

Kate Northstone
PHS, Bristol Medical School

ORCID: 0000-0002-0602-1983

ALSPAC's use of data notes to promote teamwork and discoverability in the time of pandemic

submitted for the

Open Research Prize 2022

SUMMARY (100 word abstract)

As one of the world's leading longitudinal population-based studies, the Avon Longitudinal Study of Parents and Children (ALSPAC) is at the forefront of promoting best practice. The study PI established a gateway on the [Wellcome Open Research](#) (WOR) publishing platform allowing open, peer-reviewed, free publication of "data notes"¹.

Throughout our intensive data collection during the pandemic we have rapidly described and published such data notes. This facilitates the rapid discovery of our data, ensures it is accurately described and kept up to date and enables our professional services staff to receive appropriate acknowledgement for their work in a new way.

What did you do?

ALSPAC is fully supportive of all open access procedures. We encourage that all papers published using our data are made open access. To help facilitate this, the study was at the forefront of establishing the WOR platform as a place to publish ALSPAC-based research. In addition, WOR allows the publication of data notes connecting specific datasets to targeted collection efforts, thereby describing the study's assets. We have our own [Gateway](#) where both academic papers and data notes can easily be discovered, with 64 published at the time of writing.

Since the start of the pandemic, the team have collected multiple sweeps of data collection across 3 generations of families including five questionnaires, two sets of antibody testing, linkage to Public Health England (now UKHSA) pillar testing results, a face-to-face case-control study involving deep phenotyping. As a result, Eight COVID-19 data notes have been published so far on WOR describing our COVID-19 datasets. As part of this, we have used the Open Science Framework (<https://osf.io/6jr7e/>) to publish the questionnaires that we used, the formal data dictionaries presenting the frequencies of variables and any code used to derive important variables ensuring transparency for all future data users.

Why did you do it?

As a study we have a duty to encourage, support and promote best practice and pride ourselves in leading the field in many areas. The study has been running for 30 years and we have clear procedures for documenting our data and making it available in a timely manner. However, the frequency and depth of data collection undertaken throughout the pandemic has never been seen before in the study and required a different way of reporting the data, knowing it would need to be turned around rapidly and made available in a reactive manner. WOR has been an excellent vehicle for publicising our rapid response to the pandemic in a reactive way; data notes are publicly available within a matter of days following submission and contain clear data availability statements as a matter of course.

¹ Data Notes are brief descriptions of scientific datasets that promote the potential reuse of research data and include details of why and how the data were created; they do not include any analyses or conclusions. <https://wellcomeopenresearch.org/for-authors/article-guidelines/data-notes>

How did you do it?

Whenever a COVID-19 related dataset was cleaned and documented according to ALSPAC standard procedures, we would prepare and submit a data note. As soon as the data note was published, we could refer to it in the relevant data dictionary and release the data. A huge team effort goes into collecting and producing datasets and by involving those staff who are critical to that effort their contributions are appropriately acknowledged. These staff are usually professional services and make a valid contribution to “Team Science” that often goes forgotten, with academic staff taking most of the credit. Data notes go beyond a simple description of the data, usually reserved for data dictionaries. The co-authors on these data notes have been instrumental in highlighting the strengths but, probably more importantly the limitations, of the datasets.

What barriers / challenges did you have to overcome?

The hard work in setting up a gateway had already been done and the concept of data notes had already been conceived. Any challenges we faced were therefore minimal and primarily logistic. However, publishing data notes requires a slightly different way of thinking about the context of the data. Previously, we provided detailed meta data when we released data with some details of the source of funding, the provenance of the tools used and additional meta-data where appropriate. Data notes force a more detailed description of what was done, why, when and how. Alongside all the challenges and changes that we inherently had to deal with in the early stages of the pandemic, developing a new way of describing our resource and coordinating data release with publication was a learning curve, particularly trying to do this at speed given the competing calls on our time. This is an adaptation of the way in which ALSPAC works and questions remain over how we will publish future data collections. Our standard data collections (i.e. self-completion questionnaires or face to face clinic visits) tend to span multiple sources of funding and researcher questions rather than focussing on one single global event and the resource required to generate them is not funded. Data notes are still subject to peer review and it can be a difficult and time consuming process to obtain appropriately qualified reviewers (this responsibility of identification is shared with the publisher), although data notes are immediately available online, they do not appear in PubMed searchers for example until fully reviewed.

What does it mean for you and your research?

Of key importance to me personally, is that my team of Professional Services staff get formal acknowledgement of their contribution to the data collection effort. This is so often overlooked; they are rarely acknowledged by name through the traditional academic publishing route. At the time of writing, the [first data note](#) has been viewed over 2600 times and downloaded 238 times which also supports my academic career and personal impact.

How might your findings / approach help other researchers?

By publishing data notes we are providing key references and aligned datasets for researchers for years to come. All the information about the data – from justifying its collection through to how derived variables were created - is in

one place. The COVID-19 data in particular, is likely to be used for years to come and having the data formally documented and fully discoverable is of huge benefit. WOR facilitates multiple versions of publications; this means that we will need to ensure we keep our data notes up to date should we make fundamental changes to the datasets ensuring transparency and following best practise.

The idea of peer reviewing data production (rather than just it's use) is interesting and perhaps warrants further thought at a higher level. We would encourage any type of study, large or small, cross-sectional or longitudinal, experimental or observational to go down the data note route. At the moment, WOR is the perfect vehicle for Wellcome funded research. The benefit of other platforms or a widening of WOR cannot be under-estimated and perhaps could be considered by the UKRN if it is not already on the agenda.

Additional Information

It is important to note that this case study has been submitted on behalf of the wider ALSPAC team.

The following screenshot shows the ALSPAC gateway: <https://wellcomeopenresearch.org/gateways/alspac> and illustrates two of the COVID-19 data notes.



Avon Longitudinal Study of Parents and Children (ALSPAC)

TRACK

[BROWSE](#) ABOUT THIS GATEWAY

SUBMIT

[Home](#) » [Gateways](#) » [Avon Longitudinal Study of Parents and Children \(ALSPAC\)](#)

FILTERS ▾

1-20 of 30 ARTICLES

Search this Gateway

Gateway Areas



ALSPAC Data Notes

ALSPAC Data Notes provide all researchers who have accessed ALSPAC data a route to publish their datasets (for free) and make... [Read more](#) >



ALSPAC Research

Our News

Postnatal depression has greater impact on children's development when it is persistent and severe

Postnatal depression which persists beyond six months after birth and is severe, increases the risk of children exhibiting

DATA NOTE metrics

REVISED

The Avon Longitudinal Study of Parents and Children - A resource for COVID-19 research: Antibody testing results, April – June 2021 [version 2; peer review: 2 approved]

Daniel Major-Smith, Sarah Matthews, Thomas Breeze, Michael Crawford, Hannah Woodward, Nicholas Wells, Ruth Mitchell, Lynn Molloy, Kate Northstone, Nicholas John Timpson

PEER REVIEWERS *Alicia Matijasevich; Matthias Pierce*

FUNDERS **Wellcome Trust** | National Institute for Health Research | University of Bristol | Medical Research Council | Health and Safety Executive | UK Research and Innovation

LATEST VERSION PUBLISHED 24 Jan 2022

DATA NOTE metrics

REVISED

The Avon Longitudinal Study of Parents and Children - A resource for COVID-19 research: Home-based antibody testing results, October 2020. An emphasis on self-screening at a