European Commission

Horizon 2020
European Union funding
for Research & Innovation

Case studies on Open Science in the context of ERC projects – Set 3

April 2018

This document presents the third of five sets of case studies that have been produced in the framework of the *'Study on open access to publications and research data management and sharing within ERC projects'.*

European Research Council
Established by the European Commission

Contract number:	ERCEA/A1/2016/06
Contracting authority:	European Research Council Executive Agency (ERCEA)
Contractor:	PPMI Group UAB with University of Edinburgh and Georg-August-Universität Göttingen
Authors:	Angus Whyte, Digital Curation Centre, University of Edinburgh Viltė Banelytė, Public Policy and Management Institute
Contact:	Dagmar Meyer, ERCEA
Contact email:	erc-open-access@ec.europa.eu
DOI:	https://doi.org/10.5281/zenodo.1235371

LEGAL NOTICE

These case studies have been prepared for the European Research Council Executive Agency (ERCEA). However, the views set out in this document are those of the authors and/or the interviewees only and do not necessarily reflect the official opinion of the ERCEA. The ERCEA does not guarantee the accuracy of the data included in this document. Neither the European Commission nor the ERCEA nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein.

TABLE OF CONTENTS

SUMMARIES

Our Mythical Childhood... The Reception of Classical Antiquity in Children's and Young Adults' Culture in Response to Regional and Global Challenges (OurMythicalChildhood)

Professor Katarzyna Marciniak has assembled an international team of researchers to study the intersection of classical antiquity and children's and young adults' culture in several regions across the globe. OurMythicalChildhood is developing an openly accessible supra-regional survey of classical references and engages in a myriad of science communication activities both online and face-to-face. The team is united by the same vision – to share their research results as soon as possible in order to facilitate their dissemination, immediate uptake and further use.

The Intergenerational Cumulative Disadvantage and Resource Compensation (INDIRECT)

Professor Jani Erola and his team have come to appreciate the benefits of open science practices while working on the ERC project INDIRECT. Professor Erola was not highly aware of open access issues before receiving his Consolidator Grant in 2013. Currently his team aims to ensure open access to articles and books as much as possible and have also set up an online platform to disseminate their working papers while they are undergoing peer review. These approaches provide opportunities for the team to generate greater interest in their work and achieve a wider impact through their published works.

Computational models for the discovery of the world's music (CompMusic)

In the CompMusic project, Professor Xavier Serra's Music Technology Group used an ERC Advanced Grant to develop their open science practices in a manner that helped answer their research questions, and enhance the opportunities for commercialisation of the results. Their research is interdisciplinary, developing computational approaches to support the scholarly analysis of music. CompMusic was particularly focused on five non-Western musical traditions and developed a software framework for data analysis. Lacking research data repositories and standards specialised for this relatively new domain, the Group constructed an approach that combines locally engineered open source software frameworks and open data collections. These are managed and shared across a variety of community databases. Through user-provided content and usage the open collections help with software enhancements. Using a dual-licensing model, the software can then be commercialised and generate income streams from subscribers, while also being made available under open source licenses for non-commercial use.

1. Our Mythical Childhood... The Reception of Classical Antiquity in Children's and Young Adults' Culture in Response to Regional and Global Challenges (OurMythicalChildhood)

Summary

Professor Katarzyna Marciniak has assembled an international team of researchers to study the intersection of classical antiquity and children's and young adults' culture in several regions across the globe. OurMythicalChildhood is developing an openly accessible supra-regional survey of classical references and engages in a myriad of science communication activities both online and face-to-face. The team is united by the same vision – to share their research results as soon as possible in order to facilitate their dissemination, immediate uptake and further use.

1.1. Introduction

The project *Our Mythical Childhood... The Reception of Classical Antiquity in Children's and Young Adults' Culture in Response to Regional and Global Challenges* (OurMythicalChildhood) that started in 2016 links two fields that until recently have not been studied together: classical antiquity and children's and young adults' culture. The Principle Investigator (PI) Professor Katarzyna Marciniak and her colleagues study these themes and how they are manifested in contemporary societies both in the regional and in the global context. The project is unique due to the fact that it involves a selection of partners from several countries. University of Warsaw is the hosting institution and, even though this is unusual for a Consolidator Grant, there are four more beneficiaries: University of New England (Australia), University of Yaoundé I (Cameroon), Bar-Ilan University (Israel) and University of Roehampton (UK).

- *Why has such an international and collaborative approach been chosen for the OurMythicalChildhood project?*

A collaborative approach and involvement of these countries were important for the project as classical antiquity is a particular experience in children's and young adults' culture. Everyone has gone through childhood and many people have had contacts with classical antiquity as a cultural experience. It is transmitted all over the globe and across the ages via education, international exchange of experiences, and currently through social media and the global popular culture. Hence, the ancient tradition has built a familiar code of communication understandable in local and global contexts alike. Many people around the world have awareness of classical antiquity and, hence, it can be a precious tool enabling researchers to interpret contemporary culture and contemporary transformations from quite an unusual perspective.

