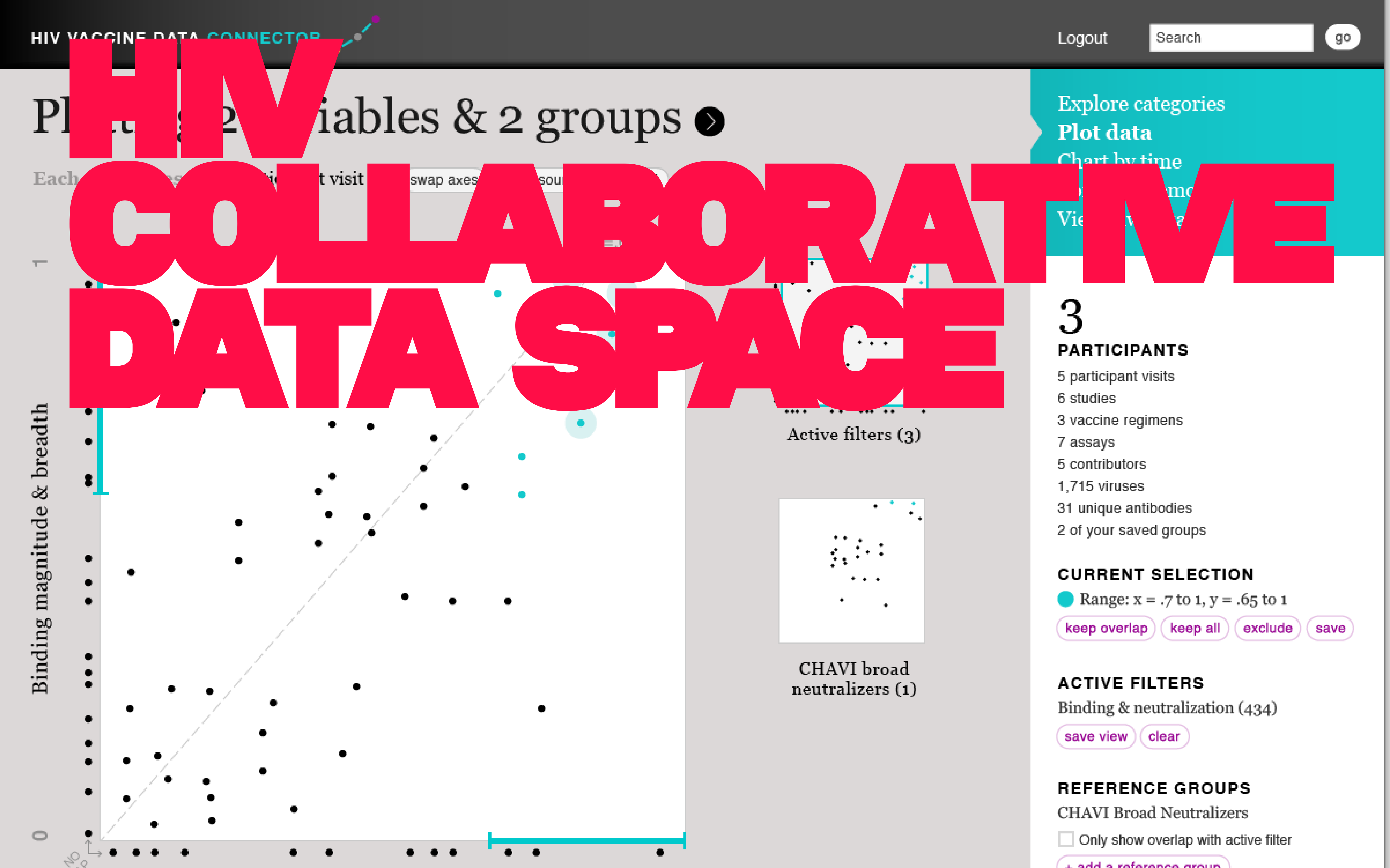
client Labkey Users Conference

project User-centric design for research tools: The CDS as a case study

date September, 2012

HIV COLLABORATIVE DATA SPACE



3

PARTICIPANTS

5 participant visits
6 studies
3 vaccine regimens
7 assays
5 contributors
1,715 viruses
31 unique antibodies
2 of your saved groups

CURRENT SELECTION

● Range: x = .7 to 1, y = .65 to 1

keep overlap | keep all | exclude | save

ACTIVE FILTERS

Binding & neutralization (434)

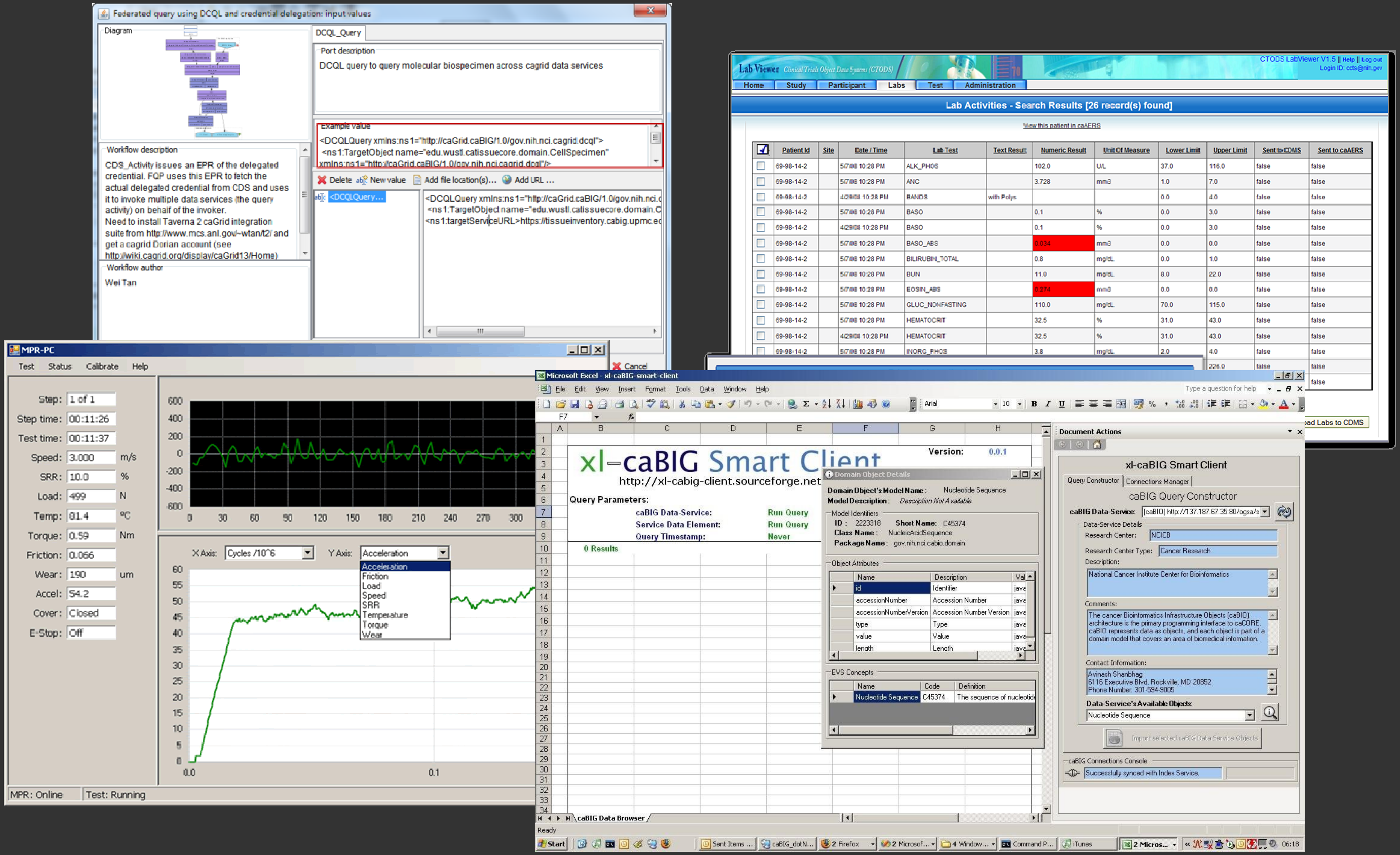
save view | clear

REFERENCE GROUPS

CHAVI Broad Neutralizers

□ Only show overlap with active filter

+ add a reference group



Port description

DCQL query to query molecular biospecimen across cagrid data services

Example value

```
<DCQLQuery xmlns:ns1="http://caGrid.caBIG/1.0/gov.nih.nci.cagrid.dcql">
 <ns1:TargetObject name="edu.wustl.catissuecore.domain.CellSpecimen"
xmlns:ns1="http://caGrid.caBIG/1.0/gov.nih.nci.cagrid.dcql"/>
```

<DCQLQuery...

```
<DCQLQuery xmlns:ns1="http://caGrid.caBIG/1.0/gov.nih.nci.c
 <ns1:TargetObject name="edu.wustl.catissuecore.domain.C
<ns1:targetServiceURL>https://tissueinventory.cabig.upmc.ed
```

Workflow description

CDS_Activity issues an EPR of the delegated credential. FQP uses this EPR to fetch the actual delegated credential from CDS and uses it to invoke multiple data services (the query activity) on behalf of the invoker.
Need to install Taverna 2 caGrid integration suite from http://www.mcs.anl.gov/~wtan/t2/ and get a cagrid Dorian account (see http://wiki.cagrid.org/display/caGrid13/Home)

Workflow author

Wei Tan

Lab Viewer Clinical Trials Object Data Systems (CTODS)
CTODS LabViewer V1.5 || Help || Log out
Login ID: ctds@nih.gov
Home
Study
Participant
Labs
Test
Administration
View this patient in caAERS
...ad Labs to CDMS

Lab Activities - Search Results [26 record(s) found]

☑	Patient Id	Site	Date / Time	Lab Test	Text Result	Numeric Result	Unit Of Measure	Lower Limit	Upper Limit	Sent to CDMS	Sent to caAERS
□	69-98-14-2		5/7/08 10:28 PM	ALK_PHOS		102.0	U/L	37.0	116.0	false	false
□	69-98-14-2		5/7/08 10:28 PM	ANC		3.728	mm3	1.0	7.0	false	false
□	69-98-14-2		4/29/08 10:28 PM	BANDS	with Polys			0.0	4.0	false	false
□	69-98-14-2		5/7/08 10:28 PM	BASO		0.1	%	0.0	3.0	false	false
□	69-98-14-2		4/29/08 10:28 PM	BASO		0.1	%	0.0	3.0	false	false
□	69-98-14-2		5/7/08 10:28 PM	BASO_ABS		0.034	mm3	0.0	0.0	false	false
□	69-98-14-2		5/7/08 10:28 PM	BILIRUBIN_TOTAL		0.8	mg/dL	0.0	1.0	false	false
□	69-98-14-2		5/7/08 10:28 PM	BUN		11.0	mg/dL	8.0	22.0	false	false
□	69-98-14-2		5/7/08 10:28 PM	EOSIN_ABS		0.274	mm3	0.0	0.0	false	false
□	69-98-14-2		5/7/08 10:28 PM	GLUC_NONFASTING		110.0	mg/dL	70.0	115.0	false	false
□	69-98-14-2		5/7/08 10:28 PM	HEMATOCRIT		32.5	%	31.0	43.0	false	false
□	69-98-14-2		4/29/08 10:28 PM	HEMATOCRIT		32.5	%	31.0	43.0	false	false
□	69-98-14-2		5/7/08 10:28 PM	INORG_PHOS		3.8	mg/dL	2.0	4.0	false	false
									226.0	false	false
											false

Step:	1 of 1	
Step time:	00:11:26	
Test time:	00:11:37	
Speed:	3.000	m/s
SRR:	10.0	%
Load:	499	N
Temp:	81.4	ºC
Torque:	0.59	Nm
Friction:	0.066	
Wear:	190	um
Accel:	54.2	
Cover:	Closed	
E-Stop:	Off	

MPR: Online | Test: Running

xl–caBIG Smart Client

http://xl-cabig-client.sourceforge.net

Version: 0.0.1

Query Parameters:

caBIG Data-Service:	Run Query
Service Data Element:	Run Query
Query Timestamp:	Never

0 Results

Domain Object's Model Name: Nucleotide Sequence
Model Description: *Description Not Available*

Model Identifiers
ID : 2223318 Short Name: C45374
Class Name : NucleicAcidSequence
Package Name : gov.nih.nci.cabio.domain

Object Attributes

Name	Description	Val
id	Identifier	java
accessionNumber	Accession Number	java
accessionNumberVersion	Accession Number Version	java
type	Type	java
value	Value	java
length	Length	java

EVS Concepts

Name	Code	Definition
Nucleotide Sequence	C45374	The sequence of nucleotide

xl-caBIG Smart Client

caBIG Query Constructor

caBIG Data-Service: [caBIO] http://137.187.67.35:80/ogsa/s

Data-Service Details

Research Center: NCICB

Research Center Type: Cancer Research

Description:
National Cancer Institute Center for Bioinformatics

Comments:
The cancer Bioinformatics Infrastructure Objects (caBIO) architecture is the primary programming interface to caCORE. caBIO represents data as objects, and each object is part of a domain model that covers an area of biomedical information.

Contact Information:
Avinash Shanbhag
6116 Executive Blvd, Rockville, MD 20852
Phone Number: 301-594-9005

Data-Service's Available Objects: Nucleotide Sequence

caBIG Connections Console

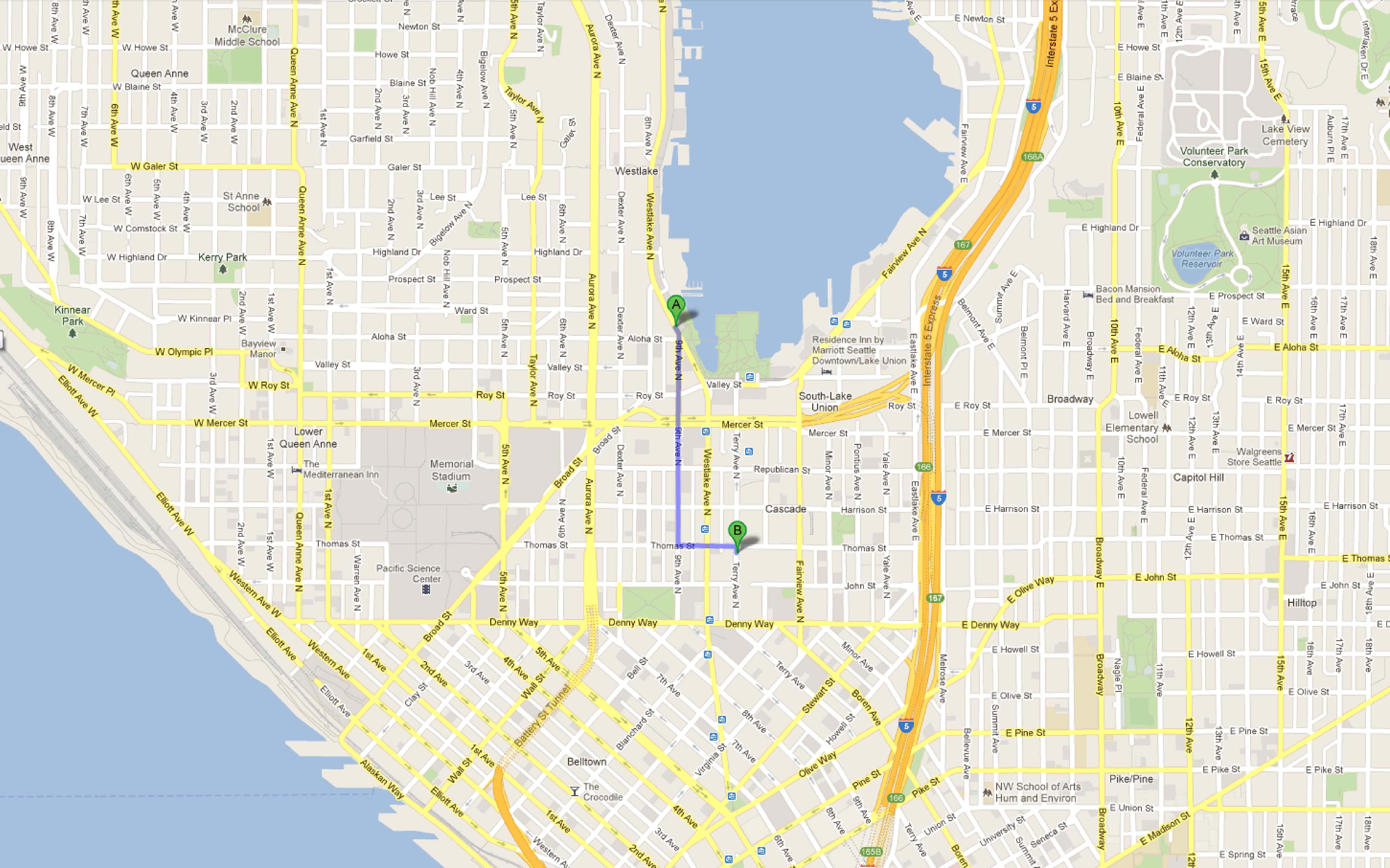
Successfully synced with Index Service.

