# NIDA Big Data Strategic Planning Workgroup

June 17, 2015

Co-Chairs:
Roger Little, Ph.D. (NIDA)
Massoud Vahabzadeh, Ph.D. (NIDA)

# Today's Agenda

- Welcome
- Review of timeline, work group priorities/activities and June 3th meeting
- Comments regarding data curation & analysis and visualization issues and opportunities homework documents
- Today's Topic – NIDA Addictome
- Other Action Items
- 5 Minute Public Comment Period
- Wrap-up and Adjourn

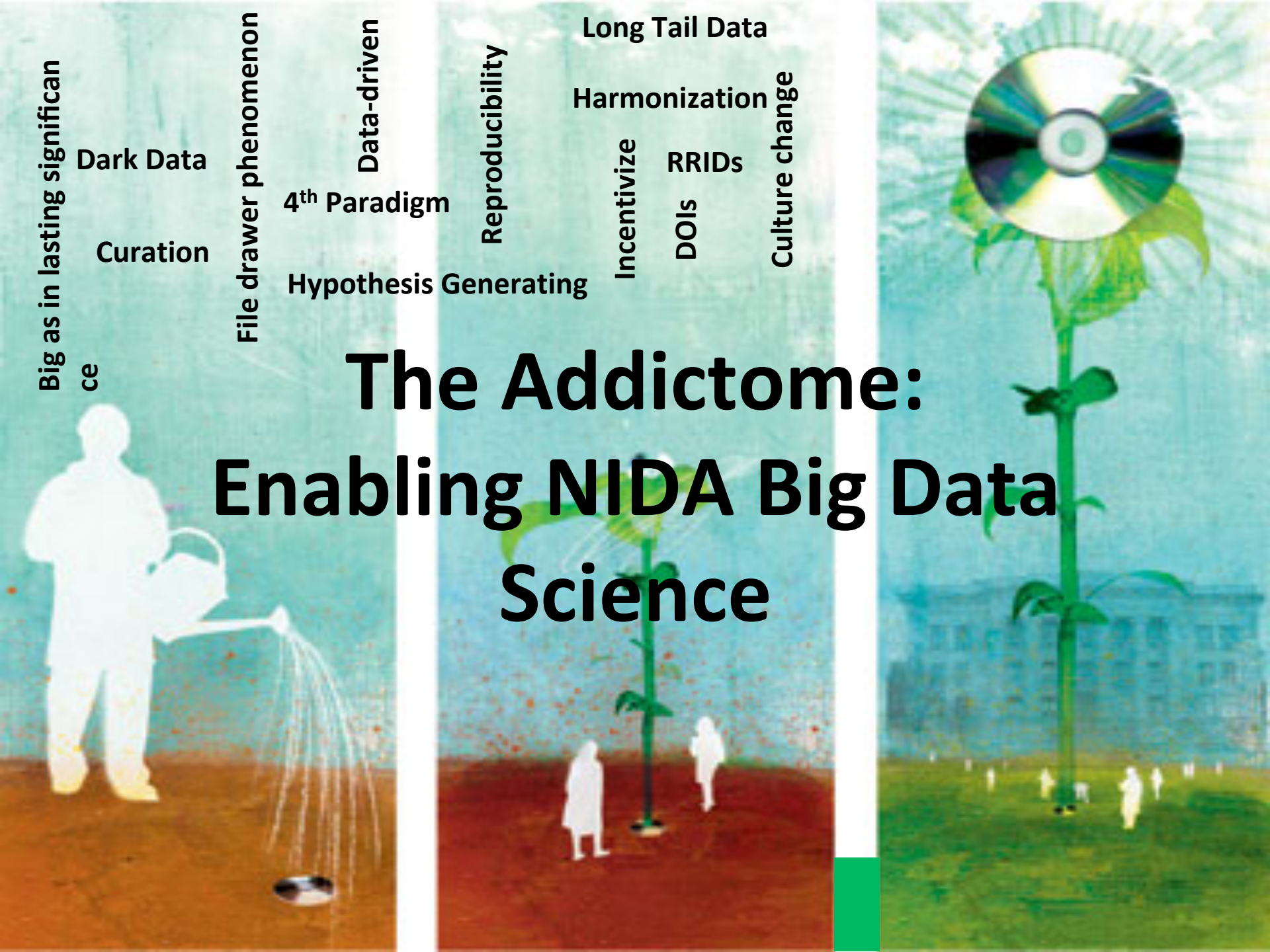
# NIDA Big Data Strategic Planning Work Group Charge and Timeline

