive systems
adrian.arroyo01@estudiant.upf.edu
[4] Masters in Sound and Music Computing
vibhor.bajpai01@estudiant.upf.edu

Abstract. In today's world the problems in the research domain do not limit themselves to a single discipline, but encompass multiple disciplines at once. This is where Interdisciplinary research comes into the picture. Interdisciplinary research is an amalgamation of research in two or more disciplines. It is a mode of research where solution to a particular problem are beyond the scope of a single discipline. Thus, it is important to study the pros and cons of this type of research so as to understand its impact and usefulness in today's research intensive environment.

Keywords: Interdisciplinary

1 Introduction

Interdisciplinarity is currently everywhere and it is used in many ways in research and at all levels of education. It is becoming increasingly popular day by day as by combining more than one research domain we can get better and more trustful results. In addition to that, many researchers and journal editors seem to agree that the key to future scientific research lies in interdisciplinarity research.

Interdisciplinarity is basically based on the education of the students because it helps them expand their understanding and achievement between all disciplines. Moreover, people learn more about working in groups and about leadership with their primary goal to synthesize disciplines in order to achieve a common target.

Some studies have been made on groups of students using Interdisciplinarity and it has been found out that the majority of students found the experience beneficial and that the students "spoke of long-term relationships and of a democratic learning environment that honored their voices and empowered them as learners.[1] The way they connect their ideas between different disciplines its the best way to deepen the learning experience since they are considering many and varied perspectives on how this ideas can be explored. Furthermore you need critical thinking skills when you are looking across different disciplines to consider all the options, compare and contrast concepts across subject areas; either you have them or you develop them.

Despite all its benefits it is very difficult to use because it is very hard to find people expert in two or more domains at the same level. You have to find people than can co-operate well with different people of different fields of studies which is at least as hard.That is why, as we have mentioned before, now it is used on the education system to develop it from an early age. In addition to that, sometimes it may get really time consuming if the people cannot co-operate well and if some people think their discipline is above others. As Rick Szostak, author of How and Why to Teach Interdisciplinary Research Practice[2] explains, the methodology of the practice of interdisciplinary is lost when a single interdisciplinary course is then considered as a major field of study. Furthermore some professors are getting focused on interdisciplinary studies and they isolate themselves from the core of their field[3]. Some more problems that are appearing are that by working in groups you have to come against situations like "lack of 'sufficient time for collaboration work'", "lack of training In group dynamics", "overlapping roles", "territorial and status conflicts", and "inadequate funding"[4]. Finally if you manage to develop a good research with interdisciplinarity the cost of it is actually much higher than typical studies and it is also harder to publish it.

To conclude, it may be hard to do a good research with interdisciplinarity but it is totally worth it at the end. For example, the chemist Willard Libby who discovered radiocarbon dating, applied his findings in Chemistry to the discipline of Archeology and won the Nobel Prize the discovery in 1960[5].

2 Research Methodology

We have chosen the Empirical methodology for the development of our investigation in interdisciplinarity research. This methodology stands for: the research based on experimentations or observations of evidences, whose final objective is to test a hypothesis in order to validate it. Its central theme is that every evidence must be empirical, which means that it has been gained by experimentation or observation.

Reasons to justify the selection of this methodology seemed clear, since we want to validate our hypothesis (whether research methodology is the way to go or not) but there is no theory or model that describes it. We are just based on some collected evidences and we want to test our hypothesis through the statistical data recorded during the investigation.

Apart from the previous reasons to pick this methodology, we have discarded the rest of the methods first. The reason of not taking the scientific method was that we are not going to test variations of our hypothesis, we are just going to validate it. As for the engineering method, we did not choose it since we are not going to improve our results after this research, as the objective is just to validate our question. Lastly, we did not pick the analytical method because we are not going to compare our results with other observations since our result is indeed based on the observations.

