

NIDA Big Data Strategic Planning Workgroup

April 22, 2015

Coo Chairs:

Roger Little, Ph.D. (NIDA) Masso
ud Vahabzadeh, Ph.D. (NIDA)

Agenda

- Welcome
- Workgroup and WebEx Logistics
- Review of HW and Feedback
 - Highest ranking 'Importance and Opportunities' and 'Major Concerns' and 'Comments & Questions' items
- Other Action Items (e.g., new work group homework assignment, etc)
- 5 Minute Public Comment Period
- Adjourn

Next Meeting

Wednesday, April 29, 2015

3:00-4:00 pm (EST)

Workgroup Logistics

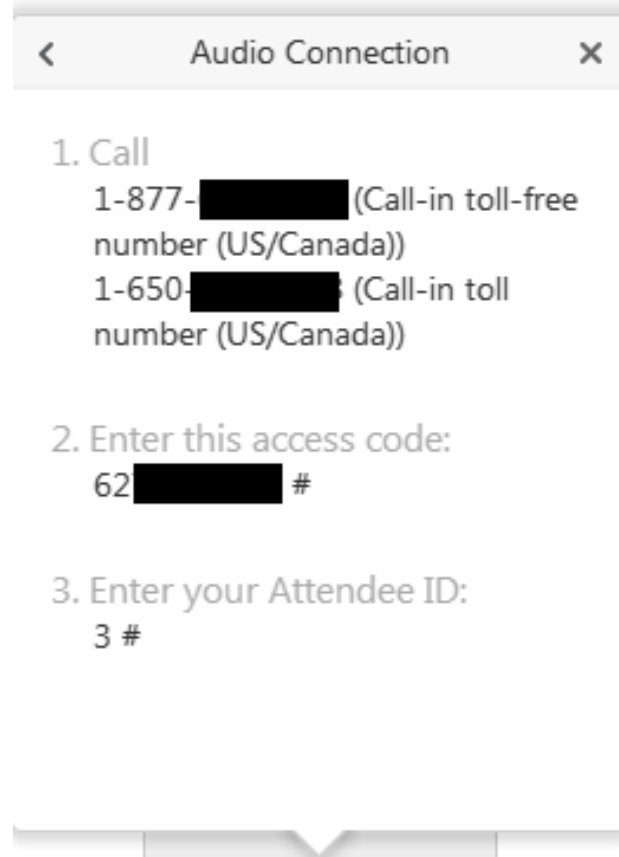
- Meet biweekly on Wednesdays from 3-4 pm EST via WebEx
- All meetings are open to the public and will be recorded
- Meeting summaries will be provided to workgroup members within 5 business days
- All correspondence should be sent to:
NIDAOSPCPlanning@mail.nih.gov
- A web-based site will be established to facilitate document sharing

WebEx Logistics

- Workgroup members have been made **panelists** in this WebEx Event
- You should have received the special panelist WebEx access url by email
- **First**, join the WebEx Event by **clicking on your panelist** url and enter all passcodes and passwords contained in the email

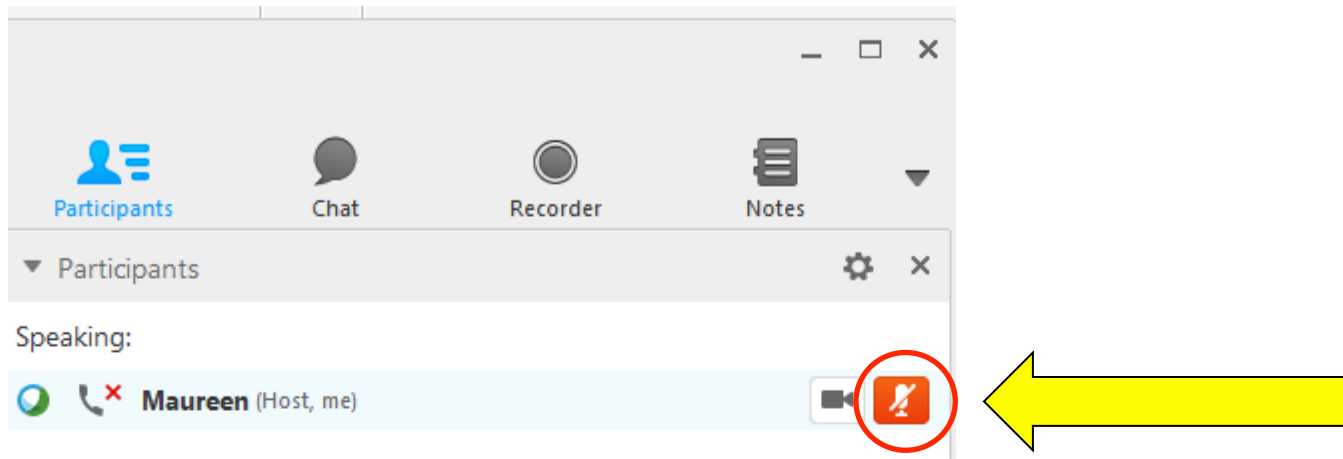
WebEx Logistics

- **Second**, join the WebEx Event by phone
 - Please dial in through the phone instead of using your computer and enter access code and attendee ID:



WebEx Logistics

- **Third**, you may mute yourself during the event to reduce background noise.
- We will mute panelists as necessary to control noise. Please **unmute to speak**.



