The use of open data as a material for learning

Journal Item

RESEARCH ARTICLE

The use of open data as a material for learning

Tim Coughlan[1]

Abstract

Open data has potential value as a material for use in learning activities. However, approaches to harnessing this are not well understood or in mainstream use in education. In this research, early adopters from a diverse range of educational projects and teaching settings were interviewed to explore their rationale for using open data in teaching, how suitable activity designs could be achieved, and the practical challenges of using open data. A thematic analysis was conducted to identify patterns and relationships in these open data-based practices that have already emerged. A document analysis of teaching materials and other related artefacts was used to augment and validate the findings. Drawing on this, common approaches and issues are identified, and a conceptual framework to support greater use of open data by educators is described. This paper also highlights where existing concepts in education and educational technology research, including inquiry-based learning, authenticity, motivation, dialogue, and personalisation can help us to understand the value and challenges of using open data in education.

Keywords Open data · Open education · Authentic learning · Inquiry-based learning · Project-based Learning · Personalisation

Introduction

Open data is "digital data that is made available with the technical and legal characteristics necessary for it to be freely used, reused, and redistributed by anyone, anytime, anywhere" (Open Data Charter 2015, para. 1). This concept has proliferated and made diverse data sets publicly available from various organisations, including government spending, research outcomes, archives, surveys, environmental sensing, and museum collections. While sharing data online is not new, the formalisation of open data has created a global movement and led to a huge growth in the breadth and depth of data available. More than 250 governments, and global and local organisations, have initiatives that create and share open data (World Bank 2017), with more than 17,000 data sets of international relevance from the World Bank alone (World Bank 2018).

✉ Tim Coughlan
   tim.coughlan@open.ac.uk

1   Institute of Educational Technology, Learning and Teaching Innovation, The Open University,
    Walton Hall, Milton Keynes MK7 6AA, UK

🊑 Springer

To date, potential uses of open data in learning are not well understood. However, several drivers could add impetus to this. Within the open data movement there is a desire to find ways to broaden public engagement (Shadbolt et al. 2012). There are also wider calls for data literacies to be considered as a key twenty first century skill set (Wolff et al. 2016). The Open Data Charter, which has been adopted by 17 national and 35 local governments to date, includes a principle to "engage with schools and post-secondary education institutions to support increased open data research and to incorporate data literacy into educational curricula" (Open Data Charter 2015, "Principle 6", para. 5). In response, this paper asks how open data could become a valuable part of education by analysing the perspectives of early adopters of open data in teaching.

Openness already has a substantial history in education (Weller 2015) which has only recently been linked with open data. It is therefore important to compare and contrast how openness has emerged in relation to data and to educational resources. Atenas and Havermann (2015) suggest that when open data is used in education it can be considered as a form of open educational resources (OER), given the similar licensing and definitions. However, they note that the use of open data in this way is under-discussed. OER is specifically defined as "free and openly licenced educational materials that can be used for teaching, learning, research and other purposes" (Creative Commons 2016, para. 1). OER emerges from educational practices and OER initiatives focus on producing resources specifically for teaching and learning. While there are common principles, there are also distinctions in the nature of these forms of knowledge and their provenance. Unlike OER, open data can be created and released with no consideration that it might be used for educational purposes, and major proponents of open data primarily highlight the economic and social value it could hold (e.g. Open Data Institute 2017; Open Knowledge International, n.d.), rather than its educational uses. Given this, it is necessary to explore whether there are challenges to educators in using open data in order for it to be useful as a material for learning.

Research suggests a lack of awareness of open data amongst educators, for whom distinctions between open data, OER, and other online content are ambiguous (Atenas et al. 2015). There is little research to suggest why or how open data should be employed in teaching practices. The contribution of this paper is provide a rich understanding of reasons for using open data, the pedagogical approaches that are considered to be relevant, and the challenges to wider use. While open data is not currently a mainstream educational resource, there are educators who make substantial use of it in their teaching. Possibilities are being explored at all levels of education. Therefore we identify and interview a diverse sample of these 'early adopter' educators in order to understand their practices, perceptions and experiences. A document analysis was also performed on materials related to the interviewed educator's practices with open data. Through a thematic analysis of this data, we develop a conceptual framework that suggests how open data could become a mainstream resource for education.

Open data in learning activities

Uses of open data in learning

The limited research on open data in teaching to date suggests a lack of awareness or coordination. Atenas et al. (2015) found that a majority of the 26 respondents to a survey on

AECT

open data use in education were not actually making use of open data. They were using data that was free, but not openly-licenced. It is concluded that there is a lack of awareness of open data amongst educators and a challenge of clarifying the distinct value of it. Similar confusion between open and copyrighted materials has been identified in findings on educator perceptions of OER (Rolfe 2012). Among the respondents who did use open data, Atenas et al. (2015) found a wide variety of sources and topics, including governmental, NGO, museum, and research project data. This suggests that a breadth of possibilities are being explored by a minority.

Case studies of teaching with open data suggest a trend towards project or inquiry-based learning: Shamash et al. (2015) ask students to explore open data sets through visualisation, in order to develop analysis skills. Gray (2014) describes assignments in which a database of court records provides the grounding for exploring themes in historical crime. Ciocola and Reggi (2015) describe an approach that leads school children to investigate the outcomes of publically-funded projects. Across these activities, the capacity for choice of direction, and real-world relevance are emphasised. Alternatively, Dunwell et al. (2016) developed a set of mini-games around nutritional data, but they still identify that a quality of open data is the extensive volume of material, used to create a different learner experience each time.

Challenges involved in using open data can also be identified. Dunwell et al. (2016) highlight difficulties with using 'live' open data, as opposed to immutable, controlled data. Love et al. (2016) report difficulties in finding convenient open data for a data mining assignment, and caution that work needed to be done upfront by educators. Reflecting on the process of creating introductory programming assignments using open data, Coughlan (2015) describe work to identify data sets that facilitated activities related to learning outcomes, test model solutions, and monitor for emergent issues, to ensure that activities worked well.

Drivers for open data use in education

Three drivers could encourage educational use of open data. Firstly, public engagement with open data is considered an important tool for democracy and empowerment of individuals in society. However, there are a lack of effective socio-technical mechanisms for this (Atenas et al. 2015). Open data could provide societal and global benefits (Manca et al. 2016). Worthy (2015) highlights the popularity of some interfaces built using UK government open data, such as 'TheyWorkForYou', which summarises the activities of elected representatives to 2–300,000 visitors a month. Meanwhile other sources, such as those on local government spending, have limited public use and are primarily aimed at businesses and experts. Who uses open data, and the mechanisms through which society is influenced by it, remain open issues where education could play a pivotal role as a mechanism for public engagement.

A second driver is that open data holds potential as a material suited to learning important skills. Data literacy broadly combines technical and statistical skills with the ability to draw meaning by posing questions, interpreting data and analyses effectively, and also in developing skills in reading and creating visualisations (Crusoe 2016). Data literacy skills can be developed through learning activities that use open data (Wolff et al. 2016), and beyond this, Atenas et al. (2015) propose that critical thinking, statistical literacies, global citizenship, teamwork, data curation and research skills can be acquired through learning with open data.

A third driver is that organisations want to encourage public engagement and enhancement of the data they release. This could develop into valuable forms of interaction between organisations, educators, and learners. Coughlan et al. (2015) describe how the development of a system for schoolchildren and the public to identify locations related to artworks was prompted by a museum choosing to share data about its collection. The additional data generated through these public engagement activities could enhance collection data while providing learning experiences. Bradley et al. (2009) describe how an educational game to teach spectroscopy can develop skills and crowdsource improvements to the data. Although these organisational drivers may not necessarily align with learning goals, there is potential for mutual benefit.