Methods used: historical data (look for patterns in historical data - quantitative method), action research (apply a research idea, evaluate results, modify idea) As for the strengths of this methodology we would like to outline the simplicity of it, as there is no model to study and we just need to focus on getting the correct observations for our question validation. The limitations are mostly coming from the fact that our hypothesis is based on observations; this means that we cannot come up with some idea or evidence directly, as we need some previous experience to develop it. In addition, it is not possible to test different variations of the hypothesis since we are just limited to the one we are given. Another important point is that there is no formal model that describes the hypothesis and we somehow have to come up with some theory in which we will put our hypothesis after it is validated.

The empirical cycle consists of 5 phases, described in the following list from our researchs point of view:

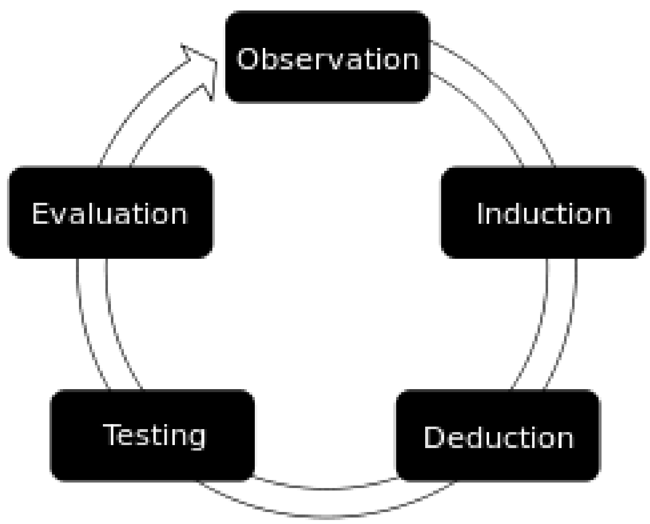

Fig. 1. This shows the cycle for Empirical research[6]

- Observation: Researching online on some interdisciplinary papers and projects.
- Induction: Process of assuming that our question is true while we are investigating about the truth of it. Similar process for the falseness of it.
- Deduction: Once we have a set of premises (observations: interdisciplinarity papers and projects) deducing by their content if our question is true or not.

– Testing: Testing the deduction previously obtained with new data.
– Evaluation: Validating our hypothesis.

3 Results

Interdisciplinary work is considered crucial by scientists, policymakers and funders but how widespread is it really, and what impact does it have? Scholars say that the concept is complex to define and measure, but efforts to map papers by the disciplines of the journals they appear in and by their citation patterns are tentatively revealing the growth and influence of interdisciplinary research.[7]

Since the mid-1980s, research papers have increasingly cited work outside their own disciplines. The analysis shown here used journal names to assign more than 35 million papers in the Web of Science to 14 major conventional disciplines (such as biology or physics) and 143 specialities. The fraction of paper references that point to work in other disciplines is increasing in both the natural and the social sciences. The fraction that points to another speciality in the same discipline (for example, a genetics paper pointing to zoology) shows a slight decline.[7]

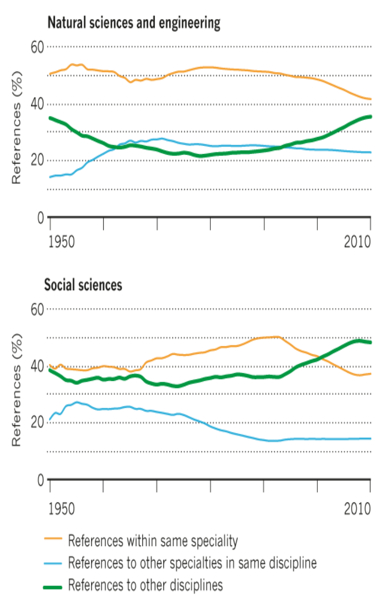

Fig. 2. Trends in Natural sciences and Social sciences[7]

Discourse about interdisciplinary research is increasing. The fraction of papers that mention interdisciplinarity in their title has fluctuated, perhaps reflecting the priorities of funders, but the twenty-first century saw that proportion reach an all-time high.[7]

