Data Literacy in The Smart City

Jane Kunze[1]

[1] Aarhus Public Libraries - Dokk1

Keywords: data, big data, data literacy, civic engagement, smart cities, Aarhus

Abstract

Aarhus Public Libraries has been working with children collecting and analyzing data about the sound of Aarhus using micro:bits. The aim is to train the students' data literacy and at the same time have them contributing to a bigger 'data story' about the city. Creating awareness around the massive amounts of data that are produced by and around us, as well as inspiring local communities to utilize this data is important, if we want the digitalization of our cities to be not only smart, but also democratic.

1 | Smart cities should be for everyone!

In 2018 Aarhus Public Libraries received funding by the Agency for Culture and Palaces for a project called Data Democracy. We set out to investigate if and how the public libraries could play a role in creating transparency and agency for citizens in the data economy and the derivative consequences on Society.

This taps into the vision of Aarhus becoming a smart city, that not only seeks the benefits of the data economy and IoT, but also values active citizenship and providing access to data about the city (https://www.smartaarhus.eu).

A vision for a smart city like this presupposes citizens understanding the 'infrastructure' of the smart city, so they can contribute to and think critically about information and decisions deriving from data within the 'infrastructure'. The digital skills required to do this are not trivial and there is a real challenge ahead, if we want to close and not widen the already existing digital divide (van Deursen & Mossberger, 2018)

1.1. Learning to speak data

Understanding and participating in a Society that is, in both private and public regards, becoming more and more data driven, requires citizens to acquire a new vocabulary and set of digital skills. As things are now, we can talk about a data divide between those who have access to data and skills to use it and those who don't, which fosters an inequality in Society (D'Ignazio, 2017), resembling the inequality of information prior to the printed press or the Internet, where public libraries also have played an active role in providing access and informal education.

Data literacy is the ability to read, work with, analyse and argue with data and to constructively engage in society through or about data (Beyond Data Literacy, 2015) and is further expanded by the impact of big data that complicates identifying (when and where data is being collected), understanding (how data is being algorithmically analysed) and weighing (real and ethical impacts of data based decisions) (D'Ignazio & Bhargava, 2015).

In the Data Democracy project, we have been prototyping different ways of doing this; from workshops and classes to data walks and meditation, but from the beginning of the project, we had an idea of combining data literacy with some kind of participatory citizenship.

1.2. Getting hands-on with data

When it comes to digital media and the use of data today there seems to be, especially among younger consumers (age 16-24), a feeling of powerlessness in regards to managing and sharing data and an inexperience in proactively using data, for instance to support community action (Me and My Big Data Report 2020).

By having citizens collect data about urban life, we wanted to facilitate a hands-on and conscious (as opposed to most of the data we unconsciously collect through our smartphones) experience with crowdsourcing data. The intention is to foster a sense of agency for participants and provide insights into how messy and ambiguous working with raw data really is (D'Ignazio, 2017).

2 | Sweet Spots: the sound of Aarhus

Together with the Alexandra Institute (https://alexandra.dk/) we have tested a prototype for crowdsourcing data about the sound of the city. Participants in our workshops, mostly school children between 12-16 years, get experience collecting, analysing, and telling stories with data. The data is simultaneously visualized on a city map (datadokken.dk), which in the longer run, we hope, can inspire a wider public conversation about life in the smart city and (big) data literacy.

The aim of the prototype is to tell an alternative story about living in Aarhus. Focusing on data on sound levels, we wanted to avoid mapping sources of noise, not because this is irrelevant, but because sensors around the city are already doing this kind of monitoring.

Figure 1: Picture from a workshop with students setting up the sound transmitter.

Instead, we wanted to collect data about sound as something that adds an extra dimension - a soundscape for our urban life, that sometimes supports a cosy or relaxed atmosphere, energizes us or provides comfort. That is something that cannot be measured on a simple scale of decibel, but requires citizens assigning emotions to the sounds at different locations in the city, like mapping emotions of the city, which is a creative method already used for citizen involvement (Panek, 2018).

2.1. Technical setup

Initially we were inspired by the smartcitizen.me (https://smartcitizen.me/) project, where you purchase your own measuring station to collect data about the environment around you, like noise, but also temperature, light, humidity, particles and air quality. It comes as a small kit, that you just assemble and connect to the internet. Afterwards you can monitor the data your station is collecting through a central webpage, where you can also access the data of other measuring stations around the world. This way you can compare your local data with data from all over the world. The idea is to support citizens and researchers in raising awareness and acting on environmental issues at the local level.

