Open Research Online

The Open University's repository of research publications
and other research outputs

Open Data & Education

Book Section

oro.open.ac.uk

Open Data & Education

Javiera Atenas (ILDA), Leo Havemann (University College London)

Key Points

- Open data can help researchers and policy makers understand the education landscape, can provide information for parents and children about education facilities and their performance, and can be used as an input into education: making a connection between open data and Open Educational Resources (OER).

- Attention must focus beyond the simple availability of education data to also question how the data is shaped, presented, and used. This should address the ways in which, without wider policy interventions, making data available about education performance may ultimately reinforce stigma and social divides.

- There has been relatively limited overlap between OER and Open Data communities, although since 2013, the Open Knowledge Education Working Group has sought to build connections between them. There are opportunities for future strengthening of these links, increasing the use of open data as a key educational resource, and supporting more applied civic education.


Introduction

According to the UN Sustainable Development Goals, Goal 4[1], states must "ensure inclusive and quality education for all and promote lifelong learning". In this chapter, we consider the ways in which open data can support the achievement of this goal. For example, the SDG indicators database provides data which can assist in assessing the state of the art in educational achievement. In a wider sense, the open data released by governments, educational institutions and national and international organisations can support a wide range of interventions, and can support stakeholders in adopting strategies to: improve the quality of education; design effective policies; and foster the development of the key literacies needed to operate and participate in today's 'datafied society'[2].

Education is a complex network of systems and practices, entangled with a series of socio-political and economic problems. Despite the efforts made by governments to promote equal access to quality education, there are still challenges to overcome, for which open data can provide a landscape perspective and a range of opportunities to better understand the core educational problems facing societies. It has also been argued that open data can be used as part of quality indicators to help people make better decisions (within their possibilities) choosing between educational centres. More overtly, open data, used as Open Educational

---

[1] United Nations (2018) Sustainable Development Goals: Education. Accessed November 14. https://www.un.org/sustainabledevelopment/education/

[2] Schafer, Mirko Tobias, & Karin van Es, eds. (2017) *The Datafied Society: Studying Culture Through Data*. Amsterdam: Amsterdam University Press. http://www.oapen.org/search?identifier=624771

Resources (OER), can be considered a key tool in developing the transversal literacies that citizens (and indeed, others) require in order to participate democratically in today's datafied societies. The figure below shows three main ways in which open data and the broad education sector intersect. You can also think of this in terms of open data use by policy makers, open data for students and parents, and open data for use by educators.

Figure 1: Open Data in Education
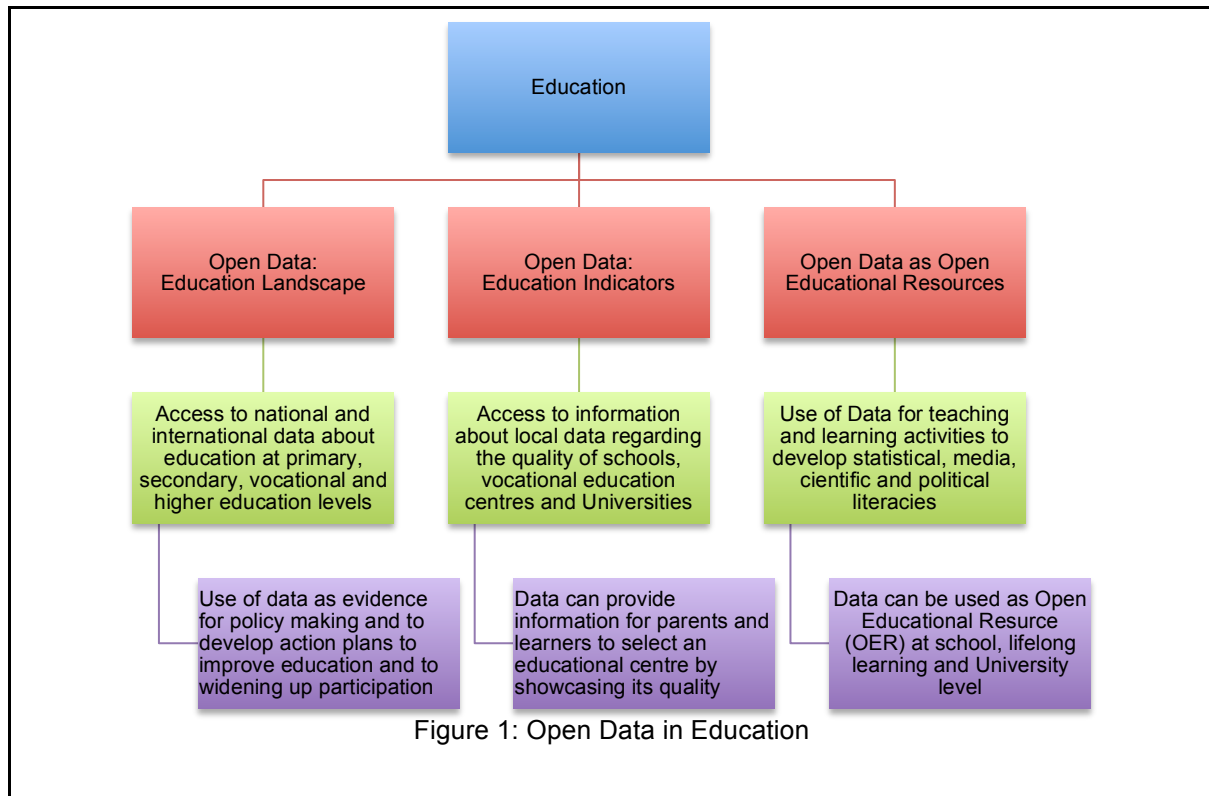
In this chapter we explore both the opportunities and challenges that Open Data presents across the educational sector, drawing upon examples from across the world, and wider critical arguments about open data[3]. We do so aware that, while open data can promote public participation and social innovation, it can also widen prejudice and stigmatise the poor and vulnerable, helping to further marginalise those who, for example, cannot choose where to live or to study. The evidence we survey suggests that, to date, the impact of open data initiatives in the education space has been mixed, and there are opportunities to substantially strenghen networks and actions around open data and education in future.

