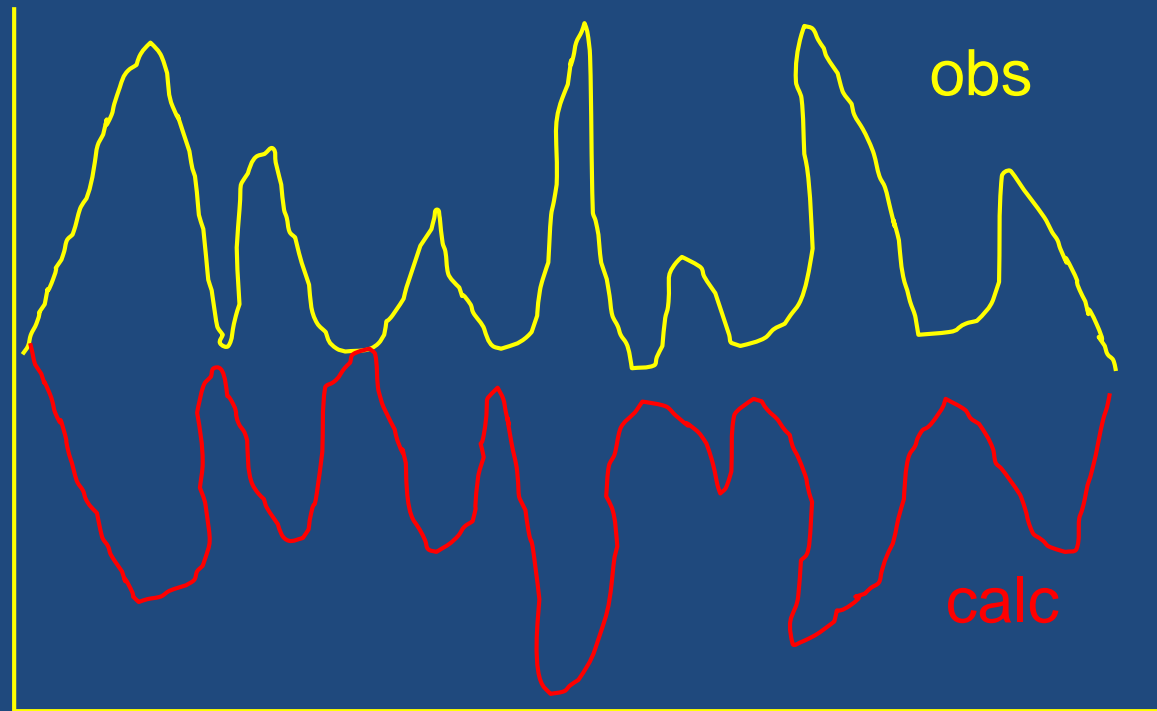


# Are your spectroscopic data being used?

*Iouli Gordon,*  
*Laurence Rothman,*  
*Jonas Wilzewski*

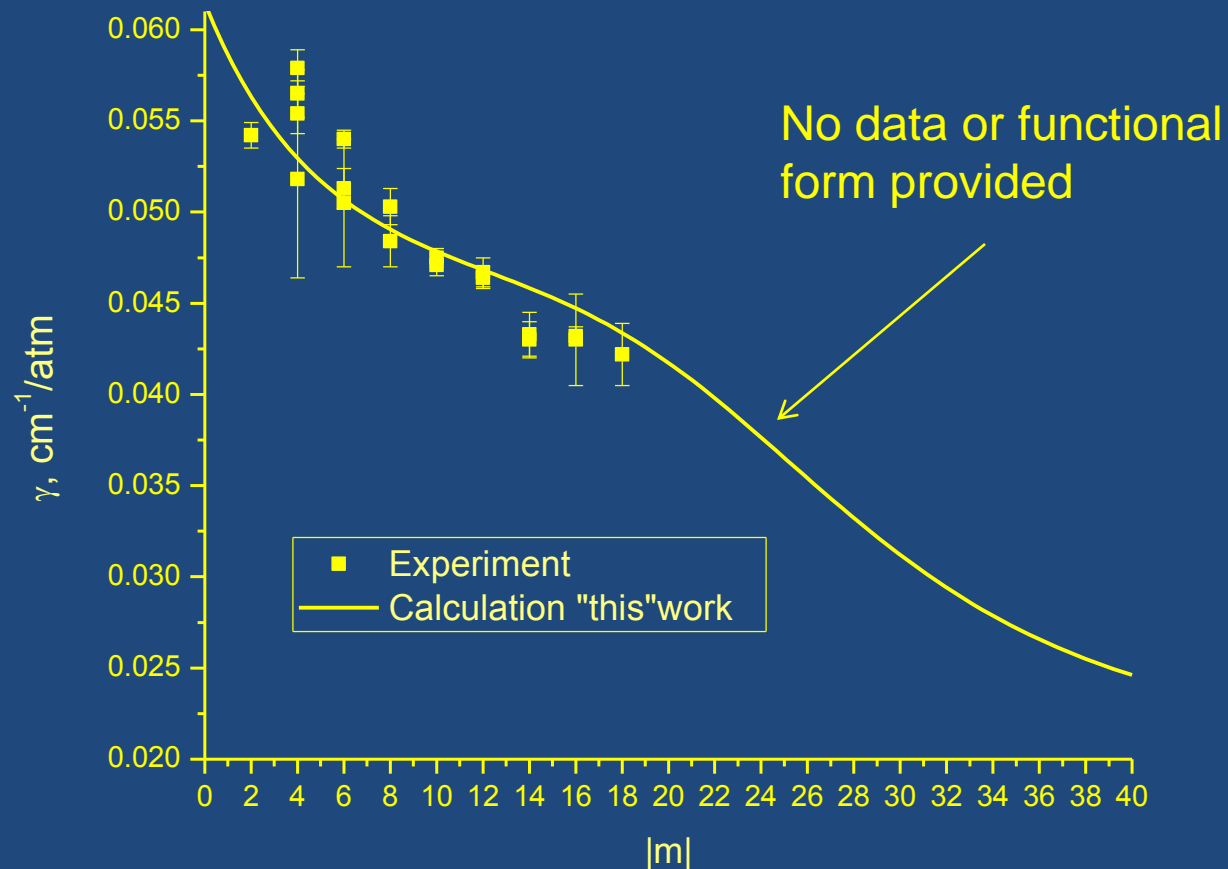
# Different levels of data presentation in scientific publication

Level 1. No data presented in the publication in any form



# Different levels of data presentation in scientific publication

Level 1. No data presented in the publication in any form



# Why provide data?



the WHITE HOUSE PRESIDENT BARACK OBAMA

★★★★★

★★★★★

Get Email Updates | Contact Us

 BLOG PHOTOS & VIDEO BRIEFING ROOM ISSUES the ADMINISTRATION the WHITE HOUSE our GOVERNMENT

Home • The Administration • Office of Science and Technology Policy

Search WhiteHouse.gov Search

 Office of Science and Technology Policy

About OSTP | Pressroom | OSTP Blog | Divisions | Initiatives | R&D Budgets | Resource Library | NSTC | PCAST | Contact Us

## Expanding Public Access to the Results of Federally Funded Research Subscribe

Posted by Michael Stebbins on February 22, 2013 at 12:04 PM EDT

 E-Mail  Tweet  Share 

The Obama Administration is committed to the proposition that citizens deserve easy access to the results of scientific research their tax dollars have paid for. That's why, in a policy memorandum released today, OSTP Director John Holdren has directed Federal agencies with more than \$100M in R&D expenditures to develop plans to make the published results of federally funded research freely available to the public within one year of publication and requiring researchers to better account for and manage the digital data resulting from federally funded scientific research. OSTP has been looking into this issue for some time, soliciting broad public input on

GIVE FEEDBACK ABOUT THIS PAGE

YOUR FEDERAL TAXPAYER RECEIPT



Launch the Receipt

# Why provide data?

1. The obvious. No one can use the data if it is not there.
2. Federally funded researchers are actually required to do so.
3. It enhances authors' professional metrics including *h*-index.

