



# BD<sub>2</sub>K

*and why we care about harmonized terminologies*

Jennie Larkin PhD

Pediatric Terminology Summit  
July 13, 2015

[illegible]

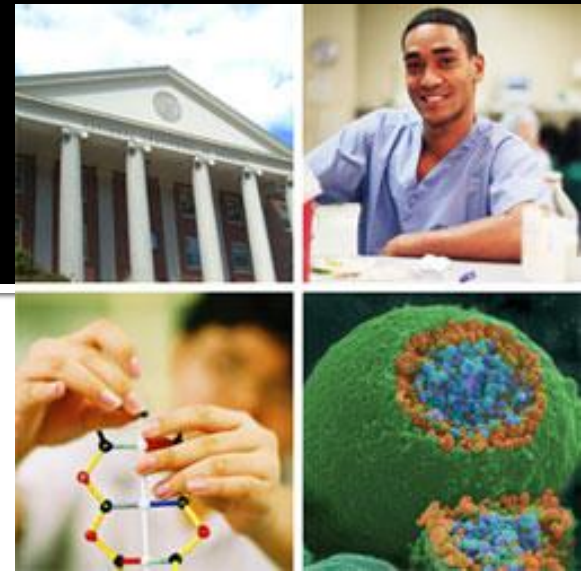
# ADDs & BD<sub>2</sub>K

**ADDs:** Office in the NIH OD that is the home to trans-NIH Data Science activities.

- ADDs: Office of the Associate Director for Data Science
- **BD<sub>2</sub>K:** trans-NIH extramural program advancing data science R&D and training
  - BD<sub>2</sub>K: Big Data to Knowledge
  - Supported and implemented by all 27 IC
- **Collaborations:** Trans-agency, international, and PPP
  - DOE, NSF, NOAA, ELIXIR, Wellcome Trust...

# Mission Statement

To use data science  
to foster an open *digital ecosystem*  
that will accelerate **efficient,**  
**cost-effective** biomedical research  
*to enhance health, lengthen life, and*  
*reduce illness and disability*



# Biomedical Big Data Opportunities

- **Maximize** the **potential** utility of existing data.
- Develop **new methods** to collect and analyze data.
- Identify and enable **new directions** for research.
- Strengthen and diversify the biomedical research workforce.



# ADDS Strategy



- **Discovery and Innovation**

*Enabling major scientific discovery and innovation through the BD2K Initiative*

- **Workforce development**

*Strengthen the ability of a diverse biomedical workforce to develop and benefit from data science*

- **Policy and process**

*Contribute to policies & processes involving data that further the NIH mission*

- **Leadership**

*Further visibility of NIH leadership in data science by the public, DHHS, USG at large, and international funders*

- **Sustainability**

*To foster a sustainable, efficient, and productive data science ecosystem*

# FY14: kickoff of BD2K funding

- **Resource Discovery**
  - Standards
  - Cloud/HPC & Commons
  - Sustainability
  - **Big Data Centers**
  - Software and Analysis Methods
  - **Training & Workforce Development**
  - Clinical Big Data
- 
- Data Discovery Index: bioCADDIE developing the prototype
  - 12 Centers focused on Big Data challenges and solutions
  - First set of training awards (R25s and K01s)

# FY15: BD2K ramping up activity

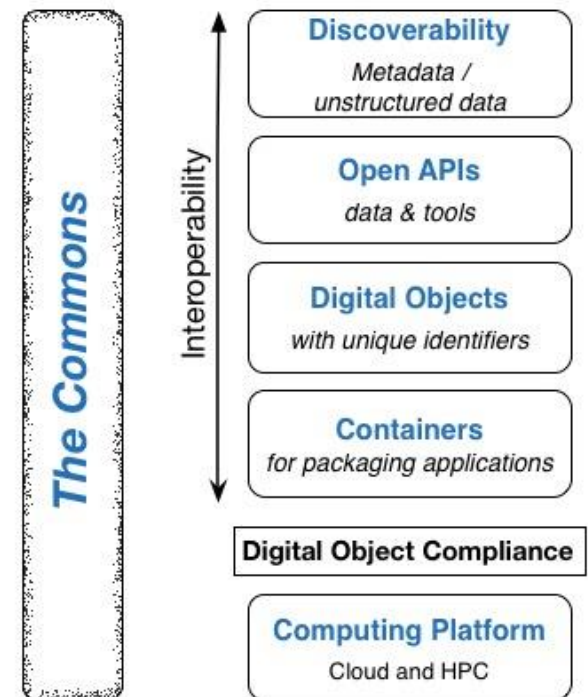
- Resource Discovery
  - **Standards**
  - **Cloud/HPC & Commons**
  - **Sustainability**
  - Big Data Centers
  - **Software and Analysis Methods**
  - **Training & Workforce Development**
  - Clinical Big Data
- Focusing efforts on developing components of the Commons. Supplements to pilot activities in use of cloud, commons components (e.g., open APIs, etc).
  - Supplements to promote more interoperable, sustainable data resources.
  - Supplements to support collaborations across BD2K awardees.
  - Released a contract for a biomedical standards coordinating center.
  - New awards on targeted software development, more training.



