

The Commons - Implementation Plans 2017 & 2018-

BD2K MCWG – September 1, 2016

Vivien Bonazzi (ADDS)

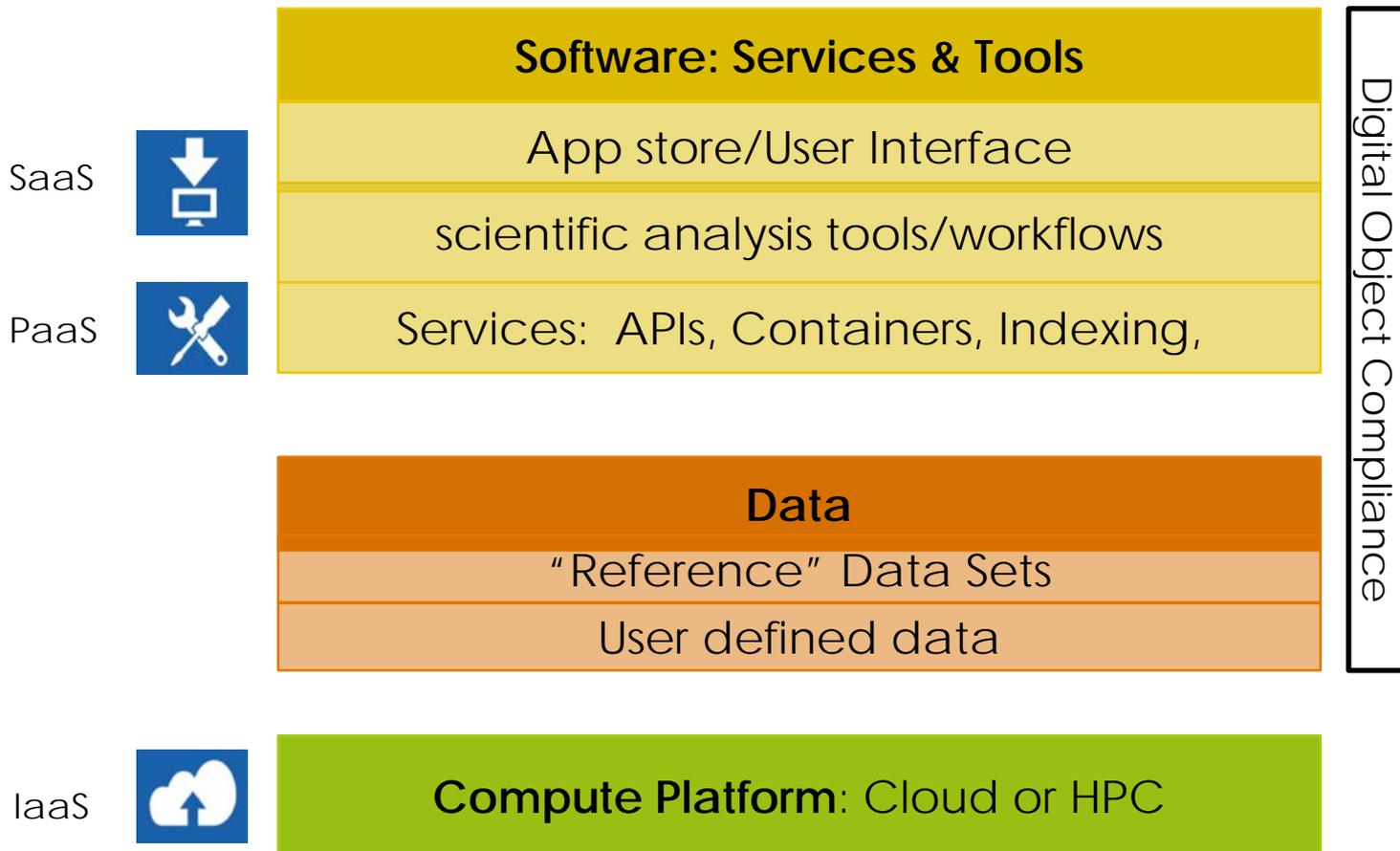
The Data Commons
is a framework
that fosters the development
of a digital ecosystem

“A **framework** is a plug and play model that allows multiple participants (*producers and consumers*) to connect to it, interact with each other and create value”

Developing a *Data Commons*

- Treats products of research – data, methods, papers etc. as *digital objects*
- These digital objects exist in a shared virtual space
 - Find, Deposit, Manage, Share, and Reuse data, software, metadata and workflows
- Digital object compliance through **FAIR** principles:
 - Findable
 - **A**ccessible (*and usable*)
 - Interoperable
 - **R**eusable

