Why should we care about datafication? Critical data literacies in higher education

OER20 Workshop Instruments and Results

_____

Authors: Bonnie Stewart, Juliana Raffaghelli

- A short workshop session offered at the OER20 [online] conference
- https://oer20.oerconf.org/sessions/o-023/
- Wed, Apr 1 2020
- Theme: Open education for civic engagement and democracy
- Recorded Session

1. Session Description

In a world of "smart" appliances and constant surveillance, teachers and learners can't engage in digital spaces – socially or educationally – without having to deal with questions of data. Whether as web tracking or as analytics collected by institutional systems, datafication surrounds us in higher education, and our knowledge-making and knowledge-dissemination systems are rife with datafication implications (Williamson, 2018). But do those of us who work in open education and

so-called higher learning understand these implications? And if we don't, what does that mean for the relationship between higher education and knowledge?

Through a hands-on Open Space conversation, facilitated via Mentimeter (even if the face-to-face version encompassed stick notes and posters), we explored the challenges datafication poses for educators in our contemporary information ecosystem, and why all of us should care. The session tried to scaffold frameworks that offer participants a critical lens to analyse their own data literacies and explore pathways to data literacy and data activism in institutions and networks. Participants contributed – with overt GDPR-compliant consent – to data visualization activities that will serve to guide future research into educators' data literacies. We also enabled contribution through hashtagged conversation on Twitter and with this document, we are also sharing the session results openly.

The session opened with a brief overview of datafication and why it matters, and to whom. We'll examine the premises and promises of big data, as well as the limitations they place on learning. We'll explore the concerns regarding surveillance, bias and exclusion connected to data-driven practices that are beginning to emerge in scholarship (Zuboff, 2015; Noble, 2018; Gilliard & Culik, 2018) and in popular media (Brown, 2017; Schwab, 2019), as well as the urgent question of what to do about non-governmental platforms, such as Facebook, that wield society-wide powers.

This Open Space "lab" is based on the authors' research into data literacy and faculty development. We'll showcase our study on how the concept of data literacy circulates in contemporary literature, and then engage participants in reflective data visualization of practices and possibilities. We're particularly interested in exploring participants' perceptions of the relationship(s) between open educational practices and data literacies, so activities will address students' data and open data within open education contexts. Our aim is to build together toward more complex and critical understandings of datafication among educators, particularly open educators, in this age of surveillance.

2. Results

The workshop was technically implemented with success, and all the interactions were undertaken without any imprevist, due to the great support given by the OER20 committee.

An amazing OER20 community got engaged with our proposal: 94 participants were visible in the chat side and actively contributing. Amongst the registered places we had: Tunisia; London UK; NYC US (2); Boston US; NJ US; Milton Keynes UK; Middle of England; Brussels, Belgium; London and Essex UK; Oklahoma, US; Lübeck, Germany; Glasgow, Scotland (2); Detroit, US; Windsor Ontario Canada; Hong Kong; Sligo, Ireland; Colorado, US; Berlin, Germany; Jacksonville, US; Dublin, Ireland; Oxford UK; Pretoria South Africa; Minneapolis US, Providence, Rhode Island, US; Adelaide, Australia; Leamington Spa, Central Scotland; Ottawa, Canada; Athens, US; Isle of Wight, UK; Nogojiwanong/Peterborough, UK; Zurich Switzerland; Norfolk, US; Swrthmore, US; Montevideo, Uruguay.

2.1. Conceptual Introduction

We briefly introduced the problem of datafication in the society and education, through two simple slides referring to each of the topics above mentioned.

We pointed out that the workshop could sound far from the COVID19 and all edtech community current concerns on how to digitise education in a hurry...or hurricane. However, the actual truth is that the problem of datafication is here more than ever, for the digital platforms we live by are collecting more data than ever, and predictive models could be fed ad infinitum for monetization.