# FY16: Focus on implementation

*FY16 will be a critical year:  
activities started in FY14 and FY15 will start to bear fruit.*

- **Cloud Broker Model:** contract will have several phases to test how cloud resources can better be provided to biomedical researchers.
- **Supplements** will result in data and tools being deployed in the cloud: testing feasibility and use.
- **New expertise and approaches:** FOAs, challenges, and prizes.
- Focus on **Clinical Data Science** challenges: partnering with PMI and other activities.



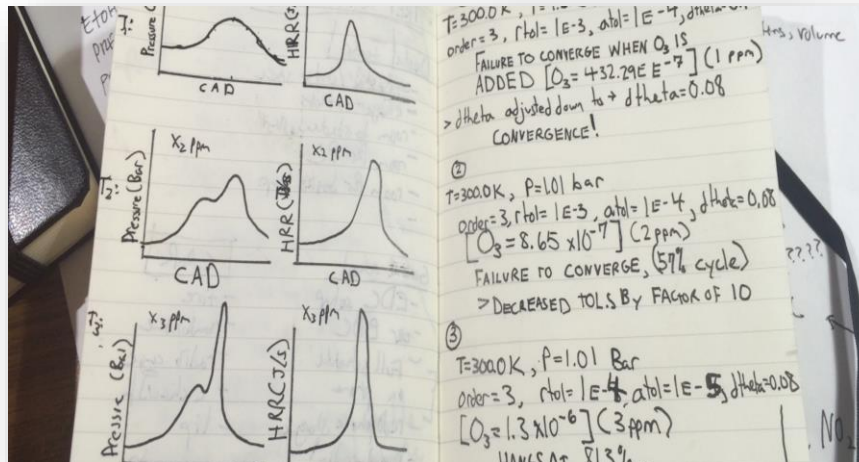
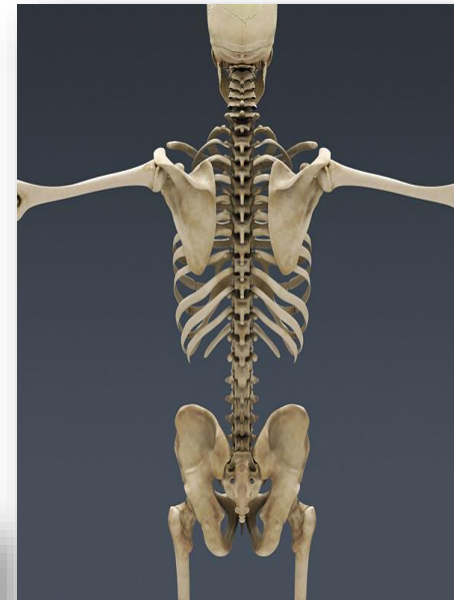
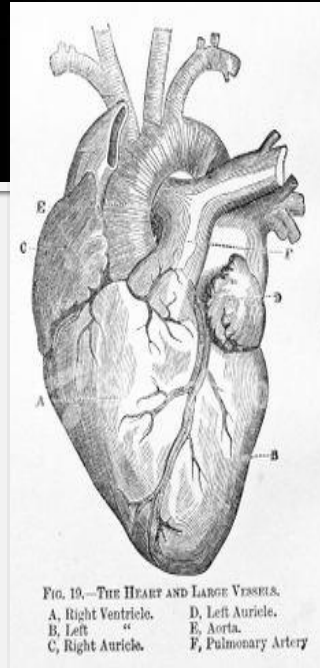
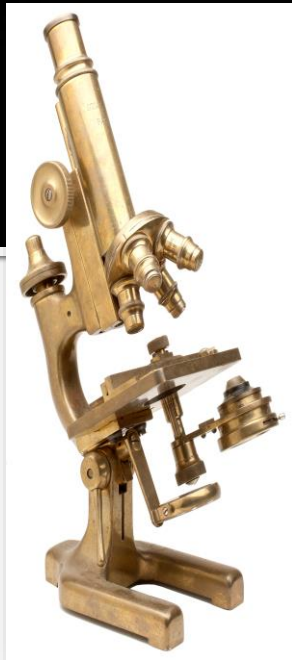
# Overall Goals by 2020

