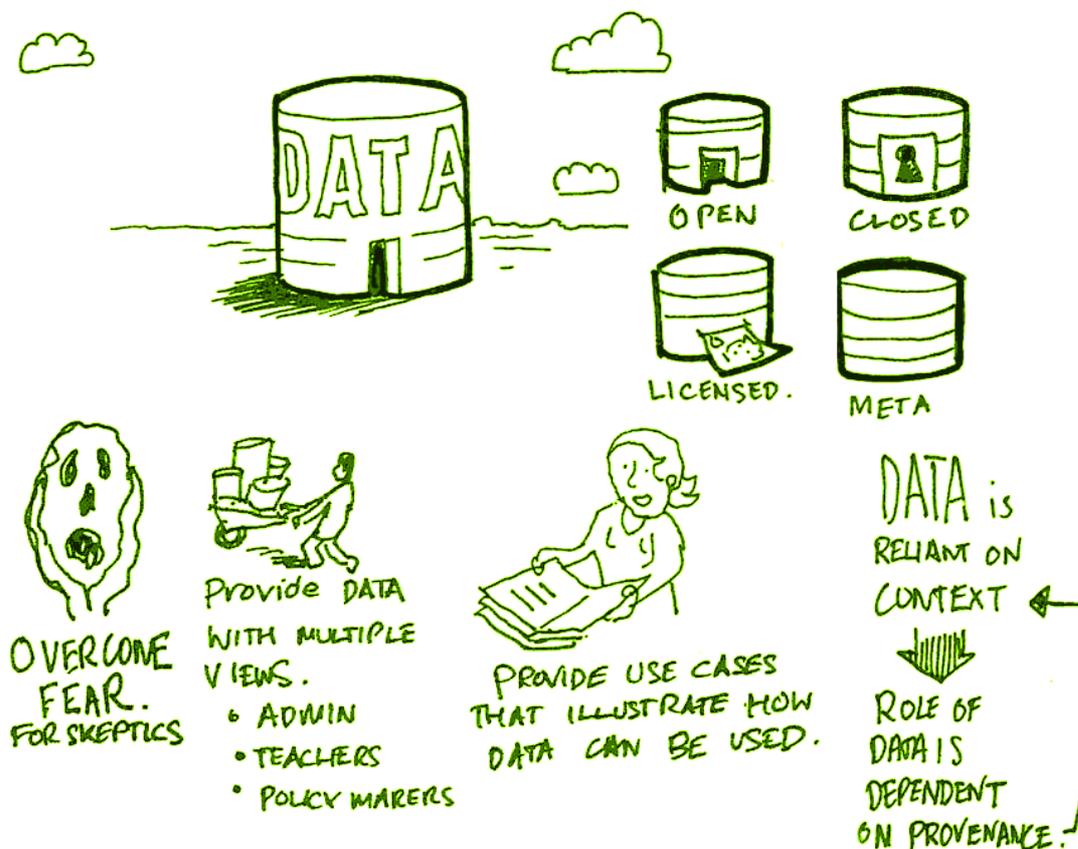
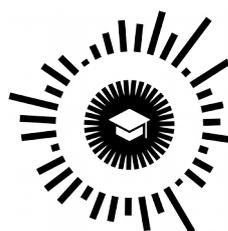


# Open Data as Open Educational Resources

## Case studies of emerging practice



Edited by Javiera Atenas & Leo Havemann



OPEN  
EDUCATION



## Open Data as Open Educational Resources

### Case studies of emerging practice

Edited by Javiera Atenas & Leo Havemann

With the support of Open Knowledge - Open Education Working Group and Universities UK

DOI: <http://dx.doi.org/10.6084/m9.figshare.1590031>

Cover image: Kevin Mears

Publishing layout and design: Santiago Martín

Cite as:

Atenas, J., & Havemann, L. (Eds.). (2015). *Open Data as Open Educational Resources: Case studies of emerging practice*. London: Open Knowledge, Open Education Working Group. <http://dx.doi.org/10.6084/m9.figshare.1590031>.



# Table of Contents

|  |    |
|--|----|
| The editors.....   | 4  |
| The scientific committee.....  | 5  |
| The authors.....   | 8  |
| Prefaces: Reflections from the scientific committee.....                                       | 12 |
| From Open Data to OER: An unexpected journey?.....   | 22 |
| A Scuola di OpenCoesione: From open data to civic engagement.....                              | 26 |
| Using Open Data as a Material for Introductory Programming<br>Assignments.....                 | 38 |
| Teaching Data Analysis in the Social Sciences: A case study with<br>article level metrics..... | 49 |
| The Alan Walks Wales Dataset: Quantified self and open data.....                               | 56 |
| Open Data for Sustainable Development: Knowledge society &<br>knowledge economy.....           | 67 |
| Acknowledgements.....  | 78 |
| Institutional Acknowledgements.....  | 81 |

# The editors

## Javiera Atenas

Is one of the co-coordinators of the Open Education Working Group of the Open Knowledge foundation and a researcher in Open Education. She holds a PhD in Education, a master's degree in Library and Information Science and an MPhil in Knowledge Management also, she is a fellow of the Higher Education Academy. She currently works co-ordinating distance learning programmes at University College London, UK and is an associate lecturer at Universitat de Barcelona at the department of Social Sciences. She is part of the board of advisors for the Lang OER programme, part of the panel of experts of the GO\_NG network, member of the UNESCO WSIS - OER community and collaborates with several open education projects at international level.

Contact: [j.atenas@ucl.ac.uk](mailto:j.atenas@ucl.ac.uk) - [@jatenas](https://twitter.com/jatenas)

## Leo Havemann

Is a Learning Technologist at Birkbeck, University of London, providing pedagogic and technical support for technology-enhanced learning at Birkbeck as well as working collaboratively with colleagues in the Bloomsbury Learning Environment consortium on cross-institutional TEL initiatives. He is a co-ordinator of the M25 Learning Technology Group (a practitioner-focused group) and of ELESIG London (a local branch of a wider community of learner experience researchers). His research interests include open educational practices, skills and literacies, blended learning, and technology-enhanced assessment and feedback. He has taught in HE in New Zealand and Australia, worked as a librarian in a London FE college, and worked in IT roles in the private sector. He has a Master's degree from the University of Waikato.

Contact: [l.havemann@bbk.ac.uk](mailto:l.havemann@bbk.ac.uk) - [@leohavemann](https://twitter.com/leohavemann)

# The scientific committee

## Marieke Guy

Is a data analyst at the Quality Assurance Agency for Higher Education (QAA). Prior to this role, she worked as a project manager for Open Knowledge and at the University of Bath. Marieke has been dealing with digital information for over 15 years and her projects include digital preservation, metadata standards, online learning, data visualisation, APIs, remote technologies and research data management. At Open Knowledge Marieke co-ordinated the Open Education working group and was involved with activities focusing on open data, open content, open licensing, open access and open GLAM.

Contact: [mariekeguy@hotmail.com](mailto:mariekeguy@hotmail.com) — [@mariekeguy](https://twitter.com/mariekeguy)

## William Hammonds

Is a public policy analyst and researcher. He developed UUK's work on open data and also manages UUK programmes on sector regulation and student experience. His previous reports include *Futures for Higher Education: analysing trends* and *Massive Open Online Courses: Higher education's digital moment?* Prior to UUK, Will has worked in a variety of policy, research and development roles in politics, local government and consultancy. He has worked with UK and European organisations ranging from Arts Council England to the European Commission. He is also a doctoral research student in government at the University of Essex.

Contact: [william.hammonds@universitiesuk.ac.uk](mailto:william.hammonds@universitiesuk.ac.uk) — [@wilamh](https://twitter.com/wilamh)

## Maria Perifanou

Is a project co-ordinator and a researcher in the field of Technology Enhanced Learning (TEL). She holds a Master degree in ICT and Foreign Language education from Ca' Foscari, University of Venice, Italy, and a PhD from the University of Athens, Greece, in the field of Applied Linguistics. Over that last ten years she has worked as a language teacher / trainer but has also collaborated with various European research institutions on several EU projects. Currently, she works at PAU Education (Spain) with the team of the Open Education Europa portal as TEL consultant and project coordinator. She also collaborates with the CONTA Lab, University of Macedonia (Greece) as well as with the Greek NGO Active Citizens Partnership (LangMOOC Erasmus+ project initiator & leading researcher). Her main practice and research concerns are Web 2.0, TEL, Open Education, MOOCs, OER, Computer Assisted Language Learning (CALL), WebQuests 2.0, Blended Learning, Collaborative Learning, Mobile Learning, Personal Learning Environments (PLEs) and more.

Contact: mariaperif@gmail.com — [@mariaperif](https://www.instagram.com/mariaperif)

## Ernesto Priego

Is a Lecturer in Library Science and Programme Director of the Library and Information Science Postgraduate Scheme at City University London. He is also the founder and editor-in-chief of *The Comics Grid: Journal of comics scholarship*, an open access journal founded in 2010, which is now published by the Open Library of Humanities. He enjoys fostering engagement and openness, facilitating collaboration and mentoring, and is passionate about is creating opportunities for discovery, creativity and innovation. He is an advocate for enhanced access to information as a public good and his main research interests revolve around library and information science, digital humanities, scholarly communications, comics and publishing scholarship, digital innovation and material culture, blogging and online journalism, intellectual rights and open science, data and education.

Contact: Ernesto.Priego.1@city.ac.uk — [@ernestopriego](https://www.instagram.com/ernestopriego)

## Anne-Christin Tannhäuser

Is a project coordinator in technology-enhanced learning and open education programmes and a consultant on educational innovation. She holds a Master's degree in Educational Sciences and Linguistics from the University of Leipzig and she was trained at the Max Planck-Institute for Human Development, Berlin, in the use of qualitative and quantitative research methodologies. In the past seven years she has managed and contributed to several TEL initiatives at national and European levels, including for the European Foundation for Quality in E-Learning, Cooperative State University Baden-Württemberg, Knowledge Information Centre Malta, Wikimedia Germany, University of Applied Sciences Ruhrwest, Linnaeus University and the Institute of Prospective Technological Studies (European Commission) in the field of open education, recognition of open learning and evaluation/communication of R&D projects. She coordinated the Open Access journal INNOQUAL, the International Journal for Innovation and Quality in Learning, for two years. She is also an associate researcher at the Berlin campus of ESCP Europe, a private business school with six locations in the EU.

Contact: [actannhauser@gmail.com](mailto:actannhauser@gmail.com) — [@actwww](https://www.instagram.com/actwww)

# The authors

## Juan Pablo Alperin

Is an Assistant Professor in the Publishing Program and a Research Associate with the Public Knowledge Project at Simon Fraser University. He is a multi-disciplinary scholar that uses computational techniques, surveys, and interviews to investigate ways of raising the scientific quality, global impact, and public use of scholarly work. He is an established scholar in both the open access and altmetrics (social media metrics) communities, having received numerous invitations to speak, and publish on these topics, both in North and Latin America, and has contributed a combination of conceptual, methodological, and empirical peer-reviewed articles and presentations, as well as edited two volumes, on issues of scholarly communications in developing regions.

Contact: [juan@alperin.ca](mailto:juan@alperin.ca) — [@juancommander](https://twitter.com/juancommander)

## Alessandra Bordini

Is a research assistant and graduate student in the Master of Publishing Program at Simon Fraser University, where her interest is in the intersection of digital media with the history of publishing. She holds an MA in Translation Studies from the University of Siena and worked in various editorial capacities with several publishing houses in Milan and Naples before moving to Vancouver. A published poetry translator and design enthusiast, she is an active member of CanaDiana, a non-profit organization devoted to cultural interchange between Canada and Italy.

Contact: [abordini@sfu.ca](mailto:abordini@sfu.ca) — [@agbordini](https://twitter.com/agbordini)

## Chiara Ciociola

Is the community manager of the project [A Scuola di OpenCoesione](#) at the Department for Cohesion Policies, Italian Presidency of the Council of Ministers. She holds a BA in Political Science, with a focus on New Media and Journalism at University of Florence and a MA in Digital Storytelling at University of Turin. In 2013 she founded [Monithon Italia](#), a civil society initiative for citizen monitoring of EU-funded projects. Since 2011 she is a contributor of [Neural magazine](#), a critical digital culture and new media arts magazine.

Contact [chiara.ciociola.ext@dps.gov.it](mailto:chiara.ciociola.ext@dps.gov.it) — [@chiaracio](#)

## Tim Coughlan

Tim is a lecturer in the Institute of Educational Technology at The Open University, UK. Previously, he was a Horizon Digital Economy Research Fellow and Lecturer in Computer Science at the University of Nottingham, UK. During that time, he created the introductory programming activities described in this book. With a background in Computer Science and Human-Computer Interaction, he researches the design of technologies and processes to support open and inclusive education, creative collaborations, and interpretation and interaction with data by learners and the wider public.

Contact: [tim.coughlan@open.ac.uk](mailto:tim.coughlan@open.ac.uk) — [@t1mc](#)

## Alan Dix

Is a Professor at the University of Birmingham, and Senior Researcher at Talis, Birmingham, and lives on a small Hebridean island, where he organises a biannual technology event Tiree Tech Wave. Alan's research interests are varied and eclectic; the majority of his work is focused around the area of human-computer interaction, and he is the author of a key international textbook in the area. In addition, he has spent time in agricultural engineering research, local government IT, submarine design, and dot.com start-ups. With a colleague at Lancaster University he invented technology to make LEDs individually controllable; this is being developed

commercially and is expected to transform cityscapes across the world. In 2013 he ran an open online HCI course now available on the open-resource site [interaction-design.org](http://interaction-design.org) and in the same year walked the perimeter of Wales, the data from which is the focus of this case study.

Contact: [alan@hcibook.com](mailto:alan@hcibook.com) — [@alanwalkswales](https://twitter.com/alanwalkswales)

## Geoffrey Ellis

Is a senior researcher in the Data Analysis and Visualisation group at The University of Konstanz, Germany. He received his Ph.D. in Computer Science from the University of Lancaster for his work on interactive visual clutter reduction through sampling. Prior to this, he was senior lecturer in computing at Huddersfield University, a freelance software developer, and research engineer at Manchester University. His expertise is in fields of information visualisation, human computer interaction, cognitive psychology and software development. His current interests include developing effective visualisation techniques, promoting user awareness of TOIs (things of interest) and mitigating cognitive biases. As well as supervising undergraduate and postgraduate project students, he is currently working on VALCRI, an EU funded R&D project developing a visual analytic-based platform for criminal analysts.

Contact: [ellis@dbvis.inf.uni-konstanz.de](mailto:ellis@dbvis.inf.uni-konstanz.de)

## Luigi Reggi

Is a public policy analyst with more than 10 years of international experience as a practitioner and researcher in the field of e-government and open government. Currently, he is a PhD student at the State University of New York at Albany and serves as an ICT policy analyst and open data specialist at the National Agency for Territorial Cohesion in Rome, Italy. He is a member of the Technical-Scientific Committee of the open government project "OpenCoesione" (Open Cohesion Policy). In 2013 he founded Monithon.it, a civil society initiative for citizen monitoring of EU-funded projects. Between 2003-2008, he worked as a public policy analyst and researcher at the National Center for IT in the Public Administration (now Agency for Digital Italy), the

Ministry of Economic Development, the University of Rome “La Sapienza” and the University of Urbino.

Contact: [luigi.reggi@gmail.com](mailto:luigi.reggi@gmail.com) — [@luigireggi](https://www.instagram.com/luigireggi)

## Katie Shamash

Is a scholarly communications analyst at JISC in the UK, currently supporting JISC's Open Access services by collecting, analysing, and presenting data, as well as by working with stakeholders to improve data quality and openness. She is also a Master of Publishing candidate at Simon Fraser University, Vancouver, researching academic publishing with a focus on humanities monographs.