Involving researchers from Australia or Cameroon in a project dealing with classical antiquity could seem unexpected at first. Professor Marciniak explains why this is a misconception: "For example, in Cameroon classical antiquity can be seen as a part of the colonisation burden. It was part of the culture imposed on the local people. However, today they can finally reinterpret this culture freely, drawing also on their African heritage. The results are amazing. If you follow the reception of classical antiquity in Africa, you can see many theatre pieces or novels inspired by African culture and classical antiquity. It turns out that the ancient heritage is a common code of communication – now even more valuable

than ever, for it enables us all to engage in a broad international dialogue and it makes us more sensitive to various cultural backgrounds."

In addition, the project team wanted to combine rather unusual themes and activities in their research, such as studying ancient myths and autism or creating animations from scenes on ancient Greek vases[1]. Such an approach requires a holistic and collaborative effort as it involves reconstructing the culture, clothing and other artefacts in order to create the scenarios for the animation. "We also collaborate with an eminent scholar who can reconstruct the ancient Greek music, as much as it can be reconstructed, and use it for the animations", says Professor Marciniak.

1.2. Successful open science practices used in the project

- *What are the main achievements of the project so far?*

Just recently the team opened the *Our Mythical Childhood Survey*[2], an online platform hosting entries on myths, books, films, computer games and other sources that contain some elements from classical antiquity. These entries are collected or prepared by researchers and collaborators in the project. Each of them goes through a double peer review process; at first within the institutional team of the author of the entry and then by another partner institution of the project. This process helps the team to ensure the best scholarly and linguistic quality of the entries. "In gathering the works for young people inspired by classical antiquity, we hope to inspire other scholars, stimulate research and cultural exchange, popularize ancient culture and both global and local contemporary cultures, and – last but not least – encourage the translators, editors, educators, and readers to make use of the material in their daily work", says Professor Marciniak. Although the survey is still being developed and fine-tuned, the team agreed to open it to the public already at this stage. By opening up the survey to the public already now they aim to share the research results immediately. In addition, the research team reached an agreement with the University of Warsaw to host the survey on their website for 10 years after the end of the project.

Another important achievement of the project that has been running for less than two years is a very active engagement in citizen science activities and broad collaboration with scholars as well as the general public. The dissemination aspect is evident in the use of social media and blogs. The project has a website[3], several social media profiles[4] and four thematic blogs[5]. The blogs are particularly attractive to young people as many entries deal with contemporary culture: "The entries cover various topics from an Australian TV series about teenage sirens, the Harry Potter exhibition, a Polish book featuring Hercules the Hare, to the newest Netflix movies with classical motifs that inspire deep reflection on our identity, dreams, challenges, behaviours, and choices in the 21st century", says

---

[1] https://www.youtube.com/watch?v=diLi8CaQaRU
[2] http://www.omc.obta.al.uw.edu.pl/myth-survey
[3] Project website (in progress): http://www.omc.obta.al.uw.edu.pl
[4] Facebook: http://www.facebook.com/OurMythicalChildhood/
YouTube: http://www.youtube.com/channel/UC6zvu9EXsI0gK5rSvgnQseQ
Twitter: http://www.twitter.com/OMChildhood
Instagram: http://www.instagram.com/omchildhood/
[5] Our Mythical Childhood Blog: https://ourmythicalchildhoodblog.wordpress.com/
Antipodean Odyssey Blog: https://antipodeanodyssey.wordpress.com/
Mythology and Autism Blog: http://myth-autism.blogspot.com/
Panoply Vase Animation Project Blog: http://panoplyclassicsandanimation.blogspot.co.uk/

Marciniak. They recently reached 400 subscribers to the 'youngest' of the four blogs which, according to the PI, indicates the interest of the public in the project and its results.

The researchers also actively engage in face-to-face dissemination activities. Examples of such activities include visiting children in schools and kindergartens and presenting the project or the topic of ancient mythology, hosting events at their institutions, participating in conferences and external events and inviting the public to the project's conferences. The team plans to display the previously mentioned Greek vase-inspired animations in the National Museum of Warsaw. According to the PI of the project, such means not only promote research but can also be used by parents, educators and teachers for discussions on the topic of classical antiquity, historic artefacts and concepts of heroism, ancient values and beauty.

Professor Marciniak provides examples of recent activities that her team organised or participated in to promote the project and their research topic: "Next week I am going to another town in Poland to meet with children who wish to study classical antiquity, and to present the project and our data. PhD students of the project also actively engage with the public and present our work. We prepared a series of events at the different partner institutions of the project for the occasion of the 10 years of the ERC. We showcased our research and our results, but we also invited high school students and prepared a session with games for them inspired by classical antiquity. We have engaged with schools in various ways. Students of an art school prepared art works which were later exhibited during a conference of our project. Currently we have announced a video competition in which young people are asked to produce short movies on classical mythology in our times. We have composed a jury that includes people with cinematographic skills and a director and they will select the best movie. Our efforts are dedicated to awaken the curiosity in young people, which is not difficult to do, if one tries".

