

The (Bioscience) Data Lifecycle....

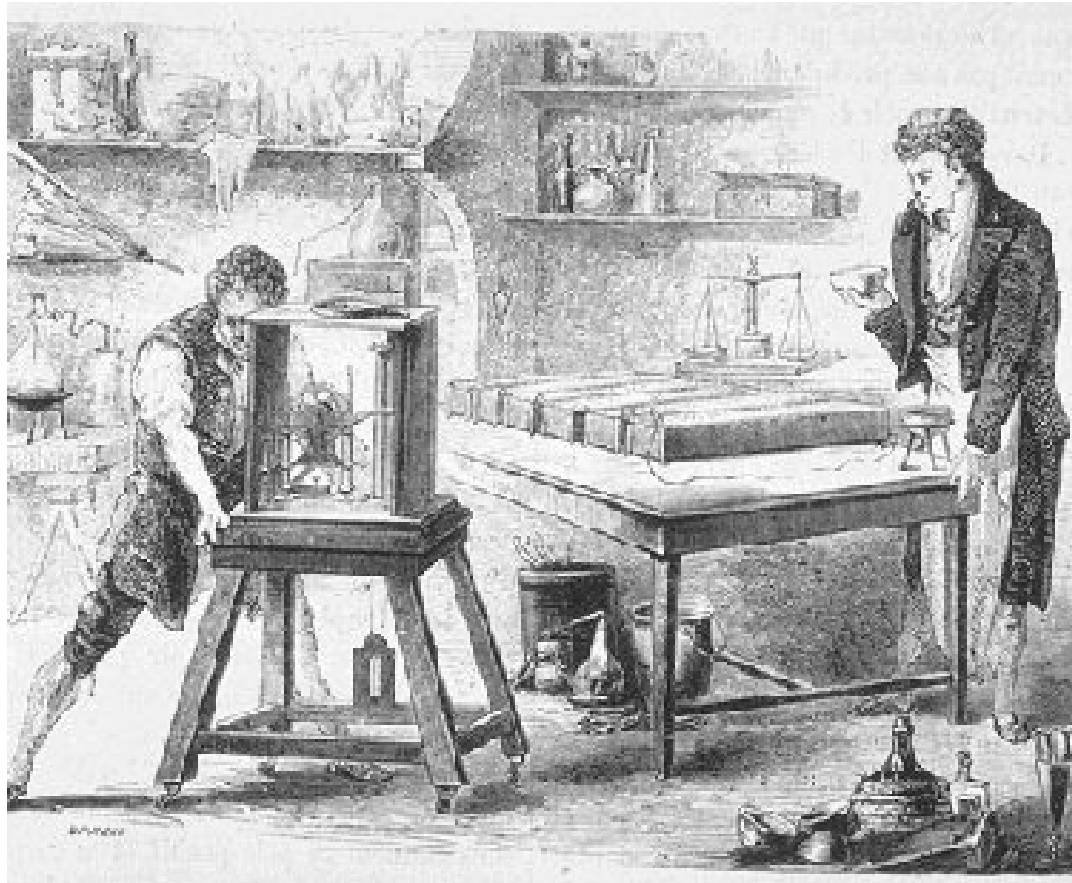
some researcher requirements

Marc Wilkins



THE UNIVERSITY OF
NEW SOUTH WALES
SYDNEY • AUSTRALIA

Until Recently, Biology was 'Data Poor'



- experiments slow and hard to do
- biological systems were mostly unknown
- studies on one gene or one protein at a time
- this was the case until the mid 1990s

The ‘–omics’ Revolution

New analytical technology and computer power has led to the –omics revolution

- Genomics – study of DNA and genes
- Proteomics – study of proteins
- Metabolomics – study of metabolites

These are all ‘global’ technologies.

Proteomic and Genomic Analysis



10,000 spectra per hour
We have 22,000 proteins in our body



Millions of sequences per run..
but each is quite small (400 bases)
The human genome is 3×10^9 bases

What Does Modern Biology Need?

Raw data storage: short term and long term

Analysis pipelines: from raw to processed data

Data comparison tools: what have I got?

Data visualisation tools: viewers, complex graphs

Data mining tools: hypothesis testing

Databases: for sharing high and low quality data

Web sites: for data browsing and dissemination

Data lifecycle?

data to information, information to knowledge

information to publications (!)

Where is it all heading?

The mass spectrometer will become the ultimate diagnostic tool (biomarkers)

The \$1000 genome will arrive at a GP near you soon

The insurance industry will have a field day!

Issues of ethics and privacy will become paramount....

Spit and send technology...

The screenshot shows a Microsoft Internet Explorer browser window displaying the 23andMe website. The browser's address bar shows the URL <https://www.23andme.com/>. The website header includes the 23andMe logo and the tagline "genetics just got personal." Navigation links include "log in", "claim codes", "blog", "we're hiring!", and "help". A search bar is also present.

The main content area features a large banner with the text "See your genes in a whole new light." and "TIME Magazine's 2008 Invention of the Year, now \$399." Below this, there are images of the 23andMe kit boxes. A "Multi-Pack Special" offer is highlighted: "Save \$100 when you order 2 kits." Three buttons are visible: "How it works", "Buy US \$399", and "Try a demo".

On the right side, there is a sidebar with two sections: "Just got your kit?" with a "Click to claim your kit" button, and "Already have an account?" with input fields for "Login name" and "Password", a "Login" button, and a "Forgot password?" link.

At the bottom, there is a "News:" section with a play/pause button and four category tiles: "Health and Traits", "Ancestry", "Sharing and Community", and "Research". The Windows taskbar at the bottom shows the Start button and several open applications, including "http://ands.org.au/r...", "23andMe - Beyond g...", "ANDS data life cycle", and "wilkins dynamic intera...". The system clock shows 8:50 AM.