Contact: [kshamash@sfu.ca](mailto:kshamash@sfu.ca) — [@katieshamash](https://www.instagram.com/katieshamash)

# Prefaces: Reflections from the scientific committee

---

**Marieke Guy, William Hammonds, Anne-Christin  
Tannhäuser, Maria Perifanou, Ernesto Priego**

---

## *A Note from the Editors*

*A collection like this is obviously not the work of its editors alone. It would, of course, have been impossible without the efforts and insights of the authors of the case studies included here. But we also wanted to acknowledge, and give a voice to, the esteemed group of colleagues who have acted as a scientific committee for this project. We invited this panel of international experts to help us shape these ideas and ensure the quality of this publication. The panel's role was to work alongside us in reviewing the proposals, selecting the cases, and supporting the authors in showcasing their educational practice. The members have diverse backgrounds and expertise but they all share an interest in, and commitment to, open educational practices. This enabled a distinctively collaborative and collegial approach.*

*Once the full draft cases were provided by the authors, the committee provided guidance via an open review model, allowing concerns, comments and feedback to be circulated as open discussions between reviewers, authors, and editors. For us, their contribution was invaluable, and we are sure the authors would join with us in providing our heartfelt gratitude. Below, the members of the panel offer their reflections on the open, collaborative process of creating this collection.*

---

## Marieke Guy

---

I am very excited to finally see the publication of *Open Data as Open Educational Resources: Case studies of emerging practice*.

This collection of case studies has been many months in the making. Javiera Atenas and Leo Havemann have been members of the [Open Knowledge Open Education Working Group](#), which I used to co-ordinate, since its inception in September 2013. They have been active members throughout: participating in events, such as our Open Education Handbook booksprint, and contributing to discussions on the mailing list and Twitter feed. In March this year they, along with Ernesto Priego from City University London, contributed a blog post on [The 21st Century's Raw Material: Using Open Data as Open Educational Resources](#), the post articulated many ideas and concepts the working group had been toying with (see the community session [What has open data got to do with education?](#) and the [Open Education data section](#) in the handbook. The authors argued that "by using real data from research developed at their own institution, multidisciplinary research projects enable opportunities to develop students' research and literacy skills and critical thinking skills by establishing ways for collaborations amongst students, researchers and academics". The post was both a reflection on the status quo and a call to action.

Three months after delivery of the post Leo and Javiera were guest speakers at the [7th Open Education Working Group call](#) (an online webinar/discussion session) which focused on Open Data as Open Educational Resources. The call also featured William Hammonds of Universities UK, who has been leading on the [Creating Value from Open Data project](#). During the call Javiera and Leo made a request for practitioners to "share academic practices and activities in the use of open data and to gather and discuss ideas that can be applied in HE, aiming to help students to develop a variety of skills, including data literacy, research methods, problem solving and citizenship skills". To solidify the objective of collecting these stories and the intention to publish them as an open e-book a [formal proposal](#) was published in mid-June.

The results of Javiera and Leo's efforts are five detailed and diverse case studies. They are cutting edge tales that show exciting efforts to try out something new, to

experiment and to learn from the results. These stories begin to touch on the potential of using Open Data as a teaching and learning resource. They show the possibilities for improvement of both digital and data literacy skills, and the benefits for student engagement and participation in activities with real world relevance. They also touch on the challenges use of open data poses: its complexity and flaws, access and format difficulties and resource implications. These case studies offer a starting point. I very much hope that other institutions and organisations will take the baton offered and begin their own experimentation and research.

---

## William Hammonds

---

My broader interests are in universities and how they work. As part of this I am interested in the role of open data in helping to support practical decision-making and innovation. I see great opportunities for universities and the sector as a whole to make better use of the data that they produce by publishing it openly. This includes the research process as well as the operational functions of a university, i.e. the services and support provided to students and staff. In addition there is a real opportunity for data skills to be embedded into the curricula of a range of subjects and the opportunity to apply these skills to real, open, data sets is particularly powerful.

The chapters in this book highlight the range of possibilities that exist within higher education to use open data from a variety of sources to help students develop a variety of relevant skills. These include practical skills associated with working with data, such as cleaning and licensing. In addition there is the skill of applying data to solve problems, make decisions and formulate conclusions. Furthermore the variety of data sets, ranging from research data produced by the same academic teams developing and delivering the lessons through to data from public bodies illustrates the breadth of data that can be used.

I also think that this book highlights the path for future development. Many of chapters, though by no means all, illustrates that examples of using open data as an Open Education Resource are more easily found in traditionally 'data literate' subjects such as computer science. The examples where other subjects, such as social science and design, have built open data working into their curricula are very powerful examples that show how this type of practice can be developed elsewhere. Being able to work with the growing volume and data is likely to be a core skill in the future, particularly in the context of greater automation.

This also illustrates the final challenge: supporting academics and teaching staff to develop their own skills and interests to use open data in this way. This includes the different types of institutional or departmental support. The majority of these examples are of committed or interested teams who already possessed the necessary skills. If there is a gap in the examples presented in this book it is probably how this

way of working can be communicated to others in ways that give them the develop elsewhere. But hopefully this book can be part of the process of highlighting the possibilities.

For me the task for a book like this is to support the authors to produce a chapter that is both useful and accessible for a reader. The main issues that I look for when doing this type of work primarily relates to identifying the main messages and trying to make sure that they are presented clearly. This involves a more developmental problem solving approach to support the author to develop their piece. In addition there needs to be care that suggestions are suggestions and do not try to distract from the findings and messages that an author wants to highlight. The initial filtering of abstracts both encourages this kind of approach by committing both the author but also the editors to the process.

---

## Anne Christin Tannhäuser

---

Javiera and Leo invited me to join the scientific committee because of my vast experience coordinating and working on Open Education projects, Open Educational Resources and Technology-Enhanced Learning at international level in the past years, but also, to provide guidance and expertise to the editorial team and scientific committee in the process of open peer review because of my experience as editor for the research publications at EFQUEL, as open peer review was at the heart of their research dissemination projects. I was excited to join them, because I grew curious on how educators are using available open data in their classrooms.

I am convinced that open data provides an excellent source for secondary and higher education teaching to tackle authentic research questions and help students to gain real knowledge from large data collections—a competence which is of much use given the vast amount of data we are exposed to nowadays—and most importantly share this knowledge with the community for more informed decision-taking.

One of the cases that called my attention *A Scuola di OpenCoesione: from open data to civic engagement* is an excellent example, relying on a clear methodology to guide young learners to engage with people outside school on topics that are relevant to their communities. I hope that the five case studies presented in this publication will inspire educators of different fields to work with open data in their classroom practice.

---

## Maria Perifanou

---

I'm really glad to participate in this inspiring publication entitled *Open Data as Open Educational Resources: Case studies of emerging practice* and I would like to thank warmly Javiera and Leo for this opportunity.

My work is strongly connected to the idea of openness, innovation, inclusion and quality in education. I am a member of the team of Open Education Europa portal, the gateway to European OERs. I see my work mostly as a passion, as an everyday challenge to facilitate a dialogue between learners, teachers and researchers who have the desire to learn and share ideas in an open learning environment embracing multiculturalism, multilingualism, openness and innovation. Two of the most recent initiatives of the OEE team are the release of the analytical list of European Repositories of Open Educational Resources (OER) as open data as well as the launch of European teacher's contest that aims to showcase every successful teaching practice across Europe!

I think that this book is highly connected with the idea of innovation and openness in teaching, and proves that educators can make a significant difference in a simple way. All what is needed is creativity, enthusiasm and of course open collaboration. Sharing is the heart of education. In this book, sharing experiences that describe the use of open data as OERs, is a powerful message to every educator who supports innovation and openness.

One excellent example of openness and innovation is the case of A Scuola di Open Coesione (ASOC) that offers a really useful service not only to students but to every Italian citizen. The OpenCoesione initiative offers to everyone open access to transparent and updated information on more than 900.000 funded project. The innovative idea of this project is the creation of the school of Open Coesione that is designed for secondary-school Italian students aiming to improve the use of public funds. This is without any doubt an example of an efficient use of open data and a step towards openness and democracy that can promote efficiency and effectiveness in government.

Another example of smart use of open data as OERs, is the case of using open data as a material with which someone can produce engaging challenges for students in order

to learn introductory programming (Python). In this case it has been shown that open data can be used as powerful material for designing learning activities. The innovation of this case lies on the creativity with which the open data has been used by educators and students producing unique and useful applications like a city-wide car park monitoring application and an e-reader application.

An interesting case study is also the Alan Walks Wales case that describes the educational use of an open dataset collected as part of a 1000-mile research walk. The innovative part in this case is that the data produced by the student (OpenStreetMap elevation data, sentiment analysis training data and more) is being fed back into the public domain and this could be very useful for other researchers in future who will be able to build on previous documented work. Another innovative characteristic is the flipped way to use open data.

By reviewing all these good practices, I got also interesting and useful information such as the tools proposed in the course “Using Open Data in Education for Sustainable Development” i.e. the infographic and interactive comparison tools provided by Education for All and World Inequality Database on Education.

I have also felt positively surprised by the power of open knowledge, like in the case of the “Teaching Data Analysis in the Social Sciences” course. In this course, it was found that when you use openly available data in the classroom this could give you a sense of belonging to a community, and a sense of pride in doing meaningful work that you want to share.

Regarding the process in which this book was developed, I would like to add that the process of open review could only be interpreted as a process of democratisation that only enrich our knowledge and open our minds.

---

## Ernesto Priego

---

This book, titled *Open Data as Open Educational Resources: Case studies of emerging practice*, is not a finished product in the sense it is not a closed output or the outcome of a project soon to be replaced by something else. This book is an open work in progress, and as such it is both the result of a process and a step in that process. Like all processes, there are here the traces of multiple signatures, both visible and invisible. It is a collective effort, initiated and spurred by the passion and determination of Javiera Atenas and Leo Havemann, who have co-ordinated a complex effort to make something important happen, out in the open, transparently and inclusively. This book is a digital object, and a series of digital objects at that. Each chapter or 'case study' here is both independent and part of something larger, not just this book, but the whole open education, OER, open data, open source and open access landscape. Each of these terms imply different, distinct phenomena, yet they are all interlinked.

A big driver for my own work is enhancing the potential of information to transform social realities through increased discoverability, access and reusability. The case studies that flesh out this book embody or ground different practices of openness in education. They do this not merely by describing these practices but by reflecting on them and by walking the walk as well as talking the talk; i.e. by sharing this work openly, making it available freely online and licensing it for reuse. Datasets and other work are fully referenced and most of them can be located and accessed by the reader, allowing them to try out the exercises presented.

The process of open peer review has been part of this effort to walk the walk of openness. We used basic, free tools and practised with an awareness of transparency and the need for pragmatic speediness, but also professionalism and respect. Authors and reviewers collaborated as equals, and followed advice promptly and generously. Everyone learned something, and a laborious task became a joy. This exercise in open peer review meant practising a different culture of collegiality, where hierarchies and power structures are interrogated by doing things differently, on a more levelled, open ground. Expertise becomes thus a performative, fluid act based on sharing, not imposing nor making scarce.

This book has the potential to be useful to many educators, students, academics, researchers, policy makers and members of the public interested in the possibilities of open data and the transformative power of OER. With a commitment to education and social transformation, information becomes knowledge, and knowledge wisdom. This is made possible through conscious and unconscious effort, labour, experience and collective generosity. What is needed is the will, the time, and especially the infrastructural affordances that enable individuals to build on the work of others. What Javiera and Leo started is an act of love, and an act of hard work too. Let's work hard ourselves so their effort keeps growing and delivering, beyond our own contexts and interests.

# From Open Data to OER: An unexpected journey?

---

**Javiera Atenas & Leo Havemann**

---

This collection presents the stories of our contributors' experiences and insights, in order to demonstrate the enormous potential for openly-licensed and accessible datasets (Open Data) to be used as Open Educational Resources (OER). Open Data is an umbrella term describing openly-licensed, interoperable, and reusable datasets which have been created and made available to the public by national or local governments, academic researchers, or other organisations. These datasets can be accessed, used and shared without restrictions other than attribution of the intellectual property of their creators<sup>1</sup>. While there are various definitions of OER, these are generally understood as openly-licensed digital resources that can be used in teaching and learning<sup>2</sup>.

On the basis of these definitions, it is reasonable to assert that while Open Data is not always OER, it certainly becomes OER when used within pedagogical contexts. Yet while the question may appear already settled at the level of definition, the potential and actual pedagogical uses of Open Data appear to have been under-discussed. As open education researchers who take a wider interest in the various open 'movements', we have observed that linkages between them are not always strong, in spite of shared and interconnecting values. So, Open Data tends to be discussed primarily in relation to its production, storage, licensing and accessibility, but less often in relation to its practical subsequent uses. And, in spite of widespread

---

1. See, eg., *Open Data Handbook: 'What is Open Data?'* (Open Knowledge). Available at [opendatahandbook.org/guide/en/what-is-open-data](https://opendatahandbook.org/guide/en/what-is-open-data).

2. See, eg., *What are Open Educational Resources (OERs)?* (UNESCO). Available at [www.unesco.org/new/en/communication-and-information/access-to-knowledge/open-educational-resources/what-are-open-educational-resources-oers](https://www.unesco.org/new/en/communication-and-information/access-to-knowledge/open-educational-resources/what-are-open-educational-resources-oers).

understanding that use of the term 'OER' is actually context-dependent, and, therefore, could be almost all-encompassing, the focus of OER practice and research has tended to be on educator-produced learning materials. The search for relevant research literature in the early stages of this project turned up sources which discuss the benefits of opening data, and others advocating improving student engagement with data<sup>3</sup>, but on the topic of Open Data as an educational resource specifically, there appeared to be something of a gap.

We should pause here and stress that we are attempting to characterise general tendencies of the OER and Open Data movements; we acknowledge that certain voices have nonetheless been speaking across the divides created by these '[open silos](#)' and our purpose in this collection is really to add to and amplify this important work. One document to tackle the interface between Open Data and OER explicitly is the collaboratively-authored [Open Education Handbook](#) from Open Knowledge, which discusses Open Education Data as both openly available data that can be used for educational purposes, and open data that is released by education institutions. Our focus in this collection tends to the former understanding: the use of Open Data in education, rather than about education (though in either case the data could inform interesting student projects). Another noteworthy source on the topic is the set of resources provided by [School of Data](#), an Open Knowledge-related project that provides guidance and support for students, educators and researchers working with Open Data in educational contexts. The School of Data provides [online self-access courses](#), and also guidelines for understanding Open Data in the form of a handbook, and for the adoption of software for [data analysis](#). Additionally, the resources from School of Data have been translated into [Spanish](#), [French](#) and [Portuguese](#), in order to assist significant numbers of educators internationally.

Open Data has been highlighted as a key to information transparency and scientific advancement. Students who are exposed to the use of Open Data have access to the same raw materials that scientists and policy makers use. This enables them to engage with real problems at both local and global levels. Educators who make use of Open Data in teaching and learning encourage students to think as researchers, as journalists, as scientists, and as policy makers and activists. They also provide a meaningful context for gaining experience in research workflows and processes, as well as learning good practices in data management, analysis and reporting. The

3. See, eg., *Making the most of data: Data skills training in English universities* (Universities UK, 2015). Available at [http://www.universitiesuk.ac.uk/highereducation/Pages/MakingthemostofdataDataskillstraininginEnglishuniversities.aspx#.Vi-pf9\\_tIBd](http://www.universitiesuk.ac.uk/highereducation/Pages/MakingthemostofdataDataskillstraininginEnglishuniversities.aspx#.Vi-pf9_tIBd).

pedagogic deployment of Open Data as OER thus supports the development of critical, analytical, collaborative and citizenship skills, and has enormous potential to generate new knowledge.

Our unexpected journey in producing this collection has been a series of steps, sometimes taken while not quite seeing the path ahead. We began with a question: was anyone actually using Open Data as OER? and sought responses using a [blog post and exploratory survey](#), which generated some initial feedback. Despite efforts to promote the survey widely, the response rate was relatively low. While all those who responded said that they were indeed making use of Open Data in their teaching practice, it became clear from the examples given that only about half of the respondents were actually using Open Data, rather than other forms of open content, such as OER, or datasets that were not actually 'open' as such. This indicated to us that there may be a lack of awareness and understanding of Open Data among educators.

