

A Decade of Progress in Open Data Practices: Insights from Biosciences at the University of Edinburgh

Haya Deeb, George Strevens, Diego Lucini de Ugarte, Suzanna Creasey, Hwee Yun Wong, Trisha Usman, Megan A. M. Kutzer, Tomasz Zieliński, Andrew J. Millar. (Centre for Engineering Biology and School of Biological Sciences, University of Edinburgh, EH9 3JF, UK.)

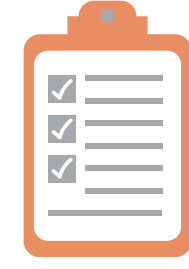
Background

- Open Data has emerged as a crucial element in scientific research, emphasizing the intrinsic value of underlying data beyond just publications.
- The FAIR guidelines were introduced to maintain the quality of shared research data (1). Additionally, the UOE incorporated the FAIR into its updated Research Data Management Policy in 2021 (2).

Objective: to analyse the progress and current standing of the University of Edinburgh in implementing Open Data Practices within biosciences research, highlighting achievements and identifying challenges.

Methods

Our evaluation covered biosciences research papers published between **2014 and 2023** at the University of Edinburgh. A total of **544** papers were selected from several institutes and assessed manually for data sharing practices, without prior knowledge of whether any data was shared. This assessment focused on **four criteria** reflecting the Openness and FAIRness of research data:



Completeness
How much?



Reusability
How?



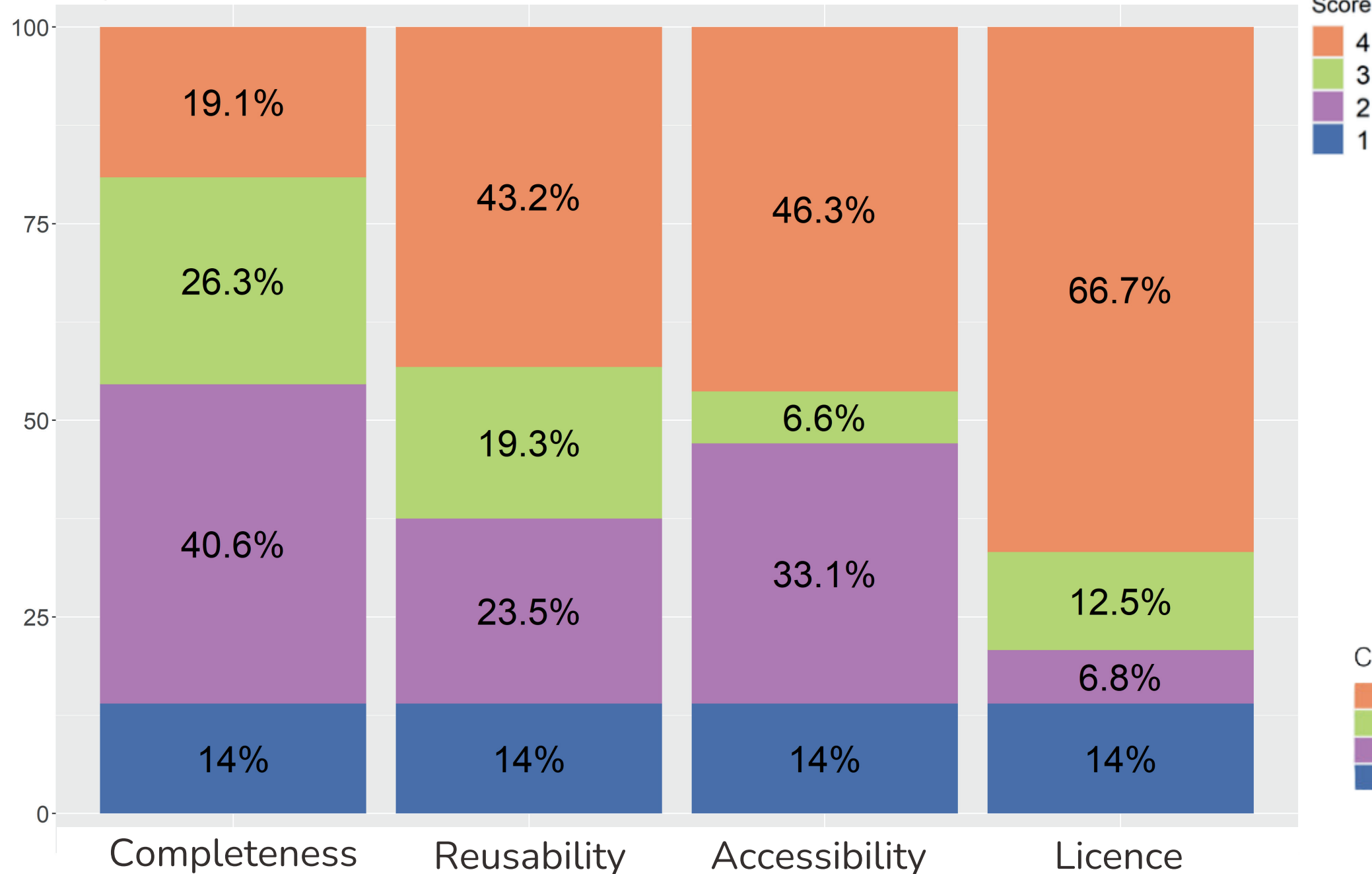
Accessibility
Where?



