

006

Education

Javiera Atenas and Leo Havemann

Key points

- Open data can help researchers and policy-makers understand the education landscape, provide information for parents and children about education facilities and their performance, and serve as a key element in the creation of open educational resources (OER).
- Attention must move beyond the simple availability of data on education to also question how the data is contextualised, presented, and used to ensure it does not result in the reinforcement of pre-existing biases and social divides.
- There has been relatively limited intersection to date between the open education and open data communities. 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 United Nations (UN) Sustainable Development Goal (SDG) 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. In the education sector, open data released by governments and educational institutions, as well as by national and international organisations, can support a wide range of interventions, including strategies to improve the quality of education, the design of effective education policies, the creation of educational resources, and the development of the key literacies needed to operate and participate in today’s “datafied society”²

The education ecosystem is made up of a complex network of systems and practices developed to address a wide range of sociopolitical and economic issues. Despite the enormous efforts made by countries to guarantee equal access to quality education, there are still challenges to overcome

for which open data can provide insight, perspective, and a wide range of tools to further our understanding of core educational problems and to support the development of solutions. It has also been argued that open data can be used as part of a series of quality indicators to help people to make better decisions related to educational opportunities and methodologies and to choose among education providers. More overtly, open data used in the development of open educational resources (OER) can be considered a key tool in promoting citizenship and democratic values and developing the transversal literacies that citizens require in order to participate in a datafied society. Figure 1 indicates three main ways in which open data and the broader education sector intersect. You can also think of this in terms of how open data use intersects with the three main education stakeholder groups: policy-makers, parents and learners, and educators.

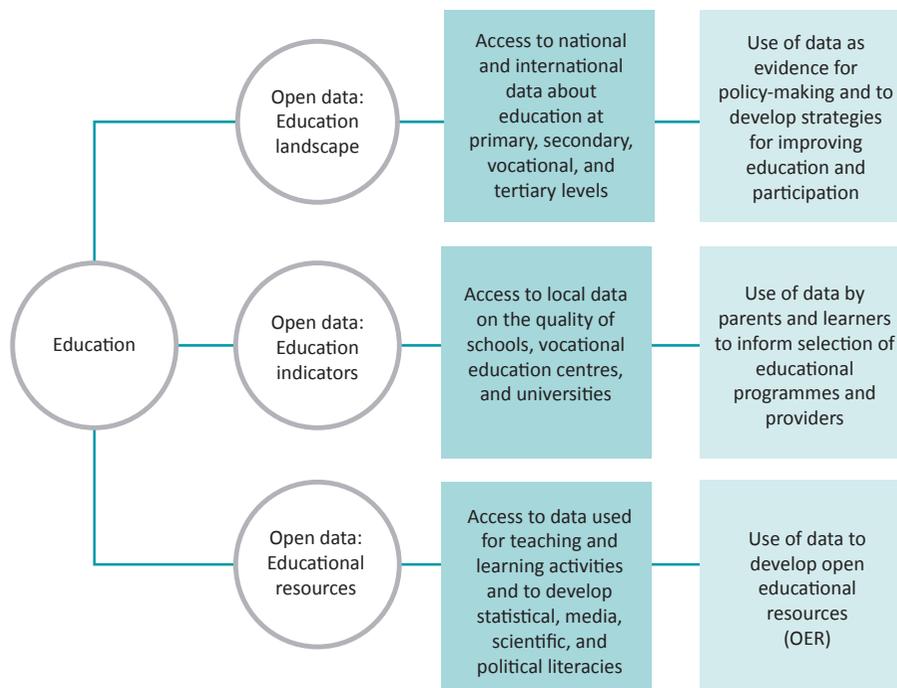


Figure 1: Open data in education
Source: Authors

In this chapter, we will explore both the opportunities and challenges that open data presents across the education sector, drawing upon examples from around the world, and wider critical arguments and studies related to open data.³ We are aware that while open data can promote public participation and social innovation, it can also reinforce pre-existing biases by connecting performance with the poor and vulnerable in an unfair manner, helping to further marginalise those who cannot choose where to live or study. The evidence we have gathered suggests that although impact to date has been mixed, there are many opportunities to substantially strengthen existing networks and activities around open data and education in the future.