It occurred to us at this point that we should focus our attention on two distinct projects. On the one hand, with Ernesto Priego, we continued to develop our ideas regarding the benefits and challenges of bringing Open Data into teaching and learning. This work became a research paper entitled '[Open Data as Open Educational Resources: Towards Transversal Skills and Global Citizenship](#)'. Here we focused on research- and scenario- based learning models, considering the application of these to open datasets as a way of developing students' transversal skills and twenty-first century literacies, alongside more traditional subject-related competencies. At the same time, we remained convinced that the question of whether educators were actually already training students in the interrogation of Open Data had surely not yet been settled by our survey. In fact, we suspected that good practices must be out there, and might be captured in a useful and approachable way through a call for case studies. The response and the results were gratifying.

And so, finally, at the core of this book we have five case studies of open educational practice, in which Open Data takes centre stage in learning. These case studies showcase diverse practice at an international level, including studies from the UK, Canada and Italy. Different approaches and disciplines are represented, reflecting different kinds of data, across different educational levels, to highlight the development of citizen engagement. The cases are more proof-of-concept than 'how-to', and we prefer to let them speak for themselves rather than describe them in detail

here. We simply wish to express our delight at the range of innovative work that has been surfaced by the authors.

By extending their commitment to open practice beyond the classroom through the act of sharing these accounts with others, our contributors have produced a set of open resources to inform, assist and inspire other educators. Our collective hope is that through the work of opening up practices as well as datasets, we are starting to create the conditions of possibility for the reuse and remixing of pedagogical ideas and learning designs around Open Data. We contend that opening content and practices is a necessary step, but one which must be taken with a higher purpose in mind. Open is not the destination, it is only the beginning.

# 1 A Scuola di OpenCoesione: From open data to civic engagement

---

**Chiara Ciociola & Luigi Reggi**  
**A Scuola di OpenCoesione, Italy**

---

**Keywords:** *Civic monitoring; citizen engagement; European policy; high schools; open spending data.*

## 1. Context and task description

Where does our money go? Who are the beneficiaries of public funding? Which projects get funded? When and how do publicly funded projects deliver concrete results? Are they effective enough? The Scuola di OpenCoesione (ASOC) is an educational challenge and a MOOC (Massive Online Open Course), designed for Italian high-school students.

It has three main objectives. First, to find out how public money is spent in a given local area or neighbourhood; second, to follow the projects and investigate how they are progressing and what challenges are they facing and third, to involve local communities in monitoring the effectiveness of public investment.

ASOC was launched in 2013 as a flagship initiative of OpenCoesione<sup>1</sup>, the open government strategy of the Italian Government on Cohesion Policy<sup>2</sup>. The ASOC team is composed of professionals with different backgrounds and skills (including Cohesion Policy experts, a community manager, a data journalist, a communications expert, training and education experts and a developer) and is actively supported by the members of the OpenCoesione team as well as by innovation experts from the Ministry of Education.

On [OpenCoesione data portal](#), anyone can find transparent information on each funded project. Open data from more than 900,000 projects has been posted so far, the data is updated every two months and it can be freely re-used and explored interactively on the portal using maps, filters and detailed data sheets on projects and recipients.

A Scuola di OpenCoesione builds on OpenCoesione data portal to actively promote the use and reuse of data as a basis for the development of civic awareness and engagement. The application of Open Data to real-life public interventions can stimulate the creation of a “monitorial citizenship” (Schudson M., 1998; Keane J., 2009). Indeed, ASOC is mainly about civics and civic education, and complements the standard high school curriculum.

---

1. ASOC was launched thanks to a partnership between the Department for Economic Development and Cohesion (now Agency for Territorial Cohesion and Department for Cohesion Policies at the Italian Presidency of the Council of Ministers) and the Ministry of Education. The project also involves the European Commission’s network of information centres, “Europe Direct”. Additional information on the team and the institutions involved can be found on the ASOC website: <http://www.ascuoladiOpenCoesione.it/team/>.

2. Cohesion policies aim at improving economic well-being in regional and local contexts, by transforming regional disparities into opportunities for growth and development. European Cohesion Policy is financed by EU Structural and Investment Funds in order to promote economic and social development in the regions lagging behind in terms of growth, most of which are located in Southern and Eastern Europe. In Italy, national policies for cohesion provide additional financial resources, concentrated in the Italian *mezzogiorno* (Southern Italy). In Italy, the overall 2007-13 cohesion budget, made up of national and European funds, amounted to approximately 100 billion Euros over a 7-year period, which on a yearly basis accounted for as much as 1% of GDP. Italy is a net contributor to the EU budget and the third biggest beneficiary country of EU structural funds (after Poland and Spain). However, Italy is also among the countries with the lowest absorption rates of the funds, and issues about misuse of the funding have been raised. This explains why cohesion policy is a particularly hot topic in Italy.



Figure 1: Participant schools

The ASOC course is organised in six main sessions, the first four consist of classes focused respectively on public policies, open data analysis, data journalism and “citizen monitoring” of public funding. The fifth involves on-site visits to the selected publicly funded projects and interviews with key players involved in the projects’ implementation, while the sixth is a final event where students meet local communities and policy-makers to discuss the findings of their investigations.

The starting point of the course is the selection of a project – based on open data from OpenCoesione – a large infrastructure project for instance, located in the local area and funded by Cohesion policies. The main criteria for project selection includes the perceived importance of the project within the local community as this can be measured, for example, by the number of local news stories covering any debate surrounding the expected results or the existing delays.

As a second step, students use quantitative and qualitative research methods to deal with data, to locate open data sources and to investigate the local socio-economic context where the selected project is operating. They create descriptive statistics,

infographics and data mash-ups to represent their territory in new, meaningful ways, and to compare the local context to other similar areas. Students also gain access to different administrative sources in order to analyse the policy objectives and administrative history of the project.

Finally, once they have gathered all available information about their projects, students develop original interview guides to conduct a “citizen monitoring” visit to the project site. They employ the methodology developed by the Italian civil society initiative “Monithon” to collect comparable, qualitative data about the progress and its results<sup>3</sup>. All the information is then published on the Monithon.it platform as a “citizen monitoring report”, including all evidence on the project’s results, which is disseminated amongst the local communities involved during the final event. The aim of the event is to stimulate an informed debate, involving the local communities and the local administrations responsible for financing and implementing the project, about the use or misuse of public funding.

Teaching activities are organized around the development of the students’ research projects and reproduce a flipped classroom setting, combining asynchronous content from the MOOC with group activities in class and online interaction with the central staff. In-class sessions share a common structure. Each class starts with one or more videos from the MOOC, followed by a group exercise.

Focused on real-life civic issues, ASOC forces the students to come up with original solutions to local challenges. Students are stimulated to acquire different skills, from working as part of a team to specific technical abilities such as analysing data or developing multimedia content.

ASOC staff throughout a series of seminars conducts and supports teacher training, and in order to have an impact on local communities and institutions, students are

---

3. The project uses the Monithon initiative’s tools and methodologies to take action and report malpractice, but also to play a role in making all these projects work, accelerating their completion and understanding whether they actually respond to local needs (Buttiglione, P. L.; Reggi, L. 2015). Monithon (literally “monitoring marathon”) and promotes the citizen monitoring of cohesion policy through the active involvement of communities and a shared methodology. The Monithon idea, initially conceived by the OpenCoesione team, was presented to the “civic hacking” community and soon transformed into an independent civil society initiative run by an enthusiastic group of developers, activists and journalists, based on a web monitoring platform operational since 2013. Within the project, Monithon’s expertise lies in monitoring not only of public funds in general, but of cohesion funds specifically. Through “monitoring marathons”, groups of citizens – sometimes under the guidance of local CSOs – embark on investigations within an area in order to gather information on specific projects of interest. By doing so, participants pool together useful material with a view to evaluating the effectiveness of public spending, and apply bottom-up methods of exercising control over public policies. The OpenCoesione-Monithon partnership reached 4th place and received the silver award at the 2014 Open Government Awards for citizen engagement.

actively supported by local associations that can contribute specific expertise in the field of open data or on other specific topics (environmental issues, anti Mafia activities, local transportation, etc.). Furthermore, the European Commission’s network of information centres, “Europe Direct” (EDIC), is involved in providing support for the activities and disseminating the results<sup>4</sup>.

The ASOC central staff systematically collects data from the teachers, the students and from local associations by means of targeted surveys<sup>5</sup>.



Figure 2: Route map

## 2. Reflection

### Initial overall results

Following the first edition of ASOC, carried out on an experimental basis in 7 pilot institutes during the 2013–2014 school year, the method was strengthened and consolidated. As a result, the total number of schools selected for the 2014–2015 school year rose to 86. The student groups formed have produced results well above expectations. The analyses and studies conducted have been effective and had a strong impact, stimulating debate at the local level, as well as dialogue between the local communities and the institutions involved.

4. During the last edition 86 schools have participated in the course. 26% of the teams were supported by both an EDIC center and an association; 33% were supported only by a EDIC; 15% by only one association.

5. Two surveys are conducted among the teachers. The first survey is to assess the characteristics of the selected schools (e.g. the teachers’ digital skills, the number of team members and their different characteristics, etc.) at the beginning of the school year. The second survey, conducted upon conclusion of the ASOC project, assesses the results.

Of the initial 97 teams, 45 completed their projects. 35 of these did so within the established time frame, allowing them to participate in the final competition. 34 of the groups completed some of the steps in the program but not all of the required activities. 18 of the selected schools, however, failed to begin the program. This was due to difficulties complying with the time frame recommended for the research, which proved to be especially difficult to respect while also keeping up with the standard curriculum – a result of the project being launched too late in the school year.

## Group work and students' motivation and interest in the project

The fact that the activities are organised in the form of a project, with ambitious goals and well-defined roles similar to those found in a newsroom<sup>6</sup>, is the aspect of ASOC that both teachers and students<sup>7</sup> most appreciate. Students are encouraged to constantly exercise interdisciplinary and soft skills, such as teamwork and respecting deadlines, as well as more specific and technical skills, like working with statistics and developing multimedia content.

The teaching method combines periods of asynchronous learning, typical of MOOCs (Massive Online Open Courses), with teacher-guided classroom activities, group exercises and online interaction with the ASOC team behind the project. Between lessons, students work independently to prepare data analysis reports and their original final project. There is also a blog dedicated to share and disseminate the students' activities on social networks.

The greatest difficulty lies in smoothly integrating these activities into the standard curriculum, since it has proven to be much more time-consuming for schools than expected. In fact, during the most recent edition, the overall hours students spent, including both classwork and independent study time, duplicated the amount of hours initially determined. This demonstrates a high level of motivation and enthusiasm for the proposed activities, but it also creates a need for better balance with respect to the implementation of standard curriculum programs.

---

6. The students participate in a "Data Expedition", which is adapted from the original model developed by Open Knowledge (see <http://schoolofdata.org/data-expeditions/>). During this exercise, the students divide into groups with specific roles: project manager; coder; analyst; social media manager; storyteller; blogger; head of research (see <http://www.slideshare.net/ascuoladioc/i-ruoli-di-a-scuola-di-OpenCoesione>).

7. 82% of the teacher has expressed favourable judgement about the efficiency of the working groups. 74% has positively evaluated the attribution of roles and the level of inclusion of the students in the activities. Also, 76% of the teachers has expressed a very positive opinion about the interaction among the students.

In the next edition, the ASOC team will reduce the overall scope of the course by selecting core contents such as data analysis exercises and civic monitoring activities. The aim is also to make the homework more focused on core data analysis and less on writing detailed stories on what the students have learned. Thanks to these changes, hopefully the teachers will be free to develop additional content according to their preferred subject and specific educational needs.

## Variety and freedom to choose how to use data

The chance to gain familiarity and interact with available open data is one of the key aspects that especially stimulated students' curiosity. Not only did students conduct critical analyses of previously processed data, but nearly all of the teams also carried out quantitative analyses and visualizations, using data not only from OpenCoesione, but also from other public portals (such as the Italian National Institute of Statistics), to assemble socio-economic statistics<sup>8</sup>.



Figure 3: Activity board

8. See the following examples of data analysis that the students have conducted:

- <http://www.ascuoladiOpenCoesione.it/i-dati-ancorati-negli-abissi-del-liri/>
- <http://www.ascuoladiOpenCoesione.it/il-dato-e-tratto/>
- <http://www.ascuoladiOpenCoesione.it/iii-lezione-analizzare-iiss-giulio-cesare/>
- <http://www.ascuoladiOpenCoesione.it/tell-me-more-about-open-data/>
- <http://www.ascuoladiOpenCoesione.it/terza-lezione-analizzare-liceo-scientifico-la-cava/>

Other teams worked on processing qualitative data, for instance by using text analysis to evaluate the recurrence of keywords in collected research materials<sup>9</sup>.

Other students came face to face with new challenges, such as the insufficiency or lack of open data at the local level, especially in small towns lacking actual “open data portals”<sup>10</sup>. They rose to meet these challenges with creative and unexpected solutions. In cases such as these, a number of teams constructed new datasets from secondary information collected on the web or from printed materials<sup>11</sup>. Others submitted FOIA requests to local administrations<sup>12</sup>.

## Civic monitoring and the active involvement of local authorities

For most of the teams, meeting with the institutions responsible for planning or implementing the projects they are researching represents a crucial moment in the program. The data show that over 70% of groups that complete the program involve the competent institutions in their project activities.

The students interview representatives of the local organizations in charge of planning or implementing the projects in question, gathering more detailed information about their positive aspects and asking for explanations regarding any issues that have emerged. Local authorities are generally especially open with students, welcoming and guiding them, organizing meetings where they can gather information from the technical departments responsible for the selected projects and providing documents that are useful for their research. In many cases, data and statements collected in this fashion are then verified by means of fact-checking and investigative journalism.

All of these activities lead up to the preparation of “civic monitoring reports” which are posted on the [monithon.it](http://monithon.it) website. These include a “brief assessment” of the project,

---

9. Content Analysis: Acropolis E-Team Alatri (see, [http://prezi.com/ya4ynbelgpmo/?utm\\_campaign=share&utm\\_medium=copy&rc=ex0share](http://prezi.com/ya4ynbelgpmo/?utm_campaign=share&utm_medium=copy&rc=ex0share)).

10. Only 24% of the top 100 Italian municipalities publish local data as open data (Between, 2014).

11. Map of the Castles, [https://a.tiles.mapbox.com/v4/soniamurruni.lm6504jg/page.html?access\\_token=pk.eyJ1Ijoic29uaWFtdXJydW5pliwiYSI6ImJvRlF0ZGMifQ.wokEt0Yv5nEvQ3YusUrWKw#10/38.2603/15.9233](https://a.tiles.mapbox.com/v4/soniamurruni.lm6504jg/page.html?access_token=pk.eyJ1Ijoic29uaWFtdXJydW5pliwiYSI6ImJvRlF0ZGMifQ.wokEt0Yv5nEvQ3YusUrWKw#10/38.2603/15.9233)

12. In 2014, the Stanislao Cannizzaro high school in Palermo requested the publication of local open data on local transportation. The local government responded and published the dataset on its website (see [http://www.comune.palermo.it/opendata\\_dld.php?id=318#.VFyhLvSG-hM](http://www.comune.palermo.it/opendata_dld.php?id=318#.VFyhLvSG-hM)). In 2015, the Veronica Gambarà high school in Brescia added new items to a public dataset on local bike lanes. The Lombardy Region then published this improved version of the dataset on its open data portal (see <https://www.dati.lombardia.it/Turismo/Ciclabili-Progetto-Garda-Bresciano/e6qe-v3v4>).

as well as detailed descriptions of activities, content from interviews, videos the students have prepared and their recommendations. The teams that reached the “Top 10” in the most recent edition focused mostly on problematic projects (out of the 10 funded projects investigated, 4 had encountered difficulties and 3 were “blocked”). Meanwhile, the groups that completed the program had chosen to research projects that were found not to have any serious issues. The most frequently researched project types were those in the spheres of culture, tourism, the environment, transportation, and energy.