Pedagogical connections

Rather than suggest that open data offers an entirely new pedagogy, it is important to identify links with existing teaching and learning concepts to guide research and practice. Four such connections are: Inquiry-based learning, Open education, Personalisation, and Authenticity.

The prevalence of inquiry or project-based pedagogy noted above brings a large body of research which, in particular, highlight the importance of providing suitable guidance to learners (Lazonder and Harmsen 2016). Project-based learning often includes first-hand data collection, but has also been based around using existing sources of data. For example, Edelson et al. (1999) used a climate data-based activity to explore inquiry learning challenges. Land and Greene (2000) analysed projects where learners searched for and used openly-available (but not openly-licensed) online data, such as grocery or airline ticket prices. A key finding was the need to scaffold the development of goal-driven strategies, while being open to some serendipitous data-driven exploration. This meant that time was not lost to ineffective searching and browsing.

There is a need to clarify how open data offers benefits above proprietary data. As the difference is primarily licencing, the benefits for open data could be similar to those of OER. Butcher (2015) summarises three transformative benefits of OER, which could be applicable to open data:

- Increased range and reduced cost of resources.
- Support for adaptation of materials, which allows educators and students to be active participants who learn by doing and creating.
- Building capacity by providing educators access to the means of production of the resources.

Open data offers potential for personalised learning, for example by supporting a student to learn with data on a topic of interest to them. Responsibility for this may be taken by the learner or teacher, or could be performed by an algorithm (Fitzgerald et al. 2017). Research suggests that personally-relevant contextualisation of activities can improve motivation, level of involvement, and learning (Cordova and Lepper 1996). Familiarity with the topic of the data can create confidence and ground critical thinking. For example, Mittelmeier et al. (2018) report use of open data from the World Bank in a learning activity and found that the use of locally-relevant content (country-level educational statistics) encouraged greater participation in intercultural student collaborations.

The potential for authentic learning experiences (Herrington and Herrington 2006) with open data should be explored and evaluated. One could compare the efficacy of simulations such as virtual laboratories with activities based on 'real' open data. Benefits to the use of a well-designed simulation include removal of the 'messiness' of activities based on real-world data collection. But losing complex authenticity—for example the possibility of mistakes in the data collection—can negatively impact on learning (De Jong et al. 2013). Extrapolating from this, the authenticity of open data could be of educational value, because it is drawn from real-world actions and professional organisational practices. However, it is argued that broader characteristics, such as community interaction and authentic assessments are needed for learning to be wholly authentic (Hung et al. 2012; Herrington and Herrington 2006). Simply using open data doesn't create authentic learning, but it offers a starting point.

In summary, disparate examples show the potential of open data for learning, and external drivers that suggest that stronger links will emerge between open data and education. However there is sa lack of mainstream understanding of open data amongst educators. Underexplored links exist with established research and practices. In the remainder of this paper, the perspectives of educators using open data are analysed to understand the potential and challenges for wider use.

Methodology

This research employs a qualitative descriptive design. This is appropriate to understand and summarise specific events experienced by individuals (Knupfer and McLellan 1996). In this case, these events are the practices of early adopters of open data in teaching. The design assumes a level of objectivity in the reports of activities by interviewees. In the phenomenological research tradition the data would be considered to represent lived experiences, rather than objective facts (Knupfer and McLellan 1996). However, the document analysis of materials related to each interviewee's practice provides further support for the reports, and the combined focus on describing the learning activities used, educator experiences, and views on the potential of open data in learning, is valuable in gaining a holistic view from early adopters.

The overarching question was: How do experienced educators conceptualise open data and its use within education? Within this, three sub-questions highlight the rationale, pedagogy, and practical issues found in the use of open data:

(1) Why would educators make use of open data in their teaching?
(2) What approaches do educators employ when using open data?
(3) How do educators incorporate open data into their teaching and what challenges exist in this?

A semi-structured interview approach was used to gather rich data from the first-hand perspectives of educators identified as having used open data in teaching through a web-based review. A process of thematic analysis was conducted with this data following the approach described by Braun and Clarke (2006). The resulting phases of this process are described in the findings section below.

A document analysis was also performed using gathered materials related to the educators practice, including course materials, project websites, and published reports of their

teaching practice. The documents were originally used to develop an initial understanding of the interviewee's practice pre-interview. After the interviews were conducted and analysed, the documents were analysed and coded according to the interview themes in order to confirm trustworthiness of the data from the interview, and to augment understanding of the themes derived from the analysis of the interviews.

Sampling and recruitment

Given limited information of open data use in education, a purposive sampling approach was deemed appropriate to recruit a sample of educators to provide insight through their experiences with open data. As preparation, a web-based review was conducted to identify projects and educators connecting open data and education. This created a base understanding of the area and identified a population of potential interviewees. 26 relevant links to courses and educational projects were identified through Google searches on relevant terms ("open data course", "open data education", "open data teaching"). In response to Atenas et al.'s (2015) finding that the meaning of open data could be ambiguous to educators, the researcher clarified that all or most of the data sources used in each of these courses or projects were openly-licenced. The research aimed to gain perspectives across diverse educational contexts, so interviewees were judged to have experiences of using open data in different teaching or educational contexts. 14 suitable contacts were identified from these 26 links. Contact was made via email to provide a project information sheet and request an interview.

Interview structure

To understand the emergence of this practice and its meaning, a constructivist viewpoint on the creation of knowledge is taken. This means that interviewees, as experts in this area, can shape the themes arising alongside the interviewer. Semi-structured interviews were used to explore experiences and perceptions through dialogue, with space for the interviewee to raise themes beyond a structure of consistent questions (see "Appendix A").

In the first stage, the interviewee provided a record of their own experiences and explained their rationale and approach to using open data. Questions then prompted further discussion of the potential, requirements, and challenges of open data in education. The themes in literature described above were used to devise these questions, but questions approached these topics generally to avoid leading the interviewee. For example, rather than asking whether they use an inquiry-based learning design, the interviewer asked what sort of approaches or learning activities they think work with open data, and whether they use particular theories of learning or pedagogy. It is also important to reflect on the influence of the background of the researcher. The researcher/interviewer has first-hand experience of using open data in their own teaching, and an interest in OER. However, the questioning focused on the interviewee's practices, to prioritise their perspectives.

Document analysis

For each of the interviewees, materials, reports, media, and websites related to their use of open data were gathered for a process of document analysis (Bowen 2009). This material was reviewed prior to the interviews for an initial understanding of the aims and practices

of each interviewee. After the thematic analysis of the interviews was performed, this material was returned to and a thematic analysis was performed to identify instances of themes proposed. This provides a suitable means to triangulate and validate the assertions made by interviewees (Bowen 2009).

Ethics

The research was approved by the Human Research Ethics Committee of The Open University, and conducted in accordance with the 1964 Helsinki declaration and later amendments. To provide an opportunity for interviewees to correct any misrepresentation of their views, a report detailing the full analysis of the data was provided to the interviewees for amendment. Informed consent was obtained from all individual participants included in the study. It was agreed that the names of interviewees, and their organisations, staff, or students are not published, in order to protect anonymity and support candid responses.

Results

14 identified people or projects were contacted and 10 interviews were conducted. 4 of the interviews were conducted in person and the remaining 6 via video conferencing. The interviews ranged from 28 to 47 min in length, with an average length of 37 min. A total of 15 artefacts were used in the document analysis with at least one artefact for each interviewee.

Table 1 describes the interviewees and the materials used for the document analysis in each case.

As a sample of interviewees in a specialised area, this provides a broad representation across teaching contexts, subject matter and location. Hereafter, ID codes are provided in brackets after quotes to identify interviewees.