---

[3] Kitchin, R. (2013) Four Critiques of Open Data Initiatives. *LSE Impact Blog*.
http://blogs.lse.ac.uk/impactofsocialsciences/2013/11/27/four-critiques-of-open-data-initiatives/

Open Data and the Education Landscape

Understanding the current state of education systems, and identifying ways to improve education, are vital tasks for policy makers Davies[4]; Niemi[5]; Burns and Köster[6] and the 2017 EU Eurydice Report[7] all argue that policy makers need better access to evidence in order to address policy issues. Data that describes achievements, attainment, enrolment, or distribution of learning are all important to determine whether the educational system is working or not. UNESCO[8] have called for grounding policy solid on reliable evidence to ensure that policies are effective, efficient and implementable. They argue for use of comparable indicators and ensuring that such data are available disaggregated by subgroups such as gender, administrative area, geographical location, socio-cultural groupings, education level and type of providers, to compare different groups and to identify those who are educationally disadvantaged.

Motivans[9], in exploring data availability to monitor the SDGs, also calls for education data that is relevant, valid, reliable, timeless, punctual, clear, transparent, comparable, accessible, affordable and consistent and have potential for disaggregation. There has been some progress on making this data available (and open), but major gaps remain. Notably, educational data from countries as Kenya, South Africa, Ecuador, or Montenegro[10] is scarce, and neither widely nor openly available, making it difficult to assess their progress in relation with SDG 4.

While some states have had standardised testing since the 1950s, it is only in the last 20 years that standard national assessments have become the norm rather than exception in Europe, and the majority of the world's population are still in countries without such testing[11]. International initiatives have stepped in to fill the gap. The best known example of performance data provided at international level is the the OECD's Programme for

---

[4] Davies, P. (1999) What Is Evidence-Based Education? *British Journal of Educational Studies* 47, no. 2 (June 1): 108–121. https://www.tandfonline.com/doi/abs/10.1111/1467-8527.00106

[5] Niemi, H. (2007) Equity and Good Learning Outcomes. Reflections on Factors Influencing Societal, Cultural and Individual Levels. The Finnish Perspective. *Zeitschrift Für Pädagogik* 53, no. 1: 92. https://www.pedocs.de/volltexte/2011/4389/pdf/ZfPaed_2007_1_Niemi_Equity_good_learning_D_A.pdf

[6] Burns, T., & F. Köster (2016) *Educational Research and Innovation Governing Education in a Complex World*. OECD Publishing. http://dx.doi.org/10.1787/9789264255364-en

[7] European Commission/EACEA/Eurydice (2017) *Support Mechanisms for Evidence-Based Policy-Making in Education. Eurydice Report.* Luxembourg: Publications Office of the European Union. https://eige.europa.eu/resources/206_EN_Evidence_based_policy_making.pdf