1. Design is more about making the right thing than how it looks.

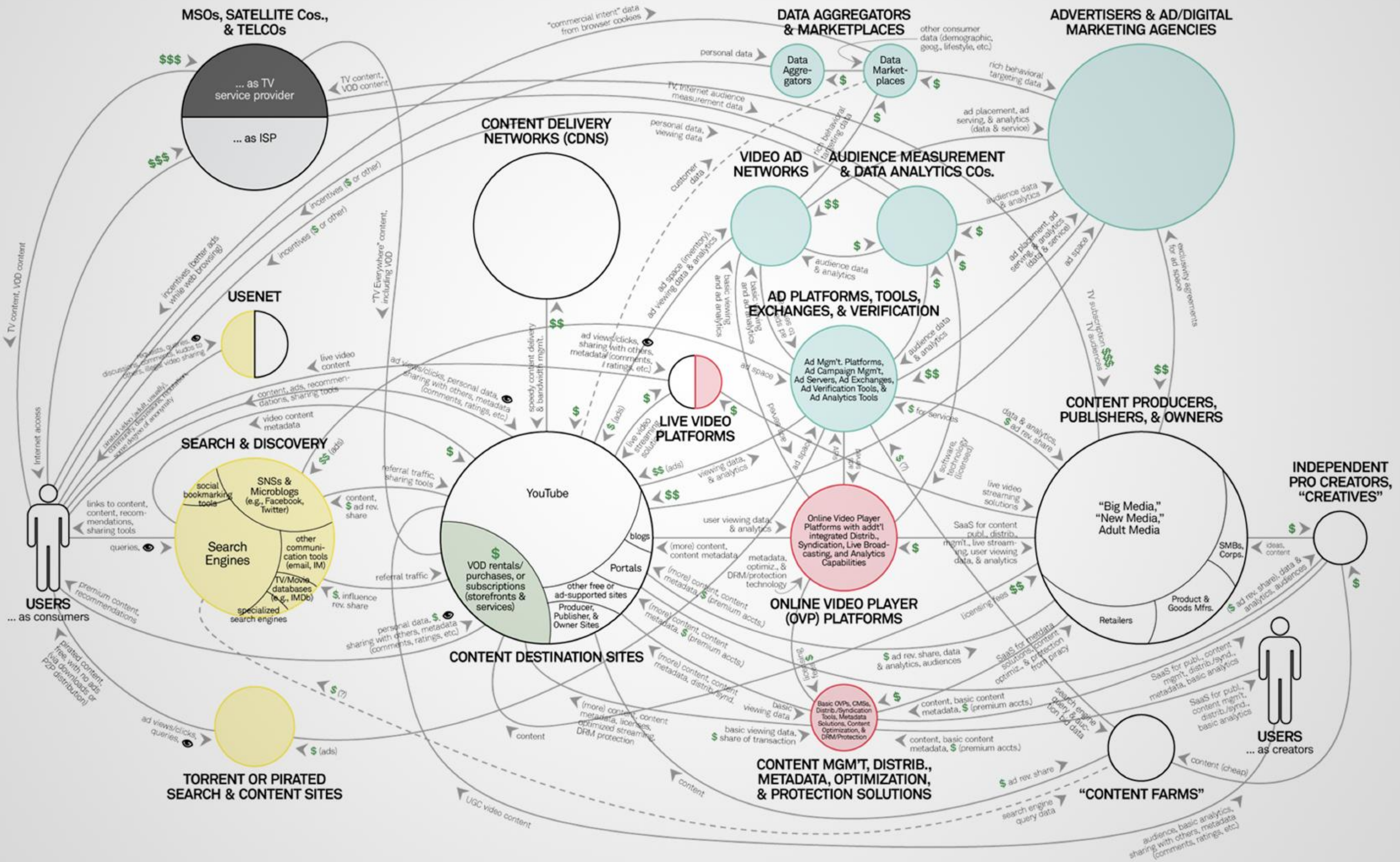
2. The process is available to you.

Which tools, programs, or resources have been successful at helping people engage/achieve goals
Where do people get health advice, support, and information from?
EXIT
CONNECTED
NEW
SOCIAL

A


A
B
McClure Middle School
Queen Anne
West Queen Anne
Kerry Park
St Anne School
Kinnear Park
Bayview Manor
Lower Queen Anne
The Mediterranean Inn
Memorial Stadium
Pacific Science Center
Westlake
South Lake Union
Residence Inn by Marriott Seattle Downtown/Lake Union
Cascade
Belltown
The Crocodile
Broadway
Lowell Elementary School
Capitol Hill
Walgreens Store Seattle
Volunteer Park Conservatory
Lake View Cemetery
Seattle Asian Art Museum
Volunteer Park Reservoir
Bacon Mansion Bed and Breakfast
Hilltop
Pike/Pine
NW School of Arts Hum and Environ
Newton St
Howe St
Blaine St
Garfield St
Galer St
Lee St
Highland Dr
Prospect St
Ward St
Aloha St
Valley St
Roy St
Mercer St
Republican St
Harrison St
Thomas St
John St
Denny Way
W Howe St
W Blaine St
W Galer St
W Lee St
W Comstock St
W Highland Dr
W Kinnear Pl
W Olympic Pl
W Roy St
W Mercer St
W Mercer Pl
Elliott Ave W
Western Ave W
Queen Anne Ave N
Taylor Ave N
Aurora Ave N
Dexter Ave N
Westlake Ave N
9th Ave N
Terry Ave N
Fairview Ave N
Eastlake Ave E
Broad St
Battery St Tunnel
Interstate 5 Express
5
167
166
168A
165B
E Newton St
E Howe St
E Blaine St
E Highland Dr
E Prospect St
E Ward St
E Aloha St
E Roy St
E Mercer St
E Harrison St
E Thomas St
E John St
E Olive Way
E Denny Way
E Howell St
E Olive St
E Pine St
E Pike St
E Union St
E Madison St
E Spring St
10th Ave E
Federal Ave E
15th Ave E
Belmont Ave E
Harvard Ave E
Broadway E
Melrose Ave
Bellevue Ave
Summit Ave
Nagle Pl
12th Ave
Boren Ave
Minor Ave
Stewart St
Olive Way
Pine St
Pike St
Howell St
Virginia St
Blanchard St
Bell St
Wall St
Clay St
1st Ave
2nd Ave
3rd Ave
4th Ave
5th Ave
6th Ave
7th Ave
8th Ave
9th Ave
Western Ave
Elliott Ave
Alaskan Way

Chat
My mibli
Downloads
Feeds
TV Stuff
moova
Rewards
Services
Log Off
STOP
RE-



MSOs, SATELLITE Cos., & TELCOs

... as TV service provider

... as ISP

CONTENT DELIVERY NETWORKS (CDNS)

DATA AGGREGATORS & MARKETPLACES

Data Aggregators

Data Marketplaces

ADVERTISERS & AD/DIGITAL MARKETING AGENCIES

VIDEO AD NETWORKS

AUDIENCE MEASUREMENT & DATA ANALYTICS COs.

AD PLATFORMS, TOOLS, EXCHANGES, & VERIFICATION

Ad Mgm't. Platforms, Ad Campaign Mgm't, Ad Servers, Ad Exchanges, Ad Verification Tools, & Ad Analytics Tools

USENET

LIVE VIDEO PLATFORMS

SEARCH & DISCOVERY

social bookmarking tools

SNSs & Microblogs (e.g., Facebook, Twitter)

Search Engines

other communication tools (email, IM)

TV/Movie databases (e.g., IMDb)

specialized search engines

USERS
... as consumers

CONTENT DESTINATION SITES

YouTube

blogs

Portals

$ VOD rentals/ purchases, or subscriptions (storefronts & services)

other free or ad-supported sites

Producer, Publisher, & Owner Sites

ONLINE VIDEO PLAYER (OVP) PLATFORMS

Online Video Player Platforms with addt'l integrated Distrib., Syndication, Live Broadcasting, and Analytics Capabilities

CONTENT MGM'T, DISTRIB., METADATA, OPTIMIZATION, & PROTECTION SOLUTIONS

Basic OVPs, CMSs, Distrib./Syndication Tools, Metadata Solutions, Content Optimization, & DRM/Protection

CONTENT PRODUCERS, PUBLISHERS, & OWNERS

"Big Media," "New Media," Adult Media

SMBs, Corps.

Product & Goods Mfrs.

Retailers

INDEPENDENT PRO CREATORS, "CREATIVES"

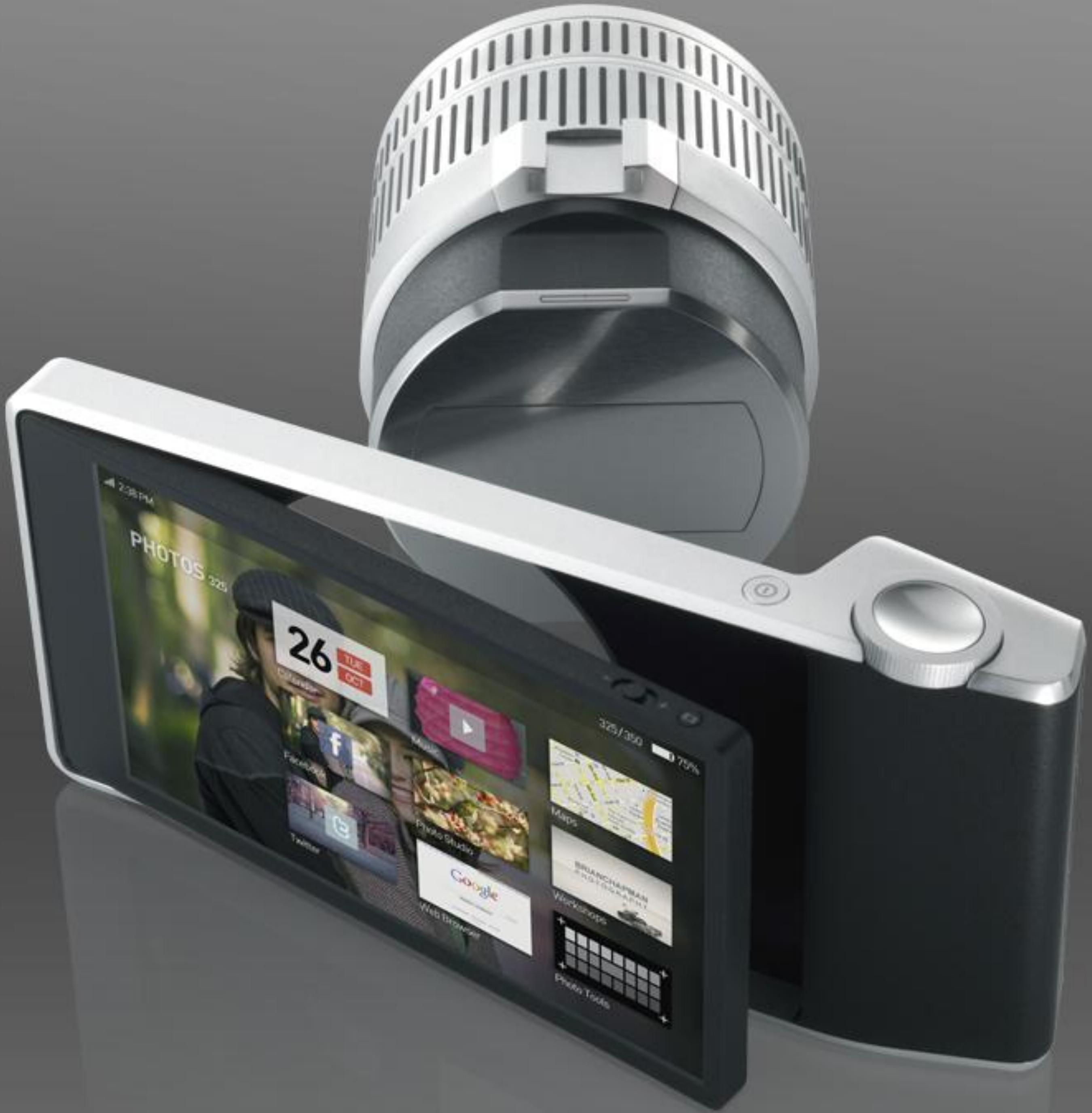
USERS
... as creators

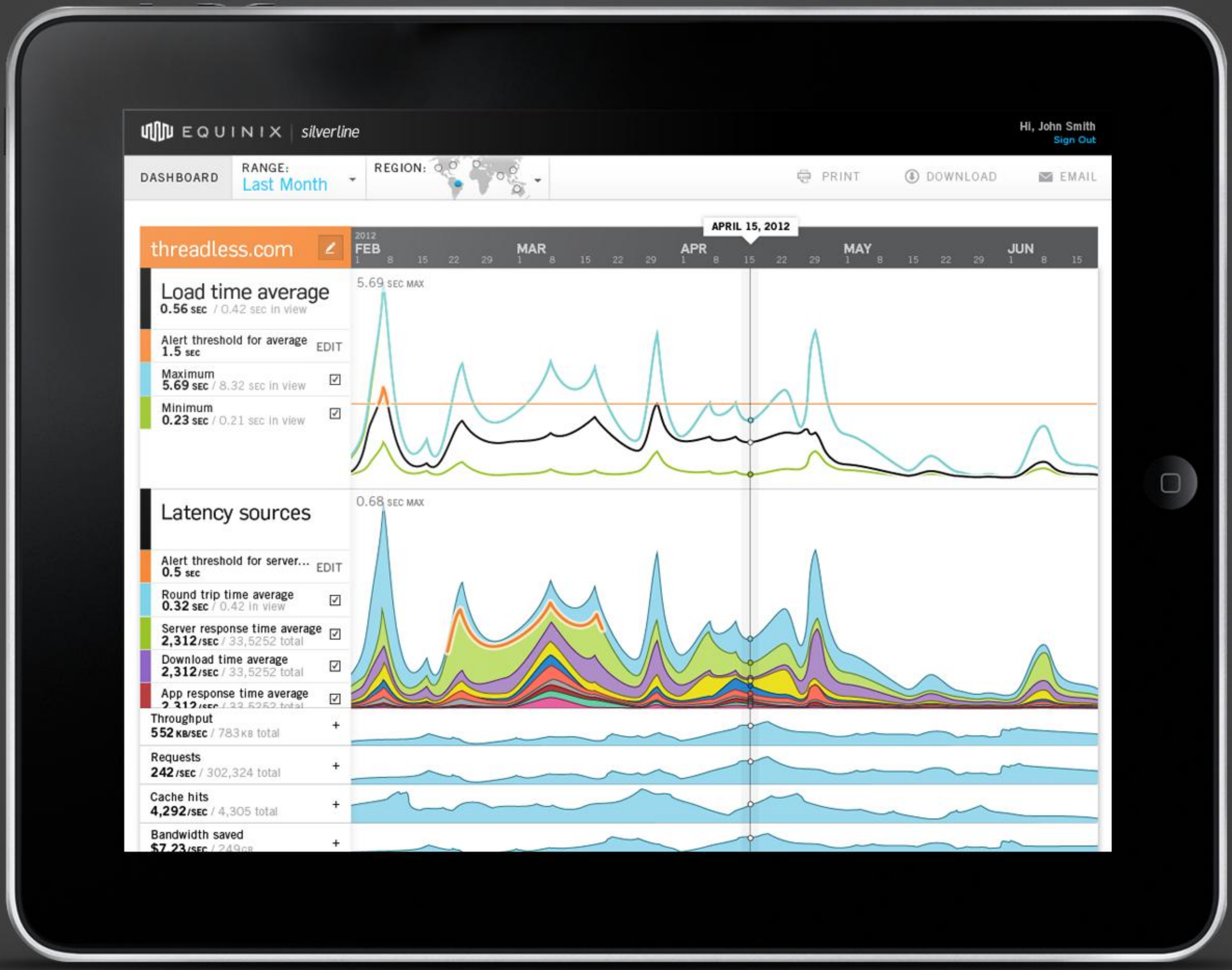
"CONTENT FARMS"

