2019 Cornell RCR symposium

Data acquisition and management

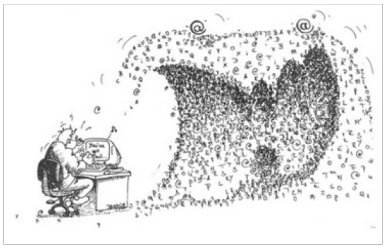
Data acquisition and management

- Key factor in reproducibility (2017 RCR symposium)
- Poor practices can lead to research misconduct, even if no malicious intent.
- Essential for effectively getting research done.

Outline

1. Intro – what are the issues?
2. What can go wrong? – examples from the literature
3. Journal and grant agency policies
4. Available resources at Cornell
5. Brief intro to the cases

Big Data



https://obsoletemedia.org/

https://obsoletemedia.org/

https://obsoletemedia.org/

Data Format Timeline

From Jacquard Loom cards to PlayStation Vita memory cards, the history of data storage on removable media. Comprising disks/discs, tape, solid-state media, ROM cartridges, and punched media.

Dates are approximate and refer to availability in the UK or Europe where known, otherwise for US or elsewhere.

1800s

Jacquard Loom card (1801 – 1990s)

1840s

Punched tape (1846 – 1980s)

1890s

Herman Hollerith's punched card design is used in the 1890 US census (1890)

Punched card (1890 – 1980s)

1940s

Aperture card (1943 – 2000s)

Open reel instrumentation and data logging tape (1949 – 2000s)

1950s

UNISERVO, the first magnetic tape drive for a commercial computer system is introduced (1951)

IBM introduces 7-track magnetic data tape (1952)

IBM introduces the hard disk drive with the IBM 305 RAMAC using 50 24-inch disks storing 5 MB (1956)

LEO tape (1958 – 1981)

2000s

The PlayStation 2, the best-selling video game console of all time, is launched. Arguably, part of its appeal is its ability to play DVD-Video discs (2000)

Sony Memory Stick / Floppy Disk Adaptor (2000 – 2001)

Sony PlayStation 2 Memory Card (2000 – 2013)

Linear Tape-Open (LTO) (2000 –)

USB flash drive (2000 –)

Sega ceases production of the Dreamcast and exits the video game console market (2001)

Microsoft enters the video game console market in the US with the introduction of the Xbox (2001)

Iomega Zip 250 (2001 – 2003)

Iomega Zip U250 (2001 – 2003)

Nintendo Game Boy Advance (2001 – 2007)

CD Cardz (2001 – 2009)

DVD+RW (2001 –)

Super DLT (2001 –)

There is concern that the discs containing the BBC Domesday Project of 1986 may already be unreadable (2002)

Pokémon mini (2002)

Iomega Zip 750 (2002 – 2003)

DataPlay (2002 – mid 2000s)

Nintendo GameCube memory card (2002 – 2007)

Nintendo GameCube (2002 – 2007)

xD-Picture Card (2002 – 2010)

DVD+R (2002 –)

Memory Stick Duo (2002 –)

Qualstar Corporation, the last manufacturer of 9-track open reel computer tape drives, announces the final shipment (2003)

Dell begins removing floppy disk drives from its PCs (2003)

Nokia N-Gage card (2003 – 2006)

miniSD (2003 – 2008)

Super Advanced Intelligent Tape (SAIT) (2003 – 2010)

https://obsoletemedia.org/

Key Concepts in Data Management

- Ownership
- Collection
- Storage
- Sharing
- Protection
- Retention
- Analysis
- Reporting

***ori**.hhs.gov/documents/rcr**intro**.pdf*

From Prof. Maurine Linder

2) What can go wrong?

From Prof. Maurine Linder

Research Articles

Research Article | Spotlight

let-7 Contributes to Diabetic Retinopathy but Represses Pathological Ocular Angiogenesis

Qinbo Zhou, Robert J. A. Frost, Chastain Anderson, Fangkun Zhao, Jing Ma, Bo Yu, Shusheng Wang

Research Article

Acetylation of Cavin-1 Promotes Lipolysis in White Adipose Tissue

Shui-Rong Zhou, Liang Guo, Xu Wang, Yang Liu, Wan-Qiu Peng, Yuan Liu, Xiang-Bo Wei, Xin Dou, Meng Ding, Qun-Ying Lei, Shu-Wen Qian, Xi Li, Qi-Qun Tang

