European Commission

Case studies on Open Science in the context of ERC projects – Set 1

December 2017

This document presents the first 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

LEGAL NOTICE

TABLE OF CONTENTS

SUMMARIES

The Epistemology of data-intensive science (DATA SCIENCE)

In the [Data Science](#) project, Professor Sabina Leonelli and her team at the University of Exeter are showing how social science and humanities researchers can embrace open science, while respecting their ethical and confidentiality commitments to research subjects and participants. Key ingredients include creative and positive thinking about what can be shared, negotiating participants' consent to share, documenting the interpretive process so others may follow it, and forward planning of the time to prepare interview materials for sharing. Openness is not just its own reward; it brings new opportunities to influence policy networks, as Professor Leonelli demonstrates.

WORDS FOR ART: The rise of a terminology in Europe (1600-1750) (LEXART)

Leading the [LexArt](#) project, Professor Michèle-Caroline Heck from the Université Paul-Valéry Montpellier had to find viable solutions for the development of a highly complex database and for high open access publishing costs for books that had not been included in the initial budget. The issues related to the database were solved by partnering with the Trier Center for Digital Humanities, who had the necessary expertise, and collaborating with them in a virtual research environment. The LexArt team also succeeded in making all their books open access and staying within the budget by choosing the Presses Universitaires de la Méditerranée, a university press, instead of other commercial publishers.

The Metallurgical Nutcracker: Probing at the Nanoscale the Structure and Properties of Hard Second Phases in Alloys and Composites (PHASENANOCRACKER)

Professor Andreas Mortensen from the Swiss Federal Institute of Technology in Lausanne (EPFL) led the [PhaseNanoCracker](#) project that was conducted in the fields of material chemistry and metallurgy. Although this is not a widespread practice in these fields of science, all the articles coming out of the project were published in open access. Professor Mortensen and his team believe that it is important to make research results accessible and available for everyone, even though it is not encouraged by the current scientific publishing enterprise.

1.   The Epistemology of Data-Intensive Science (DATA SCIENCE)

Summary

In the [Data Science](#) project, Professor Sabina Leonelli and her team at the University of Exeter are showing how social science and humanities researchers can embrace open science, while respecting their ethical and confidentiality commitments to research subjects and participants. Key ingredients include creative and positive thinking about what can be shared, negotiating participants' consent to share, documenting the interpretive process so others may follow it, and forward planning of the time to prepare interview materials for sharing. Openness is not just its own reward; it brings new opportunities to influence policy networks, as Professor Leonelli demonstrates.

1.1.   Introduction

Sabina Leonelli is Professor of Philosophy and History of Science at the University of Exeter. Being a keen advocate of open science, Professor Leonelli is a member of the Working Group on Open Access Publishing of the Philosophy of Science Association; and of the Open Science Policy Platform (OSPP), a body advising the European Commission on policy formulation and implementation in this area. Leonelli is also PI on the project *The Epistemology of Data-Intensive Science*, which began in 2014 and extends to 2019 supported by an ERC Starting Grant under the 7th EU Research Framework Programme. The project explores the philosophical assumptions underlying the choice and use of taxonomies, theories, models and experimental methods in data-centric biology and biomedicine. It also looks at how relevant data infrastructures, and their division of labour, are involved in the collective modes of enquiry and scientific modes of understanding in these fields. A further angle to the research considers how tools for data dissemination enable integration and discovery.

Three years into the project, its detailed empirical investigations are already providing crucial insights on the complex conditions under which data can be made 'open', as required by current EU Open Science policies. This applies to the data managed in the project itself, as much as to the data the project is tracking on its journey through biological and biomedical infrastructures to wider dissemination and reuse. The team are making their observations available through open data repositories, and endeavouring to apply their insights about data governance to their own data management practices. Those insights include the role of security and ethical concerns in the strategies used to integrate biomedical data; the ways in which labels, models and visualisation tools used by databases affect the interpretation of research data and their use as evidence; and the implications of the various ways in which research data are being made 'open'. The project is framing a new epistemological perspective on scientific research, which places data at its centre and helps to explain the impact and implications of data science and 'big data' for contemporary research efforts across different disciplines.

1.2.   Successful open science practices used in the project

- *What has the research accomplished so far?*

"We looked at lots of different cases of reuse of data, and how data infrastructures actually mediate that reuse", says Professor Leonelli. The Data Science project has charted the obstacles encountered by data as they travel from the lab to the outside world, via research networks and the tools, databases, and repositories that transform them on their route to reuse. Publications by Professor Leonelli and

her team demonstrate the significance of situations in which these data are incomplete, not provided at all, or provided in a form that is inaccessible for future research. The studies document how research communities are organised in order to take advantage of large datasets and related digital technologies. For example the SAIL Databank (Secure Anonymised Information Linkage) has collected substantial patient and clinical data volumes over 15 years. The databank "has real expertise in how to manage data" says Leonelli. "It is a very important case study for us on how reuse can be mediated both for ethical reasons and to make reuse more efficient." Another case is MEDMI, a collaboration of the University of Exeter, the MET Office and others in a data infrastructure that produces a mash-up of medical, health and climate data to run correlations. A third case is about somatic mutation data in cancer genomics, through the EBI COSMIC database. Each biomedical case has identified very different issues, and their diversity increases with the biological databases.