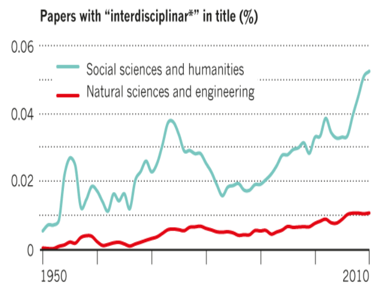

Fig. 3. Papers with Interdisciplinary in their title[7]

Whether interdisciplinary research gains more citations than disciplinary research is contentious. Over three years, papers with diverse references tend to pick up fewer citations than the norm, but over 13 years they gain more. Some studies suggest that a little interdisciplinarity is better than a lot: papers that combine very disparate fields tend to earn fewer citations. But interdisciplinary work can have broad societal and economic impacts that are not captured by citations.[7]

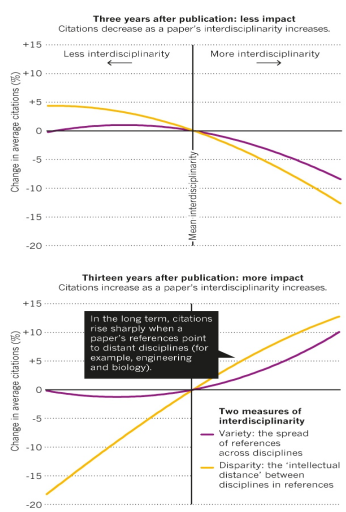

Fig. 4. Papers with Interdisciplinary in their title[7]

4 Conclusions

Looking at the graphs and the trends in the past many years, it can be said that the effect of interdisciplinary research is not evident in the first few years of the research, but it takes time to have an impact on the related fields. A course of 3 years is too short for an interdisciplinary research to show significant influence in the research but a period of 13 years shows much successful results.

We can safely conclude that interdisciplinarity in research is a far sighted approach rather than a myopic one, where we cannot expect to see the results and influence of the research as soon as it is published, but it takes time to have an impact.

References

1. Boyer, Bishop, 2004. "Young Adolescent Voices: Students' Perceptions of Interdisciplinary Teaming" RMLE, v.1. http://eric.ed.gov/data/ericdocs2sql/content_storage_01/0000019b/80/3e/a6/
2. Szostak, Rick, 2007. "How and Why to Teach Interdisciplinary Research Practice." Journal of Research Practice, v.3, i.2. http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/3e/26/b9.pdf.
3. Kanakia, Rahul, 2007. "Talks touts benefits of interdisciplinary approach, as well as some of its pitfalls." Stanford Reprot. http://news-service.stanford.edu/news/2007/february7/barr-020707.html
4. Haynes, Carolyn, 2002. Innovations in Interdisciplinary Teaching, West port, CT, American Council on Education ORYX Press.
5. Youngblood, Dawn: "Interdisciplinary Studies and the Bridging Disciplines: A Matter of Process." Journal of Research Practice, v.3, i.2. (2007)
6. Empirical Research https://explorable.com/empirical-research
7. Interdisciplinary research by the numbers http://www.nature.com/news/interdisciplinary-research-by-the-numbers-1.18349

Analysis of AcousticBrainz, Essentia and Freesound platforms[1] from an Open Science perspective.

P. Alonso-Jiménez, G. Erruz, and G. Ruiz-Marcos

Universitat Pompeu Fabra (UPF), Barcelona, Spain
{pablo.alonso02, gerard.erruz01, german.ruiz01} @estudiant.upf.edu

Abstract. Open Science is a well-known movement to make scientific research nowadays. So that, we decided to focus on an useful study of the state-of-the-art projects developed at the MTG in the recent years: AcusticBrainz, Essentia and Freesound, in order to analyse how Open Science is conceived in this platforms. On the other hand, harnessing our own results from the analysed data we could relflect about ways of solving whatever issue that might help these platforms getting better. To approach this idea, we took the Open Data analysis as the starting point of our research, basing our features selection in previous research from other domains. That left us 18 parameters of study with which we were able to analyse the projects, from an Open Data perspective, concluding with some improvements to this platforms, what could highlight even more the impact and value of this piece of research.