Research Article

RNA Binding Protein CUGBP1 Inhibits Liver Cancer in a Phosphorylation-Dependent Manner

Kyle Lewis, Leila Valanejad, Ashley Cast, Mary Wright, Christina Wei, Polina Iakova, Lauren Stock, Rebekah Karns, Lubov Timchenko, Nikolai Timchenko

Research Article

Cyclin-Dependent Kinase-Dependent Phosphorylation of Sox2 at Serine 39 Regulates Neurogenesis

Shuhui Lim, Akshay Bhinge, Sara Bragado Alonso, Irene Aksoy, Julieta Aprea, Chit Fang Cheok, Federico Calegari, Lawrence W. Stanton, Philipp Kaldis

Research Article | Spotlight

DNA Polymerase Beta Participates in Mitochondrial DNA Repair

P. Sykora, S. Kanno, M. Akbari, T. Kulikowicz, B. A. Baptiste, G. S. Leandro, H. Lu, J. Tian, A. May, K. A. Becker, D. L. Croteau, D. M. Wilson, R. W. Sobol, A. Yasui, V. A. Bohr

Research Article

Targeting Oxygen-Sensing Prolyl Hydroxylase for Metformin-Associated Lactic Acidosis Treatment

Tomoko Oyaizu-Toramaru, Tomohiro Suhara, Noriyo Hayakawa, Takashi Nakamura, Akiko Kubo, Shizuka Minamishima, Kyoji Yamaguchi, Takako Hishiki, Hiroshi Morisaki, Makoto Suematsu, Yoji Andrew Minamishima

Research Article

Bypass of Activation Loop Phosphorylation by Aspartate 836 in Activation of the Endoribonuclease Activity of Ire1

Michael C. Armstrong, Sergej Šestak, Ahmed A. Ali, Hanan A. M. Sagini, Max Brown, Karen Baty, Achim Treumann, Martin Schröder

Author Corrections

Author Correction

Correction for Zehorai and Seger, "Beta-Like Importins Mediate the Nuclear Translocation of Mitogen-Activated Protein Kinases"

Eldar Zehorai, Rony Seger

Author Correction

Correction for Lin et al., "Protein Tyrosine Phosphatase-Like A Regulates Myoblast Proliferation and Differentiation through MyoG and the Cell Cycling Signaling Pathway"

Xi Lin, Xiangsheng Yang, Qi Li, Yanlin Ma, Shuang Cui, Dacheng He, Xia Lin, Robert J. Schwartz, Jiang Chang

Author Correction

Correction for Shah et al., "A Deacetylase-Deficient SIRT1 Variant Opposes Full-Length SIRT1 in Regulating Tumor Suppressor p53 and Governs Expression of Cancer-Related Genes"

Zahid H. Shah, Shafiq U. Ahmed, Jack R. Ford, Simon J. Allison, John R. P. Knight, Jo Milner

Author Correction

Correction for Fukunaga et al., "Activation of Protein Kinase Tel1 through Recognition of Protein-Bound DNA Ends"

Kenzo Fukunaga, Youngho Kwon, Patrick Sung, Katsunori Sugimoto

Author Correction

Correction for Copetti et al., "p65/RelA Modulates *BECN1* Transcription and Autophagy"

Tamara Copetti, Cosetta Bertoli, Emiliano Dalla, Francesca Demarchi, Claudio Schneider

Author Correction

Correction for Travesa et al., "Repression of G_1/S Transcription Is Mediated via Interaction of the GTB Motifs of Nrm1 and Whi5 with Swi6"

Anna Travesa, Tatyana I. Kalashnikova, Robertus A. M. de Bruin, Sarah Rose Cass, Charly Chahwan, David E. Lee, Noel F. Lowndes, Curt Wittenberg

Retraction

Retraction

Retraction for Sorek et al., "Activation Status-Coupled Transient S Acylation Determines Membrane Partitioning of a Plant Rho-Related GTPase"

Nadav Sorek, Limor Poraty, Hasana Sternberg, Einat Bar, Efraim Lewinsohn, Shaul Yalovsky

Research Articles

Research Article | Spotlight

let-7 Contributes to Diabetic Retinopathy but Represses Pathological Ocular Angiogenesis

Qinbo Zhou, Robert J. A. Frost, Chastain Anderson, Fangkun Zhao, Jing Ma, Bo Yu, Shusheng Wang

