2020 Spring Regulatory Update and Hot Topics in Clinical Research

COVID-19: The Virus, Preparedness in the time of Crisis, and Clinical Research

PANEL 1

11:45am – 12:30pm

Research Data Management and Sharing

CTSI Clinical and Translational Science Institute at Children's National

A PARTNERSHIP WITH THE GEORGE WASHINGTON UNIVERSITY

GU · HU · MHRI · ORNL · WDCVAMC

GHUCCTS

Georgetown-Howard Universities
Center for Clinical and Translational Science

National Institutes of Health
Office of Science Policy

Spring Regulatory Update & Hot Topics in Clinical Research
COVID-19: the Virus, Preparedness in the time of Crisis, and Clinical Research

CHARTING A PATH FORWARD FOR DATA SHARING

Lyric Jorgenson, PhD
Deputy Director, Office of Science Policy
National Institutes of Health
April 21, 2020

Why Share Data?

- **Important for SCIENTISTS**
  - Enables validation of scientific results
  - Allows analyses to be strengthened by combining data
  - Facilitates reuse of hard-to-generate data
  - Accelerates future research

Why Share Data?

- **Important for SCIENTISTS**
  - Enables validation of scientific results
  - Allows analyses to be strengthened by combining data
  - Facilitates reuse of hard-to-generate data
  - Accelerates future research
- **Important for the PUBLIC**
  - Demonstrates stewardship over taxpayer funds
  - Fosters transparency and accountability
  - Maximizes research participants' contributions

NIH's Longstanding Commitment to Data Management and Sharing

- **Policies to make research data and findings more readily accessible**
  - 2003 NIH Data Sharing Policy: Data Management and Sharing Plans for Applications >$500K
  - 2008 NIH Public Access Policy: Publications made available no later than 12 months
  - 2015 NIH Genomic Data Sharing Policy: Large-scale Genomic Data
  - 2007/2016 Policies for Clinical Trials Registration and Summary Results Reporting

- **Shifting the culture to broad data management and sharing**
  - 2015 "NIH Plan" & programs begin incorporating policies
  - 2016 *Proposed* "who, what, when, where, and whys" of data management and sharing
  - 2018 *Proposed* Provisions for Future Data Management and Sharing Policy

Goals for Responsible Data Sharing

- Foster a culture of data stewardship
- Balance data management with sharing need
- Practices consistent with FAIR principles
- Respect research participants' values and consent

Current Proposal

Current Proposal

- Researchers prospectively submit a plan for managing and sharing data

Current Proposal

- **Researchers prospectively submit a plan for managing and sharing data**
- **Policy expectations:**
  - Submission of a Data Management and Sharing Plan
    - Describes how data will be managed, preserved, and shared
    - Outlines how participants' privacy, rights, and confidentiality will be protected and any potential limits to sharing
    - Indicates anticipated timelines for data preservation and access
  - Compliance with the approved Plan

Current Proposal

- **Researchers prospectively submit a plan for managing and sharing data**
- **Policy expectations:**
  - Submission of a Data Management and Sharing Plan
    - Describes how data will be managed, preserved, and shared
    - Outlines how participants' privacy, rights, and confidentiality will be protected and any potential limits to sharing
    - Indicates anticipated timelines for data preservation and access
  - Compliance with the approved Plan
- **Plans may be updated (with approval by NIH)**

Current Proposal

- **Policy is deliberately flexible,** for instance:
  - Accommodates differences in breadth, size, and diversity of data
  - Acknowledges there are considerations that may limit data sharing (e.g., legal, ethical)
  - Allows for various time constraints related to data collection and/or dissemination

Current Proposal

- **Policy is deliberately flexible,** for instance:
  - Accommodates differences in breadth, size, and diversity of data
  - Acknowledges there are considerations that may limit data sharing (e.g., legal, ethical)
  - Allows for various time constraints related to data collection and/or dissemination

NIH DEVELOPING GUIDANCE TO HELP NAVIGATE WHILE RETAINING FLEXIBILITY!

Current Proposal

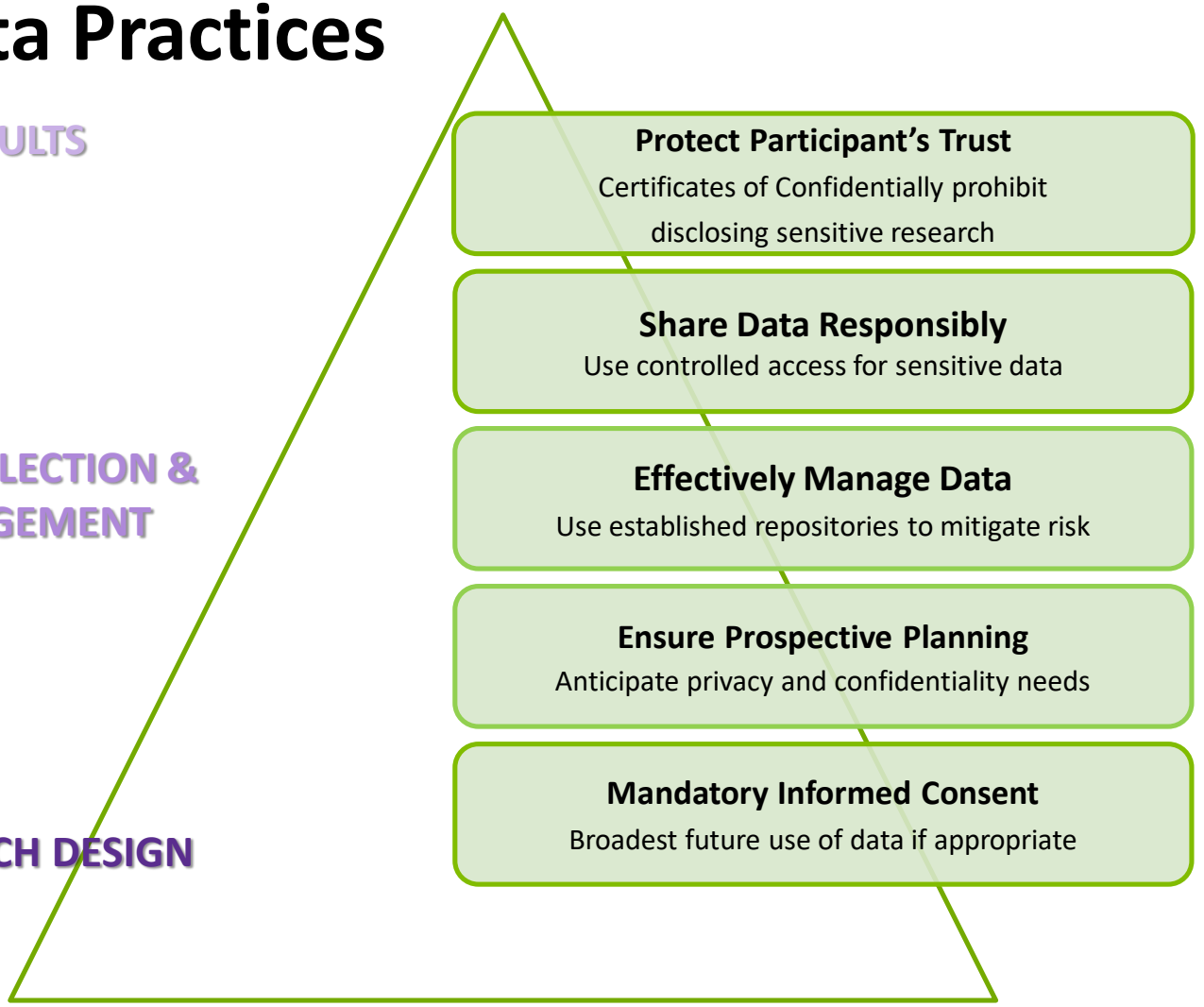
- **Research with Human Participants**
  - Existing protections (e.g., Common Rule, Certificates of Confidentiality) continue to apply
  - NIH is committed to promoting ethical data sharing
    - As with the Genomic Data Sharing Policy, practices such as honoring the consent of participants of primary studies and implementing controlled mechanisms for data access help achieve this goal
  - "NIH encourages the broadest use of scientific data resulting from NIH-funded or conducted research, consistent with privacy, security, informed consent, and proprietary issues." (*Supplemental Draft Guidance, Elements of a Data Management and Sharing Plan*)