Looking a little closer at the smartcitizen.me platform there are around 30 measuring stations around Aarhus - not many of them er online and more than half of them are placed at the university or other public institutions - crowdsourcing data at citizen level is still a narrow subject area and could remain so, if we don't address both the need for data literacy (create awareness) and the necessary skill (facilitate learning opportunities).

Therefore we wanted to work with a prototype that had a potentially broader reach, which was the reason for turning our attention to Micro:bits. A Micro:bit is a pocket-sized computer that is used to introduce children and youth to programming and computational thinking. In Denmark they have been distributed and introduced to many school children through the national programme Ultrabit (https://www.dr.dk/skole/ultrabit). By using the Micro.bit, we were hoping to tap into an already existing national initiative to build digital og technological literacy.

At that time (2019) the Micro:bit didn't quite have the functionalities needed to work as our sound transmitters, so a daughterboard was constructed with the most basic features being a microphone, gps sensor and more computing power. With the new Micro:bit coming out, this might become far more accessible, since the new version has a lot of the features we needed to add.

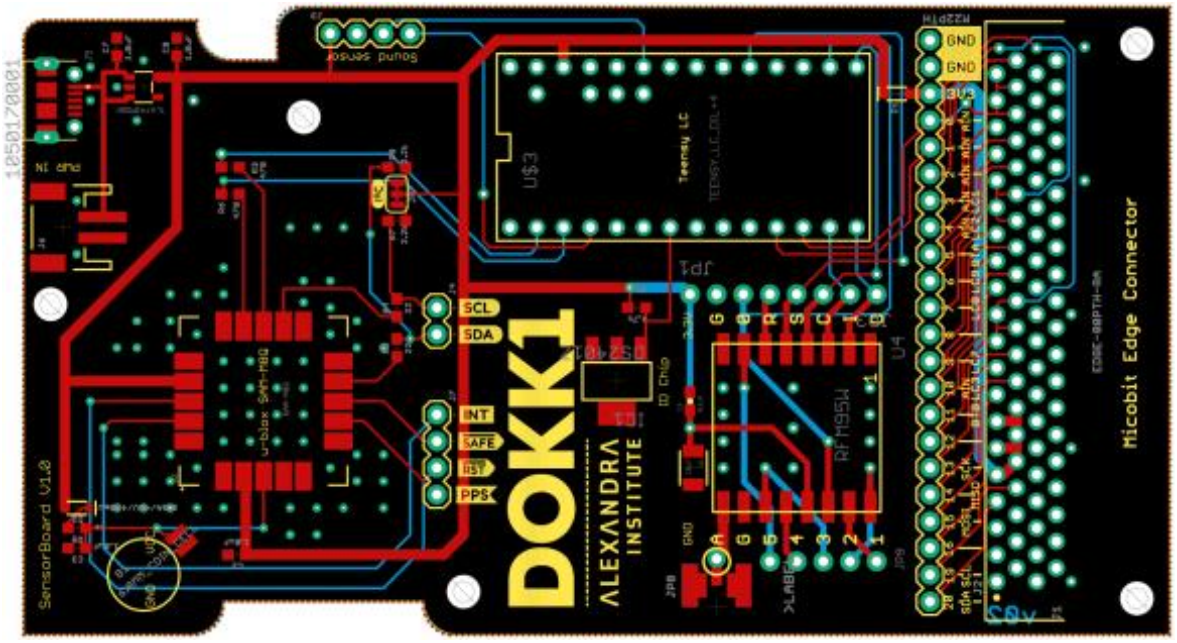

Figure 2: First draft of daugtherboard.

Working with the Micro:bit also created the opportunity for the participants to have a hand-on experience with programming the sound transmitter themselves using block programming on the Makecode platform (https://makecode.microbit.org/). We hoped that constructing the device from scratch, both assembling the hardware and programming the software, would support a sense of ownership for the participants as opposed to using for instance the smartcitizen.me kit, which is already set up for the user.

Tying the prototype to the overall idea of a smart city we wanted to transmit the data from the sound transmitters via the city's LoRa network, this way the data would also be available in real time. Using the LoRa network also made it possible to avoid possible challenges connecting to wifi, which is still not always accessible, when you are outside.