Figure 1 displays a series of nine density plots, one for each year from 2001 to 2009, showing the distribution of the number of citations for papers. The x-axis represents the 'number of citations' on a logarithmic scale, ranging from 0 to 3000. The y-axis represents the 'density' of citations, ranging from 0 to 0.5. Each plot compares two groups: 'data NOT available' (represented by an orange shaded area) and 'data available' (represented by a blue shaded area). The distributions are generally right-skewed, with most papers receiving a small number of citations (between 1 and 100). The 'data available' distributions often show a higher density at higher citation counts (above 100) compared to the 'data NOT available' distributions, which tend to be more concentrated at lower citation counts. The overall trend across the years shows a shift towards higher citation counts for the 'data available' group.

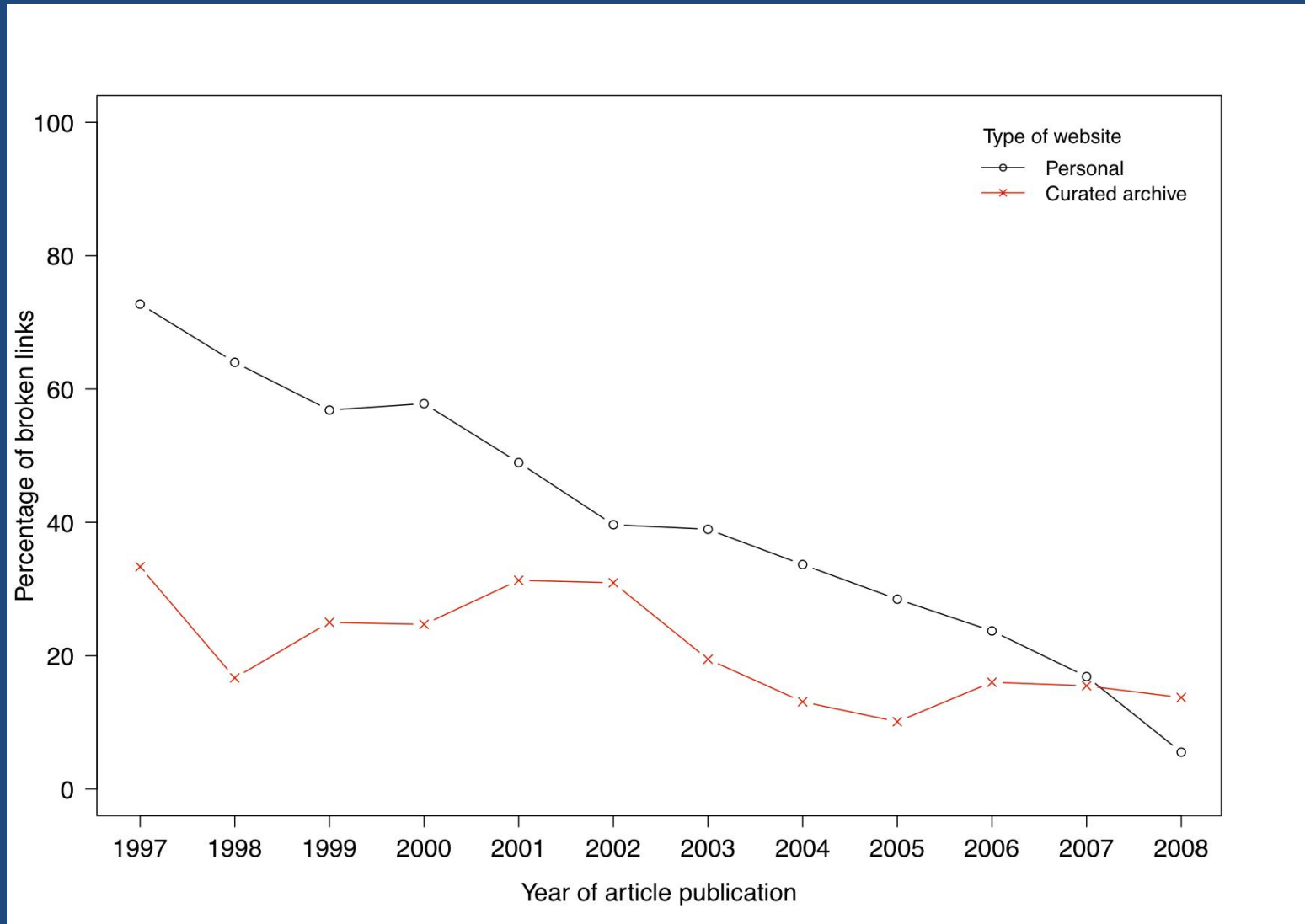
From Pepe et al  
(unpublished),  
data collected by  
ADS and the  
library at SAO