Biomedical Research Lifecycle: Prioritizing Ethical Data Practices

RESULTS

DATA COLLECTION & MANAGEMENT

RESEARCH DESIGN



Protect Participant's Trust
Certificates of Confidentially prohibit disclosing sensitive research

Share Data Responsibly
Use controlled access for sensitive data

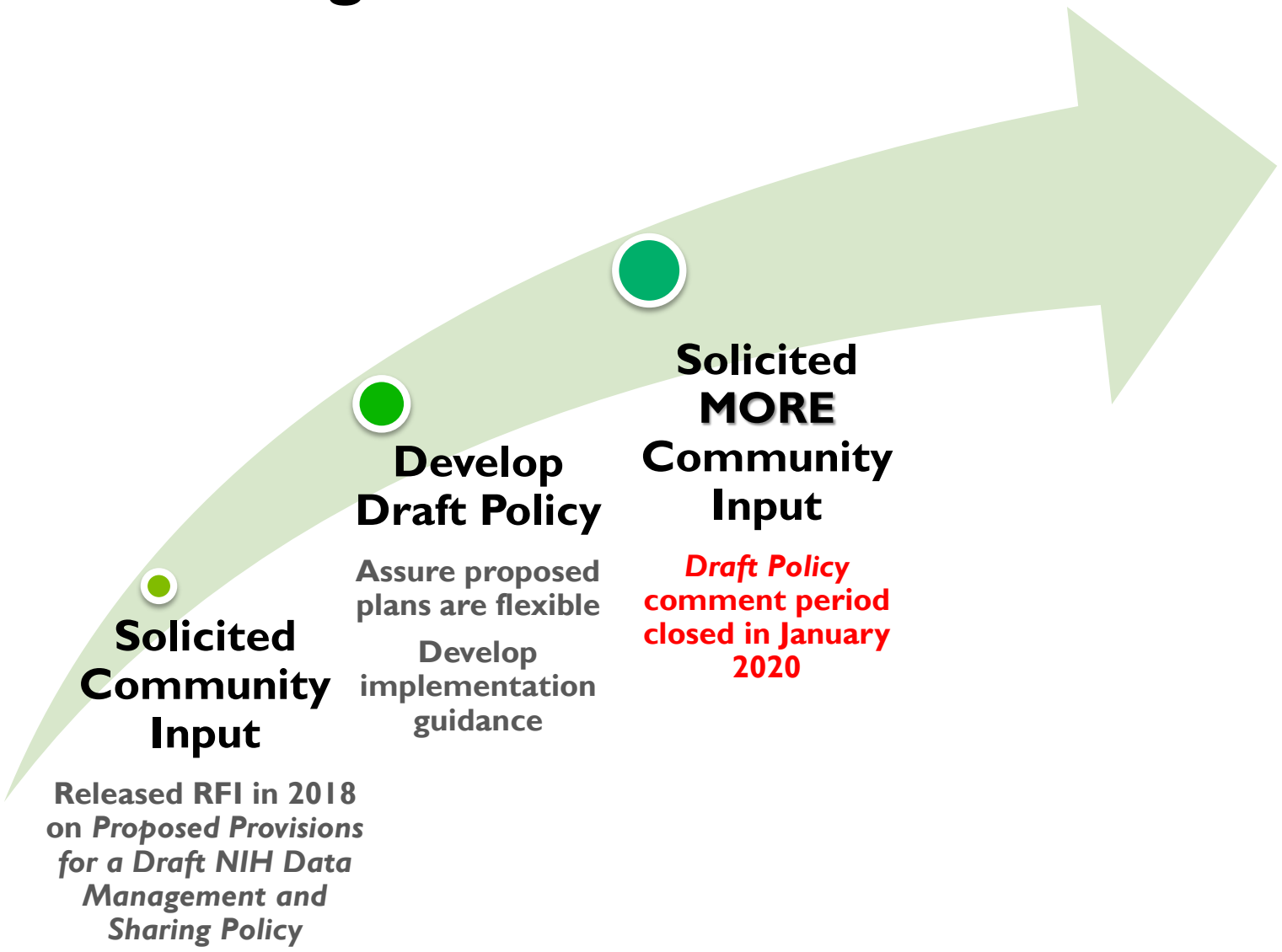
Effectively Manage Data
Use established repositories to mitigate risk

Ensure Prospective Planning
Anticipate privacy and confidentiality needs

Mandatory Informed Consent
Broadest future use of data if appropriate

DRAFT NIH POLICY FOR DATA MANAGEMENT AND SHARING

Considering All Feedback Received



Considering All Feedback Received

- **Request for Public Comment on the Draft NIH Policy for Data Management and Sharing and Supplemental Draft Guidance (November 10, 2019 – January 10, 2020)**
  - **205 responses** from both international and domestic stakeholders
  - Comments are publicly available

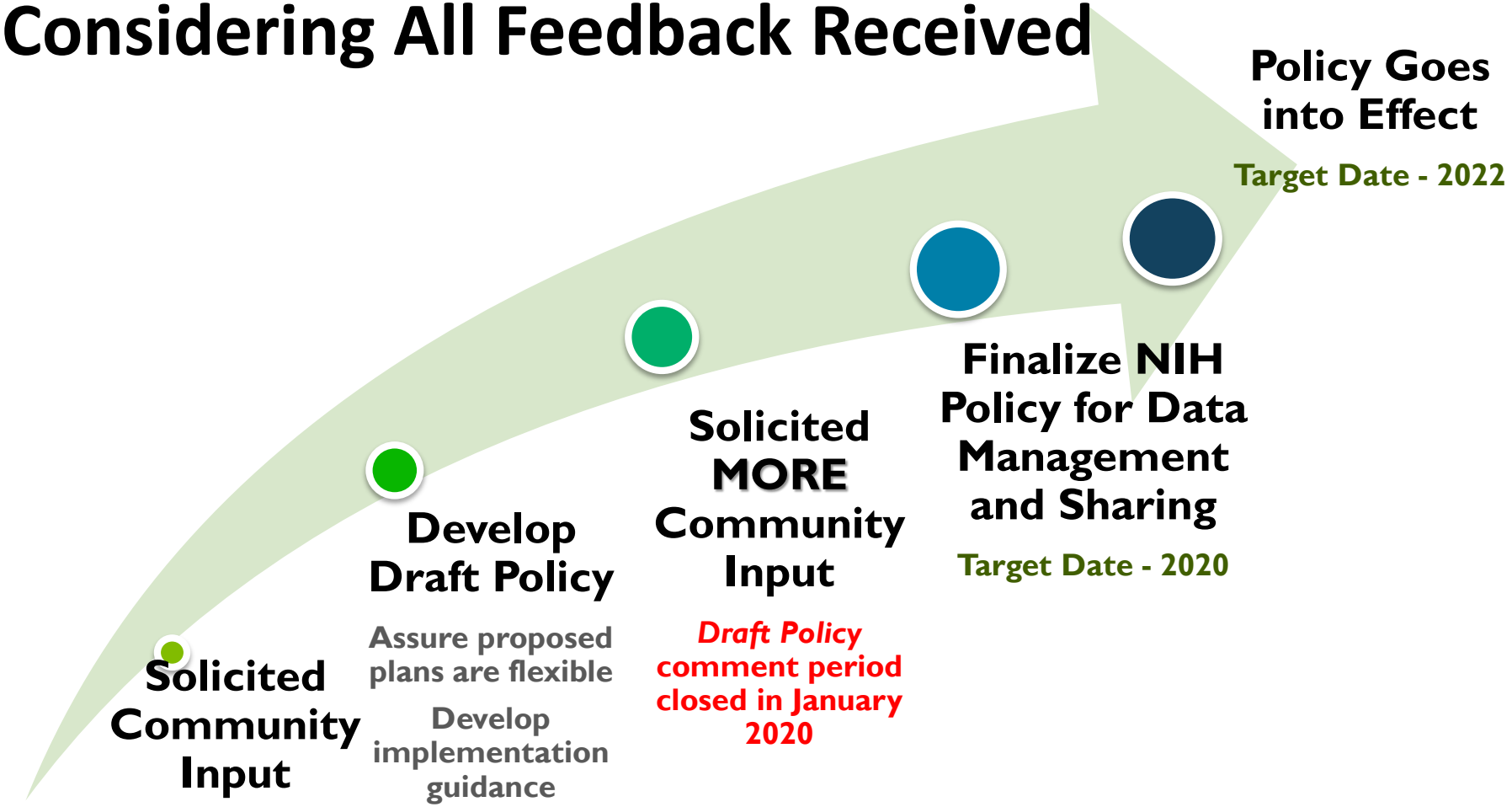
Considering All Feedback Received

- **Request for Public Comment on the Draft NIH Policy for Data Management and Sharing and Supplemental Draft Guidance (November 10, 2019 – January 10, 2020)**
  - **205 responses** from both international and domestic stakeholders
  - Comments are publicly available