[8] UNESCO Bangkok (2013) UNESCO Handbook on Education Policy Analysis and Programming, Volume 1: Education Policy Analysis; 2013. http://unesdoc.unesco.org/images/0022/002211/221189E.pdf

[9] Motivans, A. (2015) Improving Education Statistics Systems: Challenges and Opportunities. Presentation at *World Statistics: Sustainable Data for Sustainable Development*. Xi'an, People's Republic of China. https://unstats.un.org/sdgs/files/meetings/sdg-seminar-xian-2015/Presentation--3.4-Sustainable-Data-for-Sustainable-Development--UNESCO.pdf

[10] See http://kenya.opendataforafrica.org/gallery/Education, http://www.statssa.gov.za/?cat=16, https://educacion.gob.ec/estadisticaseducativas/ and http://www.mpin.gov.me/en/ministry respectively. Accessed 14th November 2018.

[11] Sandefur, J. (2018) The Case for Global Standardized Testing. Accessed November 14. https://educationinnovations.org/case-for-global-standard-testing

International Student Assessment (PISA) test,[12] initiated in 2000, and providing data about learner performance in science, mathematics and reading. The results of this test, linked to socio-demographic data, enable comparative analysis regarding differences in performance by diverse groups of learners (taking into account gender, social background, migrant learners and ethnicity). In 2015, 72 countries participated in the PISA survey, generating data that is commonly used in evidence-based policymaking, helping educational stakeholders to target specific problems guided by a clear set of information. Individual (anonymous) student results from the study are published to download as structured data.

When reasonably disaggregated data is made available openly then a wide range of actors can get involved in its analysis. Academics are clearly a major user of education data, but private consultancies and non-profit organisations have also taken advantage of available datasets. In the UK, for example, the Education Datalab[13] was established by a non-profit education services company to help policymakers to help improve educational practice. International organisations such as the Organisation for Economic Co-operation and Development (OECD), the United Nations Educational, Scientific and Cultural Organization (UNESCO) and the World Bank also use data (combined with qualitative research) to provide international collections of policies, presentations of educational data, policy tools and frameworks intended to support evidence-based policy-making. Van Schalkwyk has also drawn attention to the way in which institutions providing performance data (in this case Higher Education Institutions in South Africa[14]) take advantage of cross-institution comparisons for benchmarking, and how making more granular information available as open data has provided 'a new fuel for transformation'.[15]

However, when approaching educational data for research and policy purposes there are at least two important considerations to keep in mind. Firstly, educators' and learners' privacy must be protected when using or sharing data: particularly administrative and statistical data containing personally identifiable information. This requires careful balancing, as it to surface and address patterns of educational disadvantage it is important that educational data can be disaggregated by gender, socio-cultural background, educational level and type of school. In the UK, controversy has emerged a number of times over the intrusiveness and level of data disclosure from the National Pupil Database[16].

Secondly, it is important to consider capacity to create and use data, as well as it's availability. In this area, one project to watch the is the CapED initiative.[17] This project, active in 25 less developed countries, aims to connect national education policies,with data sources and to support states in their use of this data in developing national action plans that ensure achievement of the SDG Goal 4 (SDG 4). As each national CapED project works

---

[12] OECD: PISA test data http://www.oecd.org/pisa/data/

[13] https://ffteducationdatalab.org.uk/, Accessed 14th Nov 2018.

[14] van Schalkwyk, F., M. Willmers, & M. McNaughton (2016) Viscous Open Data: The Roles of Intermediaries in an Open Data Ecosystem. *Information Technology for Development* 22, no. sup1 (August 30): 68–83

[15] van Schalkwyk, F. (2017) Open Data on Universities – New Fuel for Transformation. *University World News*. http://www.universityworldnews.com/article.php?story=20170710104034491

[16] Wikipedia. (2018) National Pupil Database. Wikipedia. https://en.wikipedia.org/w/index.php?title=National_Pupil_Database&oldid=832938465

[17] See https://en.unesco.org/themes/education/caped

with UNESCO's Institute of Statistics to implement a data component there may be opportunities to further emphasise open data approaches.

When micro-data cannot be disclosed then the design of indicators that describe the data landscape is also of crucial importance. At national level, one example that demonstrates this is the Data Chile education indicators site[18] that provides information from the National System of Performance Evaluation (SNED). This is constructed using 6 indicators on school effectiveness, improvement, initiative, improvement of working conditions, equal opportunities, and the integration of teachers, parents and guardians. In an open data context, it is important to think about who gets involved in defining the indicators that will shape the sources of data that will be available in future.

