FAIR Data Sharing: A Data Generating Researchers Perspective

Sean D Mooney

Chief Research Information Officer, UW Medicine
Interim Director, UW Institute for Medical Data Science
Professor, Biomedical Informatics and Medical Education

UW Medicine
INFORMATION TECHNOLOGY SERVICES

UW Medicine
BIOMEDICAL INFORMATICS AND MEDICAL EDUCATION

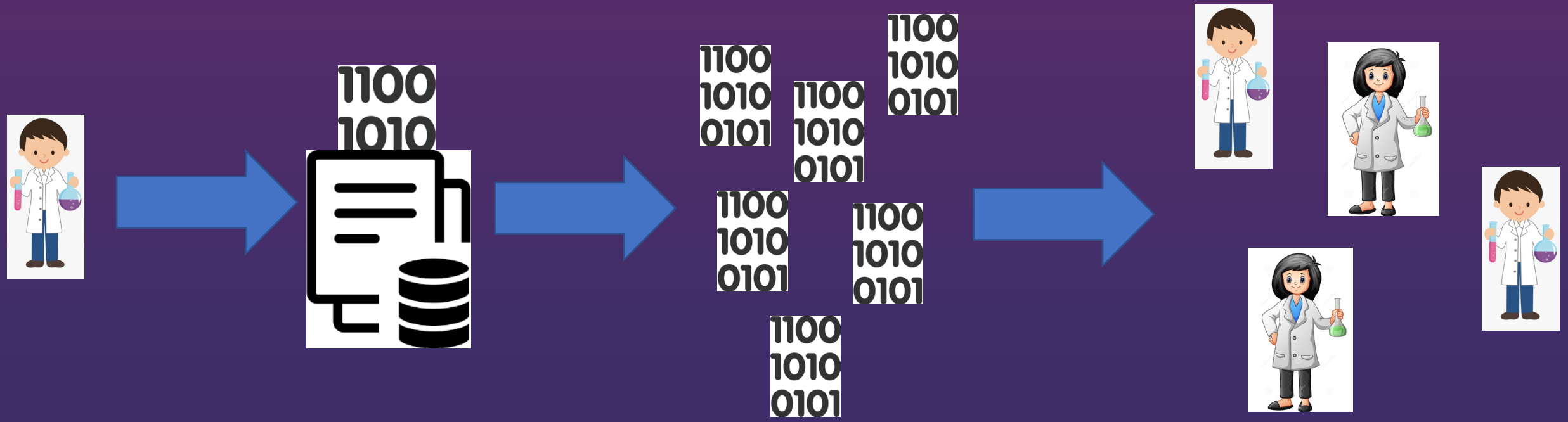
ITHS | Institute of Translational Health Sciences
ACCELERATING RESEARCH. IMPROVING HEALTH.

The investigator

- We all know that scientists are generating data at an unprecedented rate
- There's a consensus among the community that data should often be shared outside of an individual publication
- Sharing of data is maximally impactful when it has some effort to make it more useful with an ecosystem of other datasets
- There remains substantial fluency gaps by investigators on how to actually accomplish that in practice