Findings

177 initial codes were generated by the researcher in the initial stage of thematic analysis to describe concepts in the interview transcripts. These were refined and sorted into broader themes and an initial set of overarching categories. The researcher then provided interviewees with a report of this analysis to check for any misrepresentation or inaccuracies of interpretation. None were identified. Further stages of analysis explored the relationships between these themes and categories. A final stage then focused on how these categories and themes could be refined to support answering the research questions by splitting or combining themes that contained multiple concepts within them. As one example of this, a theme of "data literacies" from the earlier analysis was identified to contain distinct themes within it that reflected different questions. Firstly, that open data provided a means to teach data literacies, and secondly, the need to ensure that students and educators had appropriate data literacies to complete the learning activities. This was therefore split into a "Basis for learning data literacies" theme within the "Characteristics for Learning" category, and "Managing data literacy requirements" and "Developing educator data literacies" themes within the "Skills and Knowledge" category.

Table 1 Interviewees, the contexts of their teaching, and the artefacts used in document analysis

ID	Context	Subject matter	Country	Uses of open data in teaching	Artefacts used in document analysis
A	University	History	UK	Multiple years of using a repository of historical court records in teaching activities and student projects	Published report on teaching approach, data source website
B	MOOCs, University	Data analysis	New Zealand	Development of a MOOC with activities based around national health survey and country-level statistics open data sets	Course materials, project website for educators, open data source websites
C	School	Politics	Italy	Creating an approach and resources to student-led projects using open government spending data	Project website, lesson plans and media, published report
D	School	Data Literacy	UK	Devising and delivering learning activities with school teachers using open and locally-collected georeferenced data	Project website, video of example lesson
E	University	Information management	Canada	Assignments around visualising government open data and creating infographics	Assignment guidance, examples of student produced outcomes
F	Skills training,	Open data	France	An initiative to improve skills and engagement with all open data sources for investigative journalism and activism	Documentation of teaching approach, examples of completed activities
G	College	Biology	USA	Teaching activities with an open scientific database to develop bioinformatics and research skills	Assignment guidance
H	MOOCs, University	Programming and data analysis	UK	Teaching through a MOOC where learners engage with multiple international-level open data sets	Course materials
J	School	Data visualisation	Sweden	Creating resources and trialling activities for schoolchildren to engage with current affairs through country-level statistics open data	Description of teaching activities and example outcomes
K	Schools, skills training	Data literacy	USA	Developing tools and activities to support teachers to make greater use of a wide range of open data sets	Project website with library of data sets

AECT

The final refined analysis described below has 6 categories that contain 28 themes. The categories and themes are described in the following sections (including Tables 2, 3, 4, 5, 6 and 7) and highlighted in bold where mentioned.

The categories of themes comprise: *Characteristics for Learning*, which represents perceived strengths of open data when used in learning. *Teaching Approach*, which represents elements of the learning activities described. *Open Data in Society* themes describe ways in which education, open data, and societal issues interact. *Creating and Sharing* themes highlight social interactions in learning with open data. *Skills and Knowledge* themes represent ways in which the understanding of learners and educators interacts with the capacity to use open data in learning. Finally, *Making Data Usable* describes work that is performed in order that learning activities can occur with open data. Below we describe each category in turn and the themes within them.

The document analysis validated the existence of 25 of the 28 themes and identified 65 instances of these themes in total. Tables 2, 3, 4, 5, 6 and 7 identify the count of documents in which each theme was found. The nature of the documents analysed meant that certain themes were strongly represented. For example, 'tools that support learning with open data' are commonly introduced to students in course materials, and tasks related to 'communicating with data' were often given to students. The nature of the documents can also explain the lack of instances of the three themes that were not identified. Educators would be unlikely to describe the need for 'developing educator data literacies' in their own teaching materials. The potential for generating 'supply prompted by demand for open data' through educational activity would not be a priority to communicate in these situations. The challenge of 'finding raw data' reported by educators was implicit in the examples of data sets used, but not explicitly discussed in these documents. Beyond validation of the themes, the documents provide useful elaboration of the themes and examples of this are integrated with the interview data in the below descriptions.

Characteristics for learning

A range of valuable characteristics of open data were identified by educators. Firstly, the *broad scope of the materials* can support a wide variety of novel activities. For instance, E's assignment guidance states that:

> "Canada's Open Data portal is http://data.gc.ca/, where you can find tens of thousands of data sets from various government ministries and agencies. One goal of releasing Open Data is to… use this data to provide services or information to Canadians that the government does not have the resources or the mandate to provide." (E, Assignment Guidance).

Related to this, open data can have *relevance to the learner or to current affairs*, and this is seen to have motivational qualities. As D surmised: "You take a map with some data, the first thing (students) do is put their postcode in. 'Can I see my house?'" (D). Similarly, K took open data on test scores in local schools back into the classroom to teach statistics, and found that:

> "They have family or friends who go to different schools in their neighbourhoods. That drab learning experience around box plots was transformed into something completely different" (K).

Table 2 Themes identified within the 'characteristics for learning' category

Theme	Count of interviews where the theme was identified	Count of documents where the theme was identified	Definition	Example quote
Broad scope	8	1	Open data sets are very broad in scope, e.g. global and location-specific details	"In the weather week, my colleague uses the weather from London, but of course they can use their own local weather… or wherever they want" (H)
Relevance to the learner or to current affairs	6	3	Data or topic is familiar to the learners, including relationships with locality or current affairs	"We had a huge wave of oncoming migrants and it was very interesting to follow that… people were discussing it so much, but it wasn't based on data" (J)
Basis for learning data literacies	6	4	Open data has value as a basis to develop data literacies that are essential in the modern world	"I'm only ever seeing open data through the lens of someone else's app, or someone else's website. Never interacting with the data. If you trust your intermediary that is not such a big deal. But if your intermediary is driven by economic reasons that right away impairs some level of trust." (E)
Advanced possibilities for activities	5	1	Open data allows learning activities to reach advanced levels of complexity, such as engaging students with activities normally undertaken by professionals	"When I did my PhD I literally had to go through and count how many things there were… I could spend a day… whereas I can do it within seconds now. I can do it in a class" (A)
Authentic experiences	4	2	Data collected from real-world projects creates learning experiences that reflect the complexity of real world problems	"As much as they are used to getting fake results so that they can learn the process, they really do like the idea that this is real science… the thing I like most about it is the unpredictability." (G)

Open data offers universal access to digitised records and analyses from a range of contexts. It can therefore create *advanced possibilities for activities* where learning occurs in ways that previously would have been only the realm of professional or higher-level research activity, or would have been too time consuming to perform. Interviewees noted how the use of these databases changed the possibilities in learning activities and outcomes. For example, G teaches college bioinformatics and described how they run lab activities in which students explore plant genetics based around an open scientific database: "They pick a gene, they research it, they even do some transformations with the gene" (G). This database is designed for researchers, rather than education. However with sufficient guidance, the labs allowed students "to experience how all this data is just there and if we work together we can get better research results" (G). They were enthused by the way that open data brought students closer to the world of professional practice, and "found it is not actually that hard for them to interpret the data once they figure out how to get there" (G) by being guided through the interface. Open data can therefore support transitions to higher levels of scholarship and to close gaps between teaching and research.

Related to this, the real-world provenance of the data, and the processes through which it is collected, offers *authenticity* for the exploration of rich and complex issues, and for development of skills in real-world contexts. In the case of the plant genetics database, G described the unpredictability that arises from the database, in contrast to a contrived textbook example. When "comparing genes you would expect to have a certain outcome. But based on what gene you pick, you may not have the data line up as you would expect… (students) ask me how come this box is red rather than green. And I get to say 'I don't know. That is an open question. You get to figure that out.'" (G).