Data Chile - SNED data - Integration
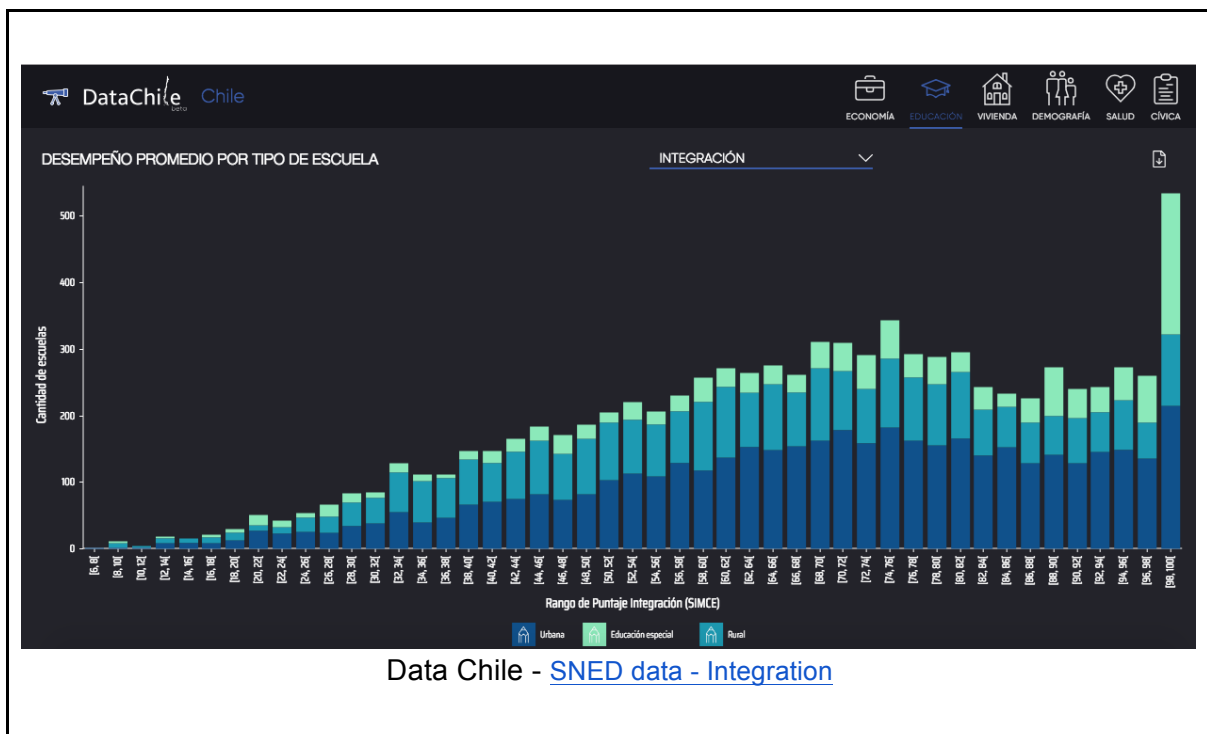
In summary: demand is high for data on the education landscape - but supply varies. When open data is available, established policy players can be joined by new actors, including entrepreneurs and journalists, to debate and shape education performance and policy. However, even in the absence of globally comparable data, or policy maker use of that data, datasets on educational institutions can also drive change through parent and pupil behaviours.

---

[18] Data Chile - Education Indicators https://es.datachile.io/geo/chile#education

Open Data about Educational Institutions

In many countries, parents or pupils have some degree of choice over educational institutions. Statistics have long played a role in that decision making, and with availability of open data, a range of interactive platforms have emerged that use school or university data to inform parents and learners, providing them with the indicators and information they may use to make informed choices about different educational centres.[19] [20] Notably, the data made available about educational centres tends to focus on performance, as, for example, in university ratings and rankings that aim to portray the quality of education using some standardised metrics.

When it comes to schools, the last decade has seen the launch of numerous portals around the world using data provided by national and local authorities regarding education quality. Some examples of this kind include the government run Identicole portal in Peru ; and MIME from the Ministry of Education in Chile, JedeSchule run by two non-profit in Germany, the mobile app-based Conozca su escuela run in Costa Rica by the Programa Estado De La Nación and Scholen Keuze and Scholen op de kaart run in the Netherlands by a private firm and a coalition of funding bodies respectively[21].