TORRENT OR PIRATED SEARCH & CONTENT SITES

$$$
$$$
TV content, VOD content
TV content, VOD content
"commercial intent" data from browser cookies
personal data
other consumer data (demographic, geog., lifestyle, etc.)
rich behavioral targeting data
$
$
$
TV, internet audience measurement data
personal data, viewing data
ad placement, ad serving, & analytics (data & service)
rich behavioral targeting data
customer data
incentives ($ or other)
incentives ($ or other)
incentives (better ads while web browsing)
"TV Everywhere" content, including VOD
internet access
$$
audience data & analytics
$
$
$$
audience data & analytics
ad space (inventory), ad viewing data & analytics
basic viewing and ad analytics
basic viewing and ad analytics
$
$
ad placement, ad serving, & analytics (data & service)
ad space
exclusivity agreements for ad space
TV subscription $$$, TV audiences
$$
requests, queries, comments, kudos to others, illegal video sharing
discussions
live video content
content, ads, recommendations, sharing tools
video content metadata
pirated video (adult, usually), community, discussions, reputation, some degree of anonymity
ad views/clicks, personal data, sharing with others, metadata (comments, ratings, etc.)
speedy content delivery & bandwidth mgmt.
$$
ad views/clicks, sharing with others, metadata (comments, ratings, etc.)
audience data & analytics
$$
ad space
$ for services
$
$
$ (ads)
live video streaming solutions
$$ (ads)
viewing data, & analytics
$$
$$ (ads)
referral traffic, sharing tools
$
links to content, content, recommendations, sharing tools
queries
content, $ ad rev. share
referral traffic
$, influence rev. share
premium content, recommendations
personal data, $, sharing with others, metadata (comments, ratings, etc.)
pirated content, free with no ads (via downloads or P2P distribution)
$ (?)
ad views/clicks, queries
$ (ads)
UGC video content
content
content
user viewing data, & analytics
(more) content, content metadata
metadata, optimiz., & DRM/protection technology
(more) content, content metadata, $ (premium accts.)
(more) content, content metadata, $ (premium accts.)
(more) content, content metadata, distrib./synd.
(more) content, content metadata, licenses, optimized streaming, DRM protection
basic viewing data
basic viewing data, $ share of transaction
$ (?)
software, technology (licensed)
data & analytics, $ ad rev. share
live video streaming solutions
SaaS for content publ., distrib., mgm't., live streaming, user viewing data, & analytics
$
licensing fees $$
$ ad rev. share, data & analytics, audiences
SaaS for metdata solutions, content optimiz. & protection from piracy
$
content, basic content metadata, $ (premium accts.)
content, basic content metadata, $ (premium accts.)
search engine query & auction bid data
$ ad rev. share
search engine query data
$
ideas, content
$
$ ad rev. share), data & analytics, audiences
SaaS for publ., content mgm't, distrib./synd., metadata, basic analytics
SaaS for publ., content mgm't, distrib./synd., basic analytics
content (cheap)
audience, basic analytics, sharing with others, metadata (comments, ratings, etc.)

WVIL


2:38 PM
PHOTOS 325
26
TUE
OCT
Facebook
Music
Twitter
Photo Studio
Web Browser
325/350
75%
Maps
Workshops
Photo Tools



RANGE: Last Month

threadless.com

Load time average
0.56 SEC / 0.42 SEC in view

Alert threshold for average EDIT
1.5 SEC

Maximum ☑
5.69 SEC / 8.32 SEC in view

Minimum ☑
0.23 SEC / 0.21 SEC in view

Latency sources

Alert threshold for server... EDIT
0.5 SEC

Round trip time average ☑
0.32 SEC / 0.42 in view

Server response time average ☑
2,312 /SEC / 33,5252 total

Download time average ☑
2,312 /SEC / 33,5252 total

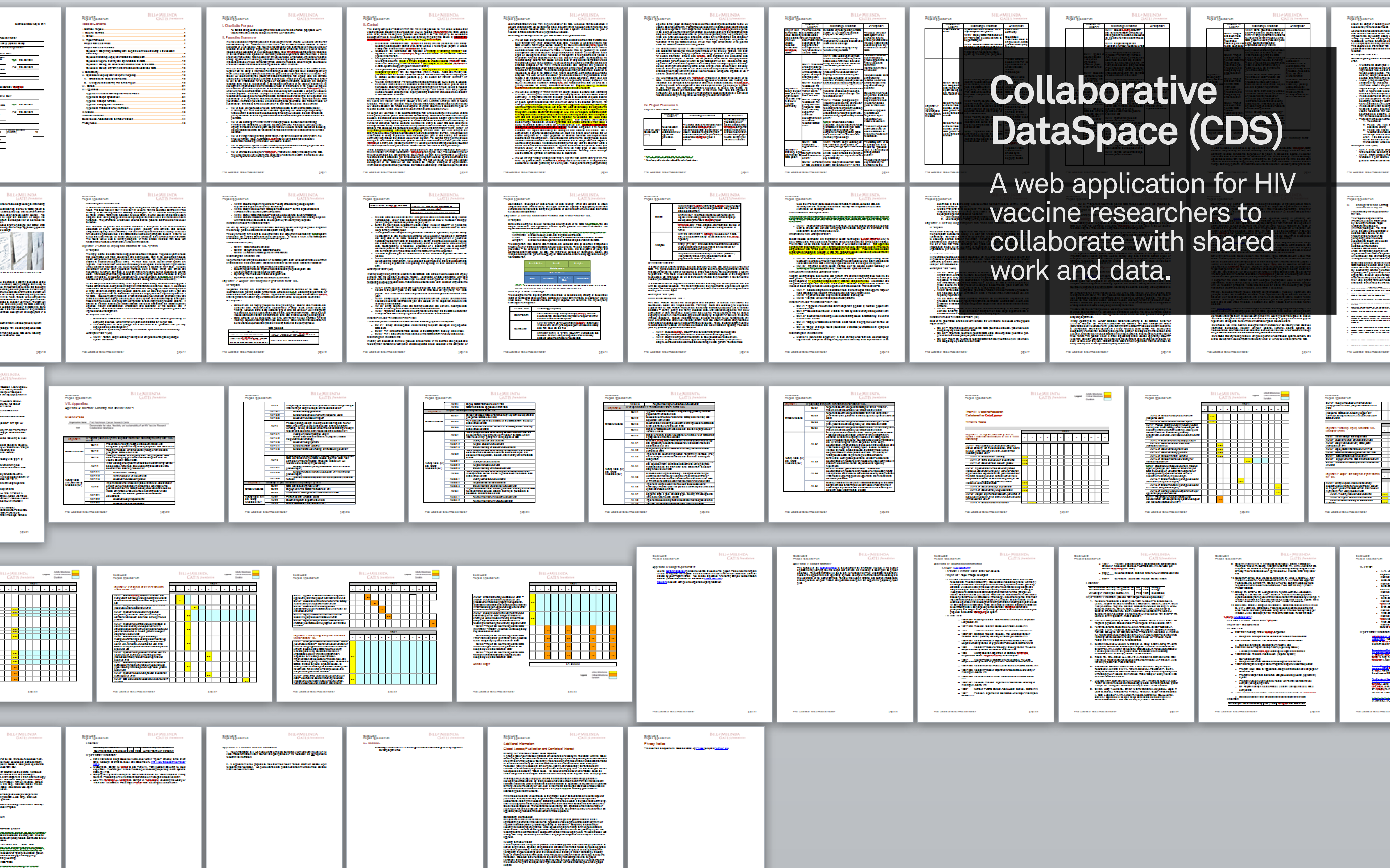
App response time average ☑
2,312 /SEC / 33,5252 total

Throughput +
552 KB/SEC / 783 KB total

Requests +
242 /SEC / 302,324 total

Cache hits +
4,292 /SEC / 4,305 total

Bandwidth saved +
$7.23 /SEC / 249GB

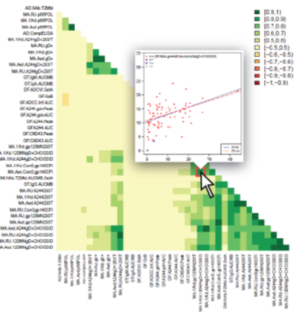


Collaborative DataSpace (CDS)

A web application for HIV vaccine researchers to collaborate with shared work and data.

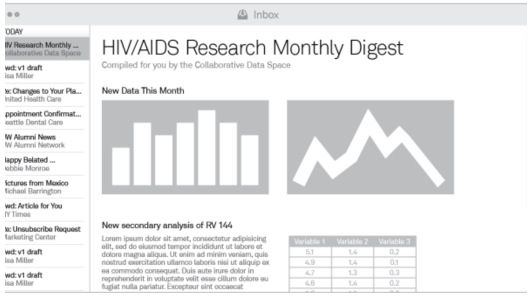
4 Drill Down Data

Supporting data can be previewed a opened in an interactive environmen



Push Communications

Stay Up-to-date About New Data
Periodic newsletter captures an editorial view of interesting updates. Users can subscribe to a search in order to get new results by mail.



ODAY

IV Research Monthly ...
ollaborative Data Space

wd: v1 draft
isa Miller

e: Changes to Your Pla...
nited Health Care

ppointment Confirmat...
eattle Dental Care

W Alumni News
W Alumni Network

appy Belated ...
ebbie Monroe

ictures from Mexico
lichael Barrington

wd: Article for You
Y Times

e: Unsubscribe Request
larketing Center

wd: v1 draft
isa Miller

wd: v1 draft
isa Miller

HIV/AIDS Research Monthly Digest

Compiled for you by the Collaborative Data Space

New Data This Month

New secondary analysis of RV 144

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint occaecat

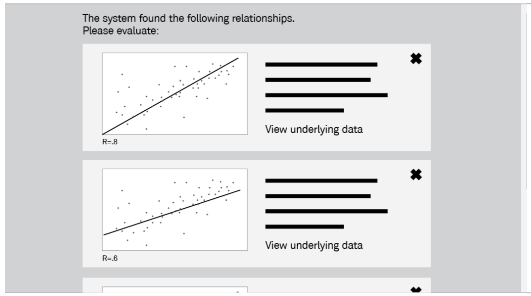
Variable 1	Variable 2	Variable 3
5.1	1.4	0.2
4.9	1.4	0.1
4.7	1.3	0.3
4.6	1.4	0.2

6 Drill Down Data

Supporting data can be previewed ar opened in an interactive environment

Automatically Identified Relationships For Your Data

Shotgun approach
The computer finds correlations and humans assess whether or r there's value.



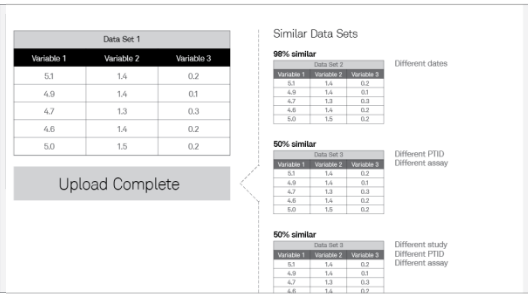
The system found the following relationships.
Please evaluate:

View underlying data

View underlying data

l Compare Data

Find Related Data
View data that is similar to that of o researchers.



Data Set 1		
Variable 1	Variable 2	Variable 3
5.1	1.4	0.2
4.9	1.4	0.1
4.7	1.3	0.3
4.6	1.4	0.2
5.0	1.5	0.2

Upload Complete

Similar Data Sets

98% similar

Data Set 2		
Variable 1	Variable 2	Variable 3
5.1	1.4	0.2
4.9	1.4	0.1
4.7	1.3	0.3
4.6	1.4	0.2
5.0	1.5	0.2

Different dates

50% similar

Data Set 3		
Variable 1	Variable 2	Variable 3
5.1	1.4	0.2
4.9	1.4	0.1
4.7	1.3	0.3
4.6	1.4	0.2
5.0	1.5	0.2

Different PTID
Different assay

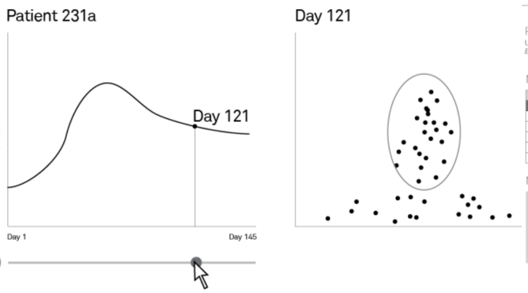
50% similar

Data Set 3		
Variable 1	Variable 2	Variable 3
5.1	1.4	0.2
4.9	1.4	0.1
4.7	1.3	0.3
4.6	1.4	0.2

Different study
Different PTID
Different assay

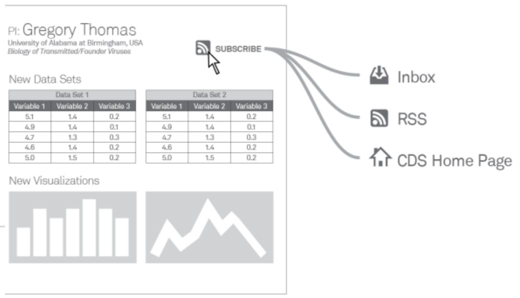
5 Interactive time visualizations

Replay and scrub nested data visualizations



2 Subscriptions

Enable Subscriptions to Researcher
New material gets pushed to users based on opt-in subscriptions.