Open data and the education landscape

Understanding the current state of education and identifying ways to improve education, are vital tasks for policy-makers. Davies,⁴ Niemi,⁵ Burns, Köster,⁶ and the 2017 EU Eurydice Report⁷ argue that policy-makers need better access to evidence in order to address policy issues. Data that describes achievements, attainment, enrolment, or the distribution of learning are all important to determine whether educational systems are working or not. The United Nations Educational, Scientific and Cultural Organization (UNESCO)⁸ has indicated the need to ground policy on reliable evidence to ensure that educational policies are effective, efficient, and implementable. They argue for the use of comparable indicators and for ensuring that data is available disaggregated by gender, administrative area, geographical location, sociocultural groupings, education level, and type of provider to enable a comparison between the different groups and to identify those who are educationally disadvantaged.

Motivans,⁹ in exploring data availability to monitor the SDGs, also calls for educational data that is relevant, valid, reliable, timeless, punctual, clear, transparent, comparable, accessible, affordable, consistent, and with potential for disaggregation. There has been some progress on making this data available (and open), but major gaps remain. Notably, educational data from countries such as Kenya, South Africa, Ecuador, or Montenegro¹⁰ is scarce and neither widely nor openly available, making it difficult to assess their progress in relation to 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 in Europe, and the majority of the world's population still resides in countries without such testing.¹¹ International initiatives have stepped in to fill the gap. The best-known example of performance data provided at the international level is the Organisation for Economic Co-operation and Development's (OECD) Programme for International Student Assessment (PISA) test¹² initiated in 2000, providing data about learner performance in science, mathematics, and reading. The results of this standard test, linked to sociodemographic data, enable comparative analysis regarding differences in performance among 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 policy-making to help educational stakeholders to target specific problems guided by clear information. Individual (anonymous) student results from the study are published in downloadable structured data formats for common statistical software.

When open data is available as disaggregated data then a wide range of actors can get involved in its analysis. Academics are clearly major users of education-related data, but private consultancies and non-profit organisations have also taken advantage of available datasets. For example, in the United Kingdom (UK), the FTT Education Datalab¹³ was established by a non-profit education services company to help policy-makers improve educational practice. International organisations, such as the OECD, UNESCO, and the World Bank, make use of data (combined with qualitative research) to contribute to the international collection of policies, presentations, policy tools, and frameworks intended to support evidence-based policy-making. Van Schalkwyk (2017) has also drawn attention to the way in which institutions providing performance data (in particular, higher education institutions in South Africa¹⁴) 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. First, the privacy of educators and learners must be protected when using or sharing data, particularly administrative and statistical data containing personally identifiable information. Surfacing and addressing patterns of educational disadvantage requires a careful balance because it is important that educational data can be disaggregated by gender, sociocultural 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

Second, it is important to consider the capacity to create and use data, not just its availability. In this area, one project to watch is the CapED initiative.¹⁷ This project, active in 25 of the least developed countries (LDCs), aims to connect national education policies with data sources, and to support states in their use of this data in the development of their national action plans to achieve SDG 4. As each national CapED project works with UNESCO’s Institute of Statistics to implement a data component, there may be opportunities to further emphasise open data approaches.

When microdata cannot be disclosed, the design of indicators that describe the data landscape is also of crucial importance. At the national level, one example that demonstrates this is the Data Chile education indicators site¹⁸ that provides information from the National System of Performance Evaluation (SNED). SNED has been constructed using six indicators: school effectiveness, improvement, initiative, improvement of working conditions, equal opportunities, and the integration of teachers, parents, and guardians (see Figure 2). 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.

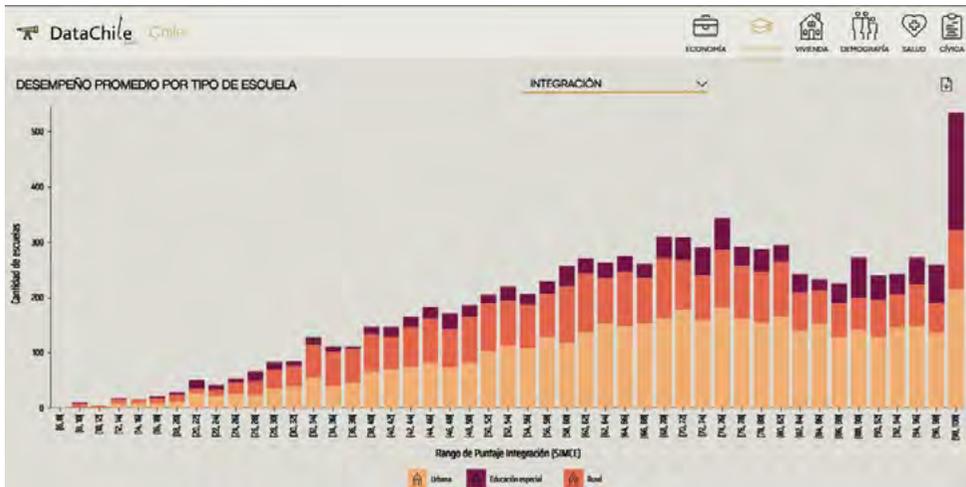


Figure 2: National System of Performance Evaluation (SNED) data: Integration
 Source: DataChile, <https://es.datachile.io/geo/chile#education>