The research has produced many more case studies than originally envisaged, and they have involved international collaboration on a broad scale. "We have worked with people in Africa, in the United States, Europe, and Australia throughout this project." This is reflected in a forthcoming edited volume, where each chapter focuses on a particular realm of practice, including social sciences and many of the natural sciences. This edited volume will be the second book to emerge from the project, following Leonelli's 2016 work *Data-Centric Biology: A Philosophical Study[1]*. A third is in the pipeline, described as "an over-arching view of what those studies teach us philosophically about what has changed in the methods of science, in the role of science in society, in how we conceptualise science, now that all these big data infrastructures are being used on a massive scale". Open access publishers are being sought for both.

1.3. Challenges faced and success achieved

- *Have there been any aspects of data management in the project itself that have proved challenging?*

Success in addressing data management challenges requires up-front investment of time early in the project, according to Professor Leonelli: "We spent a lot of the first year of the project setting up for using open data. There are big issues in open data for projects like this." She continues: "When we talk to people about *their* difficulties in managing their data, there are big methodological issues about following the data and picking up the materials you want." Some of these issues also apply to the data the team are collecting themselves. This includes substantial qualitative material from interviews and visits to laboratories and other locations where researchers are based. It extends to their personal archives and their documentation online, including their conference proceedings and publications.

The diversity and nature of this material requires forethought on how to make them open in a meaningful way. Professor Leonelli explains: "It is a very diverse set of data sources, which is very unwieldy to deal with. It is very much in the tradition of qualitative research and ethnography, where a lot depends on the point of view of the researcher and on their interpretive judgements, so there is never really anything objective about the data. However, in this particular case it is very much dependent on the interaction of the researchers and the interviewees. It is very situated and something we have seen is a problem for others and of course for us. It is a really big issue and a question of how to compile all this data to be open."

---

[1] Leonelli, S. (2016). Data-Centric Biology: A Philosophical Study. Chicago; London: The University of Chicago Press. ISBN 978-0226416335

- *How did you approach these issues about making the interview material open?*

To overcome the complexities of opening the interview data, the research team offered straightforward choices to the interviewees: an option of making their interview openly accessible, and commitment to appropriate action when they prefer otherwise. This means offering information to guide the respondents' choice, as Professor Leonelli explains: "We wanted them to think about whether they want their interview to be part of a corpus of open data, and give them the option to opt in to that. What we have done was elaborate, pilot, and then widely use a particular type of ethics consent form." This consent form[2] briefly describes the project's dissemination, and asks respondents whether they would opt in to making the data and the transcripts available online. They may choose to be anonymous, in which case references to individuals and locations are elided. They can also choose to have the transcript available online and attributable to themselves. "There are all sorts of options here and it was quite interesting to put them together" Professor Leonelli says, "interviewees are very happy to have these options that comply with Data Protection law and our own ethical requirements as researchers."

The approach has brought a corpus of open data to a field that has few available. Taking a broad view of the opportunities for data reuse has been key to the project's success in overcoming the typical barriers to working openly in the field. "A lot of people who do my kind of work think that these are such sensitive data that there is not even a question of making them open" says Professor Leonelli. "I actually think one has to be very careful around making assumptions about what it means to have open data for qualitative studies, as it is actually possible. In a study like ours, a lot of the more technical interviews for example are about people explaining in their own words how they make certain databases work for them." Making this information available tends to be less controversial for interviewees, and can be extremely useful for other groups who are interested in understanding how open data works and about its history. "You can generate a corpus of materials that can be beneficial to others. For example, others can use the material and do discourse analysis on it or analyse the user stories. It is worthwhile", says Leonelli.

That success does come with caveats, however. Professor Leonelli explains, "what typically happens is that the more interesting an interview is for our purposes, the less likely it can be made open. The simple reason is that the changes around data sharing are cultural, involving tensions and conflicts between parts of academia and beyond academia". This inevitably means some data cannot be made available openly without betraying confidentiality agreements with the interviewees.

- *How will you make the case study data available?*

The research team is preparing to use Figshare to share a large corpus of data and a dataset relating to each study. The choice of this generic repository reflects a lack of suitable alternatives that focus on qualitative social data. The team have found that Figshare offers them enough flexibility to apply metadata that fits the project aim. The aim is for each collection to be organised to help potential re-users find data relating to particular cases and themes. Considerable effort has been invested in curating the data to make it findable and accessible. Each transcript needs to be checked with the interviewee where they have specifically required this. To deal with the workload, the team has scheduled release of the data to coincide with each paper. A key point here is the need to plan the effort and resources in the grant application before starting the research, as Leonelli says "…any time that goes into this means less time to do the analysis and the research".

---

[2] The consent form is accessible from the project website, datastudies.eu, at:
https://datastudies.eu/images/downloads/other/Ethics_consent_form_DATASCIENCE.pdf

- *You have written extensively about issues in making data mobile and how some data can travel easier than others to be used in another context. Does the effort involved in describing the case study contexts sufficiently make your own data less mobile?*

Professor Leonelli's team documents the 'life history' of particular forms of research data and how they are dealt with in the everyday practices of researchers and others involved. This presents challenges in interpreting and recording sufficient details about the context to enable later re-analysis. "In the data I deal with it is absolutely obvious that data provenance is essential for any kind of re-analysis. The work I do is partly historical, and a lot of the relevant metadata will be in the published article, because that is the story of where the data has come from. So this makes it clear that data reuse needs to be aligned with reuse of the publications coming out of the project, because without that you will not be able to understand the context of data collection." This particular research is acutely sensitive to the need for information on the provenance of data underlying its findings, but the issue is familiar to data management in other disciplines. "It is not so dissimilar with what is happening in the natural sciences. What you get in some fields like economics is the feeling that you could reuse the data without much provenance information. Certainly in the biological sciences anyone who does very serious work with big data analysis is adamant that they're not entirely sure what the data mean or how to manipulate them properly without being able to disaggregate the data and dig back to where it comes from. It's important everywhere, it is just that in qualitative social science it is unavoidable. The use of numbers in other disciplines may hide the fact that you need to go behind the numbers to specific situations." As Leonelli explains, in qualitative research the chain from initial observations to findings may involve many layers of description and interpretation of the data gathering context. This makes the contextual and provenance information impossible to take out; a strength of the approach as "… it can get to things that measurement does not give you."