A number of platforms go beyond using data to encourage 'shopping around' to choose schools. For example, Mejora tu escuela in Mexico[22], created by El Instituto Mexicano para la Competitividad (IMCO) with funding from Omidyar network, places emphasis on gathering feedback from users of the platform, and equipping them to advocate for improvements to their existing schools. In the UK, School Cuts[23], created with the backing of major teachers unions, places emphasis on how funding cuts for education are impacting individual schools and was used as an advocacy tool in the last election. One of the unions funding the project claimed it helped to change '750,000 votes at the election, and resulted in the government stumping up another £1.3bn for schools in July'.[24]

The vast majority of platforms, however, focus on maps and rankings. Figure 2 shows two further examples from the UK. The first one, School Atlas, was developed by the Mayor of London and highlights the degrees of income deprivation affecting children across London. The second example is a map of schools in London developed by a private firm Locrating Ltd, cross referencing data from Ofsted (the inspector of schools) and the Department of

[19] McAuley, D., H. Rahemtulla, J. Goulding, & C. Souch (2011) How Open Data, Data Literacy and Linked Data Will Revolutionise Higher Education. In *Pearson Blue Skies: New Thinking About the Future of Higher Education*. https://www.academia.edu/1945211/The_Open_Data_Revolution_and_Data_Literacy_in_Higher_Education

[20] Guy, M. (2016) The Open Education Working Group: Bringing People, Projects and Data Together. In *Open Data for Education: Linked, Shared, and Reusable Data for Teaching and Learning*, ed. D. Mouromtsev & M. d'Aquin, 166–187. Lecture Notes in Computer Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-30493-9_9

[21] See http://identicole.minedu.gob.pe/, http://www.mime.mineduc.cl/mvc/mime/portada, https://jedeschule.de/, https://www.conozcasuescuela.ac.cr/, http://www.scholenkeuze.nl/ and https://www.scholenopdekaart.nl/ respectively. Accessed 14th November 2018.

[22] See http://www.mejoratuescuela.org/. Accessed 14th November 2018.

[23] See http://www.schoolcuts.org.uk/. Accessed 14th November 2018.

[24] Whittaker, F. (2018) NUT Spent £326k on General Election Campaign. Schools Week, March 19. https://schoolsweek.co.uk/nut-spent-326k-on-general-election-campaign/

