

Data management and integrity

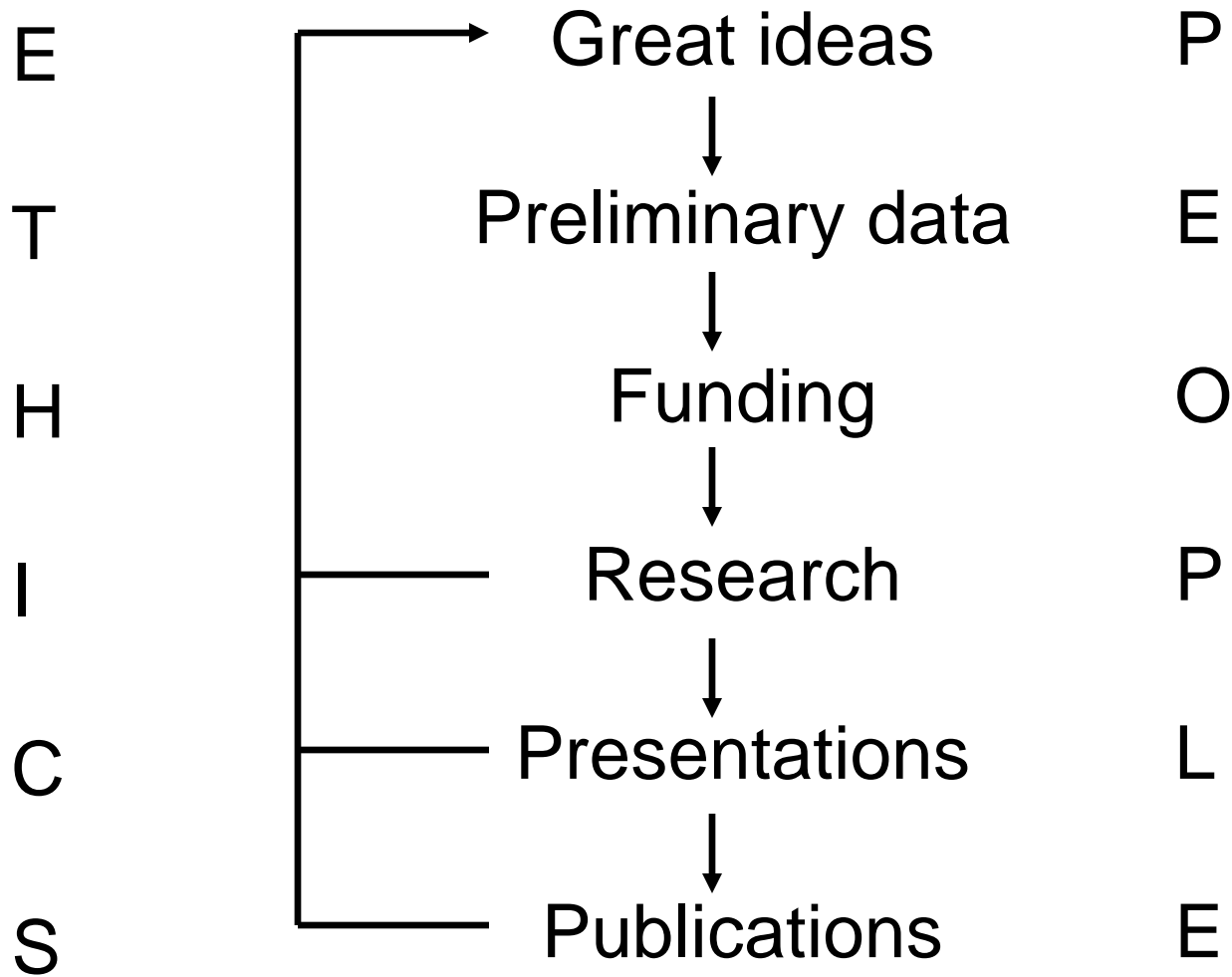
The fate of data before and after
the paper is published

Neal Stewart
nealstewart@utk.edu

Objectives

Understand the backdrop to determine best practices for:

- Data collection
- Data storage
- Data management & integrity



All of the above in research work flow must have integrity or the system breaks

Data collection, storage, management, & integrity



...back in the good ole days...