Other educators were keen to use open data as a *basis for learning data literacies*. Interviewees expressed that developing capacity to understand data was important to learners and societies. There was seen to be potential through this to address imbalances where, for example, there was a "culture where people rely a lot on what politicians say rather than the data" (F), or for E, to make important data accessible for the public to understand.

Teaching approach

A variety of themes highlighted the ways in which learning activities were designed to harness open data and to respond to the challenges of matching it to student capabilities.

Eight of ten interviewees employed forms of *project or inquiry-based learning* to harness the relevance, broad scope and authenticity characteristics described above. The lesson plans used by C were structured in project phases of "designing, exploring, analysing and telling", and the adaptation of inquiry-based approaches to use of open data was considered relatively straightforward, with D stating that:

> "In scientific experimentation, often that says 'you collect your data', but you can collect your data from an open data source. Really it is the same scaffolding." (D).

The broad scope of data sources supported students to make *choices for personalisation,* through which the potential for the characteristic of relevance to the learner could emerge. Teaching materials from C and E guide students to explore a particular data repository in order to develop an interesting or "non-trivial" (E) project, while identifying a topic and any relevant open data sources is an initial stage in the activity developed by F.

Most educators saw the capacity to *harness connections* by working across multiple data and information sources as important. This theme was prevalent in the document

Table 3 Themes identified within the 'teaching approach' category

Theme	Count of interviews where theme was identified	Count of documents where theme was identified	Definition	Example quote
Project or inquiry-based learning	8	4	Project or inquiry-based pedagogy was used. Open data replaced or complemented data collection	(The process is) "defining the problem, finding the data, getting the data, verifying the data, cleaning, analysing, presenting. It could be on pollution. Is there a link between pollution and admissions in hospitals?" (F)
Harnessing connections	8	5	Connections between multiple data and information sources can be used to develop understanding and skills	"It is not one data set. They have to match different data, for example at the local organisation of the project they have to mix two datasets. They have a key field (to link them)." (C)
Simplification	6	1	Limitations of time or skills require that the data sources and activities are simplified or structured, in order for learners not to be overwhelmed	(Students say)"'I spend so much time looking at data sets that don't do what I want' … or 'I tried out 10 data sets and couldn't find anything interesting'. There is always a tension…between this open-ended 'pick an open data set' which I think is really pedagogically valuable, or… 'here are 10 data sets that I know are really interesting'" (E)
Harnessing learner motivations	5	1	Open data-based activities can be linked with learner motivations	"It is a lot of 'what is in demand', 'what is of interest', 'what is applicable to their jobs, industries?' … primarily around things that are fundamental to today's economy" (K)
Choices for personalisation	4	3	Giving learners the capacity to choose different foci within the data sets and pursue their interests	"Some say 'OK my grandfather died of tuberculosis'. Others say 'well I want to see how my country, Mexico, compares with developed countries'" (H)
Process-based assessments	4	2	Assessments tended to consider whether students used a suitable process and sources	"It is assessed for completion rather than accuracy… I am less interested in what they get than I am in how they get there." (G)

AECT

analysis where assignments and examples of student activities showed the value of combining sources. For example, educator E provided an example to their students where border agency data on expulsions from the country was combined with other sources of data on crime and demographics to identify correlations. The examples they share of assignments created by their students include various comparisons of data from different sources. For example, analysing mortgage rates in conjunction with data on average home rental prices over time to show changes in the financial implications of home ownership or renting.

Where assessment was important, *process-based assessments* were mentioned. This focused on how well students had followed a process and used sources, rather than for accuracy or a specific goal. For example, A stated that they "judge them on… the effective use of the sources they have got. They (should) go off and find other sources." (A).

Although they were keen to support students to connect multiple sources and for the authenticity of the data to emerge, educators could also use various forms of *simplification* to reduce the burden of complexity and support the achievement of learning outcomes in the time available. This was particularly evident in school-based activities. In working with teachers to create such activities, D noted that:

> "we've got three sessions (in the school) but in reality if someone is doing a genuine data analysis of these kinds of complex data sets (they might take) five or six weeks." (D).

Simplification could also include devising small test activities before a main project activity began, and scaling up from small data sets in which each value and computation could be understood. For example, a starting activity where students were restricted to a small number of values within the data, that let pupils "see that: 'OK if I can make it work with 10 values it will also work with 12,000'" (J)

Open data in society

Rationale and use of open data was influenced by the open data movement and data-producing organisations, and the engagement of learners with open data was considered a means towards societal change. This was highlighted when interviewees expressed that their learning activities had the potential for *real-world impacts*, even if these were small or sporadic. For example that students had added information to a public research database. Open data also held potential for *infusing education with politics and governance*. Some educators such as C directly use open data as a means to engage learners with political issues like spending. Others were more concerned with *developing public engagement* with open data through education. For example, that:

> "…now that we have this base of available open data, there is the issue of getting people to use it. And getting people to understand the possibilities for their own work." (F)

In this, *supply and demand* for open data are intertwined: if open data is used by the general public through education, this can create demand for more data to be released

Table 4 Themes identified within the 'open data in society' category

Theme	Count of interviews where theme was identified	Count of documents where theme was identified	Definition	Example quote
Real-world impacts	5	2	Open data-based activities could lead to impacts through authentic engagement with society	"There was funding to create a hostel… but the local municipality didn't assign it to any subject or association. So it was very beautiful but it was empty. And after a year of action by students and teachers now they publish an application to manage this hostel." (C)
Developing public engagement	4	2	Using education to develop wider engagement with open data	"Now that we have this base of available open data, there is the issue of getting people to use it…to understand: What are the possibilities?" (F)
Supply prompted by demand for data	3	0	Greater interaction with open data through educational activities can drive processes that lead to more data being released openly	(We were exploring) "the rise of shared bike riding…. and pollution. But the data about the bikes was not open. We did a small advocacy campaign around this and it was opened… In some countries there is resistance to open the data because they don't see the demand."(F)
Infusing education with politics and governance	3	1	Politics has a strong influence on the open data movement: It can define the data that is collected and shared, and encourage open data use in education	"I hope that it will become a normal activity in high schools. (We talk with the) Minister of Education to make it more structured in the curriculum…. Every region has to compile data on the development of the projects in this system… the meta-data structure was built to talk to the EU." (C)

AECT

in usable forms. As E noted: "It is not enough to work with the open data that is there. You should also be asking questions about the data that is not there" (E).

Creating and sharing

This category captures how social and collaborative approaches were employed around open data-based activities. All ten interviewees mentioned that *Communicating with data* was seen as a key skill to be developed. There was a strong emphasis on creating stories with the data, for example that "what you are using it for is to go in and find a story…that will be compelling" (E). This theme was also identified in six artefacts in the document analysis, showing how communicating with data was embedded in the design of learning activities. For example, the teaching materials used by C and F both included the assignment of a group member to be the 'storyteller', who could be responsible for "finding interesting angles to explore and producing outputs that really speak to the intended audience" (F).

Creating data and sharing it offers another route to make contributions to society and to broaden understanding of the processes and influences on data in our lives. Learners could combine activities around existing open data with others to "collect (data) with thermometers, to geographically map where they took different temperature measurements" (J). Such activities could provide an additional perspective and raise awareness of issues such as accuracy or personal data privacy. F's organisation provides materials on 'Collecting data using Smartphones', as well as on creating maps to represent data. *Peer support* mechanisms included sharing advice as students grappled with new concepts and skills, and for different roles in collaboration.