Research Article

Acetylation of Cavin-1 Promotes Lipolysis in White Adipose Tissue

Shui-Rong Zhou, Liang Guo, Xu Wang, Yang Liu, Wan-Qiu Peng, Yuan Liu, Xiang-Bo Wei, Xin Dou, Meng Ding, Qun-Ying Lei, Shu-Wen Qian, Xi Li, Qi-Qun Tang

Research Article

RNA Binding Protein CUGBP1 Inhibits Liver Cancer in a Phosphorylation-Dependent Manner

Kyle Lewis, Leila Valanejad, Ashley Cast, Mary Wright, Christina Wei, Polina Iakova, Lauren Stock, Rebekah Karns, Lubov Timchenko, Nikolai Timchenko

Research Article

Cyclin-Dependent Kinase-Dependent Phosphorylation of Sox2 at Serine 39 Regulates Neurogenesis

Shuhui Lim, Akshay Bhinge, Sara Bragado Alonso, Irene Aksoy, Julieta Aprea, Chit Fang Cheok, Federico Calegari, Lawrence W. Stanton, Philipp Kaldis

Research Article | Spotlight

DNA Polymerase Beta Participates in Mitochondrial DNA Repair

P. Sykora, S. Kanno, M. Akbari, T. Kulikowicz, B. A. Baptiste, G. S. Leandro, H. Lu, J. Tian, A. May, K. A. Becker, D. L. Croteau, D. M. Wilson, R. W. Sobol, A. Yasui, V. A. Bohr

Research Article

Targeting Oxygen-Sensing Prolyl Hydroxylase for Metformin-Associated Lactic Acidosis Treatment

Tomoko Oyaizu-Toramaru, Tomohiro Suhara, Noriyo Hayakawa, Takashi Nakamura, Akiko Kubo, Shizuka Minamishima, Kyoji Yamaguchi, Takako Hishiki, Hiroshi Morisaki, Makoto Suematsu, Yoji Andrew Minamishima

Research Article

Bypass of Activation Loop Phosphorylation by Aspartate 836 in Activation of the Endoribonuclease Activity of Ire1

Michael C. Armstrong, Sergej Šestak, Ahmed A. Ali, Hanan A. M. Sagini, Max Brown, Karen Baty, Achim Treumann, Martin Schröder

Author Corrections

Author Correction

Correction for Zehorai and Seger, "Beta-Like Importins Mediate the Nuclear Translocation of Mitogen-Activated Protein Kinases"

Eldar Zehorai, Rony Seger

Author Correction

Correction for Lin et al., "Protein Tyrosine Phosphatase-Like A Regulates Myoblast Proliferation and Differentiation through MyoG and the Cell Cycling Signaling Pathway"

Xi Lin, Xiangsheng Yang, Qi Li, Yanlin Ma, Shuang Cui, Dacheng He, Xia Lin, Robert J. Schwartz, Jiang Chang

Author Correction

Correction for Shah et al., "A Deacetylase-Deficient SIRT1 Variant Opposes Full-Length SIRT1 in Regulating Tumor Suppressor p53 and Governs Expression of Cancer-Related Genes"

Zahid H. Shah, Shafiq U. Ahmed, Jack R. Ford, Simon J. Allison, John R. P. Knight, Jo Milner

Author Correction

Correction for Fukunaga et al., "Activation of Protein Kinase Tel1 through Recognition of Protein-Bound DNA Ends"

Kenzo Fukunaga, Youngho Kwon, Patrick Sung, Katsunori Sugimoto

Author Correction

Correction for Copetti et al., "p65/RelA Modulates *BECN1* Transcription and Autophagy"

Tamara Copetti, Cosetta Bertoli, Emiliano Dalla, Francesca Demarchi, Claudio Schneider

Author Correction

Correction for Travesa et al., "Repression of G_1/S Transcription Is Mediated via Interaction of the GTB Motifs of Nrm1 and Whi5 with Swi6"

Anna Travesa, Tatyana I. Kalashnikova, Robertus A. M. de Bruin, Sarah Rose Cass, Charly Chahwan, David E. Lee, Noel F. Lowndes, Curt Wittenberg