The Data Workflow

- The same basic workflow applies to a wide variety of situations



Researcher

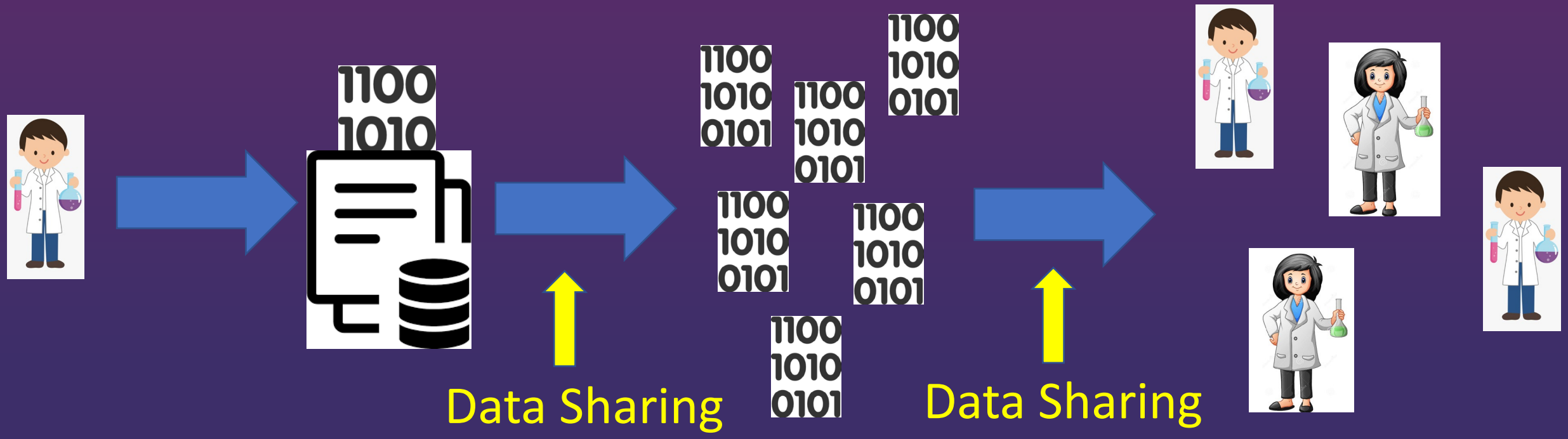
Dataset

Collection of Data
to a central repository

Scientists Using
Data

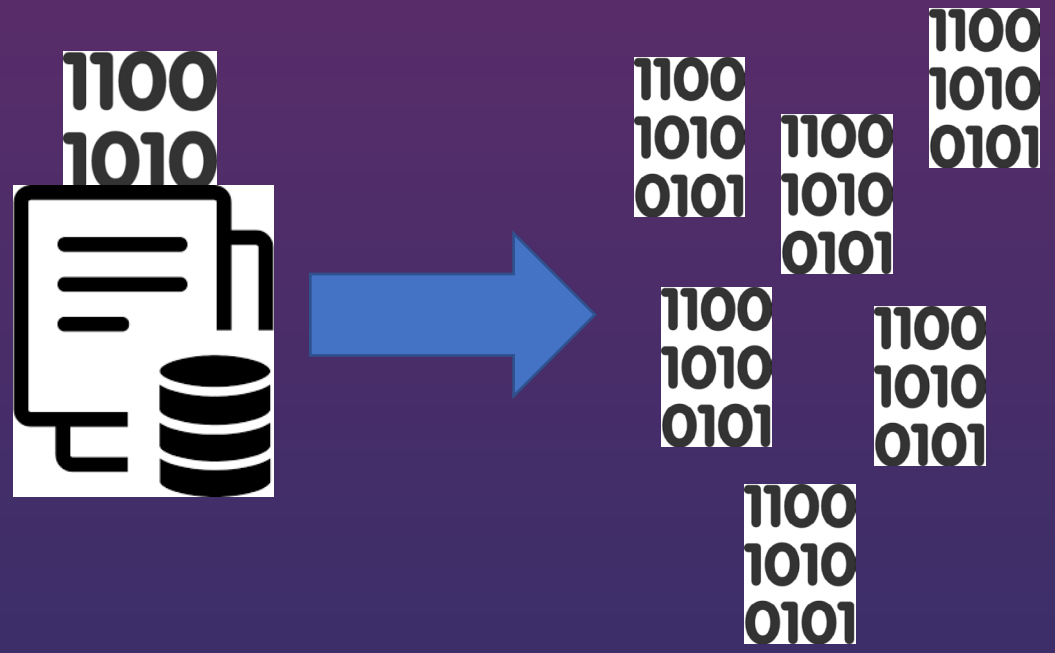
The Data Workflow

- There are two sets of data sharing events



Researchers Contributing to Data Repositories

- Investigators often have both pressure to share data and to make that data consumable by others
- In order to get there optimally, investigators must choose
  - common data elements
  - nomenclature standards/biomedical ontologies
  - standard instruments for clinical research
  - standard information for metadata



Dataset

Collection of Data
to a central repository

FAIR Data

- FAIR – Findable, Accessible, Interoperable, Reusable – is a basic framework for data sharing
- Some things are easy – adopting standard file formats, submitting data, etc.
- Some things are harder – use of biomedical ontologies, common data elements, standardized instruments, standard metadata, etc.
- Some things are very hard – generation of synthetic data, sharing protected data under a data use agreement, etc.