- Enable major scientific discovery through the BD2K initiative
- Establish and provide evidence of a **more sustainable, efficient and productive data science ecosystem** both internal and external to NIH
- Establish and provide evidence of a **well-trained and diverse workforce** able to use and develop biomedical data science tools and methods
- Build upon NIH's **leadership and reputation** in data science

# Why we care about terminologies









यै प्रवर्त्तमानं योगं नृणां वेद्यां सद्योः स्वैः ॥ पदान्ते देवपातनां विना गोमिश्र  
 मीद्व्युत्तं छंदः संपदं तांतां यो र्धमन्येत पादतः ॥ प्रायो र्थो वृत्तमित्येते पाद  
 ज्ञानस्य हेतवः ॥ विशेषः सन्ति पातेषु पूर्वपूर्वपरंपरं ॥ ३ ॥ अनुदात्ततुपादा  
 दोऽनोवर्त्तमानं विद्यते पदं पादादावनुदात्ततुपादौ यद्व्यतिरेकं दिहोदितं ॥ वश  
 स्तीक्ष्णं स्तीक्ष्णं कमानिष्टदिति च चोनेति पूर्वाणि सर्वाणि मधुच्छंदस्य तावत् ॥  
 स्तोमशब्दपरं धातुसतशब्दपरं सिधत्तुवेत्तराणां भुज्यते च पंक्तयुतं  
 तं ॥ प्रेक्ष्यतेति चैतन्मिच्छते पादौ ॥ स्तिपंचमं ॥ सर्वांनुदात्तं षट्स्य स्वादि  
 तश्च लुट् ॥ १ ॥ पादौ गान्धर्वैराजावष्टाक्षरदशाक्षरौ ॥ एकाक्षरी द्वाक्षरी  
 नौ विधात्रैष्टुभजागतौ ॥ ४ ॥ वृद्धिश्चाभिष्टयोरेषा लघुपोतममक्षरांशुर्व  
 वेतरयोस्तु तद्वत्तं छंदसां प्राहुः ॥ एतैश्छंदसां भिन्नैर्त्तैर्त्तैर्वा एयं चैरतोत्तरांशु  
 शः ॥ एतद्विकारा एवान्ये सर्वे लुट् प्रीतिताः ॥ रुमाः ॥ एक एक पदैर्द्वैतेषां  
 द्वौ पादौ द्विपदोच्चारितौ ते तु तेनैव प्रोच्येते मरूपेयस्य पादतः ॥ न दश

लिगड

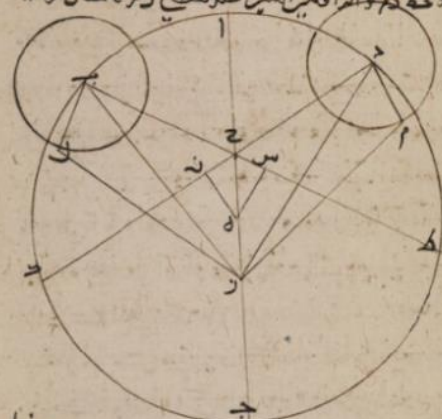
O JEHOVA! Quam ampla sunt opera Tua!