<http://star-spangledheart.blogspot.com/2011/08/nerd-stuff.html>

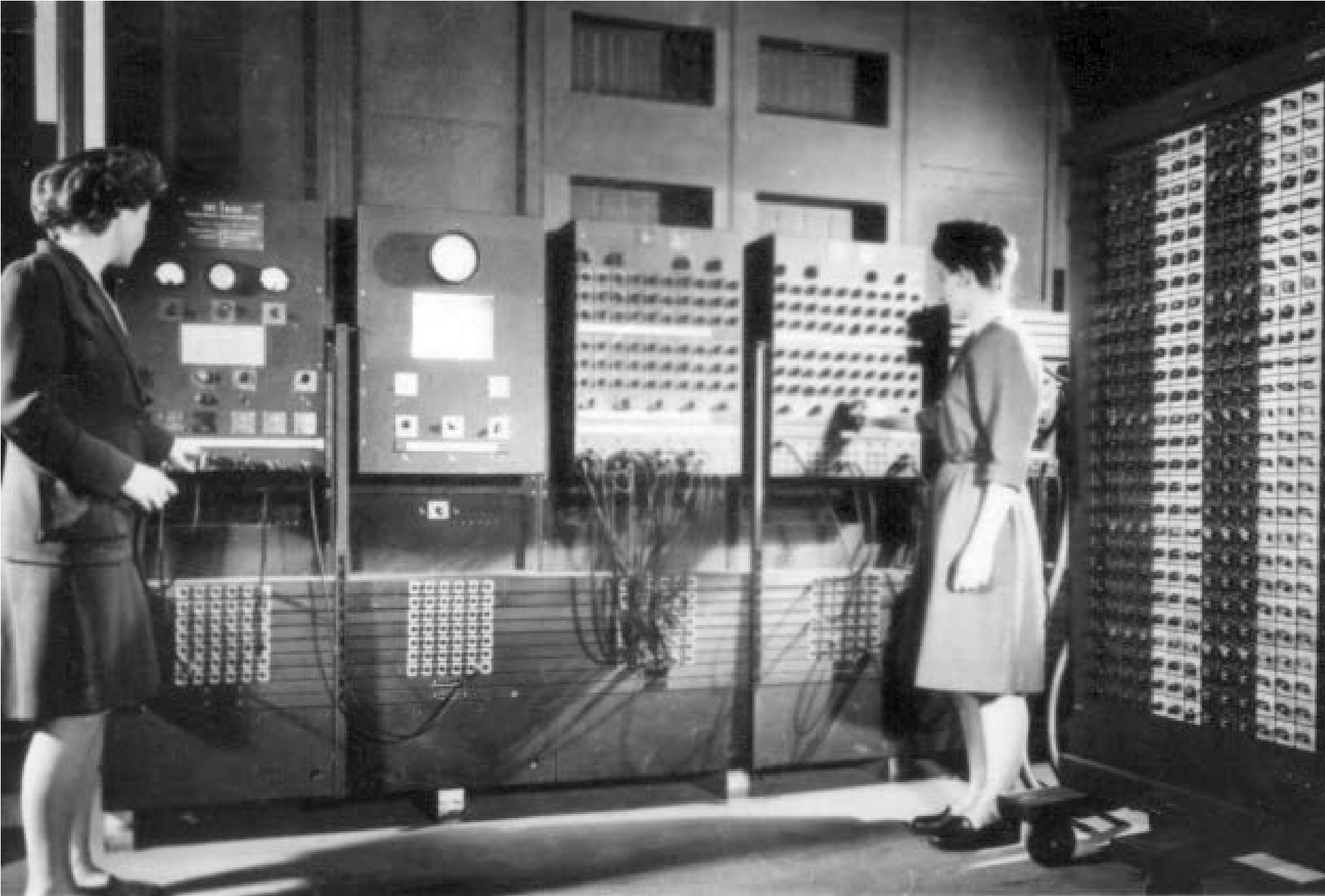
BU T

TODAY

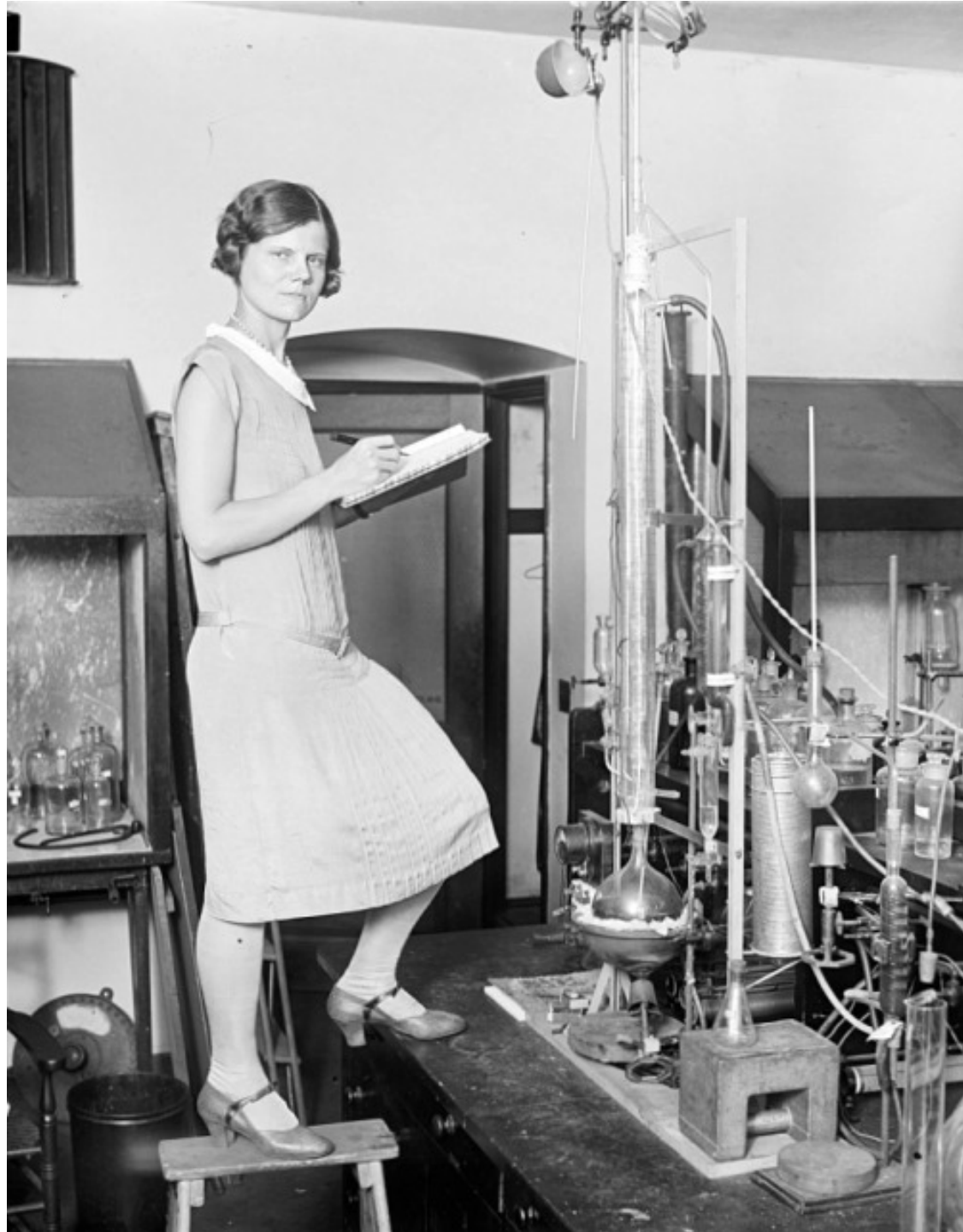
LIFE