The Data Commons Framework



FY15-FY16 BD2K Data Commons Activities

Commons Framework Pilot

- Explore feasibility of the Commons framework
- Provide data objects to populate the Commons
- Facilitate collaboration and interoperability
- *Supplements*

Cloud Credit Model

- Provide access to cloud (IaaS) and PaaS/SaaS via credits
- Connecting credits to NIH Grants
- *Contract to MITRE (FFRDC)*

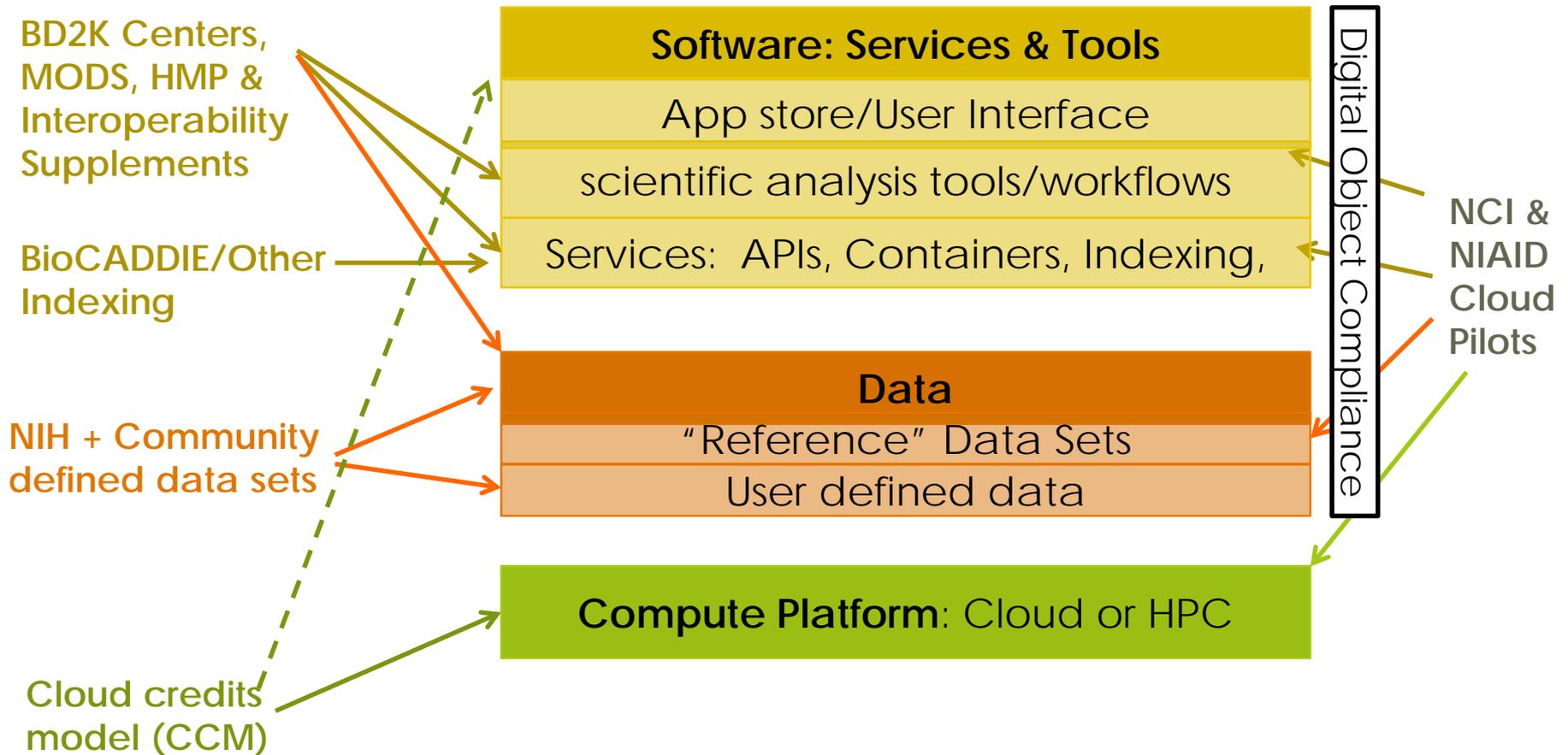
Reference Data Sets

- Making large and/or high impact NIH funded data sets and tool accessible in the cloud
- *Intersects with the Data Management Task Force & Common Fund*