Table 5 Themes identified within the 'creating and sharing' category

Theme	Count of Interviews where the theme was identified	Count of documents where the theme was identified	Definition	Example quote
Communicating with data	10	7	Visualising, storytelling, and discussion were common activities	"With a simple button click you can share…it becomes a public URL… (One activity is to) share your project… some of them really look at each other's and comment" (H)
Peer support	6	2	Collaboration and cooperation helped learners to overcome barriers, or to take roles that suited them	"It is quite a challenge for them to wrap their minds around… real information off the computer that is not pre-set… so I find it helps that they work together… collaboration is a big part of science." (G)
Creating open data	4	2	As well as learning with existing open data, activities where learners created data could be complementary	"One way to gain understanding … is activities where you are contributing to the data. So you are understanding it from both directions…that actually this data is coming from them" (D)

Skills and knowledge

Themes in this category highlight the interplay between the understanding required to enable open data-based activities, and where skills could be developed through these activities. Eight of ten educators highlighted that their teaching *combined subject knowledge and data literacies*, even where data literacies was the primary goal. For example:

(For other educators) "…the real end is understanding more about politics. For me, the real end is… (to) use an interest in politics to get them interested in data" (B). However, they also "wanted data sets where a fairly wide spectrum of people would have some idea of what it is about" (B) and felt that "you should be learning real and interesting things as part of the by product of what you are doing" (B).

They therefore chose data sets that provided scope for comparison of basic statistics about countries around the world, and around health, that would be of wide interest, as a foundational familiarity with the subject domain of the data was seen as necessary, and interest in the subject matter a motivator.

Managing data literacy requirements for students was a key issue for educators, including the technical skills required and the capacity to follow an inquiry process using data. G noted that:

Table 6 Themes identified within the 'skills and knowledge' category

Theme	Count of interviews where the theme was identified	Count of documents where the theme was identified	Definition	Example quote
Managing data literacy requirements	9	2	Technical and conceptual skills are required to understand data and use it effectively	"Learning to cope with messy data… to get it visualized or to work in the tool… was very distracting from what I was really trying to teach… (such as) 'Have you asked a sensible question?'" (D)
Combining subject knowledge and data literacies	8	3	A foundational understanding and interest in both data literacies, and of the subject matter of the data is important	"You need to understand how to work with raw data. But you also need to have some understanding of the domain… don't try to work in a context in which you are unfamiliar." (E)
Developing educator data literacies	6	0	Skills to teach with open data were needed and could be a barrier	(Teachers) "get a fear that they're going to be asked questions that they don't know the answers to… When I had a lot of time, maybe five sessions …it was much better because they know how to teach the class." (D)
Developing student digital literacies	4	1	Open data-based activities require the development of more general digital skills	(During the projects, they learn to use) "social media and online communications, video making skills because they have to make a video. And also about how to publish online" (C)

AECT

"The only way I've managed that is to take it real slow… Because they first have to grasp the computer stuff and then they have to grasp the science… I don't actually see the two as in conflict. I see tech as being a bigger part of the science nowadays." (G)

The wider *development of student digital literacies* was also raised. Skills could be developed in activities, however there could also be checkpoints to assess existing literacies. For example A noted that they get their students "au fait with the technology… very early and then we come back to it. So I get them to do mini things… so I know then that they have been able to use the search mechanism" (A).

Six interviewees described that *developing educator data literacies* was essential, but it was not clear how this would be facilitated. It was noted that: "open education and open data…(should) meet and have some sort of repository where all these (learning activities) exist, where we can share them" (G). The importance for educators to feel comfortable in their own knowledge of the data was seen to be key to making data usable.

Making data usable

A range of themes described preparation and manipulation activities around the open data. All interviewees noted that *finding good data for a purpose* could be a substantial undertaking for educators or learners. Teaching materials from F identified the 'scout' as a member of each team whose role was to find relevant data, and to find information about its quality and limitations. Educators looked for certain qualities in these sources, for example, that they offered *raw data*, rather than aggregations that were of limited value some purposes. *Refining and curating* data sources was a common undertaking to simplify the learning activity, with "a lot of curation and cleaning up and digging…before that open data set has any value in the context of learning" (K).

Four interviewees emphasised that poor choices in data *formats and documentation* could lead to the potential misinterpretation of data by students and additional challenges for educators. A lack of clarity or consistency in how data is captured and analysed may result in obscured issues of *accuracy and change* to data that educators had to respond to. Interviewee H described how, on returning to run a second presentation of their course, they were surprised that the same data points had changed and no longer matched the course materials. This was because these points was based on estimates that had altered after the inclusion of additional data. As in other tasks to make data usable, the work required to assess accuracy could become part of the learning activity, For example, the skills taught in D's project were stated as including "understanding the validity" and engaging students in assessing "anomalies, bias, and completeness" in the data used.

Six interviewees described *tools that support learning with open data*. Open data was commonly manipulated in spreadsheets and this was seen to help students to review data and formulas. However some interviewees noted that these general tools for data manipulation and analysis offered poor support to learn from mistakes. E.g.: "Most spreadsheet software, if you don't get the exact syntax …they will give you an error message and that's pretty much it" (K). The potential to simplify understanding of data using visualisation was also raised, to support students "to be able to look at a whole lot of variables at once, using graphical metaphors that are pretty easy to understand" (B). B and K were involved with projects to produce tools for learning with open data. Tools were described in eight of the

Table 7 Themes identified within the 'making data usable' category

Theme	Count of interviews where the theme was identified	Count of documents where the theme was identified	Definition	Example quote
Finding good data for a purpose	10	2	All interviewees discussed exploration and decisions made to find suitable data for their teaching	"I wanted a rich array of variables…so there is interest for different people. And that's why mainly we have two big datasets. I was talking to people across countries so country-level data is an obvious point of contact…" (B)
Refining and curating data sources	7	3	Open data is vast, diverse, and not designed for education, so refining and curation of sources is required	(I) "resampled the data to undo the weighting that (the data producers) had used. I'd undo that so that when people looked at things they weren't getting false impressions from the complexity of the data source." (B)
Tools to support learning with open data	6	8	Valued characteristics in tools included means of representing data, guiding discovery, and recovery from mistakes	"With interactive visualisations you can look at time series for example… For geography the (important features are the) possibility to view things on a map, for history and social science… to see how things evolve over time." (J)
Understanding formats and documentation	4	2	Open data takes various forms and a lack of documentation can impact on educational uses	"Some of the rows apply to women over 50 and some rows for kids under 6…There is not enough documentation. (Sometimes a student's work) ends up having a false note to it, because they have used the data incorrectly." (E)

AECT

Table 7 (continued)

| Theme	Count of interviews where the theme was identified	Count of documents where the theme was identified	Definition	Example quote
Finding the raw data	3	0	Raw data was valued yet often unavailable. Aggregated data may not supported intended learning activities	"Aggregated data where work has already been done, some assumptions have been made and maybe even some data has been removed… a lot of the learning that could be drawn from that data set has been taken away" (K)
Monitor accuracy and change	3	1	Data may have questionable accuracy, or unclear collection methods	"You have a big chain of transmission… the consolidated data at the national level open data portal (could) be a bit different to… the local level." (F)
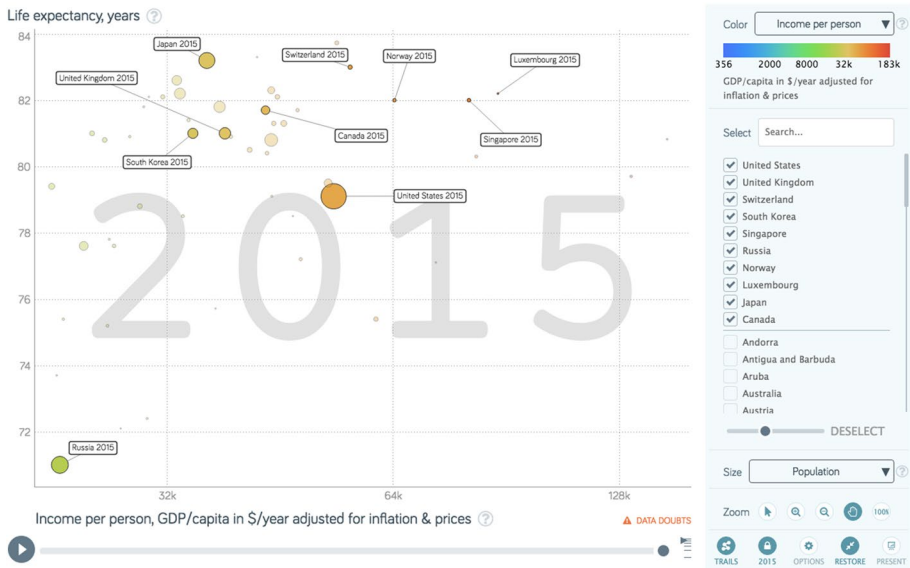
Fig. 1 Gapminder interface for exploring open data (author Gapminder.org, CC-BY)