Licence
Who owns?

Results

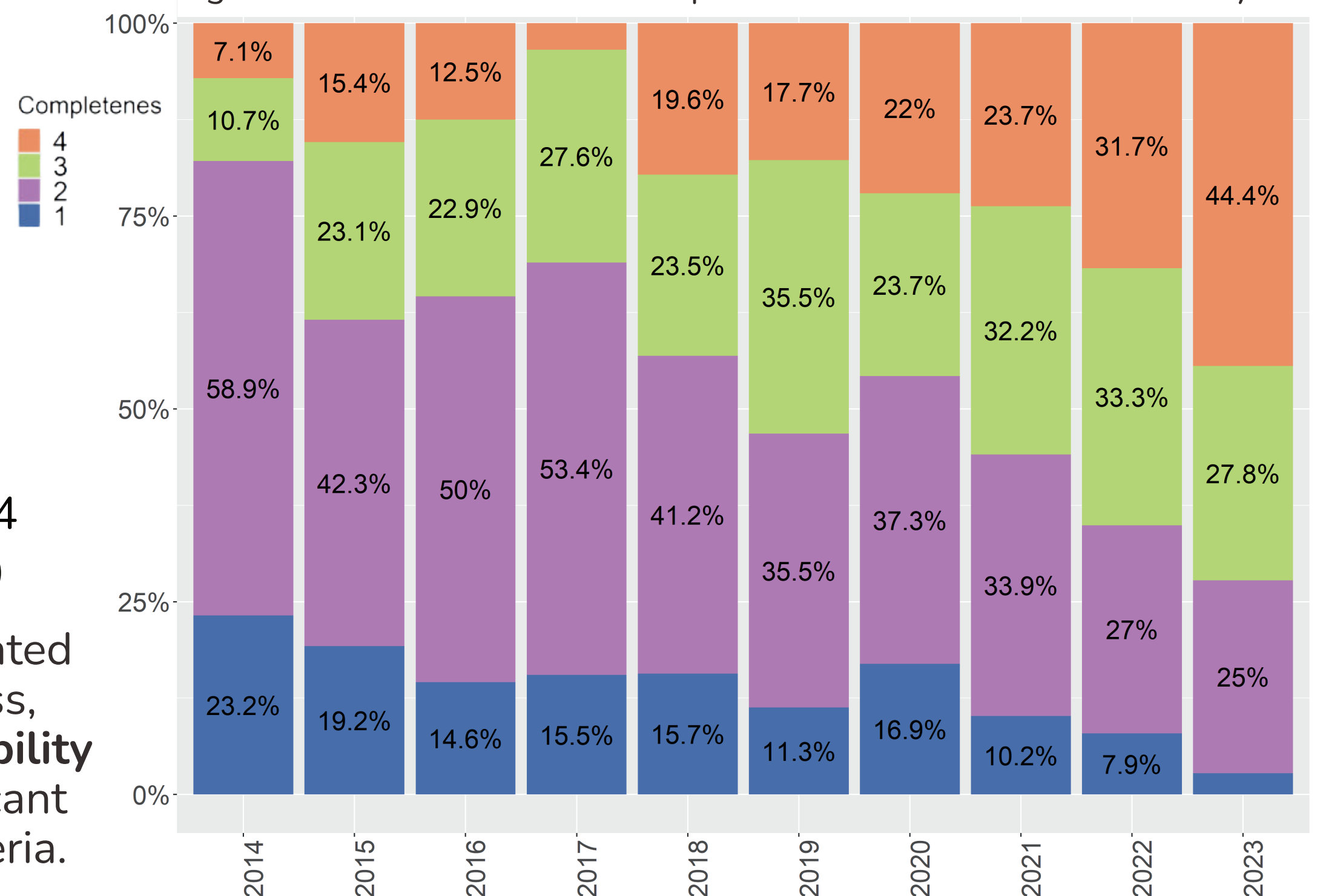
Figure1: Scores of the four criteria in all the research publications