The project has also already produced a journal publication that was published in open access. The team members plan to have all articles and books resulting from the project published in open access via the gold route, i.e. immediate open access. The PI admits that such an approach requires many resources as covering publication costs, especially for books, can be very expensive. But the team are aware of it and are planning their budget carefully.

1.3. Challenges faced and success achieved

- *What difficulties do you face while working in a collaborative way, putting emphasis on sharing results and engaging with the public?*

According to Professor Marciniak, the initial challenge was the preparation of a common template that can be used for the collection of entries to the survey. In particular, the standardisation of the template was complicated, as it had to capture not only the myths from the storytelling but also books, computer games, novels, films and TV series. The team came to a consensus to have eight templates for different entries. In addition, working with an international team based in different continents further contributed to the complexity. "As people come from very different countries, cultures and scientific backgrounds, it took some time to align everyone's ideas. But exactly this aspect is our biggest strength, for we inspire and complement each other within our expertise", says Marciniak.

The team also devoted efforts towards making the survey visually appealing and user friendly. Professor Marciniak notes: "More often than not, the usual repositories lack this component, are not very user friendly or are intended more for the specialists or a specific group of scientists. Our online platform is dedicated to everyone, including the general public; hence the user friendliness aspect was crucial to us. The characteristics and data should be accessible to researchers, and on the other hand, there should be elements that would make the data very attractive and easy to understand for the

general public". To achieve that, the team collaborated with an IT-specialist and a graphic designer who made the design of the survey attractive and easy-to-use.

Another challenge for the research team was the question of when the survey and, hence, the data should be published openly. The project is in its second year, the data are still being collected, analysed and compiled. The survey itself is not developed to its full extent and some elements, such as the search engine, are not yet functioning. The project team debated whether they should wait for the finalisation of the data collection and full-scale functioning of the platform, or whether they should go ahead and open it up already. The consensus was to open the platform, even though it still needs more inputs. This approach gives an opportunity to open and share the project data and results collected so far immediately with the public, which is very much at the core of the project and its vision. All interested parties – scholars, students, teachers, editors, cultural managers and translators – can access the information and use it for their work or information purposes with no delay. At the same time, the team are working to further develop the platform with all its functionalities.

Professor Marciniak admits that the overarching challenge in the project is time. Obtaining consent and receiving permission to use the material collected usually demands a lot of the researchers' time. Also, the activities they participate in and organise are time-consuming. On the other hand, their commitment is strong, and the team make their best efforts to engage with the public as much as possible. Marciniak explains: "We are in touch with journalists; we are currently making a movie about the project. We receive invitations to come to schools. All these activities require time. But we try to organize our work as efficiently as we can and invest the time into all the activities that need to be conducted."

- *What ethical or legal issues do you encounter in your research?*

The first issue was to find an efficient way to ensure informed consent from authors of the entries. The material for the entries is in some cases obtained from several writers, authors or informants in the field. As owners of the story they have to agree or disagree that their words, names and other personal information will be published online. The team always asks for their permission and obtains informed consent. For this purpose, the team's researchers prepared an ethical questionnaire that is filled in before publishing the data. Even though this is an extended procedure that requires time and effort, the researchers are content with the chosen approach.

In addition, to make the survey more visually appealing to the users, book covers or shots from movies are used as illustrations for the entries. However, these cannot be taken from the internet and used freely, if they are not in the public domain. After consulting a lawyer specialising in copyright law and considering the country differences, the team elaborated a procedure and follows strict rules to obtain such permissions. The permissions are collected and uploaded to an archive. Professor Marciniak notes: "If we do not currently have the permission to use an image, we use an icon instead and we try to obtain the permission in the future. So far, we have usually succeeded in obtaining the permissions. In some cases of big companies this might be more complicated as they require a fee for the use of an image, and their contacting procedures may limit the possibility of interaction to explain our idea. But these are very rare cases. Usually people are willing to contribute and are happy to share their work. We believe it is partly because the platform has a societal impact that can be easily seen and grasped by the authors, editors and storytellers. They can see that their work is important, that it is discussed, and it is analysed."

1.4. Impact of open science practices

- *What motivates you and your colleagues at OurMythicalChildhood to devote time and effort to disseminating your research and openly sharing research results?*

From the start of the project research data sharing and being open was a crucial aspect for researchers in the OurMythicalChildhood project. “We are a team of similar people who believe in similar values”, says Professor Marciniak. “Research communication and engagement of the general public are very important from my personal point of view, and everybody in the project shares this idea”. They also share the idea that it is important to disseminate the intermediate results, even though the work is still ongoing, or, in the case of the online survey platform, some improvements are still under way. Such an approach allows the team to share the peer reviewed content, which is the most important for them, and solve the technical issues simultaneously.