## The use of complex communication techniques to bolster involvement

Extremely effective communication campaigns were created to disseminate the students' results, aiming at getting the local community actively involved. These campaigns demonstrate considerable mastery of technical tools (creation of animations and 3D representations, the use of editing software and compositing skills) and narrative tools (the use of complex storytelling and narrative techniques)<sup>13</sup>. In 40% of cases, local media were actively brought in to disseminate the results of the studies.

However, the students focused the narrative of their project on the emotional aspects of working together, at the expense of a more in-depth description of the development and progress of their research. This self-referential manner of storytelling was sometimes reflected in the articles published in newspapers and in local TV news stories, which focused more on aspects of the ASOC method and program than on the students' “discoveries” regarding the local area and the use of public funds.

The final events, where the students' results were presented, incorporated numerous initiatives to raise community awareness. Some of these took the form of information dissemination (pavilions on the streets, distribution of flyers, participation in public

---

13. Some examples:

- Liceo Scientifico Peano Pellico, Cuneo: <http://liceocuneo.it/oc1e>
- Liceo Bertarelli Ferraris, Milano: <https://youtu.be/9AR6YmiTTLQ>
- Liceo V. Gambara, Brescia: <http://www.powtoon.com/embed/dPzsSkI3TuJ/>
- Istituto E. Medi, Galatone: <https://youtu.be/lC29Z66flPg>
- Liceo G. Giordani, Monte Sant'Angelo : <https://youtu.be/wdk7Ul8372M>
- Liceo G. Asproni, Nuoro: [https://youtu.be/8Z6SOV1q\\_BQ](https://youtu.be/8Z6SOV1q_BQ)

events), but others focused on listening and gathering opinions (drafting and collecting questionnaires targeting both the community at large and those most affected by the selected topics).

The turnout on the part of institutional representatives was low in some cases, where those who had been invited preferred not to deal with either the issue of the negative aspects of funding or with students' suggestions. There were few cases in which an institutional presence was completely lacking. These occurred in particularly critical situations (e.g. municipalities under extraordinary administration due to Mafia activity). We believe, however, that the initiatives where this occurred were especially important in terms of raising students' awareness and helping them to develop a critical sense.

### 3. Recommendations for good practice

#### Studying to gain independent problem solving skills

ASOC's working method is centered around specific goals, well-defined roles and decision-making. This has allowed students to independently manage every aspect of their project activities, from the choice of research methods to how to disseminate the results. After having retraced the genesis and the status of implementation of the projects, students identify problems, ask questions and propose ways in which to improve the use of public funds.

#### Employing (hyper)local data —with the greatest possible granularity

Investigating local and concrete issues, starting with information on individual projects published by OpenCoesione, is a strong motivator for students. Being able to precisely assess the status of infrastructure and other types of funded projects can act as a catalyst for the creation of surprising and creative projects, capable of effectively

translating analyses of administrative and statistical data into high-impact stories and formats.

Thanks to accurate and detailed data on hundreds of thousands of individual projects funded<sup>14</sup>, it is possible to offer practical suggestions about the projects that have been undertaken, exploiting collective intelligence and new modes of public participation. By getting local community networks structurally involved in the process, ASOC aims to connect players who are already active in the field of local policy monitoring. As the students' interest is aroused, so they arouse the interest of others, in a reciprocal process of "discovery", civic education and community building.

## Seeing it with their own eyes

Thanks to the chance to see the projects undertaken with cohesion resources with their own eyes, students became active citizens. On-site visit is often worth more than dozens of theoretical lessons. Additionally, the chance to get the students involved in the places, activities and phenomena whose data they have studied, makes them active and involved participants in the topics they have explored through open data. Indeed, teachers describe the monitoring trip as the "most engaging" moment in the program, because it provides students with a chance to compare the data they have collected with the facts on the ground.

## Building stories

Data analysis alone is not sufficient to stimulate student participation. It is also necessary to ensure they exercise skills such as storytelling and communication, the use of social media, the ability to create a dialogue and work together with local authorities (knowing how to speak multiple 'languages', ranging from administrative to journalistic). At ASOC, data are just a starting point for the development of original representations and different data mashups, leading to the collection of primary data and becoming a catalyst for informed civic debate.

---

14. Art. 115 of the EU Regulation no. 1303/2013 forces the all European Managing Authorities of EU Structural and Investment Funds to publish highly detailed data on the beneficiaries and the projects as open data.

## Ongoing monitoring

It is particularly useful to encourage ongoing civic monitoring activities, even beyond the conclusion of the students' project activities, as they provide continuing beneficial oversight and act as a constant reminder to the institutions involved. It is therefore crucial that the ASOC program be integrated into the standard educational curriculum. Ongoing monitoring over time, as part of an education in civics, could result in a level of engagement with an even greater impact.

## References

Between (2014). *Smart city index*. Retrieved October 11, 2015. Retrieved from <http://www.between.it/ita/smart-city-index.php>

Buttiglione, P. L.; Reggi, L. (2015). Il monitoraggio civico delle politiche di coesione e lo sviluppo di comunità civiche. *PRISMA Economia - Società - Lavoro*, Vol. 1. Retrieved from <https://mpra.ub.uni-muenchen.de/62101>

Schudson, M. (1998). *The good citizen: A history of American civic life*. New York: Martin Kessler Books.

Keane, J. (2009) Monitory democracy and media-saturated societies. *Griffith Review*, 24. Retrieved from <http://griffithreview.com/articles/monitory-democracy-and-media-saturated-societies/>

## Acknowledgements

The authors would like to thank Carlo Amati, William Hammonds, Damien Lanfrey and Donatella Solda for the valuable comments and suggestions.

The authors are also grateful to Jennifer Delare for her help in translating and proofreading a previous draft of this article.

# 2 Using Open Data as a Material for Introductory Programming Assignments

---

**Tim Coughlan**  
**Institute of Educational Technology,**  
**The Open University, UK**

---

**Keywords:** *Open Data; Programming, Networks; Files; Coursework; Code; Python.*

## 1. Context and task description

This case study explores why and how open data can be used as a material with which to produce engaging challenges for students as they are introduced to programming. Through describing the process of producing the assignments, and learner responses to them, we suggest that open data is a powerful material for designing learning activities because of its qualities of ease of access and authenticity. We conclude by outlining steps to take in devising and implementing open data-based assignments.

In two successive years, forms of open data were used to construct coursework assignments for postgraduate students at the University of Nottingham, UK. The rationale for using open data was to shift the focus towards an outward-looking

approach to coding with networks, files and data structures, and to engage students in constructing applications that had real-world relevance.

The students were studying for a postgraduate qualification in [computer science](#), without having previously studied computing —for example, studying [human computer interaction](#), but coming from a psychology background. As such, the course assumed no prior knowledge, and aimed to provide a basic conceptual and practical grounding in software development. Approximately 50 students took the class each year, and contact time was a mixture of one-hour lectures and two-hour computer lab sessions. The lab sessions were initially used to complete small exercises with formative feedback only, but as the coursework deadlines approached, these became a space for questions, mentoring and peer discussion of progress, supported by the lecturer and teaching assistants.

Python was chosen as the programming language. It has grown in popularity as a means to introduce programming (e.g. see Guo et. al. 2014) due to a low entry barrier to create working code, high flexibility, and also wide used in industry. It includes relatively simple functions for accessing and manipulating data from files and networks, so the barriers to creating applications with open data are relatively low.

While lab computers were available, students were encouraged to install Python on their own laptops if possible. The aim of this was that they would become comfortable with programming outside of the lab setting. In general this worked well, but it was important to specify that students all install the same version of Python as that available in the labs, and to clarify some platform-specific differences.

The assignment in the first year utilised e-book text files from Project Gutenberg<sup>1</sup>, and required students to build an e-reader application. In the next year, car park status data, which was made available in a regularly updated form by the city council through their open data initiative<sup>2</sup> was used as the basis for an assignment in which students developed a city-wide car park monitoring application.

Students were given a set of requirements that progressed from the basic creation of the application, towards challenging them to make use of the data in complex ways. A short description of each assignment follows.

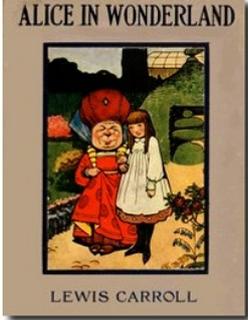
---

1. <https://www.gutenberg.org/>.

2. <http://www.opendatanottingham.org.uk/>.

Project Gutenberg > [50,180 free ebooks](#) > [32 by Lewis Carroll](#)

### Alice's Adventures in Wonderland by Lewis Carroll



ALICE IN WONDERLAND  
LEWIS CARROLL

[f](#)
[+](#)
[t](#)
[p](#)



Download
Bibrec

**Download This eBook**

| Format <span>?</span>                       | Size   | <span>?</span> | <span>?</span> | <span>?</span> |
|---|--------|----------------|----------------|----------------|
| <a href="#">Read this book online: HTML</a> | 90 kB  |                |                |                |
| <a href="#">EPUB (with images)</a>          | 536 kB |                |                |                |
| <a href="#">EPUB (no images)</a>            | 71 kB  |                |                |                |
| <a href="#">Kindle (with images)</a>        | 1.1 MB |                |                |                |
| <a href="#">Kindle (no images)</a>          | 238 kB |                |                |                |
| <a href="#">Plain Text UTF-8</a>            | 73 kB  |                |                |                |
| <a href="#">More Files...</a>               |        |                |                |                |

Figure 4: Free books on the Project Gutenberg website: <https://www.gutenberg.org/>.

would be of very little use without my shoulders. Oh, how I wish I could shut up like a telescope! I think I could, if I only know how to begin.' For, you see, so many out-of-the-way things had happened lately, that Alice had begun to think that very few things indeed were really impossible.

There seemed to be no use in waiting by the little door, so she went back to the table, half hoping she might find another key on it, or at any rate a book of rules for shutting people up like telescopes: this time she found a little bottle on it, (which certainly was not here before,' said Alice,) and round the neck of the bottle was a paper label, with the words 'DRINK ME' beautifully printed on it in large letters.

It was all very well to say 'Drink me,' but the wise little Alice was not going to do THAT in a hurry. 'No, I'll look first,' she said, 'and see whether it's marked "poison" or not'; for she had read several nice little histories about children who had got burnt, and eaten up by wild beasts and other unpleasant things, all because they WOULD not remember the simple rules their friends had taught them: such as, that a red-hot poker will burn you if you hold it too long; and that if you cut your

pg11.txt  
pg76.txt

6

Figure 5: An example application of the type that students produced for the e-reader coursework.

The E-Reader assignment required students to combine their knowledge to produce their most complex piece of work, after the guided production of many small programs and a smaller assignment. Students needed to demonstrate their ability across areas including the manipulation of data and files, and the production of a complex graphical user interface.

The basic requirements of the assignment matched the fundamentals of an E-reader, such as switching between different books, moving between pages, and changing the font size. Advanced requirements challenged the students to show their understanding of creating algorithms, and the large volumes of textual lent themselves to this. For example one requirement asked students to create functionality that displayed a list of the most common words in a particular book.

The screenshot shows the Open Data Nottingham website. At the top left is the logo 'Open Data Nottingham'. To the right is the Nottingham City Council logo and a search bar with the text 'Search for Data...' and a 'Search' button. Below the logo is a navigation menu with links: Home, Data, Publication Scheme, Licensing, About, Interact, and More. The main content area features a large image of a building facade. Below the image is the title 'Car park occupancy' and a subtitle 'Live car parking data for sites within the Nottingham ParkSmart system.' A 'Technical description' section follows, explaining that the dataset is a live dump from the internal car park management system (Parksmart) and may contain duplicate entries during maintenance. To the right of the technical description is a 'Downloads' section with three rows: CSV, JSON, and XML, each with a download icon and the text 'Car Park Occupancy'. Below the downloads is a 'Filetypes' section with the text 'More information on these filetypes is available on our about file types page.'

Figure 6: Open Data Nottingham: Car park occupancy data. Available from: <http://www.opendatanottingham.org.uk/>.

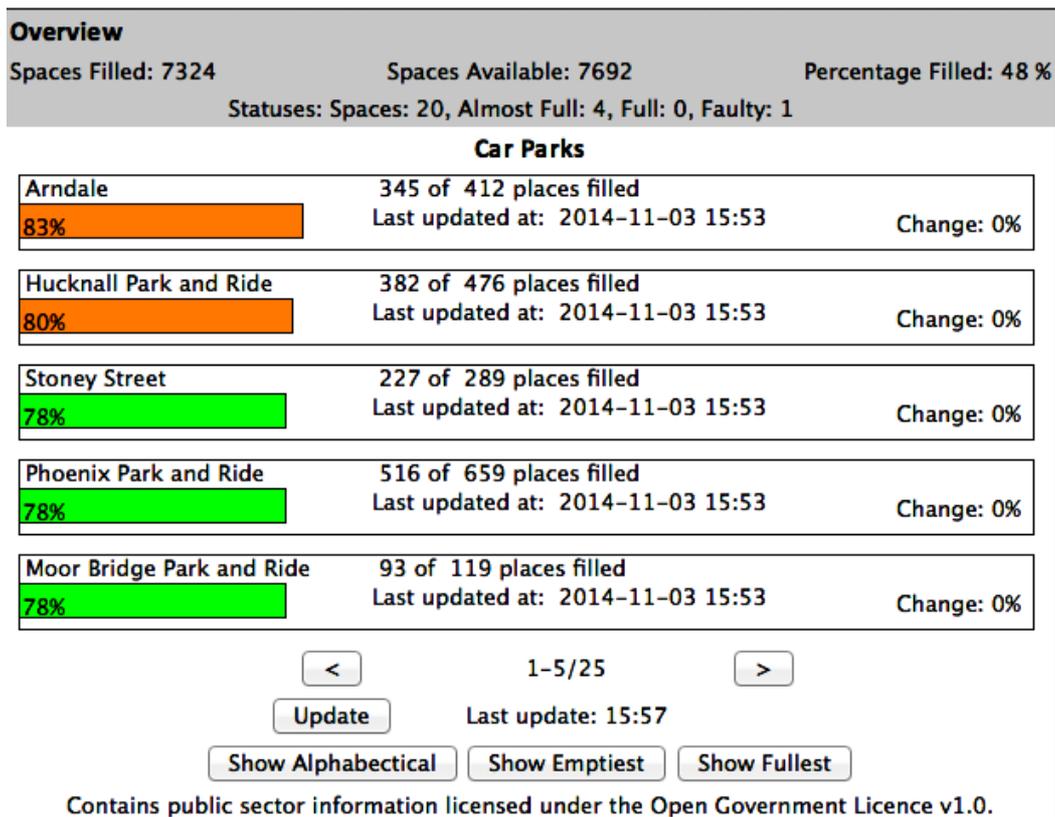


Figure 7: An example solution produced for the Car Parks Monitor coursework.

The Car Park Monitor application was again the most complex project that the cohort created. In this case, when devising the coursework the potential for local interest led to exploration of the datasets made available by Nottingham City Council as part of [Open Data Nottingham](#).

| SystemCode | Capacity | ShortDescription               | State       | Occupancy | OccupancyPercentage | LastUpdated           |
|------------|----------|--------------------------------|-------------|-----------|---------------------|-----------------------|
| 210        | 472      | Wilkinson Street Park and Ride | Full        | 461       | 97                  | 2015-08-25 15:02:46.0 |
| 211        | 119      | Moor Bridge Park and Ride      | Almost Full | 101       | 84                  | 2015-08-25 15:02:46.0 |
| 212        | 476      | Hucknall Park and Ride         | Almost Full | 396       | 83                  | 2015-08-25 15:02:46.0 |
| 911        | 250      | Huntingdon Street              | Spaces      | 14        | 5                   | 2015-08-25 15:02:46.0 |
| Am2-1      | 455      | Trinity Square                 | Almost Full | 410       | 90                  | 2015-08-25 15:02:46.0 |
| Am3-2      | 517      | Fletcher Gate                  | Spaces      | 384       | 74                  | 2015-08-25 15:02:46.0 |
| BRDMSH     | 1185     | Broadmarsh                     | Spaces      | 898       | 75                  | 2015-08-25 15:02:46.0 |
| C11001     | 412      | Arndale                        | Full        | 372       | 90                  | 2015-08-25 15:02:46.0 |
| C11002     | 427      | Mount Street                   | Spaces      | 272       | 63                  | 2015-08-25 15:02:46.0 |
| C11003     | 405      | Queensbridge Road              | Spaces      | 0         | 0                   | 2013-08-20 14:53:00.0 |

Figure 8: Excerpt of Open Nottingham Car Park Occupancy Data in CSV format. This shows only the relevant columns for the coursework. Each row represents the current status of an individual car park. Nottingham City Council, 2015. Licensed under the [Open Government Licence](#).