The data collected from the sound transmitters is accessible on a platform (datadokken.dk) much like the smartcitizen.me. From here the data can be explored geographically or through filters on volume and time. The data can also be exported in .csv and .json formats for further analysis and visualization.

The idea of a platform to visualize the data, was to display it at Dokk1 (that on average has 4000 visits per day), as another way of increasing awareness among citizens about data literacy and start a conversation on how crowdsourcing data might be purposeful in a city like Aarhus.

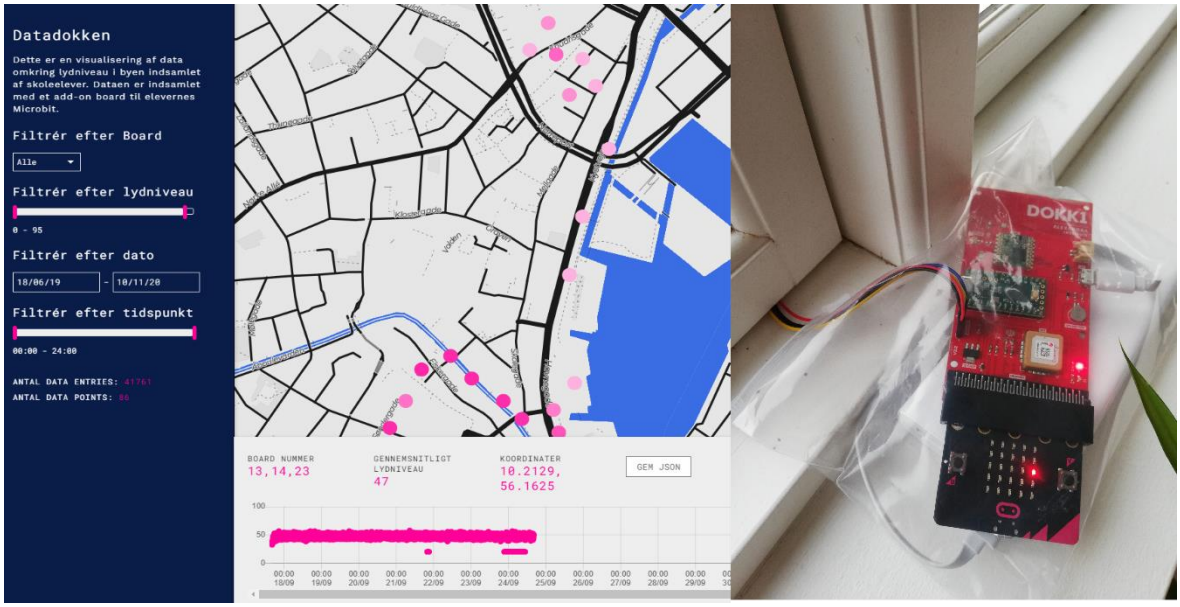

Figure 3: Screenshot from datadokken.dk. and picture of the sound transmitter in action.

Immediately, when we started collecting data, the platform also turned out to be really useful, because we were able to monitor the sound transmitters in real time, for instance to check everything was working, before the workshop participants bought the sound transmitters home with them.

2.2. Workshop setup

We wanted the participants to have a hands-on experience with the four overall elements when working with any kind of data, whether big or small; data collection, data cleaning, data analysis and data visualization (data literacy).

We also wanted them to get a more tangible experience of what big data (= big amounts of data) can look like and that it is framed and interpreted like any other kind of data and therefore should be subject to the same kind of critical assessment and evaluation (big data literacy).

Finally, we also wanted the workshop participants to have fun! Data literacy training can have a somewhat steep learning curve, because it is a rather complex field with possible detours to other fields like statistics, (visual) communication, information and computer science.

We took inspiration from the five approaches to creative data literacy; work with community-centred data, write data biographies, make data messy, build learner-centred tools and favour creative community-centred outputs over Tuftean purity (D'Ignazio, 2017).

Working with community-centred data is to create context to data and to open up to participants linking data to their everyday lives. Collecting data about sound levels of public spaces in the city you live in, is on its own community-centred, but we wanted to strengthen this even more by explicitly connecting the data collected to the feelings of the participants. That is why we named the prototype Sweet Spots, because a sweet spot is a place that is optimal for obtaining a certain desirable effect or result - in this case, a place you go because it supports a certain atmosphere or positive feeling.