This sensitivity to context and ethical challenges it presents can also limit how much of the data can be made open. But the key point for Leonelli is to share what can be shared: "the idea in open data is to focus on data sources that can be reused in other projects. They may contain interesting information about how certain resources were put together and what lay behind it. And this is the kind of information that can be very useful for other people thinking about these issues. Things like people explaining their motivations, and what were the challenges at that particular time in that particular location and in that field. This is the kind of thing that can be important to the history of curation in certain fields. The differences between fields, over time are very important as we are in a very transformative moment in science. It will be interesting to look back on how different researchers were articulating their position."

1.4. Impact of open science practices

- *In some of your publications you highlight differences between making data accessible and actually reusable, and describe capabilities people need to reuse data. Does this work help the data infrastructures themselves in understanding what they need to do to help that?*

Data infrastructures need to ensure the use of metadata standards that reflect the discipline enough to support meaningful searches of their data collections, and to help reconstruct the analysis. As Leonelli puts it, "metadata will always be partial; the question is what rationale is given for them." Domain repositories that offer support for searching using disciplinary metadata standards, such as the ISA-TAB standard used in biosciences, have a distinct advantage in that they inform users about the scope of what they may search for. Expert curation is invaluable for reuse: "The best cases of reusable data that the research team has looked at are those where knowledgeable curators are involved; people who know the data inside out – and can advise on how to formulate the search." In Leonelli's own research, however, there is a lack of metadata tools that fit the methodology well enough to allow data

to be appropriately labelled at the right time and then equip data re-users with context-specific search terms. "In our kind of social science we are stuck in thinking about simple archival labels… it will be interesting to signpost these data using keywords to help reconstruct the analysis. The approach we use is very much a Grounded Theory one, where we are open to new things coming up. For example, security concerns and what they mean for data management."

- *From your experience, does the Grounded Theory approach offer lessons for other disciplines that use big data?*

The Grounded Theory approach used in the project is a well-established approach to theory-building in qualitative social research. It involves a highly iterative and data-driven approach, developing a conceptual framework to account for the analysts' descriptions of the phenomena studied, and explanations of recurring patterns. This 'inductive-deductive' reasoning is also found in big data analysis. In Leonelli's experience however, some disciplines embracing big data are placing too little emphasis on the role of interpretive judgement in their evidence trail. "The problem with some of the other disciplines that use big data is the expectation that the patterns they spot will be acquired automatically. Certainly people see things they have not seen before, in biology and other cases, because of the algorithms they have applied. But the second important step is to verify what that means and how it can be justified, and that requires very specific judgements from the investigator". Human judgement by experts in the subject is also needed to judge whether a pattern makes sense or is just an epiphenomenon of the instruments, or the sampling approach.

- *How is that interpretive ability and context sensitivity emphasised in developing the research skills for data science?*

In Leonelli's experience "it is absolutely not emphasised". This has motivated course development that aims to improve skills in providing contextual information, to complement the technical application of algorithms in data science. "There is a huge danger in taking a wrong direction by conceptualising of data science as just a question of devising cleverer algorithms. Anyone we interviewed who does interesting work with big data does not operate like this. Of course it is important to have better software, but the fact there has to be judgement and contextuality involved in every stage of analysis is really important." The risk is that training programmes for data skills under-emphasise these less technical skills.

- *Does the ERC-funded work also feed into your work in the Open Science Policy Platform?*

The project's outcomes have informed Leonelli's view of the policy requirements for Open Science, including data science skills development. "I am in the OSPP representing the Global Young Academy, the viewpoint of established researchers not in the most senior stage of their career, and they very much have a stake in what is going on right now. I am also working as an expert for the European Commission on a *Mutual Learning Exercise*[3] developing a road map for open science in 14 countries, and the ERC project has very strongly informed that."

As well as highlighting the issues mentioned above, the project has influenced several reports by Professor Leonelli. They include a report on improving the incentives and rewards available to researchers for working openly; one of the key issues for the OSPP, and for the broader research

---

[3] Mutual Learning exercise on Open Science – Altmetrics and Rewards, 9 January 2017 to 17 January 2018.
https://rio.jrc.ec.europa.eu/en/policy-support-facility/mle-open-science-altmetrics-and-rewards

community.[4] A further report for the Swiss government describes the impact of big data on biomedical knowledge production.[5]

[4] Leonelli, S. (2017). Mutual Learning Exercise on Open Science – Altmetrics and Rewards. Incentives and Rewards to engage in Open Science Activities (Thematic Report No 3). https://rio.jrc.ec.europa.eu/en/library/mle-open-science-%E2%80%93-altmetrics-and-rewards-incentives-and-rewards-engage-open-science

[5] Leonelli, S. (2017) *Biomedical Knowledge Production in the Age of Big Data.* Report for the Swiss Science and Innovation Council, ISBN 978-3-906113-52-4.
http://www.swir.ch/images/stories/pdf/en/Exploratory_study_2_2017_Big_Data_SSIC_EN.pdf

2. WORDS FOR ART: The rise of a terminology in Europe (1600-1750) (LEXART)

Summary

Leading the [LexArt](#) project, Professor Michèle-Caroline Heck from the Université Paul-Valéry Montpellier had to find viable solutions for the development of a highly complex database and for high Open Access publishing costs for books that had not been included in the initial budget. The issues related to the database were solved by partnering with the Trier Center for Digital Humanities, who had the necessary expertise, and collaborating with them in a virtual research environment. The LexArt team also succeeded in making all their books open access and staying within the budget by choosing the Presses Universitaires de la Méditerranée, a university press, instead of other commercial publishers.

