The Burden and Benefits of ‘Long-tail’ data sharing

Adam R. Ferguson, Ph.D.
Associate Professor and Principal Investigator
Brain and Spinal Injury Center (BASIC)
Department of Neurological Surgery
Weill Institute for Neurosciences, UCSF
Principal Investigator, SFVAHCS
Disclosures

Grants:
US National Institutes of Health: R01NS088475, R01CA213441, R01AG056770, R01MH116156, U01NS086090, P30AR066262, UG3NS106899
US Veterans Affairs: 1I01RX002245, I01RX002787
US Department of Defense: SC150198, SC150177
US Department of Energy
Craig H. Neilsen Foundation
Wings for Life Foundation

Data Science Consulting (noncommercial):
Moody Project for Translational TBI
Santa Clara Valley Medical Center
Central Nervous System (CNS) Injury is Complex!
Gazillions of tiny measures of biofunction

Figure Sources: Rosenzweig et al., 2010; 2019 Nat Neurosci; Ferguson et al., 2013 PloS One; Nielson et al., 2014, J Neurotrauma; Nielson et al., 2015, Brain Res.; Friedli et al., 2015 Science TM; Rosenzweig et al., 2019 Nat Medicine
Figure Sources: Rosenzweig et al., 2010; 2019 Nat Neurosci; Ferguson et al., 2013 PloS One; Nielson et al., 2014, J Neurotrauma; Nielson et al., 2015, Brain Res.; Friedli et al., 2015 Science TM; Rosenzweig et al., 2019 Nat Medicine
Figure Sources: Rosenzweig et al., 2010; 2019 Nat Neurosci; Ferguson et al., 2013 PloS One; Nielson et al., 2014, J Neurotrauma; Nielson et al., 2015, Brain Res.; Friedli et al., 2015 Science TM; Rosenzweig et al., 2019 Nat Medicine
CNS Injury as a ‘big data’ problem?

Ferguson et al., Nat Neuro 2014; Huie et al., 2019, Curr Opin. Neurology

-volume
-velocity
-variety
CNS Injury as a ‘big data’ problem?

-volume
-velocity
-variety

Ferguson et al., Nat Neuro 2014; Huie et al., 2019, Curr Opin. Neurology
TOWARD ‘DATAFICATION’ OF BIOMEDICAL RESEARCH

Ferguson et al., 2014, *Nature Neuroscience*
Huie et al., 2019, *Current Opinion in Neurology*
Hawkins et al., 2019, *Journal of Neurotrauma*
TOWARD ‘DATAFICATION’ OF BIOMEDICAL RESEARCH

- Ferguson et al., 2014, *Nature Neuroscience*
- Huie et al., 2019, *Current Opinion in Neurology*
- Hawkins et al., 2019, *Journal of Neurotrauma*
TOWARD ‘DATAFICATION’ OF BIOMEDICAL RESEARCH

Estimated 85% Dark Data Worldwide: $240 B in Research/yr

Ferguson et al., 2014, *Nature Neuroscience*
Huie et al., 2019, *Current Opinion in Neurology*
Hawkins et al., 2019, *Journal of Neurotrauma*
TOWARD ‘DATAFICATION’ OF BIOMEDICAL RESEARCH

Estimated 85% Dark Data Worldwide: $240 B in Research/yr
Estimated Value of Dark Data>$200 B

(MacLeod et al, Lancet, 2014)

Ferguson et al., 2014, Nature Neuroscience
Huie et al., 2019, Current Opinion in Neurology
Hawkins et al., 2019, Journal of Neurotrauma
Why Most Published Research Findings Are False

John P. A. Ioannidis
Scientific Publication:

- **Methods**
  - Protocol

- **Data**
  - Data collection forms
  - Raw participant-level dataset
  - Clean participant-level dataset
  - Summary analyses