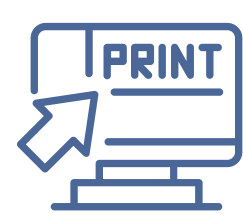
Out of 544 papers, only **76** did not share any research data (Completeness=1), whereas **104** shared all the necessary data for reproducibility (Completeness=4).

There was a **statistically significant improvement** in all the four criteria over the last 10 years. This suggests that researchers are not only sharing more data but are doing so in a more effective way.

Figure2: The distribution of Completeness scores over the last ten years



Genomic data was the most shared, while **image** data was the least shared, even compared to **human** data, indicating variations in data sharing by type.



119/544
(21.88%)



270/544
(49.63%)

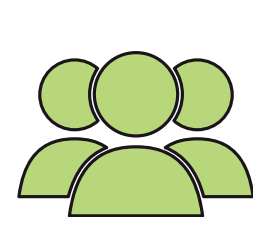
The **Preprint** was correlated only with Completeness, whereas the **Data Availability Statement** had a significant influence on all four criteria.



237/319
(74.29%)



67/346
(19.36%)



47/156
(30.13%)

Discussion

Our manual scoring method assessed the openness and FAIRness of published data, starting with the presence of data sharing. This approach, previously used (3), offers deeper insights than studies limited to already shared datasets (4).

The presence of data availability statements or preprint sharing correlates with higher scores across all criteria, especially completeness, indicating a more thorough data sharing approach with established guidelines.

Conclusion

The biosciences' move towards open data reveals progress and challenges, underscoring the effectiveness of specific guidelines in enhancing data sharing, as well as the challenges posed by certain data types like images

References:

- Musen MA, O'Connor MJ, Schultes E, Martínez-Romero M, Hardi J, Graybeal J. Modeling community standards for metadata as templates makes data FAIR. *Sci Data*. 2022 Nov 12;9(1):696.
- The University of Edinburgh. Policy for Management of Research Data [Internet]. 2011 Apr. Available from: https://era.ed.ac.uk/bitstream/handle/1842/38258/RDM-Policy_CMG-20Apr11.pdf?sequence=1&isAllowed=y
- Hamilton DG, Page MJ, Finch S, Everitt S, Fidler F. How often do cancer researchers make their data and code available and what factors are associated with sharing? *BMC Medicine*. 2022 Nov 9;20(1):438.
- Roche DG, Kruuk LEB, Lanfear R, Binning SA. Public Data Archiving in Ecology and Evolution: How Well Are We Doing? *PLOS Biology*. 2015 Nov 10;13(11):e1002295.

Funding: MRC grant MR/X009726/1



Scan me to read the preprint