2.1. Introduction

Michèle-Caroline Heck, professor of modern art history at Université Paul-Valéry Montpellier, received an ERC Advanced Grant in the Cultures and Cultural Production panel for a five-year period, starting in April 2013. Her project *Words for Art: The rise of a terminology in Europe (1600-1750)* investigates the emerging art vocabulary during the 17th century and its further development at the beginning of the 18th century in Europe north of the Alps. This is when, in theoretical art literature in England, the Netherlands, Germany, and France, an artistic terminology began to constitute itself, which addressed as much artists and art connoisseurs. "With this project we wish to propose a counter-model of the current idea that theoretic writing existed independently from artistic creation, showing on the contrary that artistic writing is indicative of art practice." – says Professor Heck.

The project brings to light artistic relations and the development of a common language for European art communities in the early modern period. It therefore studies the written word, because it plays a key role in the transfer of practices, and is the active agent in the circulation of concepts. Especially in the circulation of texts across Europe the word actually transformed its meaning according to the context and language in which it was used. The same word could thus have different meanings, or inversely different words could carry the same meaning. It is from this perspective of the analysis – the transformation of the word – that the research team situated the project. Hence, LexArt does not present words as having only one accepted meaning, but embraces the possibility of diversity, which also presides over the act of writing itself. The research concerns both the study of the word based on its explanation and contextualisation within the sources it originated from, as well as the study of these words as an instrument for the intellectual work of the theorist, as well as for the more practical activity of the artist. Professor Heck explains that "this approach opens up new perspectives: a deepening of our knowledge about manner and styles, a broadening of our consideration of the analysis of works of art, and the definition of the criteria, which permits us to resituate the creation in both the social and semantic context." LexArt thus makes it possible to comprehend the transformation and circulation of concepts and practices, as well as to observe their initial introduction into the different intellectual networks, revealing the permeability of artistic borders. The main achievement of the project is a new research apparatus, serving further research in art history. It also provides a conceptual framework, which could be reused by other research communities.

2.2. Successful open science practices applied in the project

- *What are the main outputs of the LexArt project?*

The aim of the LexArt project is to establish conceptual tools, which are necessary to stimulate and rethink art theory research in Europe, as well as to propose a new foundational research apparatus that is of synchronic and critical nature. To that end the LexArt project has three main outputs.

First, the project created a volume of theoretical essays with interdisciplinary approaches and methodological and cross-cultural questions, which is forthcoming in December 2017. Second, the LexArt team is currently working on a dictionary that contains the artistic terms and concepts of early modern Europe. They aim for publication in March 2018. Both of these publications will be published in print as well as digitally as open access, and the Dictionary will be available in French and in English.

Third, the project is developing a database of artistic terms and concepts, organized in multi-lingual entries (French, English, German, Dutch, and Latin) with semantic and conceptual cross-references. All entries are linked to a digital library that contains images and digital versions of the books the terms originated from, giving the user the possibility to have immediate access to the source and the context of each term. The database has been developed in collaboration with the Trier Center for Digital Humanities[6] in Germany. It supports linked open data, visualisation of networks, interlinked data, formal data modelling, representation of data relations, and collaborative, web-based data capturing. As Professor Heck points out, "LexArt will be able to continuously receive new information, such as new entries, images and source texts, and add it to the database even after the project's completion. Further, the formal description of the interlinking between the information entities and the underlying data modelling can be seen as a general way to capture these structures in a virtual research environment. And this opens the possibility to reuse the data model for further projects." The University of Trier is currently planning a similar project to examine historical texts from the field of jurisprudence in the same way the LexArt project team did, and some of the visualization models from the LexArt project could be reused for their research.

2.3. Challenges faced and success achieved

- *What was your experience in terms of open access? Did you face any challenges in publishing OA or in making the database publicly accessible?*

LexArt faced two core challenges. The first was that of open access publishing. At the time when Professor Heck submitted the project proposal, open access was not as relevant as it is today and she did not include open access costs in the provisional project budget. However, according to Professor Heck the costs of open access publishing in her field are very high, approximately €80-85 per page across all academic publishers. Hence, financing posed a great obstacle. The challenge was to stay within the budget and to find a publisher who would agree to publish the books in print and to also deposit them on an open access platform.

As Professor Heck recalls, the second challenge emerged in the context of developing the database of artistic terminology. There were three main issues. First, the database was supposed to be made openly accessible and the project team had to find a suitable company or a development and hosting partner that would agree to release the database as open access. They found the University of Trier, in particular the Trier Center for Digital Humanities, to be the best partner to establish a research

---

[6] http://kompetenzzentrum.uni-trier.de/en/

agreement with. "This was a very good solution because the Center specializes in Historical Linguistics and Digital Humanities, and both teams' interests converge in these fields." – explains Professor Heck.