Keywords: Open Science, Open Data, MTG, UPF.

1 Introduction

1.1 Open Science

"Open Science represents a new approach to the scientific process based on cooperative work and new ways of diffusing knowledge by using digital technologies and new collaborative tools"[1]. Likewise, the OECD defines Open Science as: "to make the primary outputs of publicly funded research results – publications and the research data – publicly accessible in digital format with no or minimal restriction" [2]. On the other hand, taking into account the amount of different definitions for Open Science available online or in books and papers, we understand Open Science as a collective concept that encompasses some specific topics related to the sharing of knowledge, resources, data, methods or tools involved on research, topics such as: open data, open source, open access, open peer review or open access, among others [3]. The aims of Open Science

[1] Three user-contributed platforms developed in the Music Technology Group of Universitat Pompeu Fabra in Barcelona.

are very diverse, but they could be summarized into: transparency, in the experimental methods and the collection of data; accessibility, in available data collection for future studies and reuse; collaboration, in scientific study and in the research process by using web-based tools [4]. As we have just said, scientific community have more than one point of view on Open Science. In fact, some authors have proposed different classifications. In his book Opening Science [5], Fecher and Friesike propose a classification based on five Open Science schools of thought: infrastructure school (which is concerned with technological architecture), public school (which is concerned with the accessibility of knowledge creation), measurement school (which is concerned with alternative impact measurement), democratic school (which is concerned with access to knowledge) and pragmatic school (which is concerned with collaborative research). Following Fecher and Friesike proposal, the scope of this project may be located under the influence of the Democratic school. Authors compromised with this interpretation are mainly concerned with the importance of the access to knowledge. Likewise, Cribb and Tjempaka [6] among others have mostly concerned about the accessibility of the research results, but there are other kind of data and user-contribution platforms comprised in the Democratic school field. On [7] J. Visions presents the idea of scientists not publishing their data sets excusing on the limitations of the printed page era. However, times have changed and the implantation of a model of fully public datasets should be addressed for both the scientific research and journal communities. In this paper, we will focus on the concept of Open Data.

1.2 Open Data

'Open Data' is a broad movement that applies for many different domains, not just research. The main idea behind Open Data is the encouragement of the aggregation of data from papers and other collections, opening it for reuse, allowing research synergies and preventing the duplication of data [5]. According to [8, 9, 10], for a correct use of Open Data you should:

- Start small, simple and fast, including your information step-by-step in an organized way.
- Engage yourself with users, developers and researchers.
- Choose the dataset you plan to open/use.
- Determine which intellectual property rights will match with your data.
- Apply an open license.
- Choose an available way of sharing your data in a useful format.
- Make sure there is a clear assertion of responsibility in the dataset.
- Be sure about the existence of a contingency plan in case the dataset ceases to operate.
- Work on the control of the quality measures of the metadata.
- Make it easy to find and work with your data.

Working according to Open Data principles will report you a big amount of benefits [11] such as transparency, so it could be known what, where and how the state-of-the-art is working on; efficiency, because access to scientific methods, tools and results can improve future research; quality and integrity, given by the evaluation and validation of a wider scientific community; economic benefits, based on the reuse of data, methods and tools, as well as knowing a priori the efficiency of previous information and techniques; public engagement, making science open and available for whole society.

1.3 Research Question

On this framework we have noticed the convenience of discussing and proposing a method to assess and scale the level fulfillment of the Democratic School of Open Science criteria. The goal of this method is not just to rank the openness of diverse institutions or research groups but also to conclude in some recommendations in order to improve its accessibility . As far as we have not found any solid previous research on this context, this paper comes up with an evaluation proposal merging the ideas of the Democratic School with some existing research outside the context of Open Science. Concretaly Medina Sánchez presents in [12] some criteria for the analysis of open data projects in the background of political science.

