

Concept Clearance:

Enhancing Diversity in Cloud-based Genomic Data Science Education (U24/R25)

Shurjo K. Sen and Valentina Di Francesco

Division of Genome Sciences

September 13, 2021



National Human Genome
Research Institute

—
The **Forefront**
of **Genomics**[®]
—

Background

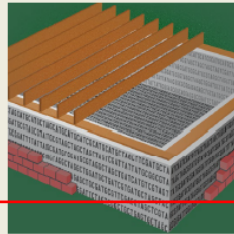
NHGRI 2020 Strategic Vision

Box 2

Sustaining and improving a robust foundation for genomics

Genome structure and function

- Enable the routine generation and analysis of increasingly complex genomic data
- Use evolutionary and comparative genomic data to maximize understanding of genome function



Genomic data science

- Develop new methods and build sustainable data resources for genomics research
- Ensure facile storing, sharing, and computing on large genomic datasets
- Develop integrated knowledgebases and informatics methods for genomic medicine

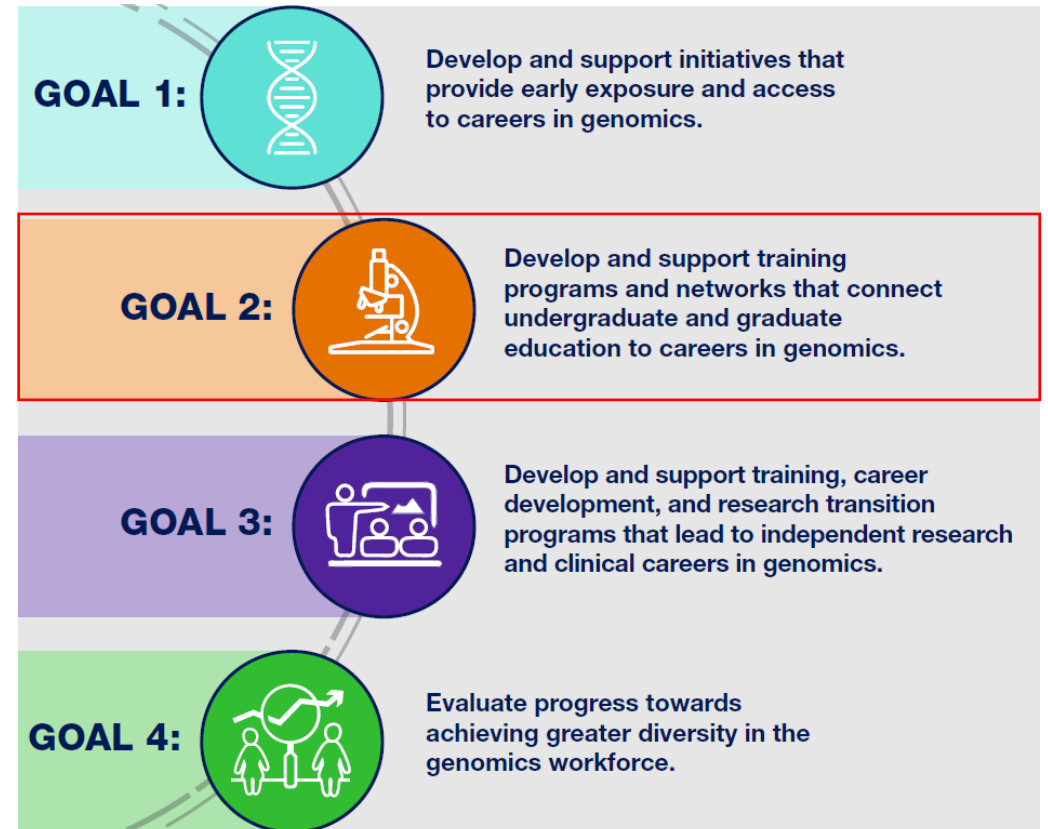
Genomics and society

- Understand the interrelationships between genomics and the social and environmental factors that influence human health
- Empower people to make well-informed decisions about genomic data and develop data-stewardship systems that reinforce their choices
- Increase the genomic literacy of all sectors of society