- Charge
  - Develop recommendations on best approaches for NIDA to utilize Big Data to advance strategic priorities, including data sharing/ access, analysis, visualization, reproducibility, negative data resource
  - Deliverable = 3-5 page summary of recommendations on leveraging 'Big Data' in the next 5 years
  - **Completion Date = Friday, June 26, 2015**

- Timeline
  - June 30, 2015 – RFI soliciting feedback from field regarding NIDA's strategic priorities closed
  - February – June 2015 – Priority area workgroups formed
    - Priority areas = Big Data, Gene x Environment x Development interactions
  - By August, 2015 – Bold Goals Challenge winner selected
  - By Summer 2015, Draft Strategic Plan out for public comment
  - By Fall 2015, Final Strategic Plan

# NIDA Big Data Strategic Planning Work Group Identified Priorities

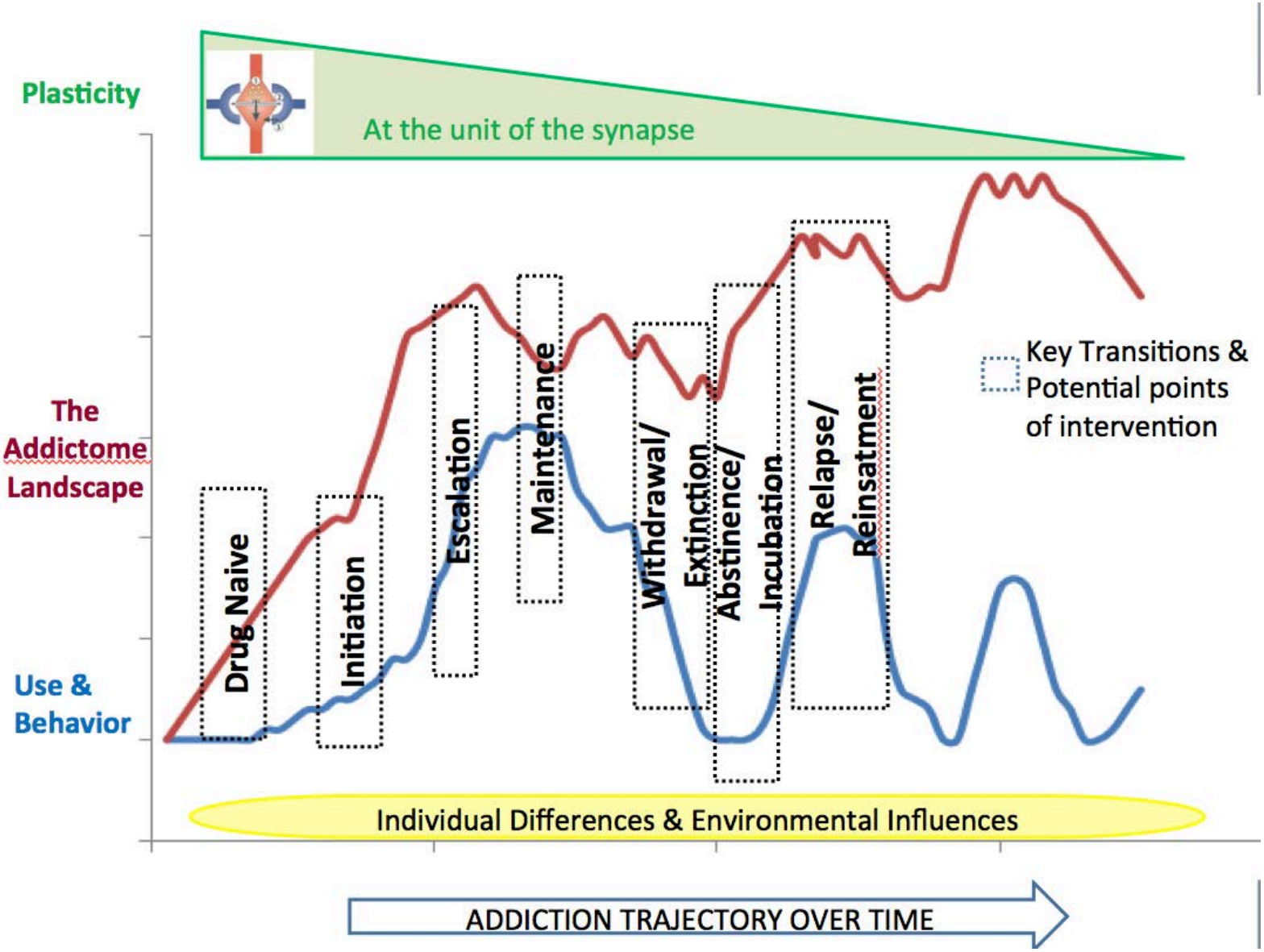
- Priority areas identified by work group
  - Data Sharing
  - Data Capture and File Formatting
  - Data Curation & Analysis, Visualization, Machine Learning
- Where are we in the document development process?
- What's next?