Algorithms and models, allegedly based on objective data. To this regard, we brought several contributions (first slides) exploring not only the issues of automatization, like the case of Weapons of Math Destruction(O'Neill, 2016), Algorithms of Opression (Noble, 2018), Automating Inequalities (Eubanks, 2019), but particularly the way in which this phenomenon is re-organizing into Surveilliance Capitalism (Zuboff, 2019). Also personal experiences are triggering our reflection in this sense. In the image reported in the first slide, a family baby picture scanned by error with Google Lens, enacts an algorithm offering products that offer a "politically corrected"

version with regard to the original version scanned. Behind this scene, the data is objectively that of a red T-SHIRT with some inconvenient or unknown faces for some cultural contexts. Therefore, the invisible layer is that of the algorithm programmer, which these days will probably come from a white, male, western cultural matrix. However, we also reported a text coming from a more "enthusiastic" view of big data, Everybody Lies (Stephens-Davidowitz, 2017), where epidemiological and sociological studies could be supported by the traces of people behavior left on the web.

Education could not escape this state of things. The second slide reported two eloquent pictures (the first one on the Internet of Toys, the second one about facial tracking and biometric data associated with learners' motivation, concentration, cognitive processes, etc.). Platforms, Apps and Internet of Things we live by. The question we pose to ourselves and the audience is: do all this paraphernalia reveal "the truth" about learning, being this a complex process? If so, at which price? Which are the trade-offs of data in education? We are not alone in this reflection since there are many researchers already contributing to this state of play (Lupton & Williamson, 2017; Markham, 2018; Perrotta & Williamson, 2018; Williamson, 2017, just to mention some works)

It looks that we as educators need to prepare ourselves for this situation and to make the right choices not only in selecting the technologies with the best affordances for our students (as we use to do ten years ago) but to think about the issues behind data collection embedded in those same platforms. Some of the social media platforms we used to love to disseminate or contribute to shared and collective knowledge. Like we once reflected on media literacy, data literacy is already a crucial part. As we showed in the third slide, we discovered in a recent research led by Bonnie and me, that this focus is far from the idea of what students and educators need, to be "data literate". We conducted a systematic review of the literature over 132 out of 386 papers on data literacy and we found much of the debate focused on technical skills and school data management, even research data management, but little about a critical perspective of datafication as a social process and problem. This phenomenon is shown in the keywords map, which terms and clusters highlight the prevalence of terms associated with data-based decision making at school, research data management, data literacy in science and technology.

Therefore, we considered the importance of sharing these ideas and opened the debate with an expert audience. Reflecting together, we will move the agenda of research and practice around this problem.
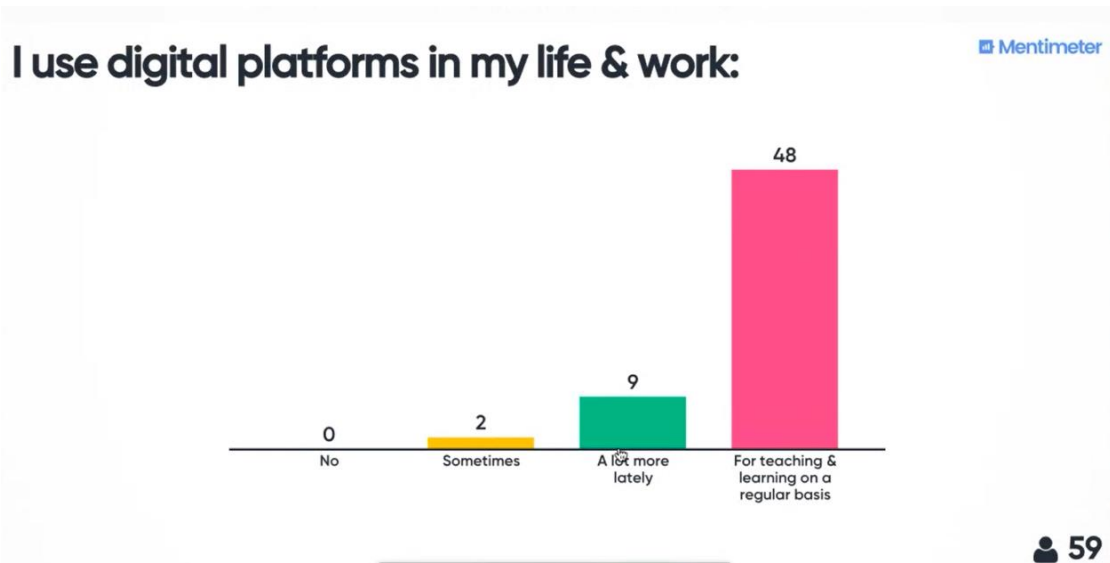
2.2. Action!