Figure 5 shows the relevant columns in the raw parking data. This also included a number of columns with no relevance to the coursework, which are not shown in this figure. Students had to work out how to take quantities such as 'Occupancy', 'Capacity', 'State' and 'Percentage' from the CSV file, and correctly translate this into a textual and visual display of the data in a working interface.

The basic requirements asked students to produce views of the data on individual car parks, and combined totals across the parking system. This required the creation of algorithms to calculate totals, and a complex interface to both show the overview information, and move through the detailed information about each of the 20+ car park sites. An advanced requirement asked the students to integrate an update button, which refreshed the display by retrieving the latest data from the website. This pushed their skills by requiring them to access the CSV file directly through their program (whereas for the basic requirements, they could have downloaded a copy of the file manually), to re-calculate the values displayed, compare the old and new data to see changes in the number of spaces filled, and to update their interface accordingly.

Documentation of the open data service was patchy. In some cases it was decided that it was part of the student's task to decipher this, for example the meaning of the column titles above. In other cases it was felt to be necessary to provide hints based on the lecturer's built understanding of the data. For example to: "Please note that the data only updates approximately every 5 minutes, so there is no point to constantly updating, you may not see a change in the values". In this case, there was a concern that students might write code that continually requested data from the service, which was of no benefit to their solution, and may cause problems if the server had limited capacity or if it detected excessive requests for data and responded by blocking access.

As part of the evolving course design, students were also asked to complete a reflective activity: To devise a testing plan for two envisaged scenarios of use: a mobile application for the public, and an analysis tool for council staff. Through this, the multiple potential uses, and users, of open data were foregrounded.

## 2. Reflection

Students responded well to both assignments and reported a high degree of satisfaction in post-module surveys. Engagement and attainment were also observed to be high. Some students reported spending much more time than was expected on the assignments, but this was then reflected in high average scores for both assessments, rather than the assignment being too difficult to pass.

Reflecting on the successes of using open data in this context, two qualities appear fundamental: Ease of Access and Authenticity.

### Ease of access

The value of Open Data in promoting ease of access is exemplified in the e-reader assignment, where Project Gutenberg's bank of public domain texts are provided in the raw .txt format. While programming was new and alien, students could find the books and the e-reader concept familiar, and could integrate books of personal interest into their created application.

In designing the assignment, we considered the course content and desired type of challenge, and used the idea of working with a repository of open content as a guiding concept. A repository that could be easily used without licence concerns, or complicated proprietary formatting, was a valuable way of framing this design process. Ideally, this content would be something of interest to the students as well. Given the thousands of books available on Project Gutenberg, students should be able to fill their e-reader with personally interesting content as they created it.

Without the Project Gutenberg texts, the standard approach of presenting e-books in complex or proprietary formats, often using digital rights management technologies, would have made this an inappropriate or impossible task at this introductory level.

## Authenticity

Authenticity is a sought after quality in teaching and learning, particularly with regard to bridging the gaps between conceptual knowledge and real-world application. Authentic learning requires contexts that reflect real knowledge use, ill-defined activities, access to expert performances, and realistic collaboration and assessment (Herrington and Herrington, 2006). These cases particularly show the potential of open data to increase elements of real knowledge use and ill-defined activity in comparison to standalone applications, or artificially-created examples of data.

While collaboration and access to expert performance are limited, a related quality is well represented: Professional software development requires working with data and systems produced by other people and organisations. Here, learners were prompted to understand data structures devised by others, and to explore obscured features (e.g. how data was calculated, and how updates would appear).

Valuable characteristics of the car parking data were its multi-faceted nature (containing text, numbers, date and time, and categorical data), and that it was updated regularly, which facilitated challenges around networking and data manipulation. In addition, there was a clear idea for how this data could be translated into a potentially useful application that did not exist: A usable interface through which to understand the changing state of parking across the city.

The use of open data produces scope for these authentic challenges to emerge. This is valuable but also brings with it a concern to stay on the right side of a boundary between authentic problem solving, and losing control of the learning activity. The creation of model solutions was used in advance to ascertain that the coursework could be completed, and to identify points of difficulty along the way. However there remains an emergent potential that is hard to fully comprehend in advance.

For example, a student paying sufficient attention might realise that because of a minor flaw in the way data on cars entering and leaving is captured, the number of spaces taken could at times exceed the total number that existed (e.g. 105 cars parked with only 100 spaces). Left alone, this would lead to a graph size greater than 100% appearing, with a subsequent effect on the application interface. Students identifying this were able to remedy the problem through their coding. The problem

was understood in advance of the coursework setting, thanks to testing model solutions.

However a further example was less visible: Overnight, a different category occurs in the car park status column, with some car parks listed as 'closed'. This was not documented, and prior to setting the coursework, model solutions were only tested during the working day, so this category was not included in the coursework specifications. It was only revealed by nocturnal students, working late into the night. This had no real ramifications, but acted as a warning that the data source was not entirely understood despite the efforts made. This was a lesson for all: Seek to understand the data source systematically under a wide variety of circumstances!

From the educator's point of view, this emergent authenticity should be largely welcome, but there are risks of more substantial problems arising. While there were some notes explaining the car park data, as with many other open data sets, documentation is limited. The data is produced by systems designed for a purpose other than public consumption or educational use. They can however become an excellent, authentic material, given sufficient investigation and an orientation towards embracing the authentic challenges that emerge.

### 3. Recommendations for good practice

The authentic nature of open data - encompassing real world relevance, complexity, and even flaws - makes it stand apart from artificial constructions or textbook examples. In this case, open data was used in response to a desire for students to apply themselves to challenges that are common to professional programming. These would be difficult to manufacture yet were readily available through building assignments around suitable open data. As it is a common goal across subjects to connect teaching with authentic professional practice, similar value could be drawn elsewhere.

There was little guidance on the use of open data in education available when these assignments were constructed. Therefore as a take-away, the following steps draw upon this case study to define a process to follow in the creation of a learning activity using open data:

1. Identify high-level goals or learning outcomes that the use of open data might achieve or improve: For example to increase realism and connections to the 'real world' processes of working with networks, files and formats, and to facilitate complex projects that are somewhat familiar to the students, because they relate to their lives.
2. Match desired challenges with data sets through envisioning potential outcomes: Take time to explore a variety of data sets to assess their potential. Concurrently, devise and evaluate ways that students could produce an outcome using these candidate data sets. For example in these cases, challenges such as searching or combining data, and creating interfaces that require multiple interacting components, were found to be particularly well matched to particular datasets and outcomes.
3. Consider emergent points of tension where openness could add authenticity, or lead to the emergence of problems. Is the dataset clearly designed, robust, provided with sufficient documentation? Is there information that you as an educator need to provide to the learners to facilitate the challenge? Or can figuring out how to use the data add to the challenge? How does the data change and will this impact upon the nature of the activity? Is a solution to the proposed activity already available and does this impact upon its suitability?
4. Create model solutions in advance to check the challenges involved: In doing this, identify points that could trouble students (e.g. where documentation is lacking). Again, decide whether these are part of the authentic challenge, or require scaffolding or workarounds.
5. Monitor the activity during implementation: Maintain awareness of any unexpected elements that emerge during the implementation, and again, consider whether to provide additional guidance, or to accept the emergent authenticity as a positive element of challenge. Finally, consider if this introduction to open data for students can lead to further open practices to become a part of their, and your, ways of working.

## References

Guo, P. (2014) Blog@CACM: Python is Now the Most Popular Introductory Teaching Language at Top U.S. Universities. Retrieved from <http://cacm.acm.org/blogs/blog-cacm/176450-python-is-now-the-most-popular-introductory-teaching-language-at-top-us-universities/fulltext>

Herrington, A., Herrington, J. (Eds.) (2006). Authentic learning environments in higher education. Information Science Pub, Hershey, PA.

Open Data Nottingham: <http://www.opendatanottingham.org.uk/>

Project Gutenberg: <https://www.gutenberg.org/>

# 3 Teaching Data Analysis in the Social Sciences: A case study with article level metrics

---

**Katie Shamash, Juan Pablo Alperin & Alessandra Bordini**  
**Publishing Program, Simon Fraser University, Canada**

---

**Keywords:** *Open Data; Altmetrics; publishing; technology; data analysis; data science; metrics; measurement; education*

## 1. Context and task description

Metrics and measurement are important strategic tools for understanding the world around us. To take advantage of the possibilities they offer, however, one needs the ability to gather, work with, and analyse datasets, both big and small. This is why metrics and measurement feature in the seminar course [Technology and Evolving Forms of Publishing](#) (Alperin, 2015a), and why data analysis was a project option for the [Technology Project](#) course (Alperin, 2015b) in Simon Fraser University's [Master of Publishing Program](#).

Yet, not everyone has these powerful tools at their disposal. To help bring them to a broader population, every student—whether of the humanities, social sciences, or other sciences—should have an opportunity to become familiar and experiment with working with data. The case study described here is the result of the experience that a group of Master of Publishing students had in these courses, and of how they learned the value and limits of working with quantitative data.

The Technology Project course, of which this case study was a part, has a substantial practical component. It is purposely designed to bring the same types of unknowns and ambiguities that characterize real-life uses of technology. In the 2014-2015 academic year, during the first half of the course, students were assigned a specific task, but in the second half (described here) they had the opportunity to choose from several different options, including analysing data from an openly available dataset. In all cases, the requirements were left intentionally vague and open-ended, so as to give students the opportunity to tailor the project to their own interests, ambitions, and experience.

The entire project description was as follows (Alperin, 2015b):

***“Data Analysis with Google Refine and APIs:** Pick a dataset and an API of your choice (Twitter, VPL, Biblioshare, CrossRef, etc.) and combine them using Google Refine. Clean and manipulate your data for analysis. The complexity/messiness of your data will be taken into account”.*

It included a description of what students would be evaluated on (Alperin, 2015b):

- 1. The value of the analysis you carry out*
- 2. The number of different types of data manipulations that you carry out*
- 3. The number of different tools you successfully employ in your analysis*
- 4. How you present the results*

A classroom session was used to allow students to self-select into projects, and to ask questions about each one. During this time, the project faced its first and biggest obstacle: fear.

Overcoming the students’ resistance was the main challenge of the open data project. It was one of five different projects presented to the class, and when the class was asked to divide into groups based on the projects that interested them, only one student volunteered for the data analysis group. When asked why they did not want to join her group, most students admitted feeling uncomfortable working with data. Many

of them had earned non-technical degrees, and had little or no experience in maths, statistics, or computing. They doubted whether they could complete the project, and felt unsure about how data analysis would help them in their future publishing careers.

In practice, as students learned more about the project, it became important to de-emphasize the first point (the value of the analysis) in favour of focusing the project around the exploratory aspects. This would allow students to feel comfortable experimenting with data analysis techniques and software, without getting caught up in finding meaningful results.

The opportunity for students to work with a dataset of their own choosing—one relevant to their area of study, publishing—was also key to overcoming their initial fear and resistance. After the first student explained her interest, others were able to see the relevance of the project to their field, and the value of learning a skill that is in high demand. One by one, three other students joined what would later be dubbed Team Commander Data.

The team eventually chose to work with the [Article Level Metrics \(ALM\) dataset](#) from the Public Library of Science (PLOS, 2014), which contained information on the social media and usage of articles of every article published by PLOS between March 2009 and September 2014. The dataset, however, did not contain information about the articles themselves, except for the articles' Digital Object Identifier (DOI). Using the DOI, the team was able to add the article's title, date of publication, and other metadata found in the CrossRef database, and available openly through their [API](#) (CrossRef, 2015). The combined dataset had a variety of data types (i.e., numeric, dates, strings) and types of variables (i.e., ordinal, numeric, categorical), which gave the students the opportunity to do a wide variety of analyses.

The next obstacle the team faced was working with a dataset that hadn't yet been cleaned or formatted to address the questions of interests to the students. The team spent the much of the five-week period—as was expected—cleaning and formatting the dataset using Google's [OpenRefine](#). Dealing with character encoding problems, missing values, inconsistent entries, and similar issues caused the students some frustration. However, they soon recognized these challenges as an inherent part of working with real data, as well as a valuable learning experience that would assist them in their future professional endeavours. Working with incomplete data also showed the team first-hand the possible errors and limitations in any dataset. The

team worked through these inconsistencies and issues, and published the resulting files, which gave them confidence and a sense of accomplishment.

Similarly, the team felt more cohesive and motivated by sharing the insights they gathered from their analysis. Having relevant data, which they improved, expanded, and learnt from, allowed them to enter into conversation with a larger community. Even though it was not required of them, the team became interested in sharing their findings and published the results they deemed most interesting on [The Winnower](#) (Alperin, Bordini & Pouyanne, 2015). To complete the open data life-cycle, the team also uploaded their augmented and cleaned dataset to FigShare (Bordini et al. 2015), where it can be reused and easily cited by others.

Although not part of the original plan, both the article published on The Winnower and this case study became part of the pedagogical experience, giving the students hands-on experience in collaborating with a professor. The publication on The Winnower, which at the time of writing had almost 900 views, brought them into conversation with other interested academics, as well as with PLOS staff who provided contextual information not available in the data. This was an unexpected, but proved an important lesson for the students, who might have otherwise felt that the data itself could provide all the answers.

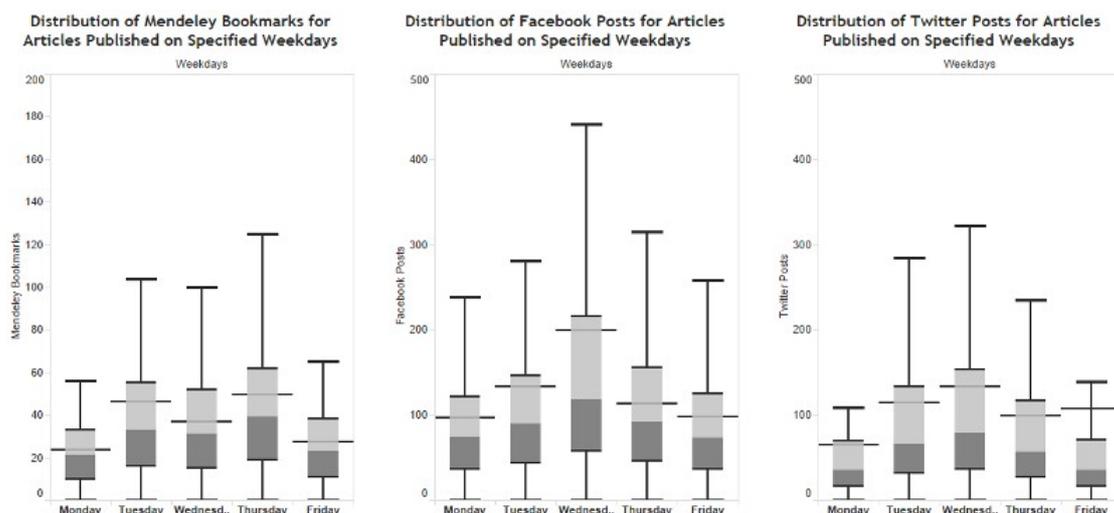


Figure 9: Some of the student's findings.