# **Different levels of data presentation in scientific publication**

Level 1. No data presented in the publication in any form

Level 2. Data are reported to be available on personal websites or through private communication

# Percentage of broken links given in the papers from four major Astrophysical journals



From Pepe et al (unpublished). Data collected by ADS and library at SAO

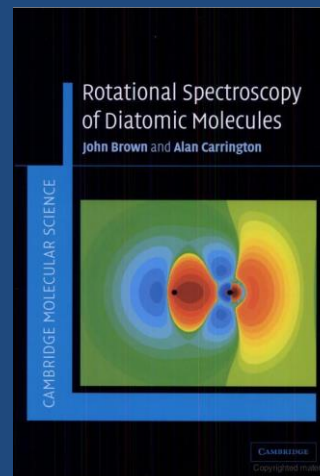
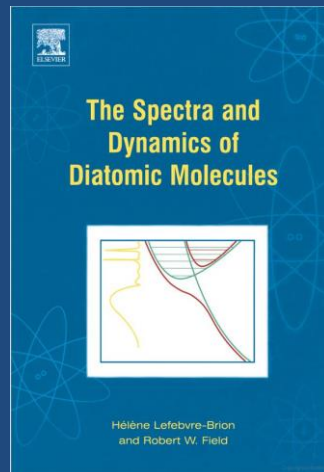
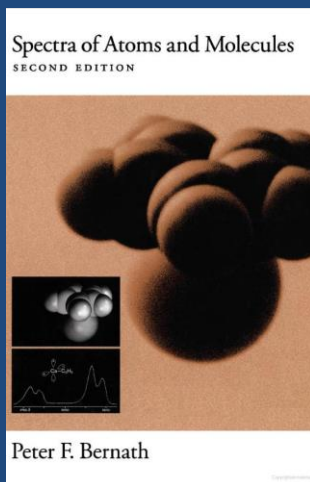
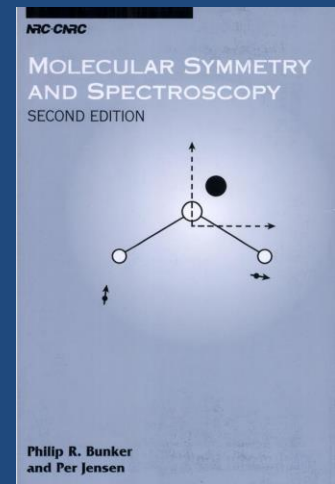
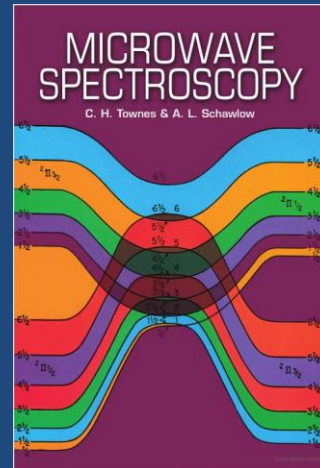
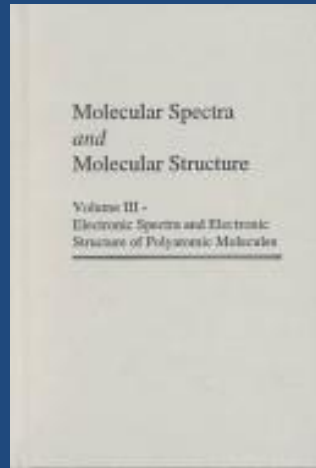
# Different levels of data presentation in scientific publications