In spite of the medium, and the initial idea of interacting using sticky notes and a poster only generating analogical interactions, we were constrained to interact using Blackboard an Mentimeter. We collected participants' reactions and are now sharing all the workshop results in the open for you to cite and re-use. Informed Consent was collected, but it was highlighted that Mentimeter does not collect any personal information and it's [GDPR compliant](#).
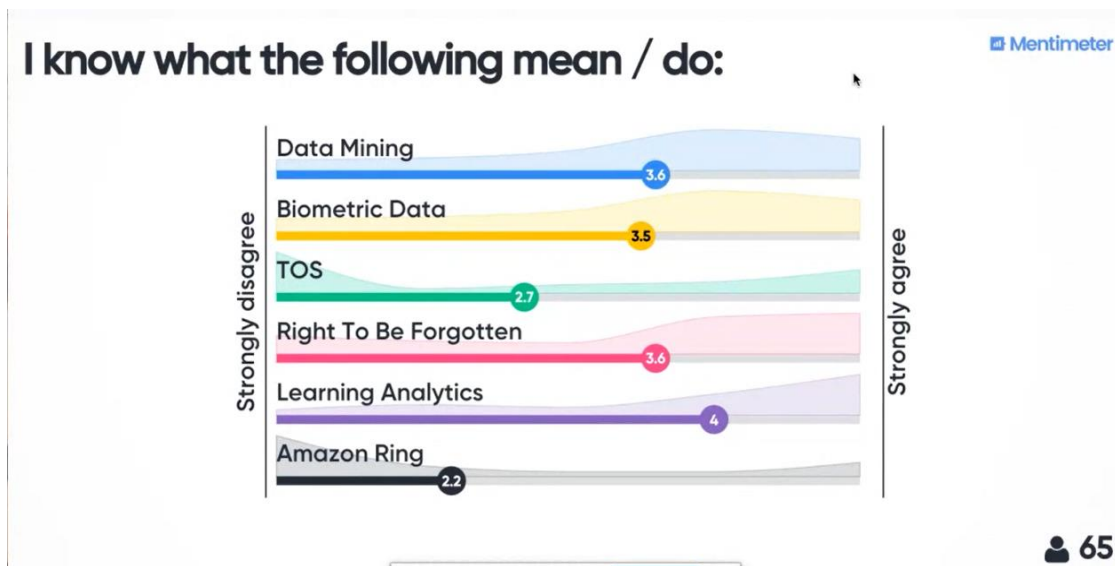
Over 101 people taking a look at the Mentimeter interactions, between 52 and 67 got engaged with the activities.

In the following, we present the interactions and comment on them.

In the **firstinteraction** our aim was to raise awareness about the "platforms we live by" that is, all those artifacts that are collecting data and we voluntarily/purposely allowed into our daily/working life. The question asked for the frequency of usage, and specified the type of platform for teaching and learning since it's supposed to be the main activity of the participants. Not surprisingly, most participants frequently use teaching and learning platforms on a regular basis (48/59).