Paramount in overcoming obstacles and exploiting the unexpected learning opportunities was the grading model laid out at the start. The assessment criteria was premised on the belief that students would be most engaged with the project, learn more, and naturally explore the best learning opportunities if they were allowed to set their own deadlines and deliverables on which they would be evaluated—as they would in a real-life scenario. To ensure students picked appropriate and realistic deliverables, they were to be “negotiated with the instructor after week 1 of the project” (Alperin, 2015b) and were told that more ambitious goals would be rewarded. As a result, students pushed themselves beyond their comfort zones, but not beyond what they could reasonably accomplish in the few weeks available.

Overall, both the students and the professor felt the project was a success. As planned, Team Commander Data learned some of the technical skills required in manipulating and analysing data, but more importantly, they overcame their fear of using data to address an informational need. Moreover, they learned the value of open data itself, and experienced first-hand how data could be used in ways that could not be foreseen or intended by those who published it. They also experienced how they themselves could, with little additional effort, contribute to this open data ecosystem and engage with the wider community.

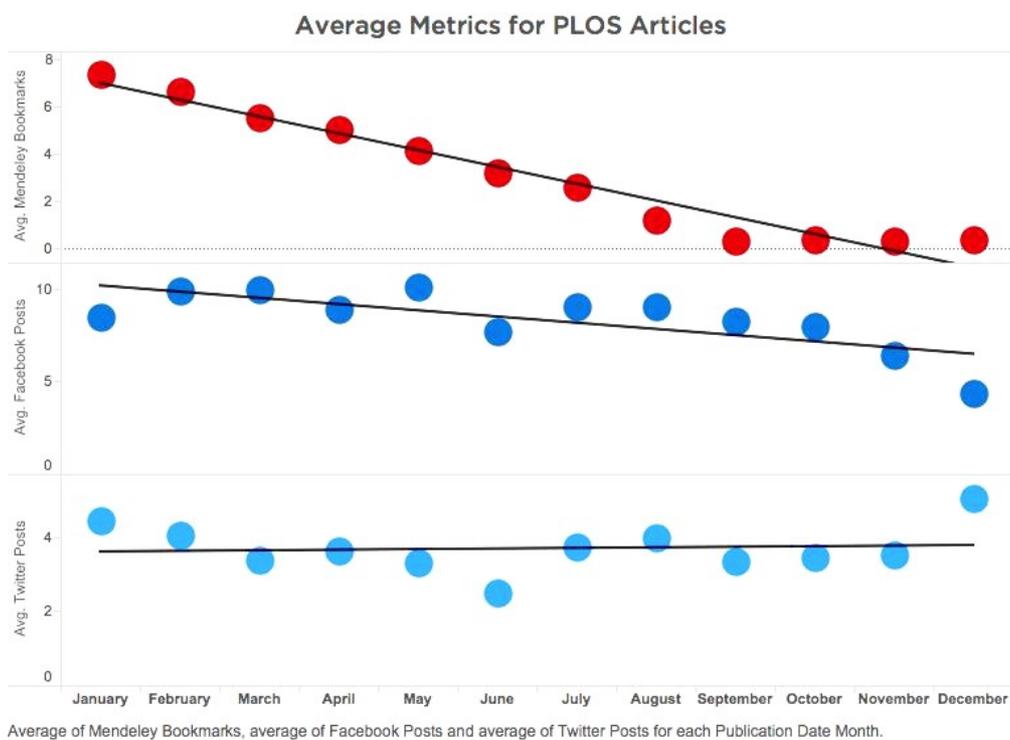


Figure 10: Some of the student's findings

## 2. Reflection

The success of the project, we feel, was due in large part to it having the following characteristics, which we recommend others incorporate into their practice:

1. The project was designed so that students felt safe. The biggest challenge the members of Team Commander Data faced was overcoming their fear of working with numbers. By choosing the most important aspects of the project themselves—the dataset, the questions asked of it, and the presentation method—they were able to find a project that was the right fit for their skill level and interest.
2. Students were encouraged to explore. The team was encouraged to perform any kind of analysis on their data, no matter how strange. This allowed them to have fun with their analysis. They could focus on the skills they wanted to learn instead of the results. Asking unusual questions has the added benefit of opening the possibility of finding unexpected results that have not previously been explored.
3. Students chose to share. One of the greatest skills Team Commander Data learned was how to explain results to a non-technical audience. Its members were encouraged to create visualizations, write summaries of their findings, and present to colleagues and faculty. The results deemed the most interesting were published in *The Winnower* and shared on social media. The dataset they produced was uploaded to Figshare with documentation. Sharing the result in as many ways as possible brought the students into conversation with their community, and encouraged further collaboration.

Although these characteristics were not all foreseen at the outset, they were all facilitated by keeping the project open ended, along with a flexible grading scheme, which motivated the students to explore new skills, and to challenge themselves. By taking the pressure off conducting rigorous statistical analysis and finding meaningful relationships, students were able to develop new skills in a more enjoyable learning environment.

More broadly, this case study shows the value of using openly available data in the classroom. While the opportunity to work with data pertaining to their chosen field of study proved essential to keeping students engaged to the project and on task till the end, it was dealing with the “real-world” messiness of the data that allowed them to gain valuable hands-on skills. Moreover, the open data ecosystem encouraged the students to share their insights and results, thus gaining a sense of belonging to a community, and a sense of pride in doing meaningful work.

## References

- Alperin, J. P. (2015a). PUB802: Technology and Evolving Forms of Publishing Syllabus for Spring 2015. Retrieved from <http://tkbr.ccsp.sfu.ca/pub802/syllabus-2015/>
- Alperin, J. P. (2015b). PUB607: Technology Project Syllabus for Spring 2015. Retrieved from <http://tkbr.ccsp.sfu.ca/pub607/syllabus-spring-2015/>
- Alperin, J. P., Bordini, A., & Pouyanne, S. (2015). PLOS, Please publish our articles on Wednesdays: A look at altmetrics by day of publication. The Winnower. Retrieved from <http://doi.org/10.15200/winn.142972.29198>
- Bordini, A., Dabrowski, P., Pouyanne, S., & Shamash, K. (2015): PLoS ALM with DOI. [Data file and description]. Retrieved from <http://dx.doi.org/10.6084/m9.figshare.1362150/>
- CrossRef. (2015). CrossRef REST API. Retrieved from <http://api.crossref.org/>
- PLOS (Public Library of Science). (2014). Article Level Metric Data Cumulative Report. [Data file and description]. Retrieved <http://article-level-metrics.plos.org/plos-alm-data/>

# 4 The Alan Walks Wales Dataset: Quantified self and open data

---

**Alan Dix<sup>(1)</sup> & Geoffrey Ellis<sup>(2)</sup>**

**(1) Talis and University of Birmingham, UK**

**(2) University of Konstanz, Germany**

---

**Keywords:** *open science; quantified self; biosensing; ECG; geodata; data cleaning; sentiment analysis; portfolio development; flipped class; learning analytics*

## 1. Context and task description

This case study describes the educational use of an open dataset collected as part of a thousand mile research walk. The content connects to many hot topics including quantified self, privacy, biosensing, mobility and the digital divide, so has an immediate interest to students. It includes inter-linkable qualitative and quantitative data, in a variety of specialist and general formats, so offers a variety of technical challenges including visualisation and data mining as well. Finally, it is raw data with all the glitches, gaps and problems attached to this.

The case study draws on experience in two educational settings: the first with a group of computer science and interaction design masters students in class-based

discussions run by the first author; the second a computer science bachelor's project (Kolb, 2015) supervised by the second author.

For the former the principal value of the open data was to make real the issues that arise when large quantities of personal information become available publicly. This was less using the data itself as using the existence of the data as a provocation for a number of discussion topics including privacy, quantified self, and the value and politics of open data in academia. It was taught in flipped mode (Magna, 2015), with students watching videos about the data before more interactive face-to-face discussions.

The second included far more in-depth analysis of the data, including data cleaning, text and data mining, linkage to other open data and visualisation. This will form the greater part of this case study. The main value here of the open data was the availability of a substantial, reasonably well-documented dataset combining quantitative and qualitative elements. In particular it gave the student exposure to real world sensor and textual data with all the problems that entails. In the words of the student:

*“Most of the exercises we do [are] with very artificial data or training data.”*

*“Interesting and challenging at the same time [not just] leaving the theory in the university but really doing it.”*

Intentionally the project student was not given a specific brief other than to use the data in whatever way he chose. This open brief both allowed significant autonomy in choosing which kinds of data to focus on, and what approaches to use, and aimed to develop the student's initiative and sense of ownership of the project outcomes.

The dataset was collected during a three-month period in 2013, when the first author walked the perimeter of Wales, a distance of approximately 1050 miles or 1700 km. As well as pursuing personal research aims, this was an open-science project with Dix offering himself as a living lab for other researchers' projects and where all the data produced is available in the public domain.

Both qualitative and quantitative data were collected<sup>1</sup>. The former were always due to explicit processes (taking a photograph, writing a blog or making an audio recording). The latter were passively collected either by devices in the first author's rucksack (GPS location tracking) or attached to his body (biosensors and accelerometer).

The raw data is supplemented by detailed technical documentation, YouTube videos (Dix, 2014) and extensive online reports and published papers (AWW, 2013; Dix, 2013; Morgan et al., 2014).

The project student worked on three main parts of the data:

1. cleaning, tidying and merging GPS traces;
2. extracting heart rate data from the ECG data;
3. performing sentiment analysis on the blog texts.

## 2. Reflection

Often the data available in student exercises is 'toy data' created to be easy to work with, or guaranteed to provide closed answers using particular techniques. The difference between a puzzle and problem is precisely the open ended nature and lack of guarantee of a closed solution. The AWW is real with all the messiness as well as engagement that this entails.

Some learning outcomes were clear from the outset: skills in data cleaning, application of data mining and visualisation techniques. However, beyond the knowledge that

---

### 1. Summary of available data:

- Daily blogs: approximately 150,000 words. The blogs are in the process of being semantically tagged, but this is incomplete, so only the raw text was available to the students.
- Photos: approximately 19,000 photos, on average more than one every 100 metres, all time stamped.
- Audio blogs: short audio notes, approximately 10 per day, again time-stamped.
- GPS: Collected by two different devices (a Garmin dedicated GPS device, and a tracking app on a mobile phone).
- ECG: Full ECG data (heart trace) for 60 days of the walk (the largest ECG trace in the public domain). Typically including one overnight for every two days to yield baseline long-term data as well as more dynamic daily data.
- EDA (Electro-Dermal Activity —previously known as GSR, galvanic skin response) as used in lie detectors to measure emotional reaction.
- Accelerometers: On both wrist and chest.

grappling with this real data was likely to be educationally valuable, the authors had limited foresight as to the full range of learning that would emerge.

This section outlines some of the pedagogic lessons we learnt during the process, both positive benefits and potential pitfalls.

## Raw data

Much of the data is raw, as collected from sensors. This is very real training but also complex and time-consuming – data cleaning alone has been estimated to take up to 80% of analysis time.

*"I would say that it was very interesting and insightful to work with this dataset as it had this real world application character and was very 'raw', however, this fact lead to most of the problems" (student)*

Problems with the data included:

- Transforming between formats —even extracting fields from JSON into CSV required hand-crafted scripts.
- Merging readings from multiple sources —in particular multiple GPS traces with different accuracies.
- Sensor glitches —for example GPS sometimes reports an occasional reading many miles from the true location.
- Manual processes —for example, forgetting to turn GPS recording on/off as can be seen by the long line in raw data in fig. 1, due to the Garmin being on during a car journey to a wedding.
- Missing or misleading readings —batteries running out, wetness affecting readings, placement of ECG sensors.

While the raw data should always be part of the open data, pre-cleaned data and part-processed versions of the data would help both in the 'getting to know the data' stage and also to allow faster bang-for-buck during project work.

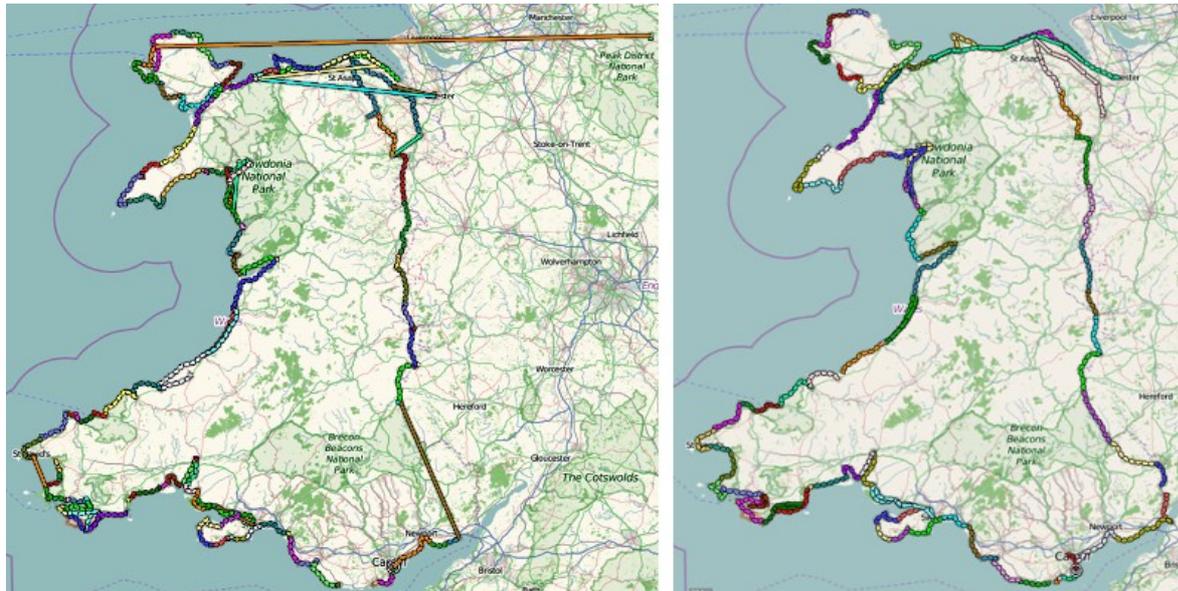


Figure 11: GPS data cleaning (Kolb, 2015): a) raw data from Garmin, b) cleaned merged data.

## Documentation

For a dataset to be useful, whether in education, research, or practical application, it needs to be very clearly documented. This is frequently a problem with government open data, either because it has been released for show, but without real buy-in from the providers, or because the raw data requires very specialised understanding<sup>2</sup>. Although the student had to do extensive desk research locating suitable open source tools, translating data formats, etc., he was able to work with the data itself based solely on the documentation. However, he did make use of the first author to interrogate aspects of the context of the data: for example "what happened at 8am in 27th July, as your heart rate hit 200bpm?". The most extensive use of this was to provide 'ground truth' to validate the sentiment analysis.

2. Here are two examples the authors' have encountered in dealing with open government data. While the kinds of 'rawness' seen in the AWW data set, are less evident in most government open data, which is already sanitized, this does not necessarily mean they are easier to use.

In 2010 one of the major issues in the UK General Election was the national deficit, and in particular the performance of the last Labour government. The first author tried to find historic data giving actual figures —it was surprisingly hard. He first looked to the National Statistics Office, the holder of official data for the UK. Undoubtedly the numbers were there, but spread over so much detailed data with terms that only an expert economist would understand. Happily this data is now available at the Guardian Data blog, which gives a simple graph and spreadsheet of deficit and national debt by year (Guardian, 2010).

In another example, the OnSupply project gathered data related to renewable energy usage and related demographics on Tiree, one of the Scottish islands (Catalyst, 2015; Simm, et al., 2015). Some data was available from government open data, but listed using region codes. While these would clearly be well understood by those dealing with the data day to day, it took the team (hardly digital novices) a considerable time to find the tables to decode these and find the right code corresponding to the area including Tiree.

## Size of data

In addition, the sheer size of the raw data was challenging. The student is studying in a group known for state of the art research in data mining and visualisation. However, he struggled at times to find tools capable of processing the larger parts of the datasets. For example, the ECG trace is sampled at 64Hz, generating 25Mb of compressed data per day. Data reduction (from ECG to heart rate) was therefore essential to be able to deal with the data. In many ways this is still small data compared with, for example, Twitter Firehose, or CERN event data; however, still problematic within the time scale of a student project.