In summary: demand is high for data across the education landscape, but supply varies. When open data is available, established policy-makers 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 the use of that data by policy-makers, datasets on educational institutions can also drive change through parent and pupil behaviours.

Open data about educational institutions

In many countries, parents and/or pupils have some degree of choice over educational institutions. Statistics have long played a role in decisions related to the selection of learning products, programmes, and providers. With the availability of open data, a range of interactive platforms have emerged that use institutional or third-party assessment data to inform parents and learners, providing them with indicators and information they can use to make informed choices.^{19,20} The data made available about educational institutions tends to focus on performance (e.g. university ratings) by using standardised metrics, but also may provide detailed information on programmes and prerequisites.

The last decade has seen the launch of numerous portals around the world that provide the means to compare the quality of education at different institutions using data provided by national and local authorities. Some examples include the Identicole portal in Peru, MIME from the Ministry of Education in Chile, JedeSchule run by non-profit organisations in Germany, the mobile app-based Conozca su escuela in Costa Rica run by Programa Estado De La Nación, and Scholen Keuze and Scholen op de Kaart in the Netherlands.²¹

A number of platforms go beyond using data to encourage “shopping around” in the selection of schools. For example, Mejora tu escuela in Mexico,²² created by El Instituto Mexicano para la Competitividad (IMCO) with funding from the Omidyar Network, places an 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 the emphasis on how funding cuts in 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 during the election and resulted in the government stumping up another £1.3 billion for schools in July”.²⁴ However, the vast majority of platforms focus on maps and rankings. Figures 3 and 4 show two further examples from the UK. The first one, School Atlas, was developed by the Mayor of London and showcases the impact of income deprivation on children in London. The second example is a map of schools in London developed by a private firm, Locrating Ltd, which places the emphasis on school quality, cross-referencing data from Ofsted (the inspector of schools) and the Department of Education (UK). It showcases schools by area, displaying school quality as “inadequate”, “requires improvement”, “good”, or “outstanding”; however, if we look at the data from a critical perspective, we can note the biases this information may portray by reinforcing preconceived notions of privilege and disadvantage.



Figure 3: Mayor of London – Schools Atlas
Source: <https://maps.london.gov.uk/schools/>

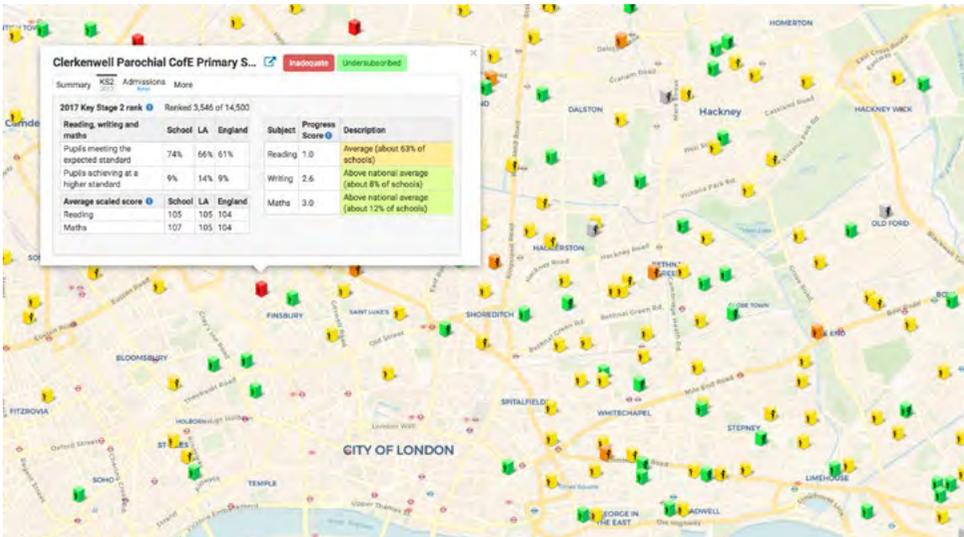


Figure 4: Examples of school information platforms
Source: Locrating, A to Z of Schools, https://www.locrating.com/all_schools.aspx