- **General Comment Themes:**
  - Strong support for advancing data management and sharing through policy development
  - Support for well justified exceptions to data sharing
  - Many requests for clarifications (e.g., which data to share, when, and where to share it)

DRAFT NIH POLICY FOR DATA MANAGEMENT AND SHARING

Considering All Feedback Received



Data Sharing in the Era of COVID-19

NIH Open-Access Data & Computational Resources on COVID-19

- **NCATS COVID-19 Data Warehouse**
  - Data sharing resource to to define the clinical natural history of COVID-19
  - Interoperable, secure clinical research data environment
- **LitCovid**
  - National Library of Medicine curated literature hub for tracking PubMed COVID-19 articles
  - Categorized by geographic location & subtopics

Additional resources from NIH
COVID-19 Data
Science Collaborations

NIH COVID-19 Data Collaborations

COVID-19 Open Research Dataset Challenge

(https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge/tasks)

- Open, machine readable COVID literature (29,000+ articles)
- Text mining key scientific questions
  - Round 1: April 16, 2020 Deadline
  - Round 2: June 16, 2020 Deadline
- Supported by Allen Institute for AI, Chan Zuckerberg Initiative, Microsoft Research, and Georgetown University, with assistance from the National Library of Medicine

Research Data Alliance (RDA) COVID-19 Working Group

- **What is RDA?**
  - International initiative with the goal of building the social and technical infrastructure to enable open data sharing

- **COVID-19 Working Group developing:**
  - Guidance related to data sharing under the present COVID-19 circumstances to help researchers maximize impact of their work
  - Strategies for policymakers to maximize timely data sharing in health emergencies

NIH COVID-19 Data Collaborations

Accelerating COVID-19 Therapeutic Interventions and Vaccines (ACTIV)

NEWS RELEASES

Friday, April 17, 2020

NIH to launch public-private partnership to speed COVID-19 vaccine and treatment options

Health agencies, leading pharmaceutical companies to join forces to accelerate pandemic response.

Participating Organizations

Government

- NIH
- HHS ASPR
- FDA
- CDC
- European Medicines Agency

Non-Profit

- FNIH

Industry

- AbbVie
- Amgen
- AstraZeneca
- Bristol Myers Squibb
- Evotec
- GlaxoSmith Kline
- Johnson & Johnson
- KSQ Therapeutics
- Eli Lilly and Company
- Merck & Co., Inc
- Novartis
- Pfizer
- Roche
- Sanofi
- Takeda
- Vir Biotechnology

Take Home Messages

- NIH is committed to making the research it funds available to advance discovery, improve health, and maintain its accountability to the public
- Responsible data sharing remains an NIH priority and we are looking to the community to help us craft a path forward

MedStar Health

Knowledge and Compassion
Focused on You

April 20, 2020

Research Data Management and Sharing

Nawar M. Shara, PhD
Director, Dept. of Biostatistics and Biomedical Informatics
Associate Professor of Medicine, Georgetown University
Director BERD-CTSA (Georgetown-Howard)
MedStar Health Research Institute
nawar.shara@medstar.net
@NawarShara

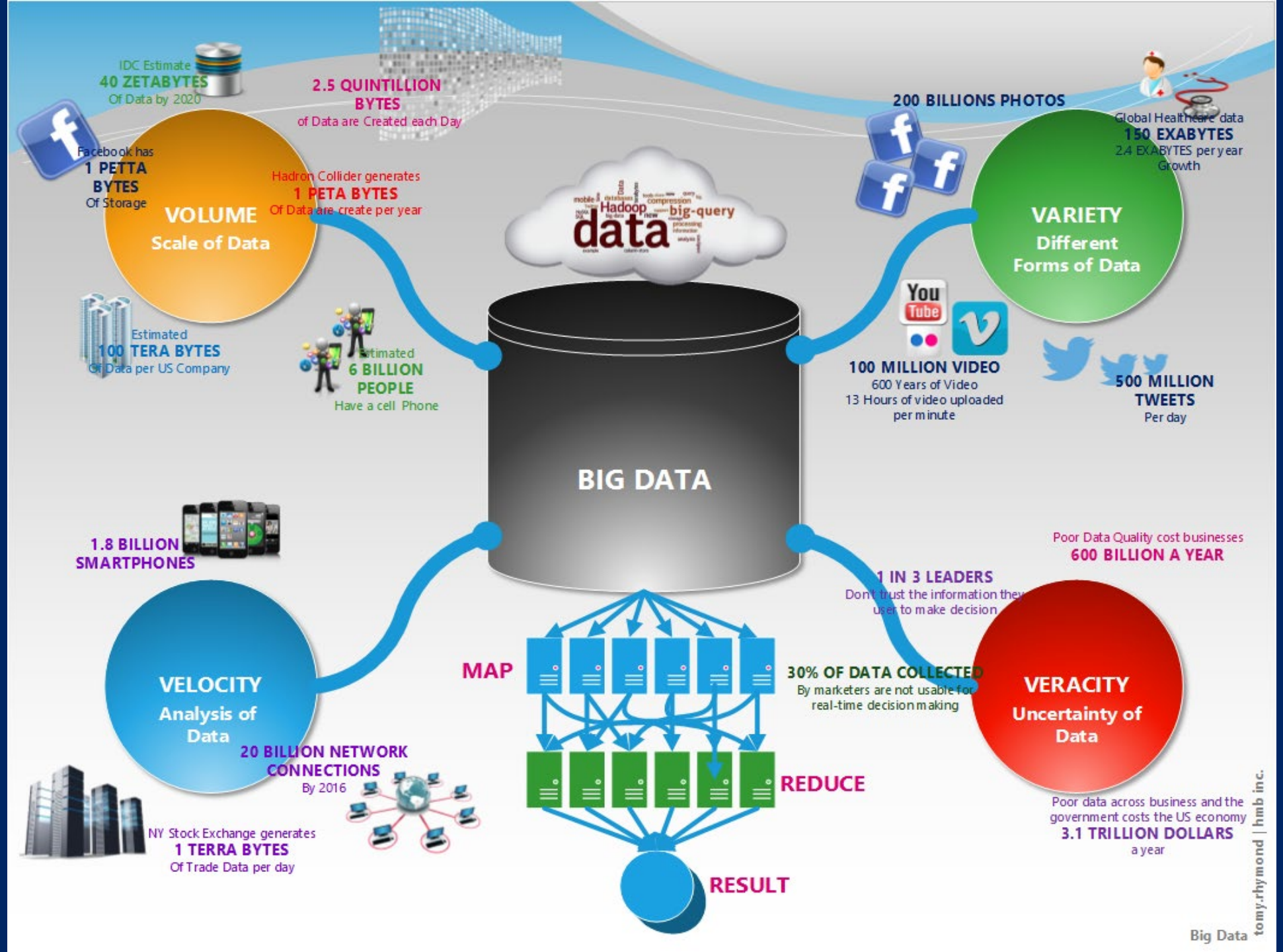
Presenter:

GU · HU · MHRI · ORNL · WDCVAMC
GHUCCTS
Georgetown-Howard Universities
Center for Clinical and Translational Science

TOPICS

- Data
- Research Data Lifecycle
- Research Data Management Activities
- FAIR Principle
- Stakeholders
- Why a Data Management Plan (DMP)?
- NIH requirements
- Sample Data Management and Sharing plans
- AI Generated Data: NIH/NCATS Study

In Healthcare, data grows exponentially and different types of data are required for research.