PI: Gregory Thomas
University of Alabama at Birmingham, USA
Biology of Transmitted/Founder Viruses

New Data Sets

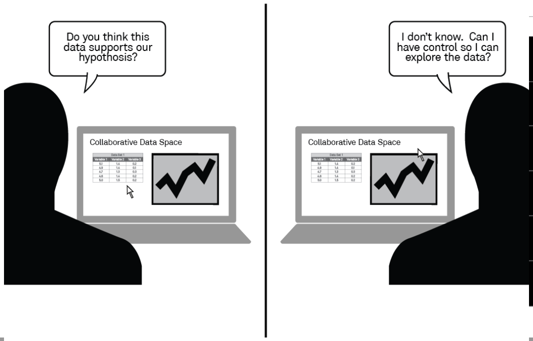
Data Set 1		
Variable 1	Variable 2	Variable 3
5.1	1.4	0.2
4.9	1.4	0.1
4.7	1.3	0.3
4.6	1.4	0.2
5.0	1.5	0.2

Data Set 2		
Variable 1	Variable 2	Variable 3
5.1	1.4	0.2
4.9	1.4	0.1
4.7	1.3	0.3
4.6	1.4	0.2
5.0	1.5	0.2

New Visualizations

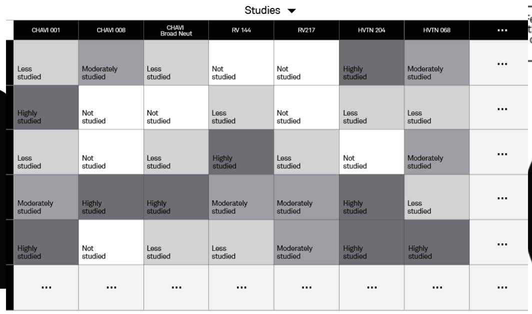
7 Remote syncronous session

Enable a shared session between tw or more parties where the data spac view in syncronized and control can t bassed back and forth


Do you think this data supports our hypothosis?
Collaborative Data Space
I don't know. Can I have control so I can explore the data?
Collaborative Data Space

Find Opportunities

Gap Analysis
Map out activity to reveal potential opportunities in less active areas.

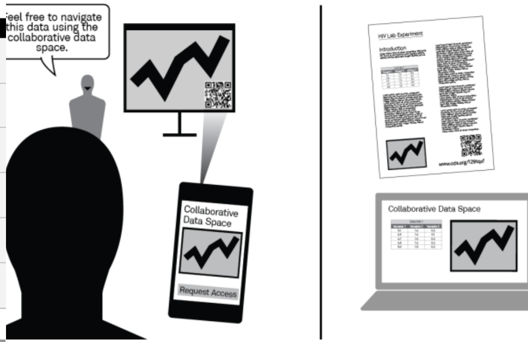


Studies ▾

CHAVI 001	CHAVI 008	CHAVI Broad Neut	RV 144	RV217	HVTN 204	HVTN 068	...
Less studied	Moderately studied	Less studied	Not studied	Not studied	Highly studied	Moderately studied	...
Highly studied	Not studied	Not studied	Less studied	Not studied	Less studied	Less studied	...
Less studied	Not studied	Less studied	Highly studied	Less studied	Not studied	Moderately studied	...
Moderately studied	Highly studied	Highly studied	Moderately studied	Moderately studied	Highly studied	Less studied	...
Highly studied	Not studied	Less studied	Less studied	Moderately studied	Highly studied	Highly studied	...
...	...	...	...	...	...	...	...

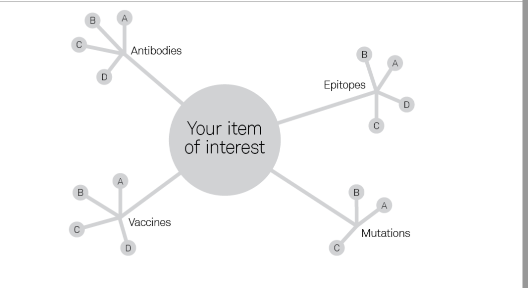
3 Physical Links To Live Data

Enable hyperlinking from presenta- tions and papers to the data space
Using QR codes or hyperlinks, allow read access to your data and finding



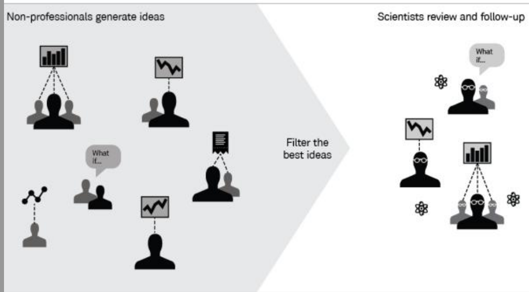
Relationship Map

Visualizations that highlight ways tha items and data are related.



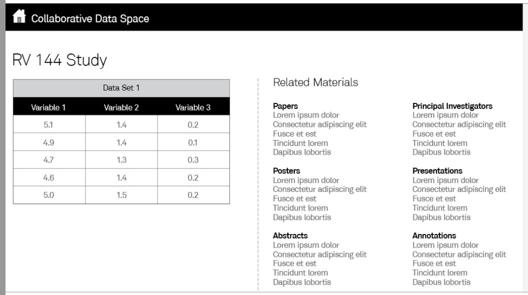
Publicly Generated Analysis

Crowdsource
Open up the data to the wider public of non-professionals so that they car generate ideas worth pursuing.


Non-professionals generate ideas
What if...
Filter the best ideas
Scientists review and follow-up
What if...

Discover Related Content

Content in CDS Leads to Related Materials
Users can easily find content that is related to the content that they view



Collaborative Data Space

RV 144 Study

Data Set 1		
Variable 1	Variable 2	Variable 3
5.1	1.4	0.2
4.9	1.4	0.1
4.7	1.3	0.3
4.6	1.4	0.2
5.0	1.5	0.2

Related Materials

Papers
Lorem ipsum dolor
Consectetur adipiscing elit
Fusce et est
Tincidunt lorem
Dapibus lobortis

Principal Investigators
Lorem ipsum dolor
Consectetur adipiscing elit
Fusce et est
Tincidunt lorem
Dapibus lobortis

Posters
Lorem ipsum dolor
Consectetur adipiscing elit
Fusce et est
Tincidunt lorem
Dapibus lobortis

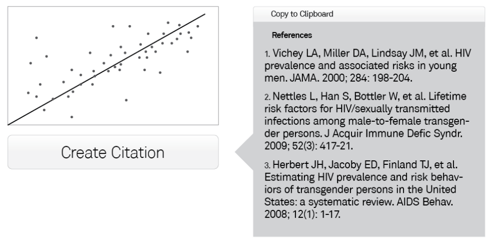
Presentations
Lorem ipsum dolor
Consectetur adipiscing elit
Fusce et est
Tincidunt lorem
Dapibus lobortis

Abstracts
Lorem ipsum dolor
Consectetur adipiscing elit
Fusce et est
Tincidunt lorem
Dapibus lobortis

Annotations
Lorem ipsum dolor
Consectetur adipiscing elit
Fusce et est
Tincidunt lorem
Dapibus lobortis

0 Generate Citations

Citations for Data Sources
Offer the ability to create lists of citations for data from multiple data sources.

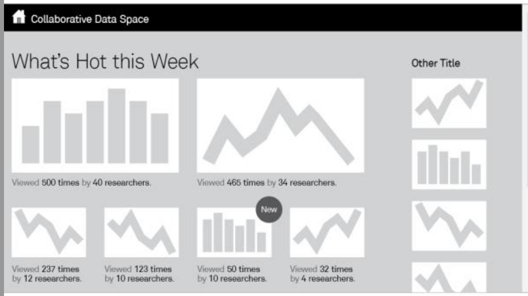


References

1. Vichey LA, Miller DA, Lindsay JM, et al. HIV prevalence and associated risks in young men. JAMA. 2000; 284: 198-204.
2. Nettles L, Han S, Bottler W, et al. Lifetime risk factors for HIV/sexually transmitted infections among male-to-female transgender persons. J Acquir Immune Defic Syndr. 2009; 52(3): 417-21.
3. Herbert JH, Jacoby ED, Finland TJ, et al. Estimating HIV prevalence and risk behaviors of transgender persons in the United States: a systematic review. AIDS Behav. 2008; 12(1): 1-17.

What's Popular

Track popularity
Bubble up the most frequently/recen accessed.



What's Hot this Week

Viewed 500 times by 40 researchers.

Viewed 465 times by 34 researchers.

Viewed 237 times by 12 researchers.

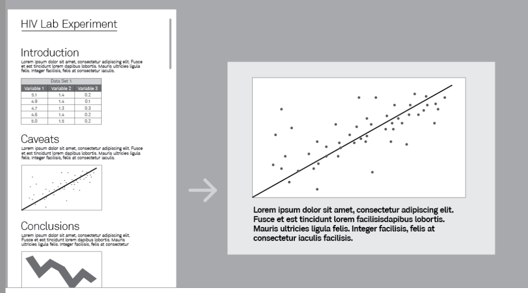
Viewed 123 times by 10 researchers.

Viewed 50 times by 10 researchers.

Viewed 32 times by 4 researchers.

Interactive Annotations

Interpretations of Data and Conclusions
Read data in the form of reports tha describe the reserachers' interpretati and conclusions.



HIV Lab Experiment

Introduction

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Fusce et est tincidunt lorem dapibus lobortis. Mauris ultricies ligula felis. Integer facilisis, felis at consectetur iaculis.

Data Set 1		
Variable 1	Variable 2	Variable 3
5.1	1.4	0.2
4.9	1.4	0.1
4.7	1.3	0.3
4.6	1.4	0.2
5.0	1.5	0.2

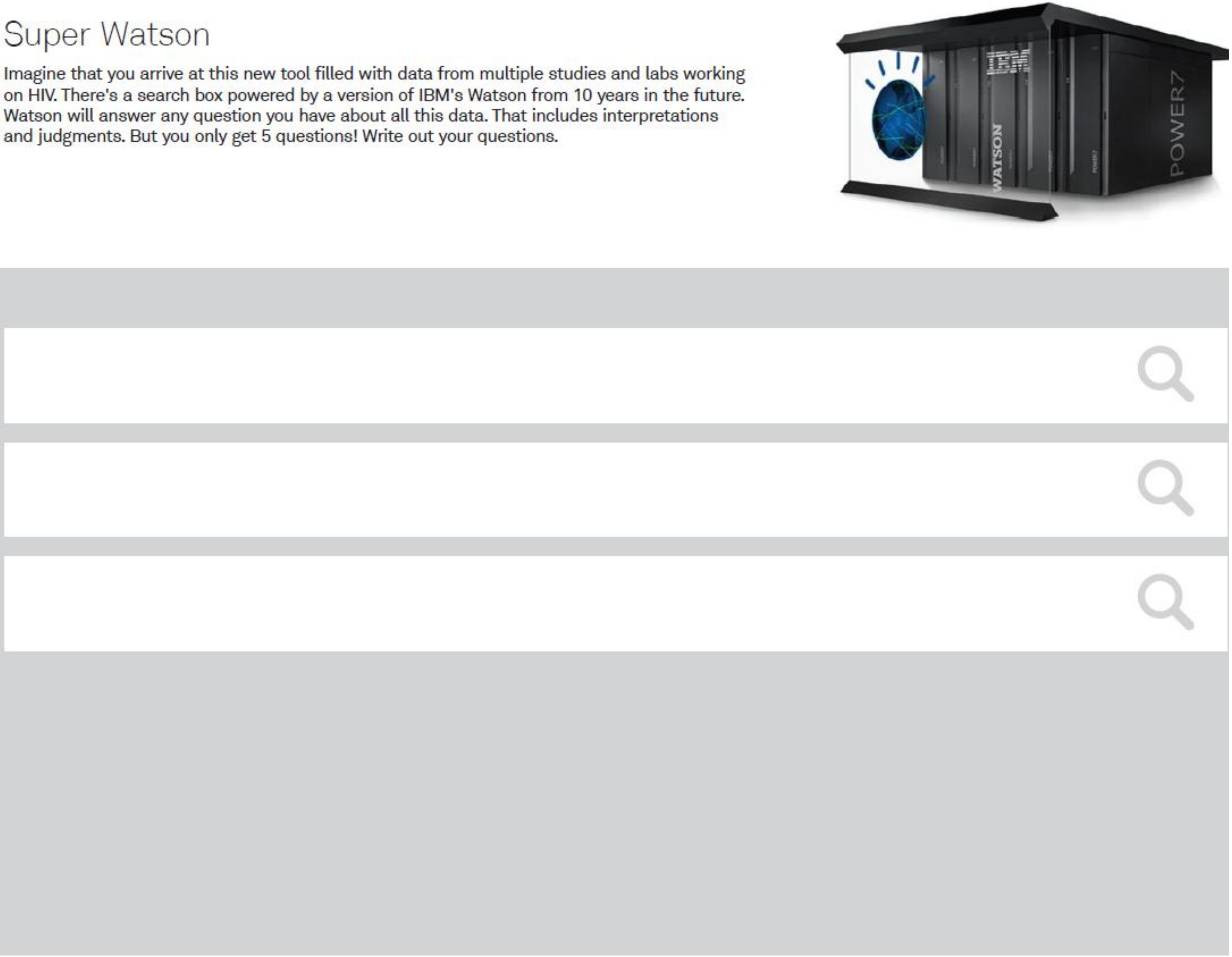
Caveats

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Fusce et est tincidunt lorem dapibus lobortis. Mauris ultricies ligula felis. Integer facilisis, felis at consectetur iaculis.