Both examples offer an illustration of how the quality of education can be portrayed, but, even with contextual data, there is a risk that such information could stigmatise pupils from schools rated as inadequate or in low-income areas. We need to consider critical ethical questions when making data available about schools or, at the very least, ensure performance data is accompanied by contextualised information about the socioeconomic challenges faced by the relevant community, such as poverty, integration, and inclusion.

While school information portals are popular and may support more informed decision-making by learners faced with a complex mix of educational opportunities, there is limited empirical evidence to date on whether they ultimately improve education as much as advocacy-oriented efforts aimed at holding governments accountable or at ensuring proper funding for quality education for the most vulnerable in our society. When it comes to data on educational institutions, we have both ample open data supply and demand, as well as active intermediaries who are able to sustain their platforms. While there may be cases of individual impact for particular learners, the net social impact is difficult to determine.

Open data as an open educational resource

The final application of open data in education is its direct use in the development, or as part, of OER. OER are defined by UNESCO²⁵ as “any type of educational materials that are in the public domain or introduced with an open license”. Open data used as OER can allow students to learn and experiment by working with the same raw data researchers, governments, civil society, international organisations, and policy-makers generate and use. They can form a key component in research- and scenario-based learning activities, and in supporting students to develop informational, statistical, scientific, media, political, and critical-thinking skills. By working with real-world data, students can develop storytelling and research skills, and can apply analytical, collaborative, and citizenship skills in using data to solve real-world problems.

This idea of using open data in education is recognised in the sixth principle of the Open Data Charter²⁶ on open data for inclusive development and innovation, which states that it is key to “[e]ngage 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 the use of open data as OER has been laid in a number of projects.

In 2015, the Open Education Working Group of the Open Knowledge Foundation, established in 2013, published *Open data as Open Educational Resources: Case studies of emerging practice*²⁷ in which a series of authors presented activities that could be adopted by educators at schools and universities to promote the use of open data in research-related activities. The book provides examples and best practices, showcasing how to use real data from research and from national and international data projects to foster educational activities to develop data literacies and critical thinking through collaborations among students, researchers, and academics. One of the practices portrayed in the book is A Scuola di OpenCoesione in Italy,²⁸ an educational challenge, designed for Italian high school students. It was funded under the open government strategy on cohesion policy in partnership with the Ministry of Education and the Representation Office of the European Commission in Italy.

Other practical examples of the use of open data as OER²⁹ can be found at the Open Data School in Russia, which provides a series of lectures and seminars from experts on open data topics. The Open Linked Data project at the Universidad Técnica Particular de Loja in Ecuador presents the results of a study on Linked Data technology for students, researchers, and educators, and Data Science Fundamentals in 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. Finally, Monithon, also from Italy, offers an example of applied learning through open data, which citizens and university students, alongside researchers and policy-makers, use to monitor development projects. However, even with these notable successes, many initiatives focused on the use of open data as OER have been relatively short-lived, and the connections between the open education and the open data communities are still relatively weak with only a few points at which the communities intersect.

Supporting use of open data as OER is closely linked to work on data literacy (see Chapter 19: Data literacy). Recently, the Latin American Initiative for Open Data (ILDA) has developed a training programme for academics in the use of open data for teaching and learning³⁰ to support them in developing the capacities needed to live and work in the datafied society, including learning to construct knowledge and analysing information critically from a wide range of data sources.³¹

Following Uhler and Schröder's argument³² that “[s]tudents may be less effectively educated and trained if they are unable to work with a broad cross-section of data”, and Davies³³ 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”, we consider that the inclusion of open data in curricular activities is key to ensuring that both educators and learners acquire the skills they need to participate in contemporary society.

Conclusion

Over the last ten years, open data availability has grown, including data about education and data that can be used within education. Looking for school performance information may have involved using tables published once a year in newspapers ten years ago, but now many countries have interactive websites offering analysis and visualisation: ranging from official government sites to private sector-managed portals. Schools and post-secondary education institutions no longer need to rely on tables in textbooks, but can go to real-world updated datasets for teaching and learning; however, many challenges remain.