Secondly, the database is highly specific and its design requires unique development solutions. It does not hold the full text for each book but instead the LexArt team chose relevant quotations for all terms. The database links the terms and quotations to the full-text books they originate from in a way that the user has immediate access to the source and the context of each entry. Regardless of the technical issues, the immediate question was how to convince the libraries, where these books are located, to provide the LexArt team with a PDF file of the digital book and to allow the team to use it by adding it to the database, and hence making it open access.

And the third major issue was faced by the Trier team. This was regarding the formalization and modelling of highly complicated networks, such as the underlying data and the relation between the terms, their conceptual fields and the excerpts from the original sources. All this had to be managed by creating an appropriate data model within the virtual research environment, the Forschungsnetzwerk und Datenbanksystem (FuD)[7] by the University of Trier, and by developing a sophisticated workflow for the team.

- *Since you did not include open access publishing costs in the grant proposal, how did you manage to provide open access to all research outputs, while still staying within budget?*

Professor Heck explains, "All LexArt publications will be made available as open access, both in French and English versions thanks to the support from the Presses Universitaires de la Méditerranée (PULM); they support the concept of open access and offered us good financial conditions, which would not have been possible with commercial academic publishers. The printed version will be published first, and the digital open access version will be made available after an embargo period of 6 months."

For all materials in the database, the contributing libraries and partners agreed on a Creative Commons CC BY-NC-SA license. The fact that several research institutes are already moving towards more open policies was particularly helpful for acquiring the books and images. For example, the Getty Research Institute was of great help for the LexArt project, supporting open access and having launched their Open Content Program[8] in 2013 eased sharing materials. Their policy allowed to provide free of charge digitisations for almost all the books the LexArt team intended to include in their collection. "The LexArt team presented the project at the Getty Institute (Los Angeles) in October 2014. The Getty Institute was entirely convinced by the value of the project and agreed to contribute by providing PDF versions of books they had already digitised, by giving access to books previously only available to the staff, and by digitising all the books in their collection (and even buying some) that the LexArt project needed. This was a first step that to a certain extent was followed by other libraries in Europe with which LexArt collaborated.", Professor Heck remembers. In addition, the LexArt project received contributions from e.g., Gallica, the French National Library (BNF Paris), the Institut national d'histoire de l'art (Paris), the University of Heidelberg, and several libraries in the Netherlands. For the materials digitised with funding provided by the German Research Foundation the approach slightly differs. For these contributions a link to the original digitisation had to be integrated into the database instead of providing the digital version directly through the database. Based on all this generous support, overall about 90% of the requested materials can be made openly accessible through the LexArt database.

---

[7] http://www.fud.uni-trier.de
[8] http://www.getty.edu/about/whatwedo/opencontent.html

- *What obstacles did you meet when developing such a complex database?*

"Looking back on how the project evolved, the development of the database was a very important tool for the project, and emerged very differently compared to what was planned in the proposal", Professor Heck says. For her it seemed essential that all issues were fixed before the actual development started. For the database, two problems had to be solved: defining its format and finding a partner to collaborate with for its realisation. A formalisation and modelling of the complicated network of terms, quotations and sources had to be developed and implemented in the database. Overall, the database holds over 1200 terms in six languages, including French, English, Italian, German, Dutch, and Latin. These terms are connected to the art literature by over ten thousand quotations.

Another concern was how to store the database to ensure long-term preservation. The database will be archived in Huma-Num[9], a service associated with the CNRS, the University of Aix-Marseille and the Campus Condorcet that archives the content of databases. However, it does not offer services for updating and maintaining operational databases. For the LexArt project, this means that for the publication of the database the next step is to find a service that preserves the database itself but also enables its maintenance to ensure that data can be added to the project continuously.

- *How did the research partnership with the Center for Digital Humanities at the University of Trier help you solve the technical challenges and manage all the data?*

"The Center for Digital Humanities at the University of Trier turned out to be a perfect fit for partnering with LexArt as they have the necessary experience and capacities to support the project," Professor Heck stresses. In this case, the historical, linguistic, and digital humanities interests of both teams and the research questions converged very productively. The Trier team helped build the three networks, in which all terms are included: 1) a linguistic network in six languages, 2) a semantic network, and 3) a conceptual network based on ten conceptual fields with about 50 subfields. Each term is linked to a quotation, which again is linked to at least one conceptual area. Besides, the quotations are also linked to the digital library of digitised texts, which is an integral part of the database. In the coming months and beyond the project duration, further information and more entries will be added by the research team.

However, there were also difficulties in bringing both teams together, as Professor Heck remarks, "One challenge was that the two teams at the universities in Montpellier and Trier do not work in the same area. One team focuses on the research questions and the other one on the data management in the centralised database. The collaboration between the two teams – the computer scientists and the LexArt team – was very important." In this case the use of the virtual research environment FuD was instrumental for LexArt. The configuration of FuD facilitates the long-distance collaboration and the data management in a centralised database, which is directly used for the online publication of the project results. FuD offers a customized work space for interdisciplinary and geographically divided collaboration; it integrates a wide range of tools for the import, export, and analysis of data (text, metadata, georeferences, etc.). Additionally, the Trier team implemented an interactive browser-based monitoring tool. This so called "export tool" offers an interactive view into the FuD database and helps very efficiently to identify inconsistencies within the data. Besides, the tool provides a graphic user interface that allows for very intuitive and easy access to the data when examining the relationships between terms, their sources and conceptual fields. This tool will further serve as the base platform for the official publication of the database.