According to Professor Marciniak, such an approach has an added value for society as research findings can be taken up immediately: “Being open is important because of the research itself and the further development of science. For example, our colleagues from Cameroon are collecting the local myths, which is something extraordinary that happened during the project and that was not planned initially. And other scholars can take these myths immediately and work with them further. There is no delay and no need to wait until final results are published. They do not need to pay for the books or wait for the inter-library exchange. The same applies to research on the use of classical myths while working with autistic children. One can wait for four or five years before publishing all the results or make them accessible from the very beginning. We took the latter approach and published the first set of materials on one of the blogs of our projects. Now it is accessible to everyone and can be used immediately by the professionals working in this area”.

Professor Marciniak points out that there is a growing understanding that engaging in research dissemination is important. The project team members receive a lot of support from their host institutions, particularly with regard to dissemination efforts. “Colleagues from our department also agreed to give their time voluntarily and participate in specific activities directed towards children, such as participating in the jury of the video competition”, says Professor Marciniak. She believes that the public needs to be aware of what the researchers are doing and why they are doing it. This way people understand the purpose of scientific research, what problems it helps to solve and how it impacts society.

2. The Intergenerational Cumulative Disadvantage and Resource Compensation (INDIRECT)

Summary

Professor Jani Erola and his team have come to appreciate the benefits of open science practices while working on the ERC project INDIRECT. Professor Erola was not highly aware of open access issues before receiving his Consolidator Grant in 2013. Currently his team aims to ensure open access to articles and books as much as possible and have also set up an online platform to disseminate their working papers while they are undergoing peer review. These approaches provide opportunities for the team to generate greater interest in their work and achieve a wider impact through their published works.

2.1. Introduction

The project *Intergenerational Cumulative Disadvantage and Resource Compensation* (INDIRECT) started in 2014 and is funded by an ERC Consolidator Grant. It is led by the Principle Investigator Jani Erola who is Professor of Sociology at the University of Turku, Finland. He and his team study intergenerational inheritance of socioeconomic attainment in households, with a focus on how resource compensation occurs in different life circumstances. Resource compensation can come into play when families lose resources, due to such situations as parental separation or unemployment, and attempt to compensate for these losses either by using other resources, which have not been lost, or by relying on help from other family members or neighbours. By extending the idea of resource compensation to a broad range of situations in which it may occur, the INDIRECT project aims to advance the theory and empirical evidence for intergenerational socioeconomic inheritance.

Resource compensation is studied in the project in three contexts:

a) Life-course changes followed by the loss of parental resources. The specific events considered in the research are parental loss, separation, unemployment and geographical mobility.

b) Changes in society that reduce resources in many families approximately at the same time. The examples include economic recession, the inflation of educational credentials due to an increasing overall level of education and changing family structures and family formation processes.

c) Structural disadvantage associated with lower level of parental resources. The forms of inequality to be analysed include the number of siblings sharing the parental resources, childhood neighbourhood and the compensation of low resources with the resources of the parents of the spouse.

The team does not collect new data but rather uses already existing household data resources. The core of the data comes from high level Finnish register panel data from Statistics Finland. The results are compared to those acquired from German socio-economic panel (SOEP) data and US-based Panel Study of Income Dynamics (PSID) data. Multiple country comparisons are conducted using European Social Survey (ESS) data. Access to the Finnish register and SOEP data is restricted, whereas the PSID and ESS data are accessible to everyone.

2.2. Successful open science practices used in the project

- *What are the main outcomes of the project so far?*

The project outputs until now include several peer reviewed journal articles, book chapters and working papers that analyse various aspects of intergenerational attainment from the perspectives of life course events (such as parental death and divorce), period changes, and structural inequalities. In addition, the project has produced one book and the PI has plans to produce another monograph, or at least a manuscript, and several more articles by the end of the project.

The articles have been published in open access when possible. Professor Erola notes that the grant money reserved for open access publishing might not be sufficient for the whole duration of the project. Thus far, however, they have had institutional funding that has been used to cover article processing charges. The Unit of Sociology, where Erola is based, has been well funded for the last year. This is because due to the research grants received from external funders, the Unit receives supplementary funding from the government. Such funding is not tied to positions and can be spent on costs such as article processing charges.

The chapters resulting from the research have been compiled into a book published by Edward Elgar publishing. The agreement between the research group and the publisher allows researchers to share their chapters of the book openly six months after publishing. Professor Erola admits that the team initially did not plan this: "We actually did not consider open access for books initially and it turned out to be a lucky coincidence. We did not have to pay anything for publishing the book either. The publisher is trying to expand their social science research publishing, and this might be a reason why they also provide good terms for publishing. To my understanding other publishers do not offer such good conditions."

In order to promote their work, Erola and his team have participated in a new initiative. The Turku Center for Welfare Research (TCWR)[6], of which their department is also a member, has set up its own working paper series *Working Papers on Social and Economic Issues (WPSEI)*[7]. Most of the papers that the researchers in the department have submitted to a scientific journal for publication are also uploaded to the working paper series website. According to Erola, this was an important development for his department.