There remains a significant knowledge gap

- Investigators, even sophisticated ones, are not fluent in what FAIR really means

- Challenges:
  - Investigators always lean toward spending money on front end data generation, not necessarily data or analysis
  - Investigators should make informed decisions on standards before the work is done (but often don't)
    - This is the informatics equivalent of biostatistical study design
  - Data sharing requirements are not strongly enforced except in specific consortia with explicit requirements up front

Dataset

Data Collections

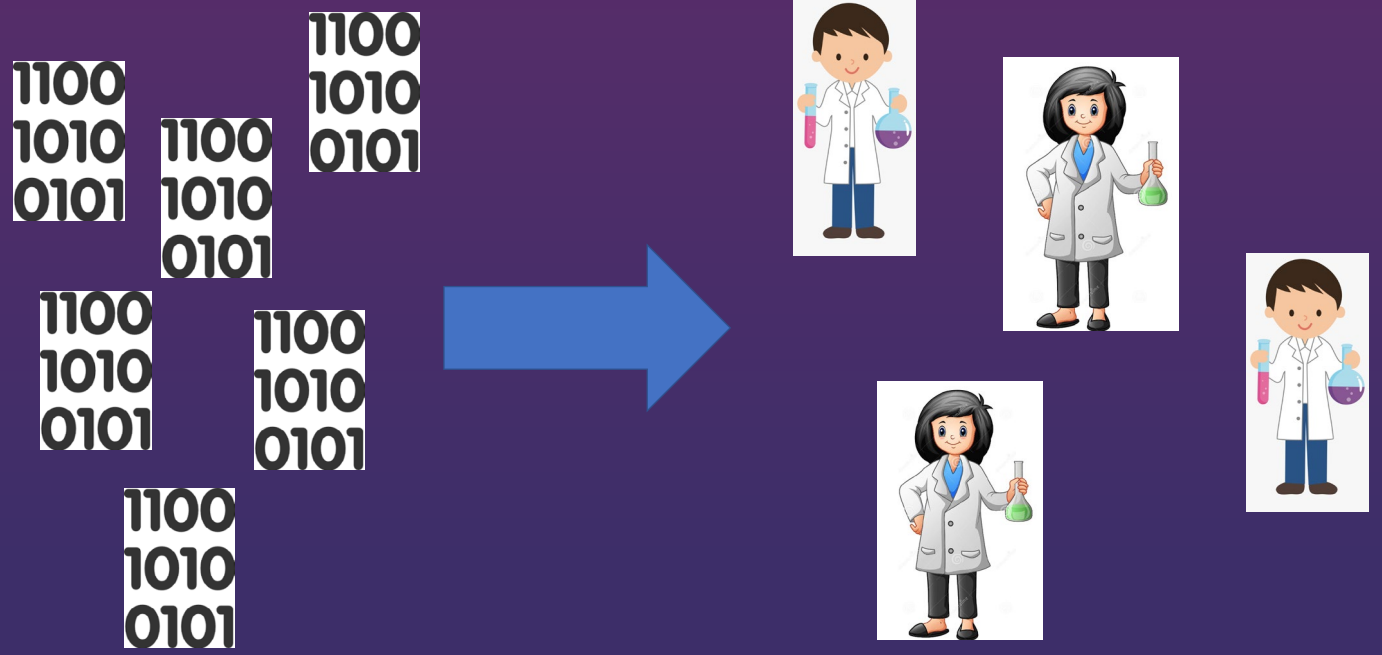
- Data collections come in many forms
  - Coordinating Centers for specific funded consortia
  - Curated knowledgebases
  - Open repositories aimed at data sharing

- Integrated data collections
  - Requires previous work be done before submitting
  - Repositories/investigators must be resourced for curation activities
  - Must have open standards available at the time of curation or those standards need to be developed at potentially substantial cost