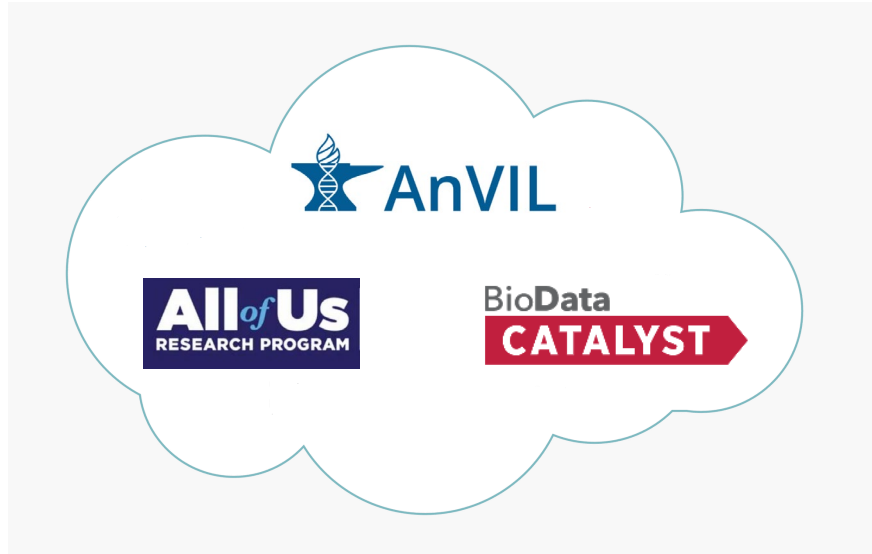
Training and genomics workforce development

- Ensure that the next generation of genomic scientists are sufficiently trained in data science
- Train healthcare providers to integrate genomics into the clinical workflow
- Foster a diverse genomics workforce

NHGRI Diversity Action Agenda



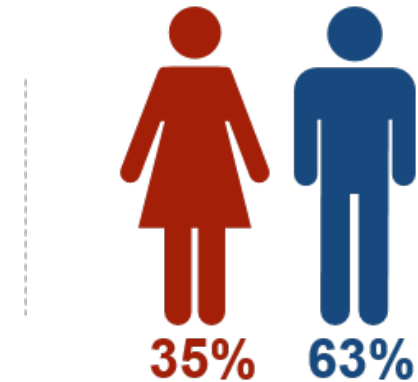
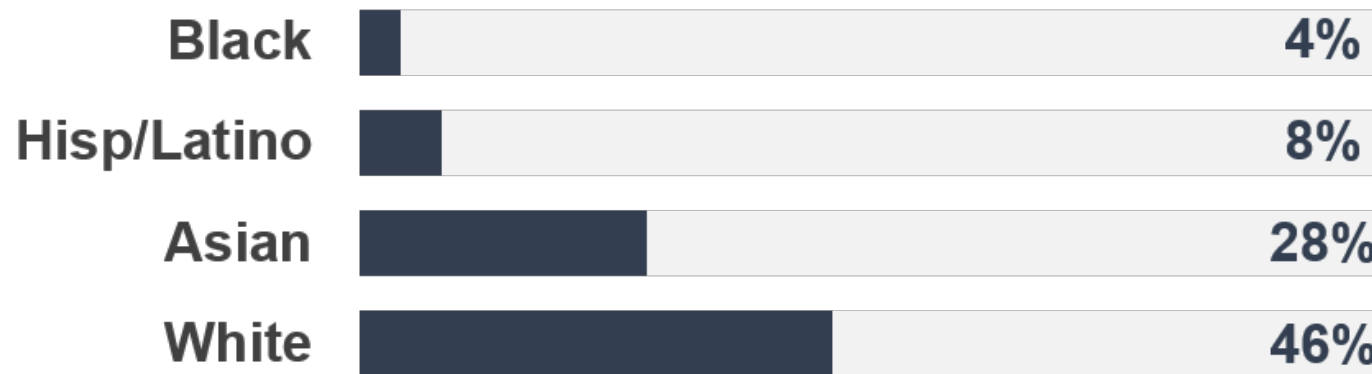
Cloud-Based Genomic Data Science Education