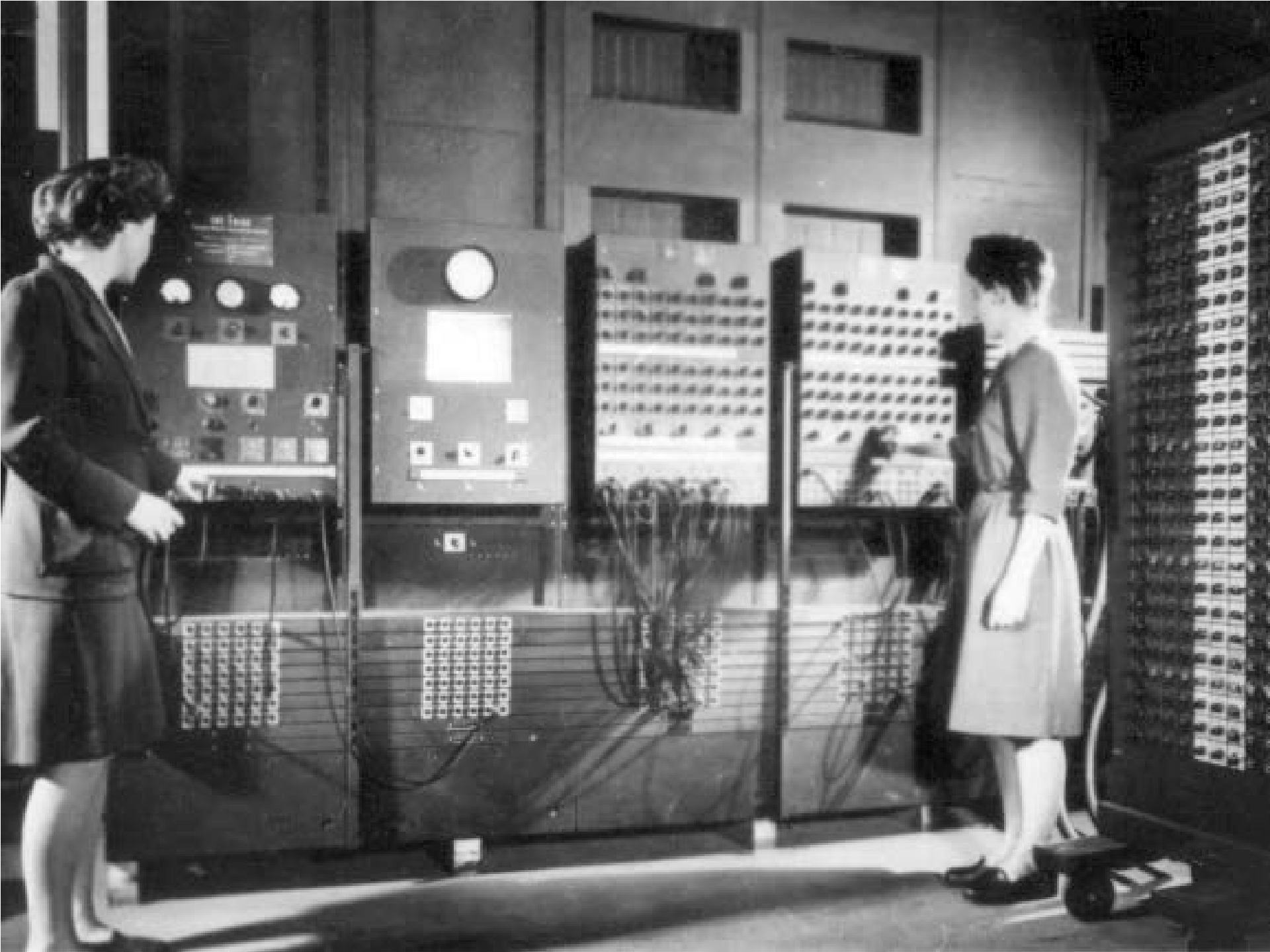
IS

COMPLICATED



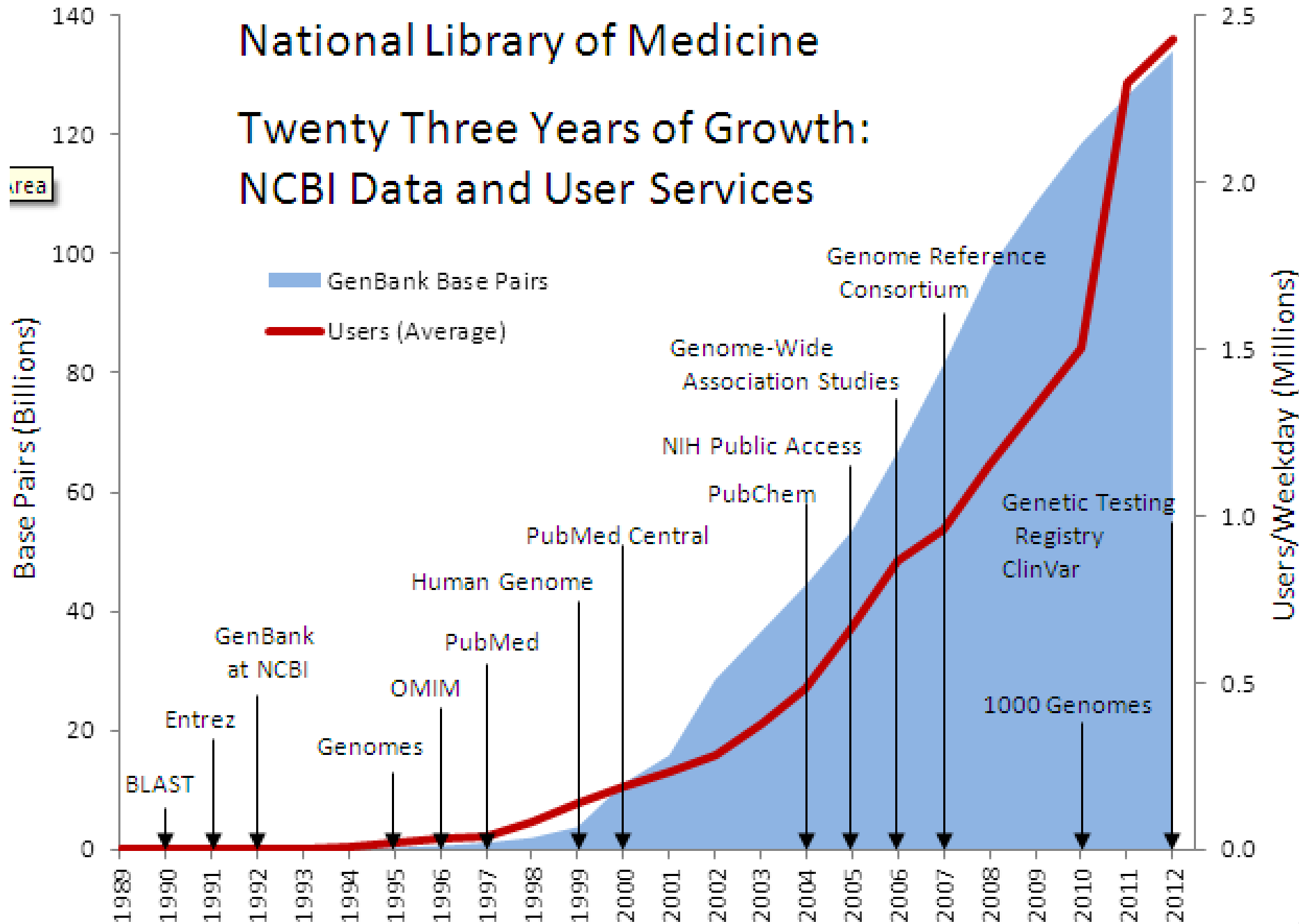
Wikipedia: ENIAC





National Library of Medicine

Twenty Three Years of Growth: NCBI Data and User Services



21st century data collection

- Written in lab notebooks by people, but also...
- Collected by robots
- Stored on computers

Wikipedia: BGI

“BGI’s 500-node supercomputer processes 10 terabytes of raw sequencing data every 24 hours from its current 30 or so Genome Analyzers from Illumina. The annual budget for the computer center is US\$9 million.”