Conclusions

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Fusce et est tincidunt lorem dapibus lobortis. Mauris ultricies ligula felis. Integer facilisis, felis at consectetur

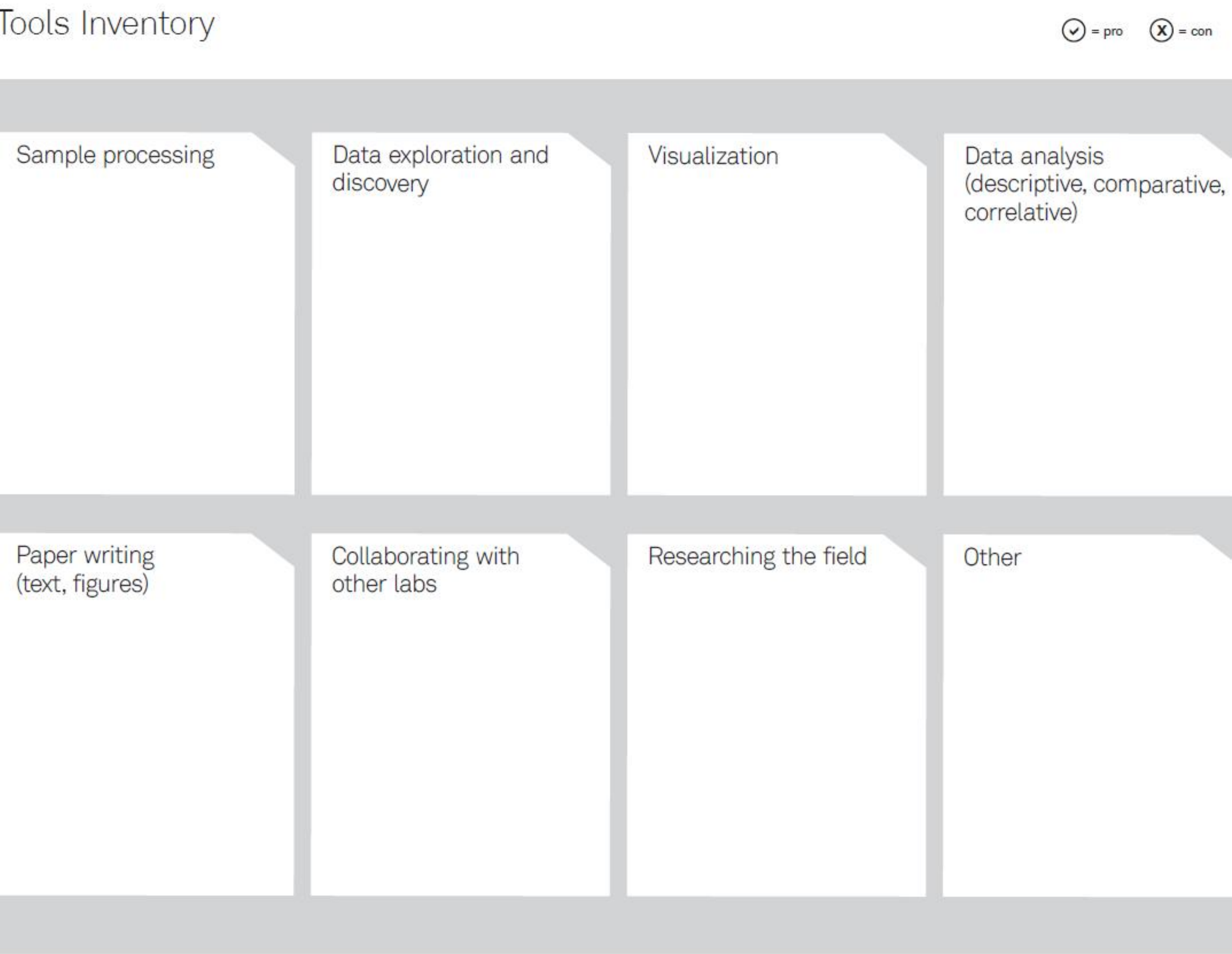
Lorem ipsum dolor sit amet, consectetur adipiscing elit. Fusce et est tincidunt lorem facilisisdapibus lobortis. Mauris ultricies ligula felis. Integer facilisis, felis at consectetur iaculis facilisis.



Super Watson

Imagine that you arrive at this new tool filled with data from multiple studies and labs working on HIV. There's a search box powered by a version of IBM's Watson from 10 years in the future. Watson will answer any question you have about all this data. That includes interpretations and judgments. But you only get 5 questions! Write out your questions.

What could you do with aggregate data that you can't do now?



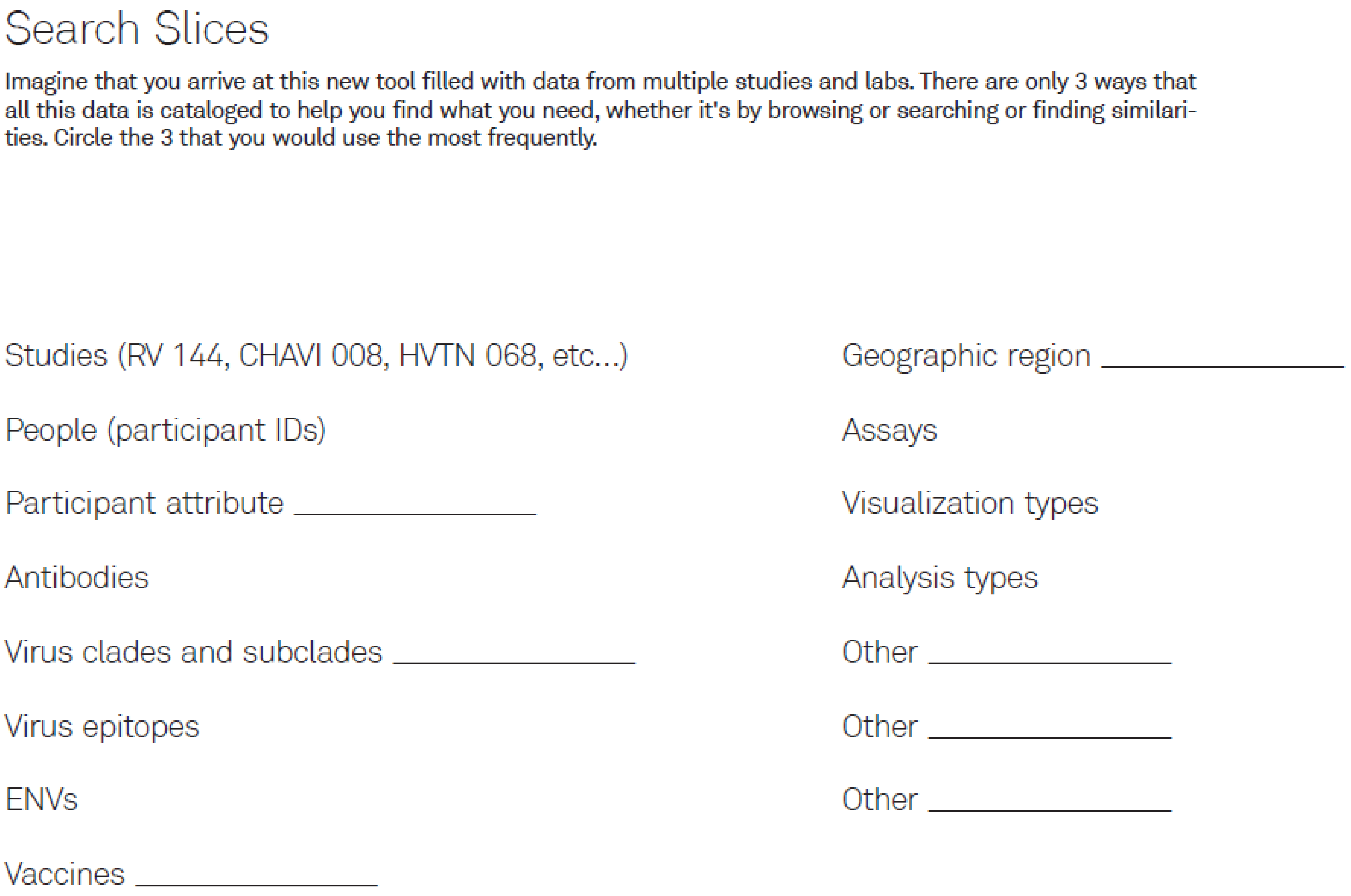
Tools Inventory

✓ = pro ✗ = con

Sample processing	Data exploration and discovery	Visualization	Data analysis (descriptive, comparative, correlative)
Paper writing (text, figures)	Collaborating with other labs	Researching the field	Other

How should this tool relate to all the other tools in use?

How should information be organized?



Search Slices

Imagine that you arrive at this new tool filled with data from multiple studies and labs. There are only 3 ways that all this data is cataloged to help you find what you need, whether it's by browsing or searching or finding similarities. Circle the 3 that you would use the most frequently.

Studies (RV 144, CHAVI 008, HVTN 068, etc...)

People (participant IDs)

Participant attribute ______________

Antibodies

Virus clades and subclades ______________

Virus epitopes

ENVs

Vaccines ______________

Geographic region ______________

Assays

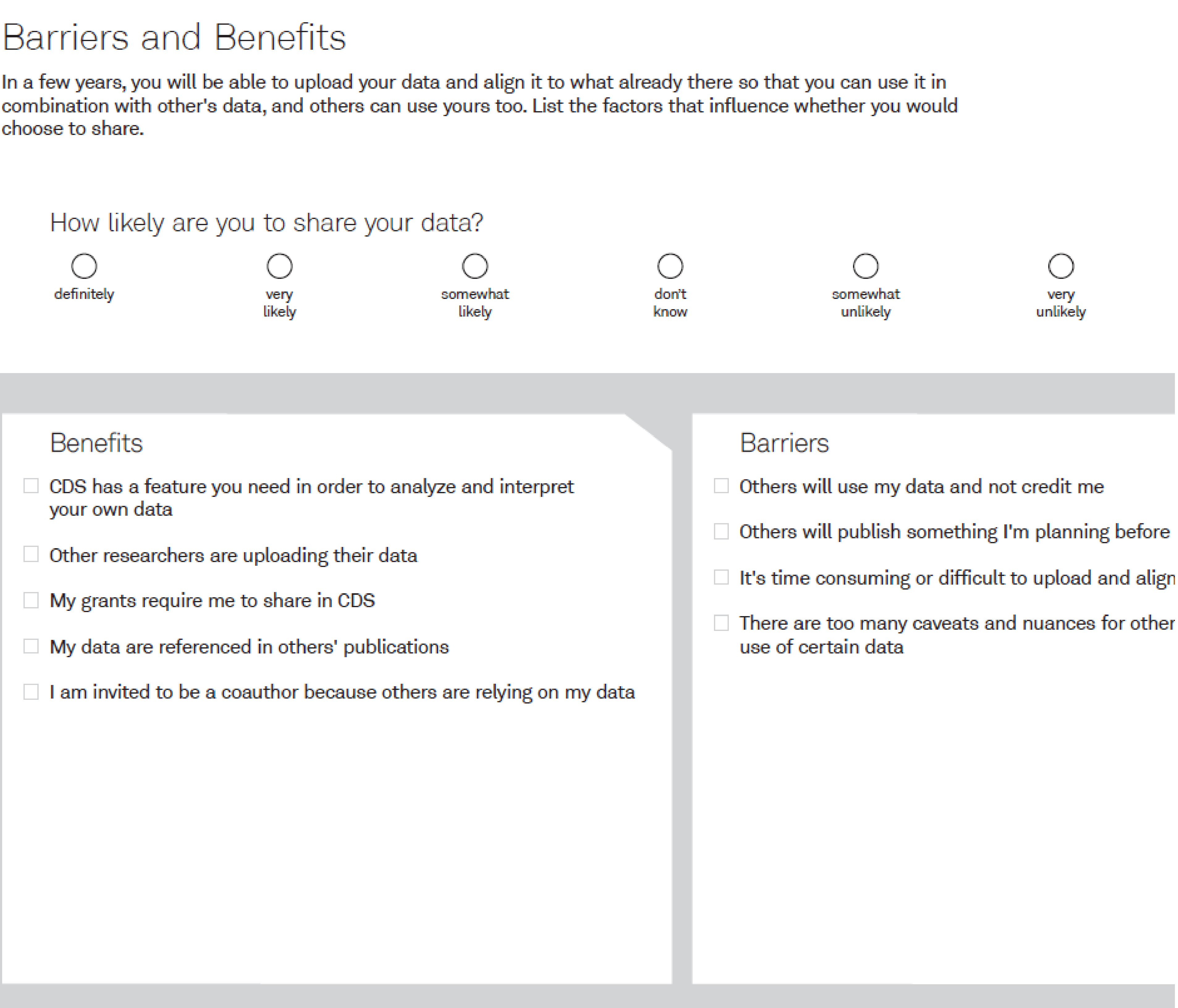
Visualization types

Analysis types

Other ______________

Other ______________

Other ______________



Barriers and Benefits

In a few years, you will be able to upload your data and align it to what already there so that you can use it in combination with other's data, and others can use yours too. List the factors that influence whether you would choose to share.

How likely are you to share your data?

○ definitely ○ very likely ○ somewhat likely ○ don't know ○ somewhat unlikely ○ very unlikely

Benefits

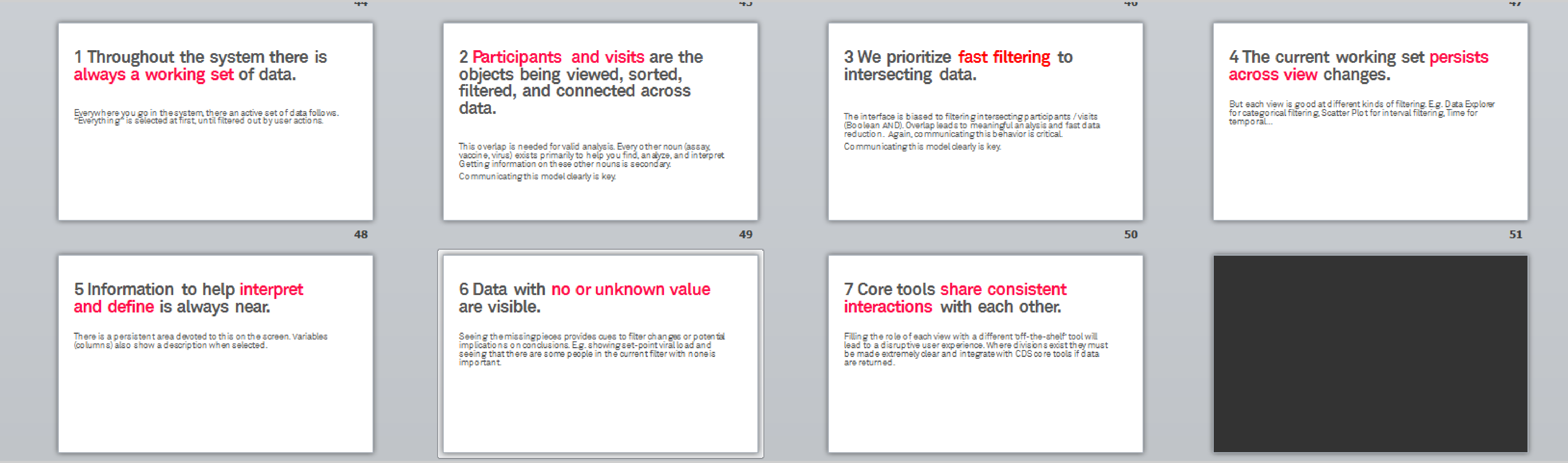
- □ CDS has a feature you need in order to analyze and interpret your own data
- □ Other researchers are uploading their data
- □ My grants require me to share in CDS
- □ My data are referenced in others' publications
- □ I am invited to be a coauthor because others are relying on my data

Barriers

- □ Others will use my data and not credit me
- □ Others will publish something I'm planning before
- □ It's time consuming or difficult to upload and align
- □ There are too many caveats and nuances for other use of certain data

What are specific privacy concerns and how can we overcome them?

Do people really want to collaborate in here? How?



1 Throughout the system there is always a working set of data.

Everywhere you go in the system, there an active set of data follows. "Everything" is selected at first, until filtered out by user actions.

48

2 Participants and visits are the objects being viewed, sorted, filtered, and connected across data.

This overlap is needed for valid analysis. Every other noun (assay, vaccine, virus) exists primarily to help you find, analyze, and interpret. Getting information on these other nouns is secondary.

Communicating this model clearly is key.

49

3 We prioritize fast filtering to intersecting data.

The interface is biased to filtering intersecting participants / visits (Boolean AND). Overlap leads to meaningful analysis and fast data reduction. Again, communicating this behavior is critical.

Communicating this model clearly is key.

50

4 The current working set persists across view changes.

But each view is good at different kinds of filtering. E.g. Data Explorer for categorical filtering, Scatter Plot for interval filtering, Time for temporal...

51

5 Information to help interpret and define is always near.

There is a persistent area devoted to this on the screen. Variables (columns) also show a description when selected.

