

*Some considerations on the nature  
of data availability statements and data citations*

*by Evgeny Bobrov*

*Feb. 17<sup>th</sup> 2022*



**Q: Where in an article do I find information about data availability?**

**A: In footnotes, methods, data availability statements, or other parts of a paper.**

**Q: What is a data availability statement?**

**A: It is a distinct section of an article indicating, expectedly, the availability of data.**

**Q: Where should I put information about data availability?**

**A: In ~~footnote, methods,~~  
data availability statement**

**Data Availability Statement:** The final dataset and accompanying code are available on the Open Science Framework, DOI [10.17605/OSF.IO/UN8JX](https://doi.org/10.17605/OSF.IO/UN8JX).

**Q: Is it always called the data availability statement?**

**A: No, sometimes it is called „data access“, „availability of data and materials“, or similar.**

## Data Access

NGS data were deposited in the NCBI Gene Expression Omnibus (GEO) under accession number GSE99398. The raw sequence tags were deposited in the NCBI Short Read Archive (SRA) under accession number SRP108249.

## Availability of data and materials

**Q: Are the data underlying  
the article available?**

**A: Data are available  
upon reasonable request.**



## **Availability of data and materials**

The datasets used in the current study are available from the corresponding author on reasonable request.

**Q: But are the data really available?**

**A: Research shows that data available upon request are difficult and sometimes impossible to obtain.**

---

**Safety and immunogenicity of an rAd26 and rAd5 vector-based heterologous prime-boost COVID-19 vaccine in two formulations: two open, non-randomised phase 1/2 studies from Russia**

**Data sharing**

Individual participant data will be made available on request, directed to the corresponding author (DYL). After approval of a proposal, data can be shared through a secure online platform.

CORRESPONDENCE | [VOLUME 397, ISSUE 10288, P1881-1883, MAY 22, 2021](#)

## Data discrepancies and substandard reporting of interim data of Sputnik V phase 3 trial

Enrico M Bucci [✉](#) • Johannes Berkhof • André Gillibert • Gowri Gopalakrishna • Raffaele A Calogero • Lex M Bouter • et al. [Show all authors](#)

Published: May 12, 2021 • DOI: [https://doi.org/10.1016/S0140-6736\(21\)00899-0](https://doi.org/10.1016/S0140-6736(21)00899-0)

“Several experts found problematic data in the published phase 1/2 results. We have made multiple independent requests for access to the raw dataset, but these were never answered.”

**Q: How should a data availability statement look like?**

**A: Indicate where data has been stored openly. Or, if restrictions apply, indicate how it can be accessed and used.**

**Data Availability Statement:** The final dataset and accompanying code are available on the Open Science Framework, DOI [10.17605/OSF.IO/UN8JX](https://doi.org/10.17605/OSF.IO/UN8JX).

### **Data availability**

Due to potential risk of de-identification of pseudonymized scRNA-seq data, the raw data will be available under controlled access in the European Genome-phenome Archive repository ([EGAS00001004772](https://ega-archive.org/studies/EGAS00001004772)). Count and metadata tables (patient ID, sex, age, cell type and quality control metrics per cell) can be found at FigShare: <https://doi.org/10.6084/m9.figshare.13200278>. In addition,

**Q: How should I cite data I reused?**

**A: Reference data in ~~footnote,~~  
~~methods,~~ reference list.**

...highly site specific, potentially limiting their wider value. However, applying the approach as conducted in this paper to data such as that presented by Barnett et al (2013) to give relative values for different organisms should provide a more generic set of 'reference data'. In taking the REML approach forward it will be beneficial to target...