Retraction

Retraction

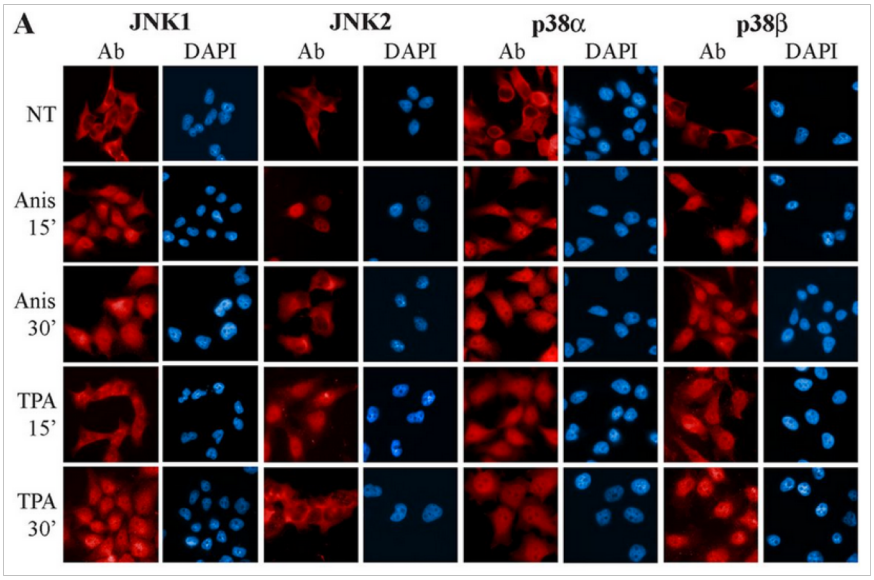
Retraction for Sorek et al., "Activation Status-Coupled Transient S Acylation Determines Membrane Partitioning of a Plant Rho-Related GTPase"

Nadav Sorek, Limor Poraty, Hasana Sternberg, Einat Bar, Efraim Lewinsohn, Shaul Yalovsky

Correction for Zehorai and Seger, “Beta-Like Importins Mediate the Nuclear Translocation of Mitogen-Activated Protein Kinases”

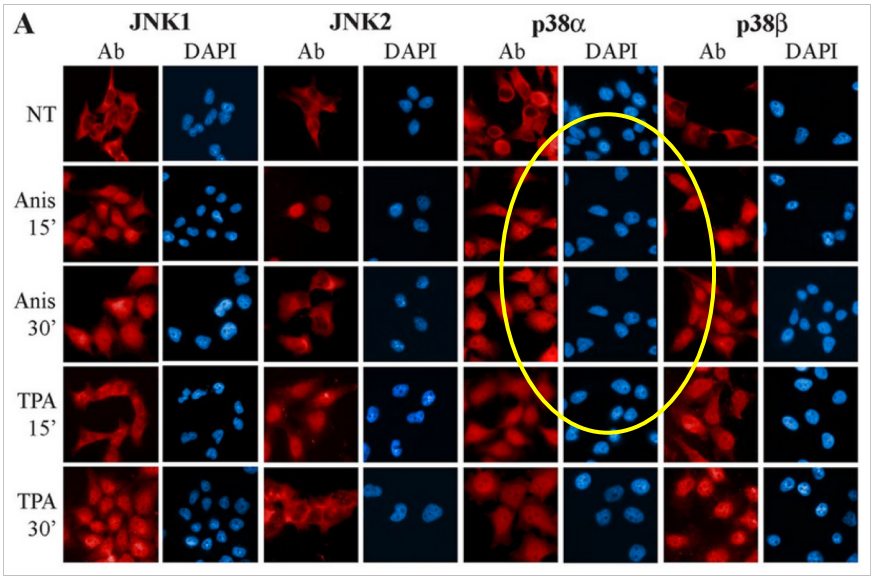
“Due to inadvertent errors in transferring images from Photoshop to Illustrator, there are **five mistakes in four figures**. We apologize for these mistakes and provide corrections below. The conclusions of the study remain unchanged by these corrections”.

What’s wrong with this picture?