IDC Estimate
40 ZETABYTES
Of Data by 2020
2.5 QUINTILLION BYTES
of Data are Created each Day
Facebook has
1 PETTA BYTES
Of Storage
VOLUME
Scale of Data
Hadron Collider generates
1 PETA BYTES
Of Data are create per year
Estimated
100 TERA BYTES
Of Data per US Company
Estimated
6 BILLION PEOPLE
Have a cell Phone
data
Hadoop
big-query
BIG DATA
200 BILLIONS PHOTOS
Global Healthcare data
150 EXABYTES
2.4 EXABYTES per year
Growth
VARIETY
Different Forms of Data
100 MILLION VIDEO
600 Years of Video
13 Hours of video uploaded per minute
500 MILLION TWEETS
Per day
1.8 BILLION SMARTPHONES
VELOCITY
Analysis of Data
20 BILLION NETWORK CONNECTIONS
By 2016
NY Stock Exchange generates
1 TERRA BYTES
Of Trade Data per day
MAP
REDUCE
RESULT
30% OF DATA COLLECTED
By marketers are not usable for real-time decision making
1 IN 3 LEADERS
Don't trust the information they use to make decision
Poor Data Quality cost businesses
600 BILLION A YEAR
VERACITY
Uncertainty of Data
Poor data across business and the government costs the US economy
3.1 TRILLION DOLLARS
a year
Big Data
tomy.rhymond | hmb inc.

Research Data Lifecycle

- Project/proposal planning
- Start up
- **Data collection**
- **Data analysis**
- **Data sharing**
- End of project
- Data archive

Research Data Management Activities

- Planning
- Documenting
- Formatting
- Storing
- Anonymizing
- Controlling Access

Research Data management and sharing according to the "FAIR Principles"

- Findable
- Accessible
- Interoperable
- Reusable

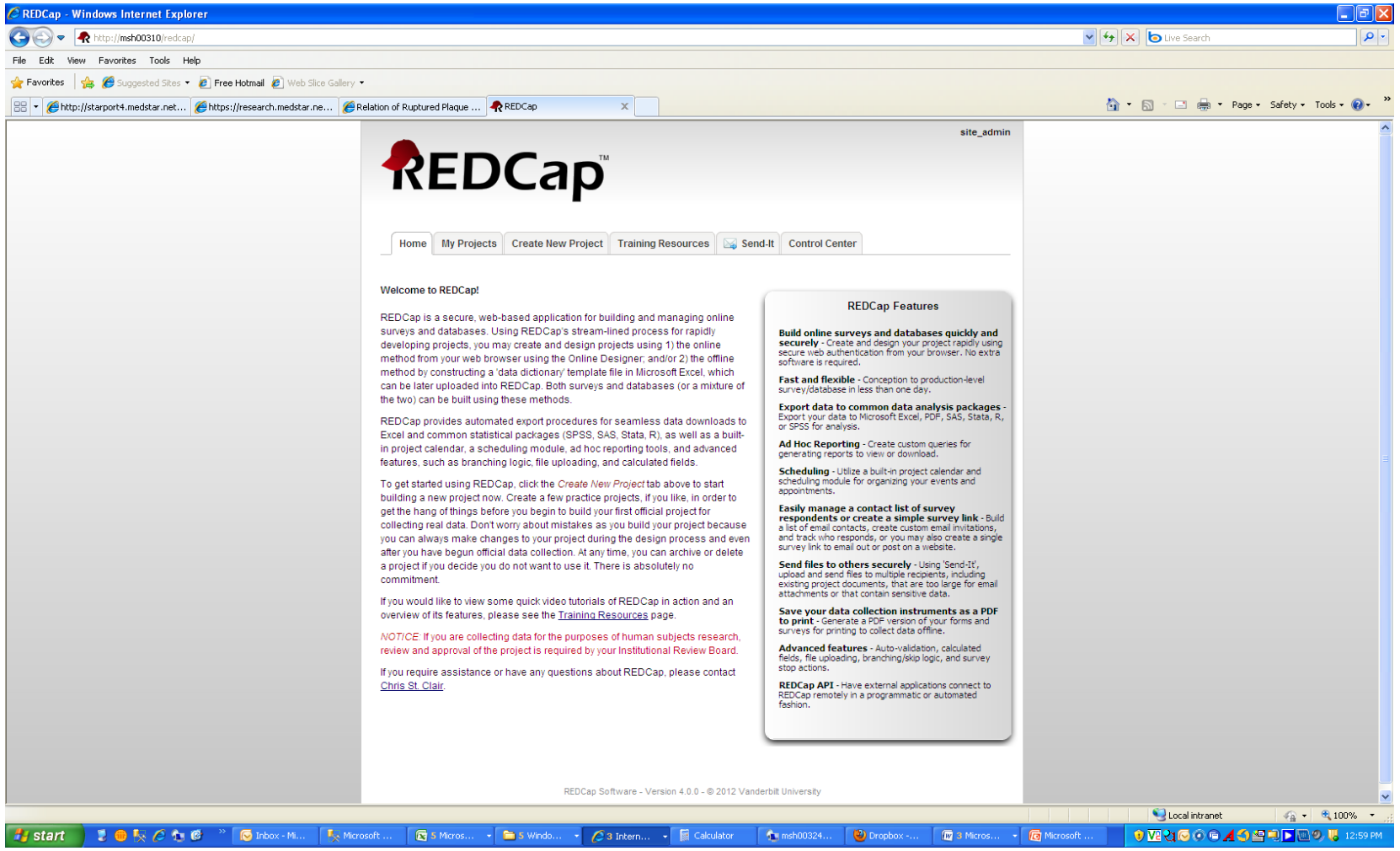
Stakeholders

- Researchers
- Institutions
- Repositories
- Funders
- Secondary users
- Publishers and Journals

What should the Data Management Plan look like?

- Everyone involved (IRB included) know what is expected at the start of the study
- To organize/create expected documents (CRF's/MOP's) so they can be produced during the course of the study
- Fulfill regulatory requirements
- Data management tasks become more visible to other groups
- Quality Assurance (QA)
- Provides continuity process and history of a project (Helpful in long term projects)

Research Electronic Data Capture (REDCAP)



http://project-redcap.org/

GU · HU · MHRI · ORNL · WDCVAMC

GHUCCTS
Georgetown-Howard Universities
Center for Clinical and Translational Science

NIH data policy

- Final Data format
- Documentation
- Analytic tools necessary to use the data
- Data sharing agreements
- How and when the data will be made available to others

NIH Proposed Methods for Data sharing

- Under the auspices of the PI
- Data Archive
- Data Enclave
- Mixed mode sharing.

https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm#archive

Sample Data Management Plan

The MHRI Department of Biostatistics and Biomedical Informatics provides the Medstar and GHUCCTS research community with customized biomedical and clinical-research informatics applications to help manage complex translational and clinical research needs. Working with the Department of Biostatistics and Biomedical Informatics gives access to the REDCap integrated data system: a user-friendly web-service for data collection linked to an encrypted MySQL database. REDCap is accessed via the Internet though secure web-applications. Data can be entered manually into the interface or uploaded in bulk from comma separated values (.csv) files. For very large datasets and data migrations it is possible to automate extract-transform-load (ETL) protocols between source databases and the REDCap MySQL backend. For custom fields, the downloadable data dictionary contains rich metadata to facilitate *post hoc* variable translation. REDCap is operating system agnostic and runs on most popular web browsers; accessing the system requires only a web browser and internet access. REDCap's powerful reporting functions allow complex interrogation of datasets. As needed the dept facilitates data visualization via custom dashboards developed using Tableau visual analytic software. Data can be bulk-exported or constrained to subsets before export to standard statistical analytics platforms for analysis (R, SAS, SPSS, etc).

Sample Research Data Sharing Plan

All data collected will be housed and maintained within the MedStar Health system on shared drives equipped with appropriate security measures behind the MedStar Health firewall. Data collection will be housed within the MedStar Health Research Institute instance of the Research Data Capture database (RedCap) file server that provides HIPAA level protection of research data. Analytics on identifiable patient safety data will be maintained within MedStar Health's research servers and conducted by MedStar Health Research Institute biostatisticians, led by Dr. Shara, whose team is used to handling sensitive and confidential data manners to maintain data privacy and security. The MedStar Health Research Institute will be responsible for maintaining the data, interventions, implementation tools and toolkits, and any other products developed from the grant activities. All research team members will be given appropriate access levels to the data to enable them to complete their portion of the project and we will ensure that all human subjects research protections are maintained by MedStar Health entities.

