BioCompute Objects

Public Workshop on Opportunities for Accelerating Scientific Discovery: Realizing the Potential of Advanced and Automated Workflows

Session on Standards, Governance, and Social Context

Raja Mazumder

The George Washington University

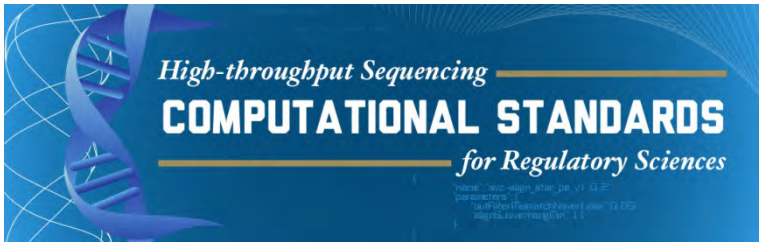
(mazumder@gwu.edu)

https://biocomputeobject.org/

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

THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC
HIVE

BioCompute Workshops



Purpose: *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 since first workshop in 2014.

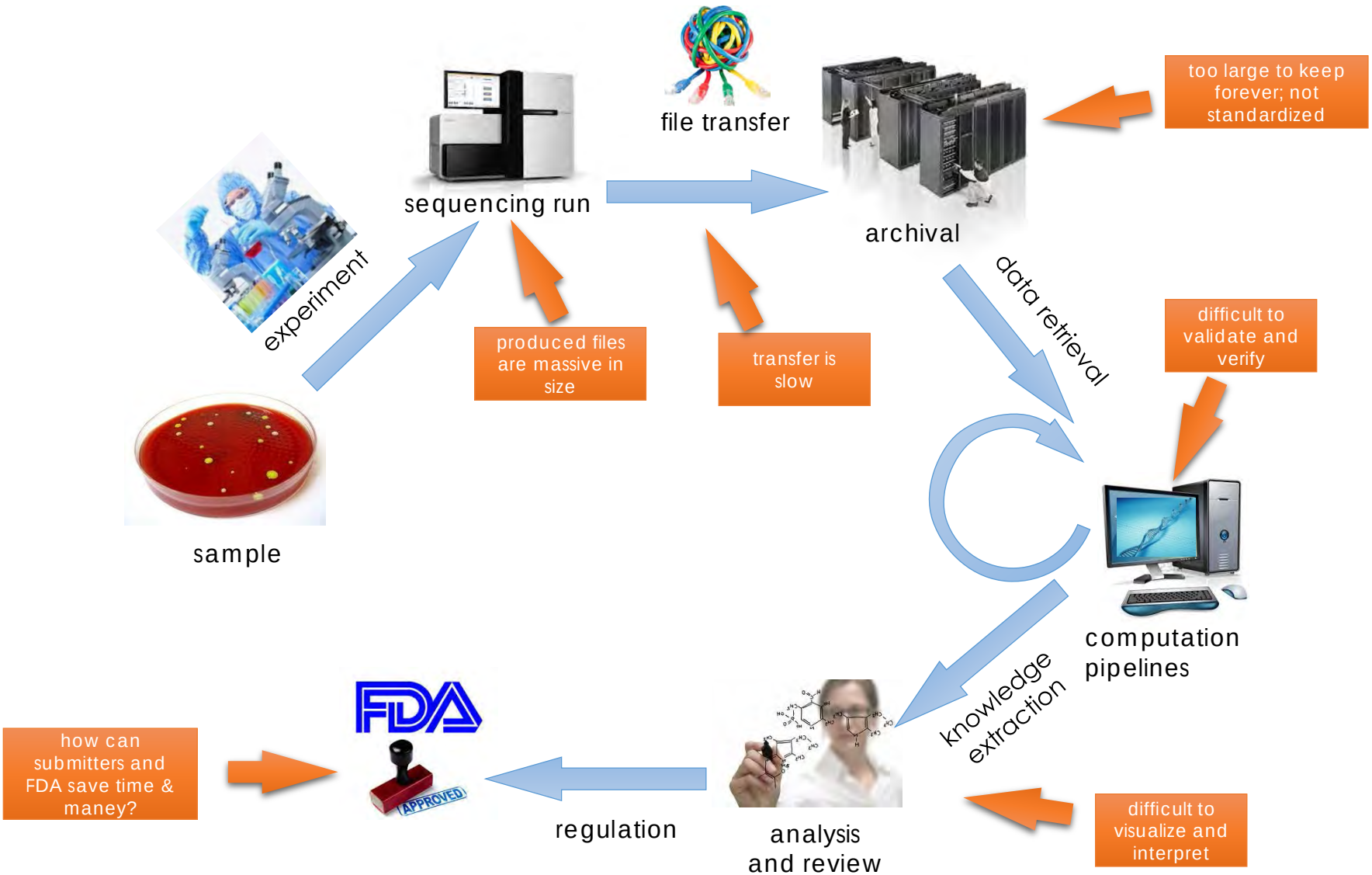
Main outcomes:

- 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.

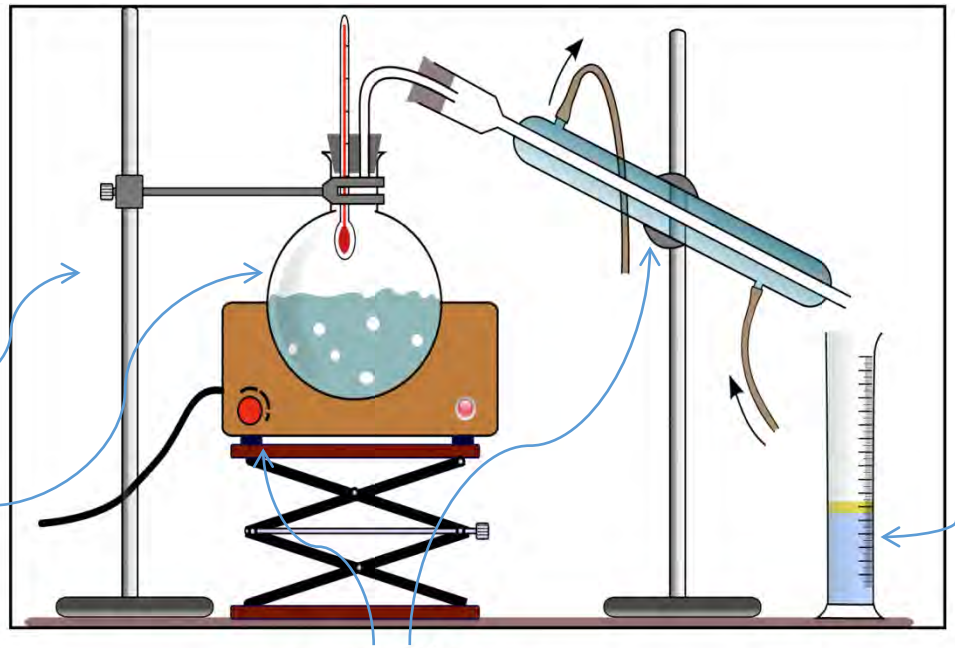
NGS lifecycle: from a biological sample to biomedical research and regulation