Write data biographies to examine, where the data you are using comes from and by who it has been collected. This is important when working with open data or data, that you have not collected yourself. In our prototype the data analysed by participants is also collected and cleaned by the

participants. In that way, it is not necessary to work with data biographies, but the meta-reflexion on the context and purpose of a dataset is still important. To stimulate this, we have the participants write a small study design sheet, articulating where and why they choose a specific spot for collecting the data and what they are expecting to find.

Making data messy happens when you take participants through a process of collecting, sorting and categorizing data to strengthen their scepticism of data - quantitative data about the physical world tends to be taken at face value and not put to the same scrutiny as qualitative data. We try to challenge this by having participants categorize and clean the data they have collected by themselves in a simple spreadsheet exercise.

Building learner-centred tools is about designing the tool to be simple and focused, but still with a potential to expand or inspire further learning. Options and choices will be limited, which is the reason why experienced users will probably choose to work with another tool. This tactic we struggled the most with implementing, since we wanted participants to have a full data processing experience (collection, cleaning, analysis and visualizing) and feel like they were contributing a small piece to a bigger puzzle.

Favour creative community-centred outputs over Tuftean purity means that the learning experience of the participants weighs more than the results of the data collection process, this is especially important when working with citizens without any prior data skill training. Big variations in the amount and quality of data that the participants collected emphasized this tactic to us during the workshops, where we had participants draw and tell simple data stories with their collected data.

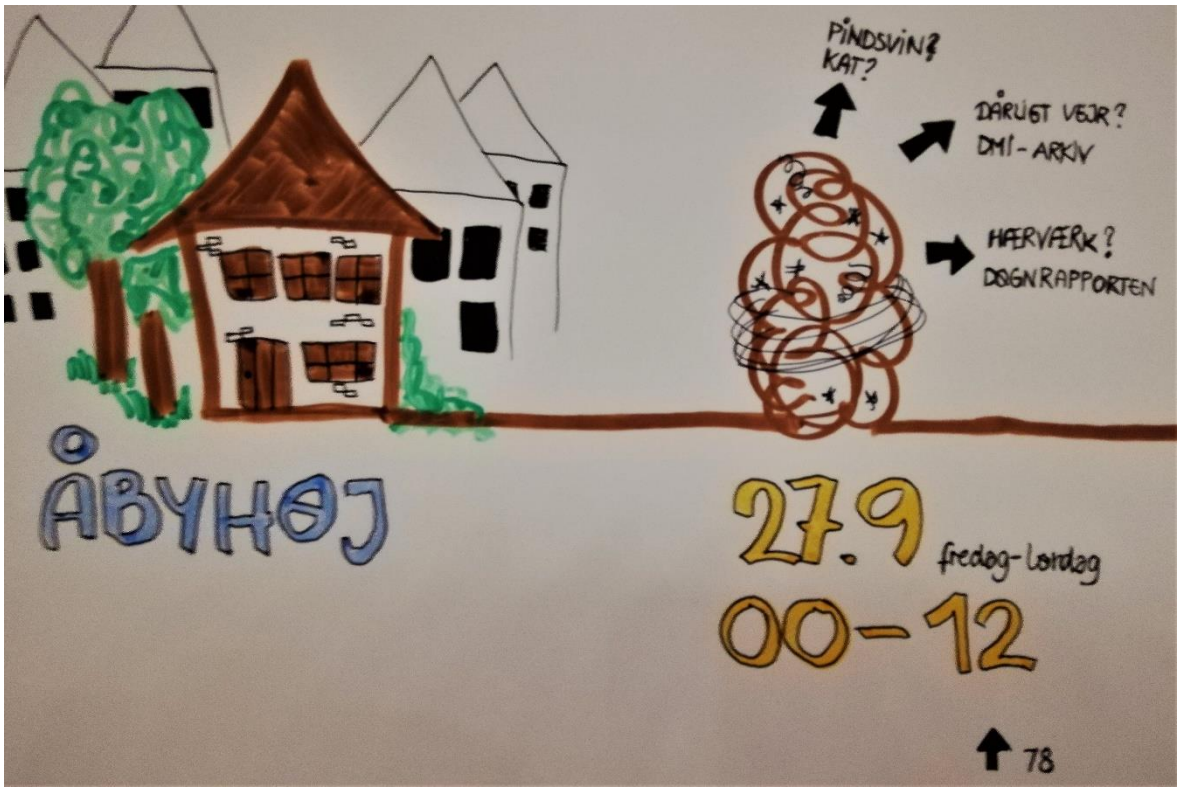

Figure 4: Example of what a datastory could look like.

