

NHGRI FUNDING FOR DATA RESOURCES

Valentina Di Francesco

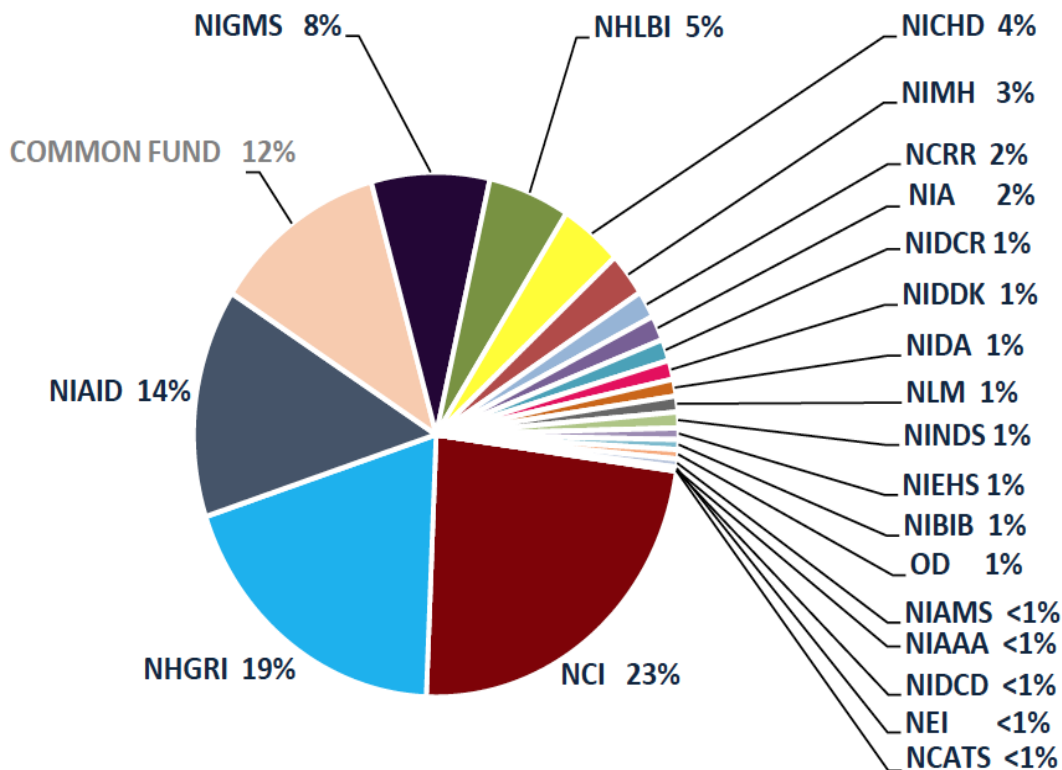
NHGRI Computational Genomics and Data Science Program

SFSP0 Meeting, Strasbourg 11/18/2016

Background Information

- FY15 funded data resources, knowledgebases and informatics platforms: ~\$55M or 15% of the NHGRI extramural budget.
 - Data resources/ knowledgebases (\$40M): MODs, Reactome, GO Consortium, Uniprot, Pathway Commons, Gencode, Clingen, etc.
 - Informatics platforms (\$15M): Bioconductor, GenomeSpace, Galaxy, GenomeBrowser, etc.
 - Not included: Data coordination centers of various funded programs.

Estimated Extramural Data Repository Spending Across NIH FY2007-FY2014, Gross Totals*



Number of Extramural Data Repositories Sampled

Gross Total Spent (Unadjusted)

Gross Total Spent (Inflation-Adjusted)

131

\$1.34 Billion

\$1.27 Billion (2014 USD)

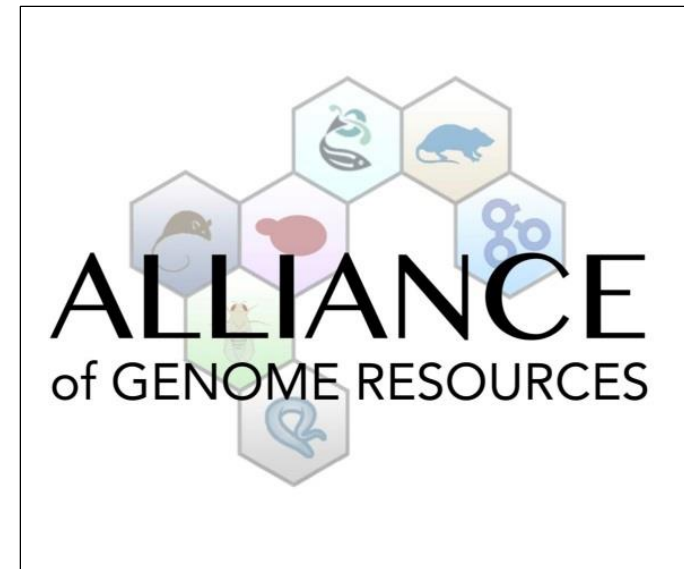
*Please note that some repositories are funded in a hybrid scheme involving multiple institutes or funding mechanisms, which confounds subtotals. Sample data last updated 10-2016