- *Why did you decide to experiment with a working paper series?*

For Professor Erola the main reason was that publishing of a scientific article takes a long time: "It takes often two years and sometimes even more from the initial submission to the final publication". Uploading the working version of a paper online gives more visibility, generates interest and can even generate citations. This is helpful from an impact point of view, as Erola explains. In addition, when they started to develop the idea, other repositories were not well-established, and this was an easy solution for Erola and his colleagues. He also notes that relatively few resources are required for running such a platform: "The website is very low-cost and people managing it are doing it voluntarily. They make sure that the articles are presented in a similar format; a series editor publishes the working paper and takes it down once the article gets published. And that is about all that the platform requires."

[6] http://www.utu.fi/fi/yksikot/soc/yksikot/sosiaalitieteet/muuta/tcwr/Sivut/home.aspx
[7] http://wpsei.utu.fi/

- *How do journal publishers react to such an approach?*

Professor Erola explains that publishers are fairly open to this solution and that the process works out fine. The article is uploaded only during the peer review process. Once the paper is accepted for publication, the working paper version is taken down from the WPSEI website and a link to the published paper is provided. "As the published version of an article is often also an improved version, it is better to have only that version available", adds Erola.

The research team also uses social media (mostly Twitter) to disseminate information about their work and to inform their peers about the new papers in the working paper series. "And people tend to read them, which is sometimes surprising to myself", says Erola, "this is working very well for me and for the project as the papers are building up audiences even before they are officially published in peer reviewed journals or books".

2.3. Challenges faced and success achieved

- *Did you face any challenges, or do you see potential challenges when publishing in open access?*

"Until now we have paid substantial amounts for open access fees for different journals including for the hybrid model. And that could sum up to a lot of money, which is of course not sustainable in the long run", says Professor Erola. He sees a large variation in open access publishing costs across the scientific fields with sociological journals being on the less expensive end of the scale. He recalls that some of his colleagues have spent 3700 USD for papers in medical journals. The highest fee he remembers having paid was 2700 USD and the lowest was around 700 USD, the latter being for a publication in a fully open access journal. Still, as Erola points out, such publishing is taking a significant amount of funds, and in his opinion is not feasible for the whole duration of the project. He adds that variation in prices suggests that the true cost of open access articles for publishers is not clear.

- *Have you considered publishing via the green open access route, i.e. by self-archiving a paper in an open access repository?*

In principle, Erola agrees with this approach but emphasises that his preferred option is to have an article published in immediate open access. He believes that this helps to have a bigger impact and to generate more citations. Also, one does not need to dig deeper and keep track of publishers' policies on self-archiving and embargo periods, which also tend to change over time and vary from journal to journal.

- *In your project you reuse data that has been collected by national registries, and access to some of these data is usually restricted. How did you obtain the permission to use these data?*

"We use openly available and restricted data sets as well as various panel datasets" says Professor Erola. "PSID, an American data set, as well as the European Social Survey are entirely openly available, the German Socio-Economic Panel has a more restricted access, and the Finnish register data is the most restricted." Before embarking on their research, the team had to obtain permissions. This was done together with the initial ethics review before the project started. Currently, the project's researchers are still constantly negotiating with Statistics Finland concerning new datasets and new extensions, which are needed for their analysis.

- *Why have you chosen, instead of collecting new data, to reuse already existing data?*

Professor Erola explains that the trend globally is that response rates to surveys are going down and survey fatigue is prevailing. Hence, in his team they are trying to exhaust the information that is already collected and available across different datasets. This is the direction in which researchers in

social sciences are heading. “We need to think of ways of obtaining data about people’s behaviour without interfering in their daily lives. For that, researchers should rely on other sources of available information, whether these are panel surveys, register data or other types of collected data. Businesses, like Facebook, know more about us and have more data than any researchers”, he adds.

According to the PI, data management activities take up rather small portions of their project time as the data is already collected and cleaned. However, they do need good tools for data processing and analysis. Currently, the team uses R and Stata programmes which they find useful and efficient for analysis of such secondary datasets.

2.4. Impact of open science practices

- *How do you think open access, open research data and other open science components will develop in the future?*

Erola admits that initially the added value of publishing in open access was not clear to him, but now he sees it in a positive light. “Things have been changing quite rapidly with the open access policies. It is becoming a more and more obvious thing to do. The first paper that we paid open access fees for is now one of our most read papers and it is likely that it will be picking up more citations during the following years”, says Erola.

In addition, Professor Erola has noticed that sharing the research data underlying an article is increasingly being demanded. “You cannot provide the register data, but you should provide the scripts so that somebody could check what you have actually done with the data. Currently they are kind of ignored, although you can learn so much from the scripts directly. In some cases, the data is already available freely online, so it is sufficient just to provide the links to the original data. I think that is exactly what we should do more and more”.