BioCompute Objects



Adapted from Vahan Simonyan

BioCompute Object need



experimental conduct

chemicals

yield

conditions

generalized experiment

inputs

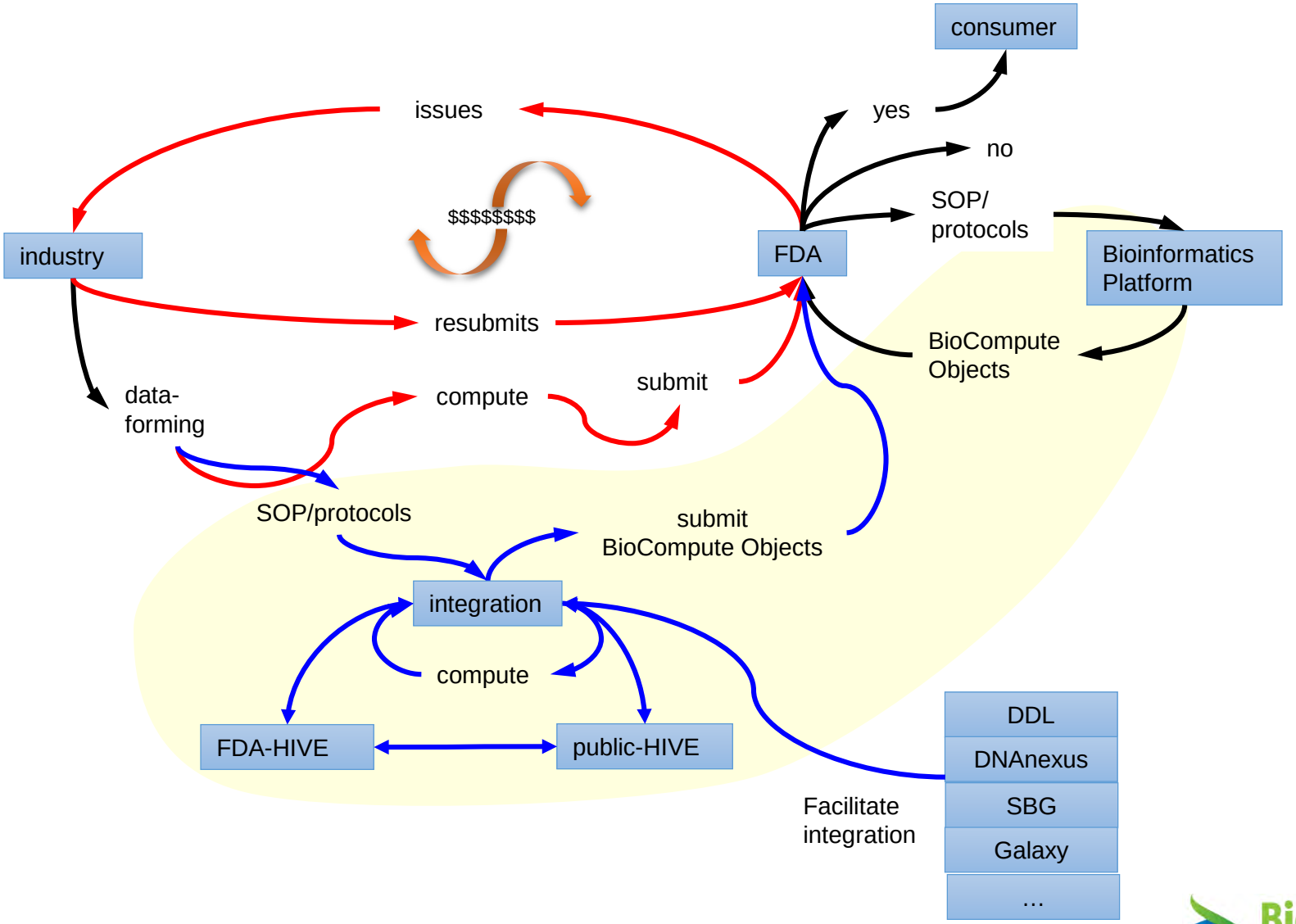
parameters

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.

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.