---

[9] http://www.huma-num.fr/

2.4. Impact of open access publishing and other open science practices

- *What will be LexArts's impact on current research practices in your field?*

Professor Heck explains: "In the course of developing the database, a highly sophisticated network of interlinked information was created. The underlying data model will be made openly available and can be reused by other projects with similar aims." The purpose of the project is connected with an innovative methodology based on a new geography in the practice itself and is also linked to other disciplines. Other teams have already contacted the LexArt team, e.g. the Getty Research Institute who have signalled interest in discussing further ways of collaboration.

Overall, the instruments created by LexArt allow a new reading of art literature and a transversal perspective in time and space. The tools – the dictionary and the database – introduce a new research practice for art historians, particularly the possibility for diachronic analysis and to connect to other research discourses. Many researchers from other disciplines have already approached Professor Heck as they intend to do things similar to the LexArt project in their field, i.e., defining relevant words and concepts and connecting them to the relevant contextual information.

- *Do you think that making the outputs of this project open access will have an impact on your career and/or on your field of research?*

Professor Heck does not think publishing open access would have an impact on her career development, since she already has an established position. However, the project will be highly relevant for the members of the research team.

She remarks, "Open access is a very good step for art history research. But three challenges have to be addressed."

First, there is the question of providing open access to the sources, i.e. books and pictures, because the researchers have to pay to publish. Regarding the books, thanks to the generosity of the Getty research Institute, the LexArt team could obtain the digitalised versions needed for their project. Although it seems that the issue of open access to research articles and books is now well advanced by libraries, open access to images still remains an open question to be solved. Many museums, even public ones, do not open up their picture collections. Professor Heck points out that the fees for use of images are frequently very high (approximately 70€ for a single photo and more for an open access publication). "It might be understandable that private collections hesitate to give away free authorisations for reproductions of their own, private collection, but this becomes a different question in the public domain. Such a situation terribly restrains academic production and the diffusion of its results in an open way. " – says Professor Heck and continues:" We need absolutely different rules for all public institutions (mainly museums and print departments of libraries) of all EU countries for example, access to material and free authorisation for reproduction 70 years after the artist's death as it is the law for writers." In theory, this already exists but when works of art are digitised once they enter the public domain after that period, oftentimes museums (or the photographers) own the photographic rights to the images of their collections themselves. Hence, distribution and reuse of these materials can still be limited even after the initial copyright protection has ceased. For this reason Professor Heck sees it as "essential to have an open access policy that obliges all museums and libraries to change their policies in a more open direction."

Secondly, digital humanities projects involve a lot of work, a lot of information needs to be managed and managed precisely, especially if it is very granular. But this requires a new responsibility for the researcher. It is necessary to be aware of the limits of what LexArt can offer with the database and of the limits to digital humanities research in general. One concern was to avoid producing research that could possibly mislead other researchers, Professor Heck explains, "for instance concluding that an

idea doesn't exist when they cannot find the word in the database." Hence, all research results must be very clearly explained, in order to avoid misunderstandings. It is essential to constantly think about how a database can open a new way of thinking in a heuristic approach.

Third, there is the challenge of the database to be recognised as a research-tool, produced by researchers and for researchers, and to grant it a status similar to that of a paper publication in order to encourage young researchers' involvement in such projects.

3. The Metallurgical Nutcracker: Probing at the Nanoscale the Structure and Properties of Hard Second Phases in Alloys and Composites (PHASENANOCRACKER)

Summary

Professor Andreas Mortensen from the Swiss Federal Institute of Technology in Lausanne (EPFL) led the PhaseNanoCracker project that was conducted in the fields of material chemistry and metallurgy. Although this is not a widespread practice in these fields of science, all the articles coming out of the project were published in open access. Professor Mortensen and his team believe that it is important to make research results accessible and available for everyone, even though it is not encouraged by the current scientific publishing enterprise.

3.1. Introduction

Andreas Mortensen is Professor and Director of the Institute of Materials at the Swiss Federal Institute of Technology in Lausanne (EPFL), where he heads the Laboratory for Mechanical Metallurgy. Since 2017 he is also the Vice President for Research at his institution. Professor Mortensen received an ERC Advanced Grant in 2011 in the Products and Processes Engineering panel. The PhaseNanoCracker project (*The Metallurgical Nutcracker: Probing at the Nanoscale the Structure and Properties of Hard Second Phases in Alloys and Composites*) aimed to probe the strength of 'second' phases in multiphase metal alloys and composites, meaning of hard particles added to strengthen a metal, or alternatively of brittle inclusions that weaken it. Such phases are ubiquitous in structural metals although not much is known of the microstructural features that govern their strength. The underlying hypothesis of the project was that defects that limit the strength of such hard second phases can be identified and then altered by processing. PhaseNanoCracker succeeded in quantifying the strength of very small particles, including alumina particles, silicon particles in aluminium and carbide particles in steel. The research team achieved this by developing their own methods and by employing such techniques as focused ion beam machining and adapted mechanical nanoprobing. According to Professor Mortensen, the project achieved very exciting outcomes: "The results of the project were quite spectacular. We showed that the strongest the particles can get is in fact amazingly strong".

3.2. Successful open science practices used in the project

- *What were the key outputs of the PhaseNanoCracker project?*

According to Professor Mortensen, the main outputs of the project were the publications, which were all published in open access (OA). Concerning the motivation for doing so, he explains: "As OA publishing was in the guidelines of the ERC, the team followed them. We devoted part of the project money and also some of the internal laboratory's money currently (as the project is finished now) towards making all the papers openly accessible. It was largely done because the ERC asked this but also because the ERC allowed to take some of the project funds to cover the publishing costs."