FAIR
Collection of Data

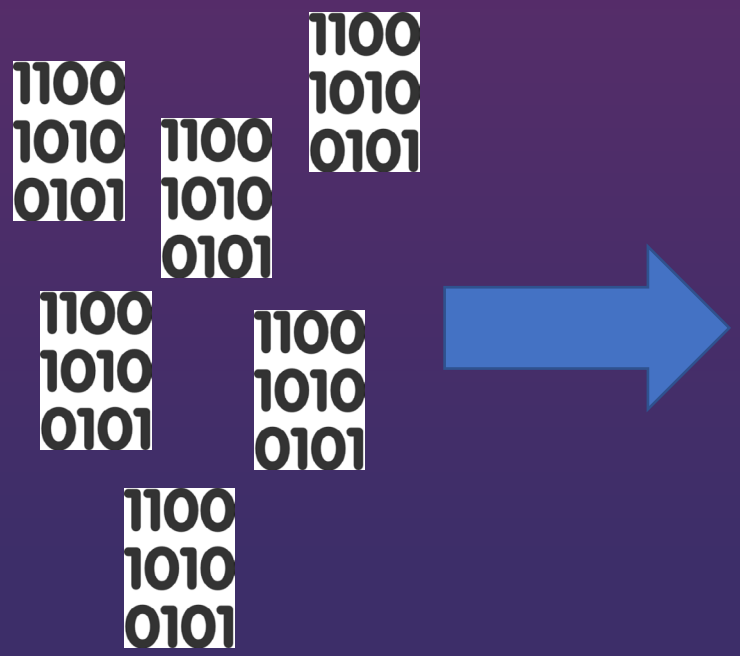
Data sharing to investigators

- Data sharing to investigators must be compliant:
  - Law (HIPAA, GDPR, etc)
  - Requirements of funding (FISMA Moderate)
  - Terms of Data Use Agreement
  - Terms of Informed Consent and IRB (In clinical research)

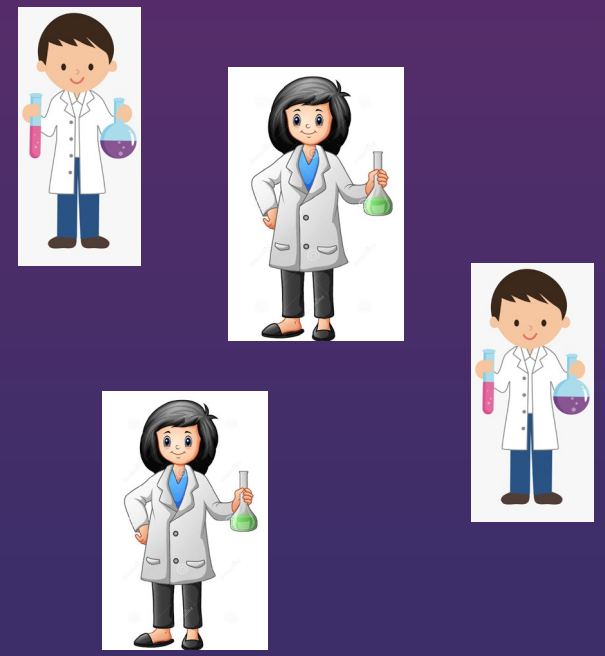


Investigator Use

- Traditional Data Sharing may:
  - Have costs associated with egress
  - Require end user data use agreement
  - Ideally has utility to a broad group
  - Persistent identifiers (e.g. DOIs) continue to be conceptually possible but tricky for investigators and some repositories