*"[...] it is hard to load 3 million points into a visualisation [...] standard line plots about 10000 values." (student)*

## Linkage

The different parts of the datasets are linked principally by time (e.g. given an interesting event on the ECG, it is possible to correlate this with EDA, GPS position, photos taken, audio blogs, etc.). Location is also highly significant, largely, but not entirely correlated with time as the blogs can refer to events and places at other points in the walk. These also allow connections to other data (weather reports, GeoNames).

Some of the student's work related solely to the AWW data, but some cross-linked it with other open data sources; for example, the sentiment analyser was trained using other walking blogs and the GPS elevation data was cross checked with OpenStreetMap (2015) elevation data. He also made extensive use of open source tools such as GPSBabel (2015) and PhysioNet (2015).

## Visualisation

As well as data manipulation, analysis and transformation, the student carried out extensive visualisation, both at the getting to know the data stage and as a way of presenting final data (see fig. 2). The ability to cross-link the individual datasets in the

collection was critical for this. However, as noted, the size of the dataset was problematic.

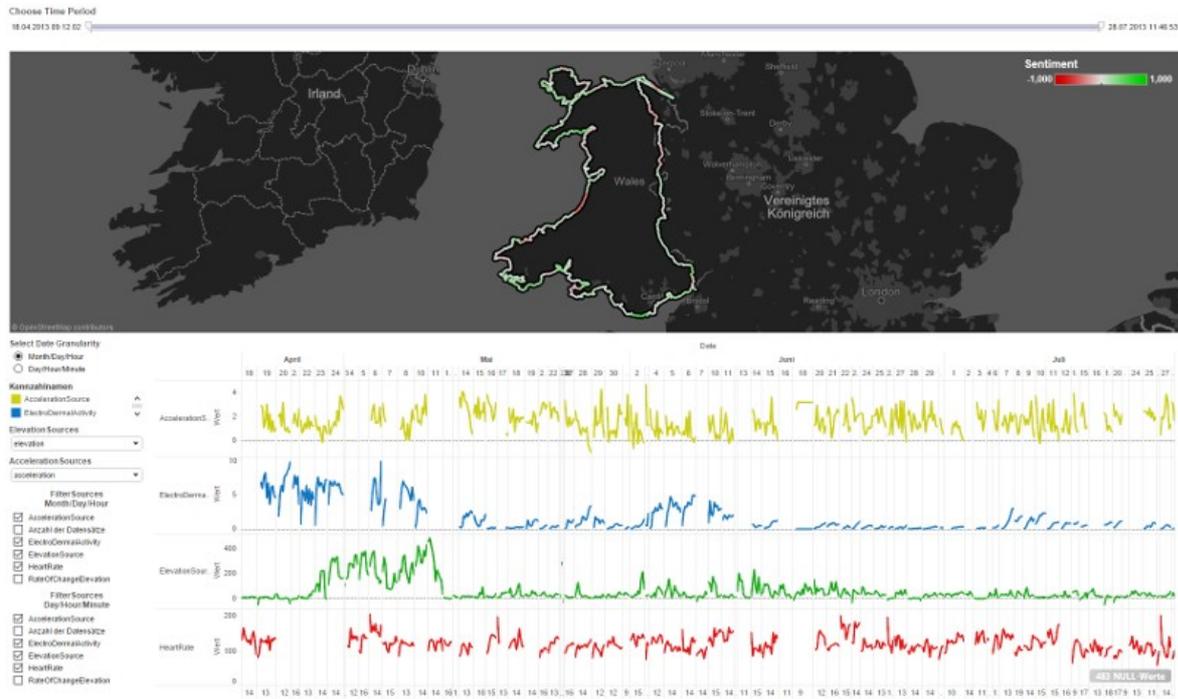


Figure 12: Visualisation of multiple quantitative data streams linked to graphic location (Kolb, 2015).

### 3. Recommendations for good practice

We will now discuss some of the broader pedagogic issues, and then summarise the core take-away advice from this and the preceding section.

#### Open brief and raw data

The open brief worked for this very capable student, who was able to explore many aspects of the data:

*"[...] you have more possibilities I think, so you don't know beforehand what you can come up with or what information lies inside the data." (student)*

However, this combination of open brief and raw data is daunting and adds a lot of uncertainty:

*"It could be that there is nothing, which wouldn't be very nice!" (student)*

Given this, a tighter brief and a cleaner dataset (alongside the raw data) might have been helpful for a less confident student. Indeed, another student was initially considering working on the dataset, but in the end decided to work with other (perhaps easier!) data.

As noted clear documentation is essential, and this was a real challenge to the first author, who had never documented data before. Compared to much open data, it appeared that the level of documentation was effective, but a wider range of data formats would be useful as would pointers to tools capable of processing the data.

This need for clean and well-documented data would be even more important for use in practical labs or flip class teaching.

## Student as data provider

One of the great successes of this project is that the data produced by the student is being fed back into the public domain. This includes tidied GPS traces, open street map elevation data, sentiment analysis training data, and heart rate data.

This benefits future students and researchers who will be able to get a head start, building on previous work rather than starting from scratch. The data will include links to the student's project report (Kolb, 2015), which documents the tools and processes used to create the derived data, essential provenance for future users of the data.

This will also benefit the student as he will be acknowledged on the data web site and can include this in his portfolio for future academic and job applications.

## Flip-classroom teaching

As noted the data was used partly in flipped mode. This seems a promising way to use open data as it is far more profitable to discuss the data in class than to lecture about it. However, this does require that the data is very well documented, which, as we

have noted, is not always the case for open data. For the AWW data, the use in flipped mode was made easier by the existing videos describing the data itself, background of data collection, and challenges it poses.

The flipped class was part of a Talis Lighthouse Pilot (Talis, 2015), using a universal media player, this meant that both video and PDFs viewed could be annotated in a uniform fashion and detailed learning analytics were available. The latter made it possible, for example, to do hot spot views of videos, or see where students dropped out when reading a document. Some of the broader flip-class lessons from this and the rest of the pilot have been reported elsewhere (Dix, 2015).

Currently, the data itself can be downloaded, but it would be very valuable to be able to have similar kinds of annotation and analytics as are available for other media. Given the wide variety of formats and types of data, this poses interesting technical challenges to develop suitable APIs so that data embedded into learning environments can become a seamless part of emerging rich pedagogy.

## Core Takeaways

- The extent and potentially interlinked nature of the AWW dataset makes it a valuable introduction to the real problems of using data.
- When dealing with rich data a range of briefs from open to more directed are needed for different student ability.
- Raw data is educationally valuable giving students experience in dealing with data cleaning and related skills, but...
- It takes a long time to 'get into' a data set, so...
- Cleaned versions of the data and subsets of large datasets may be needed for simpler projects or to help introduce students to complex data.
- Clear data documentation is essential; if (as is common with open data) the documentation is not available, make it... or set this as a student project!

- Allow students to feed back into the open data, as co-creators of the data and educational material, this benefits the student professionally and adds value to open data.
- Well documented open data is a potentially valuable resource for flipped modes of learning.
- Data should be able to fully integrate into rich online learning environments, although this is currently limited by existing technology.
- Real data is exciting, but real data is hard.

## References

AWW Video: <http://alandix.com/academic/talks/AWW-sensing-the-miles-2014/>

AWW (2013–2015). Alan Walks Wales Report, <http://alanwalks.wales/report/>

Catalyst (2015). OnSupply – Catalyst Project: Citizens Transforming Society Tools for Change. Retrieved from <http://www.catalystproject.org.uk/projects/sprints/on-supply/>

Dix, A. (2013). The Walk: exploring the technical and social margins. Keynote, APCHI 2013 / India HCI 2013, Bangalore India, 27th September 2013. Retrieved from <http://www.hcibook.com/alan/talks/APCHI-2013/>

Dix, A. (2014). Alan Walks Wales: Sensing the Miles. Video Presentation, 'Enhancing Self-Reflection with Wearable Sensors', workshop at mobileHCI 2014 Toronto, 23rd September 2014. Retrieved from <http://www.alandix.com/academic/talks/AWW-sensing-the-miles-2014/>

Dix, A. (2015). More than one way to flip a class: learning analytics for mixed models of learning. APT 2015, Greenwich, 7th July 2015. Retrieved from <http://www.hcibook.com/alan/papers/apt2015-more-than-one-way/>

Guardian (2010) Deficit, national debt and government borrowing - how has it changed since 1946? Guardian Data Blog 19th Oct. 2010. Retrieved from <http://www.guardian.co.uk/news/datablog/2010/oct/18/deficit-debt-government-borrowing-data>.

Kolb, D. (2015). Walking Wales : The Data Challenge. Bachelor Project, University Konstanz,

Germany.

Magna (2015). Flipped Classroom Trends: A Survey of College Faculty. Faculty Focus Special Report. Magna Publication. Retrieved from <http://www.facultyfocus.com/free-reports/flipped-classroom-trends-a-survey-of-college-faculty/>

Morgan, A., Dix, A., Phillips, M. and House, C. (2014). Blue sky thinking meets green field usability: can mobile internet software engineering bridge the rural divide? . Local Economy, September–November 2014. 29(6–7):750–761. DOI: 10.1177/0269094214548399

Simm, W., Ferrario, M., Friday, A., Newman, P., Forshaw, S., Hazas, M., and Dix, A. (2015). Tired Energy Pulse: Exploring Renewable Energy Forecasts on the Edge of the Grid. CHI'2015, Seoul, S. Korea, April 2015. ACM pp.1965-1974. DOI: 10.1145/2702123.2702285

Talis (2015). Lighthouse – a Talis project <http://talis.com/lighthouse>

## Software

GPSTabel. <http://www.gpsbabel.org/>

OpenStreetMap. <http://www.openstreetmap.org/>

PhysioToolkit Software. <http://www.physionet.org/physiotools/softwareindex.shtml>

## Acknowledgements

Many thanks to David Kolb, the student whose project is described here. Thanks also to dot.rural at University of Aberdeen (<http://www.dotrural.ac.uk>) for its support of the walk as part of its partner scheme.

# 5 Open Data for Sustainable Development: Knowledge society & knowledge economy

---

**Virginia Power**  
**University of the West of England (UWE), Bristol,**  
**UK**

---

**Keywords:** *Open; Data; Knowledge; Society; Statistics; Economy; Sustainable; Development; Literacy; Datasets*

## 1. Context and task description

I am a Graduate Tutor who lectures in Information and Knowledge Management (IKM) at the University of the West of England (UWE) in Bristol, United Kingdom. This IKM 15 credit module forms part of the Masters programmes module selection available to postgraduate students working towards the qualifications of MSc. Information Management<sup>1</sup> or MSc. Information Technology. The module attracts a mixed cohort of students from both discipline strands which enables a high level of discussion and debate and a rich transferability of knowledge and experience to take place. There is a high proportion of international students largely drawn from Nigeria and other African

---

1. <http://courses.uwe.ac.uk/P11012/2015#coursecontent>.

countries, as well as a number of students currently employed in information and library service organisations. Typically the module attracts 12-18 students.

Throughout the module we look at how 'making sense' of data as an individual, for an organisation and within society has major benefits for social, cultural, economic development, and why, as information and knowledge managers we have a key role to play to make this happen in a timely and effective manner. The intention is to introduce students to a number of different data sets and web sites and to provide them with opportunities to fully explore what the data can tell us in terms of useful statistics and to consider issues of reliability and validity, including potential statistical bias in what they are seeing. From this we then progress to activities involving comparative data from different sources —what picture can we build regarding the development of a knowledge economy through the use of a variety of sources?

Students are introduced to the topic through an introductory mini-lecture which sets the scene for the ensuing statistical activity. We use a wide range of activities largely based around web site statistics and visualisation tools; although students are given specific countries to research we also look at some specific case studies, academic papers, videos and student-generated research to inform our discussions. I have found that the use of statistics available through the UNESCO Education for All web site<sup>2</sup> has been an excellent way of engaging students with the potentially dry topic of knowledge economies —the web site employs data visualisation techniques which makes the data very accessible. Not all websites produce their data in this way —for example, the International Telecommunications Union<sup>3</sup> database is much more complex and requires greater effort for interpretation.

I have been using open data in one of the module lectures in particular; this lecture focuses on the role of information and knowledge management in the development of global knowledge societies in support of the knowledge economy. Without access to open datasets it would be extremely difficult for me to relate theory to practice and to extend the student's' knowledge and understanding of the key relevance and role that information and knowledge plays in the creation of the knowledge society. Through the use of open data students can interrogate and analyse trends and developments to support their learning. The students usually address this topic halfway through the module, having explored data, information and knowledge as concepts prior to

---

2. [http://en.unesco.org/gem-report/sites/gem-report/files/2015\\_report\\_dataviz/index.html](http://en.unesco.org/gem-report/sites/gem-report/files/2015_report_dataviz/index.html).

3. <http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx>.

addressing knowledge economies. I have found that this helps with the interpretation of data and contributes to a deeper understanding of how we use data to provide information and to deepen knowledge.

At UWE, our commitment to education for sustainable development (ESD) requires each module to include at some juncture an element of ESD and to report on this through our module evaluation. As a University we have an holistic approach to ESD<sup>4</sup> and our interpretation is based on guidance provided by the Higher Education Funding Council for England (HEFCE)<sup>5</sup> and the Quality Assurance Agency (QAA)<sup>6</sup> guidance on education for sustainable development. The QAA provides guidance for higher education institutions to develop students in ways that enables them to:

- Consider what the concept of global citizenship means in the context of their own discipline and in their future professional and personal lives;
- Consider what the concept of environmental stewardship means in the context of their own discipline and in their future professional and personal lives;
- Think about issues of social justice, ethics and well-being, and how these relate to ecological and economic factors;
- Develop a future-facing outlook; learning to think about the consequences of actions, and how systems and societies can be adapted to ensure sustainable futures. (QAA, 2014, p.6).

Thus in this session devoted to the development of Knowledge Societies and the Knowledge-Based Economy students spent some class time in the discussion of the knowledge economy across international countries and specific developments in case studies. Students are introduced to the concepts of a knowledge society and the knowledge-based economy using the definitions provided by the Organisation for Economic and Co-operation and Development (OECD)<sup>7</sup> Exploration of these definitions increased the student awareness of knowledge as a key commodity for economic, cultural and societal growth. In addition questions regarding the knowledge society/economy are included in the end of module examination. Students are

---

4. <http://www1.uwe.ac.uk/aboutus/visionandmission/sustainability/education/ourapproach.aspx>.

5. <http://www.hefce.ac.uk/pubs/year/2014/201430/>.

6. <http://www.qaa.ac.uk/en/Publications/Documents/Education-sustainable-development-Guidance-June-14.pdf>.

7. <http://www.oecd.org/sti/sci-tech/1913021.pdf>.

introduced to the four pillars approach to the Knowledge Economy, an approach devised by the World Bank as a Knowledge Economy Framework.<sup>8</sup>

Students were encouraged to discuss knowledge society/economy concepts through the exploration of key terms and definitions using video and documents. We used videos from GESCI<sup>9</sup> exploring African perspectives and contrasted these with other perspectives, including videos from Georgia<sup>10</sup> and the University of Waterloo in Canada<sup>11</sup>. Students were also given a range of documents and web sites to illustrate the concepts including A Primer on the Knowledge Economy by John Houghton and Peter Sheehan from the Centre for Strategic Economic Studies at Victoria University<sup>12</sup>. Although written in 2000 the primer is an excellent starting point for students. We also used the BBC channel Knowledge Economy for current stories<sup>13</sup>. Once the concepts had been explored, discussed and understood, the students were asked to work together in small groups to use open data sources to track specific countries in order to identify the characteristics, themes and issues of relevance to the development of knowledge societies. For example, what makes Denmark a strong knowledge society in comparison to India? Students were given Denmark, India, Nigeria and Turkey to research and were asked to interrogate data sets in relation to specific questions concerning ICT, Education, Economic Incentive and Institutional Regime and Innovation Systems - the four pillars identified by World Bank as 'critical requisites for a country to fully participate in the knowledge economy' (World Bank, 2013). This dictated which datasets were used.