Created by Jeremy Wilburn, licensed under CC BY-NC-ND 2.0

- **Goal:** To facilitate the development of a diverse genomics workforce with training in computational genomics and data science (CGDS).
- **Strategy:** Use NIH-funded cloud platforms as a tool to enhance CGDS education at institutions serving groups underrepresented in the biomedical and data science fields.

Diversity in data science education



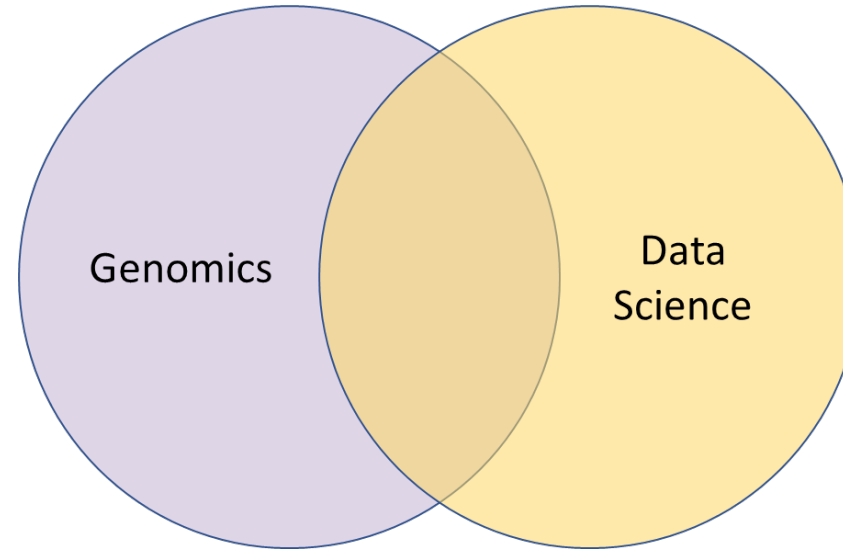
*Not all race/ethnicity are shown. Numbers rounded.

<https://towardsdatascience.com/diversity-in-data-science-a-systemic-inequality-b97a0e953f6e>

Eligible Institutions

Historically Black Colleges and Universities (HBCUs), Hispanic-Serving Institutions (HSIs), Tribal Colleges and Universities (TCUs) & Women's Colleges