3 | How did it go and what did we learn?

We had school children around 12-13 years test the prototype in workshops lasting two days. The children would come in one day to program and assemble the data transmitter, decide on their study design and test the transmitters. Then they would return about a week later to clean, visualize, and tell small data stories with their data.

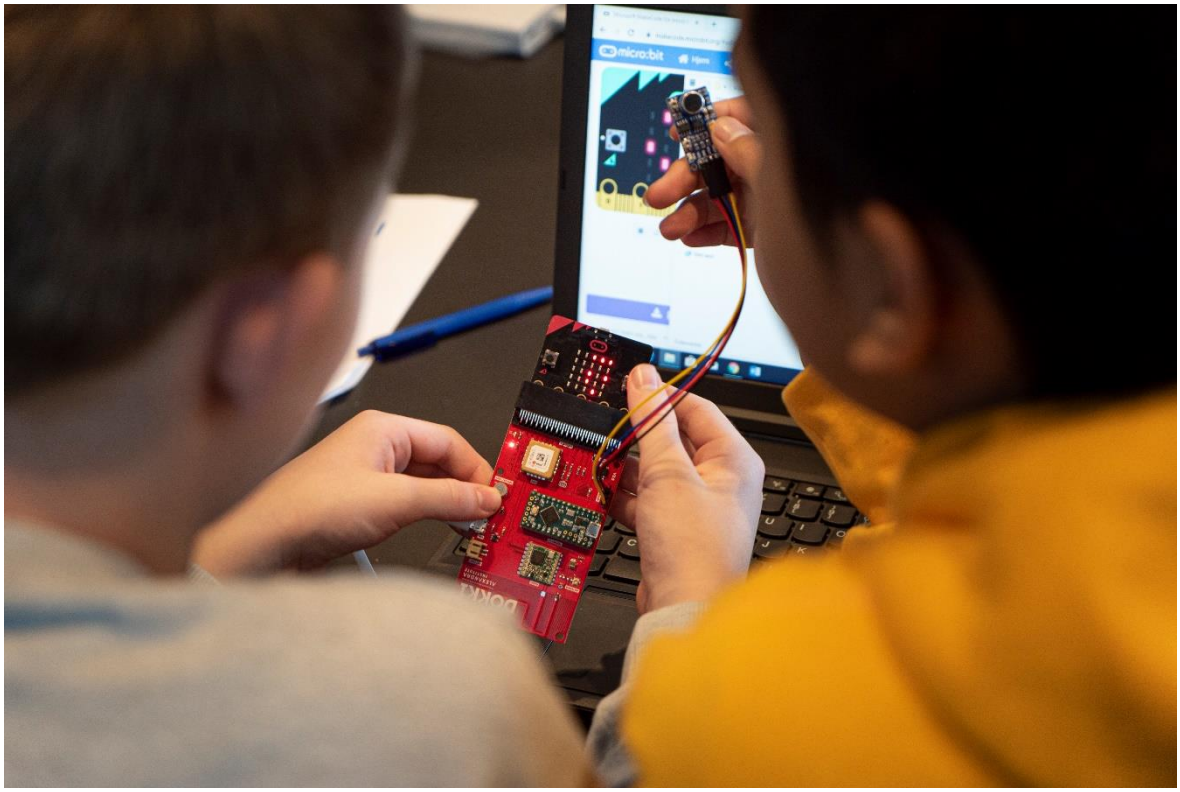

Figure 5: Students programming the Micro:bit using Makecode.

The children were mostly super excited the first day, when setting up their data transmitters, but returning to work with their collected data was more challenging - the data collected was often not what they expected.

For instance, one group had placed their data transmitter in a bedroom, not quite getting the part about a public space, but only the part about a positive atmosphere and were disappointed when the misunderstanding dawned on them. The data collected from a bedroom was as you would expect - very quiet, but while cleaning it the group discovered a long sequence of high sounds during the night and were intrigued about it; was it a monster, an intruder or?