- **Full study report**

- **Registry and results database**

- **Published primary report**

- **Conference report**

- **<10 pages**

Up to hundreds to thousands of pages

Increasing information loss and potential for selective reporting

Chan et al., 2014, *Lancet*
Scientific Publication: 
A 17th century data-sharing technology
21st Century Solution?
~60 million data-points from ~4000 basic science and human medical records; 13 centers

‘VISION-SCI’ Repository
~60 million data-points from ~4000 basic science and human medical records; 13 centers

‘VISION-SCI’ Repository

SCI Open Data Commons Initiative
~60 million data-points from ~4000 basic science and human medical records; 13 centers

‘VISION-SCI’ Repository

SCI Open Data Commons Initiative

TBI Open Data Commons

http://odc-tbi.org
~60 million data-points from ~4000 basic science and human medical records; 13 centers
~60 million data-points from ~4000 basic science and human medical records; 13 centers

‘VISION-SCI’ Repository

SCI Open Data Commons Initiative

VA Gordon Mansfield SCI Consortium

http://odc-tbi.org
FAIR Data Stewardship

- **Findable:** (meta)data is uniquely and persistently identifiable. Should have basic machine readable descriptive metadata.

- **Accessible:** data is reachable and accessible by humans and machines using standard formats and protocols.

- **Interoperable:** (meta)data is machine readable and annotated with resolvable vocabularies/ontologies.

- **Reusable:** (meta)data is sufficiently well-described to allow (semi)automated integration with other compatible data sources.

Spinal Cord Injury Preclinical Data Workshop:

Developing a FAIR Share Community

Sharing Data that is

Findable

Accessible

Interoperable

Reusable

October 4-5, 2016
6001 Executive Blvd.
North Bethesda, MD
Spinal Cord Injury Preclinical Data Workshop:

Developing a FAIR Share Community

Sharing Data that is
Findable
Accessible
Interoperable
Reusable

October 4-5, 2016
6001 Executive Blvd.
North Bethesda, MD

FAIR-SCI Ahead

SCI Preclinical Community Readiness and Next Steps

Washington DC, November 10, 2017
FAIR-SCI Ahead

SCI Preclinical Community Readiness and Next Steps

Washington DC, November 10, 2017

STREET-FAIR

SCI Team Research, Enabling Expansion and Translation of FAIR data sharing

November 4, 2018
San Diego, CA
So you now have big data…

THEN WHAT??
Syndromic Data Integration

Functional Changes

- Individual (rat, mouse, patient)

Cognitive Performance

Neurophysiology
Syndromic Data Integration

Functional Changes

Cognitive Performance

Neurophysiology
Syndromic Data Integration

- Functional Changes
- Biomarkers
- Tissue Changes
Syndromic Data Integration

Functional Changes

Biomarkers

Tissue Changes
Syndromic Data Integration

Functional Changes
Biomarkers
Tissue Changes
Syndromic Data Integration

Group 1  Group 2  Group 3

Functional Changes
Biomarkers
Tissue Changes
Syndromic Data Integration

Group 1  Group 2  Group 3

Functional Changes

Biomarkers

Tissue Changes
Precision Medicine

Patient Group 1
Patient Group 2
Patient Group 3

Functional Changes
Biomarkers
Tissue Changes
Precision Reproducibility?

Lab Group 1
Lab Group 2
Lab Group 3

Functional Changes
Biomarkers
Tissue Changes

Lab Group 1
Lab Group 2
Lab Group 3

Precision Reproducibility?
Precision Reproducibility?
Topological data analysis for discovery in preclinical spinal cord injury and traumatic brain injury

Jessica L. Nielson¹, Jesse Paquette², Aiwen W. Liu¹, Cristian F. Guandique¹, C. Amy Tovar³, Tomoo Inoue⁴, Karen-Amanda Irvine⁵, John C. Gensel⁶, Jennifer Kloke⁷, Tanya C. Petrossian⁸, Pek Y. Lum⁹, Gunnar E. Carlsson⁷,¹⁰, Geoffrey T. Manley¹, Wise Young¹¹, Michael S. Beattie¹, Jacqueline C. Bresnahan¹ & Adam R. Ferguson¹,¹²
Long-tail data archive: Multicenter animal spinal cord injury study (MASCIS) 1994-1996
Multicenter Animal SCI Study (MASCIS 1994-96)

