

Biological Databases (Part 2)

Kam D. Dahlquist, Ph.D.
Department of Biology
Loyola Marymount University

BIOL 367
February 1, 2024

Biological Databases Are Evaluated Based on Quality and Utility

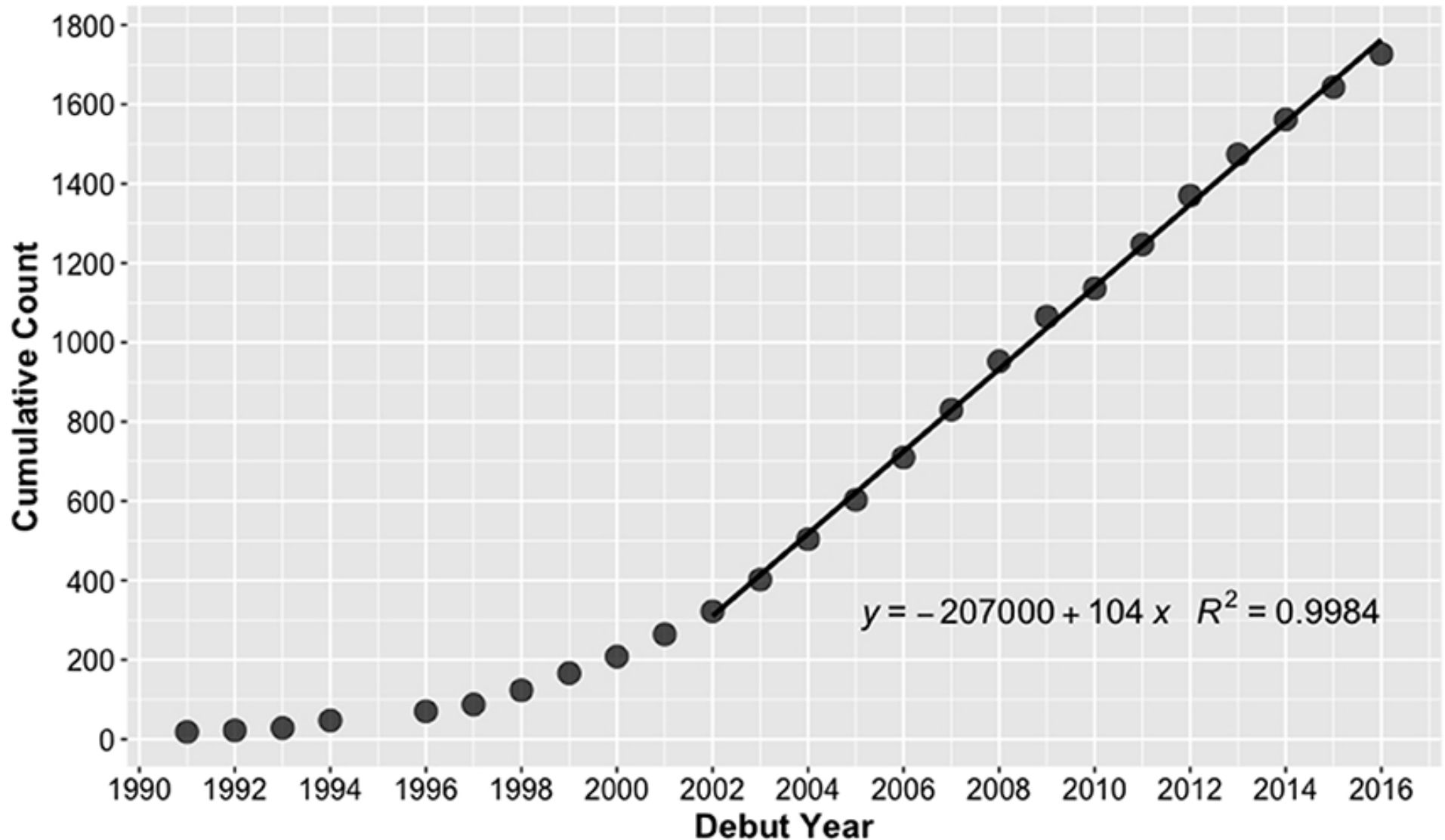
- **Scientific quality of the database**
 - **What biological information does it contain?**
 - **Does the content appear to completely cover its content domain?**
 - **What species are covered in the database?**
 - **Is the database content useful?**
 - **What biological questions can it be used to answer?**
 - **Is the database content timely?**
 - **Is there a need in the scientific community for such a database at this time?**
 - **Is the content covered by other databases already?**
 - **How often is the database updated?**
 - **When was the last update?**