RIWG Resource Index WG

- Developing Data & Software Indexing methods
- Leveraging BD2K efforts bioCADDIE et al
- Collaborating with external groups
- *Supplements and bioCADDIE award*

Mapping BD2K Activities and Commons Pilots to the Commons Framework



Detailed description of current
FY16 Data Common Pilot activities

Appendix Section

[Proposed Commons Implementation Concepts_V3.doc](#)

BD2K Data Commons

Proposed Implementation Plans

FY 2017-2018

Commons Implementation Concepts

1. Large, High-Impact Datasets in the Cloud - Populating the Data Commons
2. BISTI-BD2K Notice to Support Commons Framework Projects
3. Implementation of Commons Frameworks.
4. Making Existing Projects/Objects FAIR.

1. Large, High-Impact Datasets in the Cloud - *Populating the Data Commons*

Overview:

Large, High-Impact Datasets in the Cloud - Populating the Commons

- Make large, high impact, NIH funded data sets available in the cloud/commons
- Co-locate large datasets and compute power, to improve access, use, re-use, and sharing of data and tools
- Kick-start the Commons with Commons-compliant data and tools
 - Data must adhere to Common compliance /FAIR principles
- Provide an indexable test data sets for bioCADDIE (and other indexing efforts)

What will we learn:

Large, High-Impact Datasets in the Cloud - Populating the Commons

This pilot project will inform NIH on:

- Which Clouds are most functional, practical, and cost effective?
- What is involved in moving data resources to the Cloud?
- What will it cost?
- How to manage challenges associated with both open access and controlled access data?
- How do we find data and resources across clouds?
- How do we compute across clouds?

Proposed Components:

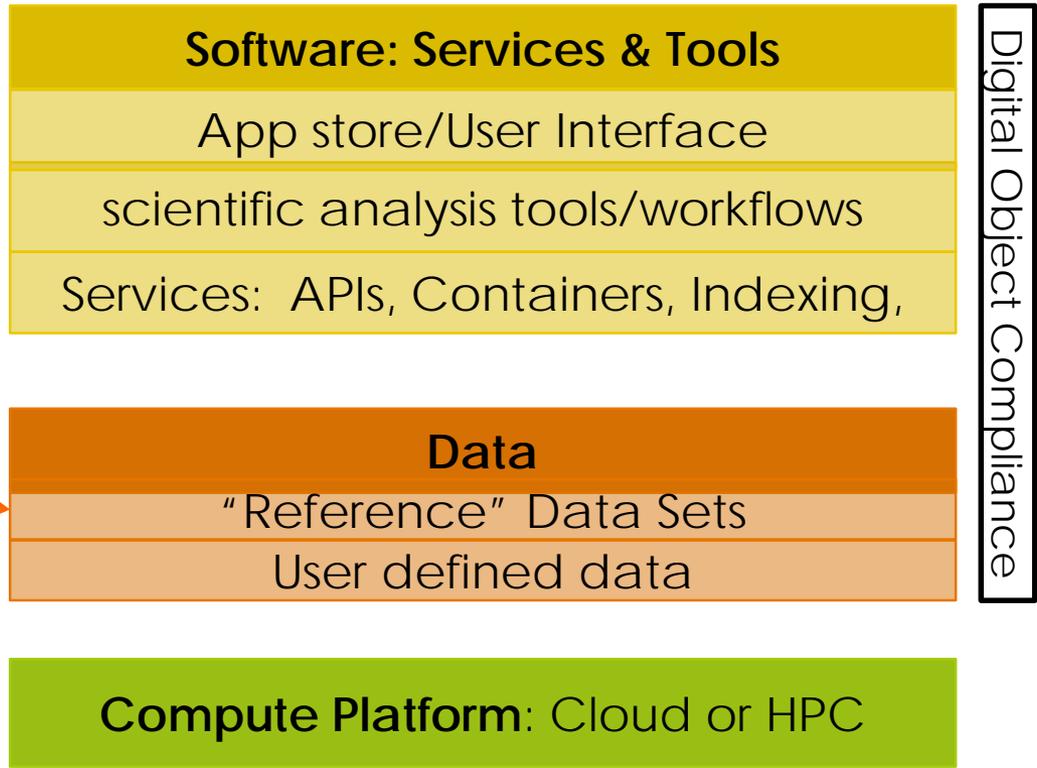
Large, High-Impact Datasets in the Cloud - Populating the Commons