They ended up with a data story saying it was snoring from the boy, who's room it was, and that placing the data transmitter on his desk, right next to his bed, would have had a huge impact on the sound levels transmitted. This was a bit of an eye opener to the class, as an example of thinking critically about raw data. As facilitators this also became a go-to example of the challenging balance, between a learner-centred data literacy tool and a platform for crowdsourcing data about our city.

Another reflection that supported this challenging balance was the need for supporting the participating school classes; their data literacy skills did not match our initial expectations. The need for support makes it very difficult to imagine engaging other kinds of citizens, who most likely would be participating in their spare time. For instance, the programme "Lyden af Danmark"

(https://lyden-af.dk/) has a much lower floor for participation, which makes it more appealing to a broader audience.

On the other hand, if citizens in the long run should crowdsource data that can be leveraged for the common good, they need to have some level of understanding for the data they are collecting and why, if not, this will be no different, then the data that is being collected from our smartphones and other devices today.

In one workshop we had a group of children, who had been collecting data from a park-like area close to their school, where they liked to hang out because it was peaceful. When they afterwards were cleaning the data, they noticed a recurring pattern of intense sound early in the morning. They could not explain the sound, although they had several plausible explanations, but it was not part of the atmosphere they knew, from using the area in the afternoon. This made them reflect on the importance of the location and timespan in their data collection and what they would do differently another time.

Ultimately the Danish outbreak of Covid19 in the spring of 2020 put an early stop to our workshops, but hopefully we can pick them up again in 2021 to get more data for our visualization and to introduce more children and youths to data collection and processing.

4 | Dreams for the future

There are some obvious challenges with the Sweet Spots prototype, both in regards to it working as a tool for data literacy training and as a platform for public data crowdsourcing, but the idea of joining data literacy and data crowdsourcing still makes sense, in our opinion, because it represents an alternative to collecting data from citizens and can provide data, that would not else be available.

We believe that mapping qualitative data like GPS coordinates or sound levels together with qualitative data like emotions or opinions on life in our local communities has the potential to motivate citizen participation. So we hope to continue working with the Sweet Spots prototype, both collecting more data and getting closer to realizing the concept of a large scale visualization at Dokk1, that can be used as a gateway for conversations with a broader section of the public, than the participants in our data collection workshops.

5 | REFERENCES

Beyond Data Literacy: Reinventing Community Engagement and Empowerment in the Age of Data (2015). Data-Pop Alliance White Paper Series at https://dspace.mit.edu/bitstream/handle/1721.1/123471/Beyond%20Data%20Literacy%202015.pdf?sequence=1&isAllowed=y [accessed 13/11/20]

D'Ignazio, Catherine (2017). Creative Data Literacy in Information Design Journal at https://www.jbe-platform.com/docserver/fulltext/idj.23.1.03dig.pdf?expires=1605015292&id=id&accname=guest&checksum=8D47191CA899973B5840E6411012D470 [accessed 10/11/20]

D'Ignazio, Catherine and Bhargava, Rahul (2015). Approaches to Building Big Data Literacy at https://www.semanticscholar.org/paper/Approaches-to-Building-Big-Data-Literacy-D%27Ignazio-Bhargava/ff9443c45b685f6ff60624f5356a531853450cb4?p2df [accessed 10/11/20]

Me and My Big Data Report 2020: Understanding citizens' data literacies: thinking, doing & participating with our data (2020) at https://www.liverpool.ac.uk/humanities-and-social-sciences/research/research-themes/centre-for-digital-humanities/projects/big-data/publications/ [accessed 19/11/20]

Panek, Jiri (2018). Emotionel Maps: Participatory Crowdsourcing of Citizens' Perceptions of Their Urban Environment in Cartographic Perspectives, Number 19 at https://www.researchgate.net/publication/329767067_Emotional_Maps_Participatory_Crowdsourcing_of_Citizens_Perceptions_of_Their_Urban_Environment [accessed 19/11/20]

Van Deursen, Alexander J. A. M. and Mossberger, Karen (2018). Any Thing for Anyone? A New Digital Divide in Internet-of-Things Skills in Policy and Internet, vol.10 issue 2 at https://onlinelibrary.wiley.com/doi/epdf/10.1002/poi3.171 [accessed 19/11/20]