I use digital platforms in my life & work:

The **second interaction** attempted to explore together the participants awareness about their own attention to the terms and techniques of a datafied society/education: Data Mining, Biometric Data, TOS, Right to be Forgotten, Learning Analytics, Amazon Ring. In a scale of 1 (Strongly Disagree) 5 (Strongly Agree) most participants (over 65 responses) expressed to be sure about their own knowledge on Learning Analytics (4/5), rather know the meaning of data mining (3.5), right to be forgotten and biometric data (3.6). But the TOS (2.7) and the Amazon Ring (2.2) remained more obscure objects.

I know what the following mean / do:

At each of the results, there were several comments by the audience, the conversations overlapped as in a complex spider web, but it was really a human net interacting with energy and

passion. One can see how much the audience learns from interacting over a presenter's idea, further and further that idea. Participants bring resources, express concerns, react emotionally and particularly get insights from the others' reactions and resources. This is beyond the presenter's original input, like a positive chain reaction.

Some of the ideas expressed where:

_____

TERMS OF SERVICE

- 00:11:38.000 --> 00:11:38.900 <jw>what is TOS
- 00:11:47.000 --> 00:11:47.900 <ke>terms of service, I think
- 00:11:51.000 --> 00:11:51.900 <jg>yep
- 00:11:56.000 --> 00:11:56.900 <a>Terms of Service ToS
- 00:12:28.000 --> 00:12:28.900 <j>Ah ok thanks. I always find acronyms tricky
- 00:15:14.000 --> 00:15:14.900 <sm>the things I click to accept but don't really read
- 00:15:38.000 --> 00:15:38.900 <c>I have before, but not typically // <gw>They're too long to read!
- <gd>Lol maybe once // <j>Bits of it // <na>I have but I'm a policy wonk
- 00:15:41.000 --> 00:15:41.900 <ap>I have once or twice, intensely boring to read!
- 00:15:47.000 --> 00:15:47.900 <mw>i read ALL of them Bonnie // <m>Rarely. Turnitin once.
- 00:16:32.000 --> 00:16:32.900 <dv>they are not written to be read (by humans) 😬
- 00:17:03.000 --> 00:17:03.900 <jg>It's mostly CYA (cover your...) language, trademark stuff, protection from legal prosecution, basically rendering themselves nigh invincible to issues // <je>Theres that site that condenses them for you
- 00:17:06.000 --> 00:17:06.900 <ap>yes, 'not really meant to be read' - usually quite legalistic in language used (not accessible to lay people)
- 00:23:45.000 --> 00:23:45.900 <v guest #3>:thumbs up on tosdr.org Use all the time

DATA MINING

- 0:13:28.000 --> 00:13:28.900 <gr>Extractive - that's a good term
- 00:13:38.000 --> 00:13:38.900 <M>Data Mining as the process and goal of an economy
- 00:14:00.000 --> 00:14:00.900 <jg>Data mining also goes into doing in-depth queries / analysis of the data you have

- 00:14:11.000 --> 00:14:11.900 <jh>The cross correlation between different sources is the really scary thing

BIOMETRICS

- 00:14:18.000 --> 00:14:18.900 <jg> The first time I ever saw data mining in-person was on biometric data at a biogenetics lab!
- 00:14:23.000 --> 00:14:23.900 <M>self-tracking
- 00:14:26.000 --> 00:14:26.900 <ke>Anyone hearing about the thermometer which sends temp to the cloud, then surveillance and COVID?
- 00:14:56.000 --> 00:14:56.900 <M>@ke: no, do you have a link to the story=
- 00:15:13.000 --> 00:15:13.900 <ke>I will see if I can find it. Not a 'scholarly' source. ☺
- 00:15:52.000 --> 00:15:52.900 <ke>https://www.nytimes.com/2020/03/18/health/coronavirus-fever-thermometers.html
- 00:21:22.000 --> 00:21:22.900 <lc>lol smart scales would be shouting at us all right now!
- 00:21:29.000 --> 00:21:29.900 <sb> i try to avoid "smart" things but it's getting harder and harder to do

RIGHT TO BE FORGOTTEN

- 00:17:52.000 --> 00:17:52.900 <tm>there is an angle in right to be forgotten which is able being able to edit your past if you have a skeleton in your cupboard...
- 00:18:30.000 --> 00:18:30.900 <gr> Should historians be able to recall the "truth" of your skeletons?