Education (England), in order to present comparative information on school quality. This is expressed in the form of an overall school rating of 'inadequate', 'requires improvement', 'good' or 'outstanding'. The schools can be viewed on a map which integrates property search data, so that parents can assess the cost of homes in the vicinity of desirable schools. While such ratings are presented as a straightforward solution to the dilemma of school choice, the reduction of complex phenomena to a single rating category obscures rather significant questions about the data that exists and that has been selected for inclusion. Consequently, the ratings expressed carry embedded biases, which act to reinforce privilege, stigmatisation and disadvantage.
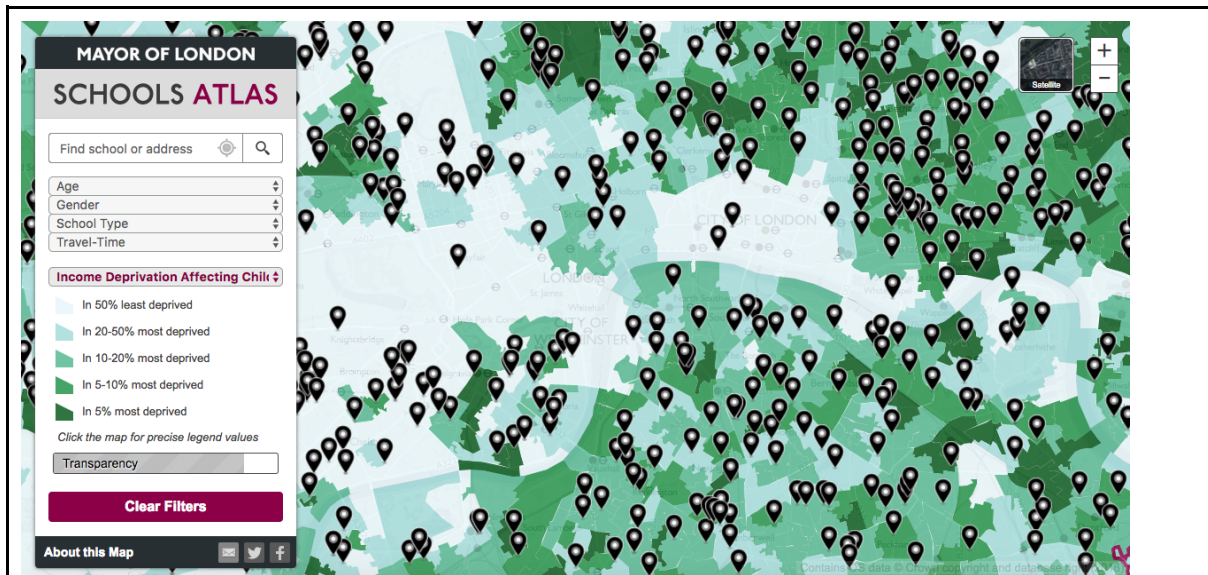
Mayor of London - London Schools Atlas
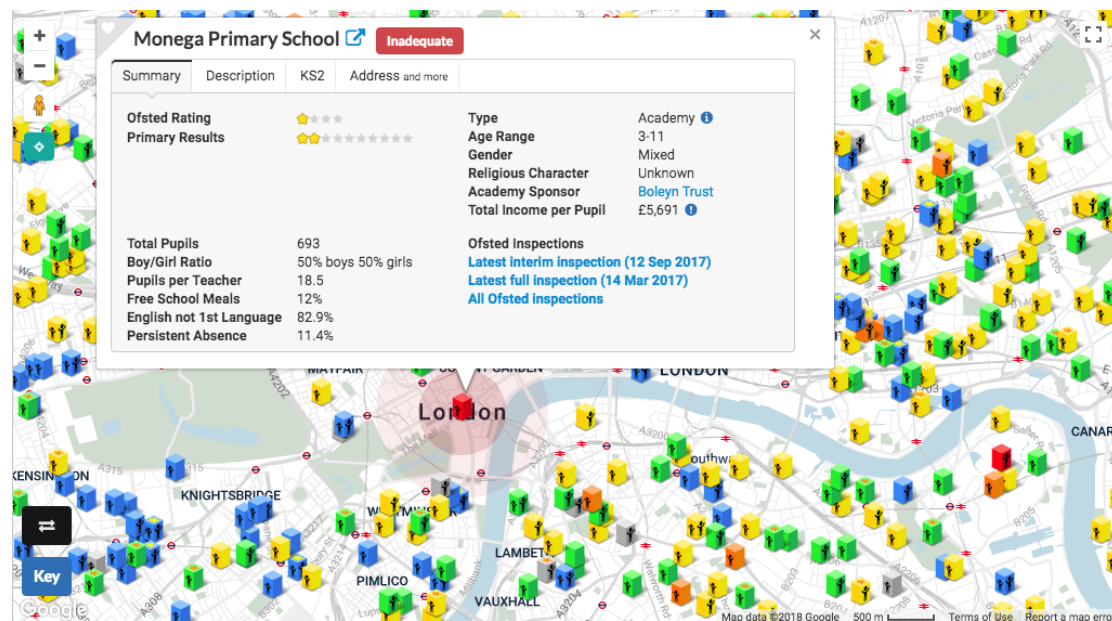
Locrating - Schools A-Z London

Figure 2: examples of school information platforms

Both examples offer some ideas of how the quality of education is commonly portrayed and how the apparent authority and reliability of statistical data can act to stigmatise pupils from schools rated as 'inadequate' or located in income-deprived areas. There is a risk that such information can be used to profile individuals as poorly educated, or affect their future prospects.

We need to consider critical ethical questions when making data available about schools, at the very least making sure performance data is accompanied with contextualised information about the socio-economic challenges faced by its community such as poverty, integration and inclusion. Whilst school information portals are an appealing concept, there is limited evidence to suggest they may ultimately improve education without being linked to wider work on participation and advocacy, and support for actors to hold governments accountable to ensure that there is proper funding for quality education for the most vulnerable children in a society. When it comes to open data on educational institutions we have both supply and demand, and active intermediaries able to sustain their platforms. Yet whilst there may be cases of individual impact for particular pupils, the net social impact is much harder to find.