Collection of Data



Scientists Using Data

Cloud Computing in Academic Medical Centers

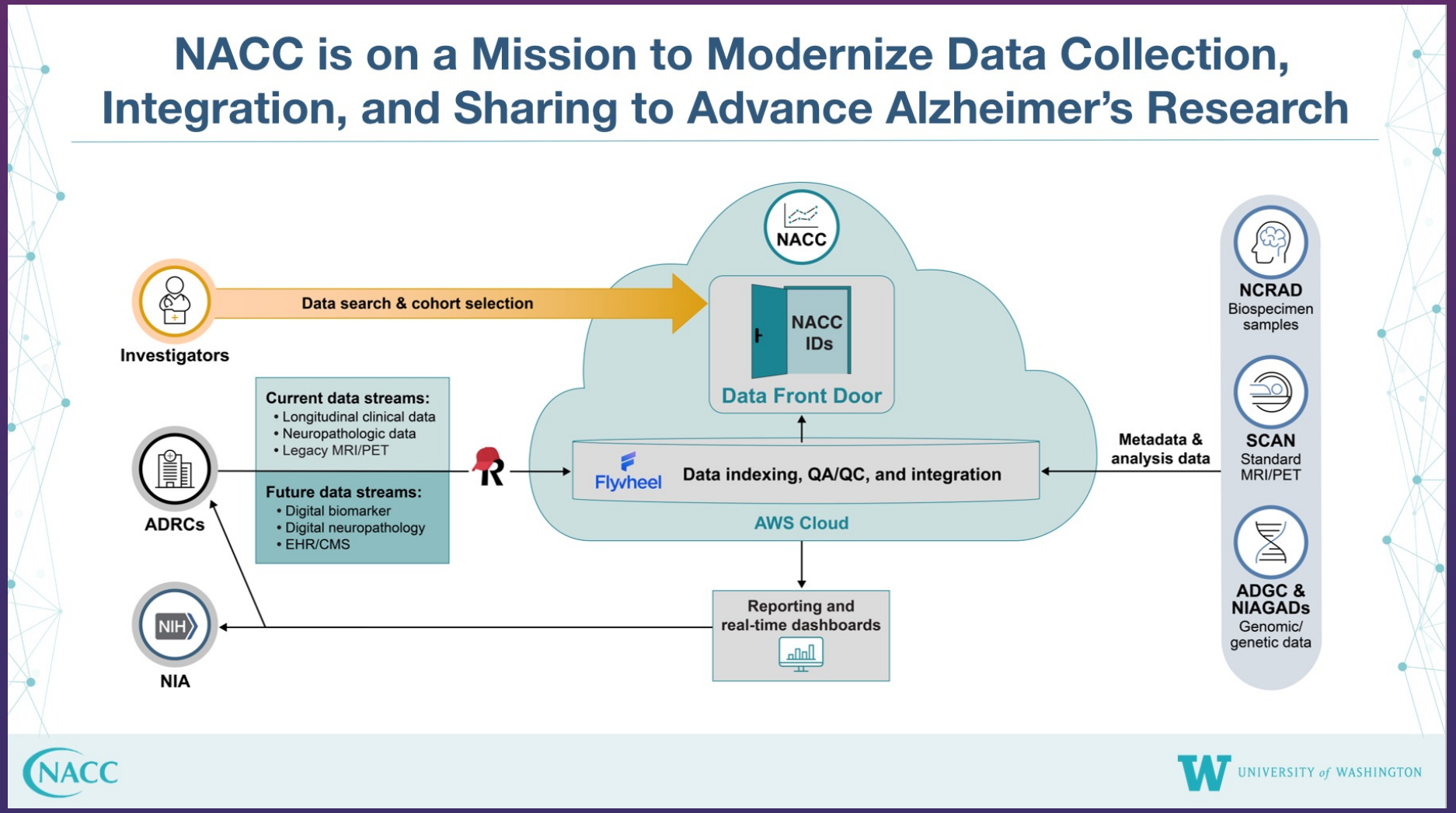
With FAIR data, increasing interest in the Cloud as a facilitator of FAIR data

These repositories require a significant amount of investment and technical resources who can develop in cloud platforms is hard

NIA funded National Alzheimer's Coordinating Center: Moving fully cloud

Uses a platform developed by the company FlyWheel.