WebEx Logistics

- **Note:**
 - All strategic planning WebEx events are open to the public and recorded
 - Non-workgroup members (attendees) will be muted throughout the meeting until the end
 - Public comments and questions will be accepted via chat and phone during the last 5 minutes
 - There is a note-taker on the call and summaries will be posed on the website:
<http://www.drugabuse.gov/about-nida/strategic-plan>

Homework – Importance & Opportunity

Please rank in order of:		Importance			Opportunity		
		Low=1	Med=2	High=3	Low=1	Med=2	High=3
1. Capture	a. data dictionaries		1	4		1	2
	b. ontologies		4			3	
	c. common data elements		2	3		2	1
	other? Please write in:						
2. Storage	a. cloud	1	2	2		1	2
	b. decentralized	2	2	1		1	1
	c. centralized	1	3	1		2	
	d. discoverability		1	3			2
	other? Please write in:						
3. File Format	a. unstructured data	2		3		1	2
	b. preprocessing	1	2	2		1	2
	c. metadata		3	2		3	
	other? Please write in:						
4. Data Curation & Analysis							
5. Sharing	a. transfer (simply too large: doesn't move very well)		2	3		1	2
	b. distribution (hard to aggregate)		3	2			3
	c. complexity (high computational expense)		3	2		1	2
	d. security	1	1	3	1	1	1
	e. privacy	2		3	1		2
	other? Please write in:						
6. Visualization	a. analysis tools	1		5			4
	b. machine learning	2	2	2		2	2
	other? Please write in:						

Homework –Importance & Opportunity

Data Curation & Analysis

5. Sharing	a. transfer (simply too large: doesn't move very well)		2	3		1	2
	b. distribution (hard to aggregate)		3	2			3
	c. complexity (high computational expense)		3	2		1	2
	d. security	1	1	3	1	1	1
	e. privacy	2		3	1		2
	other? Please write in:						

Homework – Importance & Opportunity

Other Areas of Concern

Please rank in order of:		Importance			Opportunity		
		Low=1	Med=2	High=3	Low=1	Med=2	High=3
1. Capture	a. data dictionaries		1	4		1	2
	b. ontologies		4			3	
	c. common data elements		2	3		2	1
	other? Please write in:						
2. Storage	a. cloud	1	2	2		1	2
	b. decentralized	2	2	1		1	1
	c. centralized	1	3	1		2	
	d. discoverability		1	3			2
	other? Please write in:						
3. File Format	a. unstructured data	2		3		1	2
	b. preprocessing	1	2	2		1	2
	c. metadata		3	2		3	
	other? Please write in:						
4. Data Curation & Analysis							
5. Sharing	a. transfer (simply too large: doesn't move very well)		2	3		1	2
	b. distribution (hard to aggregate)		3	2			3
	c. complexity (high computational expense)		3	2		1	2
	d. security	1	1	3	1	1	1
	e. privacy	2		3	1		2
	other? Please write in:						
6. Visualization	a. analysis tools	1		5			4
	b. machine learning	2	2	2		2	2
	other? Please write in:						

Homework – Major Concerns

Please rate your level of concern about these issues	Low=1	Med=2	High=3
Sharing my data	5	0	0
Maintaining Privacy/security of my data	2	0	3
Reproducibility: harmonizing to compare apples to apples	0	0	5
Data standardization to enable merging of disparate data sets	1	0	3
Finding a place to store my data	1	4	0
Optimal formats for long term storage	1	4	0
other? Please write in:			

Homework – Questions & Concerns

- **Comments on Best Data vis-à-vis Usage/Non-Usage**
 - Quality and provenance are critical
 - Model where all data released and users can use their own standards and needs
 - Crowd sourcing
 - Common data elements in large healthcare settings and EHRs
 - Uniform data domains, vocabulary, standards and interoperability

Homework – Questions & Concerns

- **What opportunities are there for NIDA-specific research?**
 - Curation of SUD phenotypes with exome and whole genome sequence data
 - Avoidance of random associations with big numbers
 - Data aggregation for longitudinal datasets in difficult to follow populations
 - Optimal use of digital data systems and common data elements
 - Concerns regarding reproducibility

Homework – Questions & Concerns

- **What are relevant or unique training opportunities?**
 - Development of statistical methods appropriate for complexities of SUD phenotypes
 - Training of individuals in areas of medical informatics/ bioinformatics training
 - Retraining opportunities for junior and mid-career level scientists

Agenda

- Welcome
- Workgroup and WebEx Logistics
- Review of HW and Feedback
 - Highest ranking 'Importance and Opportunities' and 'Major Concerns' and 'Comments & Questions' items
- Other Action Items (e.g., new workgroup homework assignment, etc)
- 5 Minute Public Comment Period
- Adjourn

Next Meeting

Wednesday, April 29, 2015

3:00-4:00 pm (EST)