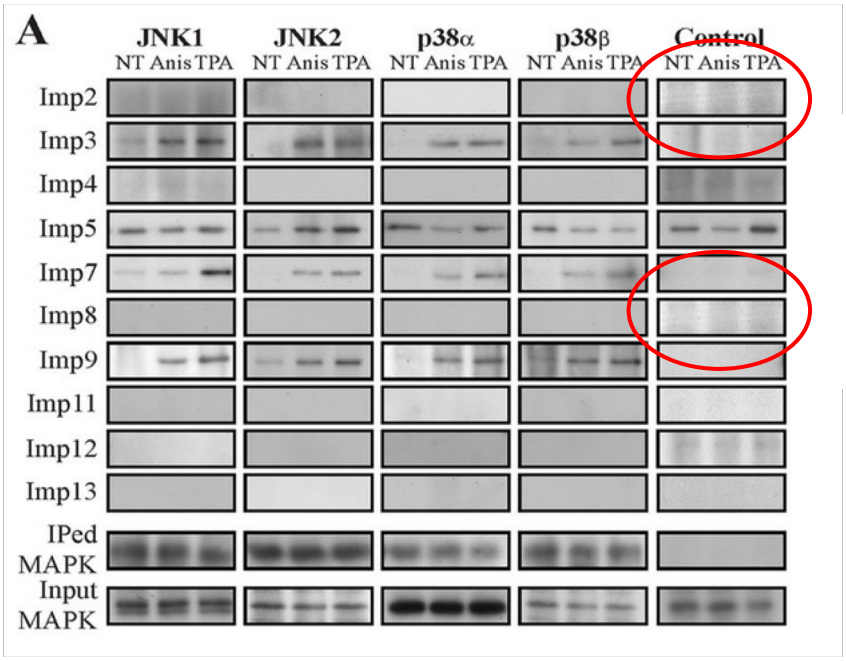


Zehorai and Seger 2014

What's wrong with this picture?



Zehorai and Seger 2014



10 months later …

Retraction for Zehorai and Seger, “Beta-Like Importins Mediate the Nuclear Translocation of Mitogen-Activated Protein Kinases”

After a recent correction of some figures (E. Zehorai and R. Seger, Mol Cell Biol 37:e00215-17, 2017), **we were made aware of additional problems, including an extra lane in the gERK panel of Fig. 1C and inappropriate splicing in Fig. 8A.** Although these errors do not change the conclusions derived from these figures, we choose to retract this paper in the interest of maintaining accuracy in the published scientific literature. We apologize for these mistakes and any inconvenience they caused.

Research Articles

Research Article | Spotlight

let-7 Contributes to Diabetic Retinopathy but Represses Pathological Ocular Angiogenesis

Qinbo Zhou, Robert J. A. Frost, Chastain Anderson, Fangkun Zhao, Jing Ma, Bo Yu, Shusheng Wang

Research Article

Acetylation of Cavin-1 Promotes Lipolysis in White Adipose Tissue

Shui-Rong Zhou, Liang Guo, Xu Wang, Yang Liu, Wan-Qiu Peng, Yuan Liu, Xiang-Bo Wei, Xin Dou, Meng Ding, Qun-Ying Lei, Shu-Wen Qian, Xi Li, Qi-Qun Tang

Research Article

RNA Binding Protein CUGBP1 Inhibits Liver Cancer in a Phosphorylation-Dependent Manner

Kyle Lewis, Leila Valanejad, Ashley Cast, Mary Wright, Christina Wei, Polina Iakova, Lauren Stock, Rebekah Karns, Lubov Timchenko, Nikolai Timchenko

Research Article

Cyclin-Dependent Kinase-Dependent Phosphorylation of Sox2 at Serine 39 Regulates Neurogenesis

Shuhui Lim, Akshay Bhinge, Sara Bragado Alonso, Irene Aksoy, Julieta Aprea, Chit Fang Cheok, Federico Calegari, Lawrence W. Stanton, Philipp Kaldis

Research Article | Spotlight

DNA Polymerase Beta Participates in Mitochondrial DNA Repair

P. Sykora, S. Kanno, M. Akbari, T. Kulikowicz, B. A. Baptiste, G. S. Leandro, H. Lu, J. Tian, A. May, K. A. Becker, D. L. Croteau, D. M. Wilson, R. W. Sobol, A. Yasui, V. A. Bohr

Research Article

Targeting Oxygen-Sensing Prolyl Hydroxylase for Metformin-Associated Lactic Acidosis Treatment

Tomoko Oyaizu-Toramaru, Tomohiro Suhara, Noriyo Hayakawa, Takashi Nakamura, Akiko Kubo, Shizuka Minamishima, Kyoji Yamaguchi, Takako Hishiki, Hiroshi Morisaki, Makoto Suematsu, Yoji Andrew Minamishima