U M,  
K, M D C

O O T

للعامة ذر وناويع كرساويه حدروك كرساويه ايضا اليه من مرز  
فلل الدولير مستاوي كرساويه حدروك كرساويه ايضا اليه من مرز  
رد اذ اسماويه



مرد اذ امساويه  
 انصار الراوية  
 درم و ذلك  
 ما قصدنا  
 لبينهم  
 ع ع ع  
 وليكن  
 انفا لسبب  
 الاصل الذي

عليه علمنا ونعطاه القطر التي يمر بها مركز الدوائر وبالنسبة الى اعدادها  
التي ولنزل ان نقطة آخر ذلك البروج ونقطه بمرز الدوائر الخارج  
المركز التي نعمل للاضلاع ونقطه ج النقطة التي تحولها  
دوير في الفلك الخارج المركز اياما في تلك الدائرة ولنجح  
ايضا خطا بة في المحيطة للمركز المستوي التي لافلك التدوير  
مقارنه الكبد وخطا ح ر ج لدوران الفلك الخارج المركز  
والمركبة العلوية المساوية في السرعة حتى يكون الزوايا التي عند  
ج ب متساوية وذلك ينفلون خط د موازيا لخط ح ر وخط بة  
موازيا لخط ج ح ولينص على خطي ح ر و ج مرز الدوائر الخارج  
ريتان عليها اللذان عليها يكون ذلك التدوير ويران مستقيمة ولسم

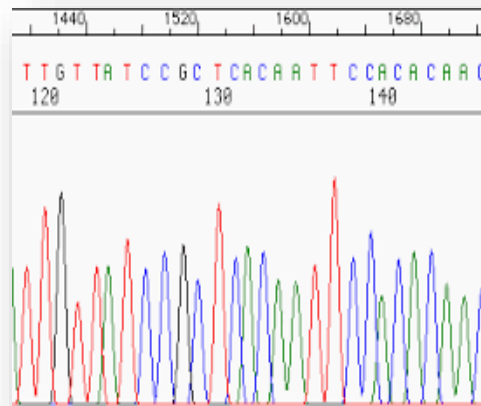
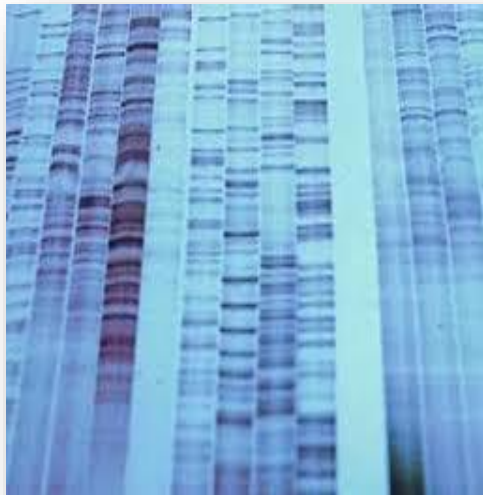
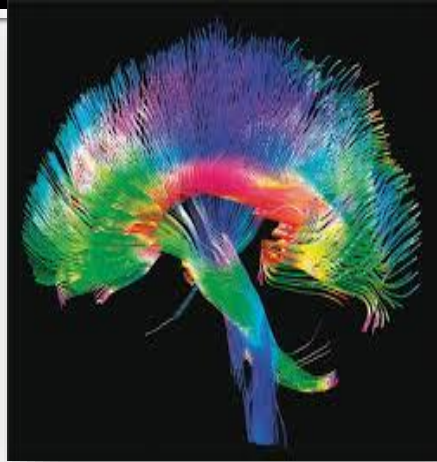
والمستقر في العلم والدين



Secret

dicuntur hinc ad hunc / locum hinc hinc hinc hinc hinc  
 sunt autem, & inueniunt quod dies viginti  
 sunt Arabiorum, sunt decem milia, & sexcenti  
 sunt, & unum dies, & est pars valde, quia quatuor  
 milia quatuordecim sunt, quia ex ea non compre-  
 henduntur, nisi in spacio maiore duobus milibus  
 sunt. Extrahunt ergo in compositionem, qui no-