LEARNING ANALYTICS

- 00:17:32.000 --> 00:17:32.900 <M #2>LEarning Analytics: use computers instead of teachers to teach
- 00:17:33.000 --> 00:17:33.900 <v gr>Data mining for learner behaviour
- 00:17:58.000 --> 00:17:58.900 <mo>usage of data input
- 00:18:07.000 --> 00:18:07.900 <jg>Learning Analytics: 1. (Common) a marketing buzzword. 2. [every other definition]
- 00:18:13.000 --> 00:18:13.900 <sb>learning analytics is such a big selling feature/advertising point of so many edtech things these days
- 00:18:19.000 --> 00:18:19.900 <ke>LA -- things that, because they are quantifiable, push us to assess things that are quantifiable.

- 00:18:23.000 --> 00:18:23.900 <M #2>Learning Analytics is conservative (only the past counts)
- 00:18:26.000 --> 00:18:26.900 <A>Some may argue that mining is seperate from analytics. Mining is the collection and analytics is the analysis
- 00:18:29.000 --> 00:18:29.900 <na>"what do we do with it and who owns it" - yes! This is the question
- 00:19:56.000 --> 00:19:56.900 <j>learning analytics are becoming more and more complex as adding all sorts of obscure monitoring to students, but clearly not every student is an object of study, only those that "need to be studied"
- 00:20:37.000 --> 00:20:37.900 <jw>how to data mine moodle would be good - its hard
- 00:21:31.000 --> 00:21:31.900 <vr>Do you think that it is possible to carry out learning analytics from the good?
- 00:22:17.000 --> 00:22:17.900 <jg>My answer is actually "it depends" here
- 00:22:21.000 --> 00:22:21.900 <sb>because the data is useful I'm assuming
- 00:22:22.000 --> 00:22:22.900 <mc>my answer there would be "assume the new platform is probably evil"
- 00:22:24.000 --> 00:22:24.900 <na>Yessss the "attentiveness" feature of zoom
- 00:23:04.000 --> 00:23:04.900 <jg>for example, would my use only affect me? If so, that's less of a concern. What kind of data would I share? What would I be okay with making the news?
- 00:23:07.000 --> 00:23:07.900 <j>@vr is hard to do them as they basically review interaction with devices, in that case we need qualitative - human data
- 00:24:18.000 --> 00:24:18.900 <vr>@jwe have already developed a national wide learning analytics project in Uruguay, I think it can be for the good
- 00:24:53.000 --> 00:24:53.900 <vr>but depends of the objectives and ideology and pedagogical concepts of the people involved
- 00:25:32.000 --> 00:25:32.900 <vr>It has to be educational research using learning analytics tools
- 00:23:42.000 --> 00:23:42.900 <A>@gr - FB does not own Zoom but they were sharing data with them - even if you did not have F
- 00:24:03.000 --> 00:24:03.900 <km>I've seen a few people tweeting about Jitsi - open source?
- 00:24:04.000 --> 00:24:04.900 <sk #3>More new edtech as a parent, than as an educator
- 00:24:05.000 --> 00:24:05.900 <ca>zoom as big issues on privacy, particularly in Mac
- <v Josh Hapern #2>The ALT forum is a wonderful place to discover new and useful EdTech aps. Join ALT

- 00:24:43.000 --> 00:24:43.900 <vsb #2>Zoom - although I think more needs to be done regarding privacy training/professional development about the tool.