analysed documents including in lesson plans, course materials, and assignment guidance. This shows the importance of guiding learners to understand tools as part of the process of using open data in learning, and again emphasised that learning with open data encompasses learning data and digital literacies alongside subject knowledge. It also highlights the need to simplify and scaffold learning through choice and guidance around tools, with A and J highlighting how simple interfaces to open data, such as Gapminder (2018), were instrumental their work,. Figure 1 shows a Gapminder tool, as used by J, which provides simple ways to compare data about countries over time.

Discussion

The analysis is now used to answer the research questions and construct a conceptual framework through which support for the use of open data in education can be understood.

Why would educators make use of open data in their teaching?

Educators identified positive characteristics that could emerge through the use of open data in learning: The broad scope for exploration of varied subject matter, and the potential relevance to the learner, were perceived to motivate learners. Open data also presented advanced possibilities for activities that reflected higher level professional or academic practices, and the potential for authentic learning experiences drawing on real-world data and processes. A final characteristic is that open data provides a basis for teaching data literacy. This was a motivation for educators engaged with the themes of the 'Open Data and

Society' category, as these data literacies are essential to developing awareness and public participation with open data.

These characteristics can link with, and extend, familiar concepts in education that interviewees implicitly or explicitly followed in their teaching. This provide avenues to inform next steps in research and practice. To highlight two such links: research evidence suggests that relevance can improve educational outcomes (Mittelmeier et al. 2018; Cordova and Lepper 1996). Authentic learning research (e.g. Herrington and Herrington 2006) would suggest that learning with open data should go beyond just accessing a 'real' data set, but include elements such as authentic collaboration and assessment of outcomes. The themes within the 'Creating and Sharing', and 'Open Data in Society' categories show how some educators developed a broader sense of authenticity by getting students to communicate, collaborate, and engage with the wider world through open data.

Using these themes, the rationale of interviewees could be classified into three groups: Those doing subject matter teaching who used open data to create advanced possibilities (A, G), those seeking to develop public engagement through open data with authentic inquiry (C, F, J), and those using the relevance and broad scope of open data to motivate learners and develop data literacies (B, D, E, H, K).

What approaches do educators employ when using open data?

Themes in the 'Teaching approach' category provides a foundation to answering this question, but other categories are connected. Project or inquiry-based activities were commonly employed, and these were a familiar pedagogy for most interviewees. This facilitated the choice of a topic and of data as a means to achieving personalisation. Research suggests that there are alternative ways to provide such personalisation, and insufficient clarity on where and why it has a positive impact (Fitzgerald et al. 2017). Educators built in encouragement to harness connections across multiple data sets and for sharing and communicating with data. In this regard, open data-based learning could be linked with Connectivism (Siemens 2005), and the principles associated with this could form a basis for activity design.

The above themes focus on harnessing positive characteristics of open data, yet teachers also design activities to overcome particular challenges. Simplification of complex activities was required. This relates to themes present in the 'Skills and knowledge' category, where learner and educator's data and digital literacies may need to be monitored and developed in tandem with subject knowledge. Research on scaffolding strategies could be particularly relevant (Belland et al. 2013), given the need for complex skill development in the context of motivation-based, authentic and ill-structured tasks.

How do educators incorporate open data into their teaching and what challenges exist in this?

The strong sense of value in using open data was tempered by practical challenges in implementation. Themes within 'Making Data Usable' highlight barriers for many educators. Finding good data for a purpose, developing awareness of formats and documentation, and refining data in order to make it usable, were not trivial tasks for educators or learners as they explored how to harness the value of open data.

Greater educator data literacy would support wider use of open data in learning, and this could be connected with a wider data-driven agenda for education, in which it is argued

that educator data literacy is essential for evaluating and improving teaching. Mandinach and Gummer (2013) identify a growth in awareness that all educators must use evidence in decision making and therefore need training in data literacies, and Kippers et al. (2018) describe a program that significantly improved the data literacy of teachers. If this wider agenda brings advances in data literacy, it could overcome some of the existing barriers to open data as a mainstream educational material.

Challenges raised by interviewees arise from gaps between producers of data and the educational users. Raw data may not be available in the form that would be most useful, and the educators or learners may not fully understand the accuracy of the data or its propensity to change. It is therefore pertinent for educators to consider how they might be influential in the wider 'Open Data and Society' space. There is potential for educators and learners to engage in the demand for open data, and therefore influence its supply. Tools that support educational use of open data, whether designed specifically for this purpose, or having appropriate characteristics, are also sought after as a means to alleviate burdens and enhance the learning potential.

Illustrating the connections between rationale, approach and challenges

To illustrate how rationale, approach and challenges are interrelated in the interviewee's activities, the overall approaches of the three differing rationale groups are now described.

Creating advanced possibilities for subject matter teaching

For A and G, open data enhances the activities they use to teach their subject. A uses open data to get students to understand how legal trials operated historically. Students can choose a single trial or set of related trials on a theme of interest. Students produce a group presentation in which they enact a trial and discuss it, and an individual written report. Assessment is process-based including the effective use of multiple sources and the cohesion of the presentation. Open data is valuable to this pedagogical design by overcoming the limitations of access and understanding of printed records. It supports inquiry across multiple sources. One source is at the core but two others are pointed to for supplementary information. A and G emphasise the ease of use of these resources, including tools built into the websites for visualisation and search, and emphasised the challenge of ensuring that students have appropriate data and digital literacies.

Develop public engagement through open data with authentic inquiry

For C, F, and J, open data is a basis for public engagement with societal and political issues. This requires the development of literacies in data and civic participation. For example, F's initiative aims to develop greater data literacy in the public to tackle local concerns. This is achieved through devising inquiry-based structures which can adapt the interests of the local groups taking part. There is an emphasis on identifying a problem that is of authentic interest and relevance. The learning process is driven by this problem, so if the existing open data is inadequate, the response is to identify or collect further datasets. Skills are developed in areas such as communicating with data in journalism or lobbying. C and J work in schools, but present a similar emphasis on basing inquiry in authentic problems and relevance to current affairs.

Use relevance and broad scope to motivate learners to develop data literacies

B, D, E, H, and K all teach topics related to data literacy, including programming for data analysis, or data visualisation. For B, the pedagogy of the MOOC provides choice to learners to personalise their data analysis activities with datasets that have been simplified to reduce the burden or potential for mistakes. B, D, and H emphasis that supporting the choice of data with personal relevance (for example, local to the student) is a tool to motivate interest where this can otherwise be lacking. Requirements for subject matter knowledge and data literacies are again a tension, but here it is the subject matter knowledge that can be missing—data cannot be understood without context. B selects data sets around widely-understood topics such as health. E guides students towards using data sets where they have some existing knowledge.