- *Did you also produce some research data and did you share it?*

"When the project started five years ago, the whole vogue of storing data and making it available was not there yet and we did not make any particular effort to put our actual raw data online", says

Professor Mortensen. There was one original practice that the research team employed in the project: Some of the methods used in PhaseNanoCracker called for numerical simulation using the finite element code. Mortensen's team openly shared on the website of his lab the portion of the code for one set of simulations. This way the researchers who want to use the PhaseNanoCracker's methods and testing techniques can download some elements of the code which can help them make the simulations. A paper that will soon be published will provide a link to this code. This will help the research team see how many people are interested and download the code. At the moment, the code is put on the website of Professor Mortensen's laboratory but if many people download it, the research team might look for a more visible site or a specific repository.

- *When the PhaseNanoCracker project received funding, publishing articles in open access was not required by the ERC. However, all publications you produced during the project were published in gold open access. Why did you choose this approach?*

"Even though open access was not mandatory when PhaseNanoCracker started and I could have saved a few thousands of project money, it is a philosophy of the funding agency and it is something that I personally agree with", says Professor Mortensen. As the ERC also provides means to implement open access, the researcher made a decision to publish articles produced in PhaseNanoCracker in gold open access. In addition, he was curious to see if publishing in open access would increase the impact of the articles. "Of course that is difficult to establish, there is no data on that. One of the papers we produced is being looked at quite a lot. However, it is difficult to say, if it is because it is OA or because the work is interesting and others are picking up on it", he explains.

- *Is publishing open access a usual practice in your lab?*

Professor Mortensen acknowledges that in general this is not a common practice in his lab. He further explains that this is mainly because it is expensive. Publishing one article in open access can cost about €3 000, with the exception of a few journals. There is one journal in material chemistry and metallurgy fields that charges only €500 per publication. But he also expresses a disappointment with the latter journal: "It is a slightly militant open access journal that was created to precisely encourage OA. We have published in this particular journal once and to some extent I regret this decision. At the time, it was a very new journal and until recently it was not visible in the Web of Science database, which is of course a drawback. It is complicated to find a good balance of publishing in OA and doing it in a good journal."

- *And is open access publishing common in your research field in general (material chemistry and metallurgy)?*

According to Mortensen, publishing in open access is not so simple, since many journals in his field are what is called "double-dipping" or "hybrid" journals. This means that the publisher earns money twice: on the one hand, libraries have to pay for a subscription to the journal, on the other hand authors who want to make single articles within the journal openly accessible have to pay an open access fee. "In my field the journals of Elsevier and Springer are kings. In the material science field, many journals belong to Elsevier or to the Springer Nature group. The publishing in gold open access that I did in the frame of the ERC grant was mostly in Elsevier journals. And this required accepting that I would pay for double-dipping journals", says Professor Mortensen. He notes that generally he practices rather green open access for publications coming out of other projects, where he uploads a pdf file of a post-reviewing and pre-publication version in his organisation's repository. Publishing in gold open access in PhaseNanoCracker was largely done to fulfil the expectations of the sponsoring agency, and because the agency allowed it to be done by providing funding. "I also think it is a good thing to do", he adds, "even though there are a number of issues, for example double dipping, that are problematic."

Other funders, like the Swiss National Science Foundation (SNSF), do encourage and promote OA. But they do not reimburse open access fees for publications in hybrid journals. What Professor Mortensen was able to put as project costs for gold OA publishing under the ERC grant, he could not get that funded by the SNSF. If researchers receive SNSF funding, they have to publish in subscription journals and then practice green open access, or in fully gold open access journals. According to Professor Mortensen, it is a rather strict policy and, "it is not very pragmatic in my point of view, but it is their policy and I accept it."

3.3. Challenges faced and success achieved

- *Since publishing in open access was not a common practice in your lab before the ERC grant, did you face any challenges in publishing the articles in open access in your institution?*

"No, we did not face any particular challenges at our institution. I became the Vice President for Research at my university and I am working on various issues, including OA publishing", states Professor Mortensen. For him, the main problem is double-dipping which they are working to address. The library at EPFL is counting all the articles published in gold open access. This information can then be used to negotiate with publishers when journal subscriptions are to be renewed and the subscription fees are to be paid. Apart from the issue of the institution being charged twice, he does not see any negative points about publishing in open access.

Regarding copyright issues, the researchers at EPFL are very well informed by the library whose staff provide advice on which licence to use when publishing articles. Also, although some of the outcomes of Professor Mortensen's research in other projects were patented, this was not the case in PhaseNanoCracker. Hence, the team did not encounter any problems related to intellectual property questions.

According to Professor Mortensen, the copyright law as applied now is at the root of the problem of the scientific publishing enterprise. Scientific publishing is run under a legislation that is not designed for it: the copyright law is not appropriate for academia but rather is designed to protect the authors of novels. He states that many people also ignore an important aspect of scientific publishing – it is a service that costs and has to be paid for. "What we buy when we publish and what we buy when we read scientific journals, is a service. We do not buy the content, we buy the service. Hence, from my point of view, the whole model of scientific publishing today is built up from a wrong starting point", says Professor Mortensen. He thinks that in order to correct it, it is important to go to the core of the problem, namely the copyright law. However, large publishers object strongly to any changes to the law, as this is their soft spot. As long as this problem is not fixed, one should try to make the most of a situation that is inherently flawed. "And if paying $3 000 for publishing services is what it takes, so be it. It is costly and that is a barrier and a significant one. That is why I do not always publish gold open access as I find it a luxury for the time being. But it is inherently right and smart to make your paper more openly accessible", he explains.