Research Article

Bypass of Activation Loop Phosphorylation by Aspartate 836 in Activation of the Endoribonuclease Activity of Ire1

Michael C. Armstrong, Sergej Šestak, Ahmed A. Ali, Hanan A. M. Sagini, Max Brown, Karen Baty, Achim Treumann, Martin Schröder

Author Corrections

Author Correction

Correction for Zehorai and Seger, “Beta-Like Importins Mediate the Nuclear Translocation of Mitogen-Activated Protein Kinases”

Eldar Zehorai, Rony Seger

Author Correction

Correction for Lin et al., “Protein Tyrosine Phosphatase-Like A Regulates Myoblast Proliferation and Differentiation through MyoG and the Cell Cycling Signaling Pathway”

Xi Lin, Xiangsheng Yang, Qi Li, Yanlin Ma, Shuang Cui, Dacheng He, Xia Lin, Robert J. Schwartz, Jiang Chang

Author Correction

Correction for Shah et al., “A Deacetylase-Deficient SIRT1 Variant Opposes Full-Length SIRT1 in Regulating Tumor Suppressor p53 and Governs Expression of Cancer-Related Genes”

Zahid H. Shah, Shafiq U. Ahmed, Jack R. Ford, Simon J. Allison, John R. P. Knight, Jo Milner

Author Correction

Correction for Fukunaga et al., “Activation of Protein Kinase Tel1 through Recognition of Protein-Bound DNA Ends”

Kenzo Fukunaga, Youngho Kwon, Patrick Sung, Katsunori Sugimoto

Author Correction

Correction for Copetti et al., “p65/RelA Modulates *BECN1* Transcription and Autophagy”

Tamara Copetti, Cosetta Bertoli, Emiliano Dalla, Francesca Demarchi, Claudio Schneider

Author Correction

Correction for Travesa et al., “Repression of G_1/S Transcription Is Mediated via Interaction of the GTB Motifs of Nrm1 and Whi5 with Swi6”

Anna Travesa, Tatyana I. Kalashnikova, Robertus A. M. de Bruin, Sarah Rose Cass, Charly Chahwan, David E. Lee, Noel F. Lowndes, Curt Wittenberg

Retraction

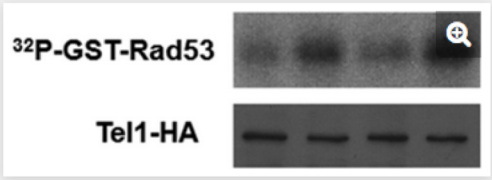
Retraction

Retraction for Sorek et al., “Activation Status-Coupled Transient S Acylation Determines Membrane Partitioning of a Plant Rho-Related GTPase”

Nadav Sorek, Limor Poraty, Hasana Sternberg, Einat Bar, Efraim Lewinsohn, Shaul Yalovsky

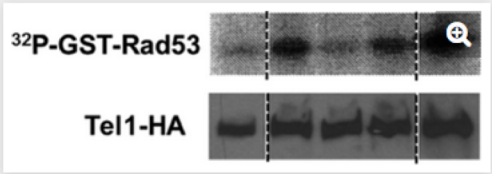
Correction for Fukunaga et al., "Activation of Protein Kinase Tel1 through Recognition of Protein-Bound DNA Ends"

Page 1962, Fig. 2A: An incorrect Western blot image was inadvertently used. The correct version is shown below. Note that the Tel1-HA lanes are identical to the ones used in Fig. 1A and C.



Open in new tab | Download powerpoint

Page 1962, Fig. 2C: A Western blot image was inadvertently duplicated and presented. Since we could not locate the original image, a set of data directly relevant to the experiment is shown below.



3) Journal/funding agency policies

AMERICAN SOCIETY FOR MICROBIOLOGY | *Journals*

ASM Ethical Publication Checklist for Authors

AVOID Authorship Disputes

- ☐ Authors included on the byline meet the ICMJE criteria for authorship.
- ☐ Everyone who meets the ICMJE criteria for authorship has been included in the byline.
- ☐ All authors agree to the order of authorship.
- ☐ All authors have reviewed the manuscript, whether primary submission or resubmission, and have approved submission to this journal.
- ☐ The corresponding author and/or the primary investigator(s), if different, have examined the raw data represented in the manuscript, affirm that such representations accurately reflect the raw data, and will ensure that the original data are preserved and retrievable for at least six years following publication.