The RING

- 00:13:35.000 --> 00:13:35.900 <ke>@nicole did you come across some of the craziness with Ring/police/surveillance? // <j>That Ring is scariest than the japanese film
- 00:19:29.000 --> 00:19:29.900 <v gr>Any employer ever look at your house on Street View?
- 00:19:34.000 --> 00:19:34.900 <lr>Amazon Ring is the net 2.0 of Castle Doctrine. ☺
- 00:19:53.000 --> 00:19:53.900 <tr>But then sometimes you catch your neighbors running naked through your yard in the middle of the night 😄
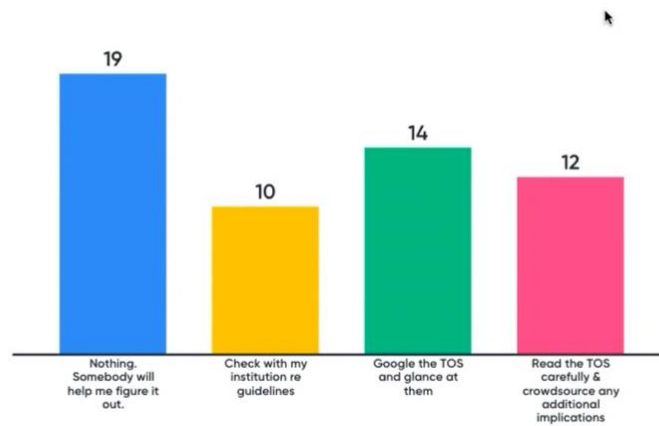
...AND MORE ON ARTIFICIAL INTELLIGENCE

- 00:21:34.000 --> 00:21:34.900 <tr>Also have Ecobee, Alexa, Firestick
- 00:21:41.000 --> 00:21:41.900 <ap>I won't have alexa or similar in my home, feels too intrusive (and colleagues set off each other's alexas in online #WFH meeting last week) // <jg>Smart thermostats are irresistible to me being in a warmer climate
- 00:21:44.000 --> 00:21:44.900 <ke>lightbulbs, lock, thermostat, house is smarter than I am

There is no need to punctualize further the concern demonstrated by the participants, as educators for the lack of ethics in artificial intelligence, data and algorithms. The concern about the naïve idea of using learning analytics to take complex decisions about the students. As a very interesting element, the participants realized their lack of engagement with the ToMs as an obscure place where the conditions about users' data could be placed. However, the experiences shared also pointed out the opacity of ToMs and the impossibility for the users to quickly engage with them in an agentic way (taking the right decision).

The perspective was completed by a wordcloud on the actual devices the participants owned. For sure the mobiles take the most, with thousands of available apps tracking data. But artificial intelligence seems to have entered participants' life through other objects supporting the IoT (Internet of Things), particularly at home. How could we escape datafication in this landscape?

Which is the least to have more control over this situation? Maybe, to decide and pick or stay away from specific technologies after reading the ToMs.

Our next step was to ask about the instruments used as "edtechies". After reflecting about the invisibilities of datafication in our daily life, the time to analyse the situation about learning platforms arrived.

We appreciated the honesty of those 19/55 that expressed that we mostly use what is convenient for our professional purposes with little reflection about how data is dealt with. Expecting that the decision could be taken at institutional level, a strategy that will probably fail since institutions make contracts with companies offering convenient or even "glossy" services to the eyes of the students. The group of 10 telling that they are reliant on the institution support somehow align in the sense that giving up control over the tools they use.