1.4 The user-contributed platforms

In order to define some existing platforms to test our evaluation method , we have focused on three projects from the Music Technology Group of Universitat Pompeu Fabra in Barcelona. The chosen projects are: Freesound (based on a sound repository)[13], Essentia (based on a code repository)[14] and AcousticBrainz (based on a feature repository)[15], three user-contributed platforms related to music information retrieval and knowledge sharing. These are projects of collaborative nature from their initial conception. The originality of this article lies on the broad and definite analysis of these kind of platforms as well as the methodology used for this purpose. Also, it will be fantastic to see other researchers applying similar approaches to other platforms, helping to grow community awareness on this topic.

2 Research Methodology

2.1 Methodology

This study adopts a qualitative approach. In order to assess the accessibility of different services developed in the Music Technology Group, several descriptive features are gathered and compared.

With all these in mind, we have decided to answer our research question using an Experimental method basing our study in an Observational model, focussing in a Project monitoring.

2.2 Feature selection

The process of feature selection involved the study of different backgrounds, like the open government data, whose applications in relation with open science have barely been assessed. A series of features are porposed on [12] in order to evaluate the user-friendliness of different web sites. So that, we will be working with Qualitative data collection (from observation and log files). These features have been adapted for the scope of this project as follows:

Common features

Access

(common1) Registration: is it necessary to register in order to access the portal or to consult the information about it?

Contact

(common2) Contact form: is it possible to identify easily and clearly some kind of contact through forms, e-mail, tickets system, among others?

(common3) Response time: how do we evaluate the response time obtained after asking the question? Our evaluation: 1 day: excellent; 2-3 days: optimal; 4-7 days: average, and more than 7 days: unfavorable.

Catalog

(common4) Number of datasets: how many datasets could we find?

(common5) Types of search: do we find any type of search?. What type of search can we use: simple, advanced, among others?

(common6) Formats available: in how many formats can we find the datasets available in the portals?

(common7) Content of the datasets (within each set of data)

Specialized users

Legal framework

(special1) Terms of use: do we find a section on the conditions of use and the type of license for the use of datasets?

(special2) Terms of use / datasets: is each dataset contained in the portal identified with the license type?

Extra information

(special3) Documentation API: do we find the API documentation of the different data sets that we find in the portals to help generate the applications?

(special4) What is new in datasets: could we notice what is new in datasets when we enter the platforms?

General users

Participation

(general1) Profile in social networks of the portals: are the profiles identified in social networks and are they updated?

(general2) Sharing datasets: can you share datasets or information related to the main project across major social networks such as Facebook and Twitter?

(general3) Discussion system: do we find discussion systems in the portal about the contents?

Value added information

(general4) Explanation of key concepts: is a section identified with the explanation of key concepts related to open data?

(general5) Objectives of the portal: do we find the objectives of the portal well-defined?

(general6) Multi-language portal: can we find the information in at least two languages?

Level of reuse

(general7) Which is the “stardata value” based on the Linked Data model created by Tim Berners-Lee [16]?

3 Results

This section presents the obtained results after applying the methodology to each platform. Table 1 shows what aspects are covered in each one and the possible improvements. It also serves as a comparison between them. Finally, we present some analysis on the results for each particular platform and some ideas to improve each of the projects based on our analysis:

AcousticBrainz presents polarised results. As we can see in Table 1, the main points of the common features are not accomplished. This is easy noticed when entering the website: there is no search engine for browsing the datasets nor contact form for general issues. For getting the information of certain music file, the user has to enter its dataset and select it directly. A better browsing system is advisable to improve accessibility, specially for common users. Also, the methods for contributing to the project require high computational skills. A user-friendly web interface for introducing missing data of new datasets would be advisable. In terms of the actual data, there is no explanation of the classification concepts (e.g. Timbre, Electronic classification, Tzanetakis model, ...). Some improvements could be done in this area, like a basic conceptual explanation page in the FAQ section, for example. And an introductory section to Open Data basic terminology. The site is clearly aimed to advanced users and excels in the license and automatic accessibility aspects. All the datasets are available in JSON format and the systems supports API integration. All the datasets are published using the CC0 public license. In general, AcousticBrainz platform lacks the intention to be accessible to the general public.