Although open data can provide evidence about problems that need to be addressed at the policy level, it can also be a key component in the development of the literacies needed in a datafied society, as well as in enhancing and promoting civic participation and understanding of the media and the sciences. However, it cannot be considered as the panacea for all educational problems.

Data is never neutral and it is ultimately a political instrument. Data and the algorithms used to analyse it can prompt stigmatisation, segregation, and discrimination. Mainstream narratives may place the blame for poor quality education on the children that perform poorly on standardised tests based on their economic or social background, instead of pointing at the authorities who have failed to provide the policies, programmes, and funding needed to improve the schools those children attend.

Arguments for opening data in education have tended to focus simply 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 for data collection and presentation is opaque or where data “consumers” lack the critical thinking skills needed to interpret the data. They often also ignore the impact of trends toward the marketisation of education. We do not believe that it helps to approach open data as innocuous and benign per se. As Kitchin³⁴ 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, with examples like SchoolCuts.org, it is not only private interests that can deploy data for implicit or explicit political ends and there is potential for critical action.

Ultimately, while there are many challenges around the use of open data for education, 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 that could be applied to OER can be part of this. In conclusion, we recommend that:

- In the use and development of education indicators, it is important to prevent analysis exclusively through the use of algorithms as these may reflect biases and can foster the stigmatisation of vulnerable students.
- When governments open up educational data, they must ensure that it is anonymised to prevent the identification of individuals and collectives and, in addition, consider the potential uses of this data by public and private stakeholders to prevent this data from being used unethically.
- When institutions, civil society, and private sector organisations build tools using educational data, they need to consider the potential impact and use for students, educators, and educational communities.
- And finally, to foster data and citizenship literacies, the open education, open data, and open science communities must collaborate to develop educational materials and curricula to support educational institutions and programmes at all levels, including training for educators and educational communities.

Further reading

Atenas, J. & Havemann, L. (Eds.). (2015). *Open data as Open Educational Resources: Case studies of emerging practice*. London: Open Knowledge, Open Education Working Group. <https://s3-eu-west-1.amazonaws.com/pfigshare-u-files/2396976/BookOpenDataasOpenEducationalResources.pdf>

Mandinach, E.B. & Gummer, E.S. (2013). A systemic view of implementing data literacy in educator preparation. *Educational Researcher*, 42(1), 30–37. <https://journals.sagepub.com/doi/abs/10.3102/0013189x12459803>

Schafer, M.T. & Van Es, K. (Eds.). (2017). *The datafied society: Studying culture through data*. Amsterdam: Amsterdam University Press. <http://www.oapen.org/search?identifier=624771>

About the authors

Javiera Atenas is the Education Lead at the Latin America Open Data Initiative (ILDA) and co-coordinator of the Open Knowledge Open Education Working Group. She holds a PhD in Education and her research focuses on educational policies, open educational practices, data ethics, and data literacies. She is an active researcher and advocate of open practices. You can follow Javiera at <https://www.twitter.com/jatenas>.

Leo Havemann is a Digital Education Advisor at University College London and a postgraduate researcher at the Open University. His interests focus on the use of technology in higher education, open educational practices, and learning literacies. Follow Leo at <https://www.twitter.com/leohavemann>.

How to cite this chapter

Atenas, J. & Havemann, L. (2019). Open data and education. In T. Davies, S. Walker, M. Rubinstein, & F. Perini (Eds.), *The state of open data: Histories and horizons* (pp. 91–102). Cape Town and Ottawa: African Minds and International Development Research Centre. <http://stateofopendata.od4d.net>.



This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) licence. It was carried out with the aid of a grant from the International Development Research Centre, Ottawa, Canada.