I want to work with a new digital platform. I do the following re data:

Mentimeter

- 19 — Nothing. Somebody will help me figure it out.
- 10 — Check with my institution re guidelines
- 14 — Google the TOS and glance at them
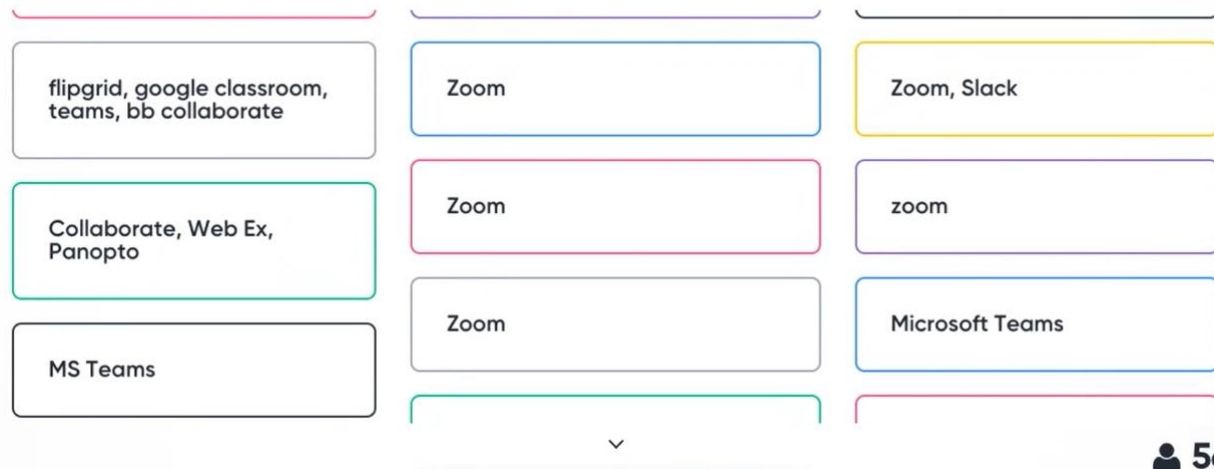- 12 — Read the TOS carefully & crowdsource any additional implications

👤 55

The rest of the group (28/55) try to exert some control over the platforms, in a more direct way (12) or by basing on the collective knowledge about the tools (Googling, 14).

And yes! There is the turn of COVID19 which also appeared here in the sense we introduced: are we connecting the use of certain tools to the data we give for free to someone?

Edtech tools I've been trying out or using more due to COVID-19 include:

Mentimeter

- flipgrid, google classroom, teams, bb collaborate
- Collaborate, Web Ex, Panopto
- MS Teams
- Zoom
- Zoom
- Zoom
- Zoom, Slack
- zoom
- Microsoft Teams

👤 56

The discussion had already appeared in the initial phase of exploration of tools: What about Zoom and the practices of analizing "attentiveness". In spite of these features, there was a

stunning prevalence of Zoom (26/56) over BBCollaborate (7/56), as just a simple example of the fact that maybe we are not considering those terrible ToMs in depth. Jitse and Moodle, two tools that enable user to exert more control over data and are Open Source, are less preferred instead (Jitsi: 3/56; Moodle: 2/56). The transparency of algorithms and procedures is still an idea in the future. Should a critical approach to the technologies consider this issues? Is this affordable for the educators' life within institutions? Are all forms of data collection by the companies offering services in education a bad thing?

These seem to be reflections for the future. But it should be the near future.

The participants left interesting thoughts by the end.

- 00:26:33.000 --> 00:26:33.900 <S>letting them know that everyone is on the same learning curve and that we are all experiencing something new at the same time. Also, if we take the approach that we can teach students digital literacies if we improve and understand digital literacies ourselves.
- 00:26:39.000 --> 00:26:39.900 <Mi>Seeing a lot more superb critical engagement now the decision is *what* to use, rather than *whether* to use
- 00:26:43.000 --> 00:26:43.900 <tr>Ethics is important; it would also help to frame the importance of data in a more applied way--how does it help? // <v ALT Events>Get them to become an online student !
- 00:27:00.000 --> 00:27:00.900 <tm>we need our eyes open!
- 00:27:08.000 --> 00:27:08.900 <jg>@S is so right. I always feel like I'm the least data-literate in any room, and knowing that more of us are on the same page is important.
- 00:27:12.000 --> 00:27:12.900 <A>I worry some of us will be forced into data literacy the hard way
- 00:27:38.000 --> 00:27:38.900 <tr>@Jeff nope, Idefinitely the least data literate here!
- 00:27:39.000 --> 00:27:39.900 <tr>@jg nope, I'm definitely the least data literate here!
- 00:27:42.000 --> 00:27:42.900 <dvo>The aspiration to enable "Personalized learning" seems to be driving data extraction from student to a considerable extent. Here an activity to help educators reflect on the risks https://eduhack.eu/course/area-4/activity-3/