Essentia is an open source library published by the MTG under Affero GPLv3 [17] license. The fact that universities understand the convenience of publish research results is already a positive result in terms of Open Science. However, after the evaluation under the proposed criteria it has been shown that there are some possible improvements for the platform. First, the search engine is not very accessible on the website and only allows to make general search on the documentation of the algorithms of the Library, which could be improved with examples and links to the source code. It would be convenient to extend the search capabilities to the whole site and provide advanced search features. Although it is not very common in this kind of sites, the website could implement multilingual functionalities, at least on the main page. In order to extent the diffusion capability of the platform it should open profiles in the main social networks.

In regards to Freesound, according to the results obtained, we could conclude that is a really good Open Science project where almost all the study parameters were positive. Likewise, it is also important to highlight that Freesound.org has a valid and up-to-date SSL certificate that expires on January 4, 2018. On the other hand, some changes could be proposed in order to improve the project: give the option of finding information in different languages, improve the software in order to get a five-stars-value and, finally, minify and compress HTML content.

4 Conclusions

Data is a classic example of a public good, in that shared data do not diminish in value. To the contrary, shared data can serve as a benchmark that allows others to study and refine methods of analysis, and once collected, they can be creatively repurposed by many hands and in many ways, indefinitely. For this reason, many voices in recent years have advocated for the removal of barriers to the availability and reusability of scientific data.

So that, in this specific paper, we decided to analyse the state-of-the-art projects developed at the MTG from a qualitative point of view, basing our analysis on the features of each of the platforms according to an Open Government type of study adapted for our concrete domain of research. Thus, we obtained a general idea of those project in that Open Data regards, as well as we proposed some improvements for each platform.

- AcousticBrainz project was conceived as holding public license datasets from the beginning, but basically lacks on common user accessibility. Improvements on browsing engines and web interface solutions for collecting missing data are strongly advisable, among others.
- Essentia is a top-ranked open source library well known on its field. However, in terms of open data, there are fields that could be improved as a more detailed documentation, a better search engine and some presence on the main social networks.
- Freesound aims to create a huge collaborative dataset of audio released under Creative Commons licenses that provides the opportunity of reusing data. As it can be seen in the previous analysis, Freesound evinces a great Open Science format and only a few improvements could be proposed.

Finally we would like to state some points about the results of this work. It is easy to notice that most of the proposed improvements would be relatively easy to implement. Furthermore this improvements could carry benefits for both, the MTG, by the improvement on the relevance of its platforms, and researchers around the world, by having a more accessible way to work with them.

The authors of this article would be grateful to see a growing awareness on this issue. Following the pattern presented in this article, similar analysis can be carried out in other scientific/technological platforms. The Open Data community has to be active and fully working for more accessible research projects, both for common users and advanced ones. We claim for public initiatives to lead this change and also making efforts to extend this practices to the industry and private projects. A collaborative and inclusive scientific community would take advantage of this synergy in a better research paradigm.

Bibliography

[1] Directorate-General for Research and Innovation. *Open innovation, open science, open to the world*. European Union, 2016.

[2] Industry Policy Papers. Making Open Science a Reality OECD MAKING OPEN SCIENCE A REALITY. (25):1–108, 2015.

[3] Open Science at the Core of Libraries: What is Open Science? Introduction. Technical report. https://www.fosteropenscience.eu/content/what-open-science-introduction.

[4] The OpenScience Project: What, exactly, is Open Science? Technical report. http://www.openscience.org/blog/?p=269.