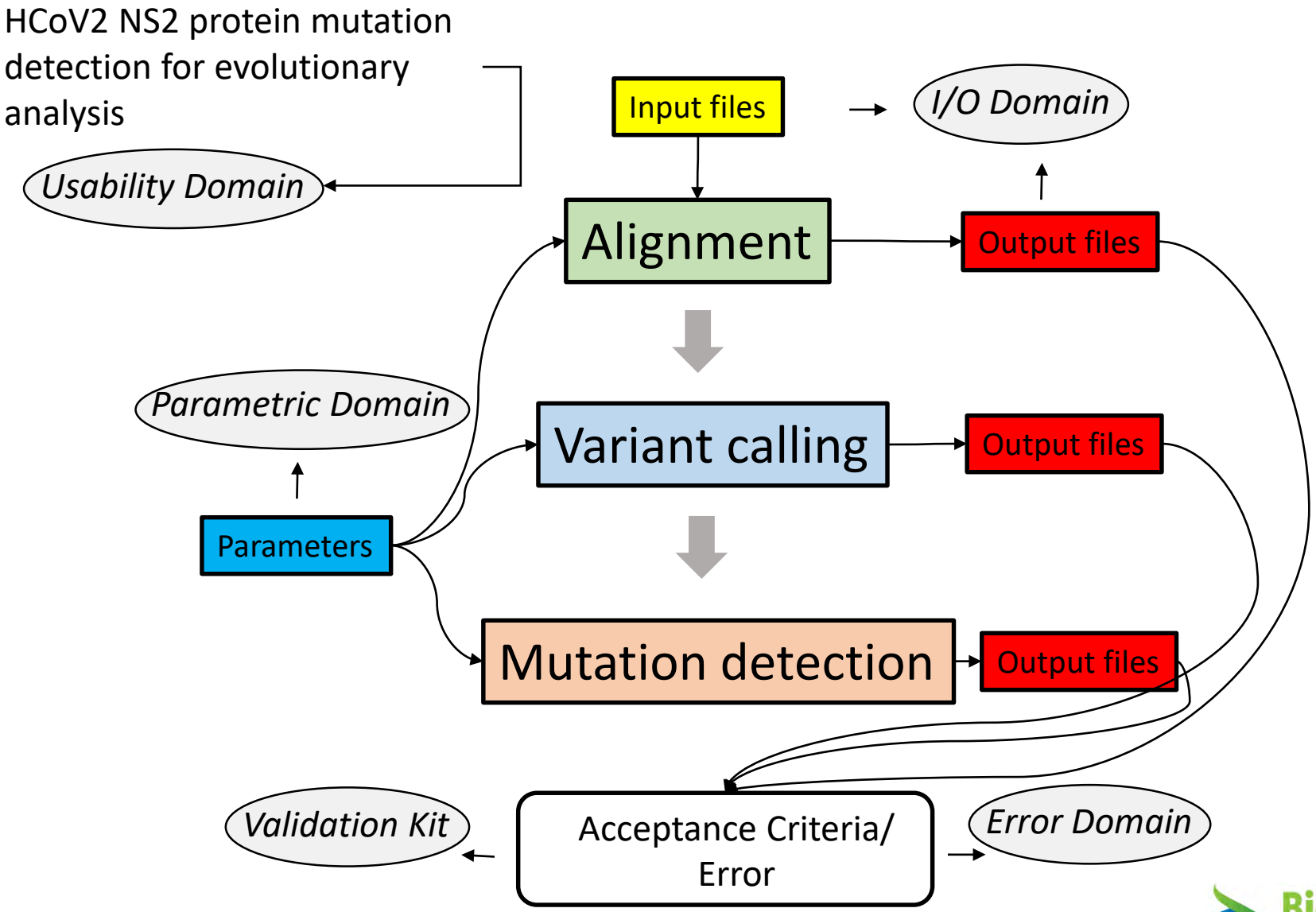
BioCompute Objects

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?