- **Biomedical data resources and tools.** Will support 3-5 awards to migrate large, high-impact datasets and associated tools into multiple cloud providers.
- **Cloud Infrastructure.** 2-4 cloud providers will supply infrastructure to support the Data Commons.
 - Each data resource will be replicated, as much as possible, in each the cloud.
- **Coordination.** One coordination award will facilitate activities across the biomedical data resources and cloud providers.
 - will track metrics of success and impact of the overall project

Mapping to the Commons Framework

Large, High-Impact Datasets in the Cloud - Populating the Commons

Large, High-Impact
Data Sets in the
Cloud



2. BISTI-BD2K Notice to Support Commons Framework Projects

BISTI-BD2K Notice to Support Commons Framework Projects

- BISTI Notice would allow NEW applications to be received from the community
 - Two R01 mechanisms, as well as SBIR and STTR programs
- Can be developed and released quickly (2-4 months)
- Uses current NIH application receipt and review mechanisms
- Engages many more ICs in the Commons
 - ICs support applications relevant to their specific mission

BISTI-BD2K Notice to support Commons Framework Projects

- BISTI Program Announcements support investigator-initiated applications
- This Notice would solicit tools and services that align with the Commons Framework.
 - New indexing methods
 - Workflows that employ API's to access the data
 - Ability to compute across many nodes
 - Improved technologies and approaches for containerization of tools and workflows to work efficiently and be more easily deployed in the cloud.
 - PaaS (Platform as a Service) or SaaS (Software as a Service) approaches that provide an entire suite of tools and services that operate in the cloud.
- Budget: This requires no BD2K set-aside funds.

3. Implementation of Commons Frameworks

FOA: Implementation of the Commons Frameworks (FY18)

- Would support investigator-initiated projects to further develop the Data Commons.
 - Propose either an Cooperative Agreement (U01) mechanism or Other Transactional
- Leverages and expands upon resources developed in “Large, High-Impact Datasets in the Cloud - Populating the Commons” .
- 6-10 awards in FY 18 and FY19
 - Two rounds of competition
 - Awards 2-4 years in duration.

4. Making Existing Projects/Objects FAIR

FOA: Making existing data and tools Commons Compliant/FAIR (FY18)

- Competitive Supplements to existing NIH Awards.
 - Applications would be reviewed by CSR
- Would provide support to existing projects to make their current digital resources FAIR & Commons Compliant.
 - Digital resources would include: data, analytical software, or workflows.
 - Digital resources may be available either in a cloud environment or accessed via an API.

Commons Compliance/FAIR guidelines: Proposed Commons Implementation Concepts_V3.doc

FOA: Making existing data and tools Commons Compliant/FAIR (FY18)

- Grantees in this program would be expected to work with other BD2K components
 - Such as those developing useful tools or resources to support for indexing, metadata, or standards.
- Expect to support 8 - 12 competitive supplements in FY18
 - Awards 1- 2 years in duration
- Program would support current data and tools resources supported by ICs into the Commons.

Thank you

- ADDS Office: Jennie Larkin, Phil Bourne, Michelle Dunn, Mark Guyer, Allen Dearry, Sonynka Ngosso, Tonya Scott, Lisa Dunneback, Vivek Navale (CIT/ADDS)
- NCBI: George Komatsoulis
- NHGRI: Valentina di Francesco
- NIGMS: Susan Gregurick
- CIT: Debbie Sinmao, Andrea Norris
- NIH Common Fund: Jim Anderson, Betsy Wilder, Leslie Derr
- NCI Cloud Pilots/ GDC: Warren Kibbe, Tony Kerlavage, Tanja Davidsen
- Commons Reference Data Set Working Group:
Weiniu Gan (HL), Ajay Pillai (HG), Elaine Ayres, (BITRIS), Sean Davis (NCI),
Vinay Pai (NIBIB), Maria Giovanni (AI), Leslie Derr (CF), Claire Schulkey (AI)
- RIWG Core Team: Ron Margolis (DK), Ian Fore, (NCI), Alison Yao (AI),
Claire Schulkey (AI), Eric Choi (AI)
- OSP: Dina Paltoo, Kris Langlais, Erin Luetkemeier, Agnes Rooke,

Vivien Bonazzi

bonazziv@mail.nih.gov

Stay in Touch



QR Business Card



LinkedIn



Slideshare



@Vivien.Bonazzi



Blog
(Coming soon!)

Populating the Cloud Commons
Trans NIH Data Management Task Force
Google doc: <http://bit.ly/2adZgBk>

FY17 Overall Commons Implementation Plans
Proposed Commons Implementation Concepts_V3.doc