Non human subjects, generally, which may change

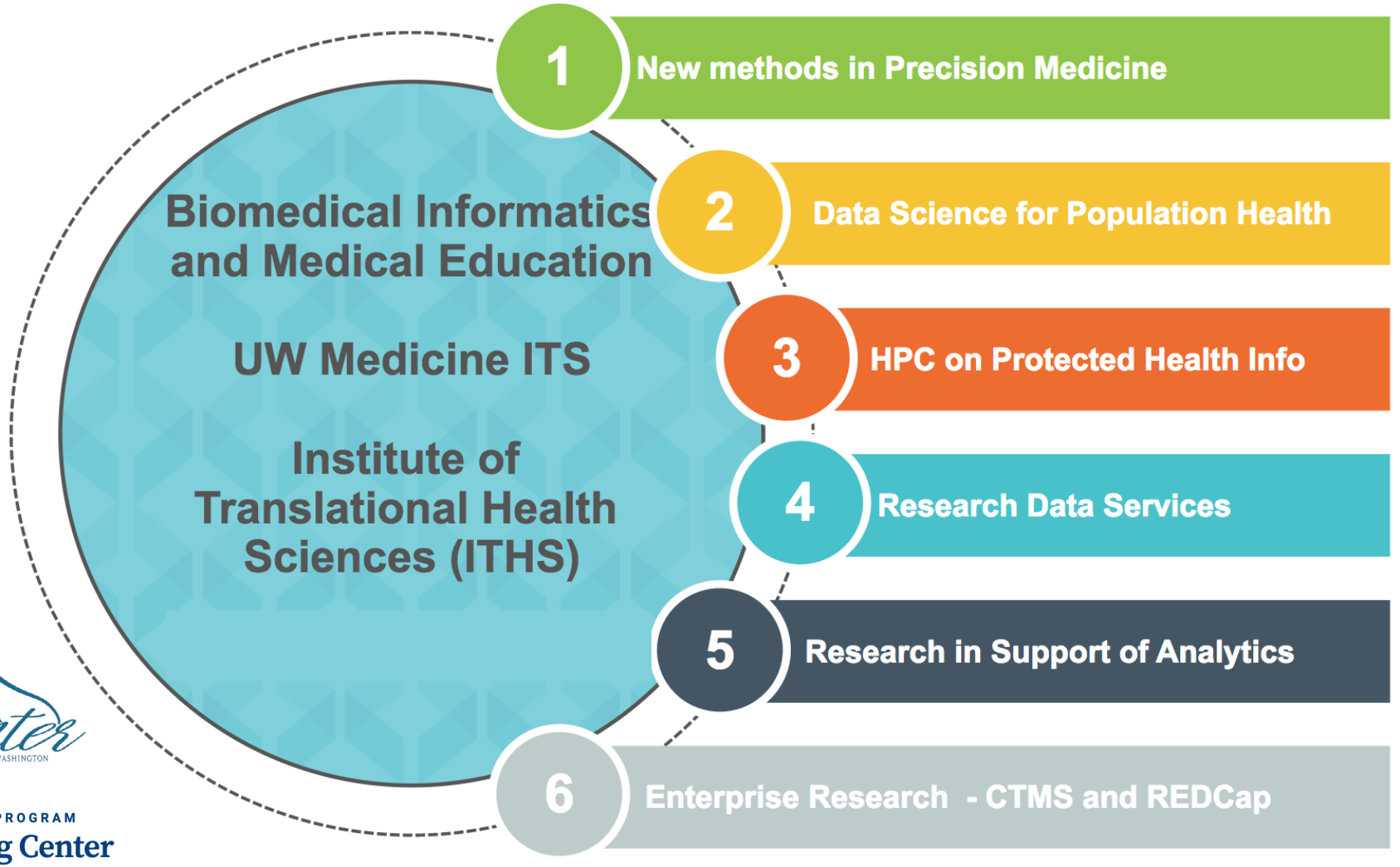


RIT

UW Medicine

Research Information Technologies

Katie Avril, Director of Research IT, UW Medicine



ITHS | Institute of Translational Health Sciences
ACCELERATING RESEARCH. IMPROVING HEALTH.

KIDNEY PRECISION MEDICINE PROJECT

CLEAR center
UNIVERSITY of WASHINGTON

NACC
THE NIA ALZHEIMER'S DISEASE RESEARCH CENTERS PROGRAM
National Alzheimer's Coordinating Center

Medical Data Science Symposium

Register today!

We are hosting a regional workshop/conference
On February 27th and 28th, 2023
UW Hub

Request for Abstracts and Registration for short talks and posters is here:
https://bit.ly/UWMDS2023

MDS
MEDICAL DATA SCIENCE

SCHOOL OF MEDICINE | COLLEGE OF ENGINEERING | SCHOOL OF PUBLIC HEALTH