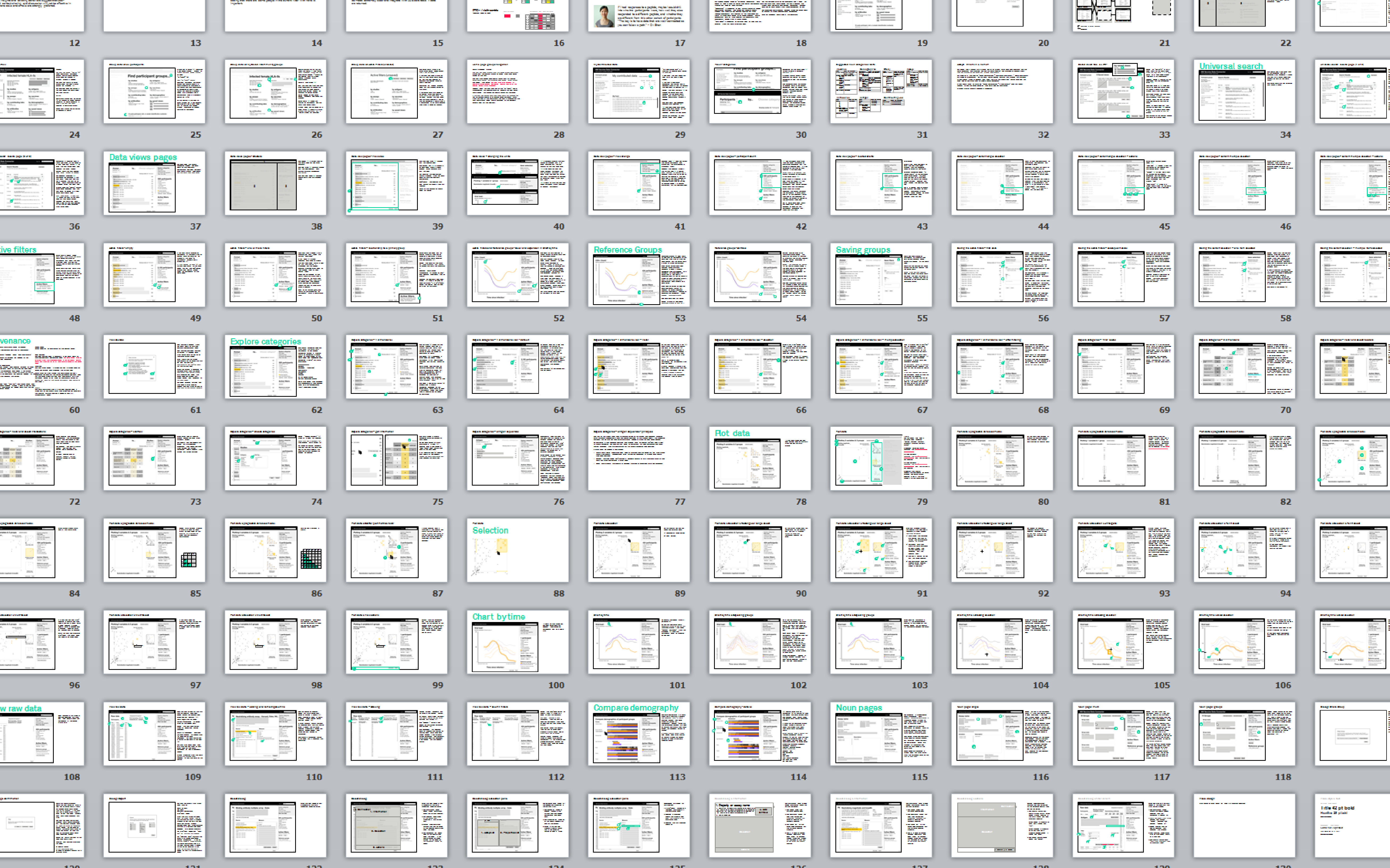
6 Data with no or unknown value are visible.

Seeing the missing pieces provides cues to filter changes or potential implications on conclusions. E.g. showing set-point viral load and seeing that there are some people in the current filter with none is important.

7 Core tools share consistent interactions with each other.

Filling the role of each view with a different 'off-the-shelf' tool will lead to a disruptive user experience. Where divisions exist they must be made extremely clear and integrate with CDS core tools if data are returned.

Can people use this?



12 13 14 15 16 17 18 19 20 21 22

24 25 26 27 28 29 30 31 32 33

Universal search

34

36

Data views pages

37 38 39 40 41 42 43 44 45 46

ive filters

48 49 50 51 52

Reference Groups

53 54

Saving groups

55 56 57 58

venance

60 61

Explore categories

62 63 64 65 66 67 68 69 70

72 73 74 75 76 77

Plot data

78 79 80 81 82

84 85 86 87

Selection

88 89 90 91 92 93 94

96 97 98 99

Chart bytime

100 101 102 103 104 105 106

w raw data

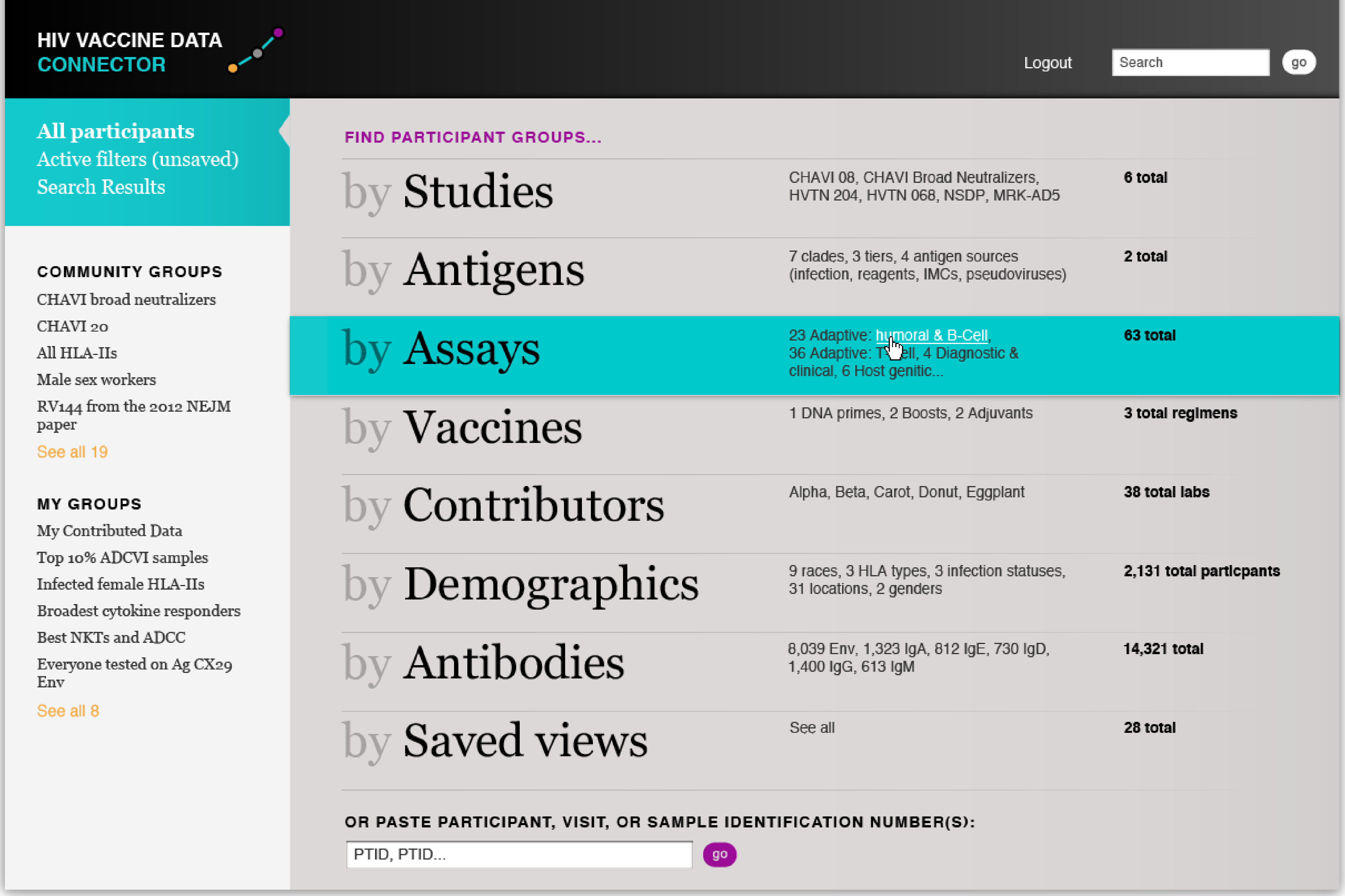
108 109 110 111 112

Compare demography

113 114

Noun pages

115 116 117 118



All participants
Active filters (unsaved)
Search Results

COMMUNITY GROUPS

CHAVI broad neutralizers
CHAVI 20
All HLA-IIs
Male sex workers
RV144 from the 2012 NEJM paper
See all 19

MY GROUPS

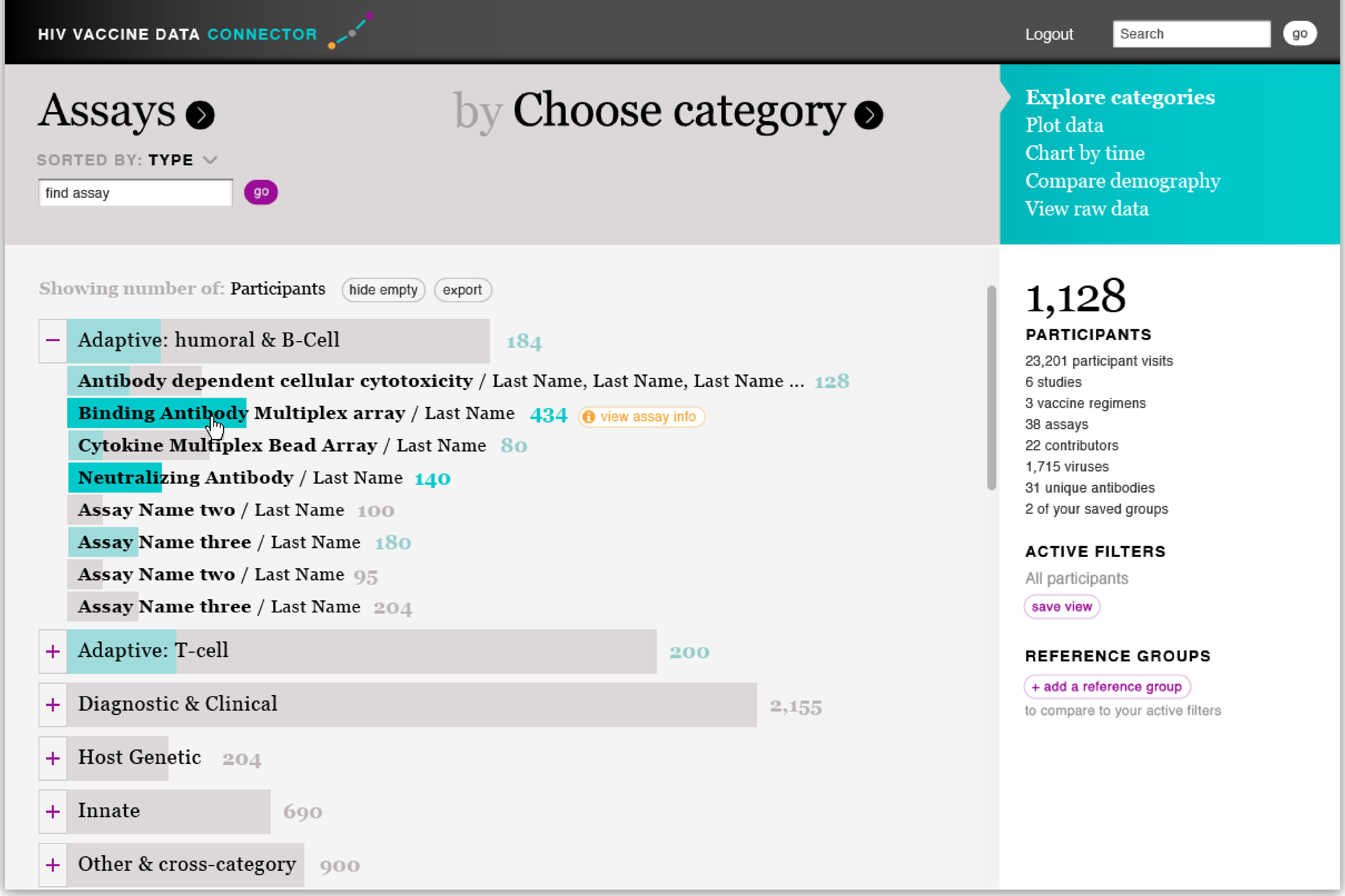
My Contributed Data
Top 10% ADCVI samples
Infected female HLA-IIs
Broadest cytokine responders
Best NKTs and ADCC
Everyone tested on Ag CX29 Env
See all 8

FIND PARTICIPANT GROUPS...

by Studies	CHAVI 08, CHAVI Broad Neutralizers, HVTN 204, HVTN 068, NSDP, MRK-AD5	6 total
by Antigens	7 clades, 3 tiers, 4 antigen sources (infection, reagents, IMCs, pseudoviruses)	2 total
by Assays	23 Adaptive: humoral & B-Cell, 36 Adaptive: T-cell, 4 Diagnostic & clinical, 6 Host genitic...	63 total
by Vaccines	1 DNA primes, 2 Boosts, 2 Adjuvants	3 total regimens
by Contributors	Alpha, Beta, Carot, Donut, Eggplant	38 total labs
by Demographics	9 races, 3 HLA types, 3 infection statuses, 31 locations, 2 genders	2,131 total particpants
by Antibodies	8,039 Env, 1,323 IgA, 812 IgE, 730 IgD, 1,400 IgG, 613 IgM	14,321 total
by Saved views	See all	28 total

OR PASTE PARTICIPANT, VISIT, OR SAMPLE IDENTIFICATION NUMBER(S):

PTID, PTID...