A framework to support greater use of open data in learning

A conceptual framework using this analysis is presented in Fig. 2. This is grounded in the themes raised in the interview analysis, but elaborates on the relationships between the categories which suggest how further practice, research and development could support open data to become a mainstream part of learning activities. This framework can guide thinking around the construction of learning activities using open data, or could inform wider initiatives to engage educators with open data.

The framework highlights the interconnected nature of each category. Here we highlight some of the interconnected relationships across these that should be a focus for further research and development.
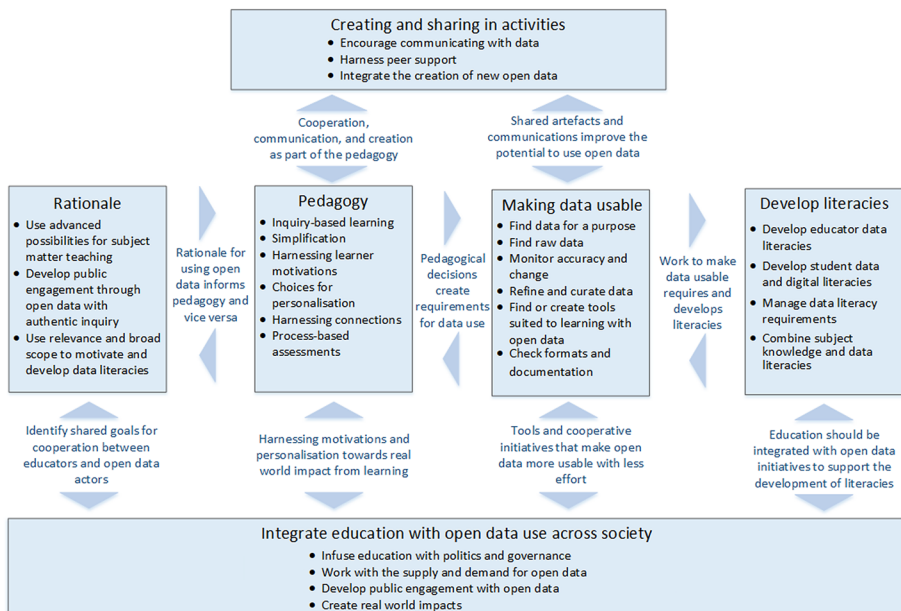

Fig. 2 Framework for using open data in learning activities

Identify shared goals for cooperation between educators and open data actors, leading to tools and initiatives that make open data more usable with less effort

Educators could engage large numbers of people with relevant data, and activities can combine use of existing open data with sharing and advocacy for release of data. The responsibility for making open data usable for education is unclear and mutually beneficial models of cooperation could support the tasks of 'Making data usable', increasing the potential for educators and learners to engage.

Harnessing motivations and personalisation towards real world impact from learning, to create shared artefacts and communications that improve the potential to use open data

Open data is grounded in societal and organisational practices and can support authentic learning experiences. However open data are rarely devised for educational purposes, and are often poorly specified. Sharing the structures, created data, and outcomes devised in educational activities using open data could provide a means to reduce these burdens.

Work to make data usable requires and develops literacies, education should be integrated with open data initiatives to support the development of literacies

For educators and learners, data literacies are both a requirement and potential learning outcome. At the same time, pedagogical expertise from the education sector would be beneficial in reducing barriers to public engagement with open data, which is a goal of many open data initiatives.

Limitations and future work

In line with many qualitative studies, this research has focused on in-depth analysis of rich data from a small sample. The limited mainstream use of open data in education to date constrained the sample of interviewees and projects. The research was also limited to English language speakers and future research should rebalance this. Early adopters may act in some ways that are not generalizable to mainstream educators. This research has identified a set of reasons and approaches to using open data in teaching. It did not seek to measure the effectiveness of open data in a specific pedagogical approach. Future work should seek to measure whether specific pedagogical designs using open data are effective for students, building on the views of educators and the links to related pedagogical research described here.

Conclusions

By working with early adopters this research identifies how open data has potential as an authentic and relevant material that expands the possibilities for learning activities. It can be used to develop public engagement with data. It offers advanced possibilities

for learning activities across a range of subject matters, and can utilise the personal interests of learners to motivate learning of data literacies. Similar themes arose from interviewees working in diverse educational contexts, from schools to MOOCs.

Limited mainstream use in education can be attributed to challenges identified here. Most open data sets are being released from real-world professional practices, such as scientific research or policy-making. Learners and educators may not have the literacies or resources to make them useful to them. Tools that simplify some of the complexities of working with open data are emerging, but there remain challenges to making data usable. These challenges are not insurmountable, and their repetition across interviewees suggests that shared solutions would be valuable. The framework in Fig. 2 highlights relationships between the themes that would be part of these solutions.

Data-sharing organisations could be actively engaged in providing suitable forms of data and support. Identifying mutual benefits for these organisations to engage with learners is one way forward. Partnerships in which educators and learners engage with the data producer could also increase the authenticity of learning experiences. However, open data can be used and adapted independent of the data sharing organisation. A second approach therefore relies on educators, learners, and educational technologists to build a layer of support and translation on top of original data sources that makes them suitable to the teaching approaches described above. This would be appropriate to working across data sources and utilising connections between these. Some evidence for the emergence of both of these approaches could be found in the interviews, and they are not mutually exclusive.

To increase effective use of open data in learning, there is a need to understand how to harness the qualities identified here effectively in the design and implementation of learning activities, and of supporting approaches and technologies. The adaptation of existing research and pedagogy into this endeavour should be fruitful, and the analysis and framework presented here offers a conceptual starting point through which to direct these efforts.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Appendix A: Interview structure

Exploring the potential of open data as a material for learning: Interview outline

The purpose of these interviews is to gain a more detailed understanding of how open data is being used in teaching and the production of educational content and activities. The questions have been devised from an initial review of the ways open data is being used in

education and the media, literature, first-hand experiences, and consideration of the wider open data agenda.

Personal interest and experience

Please describe how you have used openly available data in relation to teaching or producing educational materials

(Highlighting a particular example known or mentioned) What were the learning outcomes you aimed to achieve in (example)?

How does the use of open data help to achieve learning outcomes such as these?

How did you become interested in openly-available data?

Potential of, and connections between open data and learning

Are there particular subject areas where you see opportunities to use open data that we haven't exploited fully?

What sort of approaches or learning activities do you think work well with open data? Why?

Are there particular theories of learning or pedagogical approaches that you think apply to learning with open data?

How open data is used in learning

Does the nature of the raw data cause problems or work for you? What tasks do you have to do in order to get the data in the form you want it to be? What tasks do you expect the learner to do?

Do you see a tension between this cleaning or wrangling work, and the achievement of learning outcomes for the students? Or are there opportunities in it?

Is visualisation important in the way you approach learning with open data? If so, how is visualisation used?

Is making the data have relevance to the audience important? If so, how do you make this possible?

Is feedback important for the learners? If so, how does this occur and are there common themes where learners benefit from it

Is collaboration or discussion amongst the students important? If so, how do you support it?

Learner participation and awareness

Do you think the activities of learners could have value for the wider use of open data?

Do you see potential for your learners to participate in extending the uses of open data during your teaching?

What are the challenges to them actively participating in a meaningful way?

Do you think that learners have a greater awareness or involvement with open data as a result of your teaching?

In an ideal world, what would your learners do with open data as a result of your teaching?

AECT

Conclusion

Do you have any other comments?
Thank you for your time.

References

Atenas, J., & Havemann, L. (Eds.). (2015). *Open data as open educational resources: Case studies of emerging practice*. London: Open Knowledge: Open Education Working Group.