Biological Databases Are Evaluated Based on Quality and Utility

- **General utility of the database to the scientific community**
 - **Are there links to other databases?**
 - **Which ones?**
 - **Is it convenient to browse the data?**
 - **Is it convenient to download the data?**
 - **In what file formats are the data provided? (standard or non-standard)**
 - **User-friendliness—can a naive user quickly navigate the website and gather useful information?**
 - **Is the web site well-organized?**
 - **Does it have a help section or tutorial?**
 - **Are the search options sensible?**
 - **Do sample query results make sense?**
 - **Would you direct a colleague unfamiliar with the field to use it?**
 - **Access—is there a license agreement or any restrictions on access?**

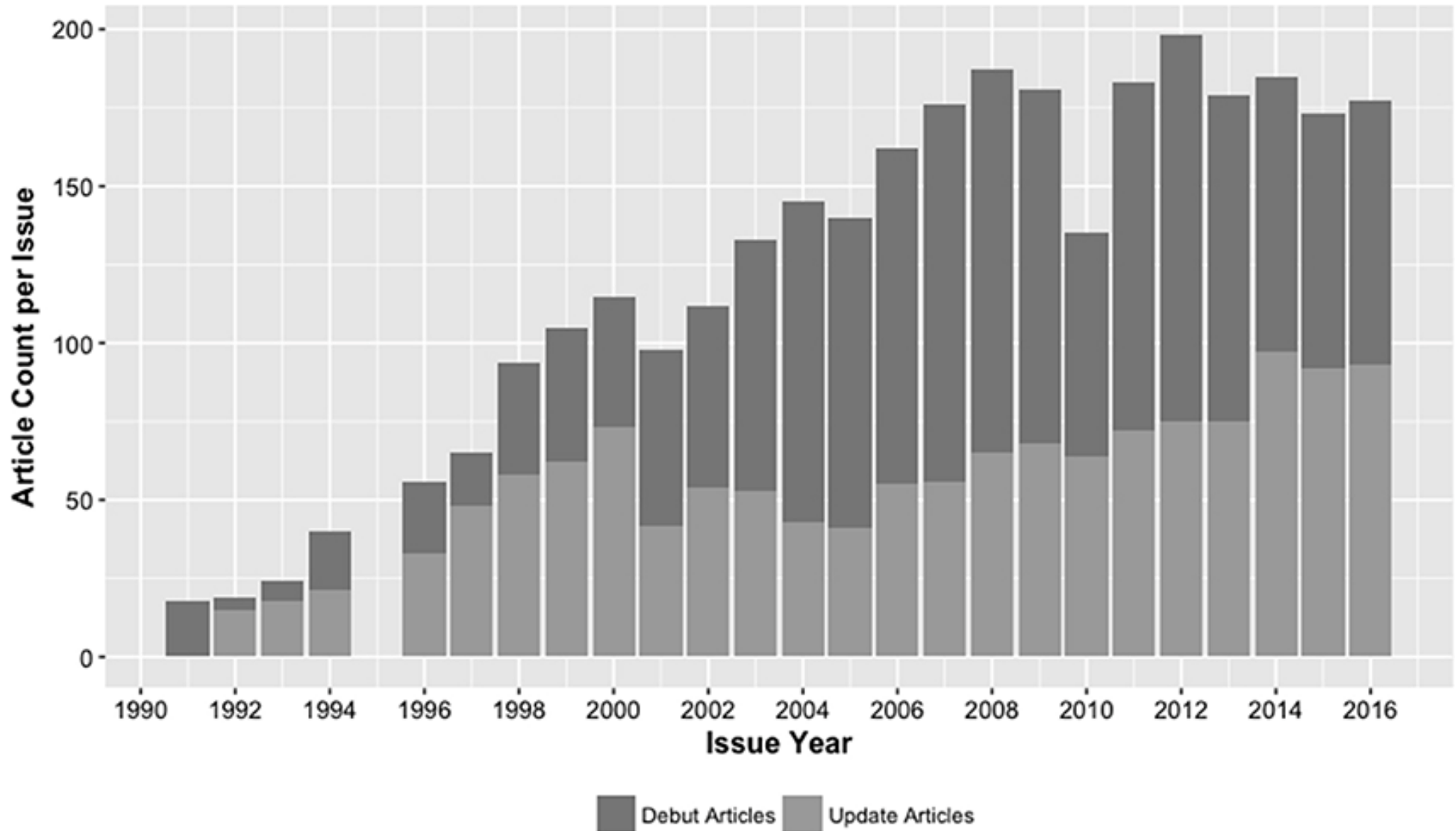
The Number of Databases Themselves Is Growing