# The Addictome: Enabling NIDA Big Data Science

# Why create an addictome data resource?

- To enable Big Data science
- Data reproducibility
- Place for negative data
- Quality metrics and standards

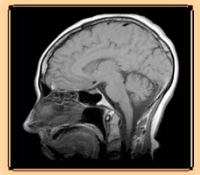

Plasticity
At the unit of the synapse
The Addictome Landscape
Use & Behavior
Drug Naive
Initiation
Escalation
Maintenance
Withdrawal/ Extinction
Abstinence/ Incubation
Relapse/ Reinsatment
Key Transitions & Potential points of intervention
Individual Differences & Environmental Influences
ADDICTION TRAJECTORY OVER TIME

# Addictome.org

Genomic & Epigenomic

Other 'Omics

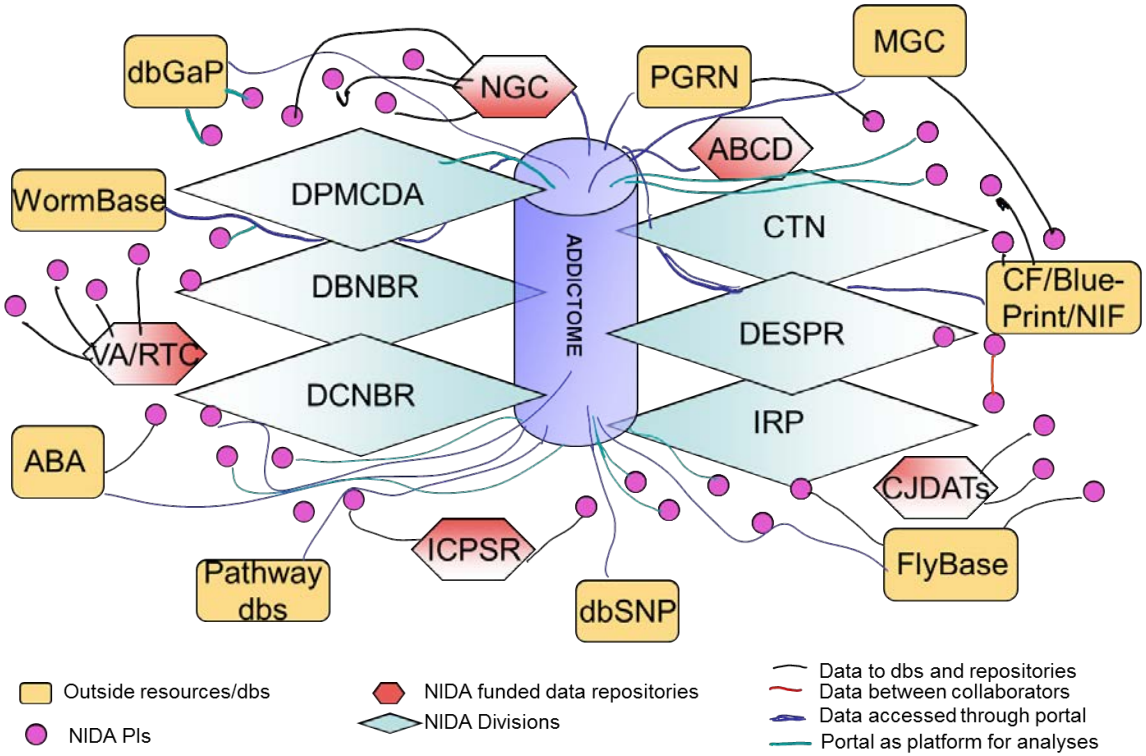
Developmental



Imaging

Phenotypic

Clinical

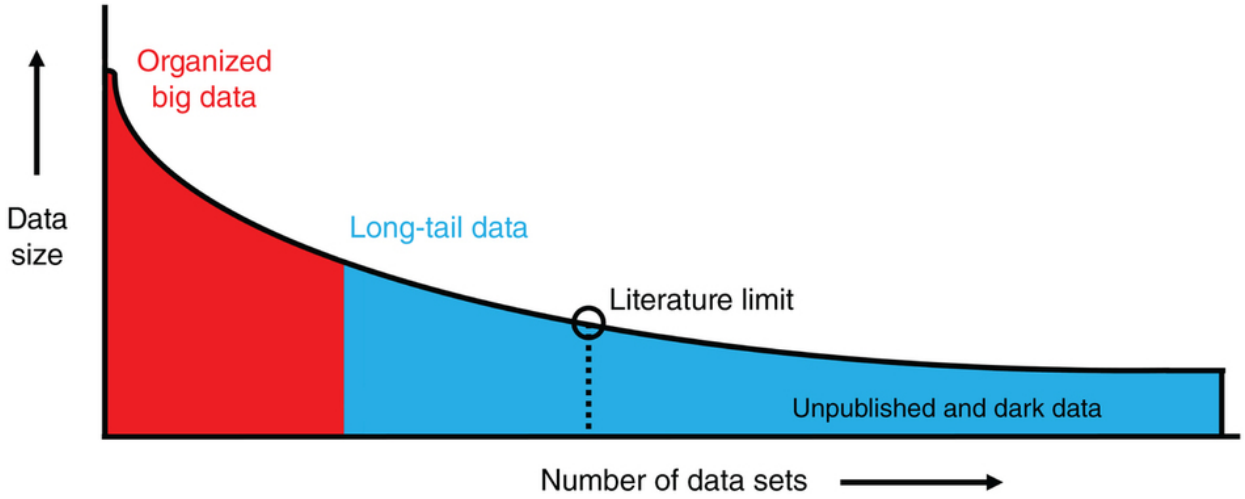



Exposure

Organism

Familial

# Successful long-tail data sharing




- IMPACT - 20 years of clinical TBI (43,243 patients with TBI, 62 publications)
- MIASCI and VISION-SCI – retrospective pre-clinical
- “CRCNS Data Share” for computational modeling – 10 yrs, 40 datasets (5 – 450 GB), 37 secondary analysis publications!
- small amounts of adequately characterized, focused data are preferable to large amounts of inadequately defined and controlled data stored in a random repository. Gardner et al. 2003

# Establishing the Addictome.org

- **Research Areas:** Pre-clinical electrophysiology, relapse stage of addiction
- **Goal:** Convene investigators in these areas of research with exp erts in data to establish a data sharing framework elements, and standards.
- **Deliverables:**
  - A minimum set of common data elements for addiction electrophysiology
  - Create a process to expand the scope of the Addictome to other data types
  - Process and workflows
  - Metadata and labeling of datasets