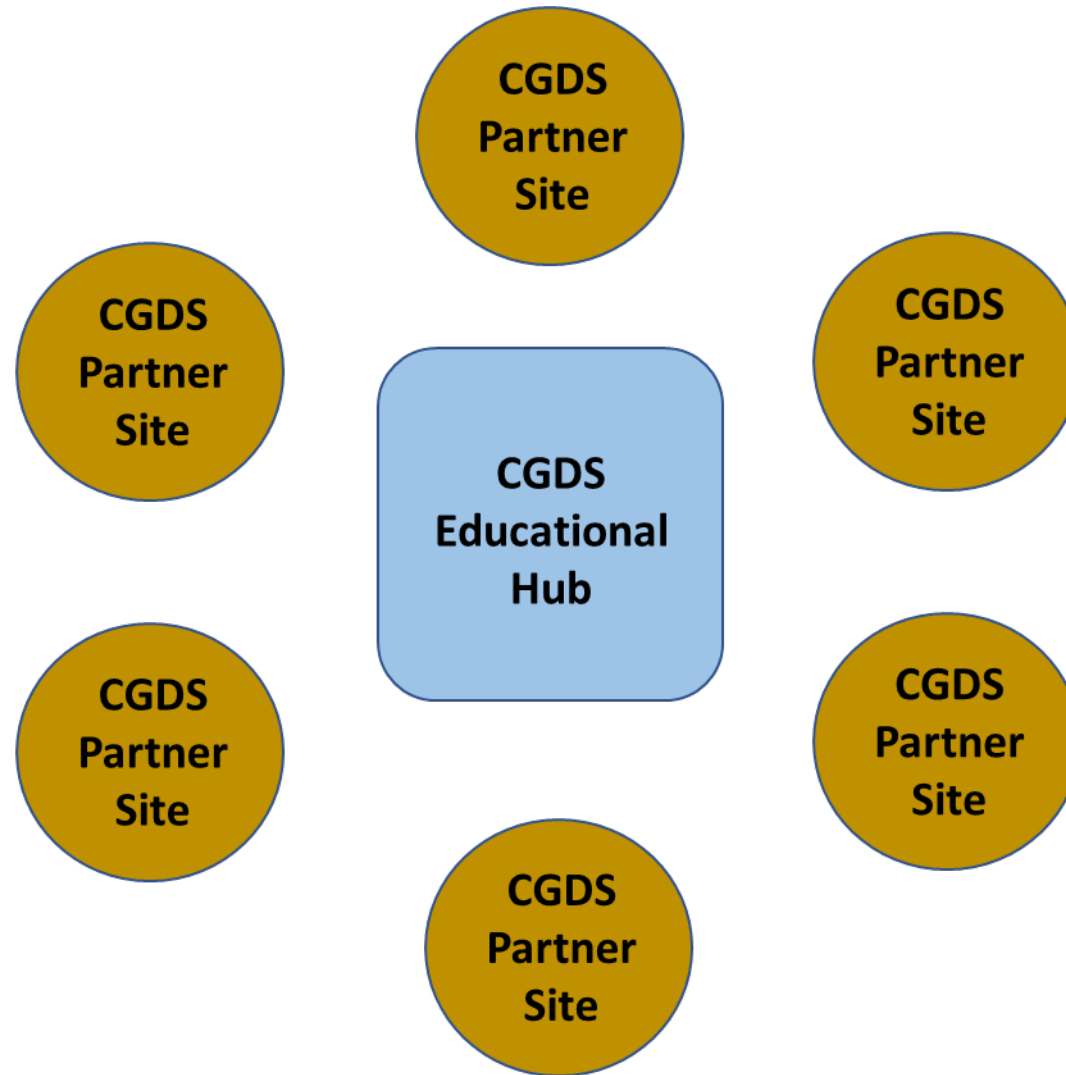
Genomic data science education



Barriers to CGDS education

- Lack of access to computing resources and genomics data sets
- Need for reciprocal expertise in genomics and data science