How to manage and share AI generated data

Knowledge and Compassion **Focused on You**

MedStar Health

Ethical dilemmas in AI

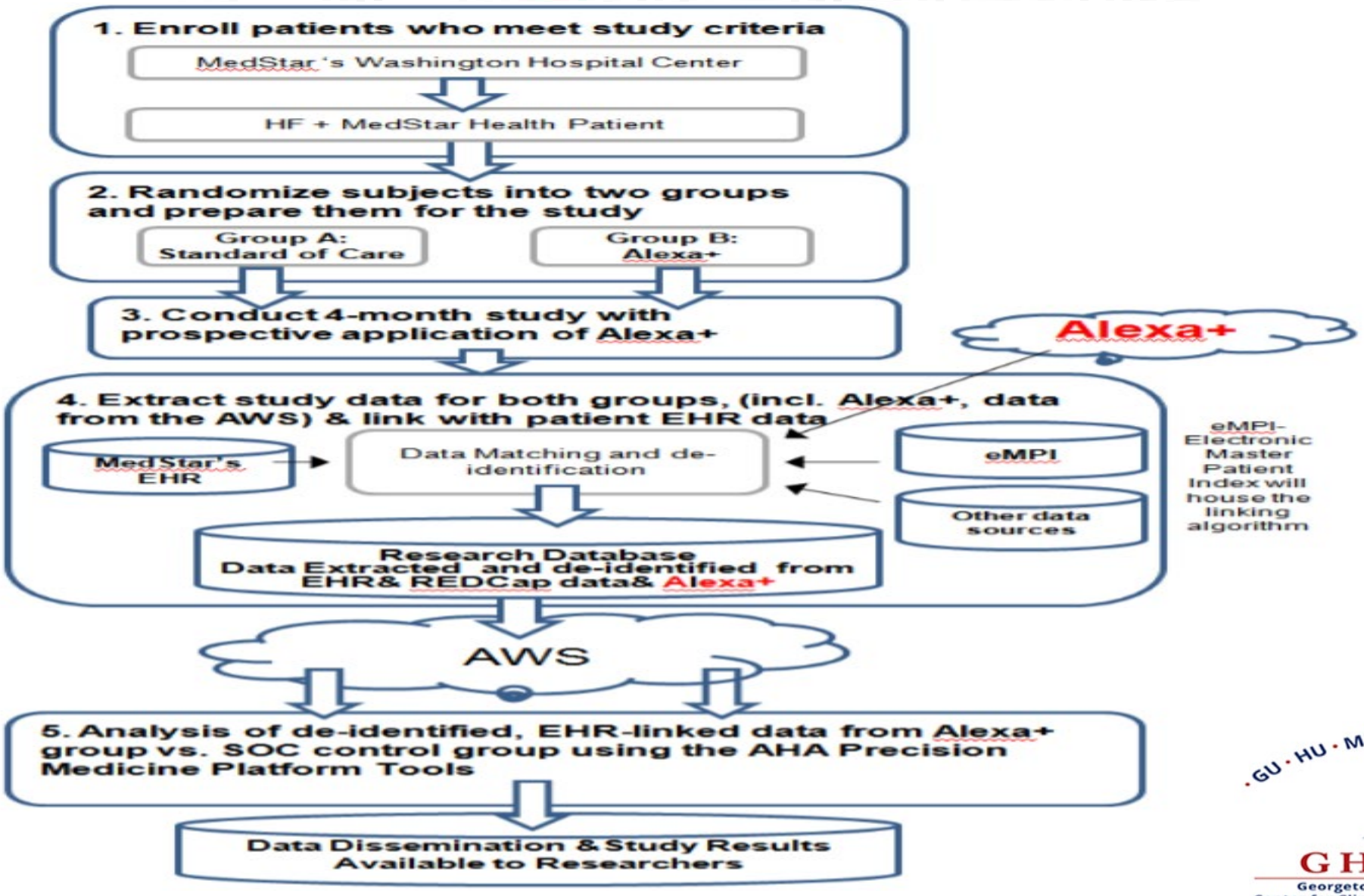
- lack of transparency (eg. algorithms)
- Who is responsible if mistakes happen (eg. Autonomous car accidents)
- Fake information can negatively impact entire population
- Unfairly discriminate (bias towards certain race or/and gender)

NIH/NCATS Funded Study:
Alexa, Treat my HF

- To design and test a customized and interactive chronic heart failure disease specific functionality (skill kit) within a voice activated technology (Amazon's Echo Dot) as a tool for management of patients at home.

- We will leverage the patients' electronic medical record to assist in future design of the tool.

Study Flow and Timeline

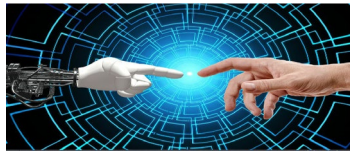


GU · HU · MHRI · ORNL · WDCVAMC

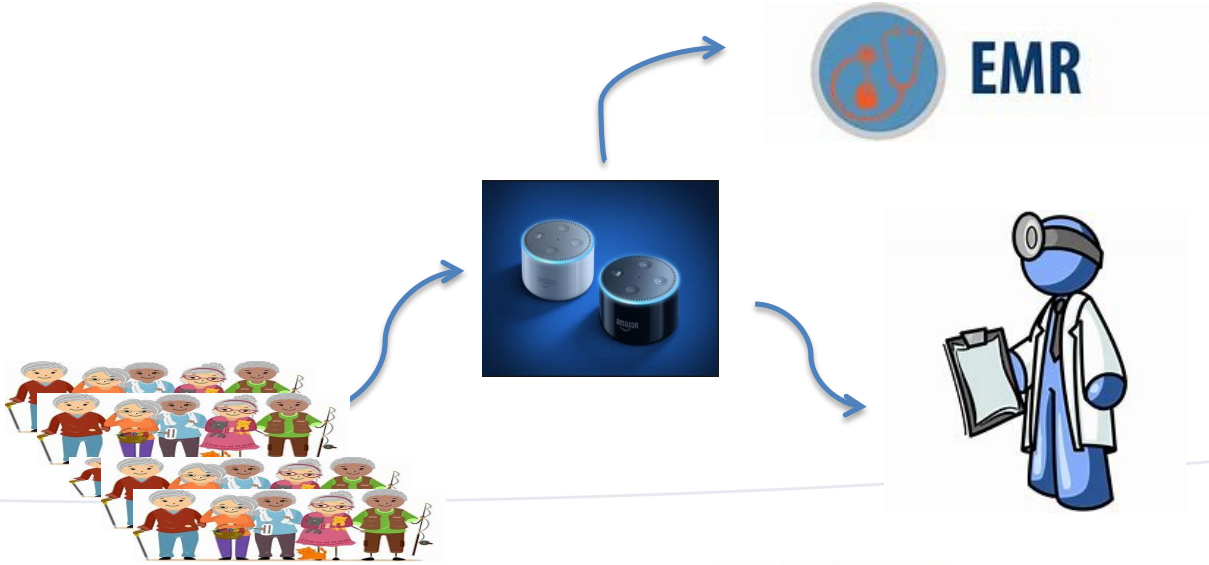
GHUCCTS

Georgetown-Howard Universities
Center for Clinical and Translational Science

Voice Enabled Devices / Avatars



- **Voice enabled, internet connected devices:**
  - **Interactive actions through a virtual personal assistant**
  - **Connect patients with health care providers**
  - **Integration with MedStar suite of EHR**



Case Study (NCATS CTSA Supplement)
PI: Nawar Shara, PhD
AI Voice activated Technology (Alexa) to Manage Heart Failure

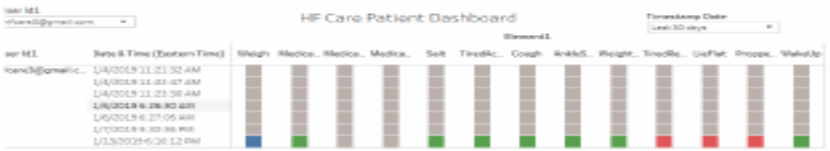
Research Methods

HF Patients Randomized into 2 Groups (SOC & ALEXA)