Locomotor Recovery

Nielson et al., 2015
Nature Communications
Multicenter Animal SCI Study (MASCIS 1994-96)

Locomotor Recovery

Blood Pressure During Surgery

Nielson et al., 2015
Nature Communications
Multicenter Animal SCI Study (MASCIS 1994-96)

Locomotor Recovery

Blood Pressure During Surgery

Nielson et al., 2015
Nature Communications

Jessica Nielson, PhD
Dark data see the light

Ellen P. Neff

LabAnimal  46  Volume 47, No. 2 | FEBRUARY 2018
Dark data see the light

Ellen P. Neff

LabAnimal

46 Volume 47, No. 2 | FEBRUARY 2018

Predictable response:
Finding optimal drugs and doses using artificial intelligence

By Shraddha Chakradhar

NATURE MEDICINE VOLUME 23 | NUMBER 11 | NOVEMBER 2017
UCSF Weill Institute for Neurosciences
Brain and Spinal Injury Center

Zuckerberg San Francisco General Hospital and Trauma Center

VA Department of Veterans Affairs
FERGUSON LAB:
Carlos Almeida, BS, MA
Carla Arellano, DPT
Daniel Fong, DPT
Jenny Haefeli, PhD
J Russell Huie, PhD
Sheena McCormack, DPT
Kazuhito Morioka, MD, PhD
Jonathan Namnath, BS
Dave Namnath, BS
Jessica L Nielsen, PhD
Cleopa Omondi, BS, MS
Ellen Stuck, DPT student
Jeffery Sacramento, BA
Dolores Torres, BS, MS2
Christina Tsutsumi, DPT student
Jennifer Truong, DPT student
Lauren VanCitters, DPT
Nadine Joseph, BS (cand.)

Collaborators
UCSF Beattie/Bresnahan Lab:
Michael S. Beattie, PhD
Jacqueline C. Bresnahan, PhD
Sang Mi Lee, PhD
Nikos Kyritsis, PhD
Xiaokui Ma, MD

UCSF Manley Lab:
Tomoo Inoue, MD, PhD
Geoff T Manley MD, PhD
Mary Vassar, RN, MS
John Yue, BA, MS2

UCSF/PT Rehab:
Susanna Rosi, PhD

UCSF Biostatistics:
Mark R Segal, PhD

UCSF Neurology/SFVA:
Ray Swanson, MD
Steve Massa, MD/PhD
Raquel Gardner, MD

UCSF/Anesth & Periop Care:
Jonathan Pan, MD/PhD
Mervyn Maze, MD
Hua Su, MD

UCSF/Radiology:
Esther Yuh, MD/PhD
Pratik Mukherjee, MD/PhD
Jason Talbott, MD/PhD
Sharmila Majumdar, PhD

UCSF Orthopedics:
Jeff Lotz, PhD
Chelsey Bahney, PhD
Ralph Marcubio, PhD

University of Minnesota
Jessica Nielson, PhD
Sophia Vinogradov, MD
Sisi Ma, PhD
Erich Kummerfeld, PhD

Ohio State University:
Phillip G. Popovich, PhD
Dana M. McTigue, PhD
Jan Schwab, MD, PhD
Michele Basso, EdD

Texas A&M University:
Jim W. Grau, PhD
Michelle A. Hook, PhD

Funding
US National Institutes of Health:
R01NS088475, R01CA213441, R01AG056770, U01NS086090, R01MH116156, UG3NS106899

US Veterans Affairs: 1101RX002245, I01RX002787
US Department of Defense: SC150198, SC150177

Craig H. Neilsen | Wings for Life | International Spinal Research Trust
Questions?