Open Data as Open Educational Resources

The final application of open data in education we survey is the pedagogic use of open data (on any topic) as Open Educational Resources (OER), which are defined by UNESCO[25] as 'any type of educational materials that are in the public domain or introduced with an open license'. Open data, used as OER, allows students to learn and experiment by working with the same raw data that is generated and used by researchers, governments, civil society, international organisations, and policymakers. Such datasets can form a key component in research- and scenario- based learning activities, and in supporting students to develop information, statistical, scientific, curation, media, political and critical thinking skills. Through working with real-world data, students can develop storytelling and research skills, and can apply analytical, collaborative and citizenship skills in using data to investigate real-world problems.

This idea of using open data in education is recognised in the sixth principle of the Open Data charter[26] on open data for inclusive development and innovation, which states (point 5.d) that it is key to "Engage with schools and post-secondary education institutions to support increased open data research and to incorporate data literacy into educational curricula". Although it is not clear how much emphasis has been placed to date on this point by countries and cities adopting the Charter, the groundwork to support use of open data as OER has been laid in a number of projets.

---

[25] UNESCO (2017) Open Educational Resources (OER). *UNESCO*.
https://en.unesco.org/themes/building-knowledge-societies/oer
[26] Open Data Charter (2015) *International Open Data Charter*. https://opendatacharter.net/principles/

In 2015, the Open Education Working Group of Open Knowledge International (established in 2013) published a book of case studies of open data as OER,[27] in which a series of authors present activities that can be adopted by educators at school and university level promoting the use of open data in research-led activities. The book provides examples and highlights good practices in order to showcasing how, by using real data from research outcomes, and from national and international data-led projects, educational activities can be fostered to develop data literacies alongside critical thinking skills by bridging collaborations amongst students, researchers and academics. One example of good practices portrayed in the book is A Scuola di Open Coesione,[28] which is an educational challenge designed for Italian high-school students. It was funded through the open government strategy on cohesion policy, in partnership with the Ministry of Education and the Representation Office of the European Commission in Italy.

Some other good practices in the use of open data as OER[29] can be found from Open Data School in Russia which provides a series of lectures and seminars from experts on open data topics.The UTPL Open Linked Data project in Ecuador presents the result of the study and application of Linked Data technology for students, researchers and educators. And Data Science Fundamentals from Palestine offers an online tool to enable students to follow the Foundations of Data Science training course developed by students and academics from Birzeit University. Lastly, Monithon, also from Italy, offers an example of applied learning through open data, which sees engaged citizens and university students work alongside researchers and policymakers to monitor development projects. It is notable, however, that many initiatives on open data as OER have been relatively short-lived, and that the connections between Open Education and Open Data communities are relatively weak, with few points at which the communities overlap.

Supporting use of open data as an OER is closely linked to work on data literacy (see Data Literacy chapter in this volume). Recently, the Latin American Initiative for Open Data (ILDA), have developed a training programme for academics in the use of Open Data for teaching and learning to support them in developing students' capacities for analysing information critically from wide range of sources, including data,[30] developing the literacies needed to understand and interpret data, and therefore, living and working in the datafied society.

Following Uhlir & Schröder's argument[31] that "Students may be less effectively educated and trained if they are unable to work with a broad cross-section of data" (p.201), and Davies'[32]

---

[27] Atenas, Javiera, & Leo Havemann, 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

[28] See http://www.ascuoladiopencoesione.it/about-opencoesione-school/. Accessed 14th November 2018.

[29] See http://opendataschool.ru/anketa/, http://data.utpl.edu.ec/, https://github.com/abedkhooli/ds1 and http://www.monithon.it/

[30] Atenas, J., L. Havemann, & E. Priego (2015) Open Data as Open Educational Resources: Towards Transversal Skills and Global Citizenship. *Open Praxis* 7, no. 4 (November 26): 377–389 http://dx.doi.org/10.5944/openpraxis.7.4.233

[31] Uhlir, P.F., & P. Schröder (2007) Open Data for Global Science. *Data Science Journal* 6: OD36–OD53 https://www.jstage.jst.go.jp/article/dsj/6/0/6_0_OD36/_pdf

assertion that "there will be greater need in future for capacity both in state and society to be able to debate the meaning of data, and to find responsible ways of using open data in democratic debate" (p. 5), we consider that the inclusion of open data in curricular activities is key to ensure that both educators and learners acquire the skills they need to participate in contemporary society.

Conclusions

Over the last 10 years, open data availability has grown: both data about education, and data that can be used within education. Whilst 10 years ago, looking for school performance information may have involved using tables published once a year in newspapers, now many countries have interactive websites offering analysis and visualisation, ranging from official government sites, to privately-operated portals. Schools and higher education institutions no longer need to rely on tables in textbooks, but can go to real-world updated datasets for teaching and learning. However, whilst this can look like great progress, challenges remain.