These expressions (and the excellent resources!) let us understand the part of the learning curve we are placed so far: the very beginning.

5. Conclusions

In this short 30 minutes session we attempted to reflect about the literacies needed in a context that is increasingly "datafied": we freely decide to use platforms, apps, IoT devices that end up in opaque practices connected to the monetization of data. The fact is that more agentic practices seem to require further awareness and as far as one of the participants expressed:

> ...if we take the approach that we can teach students digital literacies if we improve and understand digital literacies ourselves.

Initially a very good way to approach the problem is to include practices of auto-ethnography as educator, or at least to open a reflective practice with regard to what is behind the platforms that we live by. And particularly the platforms we work with.

As another participant pointed out, the risk of not engaging critically with data tracking practices is that:

> I worry some of us will be forced into data literacy the hard way

And, no, this shouldn't be the idea if faculty engages in forms of activism starting by trying, exploring, supporting others in making good technological choices. In other participant's words:

> Seeing a lot more superb critical engagement now the decision is *what* to use, rather than *whether* to use

Not to worry about one's initial lack of knowledge or awareness, for against an emerging problem, we all are inventing the next future:

> I always feel like I'm the least data-literate in any room, and knowing that more of us are on the same page is important.

Thank you for being with us!

Resources

- <u>Slide deck</u>
- Mentimeter Results

References

Lupton, D., & Williamson, B. (2017). The datafied child: The dataveillance of children and implications for their rights. *New Media & Society*, *19*(5), 780–794. https://doi.org/10.1177/1461444816686328

Markham, A. N. (2018). Critical Pedagogy as a Response to Datafication. *Qualitative Inquiry*, 107780041880947. https://doi.org/10.1177/1077800418809470

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism by Safiya Umoja Noble. NYU Press*. https://doi.org/10.15713/ins.mmj.3

O'Neil, C. (2016). *Weapons of math destruction : how big data increases inequality and threatens democracy*. New York: Penguin.

Perrotta, C., & Williamson, B. (2018). The social life of Learning Analytics: cluster analysis and the 'performance' of algorithmic education. *Learning, Media and Technology*, *43*(1), 3–16. https://doi.org/10.1080/17439884.2016.1182927

Stephen-Davidowitz, S. (2017). Everybody Lies. Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are. New York: Harper Collins.

Schwab, K. (January 6, 2019). IoT security is so bad, many companies can't tell when they're hacked. Fast Company. Retrieved from

https://www.fastcompany.com/90292568/iot-security-is-so-bad-many-companies-cant-tell-when-theyre-hacked

Williamson, B. (2017). Educating Silicon Valley: Corporate education reform and the reproduction of the techno-economic revolution. Review of Education, Pedagogy, and Cultural Studies, 39(3), 265–288. https://doi.org/10.1080/10714413.2017.1326274

Williamson, B. (2018). The hidden architecture of higher education: building a big data

infrastructure for the 'smarter university.' International Journal of Educational Technology in Higher Education, 15(1), 12. https://doi.org/10.1186/s41239-018-0094-1

Zuboff, S. (2015). Big other: Surveillance Capitalism and the Prospects of an Information Civilization. Journal of Information Technology, 30(1), 75–89. https://doi.org/10.1057/jit.2015.5