DISCLOSE Conflicts of Interest

- ☐ All funding sources and nonfinancial support (including equipment, reagents, and services) have been acknowledged in the manuscript.
- ☐ All financial and personal interests that may be affected by the publication of this study have been declared in both the manuscript submittal letter and in the acknowledgment section of the manuscript.

AVOID Duplicate Submissions and Publications

- ☐ This manuscript is not currently under consideration for publication by another journal.

If this manuscript is a resubmission of a previous submission to this or any other ASM journal:

- ☐ The former manuscript control number has been included in the cover letter and on the submission form.
- ☐ Point-by-point responses to reviews of the previously submitted manuscript have been uploaded as a separate file.
- ☐ A compare copy of the revised manuscript, showing all changes, has been included as a marked up manuscript.

AVOID inappropriate image manipulation

- ☐ Submitted figures and tables show original data.
- ☐ Linear adjustments to contrast, brightness, and/or color, if used, are applied to the entire image.
- ☐ Unless fully disclosed, explained and demarcated (see the following bullet points), no portions of any figures have been:
  - Removed or deleted
  - Concealed
  - Duplicated (cut and pasted)
  - Added
  - Selectively enhanced
  - Repositioned
- ☐ Nonlinear adjustments made to images (e.g., changes to gamma settings) are fully disclosed and explained in the figure legend.
- ☐ Appropriate "tooling" has been used to demarcate individual images that have been compiled to make a figure. Descriptions and explanations for any such compilation are described in the figure legend.

Journal policies – ASM / Molecular and Cellular Biology

- **Image manipulation.** Digital images submitted for publication may be inspected by ASM production specialists for any manipulations or electronic enhancements that may be considered to be the result of scientific misconduct based on the guidelines provided.

PLoS Journals “Data availability” guidelines

- Acceptable Data-Sharing Methods
- Unacceptable Data Access Restrictions
- Explanatory Notes and Guidance
- Recommended Repositories
- Repository Inclusion Criteria
- FAQs for Data Policy
- PLOS Data Advisory Board

https://journals.plos.org/plosbiology/s/data-availability

PLoS Journals “Data availability” guidelines

- Acceptable Data-Sharing Methods
- Unacceptable Data Access Restrictions
- Explanatory Notes and Guidance
- Recommended Repositories
- Repository Inclusion Criteria
- FAQs for Data Policy
- PLOS Data Advisory Board

Cross-disciplinary repositories

- Dryad Digital Repository
- figshare
- Harvard Dataverse Network
- Open Science Framework
- Zenodo

Repositories by type

Biochemistry	**Neuroscience**	**Social Sciences**
Biomedical Sciences	**Omics**	**Structural Databases**
Marine Sciences	**Physical Sciences**	**Taxonomic & Species Diversity**
Model Organisms	**Sequencing**	**Unstructured and/or Large Data**

https://journals.plos.org/plosbiology/s/data-availability

Biomedical Sciences

- The Cancer Imaging Archive (TCIA)
- ImmPort
- Influenza Research Database
- National Addiction & HIV Data Archive Program (NAHDAP)
- National Database for Autism Research (NDA
- PhysioNet
- SICAS Medical Image Repository

Marine Sciences

- SEA scieNtific Open data Edition (SEANOE)

Model Organisms

- The Arabidopsis Information Resource (TAIR)
- Eukaryotic Pathogen Database Resources (EuPathDB)
- FlyBase
- Mouse Genome Informatics (MGI)
- Rat Genome Database (RGD)
- SmedGD
- VectorBase
- WormBase
- Xenbase
- Zebrafish Model Organism Database (ZFIN)

Neuroscience

- Functional Connectomes Project International Neuroimaging Data-Sharing Initiative (FCP/INDI)
- German Neuroinformatics Node/G-Node (GIN)
- NeuroMorpho.org
- OpenNeuro

Omics

- ArrayExpress
- Biological General Repository for Interaction Datasets (BioGRID)
- Database of Interacting Proteins (DIP)
- dbGAP
- The European Genome-phenome Archive (EGA)
- Gene Expression Omnibus (GEO)
- GenomeRNAi
- GPM DB
- IntAct Molecular Interaction Database
- MetaboLights
- Metabolomics Workbench
- NURSA
- PeptideAtlas
- ProteomeXchange
- Proteomics Identifications (PRIDE)