Voice Activated HF Survey Administered to ALEXA Group Daily for 3 Months

Patients Invoke ALEXA and answer Health Maintenance Questions (Y/N)

Responses are Recorded and Stored in AWS Cloud and shown on a dashboard to monitor participants responses



Results will be Integrated with EHR Data to Assess outcomes (e.g. readmission, technology uptake)

Regulatory

AI Assistant Programmed to Record Only Yes or No Responses

Data Secured on AWS

Required inclusion criteria of a private home with a WiFi system

Option to Turn-off Listening by locking the ALEXA unit

Recording Stops When Locked, Sleeping, or Force-quit

In summary,

- Proper research data management and data sharing plans will provide a better and increased potential for funding.
- Transparency and better collaborative efforts.
- Reproducible research, scientific integrity, and validation of the findings/results.
- Adoption by the scientific community

BioCompute Objects

CTSA 2020

Panel on Research Data Management and Sharing

Raja Mazumder

The George Washington University

(mazumder@gwu.edu)

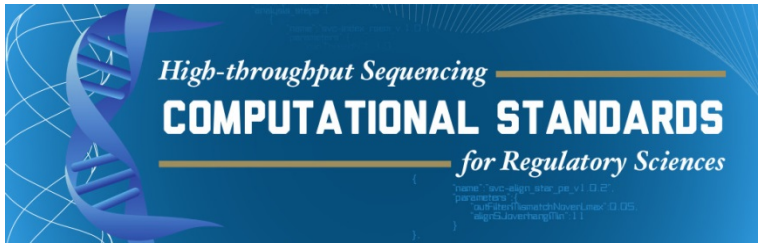
https://biocomputeobject.org/

Copyrightable parts of slides: Creative Commons Attribution (CC BY 4.0) License

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

H I V E

BioCompute Workshops



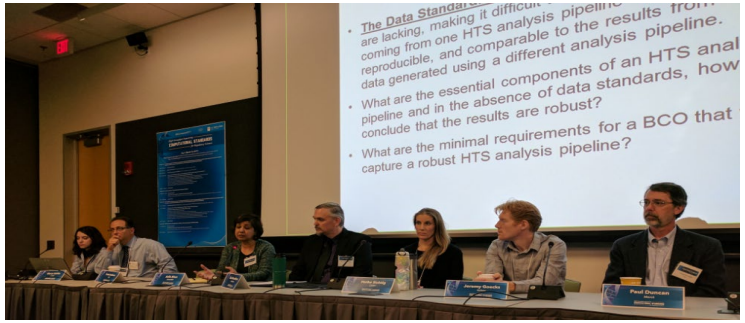
<u>Purpose:</u> *Community input on creating a standardized framework for computational analyses of HTS (NGS) data for FDA submission.*

This FRAMEWORK is known as a BioCompute object (BCO). We had hundreds workshop participants in yearly workshops. 1st FDA workshop in 2014.

Main outcomes:

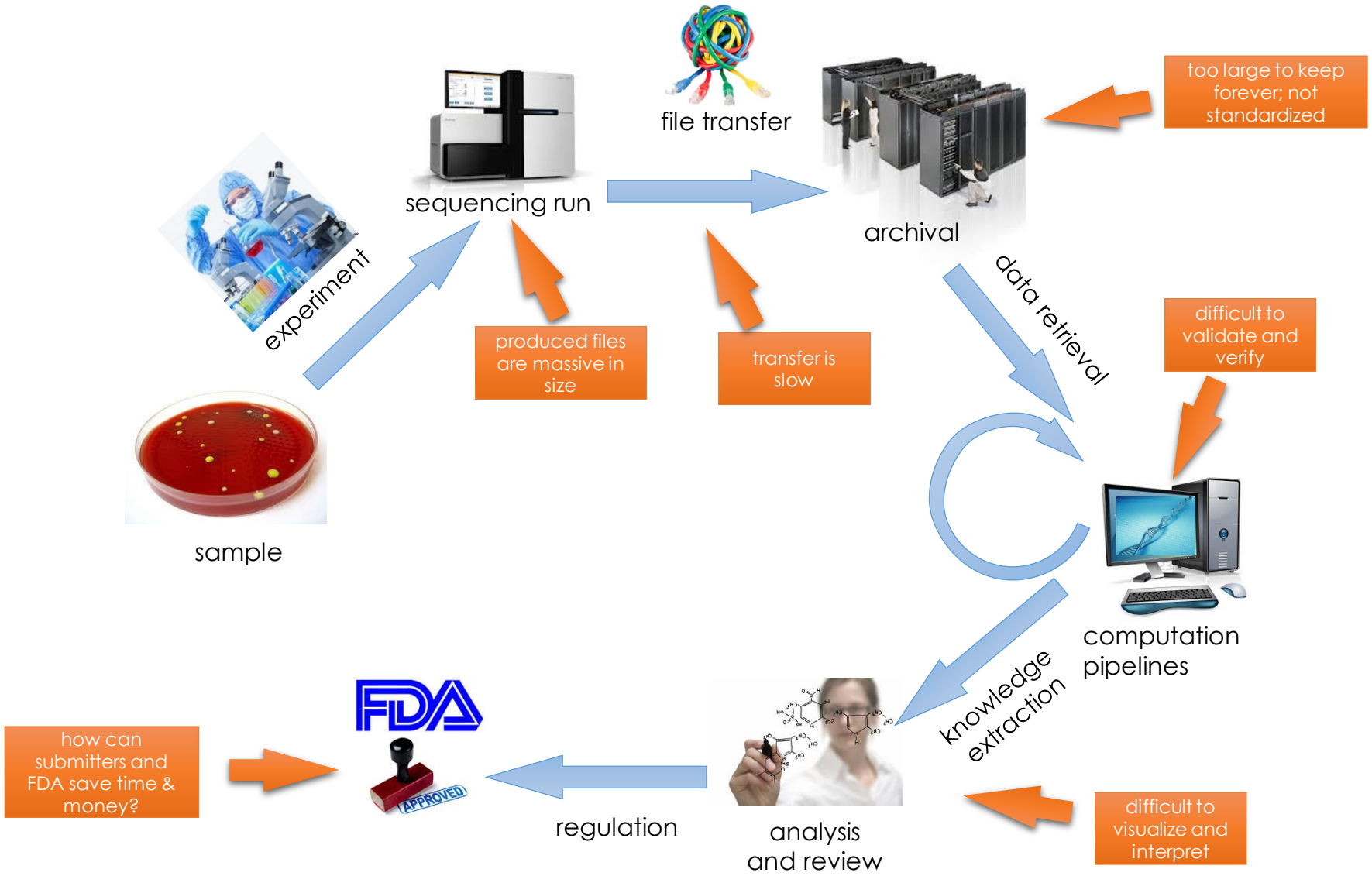
- Creation of the BCO specification document
- Public-private BCO-spec working groups with regular meetings
- BCO demonstration projects (for submitters and software platform developers) with community stakeholders.

Accurate **communication** is critical for regulatory evaluation of HTS (NGS) based products.