## 2. Reflection

The students responded to the use of open data very well and appeared to enjoy the practical application of the theory; this is not an easy subject to convey to students! The preliminary discussion of definitions and concepts, together with an understanding of the four pillars approach to the Knowledge Economy was vital so that students could engage with the group discussions. Each group were given a set of questions to

---

8. [https://en.wikipedia.org/wiki/Knowledge\\_Economic\\_Index#The\\_4\\_pillars\\_of\\_the\\_Knowledge\\_Economy\\_framework](https://en.wikipedia.org/wiki/Knowledge_Economic_Index#The_4_pillars_of_the_Knowledge_Economy_framework).

9. <http://www.gesci.org/>.

10. <https://www.youtube.com/watch?v=oBC7bA-PeSl>.

11. <https://www.youtube.com/watch?v=4mWYA3Z-YJ0>.

12. <http://www.cfses.com/documents/knowledgeeconprimer.pdf>.

13. <http://www.bbc.co.uk/news/business-12686570>.

investigate such as education levels, employment in knowledge-intensive services, Internet and broadband access, life expectancy and innovation practices.

Preparation for an intensive session like this is inevitably time-consuming and this was not made any easier through the lack of information regarding if, and how I could use the data within a classroom setting. Most of the notices (tucked away at the bottom of the home pages) talk about personal, non-commercial use and I felt it was therefore permissible to use the data as illustration for instruction; however World Bank with their open data approach, together with their Open Knowledge Repository<sup>14</sup> are to be commended for the clarity and explicit approach to their data use.

Prior to the session I had researched each dataset to ensure that the data could be found and I was confident that the students would not encounter problems; however what quickly became clear was that the level of data literacy, particularly in terms of data interpretation was very low which I found surprising for a postgraduate cohort. As Phil Richards, Jisc Chief Innovation Officer has observed “the ability to process, manage and take tangible meaning from data is becoming increasingly important for all industries and sectors. What this creates is a new requirement to develop a strong analytical talent pipeline, in order that the UK can take advantage of the valuable opportunities that come with smarter data analysis”.<sup>15</sup>

Universities UK (UUK) have just released their 2015 report —Making the most of data: Data skills training in English universities<sup>16</sup> as a means of highlighting how universities can ensure that they are giving graduates the requisite skills in order for them to operate in an ever-increasing digital data-driven environment. In addition two further 2015 reports from Nesta<sup>17</sup> and the British Academy for Humanities and the Social Sciences<sup>18</sup> have also highlighted the issues of data skills competences for the future.

---

14. <https://openknowledge.worldbank.org/>.

15. <https://www.jisc.ac.uk/news/data-literacy-and-skills-development-vital-to-uk-economic-health-13-jul-2015>.

16. <http://issuu.com/universitiesuk/docs/makingthemostofdata>.

17. <http://www.nesta.org.uk/publications/skills-datavores-talent-and-data-revolution>.

18. [http://www.britac.ac.uk/policy/count\\_us\\_in\\_report.cfm](http://www.britac.ac.uk/policy/count_us_in_report.cfm).

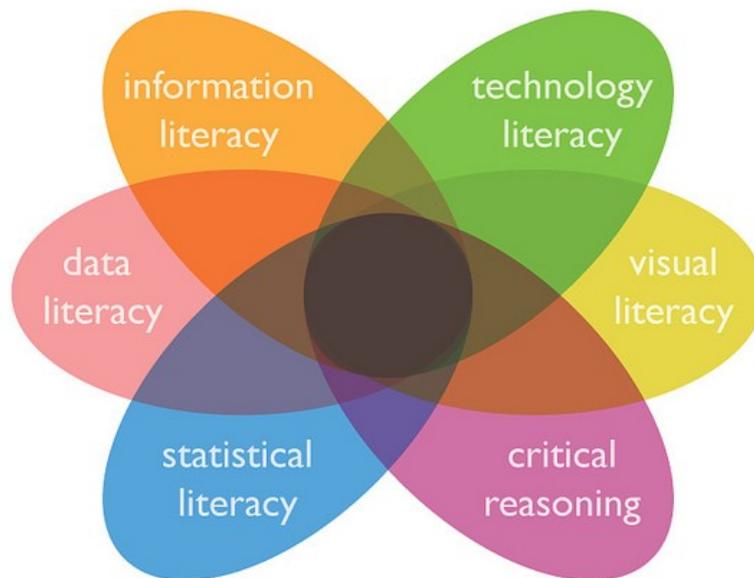


Figure 13: Literacy rainbow. CC BY-SA 2.0 (Justin Grimes).

This has made me reflect on, and review how I could make the process easier for the students, whilst also developing their data literacy skills and I will be looking to develop a simple online data literacy activity to enable students to hone their skills prior to the lecture.

Currently I believe that there are limited resources available for teaching data literacy, although the reports previously mentioned go some way to suggest potential methodologies. Nevertheless many of the existing literacy resources for University focus on data and research management, information literacy or digital literacy. I particularly like the definition of data literacy in the Data Journalism Handbook 'the ability to consume for knowledge, produce coherently and think critically about data. Data literacy includes statistical literacy but also understanding how to work with large data sets, how they were produced, how to connect various data sets and how to interpret them.' (Gray, J. et al., 2012)<sup>19</sup> Therefore, as a University lecturer I believe that this is part of my role to enable students to develop their digital skills competences and I need to convey this to my students.

Nevertheless the data that we were able to find and use was sophisticated and of good quality and quantity to enable us to make some basic observations relevant to the

---

19. <http://datajournalismhandbook.org/1.0/en/index.html>.

session. Of most benefit were the visual tools such as infographics and interactive comparison tools (Education for All<sup>20</sup> and World Inequality Database on Education (WIDE)<sup>21</sup> are particularly good examples) as this brought the data alive to the students and certainly provided an excellent foundation on which to build lively discussion. It is probably fair to say that some datasets were easier to understand than others —some of the tables were quite complex and needed a lot of interpretation such as those available from the International Telecommunications Union<sup>22</sup>—, but I was able to demonstrate with the students the complexities of making sense of data in order to arrive at useful information in support of their learning objectives.

I would like to see more educational support materials like this made available for teachers and students to make data accessible; the Ordnance Survey<sup>23</sup> for example have some excellent materials available for learning and teaching. In addition School of Data<sup>24</sup> advocates learning data skills through an expedition approach which could easily be adapted to a classroom setting. The UK Data Service<sup>25</sup> does provide some general advice and guidance about the use of data in learning and teaching and some great ideas.

Having such sophisticated datasets available as open data and literally at our fingertips was so useful and is definitely something that I shall continue to use and exploit for learning and teaching. I am very excited to see more and more data being released as open, for example the United Kingdom Department for Environment, Food and Rural Affairs (DEFRA)<sup>26</sup>, as this provides high quality and extensive learning and teaching resources in a vast array of subject areas —all for free!

---

20. [http://en.unesco.org/gem-report/sites/gem-report/files/2015\\_report\\_dataviz/index.html](http://en.unesco.org/gem-report/sites/gem-report/files/2015_report_dataviz/index.html).

21. <http://www.education-inequalities.org/>.

22. <http://www.itu.int/en/ITU-D/Statistics/Pages/default.aspx>.

23. <http://www.ordnancesurvey.co.uk/education-research/index.html>.

24. <http://schoolofdata.org/data-expeditions/>.

25. <http://ukdataservice.ac.uk/media/398744/learningteaching.pdf>.

26. <https://defradigital.blog.gov.uk/category/open-data/>.

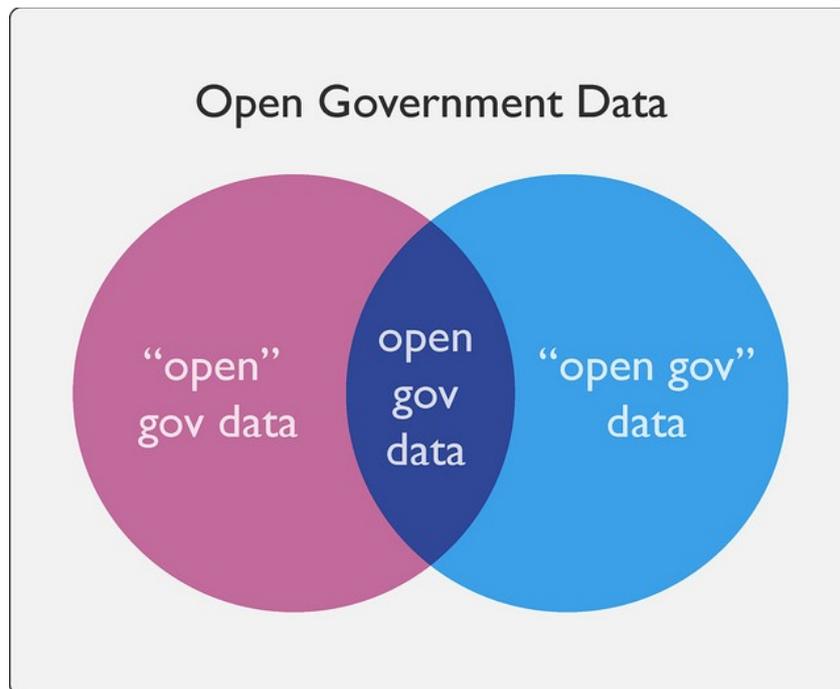


Figure 14: Open Government Data. CC BY-SA 2.0 (Justin Grimes).

Whilst this session was not attached to a particular assignment, the topic of knowledge societies and economies is addressed within the written examination paper at the end of the semester. However, having piloted the approach I am considering making this an assessed piece as I do believe that it gives the students an excellent opportunity to work with different datasets to enhance their knowledge and understanding, together with the chance to build on their data literacy skills. I will therefore be developing a learning activity and assessment piece to enable students to fully comprehend the information that can be gained from open data and to successfully interrogate the data to provide meaningful and carefully considered observations with regard to knowledge societies and the knowledge economy.

I work within the Faculty of Environment and Technology (Computer Science and Creative Technologies) and we frequently have opportunities to talk together about our research and our learning and teaching practice. One such opportunity is our SALT meetings (Sharing Approaches to Learning and Teaching) and this case study will be presented as part of that activity. In addition we hold regular Faculty Forums and the opportunity will be taken to disseminate my findings and my approach further, and to develop effective practice with colleagues who may also be conducting similar

activities, not necessarily with open data. I will also disseminate through Open Education networks!

### 3. Recommendations for good practice

As a result of this activity I would suggest that the following recommendations will help in using open data as part of learning and teaching:

- Read the latest publications on data literacy identified in this case study —there are valuable lessons to be learned and support on how to approach this essential skill as a learning and teaching practitioner;
- Sign up to the Open Data Institute newsletter and visit their website for information on open data and current developments with new datasets;
- Check student understanding of data interrogation; how data literate are your students?
- Provide an opportunity for some data exploration, or a self-study resource prior to any sustained use of datasets;
- Investigate and explore the open datasets that are out there —they are growing and developing all the time and are a huge, high quality and free resource, often with additional educational materials to support their use;
- Be clear about what you can and can't do with the data you are using - not every site is clear about their usage policy;
- Don't underestimate the time it takes to prepare a session involving open data —don't set your students up to fail!
- If you use a dataset and have some learning and teaching materials to share — please do. The provider of the dataset would love to have something to showcase and you will be helping the whole educational community!

# References

- Bakhshi, H., Mateos-Garcia, J. & Windsor, G. (2015). *Skills of the Datavores: Talent and the data revolution*. Retrieved from <http://www.nesta.org.uk/publications/skills-datavores-talent-and-data-revolution>.
- Diamond, I., Davies, R., Richards, P., Shah, S. & Smith, A. (2015). *Making the most of data: Data skills training in English universities*. Retrieved from <http://www.universitiesuk.ac.uk/highereducation/Documents/2015/MakingTheMostOfDataDataTrainingSkillsInEnglishUniversities.pdf>.
- Gray, J., Bounegru, L. & Chambers, L. (2012). *The Data Journalism Handbook*. Retrieved from <http://datajournalismhandbook.org/1.0/en/index.html>.
- Mansell, W. (2015). *Count us in: quantitative skills for a new generation*. Retrieved from <https://gss.civilservice.gov.uk/wp-content/uploads/2015/08/Count-Us-In-Full-Report.pdf>.
- Newman, A., Newman, A., Kowalczyk, P., Newman, A., Richardson, J., & Coley, A. (2015). *Defra digital*. Defradigital.blog.gov.uk. Retrieved from <https://defradigital.blog.gov.uk/>.
- Open Data Institute (2015). *Open Data Institute*. Retrieved from <http://theodi.org/>.
- Ordnance Survey (2015). *Ordnance Survey adds four new products to its open data portfolio*. Retrieved from <http://www.ordnancesurvey.co.uk/about/news/2015/four-new-os-open-data-products.html>.
- School of Data (2015). *Data Expeditions*. Retrieved from <http://schoolofdata.org/data-expeditions/>.
- UK Data Service (2015). *Data-driven learning and teaching*. Retrieved from <http://ukdataservice.ac.uk/media/398744/learningteaching.pdf>.
- UNESCO (n.d.). *Building inclusive knowledge societies*. Retrieved 24 July 2015, from <http://en.unesco.org/post2015/building-inclusive-knowledge-societies>.
- World Bank (2013). *The Four Pillars of the Knowledge Economy*. Retrieved from <http://go.worldbank.org/5WOSIRFA70>.

## Case study data sources

Data was taken from a variety of sources; only World Bank data is explicitly open licensed, although all the data is publicly available:

Global Innovation Index (2015). *Global Innovation Index 2015*.

[http://www.wipo.int/econ\\_stat/en/economics/gii/](http://www.wipo.int/econ_stat/en/economics/gii/).

International Telecommunications Union (2014). *Statistics*. Retrieved 24 July 2015.

<http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx>.

UNESCO (2015). *Education for All Global Monitoring Report*. Retrieved 24 July 2015.

<http://www.unesco.org/new/en/education/themes/leading-the-international-agenda/efareport/statistics/>.

UNESCO (2015). *Education for All Data Viz Tool*. Retrieved 24 July 2015.

[http://en.unesco.org/gem-report/sites/gem-report/files/2015\\_report\\_dataviz/index.html](http://en.unesco.org/gem-report/sites/gem-report/files/2015_report_dataviz/index.html).

UNESCO (2015). *UNESCO: Institute for Statistics*. Retrieved 24 July 2015.

<http://www.uis.unesco.org/Pages/default.aspx>.

WIDE (2015) *World Inequality Database on Education*. Retrieved 24 July 2015.

<http://www.education-inequalities.org/>.

The World Bank (2015). *Data*. Retrieved 24 July 2015. <http://data.worldbank.org/>.

# Acknowledgements

The editors would like to thank our colleagues and friends for their support, sympathy, advice and encouragement. Without them this book might have never been published.

We are thankful to Marieke, William, Maria, Anne-Christin and Ernesto, our fantastic scientific committee, for joining us in this adventure and generously sharing their time and expertise.

We also want to thank to the authors, for sharing their practices and for trusting us.

Paul Bacsich

Open Knowledge Foundation, Open Education Working Group

Elena Stojanovska

Open Knowledge Foundation, Open Education Working Group

María José Rubio

Universitat de Barcelona

Antonio Moneo

Banco Interamericano de Desarrollo

Geraldine García

Banco Interamericano de Desarrollo

Lorna Campbell

University of Edinburgh

Paul Ayris

University College London

Mehmet Izbudak

SOAS, University of London

Joana Barros

Birkbeck, University of London

Nelson Piedra

Universidad Técnica Particular de Loja

Virginia Rodés

Universidad de la República

Manuel Caeiro

Universidad de Vigo

Eleni Zazani

Imperial College

Elizabeth Charles

Birkbeck, University of London

Fabrizio Scrollini

Iniciativa Latinoamericana por los Datos Abiertos

Dawn Marsh

University of Waikato

Santiago Martín

University College London

And the members of...

the OKFN Open Education Working Group and the ALT Open Education SIG.

# Institutional Acknowledgements

This book has been made with the support of