Unique Databases Debuted between 1991-2016



Imker, H. (2018). 25 Years of Molecular Biology Databases: A Study of Proliferation, Impact, and Maintenance. *Frontiers in Research Metrics and Analytics*, 3, 18. <https://doi.org/10.3389/frma.2018.00018>, Figure 2

Growth of Articles in NAR Database Issues 1991-2016



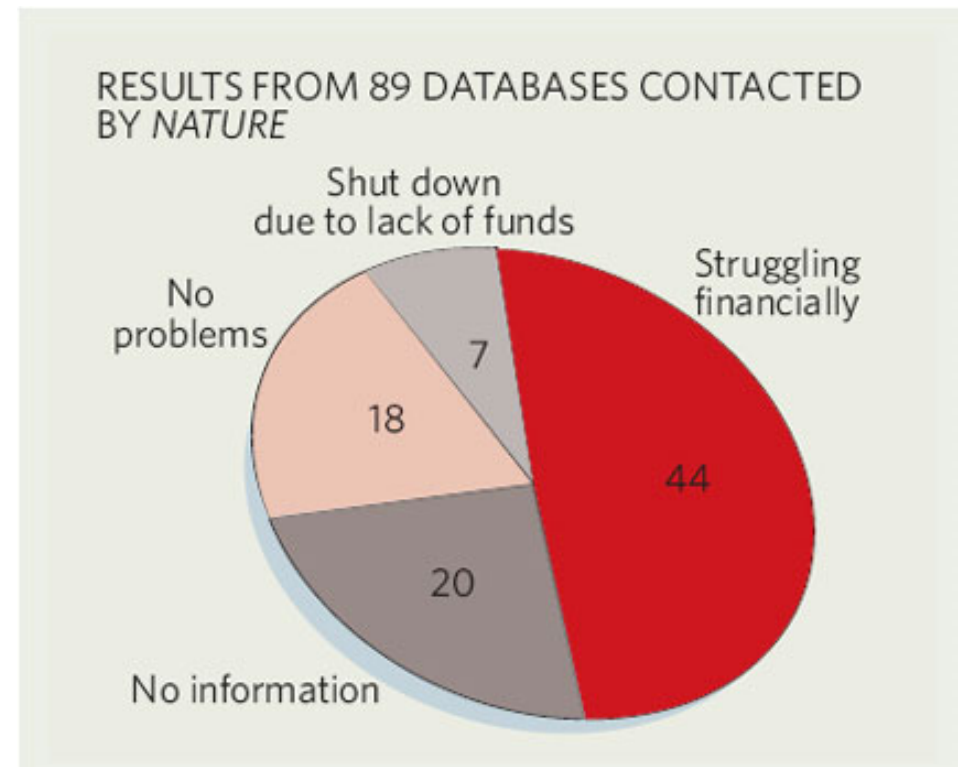
Imker, H. (2018). 25 Years of Molecular Biology Databases: A Study of Proliferation, Impact, and Maintenance. *Frontiers in Research Metrics and Analytics*, 3, 18. <https://doi.org/10.3389/frma.2018.00018>, Figure 1