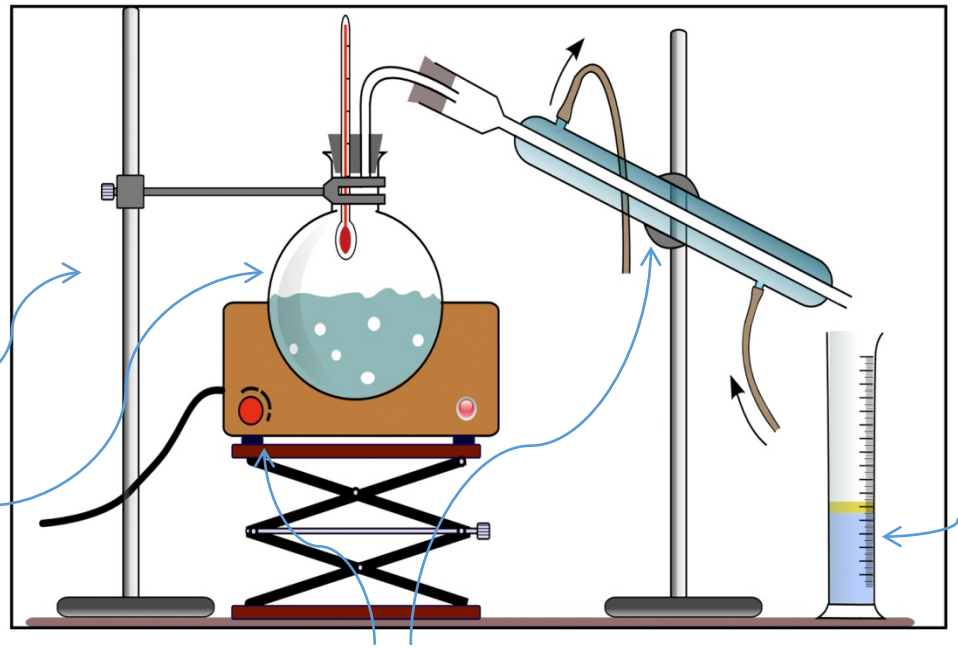
NGS lifecycle: from a biological sample to biomedical research and regulation

BioCompute Objects



Adapted from Vahan Simonyan

BioCompute Object need



experimental conduct

chemicals

yield

conditions

generalized experiment

inputs

parameters

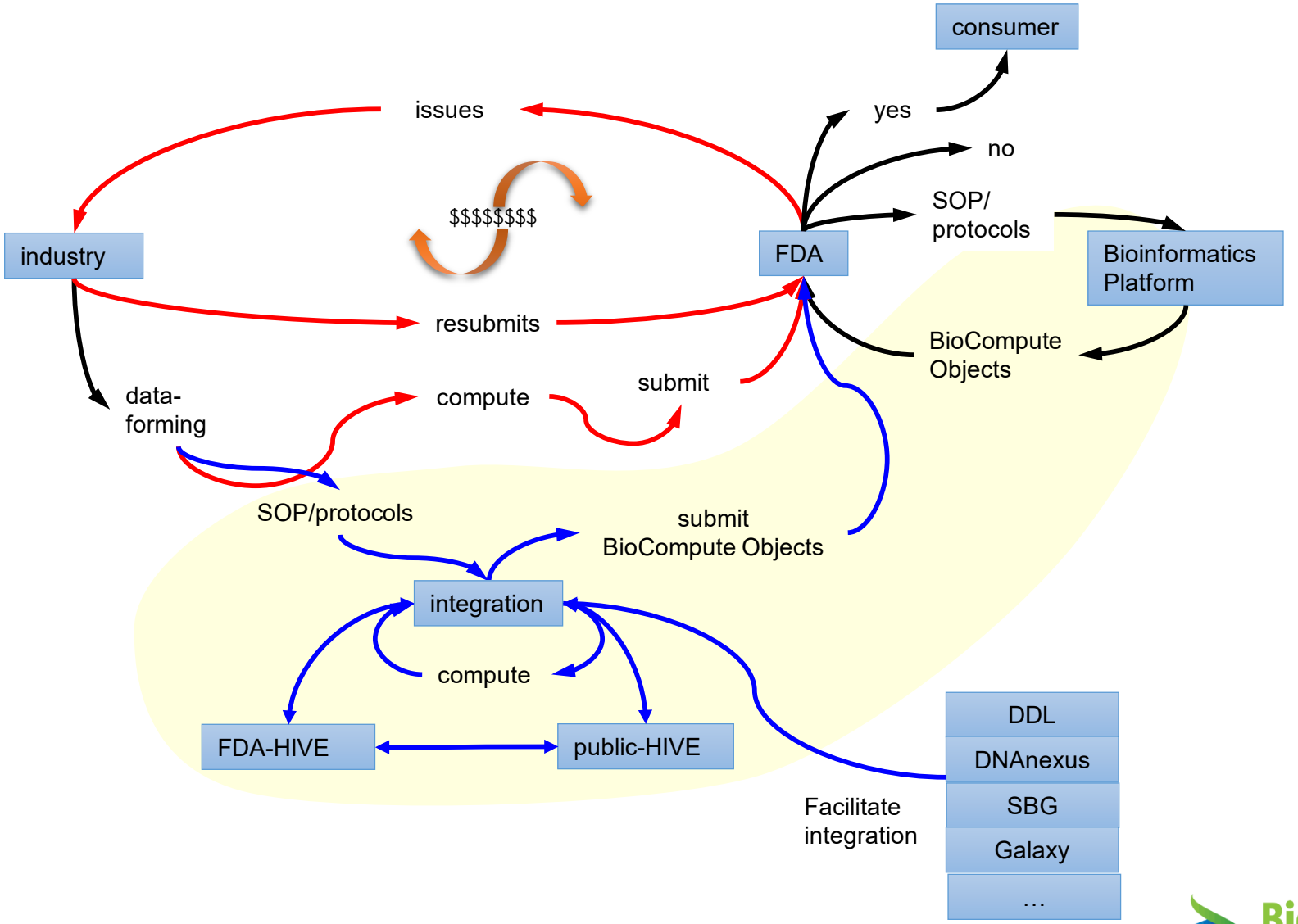
outputs

```
> my_program -i input_file1 -parameter1 value1 -parameter2 value2 -o out_file
```

Adapted from pixabay.com/en/chemistry-distillation-experiment-161575/ and Biocompute Objects-A Step Towards Evaluation and Validation of Biomedical Scientific Computations. PMID: 27974626.

Adapted from Vahan Simonyan

BioCompute Object need



BioCompute Objects

A solution should …

- Be **human readable**: like a GenBank sequence record
- Be **machine readable**: like a GenBank sequence record. Structured information with predefined fields and associated meanings of values
- Contain enough information to interpret information, understand the computational pipelines, maintain records, and reproduce experiments
- Have a way to be sure the information has not been altered: immutable

BioCompute Objects

802.11 Analogy

- More commonly called “WiFi.” Does not standardize the platform that information is generated on, the applications that use the information.
- The only thing that it standardizes is how information is collected and communicated between two devices. From there, you can do whatever you want with it.