[5] B. Fecher and S. Friesike. Open Science. One term, five schools of thought. In S. Bartling and S Friesike, editors, *Opening Science. The Evolving Guide on How the Web is Changing Research, Collaboration and Scholarly Publishing*, pages 213–224. 2014.

[6] J. Cribb and S. Tjempaka. *Open Science: sharing knowledge in the global century*. Csiro Publisher, 2010.

[7] Todd J. Vision. Open Data and the Social Contract of Scientific Publishing. *BioScience*, 60(5):330–331, 2010.

[8] HOW TO OPEN DATA: A QUICK PRIMER ON HOW DATA HOLDERS CAN OPEN UP THEIR DATA. HOW TO START SMALL BUT THINK BIG. Technical report. https://okfn.org/opendata/how-to-open-data/.

[9] Panton Principles: Principles for Open Data in Science. Technical report. http://pantonprinciples.org/.

[10] How to Publish: Data Guidelines. Technical report. https://f1000research.com/for-authors/data-guidelines.

[11] Open Science at the Core of Libraries: What are the benefits of Open Science? Technical report. https://www.fosteropenscience.eu/node/1422.

[12] Vanessa Medina. Análisis de proyectos 'open data' en españa: propuestas de mejora. pages 53–72, 2015.

[13] Frederic Font, Gerard Roma, and Xavier Serra. Freesound Technical Demo. In *ACM International Conference on Multimedia (MM{\textquoteright}13)*, pages 411–412, Barcelona, Spain, 2013. ACM, ACM.

[14] D Bogdanov, Nicolas Wack, Emilia Gómez, Sankalp Gulati, P Herrera, O Mayor, G Roma, J Salamon, J Zapata, and Xavier Serra. ESSENTIA: An audio analysis library for music information retrieval. *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 493–498, 2013.

[15] Alastair Porter, Dmitry Bogdanov, Robert Kaye, Roman Tsukanov, and Xavier Serra. AcousticBrainz: a community platform for gathering music information obtained from audio. In *16th International Society for Music Information Retrieval Conference (ISMIR 2015)*, pages 786–792, Malaga, Spain, 2015.

[16] 5 stars DATOS ABIERTOS. Technical report. http://5stardata.info/es/.

[17] SPSS Inc. GPL reference guide. *Psychology and aging*, 21(2):141, 2005.

5 Annex

Table 1. Results of our analysis based on the questions included in the methodology.

Reference	AcousticBrainz	Essentia	Freesound
common1	No	No	Yes
	N/A	Not necessary in order to download and use the library	Necessary for posting and downloading
common2	No	Yes	Yes
	N/A	Email, forum and ticket system based on GitHub	Forum, help and blog tags in the main page
common3	N/A	1-2 days	1-3 days
common4	52	1	24
	N/A	N/A	Collected in 5 categories
common5	No	Yes	Yes
	N/A	Simple search	By file names, tags, users names, similar sounds, geotags, packs, sound Ids and descriptions
common6	2, web summary and JSON	1	4-5, depending on the category
common7	Acoustic, rhythm, tonal, moods, genres, vocals and music type information	Source code repository	User that posts the sound, date of posting, description, tags, comments and sound characteristics: type, duration, filesize, samplerate, bitdepth and channels
special1	Yes	Yes	Yes
	CC0 license (public domain)	General use of the website and source code terms of use	General use of the website, Intellectual Property Rights, User Account, Interactive services and User content, IPR in the Interactive Services And User Content, Liabilities and Privacy policy
special2	Yes	Yes	Yes
	Not specifically but the public license is general	Affero GPLv3	
special3	Yes	N/A	Yes
special4	Yes	Yes	Yes
general1	Yes	No	Yes
	Twitter		In 4 platforms
general2	No	No	Yes
general3	Yes	Yes	Yes
	Blog and Bug Tracker	Technical forum based on GitHub	Comments, forum and blog
general4	No	Yes	Yes
general5	Yes	Yes	Yes
general6	No	No	No
general7	5 stars-value	3 stars-value	3 stars-value