Level 1. No data presented in the publication in any form

Level 2. Data is reported to be available on personal websites or through private communication

Level 3. Data analyzed, fitted and corresponding spectroscopic constants reported

# Spectroscopists path on working with complicated spectra



# Sophisticated interaction matrix was derived Spectrum fitted Constants reported

|     |    |   |    |    |   |   |
|-----|----|---|----|----|---|---|
| 2♠♦ |    |   |    |    |   |   |
|     | ♠Υ |   |    |    |   |   |
|     |    | ♣ | ₪9 |    |   |   |
|     |    | ⚡ | 鯨  | ☾  |   |   |
|     |    |   | 0  | ◌̇ |   |   |
|     |    |   |    |    | ‰ |   |
|     |    |   |    |    |   | ☄ |

Congratulations!!!

No one else can use it if no data or even a program given

# Problems with this approach

1. The theory is too complicated for the user.
2. If user actually manages to use reported constants they often do not get same results.
3. Impossible to do the global fit if complimentary data becomes available.
4. Substantially increases room for error. There are cases when even if one does use same constants and program experimentally observed parameters are not reproduced.

# Suggested approach

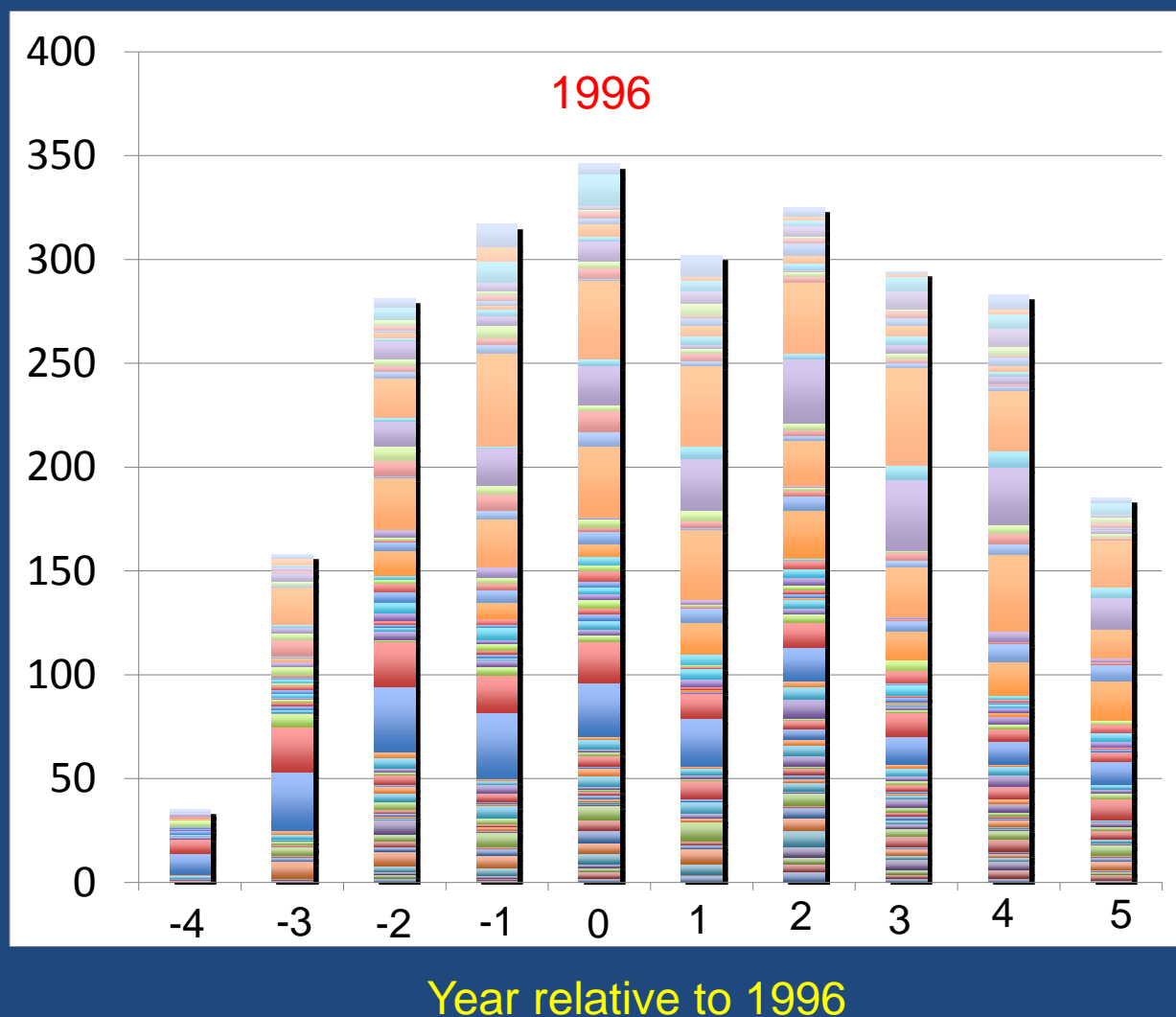
1. Report experimental data.
2. Report constants.
3. If your program is too custom, make it available. There are now ways to make your codes citable. Zenodo initiative from CERN.
4. Alternatively or additionally, do calculate the linelist.
5. Data and linelists should be placed in Journal archives!

# Does depositing to spectroscopic databases steal citations from your articles?

1. Intuitively, YES, as some users do not look for or cite the original publications.