In general, the open access approach is gaining more and more support in Finland. There are discussions between researchers and universities on the intellectual property rights of research products. There are opposing opinions – on the one hand, researchers should be free and own their products, on the other hand, the universities want to keep the IPR as well. In addition, in Finland there is a tradition of publishing and translating works to the local language and this process is subsidised by public funds. Public funders are now asking to have these works available openly as well. Hence, there seems to be a growing understanding that research should become more transparent and open.

Scientific research is also funded by public funding and publishing the results openly is a way to give the favour back to the public. “It is possible that non-open access publishing would perhaps provide somebody a competitive edge. But evidence for that is rather sketchy and there is better evidence that open access publishing is pushing the science and research forward faster than traditional publishing. And I do not think that researchers in sociology have that much of an economic interest even in the traditional publishing system to justify its continuation”, notes Professor Erola.

3. Computational models for the discovery of the world's music (CompMusic)

Summary

In the CompMusic project, Professor Xavier Serra's Music Technology Group used an ERC Advanced Grant to develop their open science practices in a manner that helped answer their research questions, and enhance the opportunities for commercialisation of the results. Their research is interdisciplinary, developing computational approaches to support the scholarly analysis of music. CompMusic was particularly focused on five non-Western musical traditions and developed a software framework for data analysis. Lacking research data repositories and standards specialised for this relatively new domain, the Group constructed an approach that combines locally engineered open source software frameworks and open data collections. These are managed and shared across a variety of community databases. Through user-provided content and usage the open collections help with software enhancements. Using a dual-licensing model, the software can then be commercialised and generate income streams from subscribers, while also being made available under open source licenses for non-commercial use.

3.1. Introduction

Many of us have direct experience of computational music technology through various software applications to help us identify music played near our smartphones, or to get recommendations of new music based on music we have listened to previously. But what if we were to go beyond the popular Western genres, to the traditional music of Turkey, India, China or the Maghreb? And what if, rather than just sampling the music being played to find an identical match in an online database of recordings, we were conducting scholarly analysis to explore similarities across different musical traditions? Musicological research is the scholarly study of music and extends beyond analysis of patterns evident in the music itself, for example to offer insights on how such musical patterns are shaped by the cultural context. The Music Technology Group led by Professor Xavier Serra at the University Pompeu Fabra in Barcelona is a leader in the relatively young field of music technology, which covers both applied and basic research. The applications include new software tools and digital infrastructures for musicologists to use in their work, as well as commercial applications for music listeners of all kinds. The basic research advances topics such as audio signal processing, music information retrieval, and musical interfaces.

The project *Computational models for the* discovery *of the world's music* (CompMusic) was an ERC Advanced Grant that exemplified the Music Technology Group's interdisciplinary approach, based firmly on an open science model that shared as much as possible, as widely as possible. Completed in June 2017, the project promoted a multicultural perspective in Music Computing. It also advanced the formalization of musical description, enabling computational approaches that take the field further beyond descriptions of audio signals, towards helping machines to deal with semantically meaningful music concepts. Information modelling techniques were developed for some non-Western music repertoires, including representation of culture-specific music contexts, and the Group has explored the commercialization of these results in two ERC Proof of Concept projects. The CompMusic team's experience offers lessons for emergent digital research fields that cross disciplines, have interconnected software and data management needs, and have commercial influences as a major factor.

3.2. Successful open science practices used in the project

- *What was the motivation to start a multicultural project that would involve open science practices?*

The commitment to an open approach was rooted in Professor Serra's ambition to make data and software tools as widely accessible as possible to the musical cultures that were also the source of the musical recordings used in their creation. That aim stems from the initial catalyst for the project, and its focus on non-Western music. This came about through an invited lecture visit to India, where Professor Serra gave a series of talks on the state of the art in music technology. The visit brought the realization that the domain had (until then) worked with Western musical genres and cultural assumptions by default. As a result, its methods and techniques could not be usefully applied to Indian music.

The Indian example inspired Serra to find other examples of musical traditions poorly served by music technology. Working with musicologists, the CompMusic proposal selected five music traditions: Hindustani (North India), Carnatic (South India), Beijing Opera (China), Turkish Makam (Turkey), and Arab Andalusian (Maghreb). By the end of the project, Professor Serra's team had compiled and made openly available corpora of 24 datasets developed for specific experiments around these traditions. The project team conceived computational methodologies to process the audio recordings in the corpora, applying the Group's open source software libraries to conduct the analysis. These led to methodologies that provided features and models for the study of melodic and rhythmic characteristics of the different music repertoires.

The project has exemplified the considerable success the Music Technology Group has found through open practices. The group has established norms for sharing their results as openly and widely as possible. More than 150 publications resulting from the project have been made openly accessible. All the software code is open source, and all the data generated is available under open licenses. Professor Serra and the CompMusic team developed an approach they now apply across the Music Technology Group. The CompMusic project has enabled the Music Technology group to formalize its data management and open research practices, while also ensuring the open resources produced in the project are successfully commercialized where appropriate.