# The Commons

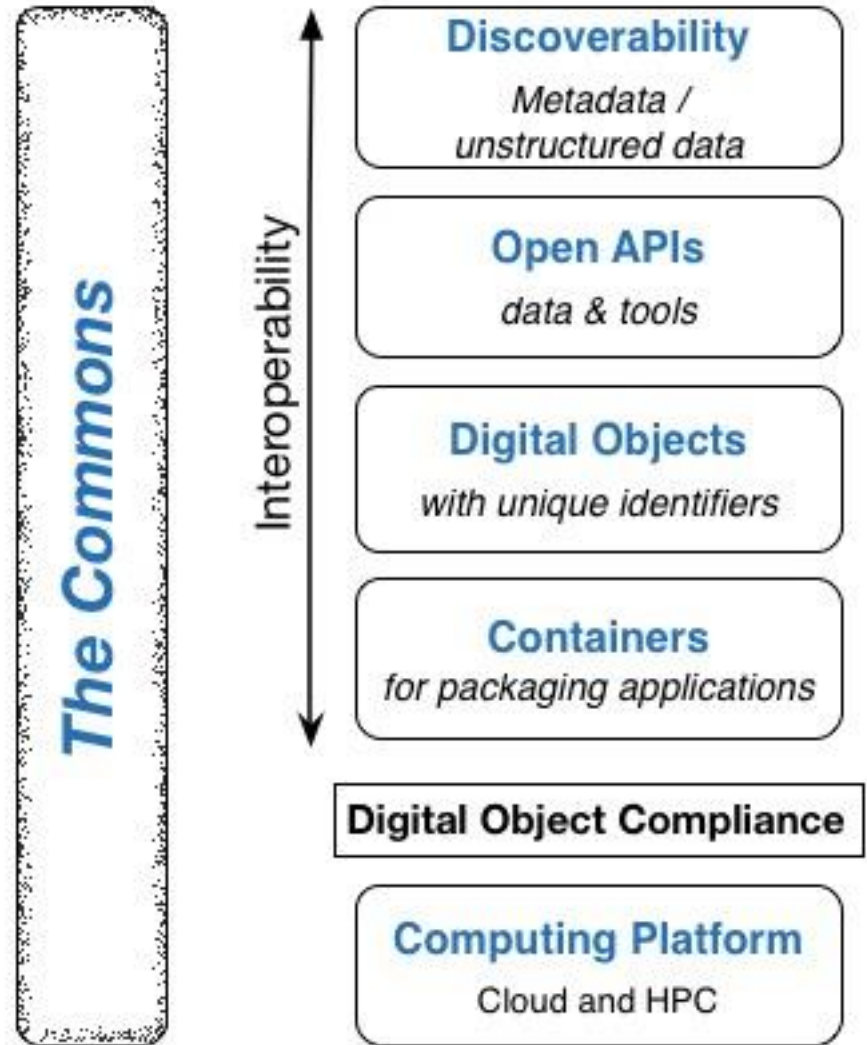
## *Concept*

- *The Commons* is a shared virtual space where scientists can:
  - Find
  - Deposit
  - Manage
  - Share
  - Reuse data, software, metadata and workflows
- An environment to find and catalyze the use of shared digital research objects

# The Commons

## Components

- **Computing environment**
  - cloud or HPC (High Performance Computing)
  - supports access, utilization, and sharing of objects.
- **Methods for Interoperability**
  - enables connectivity, sharing, and access between digital objects.
- **Digital object compliance**
  - describes the properties of digital objects to be discoverable and shareable.



# Metadata

- “data about data”
  - Makes data more discoverable & usable

## UCSC Genome Bioinformatics

[Genomes](#)[Genome Browser](#)[Tools](#)[Mirrors](#)[Downloads](#)[My Data](#)[Help](#)[About Us](#)

### Track Search

The Genome Browser's Track Search feature allows users to find and display tracks of interest quickly and easily.

#### Searching

The track search feature provides users with two search options, "Search" and "Advanced." If multiple terms are entered, only tracks with all terms will be part of the results. We do not currently support either/or searching (e.g. a OR b).

**Search tab:** "Search" is the default tab of the track search feature. To search for a track, enter search terms into the box and click search. This search looks for the user's terms in the track names, descriptions, groups, and metadata. Metadata is information, such as cell line, experiment type, and treatment, about ENCODE data.

**Advanced tab:** The "Advanced" tab allows users to search for terms in a specific aspect of the track (e.g. the track name, the description, and the track group) and can be especially helpful for finding ENCODE tracks by setting specific criteria in the "ENCODE terms" section of this tab.

# Harmonized Terminologies

- Help assure more consistent metadata
- Ontologies and controlled terminologies
- Common Data Elements (CDEs)





Questions??

**Thank you**

Jennie Larkin: [jennie.larkin@nih.gov](mailto:jennie.larkin@nih.gov)