Assays by Choose category

Showing number of: Participants

Assay	Participants
− Adaptive: humoral & B-Cell	184
Antibody dependent cellular cytotoxicity / Last Name, Last Name, Last Name ...	128
Binding Antibody Multiplex array / Last Name	434
Cytokine Multiplex Bead Array / Last Name	80
Neutralizing Antibody / Last Name	140
Assay Name two / Last Name	100
Assay Name three / Last Name	180
Assay Name two / Last Name	95
Assay Name three / Last Name	204
+ Adaptive: T-cell	200
+ Diagnostic & Clinical	2,155
+ Host Genetic	204
+ Innate	690
+ Other & cross-category	900

1,128

PARTICIPANTS

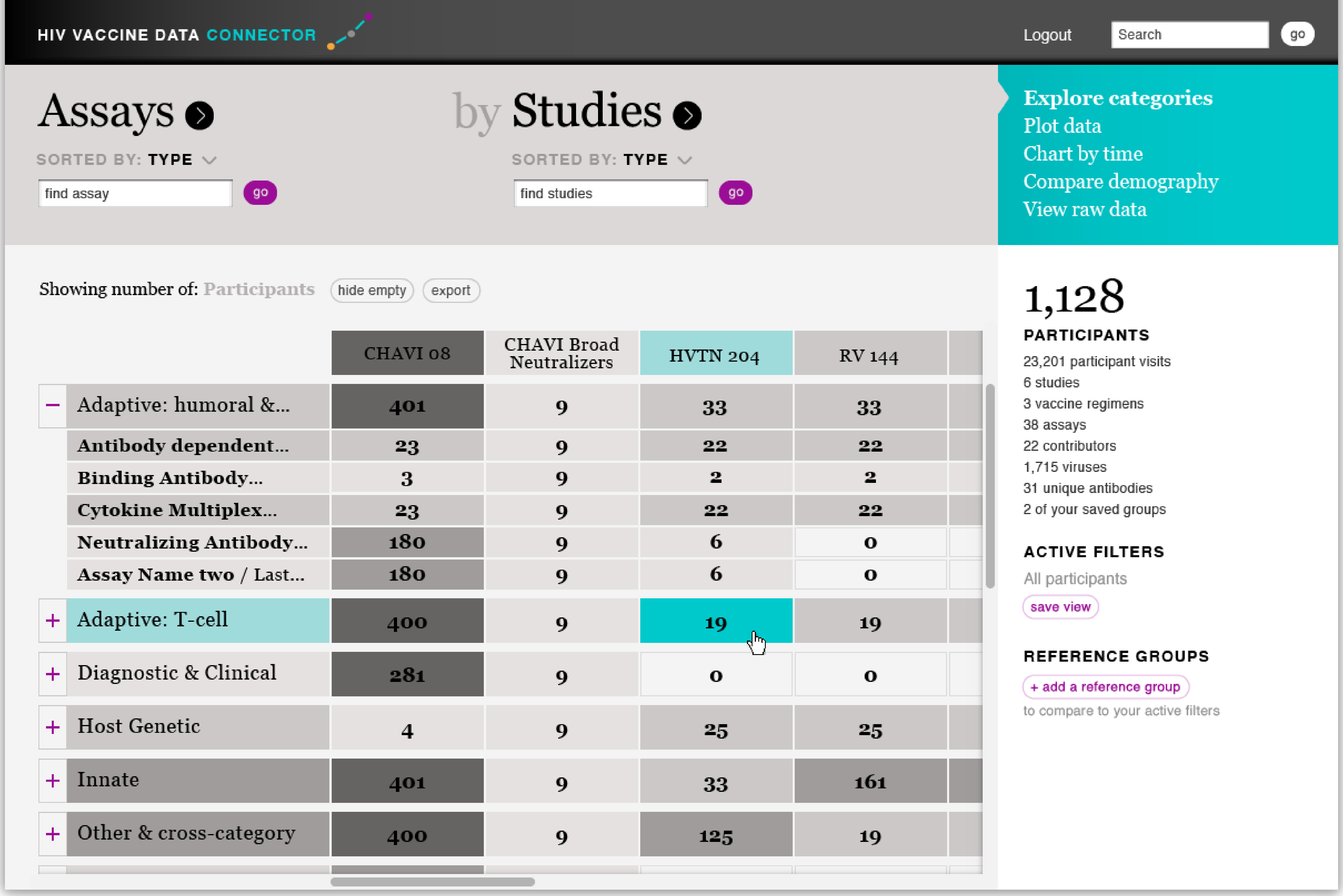
23,201 participant visits
6 studies
3 vaccine regimens
38 assays
22 contributors
1,715 viruses
31 unique antibodies
2 of your saved groups

ACTIVE FILTERS

All participants

REFERENCE GROUPS

to compare to your active filters



HIV VACCINE DATA CONNECTOR

Logout

Search

go

Assays

by Studies

SORTED BY: TYPE

find assay

go

SORTED BY: TYPE

find studies

go

Explore categories

Plot data

Chart by time

Compare demography

View raw data

Showing number of: Participants

hide empty

export

	CHAVI 08	CHAVI Broad Neutralizers	HVTN 204	RV 144
− Adaptive: humoral &...	401	9	33	33
Antibody dependent...	23	9	22	22
Binding Antibody...	3	9	2	2
Cytokine Multiplex...	23	9	22	22
Neutralizing Antibody...	180	9	6	0
Assay Name two / Last...	180	9	6	0
+ Adaptive: T-cell	400	9	19	19
+ Diagnostic & Clinical	281	9	0	0
+ Host Genetic	4	9	25	25
+ Innate	401	9	33	161
+ Other & cross-category	400	9	125	19

1,128

PARTICIPANTS

23,201 participant visits

6 studies

3 vaccine regimens

38 assays

22 contributors

1,715 viruses

31 unique antibodies

2 of your saved groups

ACTIVE FILTERS

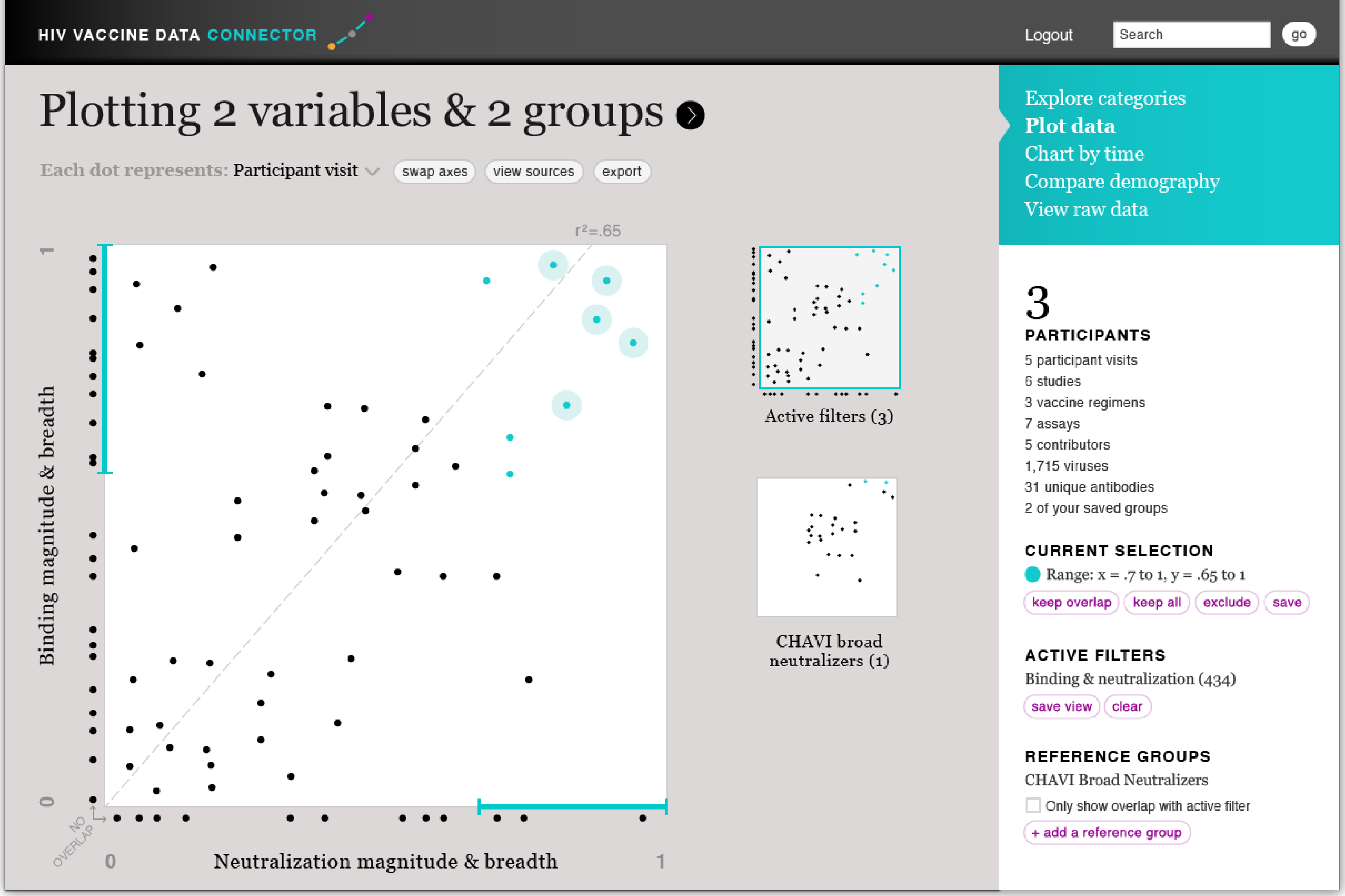
All participants

save view

REFERENCE GROUPS

+ add a reference group

to compare to your active filters



Plotting 2 variables & 2 groups

Each dot represents: Participant visit

3

PARTICIPANTS

5 participant visits
6 studies
3 vaccine regimens
7 assays
5 contributors
1,715 viruses
31 unique antibodies
2 of your saved groups

CURRENT SELECTION

● Range: x = .7 to 1, y = .65 to 1

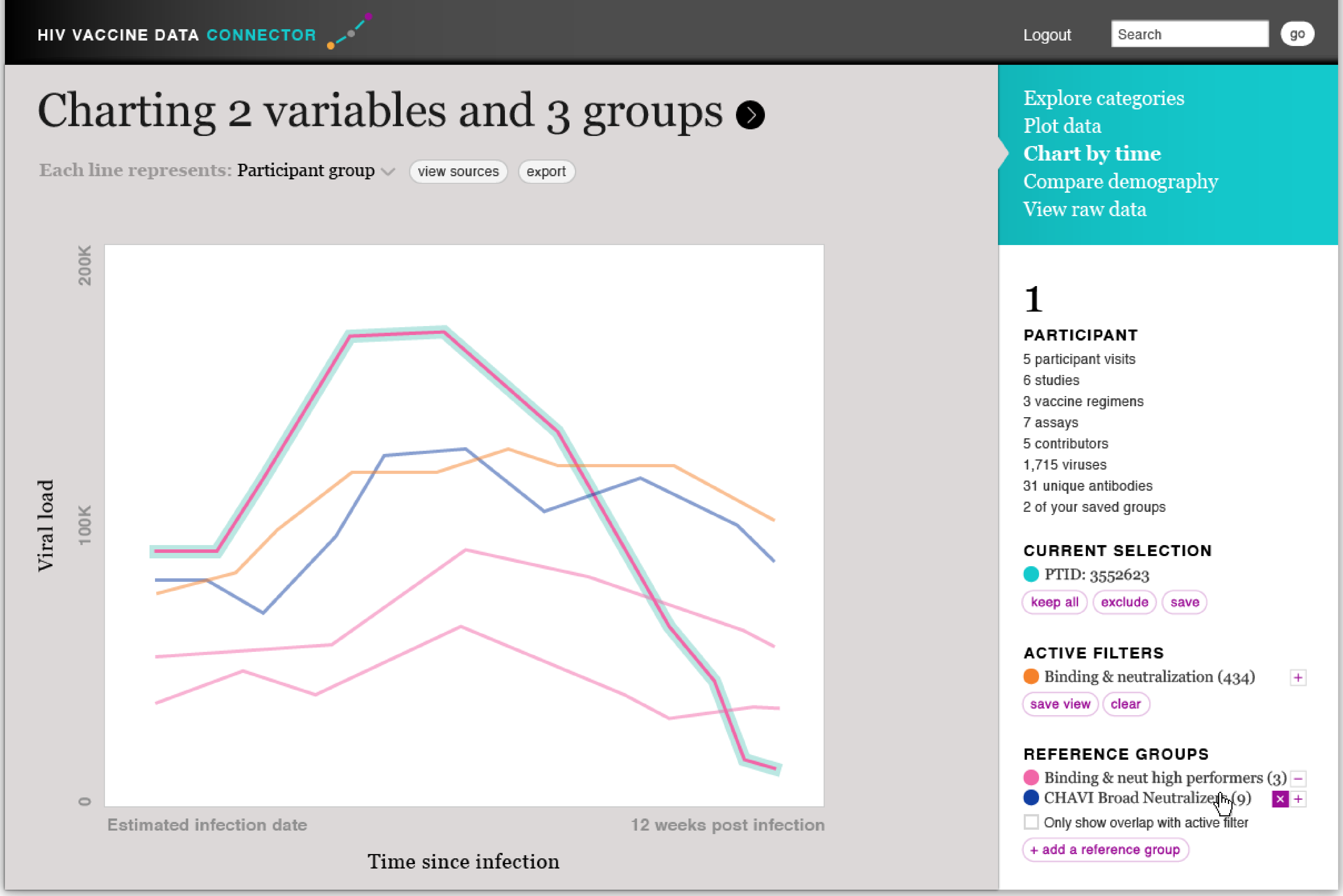
ACTIVE FILTERS

Binding & neutralization (434)

REFERENCE GROUPS

CHAVI Broad Neutralizers

□ Only show overlap with active filter



Charting 2 variables and 3 groups

Each line represents: Participant group

1

PARTICIPANT

5 participant visits
6 studies
3 vaccine regimens
7 assays
5 contributors
1,715 viruses
31 unique antibodies
2 of your saved groups

CURRENT SELECTION

● PTID: 3552623

ACTIVE FILTERS

● Binding & neutralization (434) +

REFERENCE GROUPS

● Binding & neut high performers (3) −
● CHAVI Broad Neutralizers (9) ☒ +

□ Only show overlap with active filter

1. Design is more about making the right thing than how it looks.

What is CDS not?

Dropbox / Atlas: directory and file-based sharing without added value

Completely public to 7 billion people

A specialized Wikipedia

The end of clarifying phone calls and emails

A replacement for statisticians

A replacement for new lab work

A source for HIV research news

“Shotgun science”

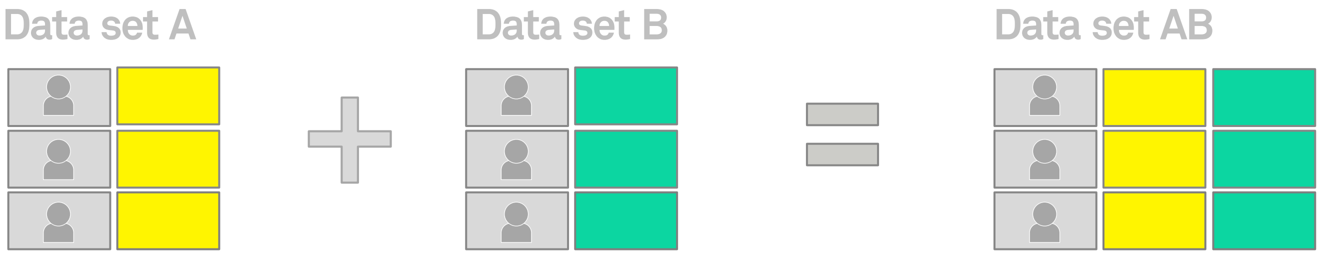
A new interactive paper format

A way to administer and evaluate study execution

Webex – live synchronous collaboration

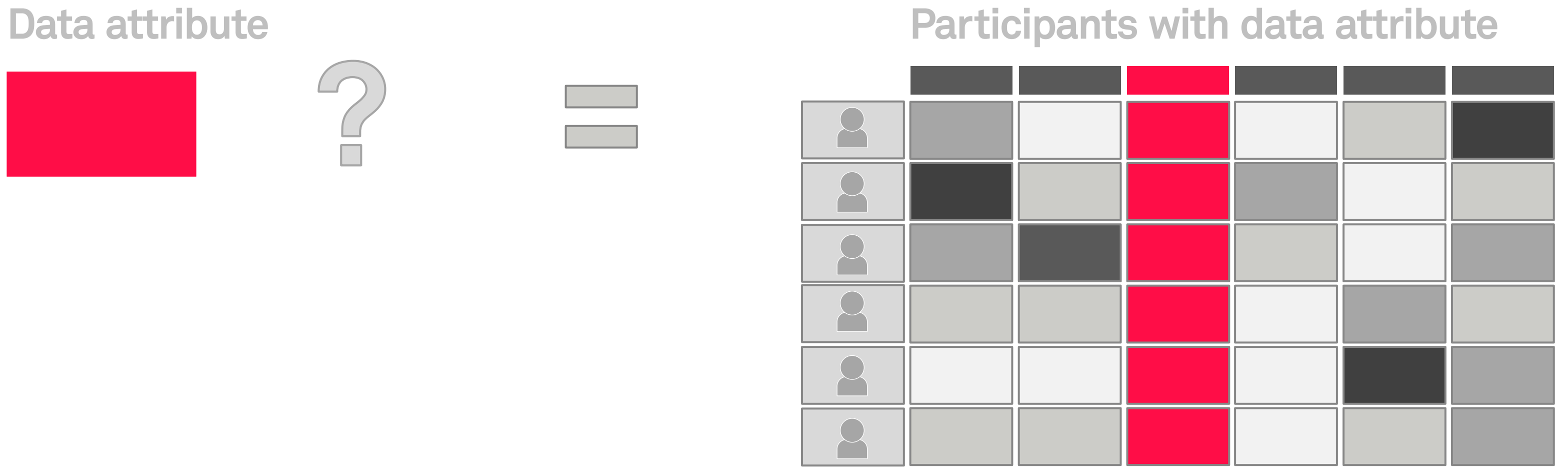
Dataset? Datacube.