CGDS Educational Hub and Partner Sites



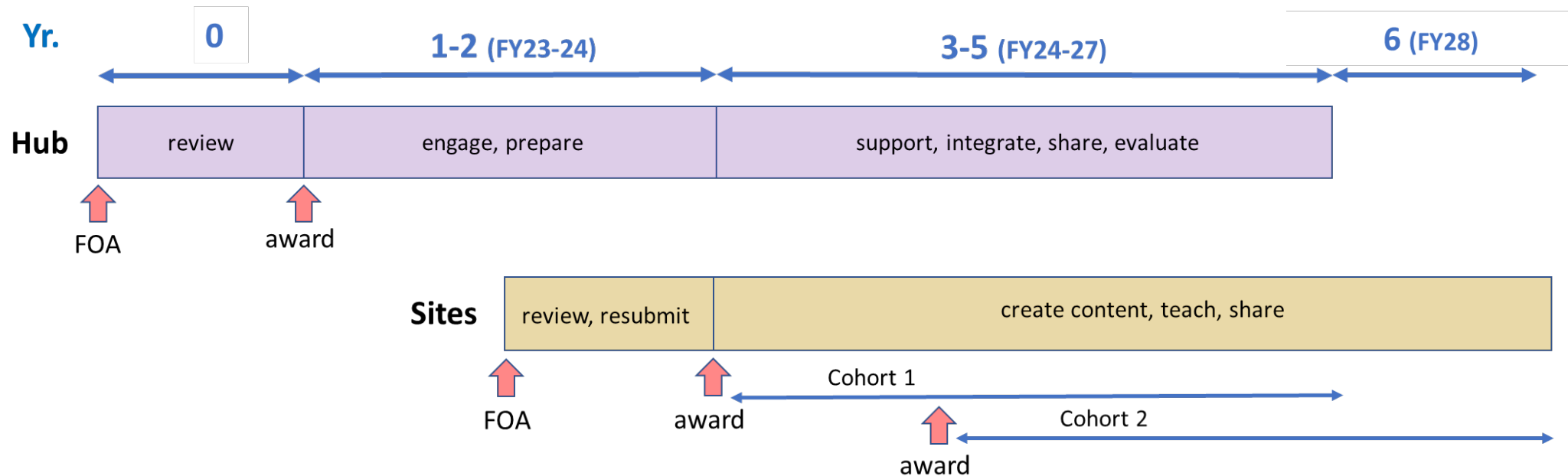
CGDS Educational Hub

- **Scope – facilitator/coordinator for cloud-based CGDS education**
- Host CGDS and cloud computing seminars (incl. travel support)
- Engage stakeholders interested in CGDS education
- Collect feedback on challenges faced in teaching CGDS
- Train-the-trainer and disseminate educational content
- Support funds for student research projects at Sites
- Possible NIH Office of Data Science Strategy co-funding (FY23-27)

CGDS Educational Partner Sites

- **Scope – support faculty at Eligible Institutions to create and adopt CGDS educational content**
- Focus on undergraduate & master's degree levels
- Curricula including classroom courses and hands-on labs
- Use of AnVIL and other NIH cloud platforms by students
- Conduct independent research projects in FY26-28
- Share content developed with educational community
- Supported by NHGRI and partner Institutes/Centers (ICs), FY25-28

Implementation timeline



- Sites supported by Hub in creating CGDS content and using the cloud
- In Years 4-5, Hub integrates and disseminates CGDS content from Sites

Budget

Educational Hub: U24 mechanism, \$1.5M- \$1.8M total costs/yr

Includes opportunity funds in FY26-27

Partner Sites: R25 mechanism, \$150K total costs/yr, 5-8 awards

Limited to three years of support

	FY23	FY24	FY25	FY26	FY27	FY28
Hub	1.5	1.5	1.5	1.8	1.8	
Sites			0.6	1.2	1.2	0.6
total	1.5	1.5	2.1	3	3	0.6
dollars in millions	Grand Total = \$11.7M*					

*Cost for full program; HG contribution may be reduced through co-funding by other NIH Institutes, Centers or Offices

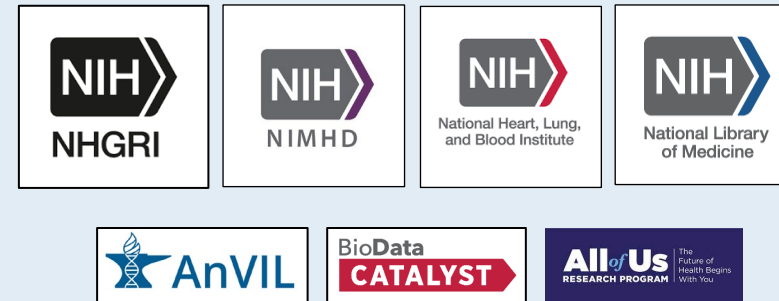
Potential co-funding from other ICs

CGDS Education Hub (U24)



CGDS Partner Sites (R25)