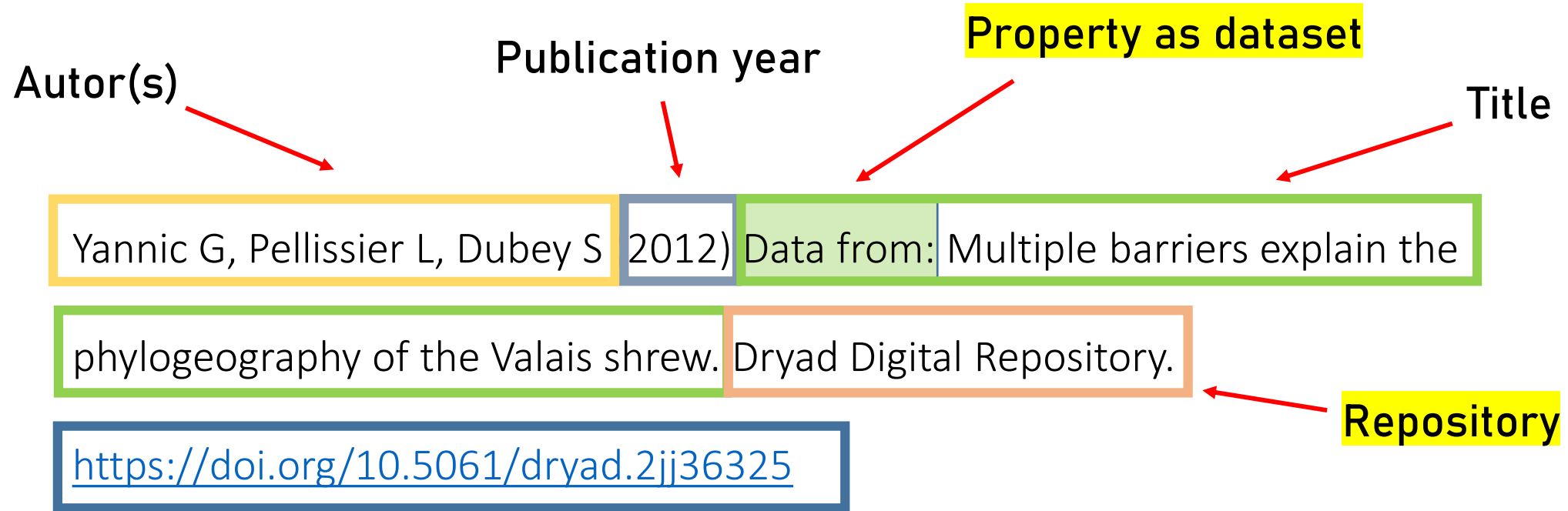
## References

Barnett et al., 2013 C. Barnett, N.A. Beresford, L.A. Walker, M. Baxter, C. Wells, D. Copplestone  
Element and radionuclide concentrations in representative species of the ICRP's reference animals and plants and associated soils from a forest in North-west England.  
NERC - Environmental Information Data Centre (2013) <http://doi.org/10.5285/e40b53d4-6699-4557-bd55-10d196ece9ea>



**Q: How does a data citation look like?**

**A: Like an article citation...  
...with a few things to consider.**



<https://doi.org/10.5061/dryad.2jj36325>

**Identifier (typically Digital Object Identifier = DOI)**

- + optional:**
- Type of resource
  - Version

**Q: Should I cite my own data the same way?**

**A: Yes. Cite it, as you would other publications of yours.**

**Q: What is the benefit of citing datasets?**

**A: It increases findability and thus reuse. It also recognizes the data contributors by allowing citation analysis.**

**Q: What is your connection to the topic?**

**A: I am project leader Open Data and Research Data Management at the QUEST Center of the Berlin Institute of Health.**

**I consult Charité researchers on research data management and offer related courses.**

# Contact

DR. EVGENY BOBROV

Berlin Institute of Health at Charité (BIH)

QUEST – Quality | Ethics | Open Science | Translation

BIH Center for Responsible Research

*Project Leader Open Data and Research Data Management*

Tel. +49 (0)30 450 543 069

[evgeny.bobrov@bih-charite.de](mailto:evgeny.bobrov@bih-charite.de)

[www.bihealth.org/de/quest](http://www.bihealth.org/de/quest)

The logo for BIH QUEST, Center for Responsible Research. It features the text "BIH QUEST" in a bold, teal, sans-serif font. To the left of "BIH" is a vertical teal bar. Below "BIH QUEST" is the text "Center for Responsible Research" in a smaller, teal, sans-serif font.

**BIH QUEST**  
Center for Responsible Research