Atenas, J., Havermann, L., & Priego, E. (2015). Open data as open educational resources: Towards transversal skills and global citizenship. *Open Praxis, 7*(4), 377–389. https://doi.org/10.5944/openpraxis.7.4.233.

Belland, B. R., Kim, C., & Hannafin, M. J. (2013). A framework for designing scaffolds that improve motivation and cognition. *Educational Psychologist, 48*(4), 243–270.

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal, 9*(2), 27–40.

Bradley, J. C., Lancashire, R. J., Lang, A. S., & Williams, A. J. (2009). The spectral game: Leveraging open data and crowdsourcing for education. *Journal of cheminformatics, 1*(1), 9. https://doi.org/10.1186/1758-2946-1-9.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101.

Butcher, N. (2015). A basic guide to open educational resources (OER). Burnaby, Canada: Commonwealth of Learning (COL). Retrieved April 15, 2019, from http://unesdoc.unesco.org/images/0021/002158/215804e.pdf.

Ciociola, C., & Reggi, L. (2015). A Scuola di OpenCoesione: From open data to civic engagement. In J. Atenas & L. Havemann (Eds.), *Open data as open educational resources: Case studies of emerging practice* (pp. 26–37). London: Open Knowledge: Open Education Working Group.

Cordova, D. I., & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology, 88*(4), 715–730.

Coughlan, T. (2015). Using open data as a material for introductory programming assignments. In J. Atenas & L. Havemann (Eds.), *Open data as open educational resources: Case studies of emerging practice* (pp. 39–49). London: Open Knowledge: Open Education Working Group.

Coughlan, T., Carletti, L., Giannachi, G., Benford, S., McAuley, D., Price, D., Locatelli, C., Sinker, R., & Stack, J. (2015). ArtMaps: interpreting the spatial footprints of artworks. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 407–416). ACM.

Creative Commons. (2016). What is OER?. Retrieved April 15, 2019, from https://wiki.creativecommons.org/wiki/What_is_OER%3F.

Crusoe, D. (2016). Data Literacy defined pro populo: To read this article, please provide a little information. *Journal of Community Informatics, 12*(3), 27–46.

De Jong, T., Linn, M. C., & Zacharia, Z. C. (2013). Physical and virtual laboratories in science and engineering education. *Science, 340*, 305–308. https://doi.org/10.1126/science.1230579.

Dunwell, I., Dixon, R., Bul, K. C., Hendrix, M., Kato, P. M., & Ascolese, A. (2016). Translating open data to educational minigames. In *11th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP),* (pp. 145–150). New York, NY: IEEE.

Edelson, D. C., Gordin, D. N., & Pea, R. D. (1999). Addressing the challenges of inquiry-based learning through technology and curriculum design. *Journal of the Learning Sciences, 8*(3–4), 391–450.

FitzGerald, E., Kucirkova, N., Jones, A., Cross, S., Ferguson, R., Herodotou, C., et al. (2017). Dimensions of personalisation in technology-enhanced learning: A framework and implications for design. *British Journal of Educational Technology, 49*(1), 165–181. https://doi.org/10.1111/bjet.12534.

Gapminder. (2018). For Teachers, Retrieved April 15, 2019, from https://www.gapminder.org/for-teachers/.

Gray, D. (2014). Putting undergraduates on trial: Using the Old Bailey Online as a teaching and assessment tool. *Law, Crime and History, 4*(1), 104–113.

Herrington, A., & Herrington, J. (2006). *Authentic learning environments in higher education*. Hershey, PA: Information Science Publishing.

Hung, D., Lee, S. S., & Lim, K. Y. (2012). Authenticity in learning for the twenty-first century: Bridging the formal and the informal. *Educational Technology Research and Development, 60*(6), 1071–1091. https://doi.org/10.1007/s11423-012-9272-3.

Kippers, W. B., Poortman, C. L., Schildkamp, K., & Visscher, A. J. (2018). Data literacy: What do educators learn and struggle with during a data use intervention? *Studies in Educational Evaluation, 56,* 21–31.

Knupfer, N. N., & McLellan, H. (1996). Descriptive research methodologies. In D. Johnassen (Ed.), *Handbook of research for educational communications and technology* (pp. 1196–1212). New York: Macmillan.

Land, S. M., & Greene, B. A. (2000). Project-based learning with the World Wide Web: A qualitative study of resource integration. *Educational Technology Research and Development, 48*(1), 45–66. https://doi.org/10.1007/BF02313485.

Lazonder, A. W., & Harmsen, R. (2016). Meta-analysis of inquiry-based learning: Effects of guidance. *Review of Educational Research, 86*(3), 681–718.

Love, M., Boisvert, C., Uruchurtu, E., & Ibbotson, I. (2016). Nifty with data: can a business intelligence analysis sourced from open data form a nifty assignment? In *Proceedings of the 2016 ACM Conference on Innovation and Technology in Computer Science Education,* (pp. 344–349). New York, NY: ACM.

Manca, A., Atenas, J., Ciociola, C., & Nascimbeni, F. (2016). Critical pedagogy and open data as a means for educating to social cohesion. *Italian Journal of Educational Technology, 25*(1), 111–115.

Mandinach, E. B., & Gummer, E. S. (2013). A systemic view of implementing data literacy in educator preparation. *Educational Researcher, 42*(1), 30–37.

Mittelmeier, J., Rienties, B., Tempelaar, D., Hillaire, G., & Whitelock, D. (2018). The influence of internationalised versus local content on online intercultural collaboration in groups: A randomised control trial study in a statistics course. *Computers & Education, 118,* 82–95.

Open Data Charter. (2015). Principles. Retrieved April 15, 2019, from https://opendatacharter.net/principles/.

Open Data Institute. (2017). What is open data and why should we care?, Retrieved April 15, 2019, from: https://theodi.org/what-is-open-data.

Open Knowledge International. (n.d.). Why Open Data? Retrieved April 15, 2019, from: https://okfn.org/opendata/why-open-data/.

Rolfe, V. (2012). Open educational resources: staff attitudes and awareness. *Research in Learning Technology, 20,* 1–13. https://doi.org/10.3402/rlt.v20i0/14395.

Shadbolt, N., O'Hara, K., Berners-Lee, T., Gibbins, N., Glaser, H., & Hall, W. (2012). Linked open government data: Lessons from data.gov.uk. *IEEE Intelligent Systems, 27*(3), 16–24.

Shamash, K., Alperin, J. P., & Bordini, A. (2015). Teaching Data Analysis in the Social Sciences: A case study with article level metrics. In J. Atenas & L. Havemann (Eds.), *Open data as open educational resources: Case studies of emerging practice* (pp. 49–55). London: Open Knowledge: Open Education Working Group.

Siemens, G. (2005). Connectivism: A learning theory for the digital age. *International Journal of Instructional Technology and Distance Learning, 2*(1), 3–10.

Weller, M. (2015). *The Battle for Open: How openness won and why it doesn't feel like victory*. London: Ubiquity Press.

Wolff, A., Cavero Montaner, J. J., & Kortuem, G. (2016). Urban data in the primary classroom: Bringing data literacy to the UK curriculum. *Journal of Community Informatics, 12*(3), 57–82.

World Bank. (2017). Open Data in 60 Seconds, Retrieved April 15, 2019, from http://opendatatoolkit.worldbank.org/en/open-data-in-60-seconds.html.

World Bank Data Catalog. (2018). Retrieved April 15, 2019, from https://datacatalog.worldbank.org/.

Worthy, B. (2015). The impact of open data in the UK: Complex, unpredictable, and political. *Public Administration, 93*(3), 788–805.

Tim Coughlan is a lecturer in the Institute of Educational Technology of The Open University. His research explores inclusive and open learning from a Human–Computer Interaction perspective.

AECT