2. In reality, NO! There is a competitive process of increased visibility especially with a new online tools. See, for instance, [hitran.iao.ru](http://hitran.iao.ru) and [www.hitran.org](http://www.hitran.org) .

# Examples of the citations to the articles that were used in compilation of HITRAN1996 but published 3 and 4 years prior

Total citations per year (with contributions from individual articles color coded)



# Acknowledgements

Many users of reference  
spectroscopic data

SAO library and ADS system

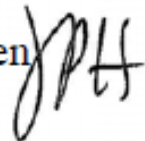
Thank you for your attention

EXECUTIVE OFFICE OF THE PRESIDENT  
OFFICE OF SCIENCE AND TECHNOLOGY POLICY  
WASHINGTON, D.C. 20502

February 22, 2013

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: John P. Holdren  
Director



SUBJECT: Increasing Access to the Results of Federally Funded Scientific Research

**1. Policy Principles**

The Administration is committed to ensuring that, to the greatest extent and with the fewest constraints possible and consistent with law and the objectives set out below, the direct results of federally funded scientific research are made available to and useful for the public, industry, and the scientific community. Such results include peer-reviewed publications and digital data.

Scientific research supported by the Federal Government catalyzes innovative breakthroughs that drive our economy. The results of that research become the grist for new insights and are assets for progress in areas such as health, energy, the environment, agriculture, and national security.

Access to digital data sets resulting from federally funded research allows companies to focus resources and efforts on understanding and exploiting discoveries. For example, open weather