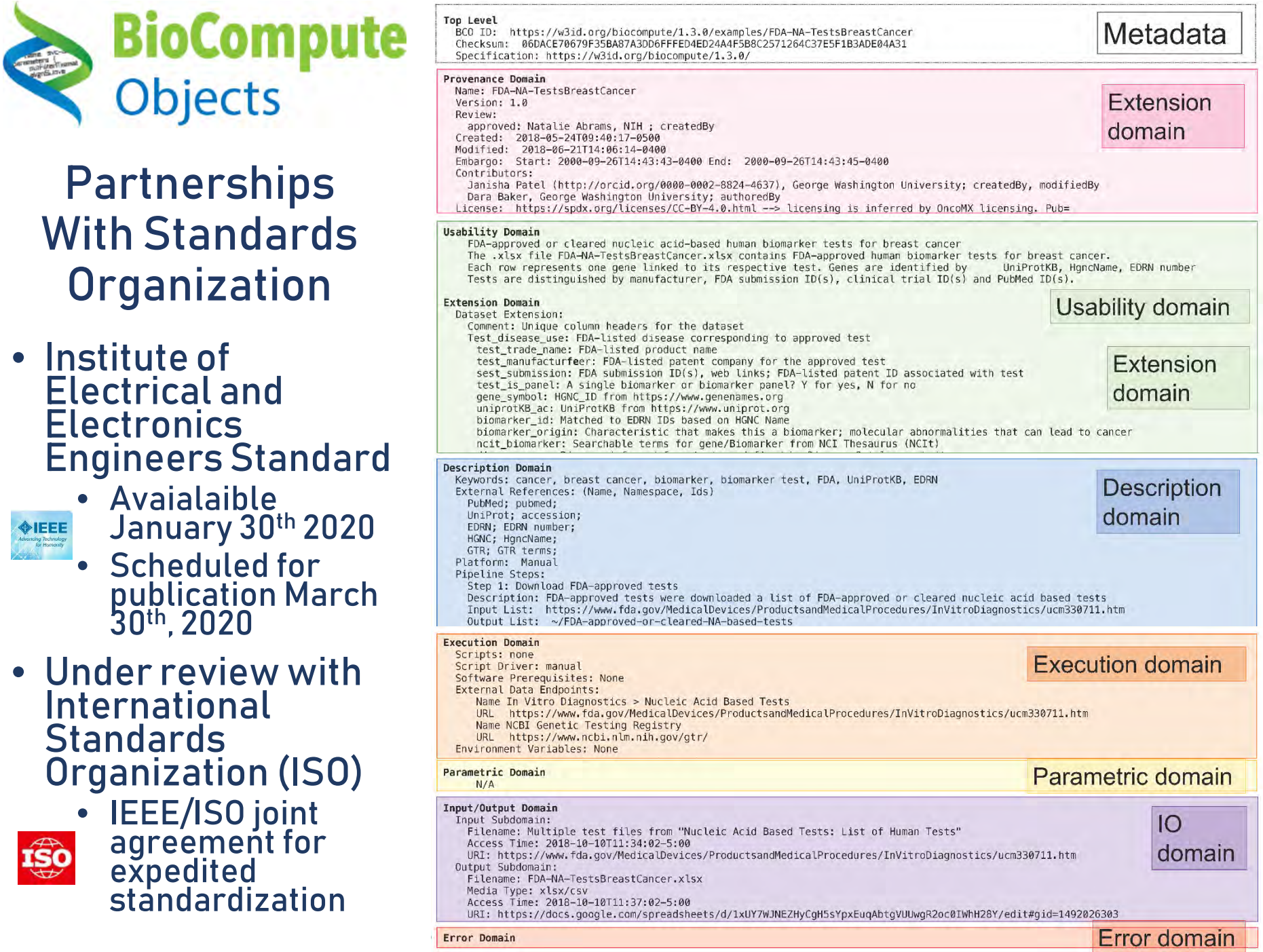


BioCompute Objects

BioCompute Objects

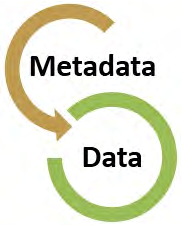
Partnerships With Standards Organization

- 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



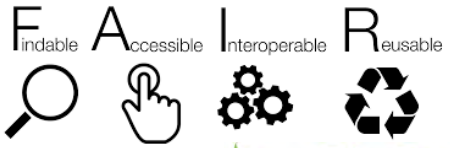
- Provides seamless communication and collaboration opportunities

- 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 (training); NSF/Internet2 E-CAS (AWS portability Exploring Clouds for Acceleration of Science)