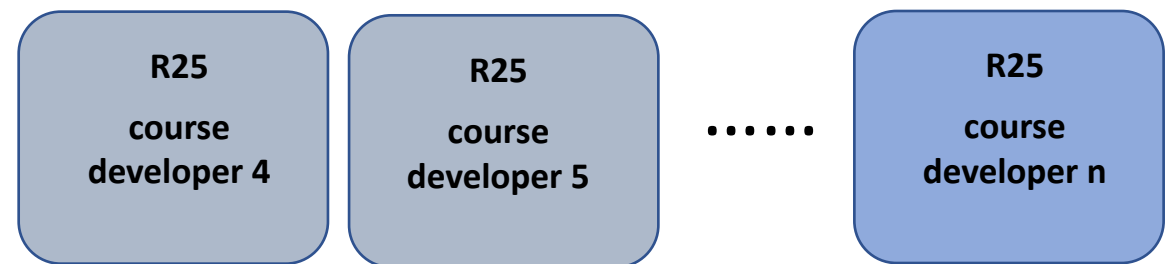
NIH ICs and cloud platforms



IC 1



IC 2



IC_n

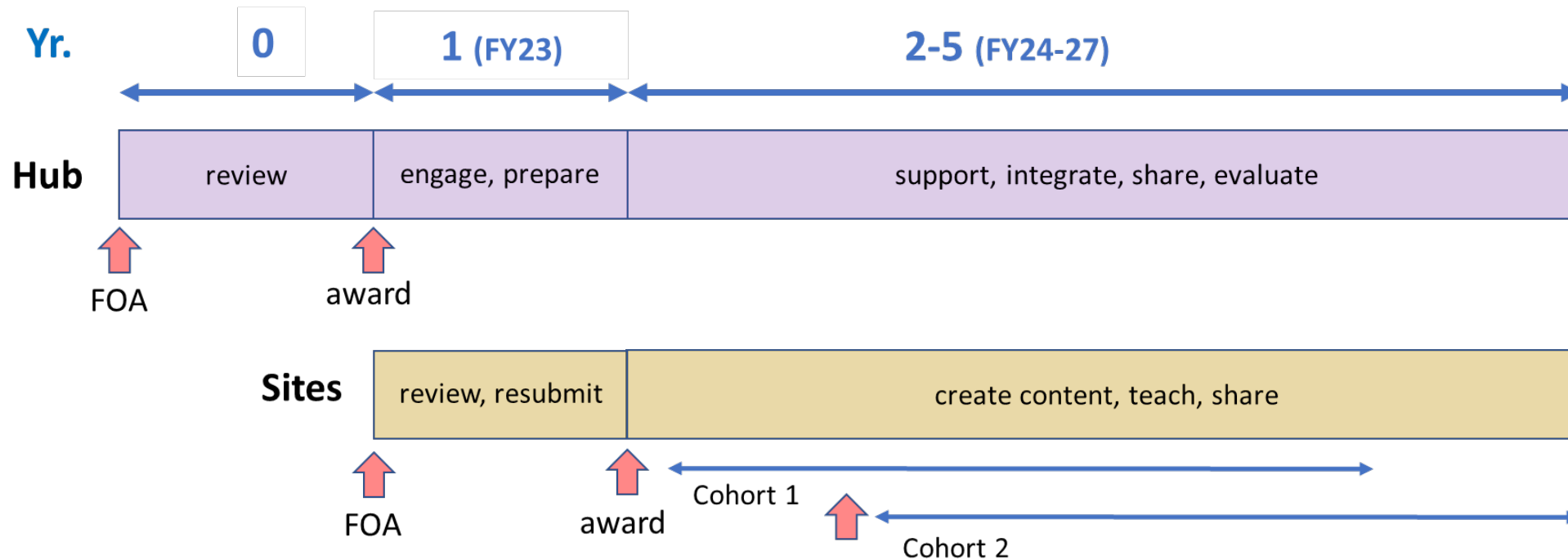
Acknowledgements

Luis Cubano, Tina Gatlin, Carolyn Hutter and ERP Training Team

Questions?



Alternative implementation timeline



- Sites supported by Hub in creating CGDS content and using the cloud
- In Years 4-5, Hub integrates and disseminates CGDS content from Sites