Flash forward to *NAR* Database Issue 2024

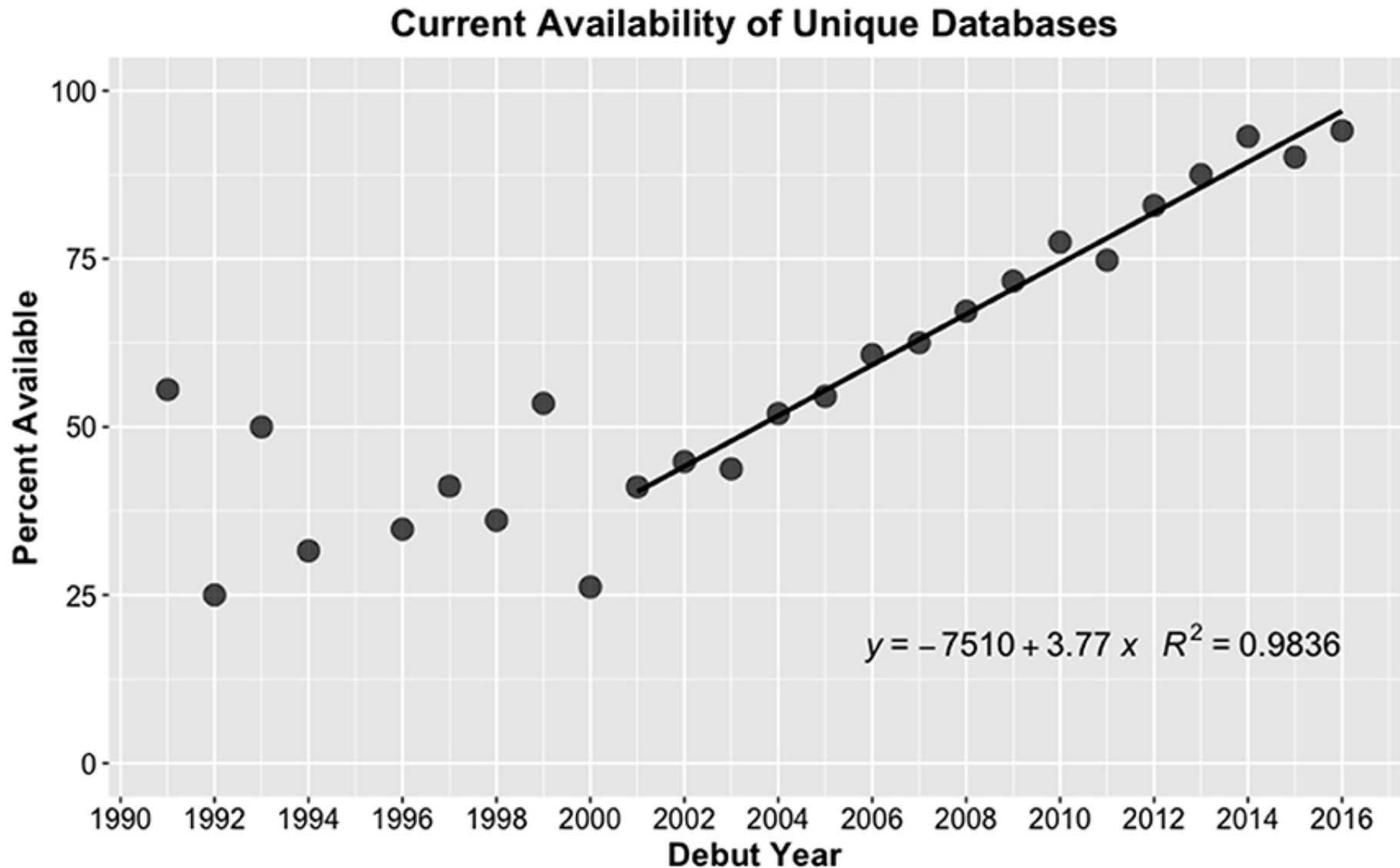
- **90 papers on new databases (record tied with 2023!)**
- **83 updates from previously published databases**
- **NAR Molecular Biology Database Collection**
 - 1060 entries reviewed in 2023
 - 97 new resources added
 - 388 eliminated (URLs discontinued)
 - Total 1959 databases in the collection

Biological Databases Are a Moving Target

- **Regular updates can be both a blessing and a curse**
 - blessing: always having access to up-to-date data
 - curse: always having to keep up with up-to-date data
- **Databases can change locations or formats**
 - breaking scripts used for automated data analysis pipelines
 - affecting interoperability with other databases
- **They can disappear completely due to lack of funding**
- **Garbage in = garbage out**
 - curation issues
 - error propagation



Current Availability of Databases that Debuted between 1991 and 2001 Averages to 39.5% (105 Databases)



Imker, H. (2018). 25 Years of Molecular Biology Databases: A Study of Proliferation, Impact, and Maintenance. *Frontiers in Research Metrics and Analytics*, 3, 18. <https://doi.org/10.3389/frma.2018.00018>, Figure 4

The Requirement for Data Sharing is Not Universal

- **From the beginning, when new sequences or structures were published, authors were required to submit the data to a public database**
- **However, there are no other uniform policies governing other types of data (although special interest groups are working on this)**
- **Open Access vs. traditional publishing**