21st century data storage

- Lab notebooks
- Computers
 - Desktops, laptops, USB sticks, backup hard drives
 - T-Storage
 - Cloud/Dropbox, etc.

21st century data management and integrity

Who owns data collected at UT?

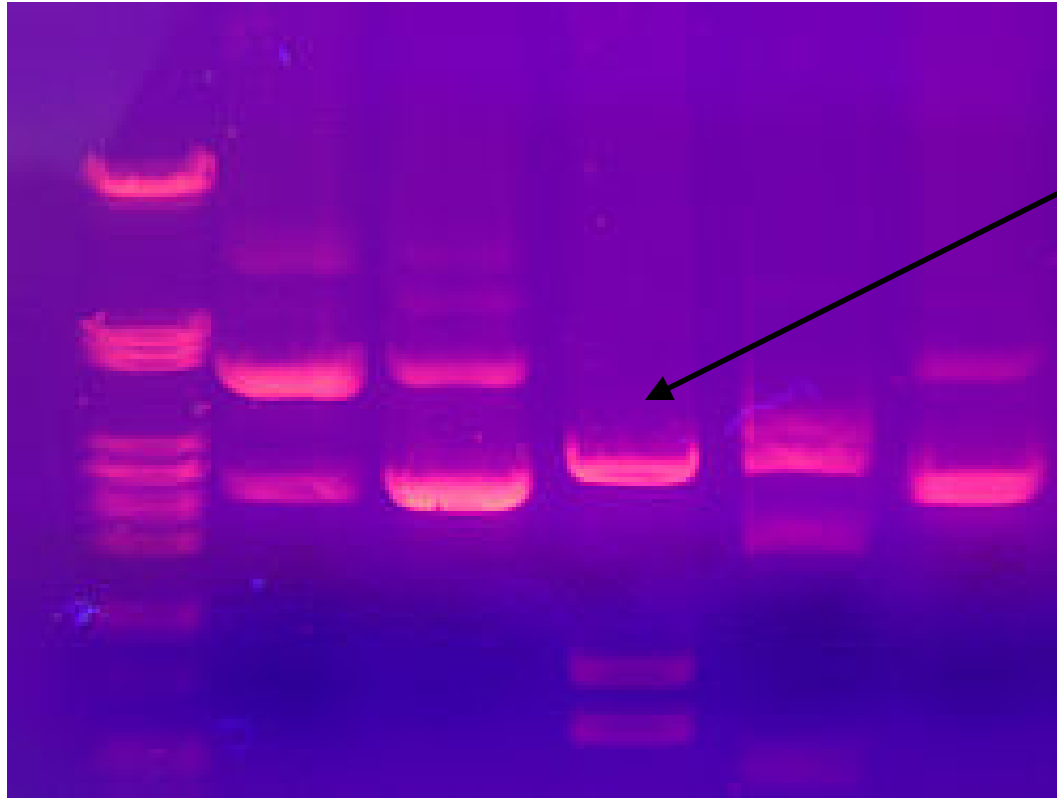
- UT
- Management and integrity mandated by funders
- PI is the steward
 - Publications
 - Material: MTAs vs third party distribution
 - Responsible for archiving and making data available on request
 - Huge help to exonerate parties accused of misconduct
- UTRF manages intellectual property
- Case studies!

1. The case of the 'data burglary'

- Judy Mikovits: chronic fatigue syndrome researcher and former Director of the Whittemore Peterson Institute
- Fired in September 29 2011
- Arrested in November 2011 after she 'stole' 18 lab notebooks→jailed
- The Ventura County, California Sheriff's Department web site lists Mikovits under [booking number 1259336](#), charged with a felony violation of California Penal Code section 1551.1, "Fugitive From Justice."
- Charges were dropped after she returned the notebooks to WPI

Case 2: The case of the funny gels

-
↓
+



Future trends

- “Future events will affect you in the future”
- More and more data in less and less time
- Data curation will become more difficult and easier, e.g., LIMS
- Funders will require proof of data curation and availability
- Computers become smaller and hacking will get worse