National Institutes of Health
Office of the Director
Data Science at NIH

Alliance of Genome Resources

- Members of the AGR
 - NHGRI-funded: MGD, FlyBase, SGD, ZFIN, WormBase, GO Consortium
 - NHLBI-funded: RGD
- Supplement request submitted in July 2016 to establish the AGR under the WormBase parent grant with subcontracts to other resources
- Awarded:
 - \$2.6M TC in Y1
 - \$2.5M in Y2 in Y3



NHGRI's Goals of the MODs Reorganization

- Facilitate access and use of these resources
- Continue to support the value and services provided by the individual resources
- Transition the resources to a more effective and sustainable funding model
- Repurpose NHGRI funds towards other needed genomics-based data science research.

Prioritization and Tradeoffs Needed

- NHGRI provided \$24.8M to the MODs, the GO Consortium and the AGR. That corresponds to ~50% of HG budget set aside for data and informatics resources.
- Prioritize the long term sustainability of the resources supported by NHGRI
- Tradeoff between resources sustainability and establishing new funding opportunities

Funding Plan for MODs and AGR

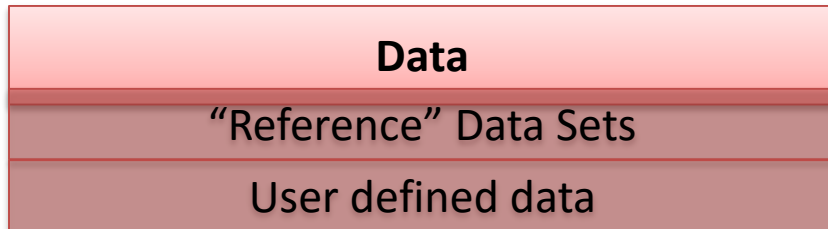
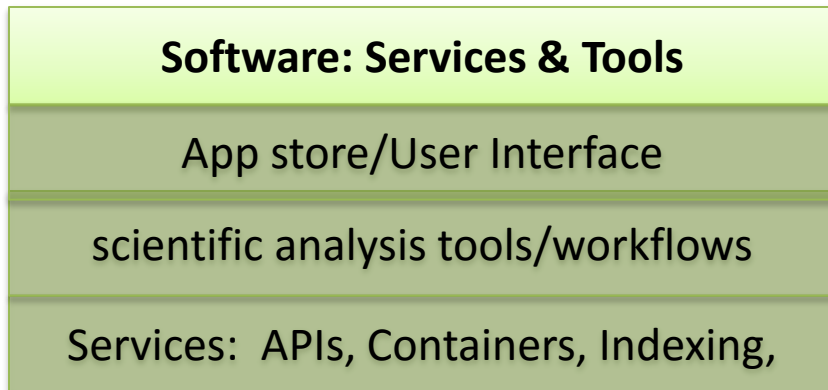
NHGRI awarded

- in FY16 \$2.6M in Y1, \$2.5M in Y2 (FY17) and Y3 (FY18)
 - Primarily personnel costs for 10 FTEs
 - Expected savings in software development and maintenance of approximately 12 FTEs at the end of the 3-year period of support
 - FY16: Total # of FTEs in 6 resources: ~120

NHGRI Plans

- FY18 and beyond: Open competition
- In FY17 budget cuts start and increase gradually over 4 years up to 30% budget reduction in FY20.

The NIH Data Commons Framework



Digital Object Compliance

This text is written vertically inside a thin black rectangular border, positioned between the 'Software' and 'Data' layers.

- Treats products of research - data, software, metadata, workflows, papers etc. - as digital objects
- Digital objects exist in a shared virtual space
 - Deposit, Manage, Find, Share, and Re-Use digital objects
- Conforms to **FAIR** principles:
 - F**indable
 - A**ccessible
 - I**nteroperable
 - R**eusable

<https://datascience.nih.gov/commons>

AGR and BD2K

AGR as a pilot for the NIH Commons and data resources sustainability

- The AGR will embrace Cloud-based technologies and ensure its data and sharing practices conform to FAIR principles for research resources.
- NIH Commons
 - The AGR aims to support findable, accessible, interoperable, and reusable (FAIR) data on a shared Cloud-based platform
- Sustainability of NIH data resources
 - These resources support biomedical research across NIH
 - They also support international biomedical research
- Received approval by the NIH BD2K Multi Council Working Group for 50% co-funding by BD2K in FY17 and FY18

Acknowledgements

- **NHGRI Informatics Group**
 - Peter Good
 - Chris Wellington
 - Jeff Schloss
 - Kevin Lee
- **ADDS/BD2K**
 - Vivien Bonazzi – Commons
 - Allen Dearry & Susan Gregurick – Sustainability WG
 - Jennie Larkin and Phil Bourne