Although open data can provide evidence about problems that need to be addressed at the policy level, and can be a key component in the development of the literacies needed in a datafied society, enhancing and promoting civic participation and understanding of the media and the sciences, open data cannot be considered as the panacea per se for educational problems, most particularly because data generation, collection and analysis can be heavily biased.

Data is never neutral and it is ultimately a political instrument.The algorithms used to analyse it can prompt stigmatisation, segregation and discrimination, by portraying people as of 'less quality'. For example, data about quality of schools may lead to the gentrification of so-called cheaper areas with good schools, pushing up the housing prices and pricing the less privileged out. Mainstream narratives may place the blame for apparent poor quality on the children that do not score as highly in standardised tests, instead of pointing at the authorities who have failed to provide policies and funding needed to improve schools.

Arguments for opening of data in education have tended to focus on the importance of access to data. Such arguments can gloss over the non-neutrality of data and the potential threats inherent in data-driven decision making, where the context or agenda of data collection and presentation is opaque or where data 'consumers' lack sufficient criticality towards the data on offer. They often also ignore trends towards marketisation of education, and the wider socio-economic context. We do not believe that open data advocacy should seek to present data as innocuous and benign. As Kitchin[33] states, 'if open data merely serves the interests of capital by opening public data for commercial re-use and further empowers those who are already empowered and disenfranchises others, then it has failed to make society more democratic and open'. However, as we have seen above, in examples

---

[32] Davies, T. (2010) Open Data, Democracy and Public Sector Reform: A Look at Open Government Data Use from Data.Gov.Uk. University of Oxford. http://www.opendataimpacts.net/report/wp-content/uploads/2010/08/How-is-open-government-data-being-used-in-practice.pdf

[33] Kitchin, R. (2013) Four Critiques of Open Data Initiatives. *LSE Impact Blog*. http://blogs.lse.ac.uk/impactofsocialsciences/2013/11/27/four-critiques-of-open-data-initiatives/

like SchoolCuts.org, it is not only private interests that can deploy data for implicit or explicit political ends; there is potential for critical action.

Ultimately, whilst there are many challenges in the ways open data about education may be used, it is through wider education about the creation and use of open data that these risks can be best addressed. The wealth of open data on all topics, when deployed as OER, can be part of this. In conclusion, we recommend that:

- In the use and development of education indicators, it is important to prevent its analysis exclusively by algorithms, as these tend to contain bias and can foster stigmatisation of vulnerable students.

- When governments open up educational data available, they must keep it anonymised ensuring that individuals and collectives cannot be identified, and also, consider the potential uses of this data by public and private stakeholders, to prevent this data to be used unethically.

- When private stakeholders and organisations (including the civil society) build tools using educational data they need to consider the potential impact that their activities can have for the students, educators and educational communities, foreseeing if their project can benefit or punish a group of people.

- And finally, to foster data and citizenship literacies, the Open Education, Open Data and Open Science communities must partner up to develop educational programmes and curricula to provide training at all educational levels including training for educators and educational communities.

Further reading

Atenas, Javiera, & Leo Havemann, eds. (2015) Open Data as Open Educational Resources: Case Studies of Emerging Practice. London: Open Knowledge, Open Education Working Group.

Schafer, Mirko Tobias, & Karin van Es, eds. (2017) The Datafied Society: Studying Culture Through Data. Amsterdam: Amsterdam University Press.

Mandinach, E.B., & E.S. Gummer (2013) A Systemic View of Implementing Data Literacy in Educator Preparation. Educational Researcher 42, no. 1 (January 1): 30–37

About the authors

[Javiera Atenas](#) (ILDA)
Javiera is the Education lead at ILDA (The Latin America Open Data Initiative) and co-coordinator of the Open Knowledge Open Education Working Group. She is an active researcher & advocate of open practices.

[Leo Havemann](#) (University College London)
Leo is a postgraduate Researcher at the Open University and is Digital Education Advisor at UCL (and formerly, learning technologist at Birkbeck). His interests focus on use of technology in higher education, open movements, open educational practices and learning literacies.

Suggested citation

Atenas, Javiera; Havemann, Leo (2018) "Open Data & Education" In *The State of Open Data* edited by Tim Davies, Stephen Walker, Mor Rubinstein, and Fernando Perini (forthcoming in 2019)