Funding agency requirements

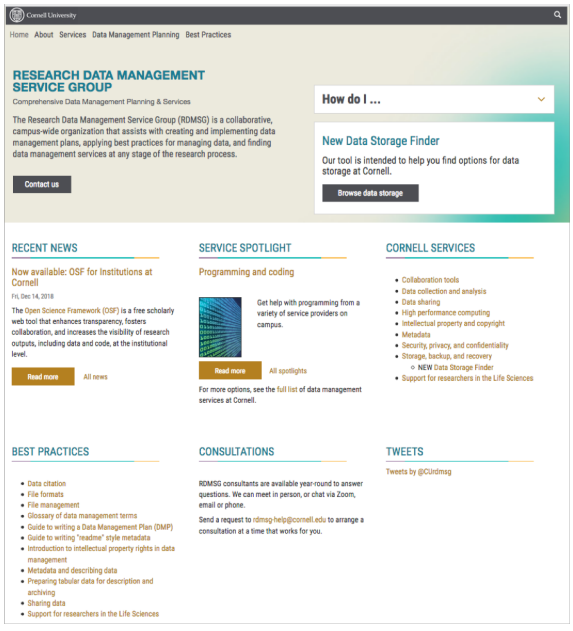
'In discussing retention requirements, one often hears that institutions must retain records on federal awards "for three years". There is just enough truth in that statement to get one into serious trouble. The key question is "three years from when?"'

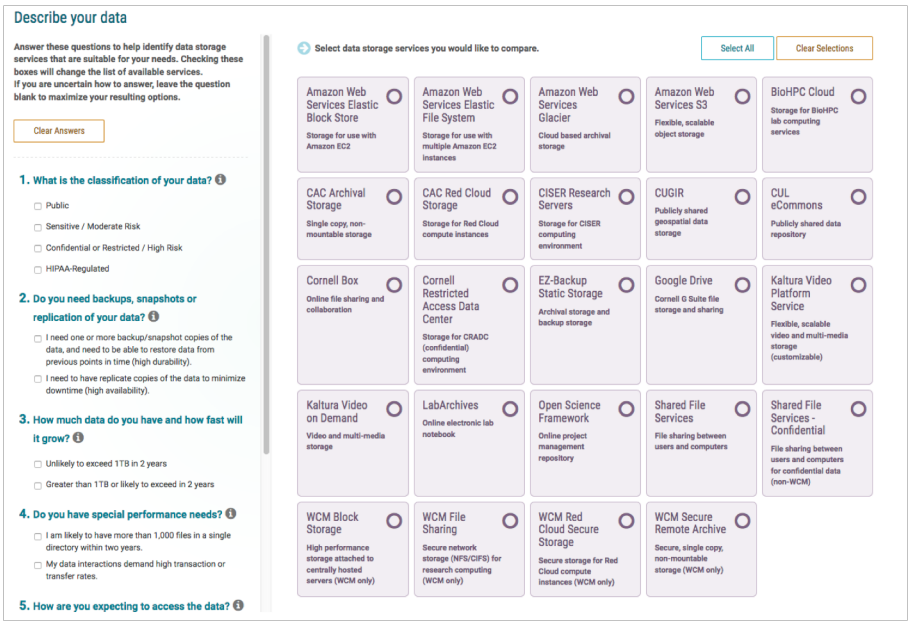
NIH: "the retention period is three years from the date the final financial report is submitted".

NSF: "records must be retained for three years after the submission of all required reports (research and other special reports)".

https://ori.hhs.gov/education/products/rcradmin/topics/data/tutorial_11.shtml

4) Cornell resources: data.research.cornell.edu





5) Intro to case studies

Case 1: Managing research data.

- Prof Carr and his unavailable image analysis computer program.
- “How does this general scenario apply to your own type of research”?
- Unspecified: how old is the paper?
- Unspoken: Involves student/advisor relationship.

Case 2: Data sharing.

- Henryk and his large database of infectious agent outcomes.
- "How does this general scenario apply to your own type of research"?
- Unspecified: How many others contributed to the database?
- Unspoken: involves postdoc/advisor relationship.

Thank you! CU Office of Research Integrity and Assurance

Amita Verma
Myles Gideon
Bridget MacRae
Guilaine Senecal