Endnotes

- 1 UN (United Nations). (2018). Sustainable Development Goals: 4. Education. <https://www.un.org/sustainabledevelopment/education/>
- 2 Schafer, M.T. & Van Es, K. (Eds.). (2017). *The datafied society: Studying culture through data*. Amsterdam: Amsterdam University Press. <http://www.oapen.org/search?identifier=624771>
- 3 Kitchin, R. (2013). Four critiques of open data initiatives. *LSE Impact Blog*, 27 November. <http://blogs.lse.ac.uk/impactofsocialsciences/2013/11/27/four-critiques-of-open-data-initiatives/>
- 4 Davies, P. (1999). What is evidence-based education? *British Journal of Educational Studies*, 47(2), 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(1), 92–107. https://www.pedocs.de/volltexte/2011/4389/pdf/ZfPaed_2007_1_Niemi_Equity_good_learning_D_A.pdf
- 6 Burns, T. & Köster, F. (2016). *Educational research and innovation: Governing education in a complex world*. Paris: Organisation for Economic Co-operation and Development 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. (2013). UNESCO handbook on education policy analysis and programming. Volume 1: Education policy analysis. Bangkok: United Nations Educational, Scientific and Cultural Organization. <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, 20–22 October 2015, Xi'an, People's Republic of China*. Montreal: UNESCO Institute for Statistics. <https://unstats.un.org/sdgs/files/meetings/sdg-seminar-xian-2015/Presentation--3.4-Sustainable-Data-for-Sustainable-Development--UNESCO.pdf>
- 10 <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>
- 11 Sandefur, J. (2018). The case for global standardized testing. *Center for Education Innovations* [Blog post], 5 May. <https://educationinnovations.org/case-for-global-standard-testing>
- 12 See the OECD's PISA test data <http://www.oecd.org/pisa/data/>
- 13 <https://ffteducationdatalab.org.uk/>
- 14 Van Schalkwyk, F., Willmers, M., & McNaughton, M. (2016). Viscous open data: The roles of intermediaries in an open data ecosystem. *Information Technology for Development*, 22(1), 68–83.
- 15 Van Schalkwyk, F. (2017). Open data on universities: New fuel for transformation. *University World News*, 14 July. <http://www.universityworldnews.com/article.php?story=20170710104034491>
- 16 UK Data Service. (2019). National pupil database. <https://beta.ukdataservice.ac.uk/datacatalogue/series/series?id=2000108>
- 17 <https://en.unesco.org/themes/education/caped>
- 18 <https://es.datachile.io/geo/chile#education>
- 19 McAuley, D., Rahemtulla, H., Goulding, J., & Souch, C. (2011). How open data, data literacy and linked data will revolutionise higher education. In L. Coiffait (Ed.), *New thinking about the future of higher education* (pp. 88–93). London: Pearson. 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 D. Mourmontsev & M. d'Aquin (Eds.), *Open data for education: Linked, shared, and reusable data for teaching and learning* (pp. 166–187). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-30493-9_9

- 21 <http://identicole.minedu.gob.pe/>, <http://www.mime.mineduc.cl/mvc/mime/portada>, <https://jedeschule.de/>, <https://www.conozcasuescuola.ac.cr/>, <http://www.scholenkeuze.nl/>, and <https://www.scholenopdekaart.nl/>, respectively.
- 22 <http://www.mejoratuiescuola.org/>
- 23 <http://www.schoolcuts.org.uk/>
- 24 Whittaker, F. (2018). NUT spent £326k on general election campaign. *Schools Week*, 19 March. <https://schoolsweek.co.uk/nut-spent-326k-on-general-election-campaign/>
- 25 UNESCO (United Nations Educational, Scientific and Cultural Organization). (2017). Open Educational Resources (OER). <https://en.unesco.org/themes/building-knowledge-societies/oer>
- 26 Open Data Charter. (2015). Principles: International Open Data Charter. <https://opendatacharter.net/principles/>
- 27 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>
- 28 <http://www.ascuoladiopencoesione.it/about-opencoesione-school/>
- 29 <http://opendataschool.ru/anketa/>, <http://data.utpl.edu.ec/>, <https://github.com/abedkhooli/ds1>, and <http://www.monithon.it/>
- 30 <https://idatosabiertos.org/en/investigaciones-2/modelo-docente-y-datos-abiertos/>
- 31 Atenas, J., Havemann, L., & Priego, E. (2015). Open data as Open Educational Resources: Towards transversal skills and global citizenship. *Open Praxis*, 7(4), 377–389. <http://dx.doi.org/10.5944/openpraxis.7.4.233>
- 32 Uhler, P.F. & Schröder, P. (2007). Open data for global science. *Data Science Journal*, 6, OD36–OD53, p. 201. https://www.jstage.jst.go.jp/article/dsj/6/0/6_0_OD36/_pdf
- 33 Davies, T. (2010). Open data, democracy and public sector reform: A look at open government data use from data.gov.uk. Master's Thesis. University of Oxford, UK, p. 5. <http://www.opendataimpacts.net/report/>
- 34 Kitchin, R. (2013). Four critiques of open data initiatives. *LSE Impact Blog*, 27 November. <http://blogs.lse.ac.uk/impactofsocialsciences/2013/11/27/four-critiques-of-open-data-initiatives/>