The project's contribution of novel data sources and state-of-the-art software to highly visible repositories has also helped generate opportunities for collaboration and commercialisation. Professor Serra notes: "As a result of all this effort we have had Proof of Concept projects, and there are now quite a number of collaborations at the industrial level. We created a company in India, now leading to a number of other initiatives based on that."

3.3. Challenges faced and success achieved

- *How did CompMusic build on your group's expertise, and what new data management challenges did it bring?*

CompMusic came from several decades of pioneering work in music technology by Professor Serra's group. This inter-disciplinary work has parallels with other fields in computer science and humanities that work with digital corpora, to identify and model their properties and build techniques to analyse the contexts of their production. Corpus linguistics, for example, uses collections of recorded speech. Research in information retrieval uses collections of text and recordings of human interaction in all its varieties. Music is however a relative newcomer. Other fields have established data management standards for describing the corpora and contextual information useful for research. Many have repositories specific to their domain, where researchers expect to find open resources and share those

they have produced. Music is an exception to these, lacking well-established metadata standards that would allow researchers to create machine-actionable music descriptions that help reuse corpora. There is also a lack of specialized repositories to support musicology.

The CompMusic project has nevertheless demonstrated successes in its approach to data management and open research. It had two strands; firstly, leveraging the Music Technology Group's in-house open source software tools to support the analysis of novel data sources; and secondly contributing to generic repositories and international initiatives that facilitate community-specific collections of data, metadata, and software. As a result, researchers in this new inter-disciplinary domain have begun to reuse the Group's resources to serve the computational aspects of the research, e.g. in machine-learning research. CompMusic project outputs have also laid the groundwork for researchers in the musicology domain to benefit from the technology now at their disposal.

The CompMusic project gave Serra a ramp for establishing good practice in managing software and data products created using its existing open source platforms. These include several major initiatives the Music Technology Group has taken towards using these as platforms for open data, motivated by the need to overcome the major legal barriers presented by intellectual property rights. These are Freesound,[8] and AcousticBrainz[9]. Freesound is a platform for tagging audio clips and sharing them using Creative Commons licenses to allow others to reuse them with few conditions. AcousticBrainz is a platform for making information about music, for example the acoustic characteristics, and information about genres, moods, keys, and scales, openly accessible. Much of this information is produced using the Music Technology Group's *Essentia* open source software library, a set of algorithms for common audio analysis processes and audio and music descriptors.[10]

- *What was different about the data analysis framework you developed in CompMusic?*

CompMusic enabled a fundamental re-think of the data representation of musical concepts, and how these are semantically related. "What 'similarity' means to an Indian person looking for similar musical recordings is very different from what 'similarity' means to us when we use technologies like Spotify to get musical recommendations", says Serra. The project has enabled more culturally sensitive recommendations. He draws parallels between the role of the piano and violin in musical traditions. "The violin", he points out, "has been widely adopted by a variety of musical cultures across the world, while the piano has proven less adaptable, staying largely within Western classical and popular music genres and acting as a 'carrier' of Western culture". The characteristics of music that vary radically between traditions, such as rhythm and the unfolding of melody, are well studied by musicologists, who in effect establish the cultural boundaries that Serra's team works within.

Initially, CompMusic adopted the working assumption that each of the five musical cultures it was working with would require its own approach to represent musical concepts, rather than assuming that a universal conceptual framework could be applied. By the end, Serra's team arrived at a common framework, named Dunya, for analysing the corpora.[11] Dunya comprises five corpora, one for each musical tradition (Carnatic, Hindustani, Makam, Jingju, and Andalusian), and the software tools to explore their characteristics. For example, the team developed automatic analysis methods to derive rhythm similarity measures and established a state-of-the-art automatic meter analysis of particular

[8] Freesound: https://freesound.org/
[9] AcousticBrainz: https://acousticbrainz.org/
[10] Essentia: http://essentia.upf.edu
[11] Dunya: http://dunya.compmusic.upf.edu/

music traditions. Through this development of data analysis techniques that can be extended to other cultures, the project has made an essential difference to the Music Technology Group's framework for musicology. This is now being applied to Western musical genres, e.g. jazz.

- *How did the approach to data management and sharing change over the course of the project?*

The project formalized the Group's practices around data and software documentation, as well as the deposition of these to repositories. That more formal approach was required because of the project's size and need to collaborate. It was also shaped by the increasing emphasis on reproducibility in the project's field. "We needed a clearer idea of how to make the code reproducible, which was not there at the start of the project, but by the end we had an understanding of how much explanation was necessary". The support by the ERC grant was essential, says Serra. "It was only with the ERC support for this large project that we were able to move beyond the approaches of individual PhD students".

This more formal approach also involved adopting Github [12] for software versioning and documentation, and making more informed collective choices of the repositories and community databases to use for completed work. The team chose Zenodo[13] as its main repository. In addition to its support for open licensing, this offered several benefits including the ability to inter-link software packages they archived in Zenodo with the working versions in Github, and the ability to link open access articles to the relevant data and software packages. The over-riding benefit from the project team's perspective was that, compared with their institutional repository, Zenodo offered the group more control of the curation practices they were developing.