Data set-centric



PTID- / visit-centric

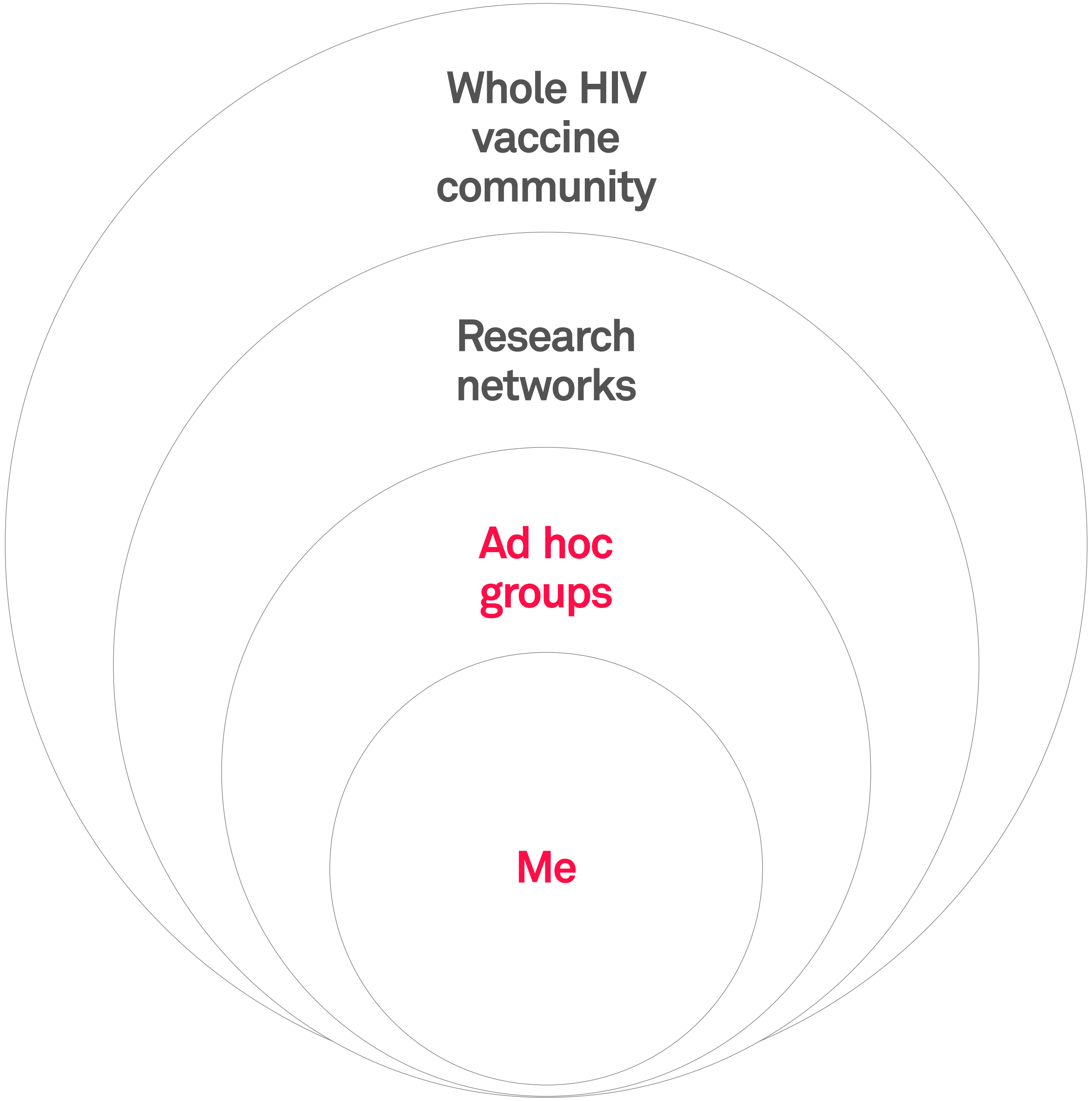
(additional power of CDS)



Open? Mixed.

Private workspaces cannot go away. But fresh public data is critical.

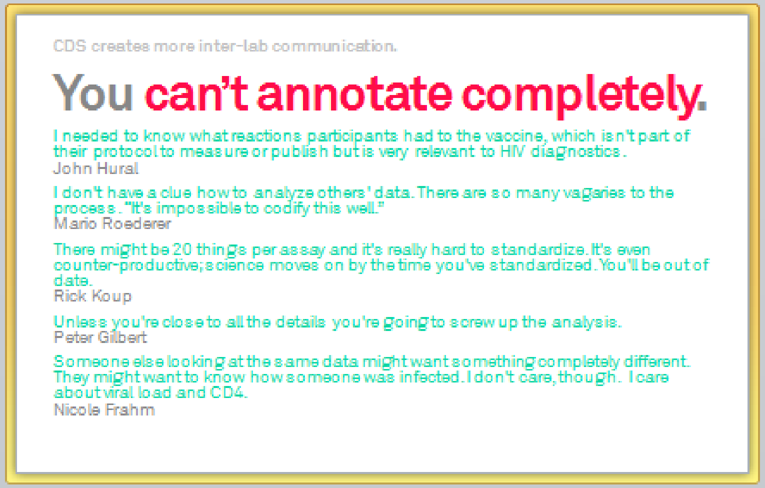
Envisioned levels of data access



Collaborative? Communicative.

Ironically, the Collaborative DataSpace won’t be a place for rich community collaboration after all.*

***But…**



CDS creates more inter-lab communication.

You can't annotate completely.

I needed to know what reactions participants had to the vaccine, which isn't part of their protocol to measure or publish but is very relevant to HIV diagnostics.
John Hural

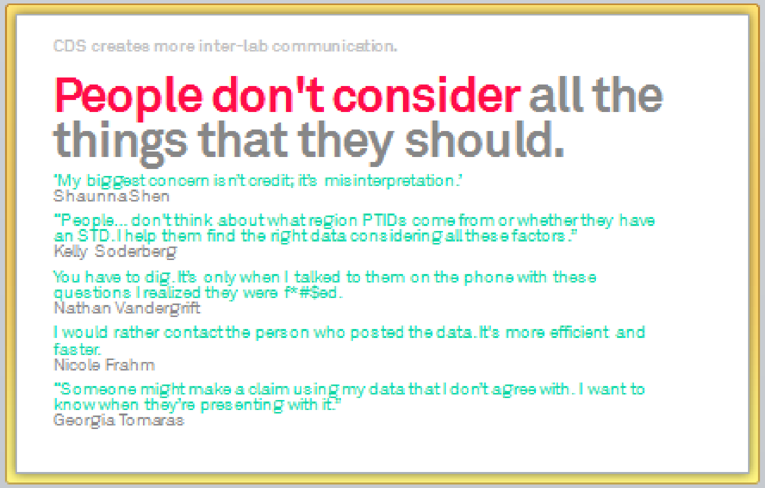
I don't have a clue how to analyze others' data. There are so many vagaries to the process. "It's impossible to codify this well."
Mario Roederer

There might be 20 things per assay and it's really hard to standardize. It's even counter-productive; science moves on by the time you've standardized. You'll be out of date.
Rick Koup

Unless you're close to all the details you're going to screw up the analysis.
Peter Gilbert

Someone else looking at the same data might want something completely different. They might want to know how someone was infected. I don't care, though. I care about viral load and CD4.
Nicole Frahm

39



CDS creates more inter-lab communication.

People don't consider all the things that they should.

'My biggest concern isn't credit; it's misinterpretation.'
Shaunna Shen

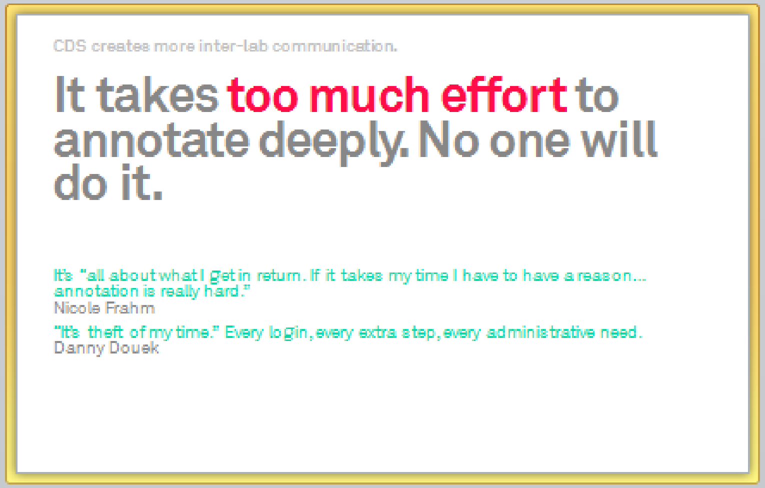
"People... don't think about what region PTIDs come from or whether they have an STD. I help them find the right data considering all these factors."
Kelly Soderberg

You have to dig. It's only when I talked to them on the phone with these questions I realized they were f*#$ed.
Nathan Vandergrift

I would rather contact the person who posted the data. It's more efficient and faster.
Nicole Frahm

"Someone might make a claim using my data that I don't agree with. I want to know when they're presenting with it."
Georgia Tomaras

40



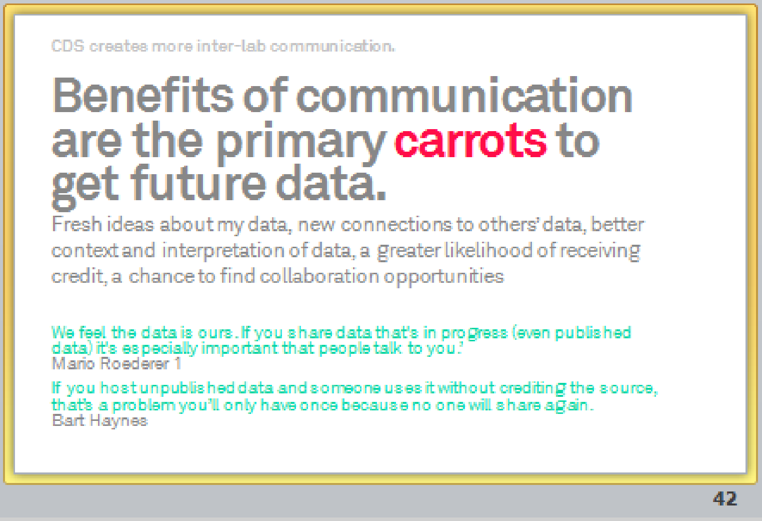
CDS creates more inter-lab communication.

It takes too much effort to annotate deeply. No one will do it.

It's "all about what I get in return. If it takes my time I have to have a reason... annotation is really hard."
Nicole Frahm

"It's theft of my time." Every login, every extra step, every administrative need.
Danny Douek

41



CDS creates more inter-lab communication.

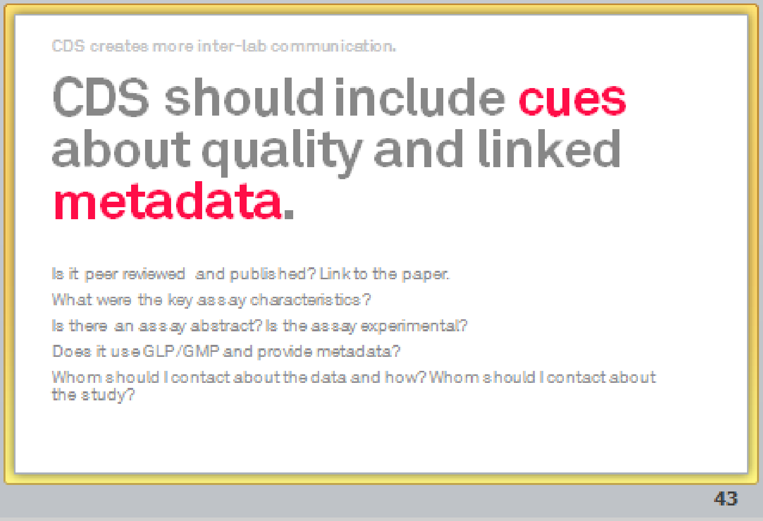
Benefits of communication are the primary carrots to get future data.

Fresh ideas about my data, new connections to others' data, better context and interpretation of data, a greater likelihood of receiving credit, a chance to find collaboration opportunities

We feel the data is ours. If you share data that's in progress (even published data) it's especially important that people talk to you.'
Mario Roederer 1

If you host unpublished data and someone uses it without crediting the source, that's a problem you'll only have once because no one will share again.
Bart Haynes

42



CDS creates more inter-lab communication.

CDS should include cues about quality and linked metadata.

Is it peer reviewed and published? Link to the paper.

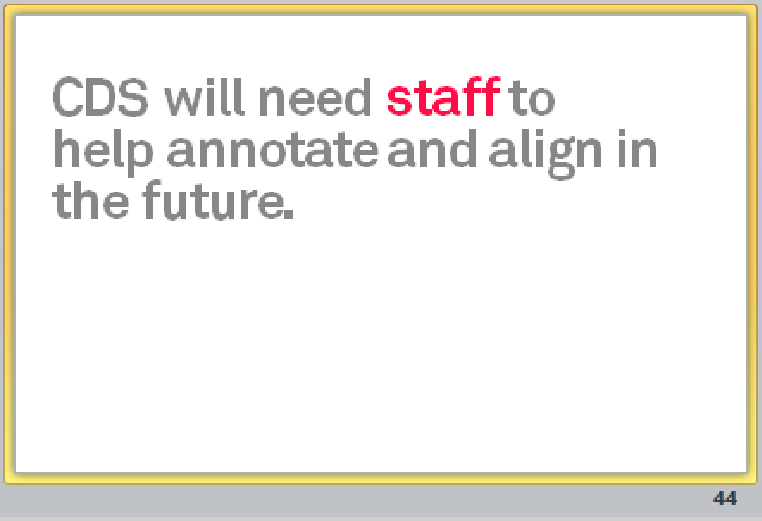
What were the key assay characteristics?

Is there an assay abstract? Is the assay experimental?

Does it use GLP/GMP and provide metadata?

Whom should I contact about the data and how? Whom should I contact about the study?

43



CDS will need staff to help annotate and align in the future.

44

2. The process is available to you.

Look for context: talk and observe before you start building

Validate assumptions: is the explicit ask really what’s needed?

Iterate at low fidelity: Test, fail, learn early and often

Prioritize: optimize for key tasks rather than exposing everything

Dialog