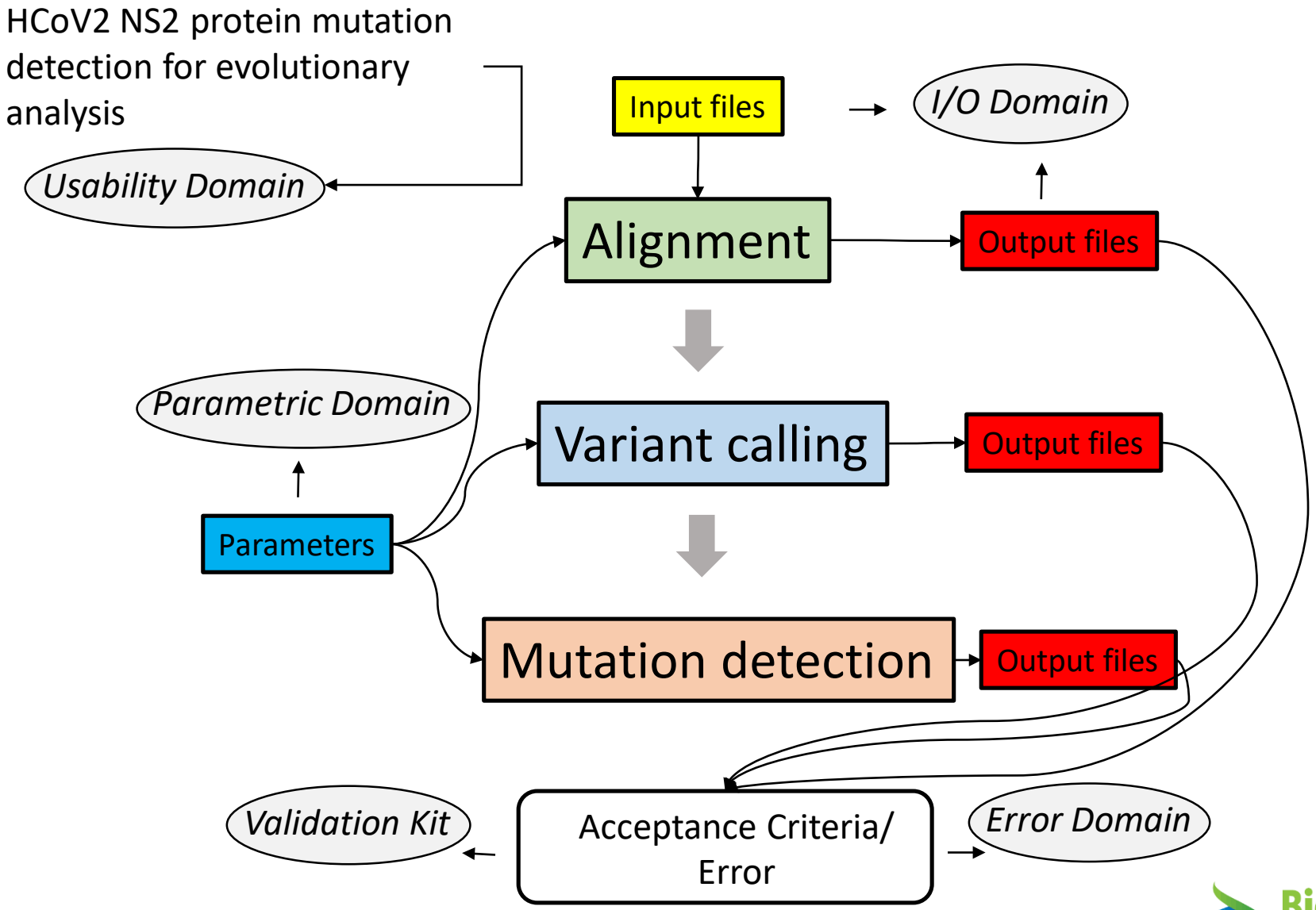
BioCompute Objects

BioCompute

- Standard for communicating computational analysis workflows
- Acts like an envelope for entire pipeline
  - Can incorporate other ontologies, standards (e.g. CWL, RO, DO, GA4GH, SEQC2…)
- Built with input from FDA, academia and industry
- Human and machine readable
  - Written in JSON
  - Unique IDs for versioning
- Categorized by domains (Usability, Execution, Error Domain etc.); Interoperable; Adaptable; …
- IEEE approved standard for communicating genomic analysis workflows (https://standards.ieee.org/standard/2791-2020.html)

BioCompute Objects

What is a BioCompute Object?

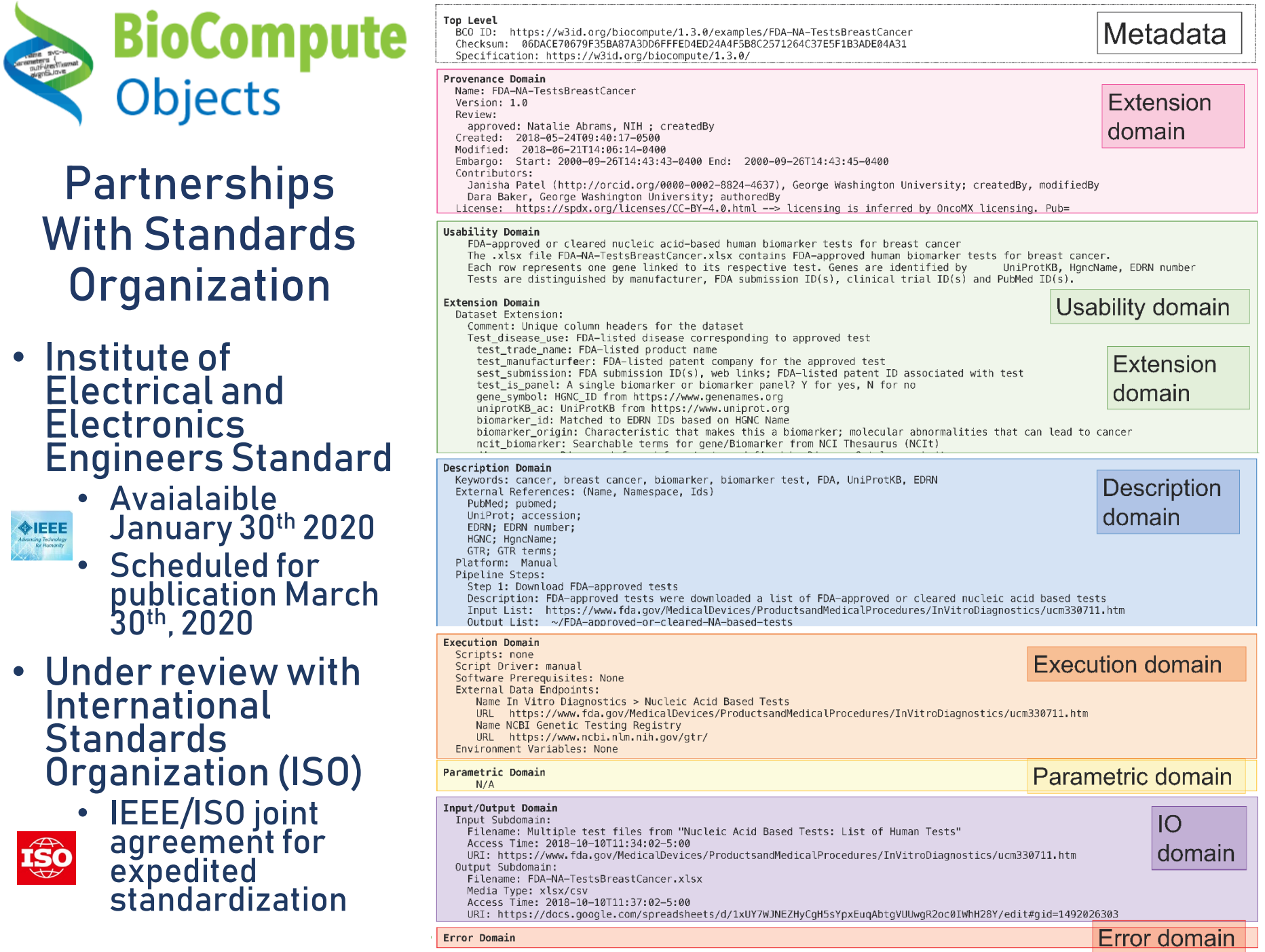


BioCompute Objects

BioCompute Objects

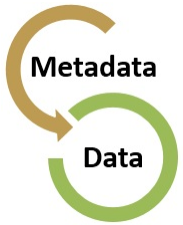
Partnerships With Standards Organization

- Institute of Electrical and Electronics Engineers Standard
  - Avaialaible January 30th 2020
  - Scheduled for publication March 30th, 2020
- Under review with International Standards Organization (ISO)
  - IEEE/ISO joint agreement for expedited standardization



Salient Features of BioCompute

- Provides important metadata required for reproducing data



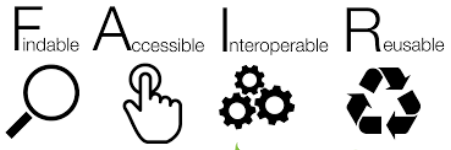
- Provides seamless communication and collaboration opportunities

- Provides appropriate attribution

- Available in machine readable and human readable format

- Provides license details that helps others to use the data accordingly

- Data and metadata can follow FAIR principles



BioCompute Objects

Acknowledgements

Vahan Simonyan
BioCompute Co-founder

Jonathon Keeney
BioCompute Lead

Hadley King
BioCompute
Technical Lead

Janisha Patel
BioCompute
Training Lead

PDA J Pharm Sci Technol, 71 (2), 136-146 Mar-Apr 2017

Biocompute Objects–A Step Towards Evaluation and Validation of Biomedical Scientific Computations

Vahan Simonyan [1], Jeremy Goecks [2], Raja Mazumder [3]

PLOS BIOLOGY
FIFTEENTH ANNIVERSARY

BROWSE PUBLISH ABOUT

OPEN ACCESS

COMMUNITY PAGE

Enabling precision medicine via standard communication of HTS provenance, analysis, and results

Gil Alterovitz, Dennis Dean, Carole Goble, Michael R. Crusoe, Stian Soiland-Reyes, Amanda Bell, Anais Hayes, Anita Suresh, Anjan Purkayastha, Charles H. King, Dan Taylor, Elaine Johanson, Elaine E. Thompson, [...], Raja Mazumder [view all]

Funding: FDA HHSF223201510129C; NSF/Internet2 E-CAS (AWS portability Exploring Clouds for Acceleration of Science)