- *Do you see issues that hinder wider OA take-up in your field?*

Professor Mortensen considers that there are a few things that could be broadly viewed as obstacles to wider OA uptake in his field. First, he mentions the absence of strong open access journals in his field that would not involve double-dipping. There are good journals but not the ones that are regarded as the top ones by the scientific community. In his opinion, it is not fair to ask a researcher, the PI or even more importantly the young people working on the project to choose journals of lower quality just to be militant about publishing in OA and ignore the journal ranking. He admits that he has done it once and regrets this decision. Professor Mortensen remarks, "there are various publishers and some have a more proactive OA policy. At our university we have a deal with one of them. If the university is

subscribed to a journal of this publisher, a researcher can publish in gold open access at a reduced rate. And this deal helps us reduce the costs of publishing at least to some extent." Still, the challenge of a lack of high-ranking OA journals in his field remains.

Another issue raised by Professor Mortensen is that the PhaseNanoCracker project ended in April 2017. Some of the work is still ongoing and scientific papers are still being produced. However, the team is not able to charge fees of the publications to the project, as the grant is over. He remarks, "I think this is incidentally something that the ERC should think about and maybe they already have. In all logic, there should be a possibility for the grantees to charge publication fees to the grant retrospectively, after the project has ended. After all, the publication fees will comparatively not be large sums of money out of the total project budgets but they can make a significant difference." In Switzerland, the cost of a gold OA paper is half a month of a PhD studentship and in other countries it could be a whole month. Professor Mortensen notes that in countries where the resources dedicated to research are scarcer, a post-grant fund would be particularly beneficial.

- *What resources did you use to cover the post-grant publication costs of open access publishing?*

Professor Mortensen could cover the additional publishing costs from his laboratory's own budget. Laboratories at EPFL receive some base funding that can be used to fund article publishing costs. The papers that have been published during the project were charged to the PhaseNanoCracker project's budget. Currently, the other three papers that are coming out (that will in total cost about €10 000) will be charged to his laboratory's budget. Professor Mortensen thinks that if a fund was available for publications related to the project that are published up to 1 or 2 years after the project has ended, it would definitely help researchers to publish OA more, particularly in countries with lower research funding.

3.4. Impact of open access publishing and other open science practices

- *What was your motivation to publish OA?*

"Overall, I think that what the ERC are doing is the right thing. I appreciate the grant I received and I had a very good experience of working with the agency. If there is a general statement, I was glad to follow the guidelines and do the right thing. And I would be surprised that there would be top-notch people who receive grants from the ERC and would not publish in open access. This would be surprising to me as this is rather straightforward", says Professor Mortensen. He adds that a more unusual practice was to put the underlying code online as well as openly accessible. Even though many people do not currently do this, he thinks that it will develop a lot and will be taken up more in the future.

Although, in general, he could not see negative things that could come out of publishing in open access, there was one thing his research team questioned during their work in PhaseNanoCracker. It was not directly related to scientific papers, but to the additional material that was produced in their lab. The team members were not sure if they should post this additional material (e.g. videos) on the website of the journal where the articles were published. They debated if the material should be given away to the journal or kept within the team. "I think in the end we decided to put the films on the journal's website, even though that meant foregoing to some extent control of what we have", he says. The team's main motivation for this decision was the aim to disseminate their research as widely as possible.

- *What impact do you think publishing in OA will have for your current work and the research field you work in?*

Professor Mortensen thinks that it is still too early to say. The field he works in is much smaller than for example health sciences. If one looks at article downloads and similar indicators, it is still hard to say whether these have been generated because an article was published in OA or because it was an interesting piece of work. He notes that one of the PhaseNanoCracker papers seems to be 'ahead' of the other publications. It is quoted substantially more than his average production. Also, this article was the second most downloaded article in that journal in the month that it appeared. But he remarks that the numbers are small in his field, "If you get cited ten times within two or three years, that is quite a lot." Hence, it is really hard to tell if publishing articles in open access did make a difference or not, even more so as the articles that were produced within the PhaseNanoCracker project are related to a topic that he and his team have not worked on outside the ERC grant. "I could say we were a little bit alone working in that area and the readership is not that broad. We will need to wait for perhaps ten years to see if this work has made a difference", says Professor Mortensen.

- *Do you think your articles published in open access under the ERC grant will have a wider societal impact than articles you published in other projects?*

Professor Mortensen is not sure to what extent open access increases dissemination and impact. He explains, "There are people who contest the idea that publishing in OA increases readership even though it does sound intuitively appealing. It also depends on how you count impact. If we just count downloads, then I think it is clear that OA will increase this kind of measurement of impact. But then you have green open access, you have illegal sharing of papers, do we count those? Something can be completely illegal in terms of copyright laws and have a lot of impact. But by and large, as far as publishing goes, to me it is simple. One should just be able to click and not be charged $30 for reading the results from a publicly funded scientific study."

He further remarks that OA publishing (and not open, reusable and verifiable data, which could receive opposition as it is a more problematic issue), is rather straightforward and it should be done. He does not see researchers thinking of making a financial profit from the copyright law as scientists give away their articles to the journals. Apart from the issue of paying $3 000 for gold open access, he states that he cannot find arguments why OA should not be done.