3.4. Impact of open science practices

- *How did the project affect your local policies on open science?*

As a result of CompMusic, the Group now has guidelines for researchers on open access publication and where to deposit data and software related to publications. The effort to improve practices around documenting and depositing data and software has had a broader impact on local guidelines, which have first been generalized from CompMusic to the Music Technology Group, then to the Information Technology department. Professor Serra believes this has helped the University to avoid imposing top-down solutions such as mandatory deposit in the institutional repository, in favour of their locally-devised approach. However, this has been challenging: "Finding a workable approach to openness at the Group level required a great deal of thought about the centrality of openness in our research philosophy and its relation to research questions and everyday ways of working. Such changes also require a degree of continuity in the new practices. For many groups this will not be a priority, unless ERC or other funders mandate that they must do something", says Serra. According to him, change needs both, the bottom-up community-driven solutions and a top-down policy.

- *What about the impact on individual careers – are young researchers rewarded for the effort they spend on managing and sharing data and software products when they could be writing articles?*

Professor Serra believes there is a strong demand for relevant data management and software development skills to create and manage multi-cultural resources. This outweighs the potential negative impact on the time that young researchers have available to further their careers by publishing articles. This is partly due to the Music Technology Group's strong reputation for

[12] a Git version control repository hosting service which is mostly used for computer code: https://github.com/
[13] https://zenodo.org/

innovation in the field, and partly, he believes, because the field is "not as competitive as some other fields".

Training for musicologists is one of the follow-on initiatives that seek to broaden the impact of CompMusic. While Professor Serra believes the project's results are clearly influencing computational research domains and the online music world, they are more challenging to use for musicologists. The project has created interest among them in non-Western music, but this has yet to translate into a specialised body of work. By offering training in the use of the project's framework of software tools, and specifically the Dunya API (application programming interface) the uptake among musicologists and digital humanists will be improved.

- *Have the resources developed in CompMusic had impacts through reuse of the corpora and the tools?*

On a broader level the impact of CompMusic has improved the multi-cultural representation of online music databases such as MusicBrainz[14] that are very widely used beyond the academic community. As mentioned above, the music technology community lacks specialized domain repositories and the Music Technology Group has not used metadata standards developed specifically for musicology. There are, however, online music data libraries, particularly MusicBrainz, that Professor Serra aims to influence by working with them to broaden the range of their holdings. The site is a database of music metadata derived from Western pop music, which keeps metadata on all available recordings. The Music Technology Group's collaboration with them succeeded in making MusicBrainz a less Western centric database. Similarly, the group has worked with MuseScore, a repository that makes music scores available, to support scores that are not in traditional Western notation.[15]

There are also a number of well-established online resources used in the broader data science domains, that Professor Serra's Music Technology Group draws from and impacts upon. For example, the CompMusic project benefited from participation in Kaggle[16], a platform for the 'grand challenge' competitions the machine-learning community uses to evolve better algorithms for information retrieval tasks. The team contributed the first dataset consisting of audio files to be used in such a large challenge competition.

- *How has working openly benefited your group considering the commercial interest and licensing issues?*

Professor Serra views open licensing as a useful exploitation approach, "We are conscious that not every exploitation strategy can be based on open science", he says, adding, "for example in the Indian company the data collected from users singing is completely closed. But I think we managed to find a good compromise that brings together open science with industrially-relevant situations". Licensing challenges demand workable approaches to dealing with the Intellectual Property Rights applicable to large volumes of individual data records (the corpus), and then to ensure that the additional intellectual property rights that are created through the research are not a barrier to others who want to reuse the corpus or reproduce the research by using software created in it. Uncertainties about the copyright applicable to particular digital works created 'upstream' of the project can present insurmountable barriers to 'downstream' reproducibility and reuse. On the other hand, it is also

[14] https://musicbrainz.org/
[15] https://musescore.com/mtg/sheetmusic
[16] https://www.kaggle.com/

challenging to secure community investment of time and money for open source software maintenance and further development.

The Music Technology Group has applied to the results from the CompMusic project their practice of dual-licensing of software tools. This means that non-commercial use is covered by an open license, with separate licensing for commercial applications. In doing so the Group has demonstrated that open research practices are no barrier to successful open innovation. Through this innovation, CompMusic has contributed to the financial sustainability of the digital infrastructure the Group produces.

Overall, Professor Serra believes the CompMusic project has demonstrated by example that open science can be successful, influencing other groups in institutions around the world. There is, nevertheless, much further work that could be done to promote a greater level of dialogue on the relation between open science, industrial exploitation and open innovation. “Openness promotes collaboration and innovation of services in which, where commercially necessary, some elements can be closed. Similar success stories and good practices need to be communicated more often”, says Serra.