  - Format: Data types, cost and benefit of sharing raw vs. processed data
  - Costs and funding of data curation and storage, storage repositories
  - Incentivized and/or required data sharing
  - Providing the data in a findable, searchable system

# Addictome Meeting Series Agenda

1. Culture Change, convey rationale for establishing the Addictome
2. Use Cases: Investigator applications for big data science
3. Data Formats, Types and Sizes, Raw vs. processed data, Common Data Elem ents (CDEs)
4. Minimum information for a drug addiction (electrophysiology) experiment Metadata
5. Investigator Recommendations towards data curation and storage
6. Providing the data in a findable, searchable system
7. Workflow of producing and sharing research data
8. Workflow of finding and reusing shared research data
9. Ongoing communications, extensibility
10. Plan for expansion into additional DBNBR and NIDA research domains

# Thank you to
# Big Data WG Members

| NAME | AFFILIATION |
|---|---|
| WORKGROUP CHAIRS | |
| Roger Little, PhD | NIDA |
| Massoud Vahabzadeh, PhD | NIDA |
| EXTRAMURAL WORKGROUP MEMBERS | |
| Christopher Chute, MD, DrPH | Johns Hopkins University |
| Michael Neale, PhD | Virginia Commonwealth University |
| Eric Nestler, MD, PhD | The Mount Sinai Hospital |
| Michael Milham , MD, PhD | Child Mind Institute |
| Arthur Toga, PhD | University of Southern California |
| Maryanne Martone, PhD | University of California at San Diego |
| NIH STAFF | |
| Philip Bourne, PhD | NIH Associate Director for Data Science |
| Maureen Boyle, PhD | NIDA |
| Ericka Boone, PhD | NIDA |
| Udi Ghitza, PhD | NIDA |
| Steve Gust, PhD | NIDA |
| Vani Pariyadath, PhD | NIDA |
| Tom Radman, PhD | NIDA |
| Joni Rutter, PhD | NIDA |
| Tisha Wiley, PhD | NIDA |