April 17, 2024

Trends from the Trenches April 2024

Bio-IT World 2024
4/17/2024
Ari E. Berman, Ph.D.

bioteam.net

BioTeam
Accelerate Science

Who is BioTeam? (one and ONLY plug)

- Consulting company of deep experts
- Integrating researchers, data, and state-of-the-art technologies to help foster scientific advancements
- Bringing science and IT together to make science go faster
- Booth #404

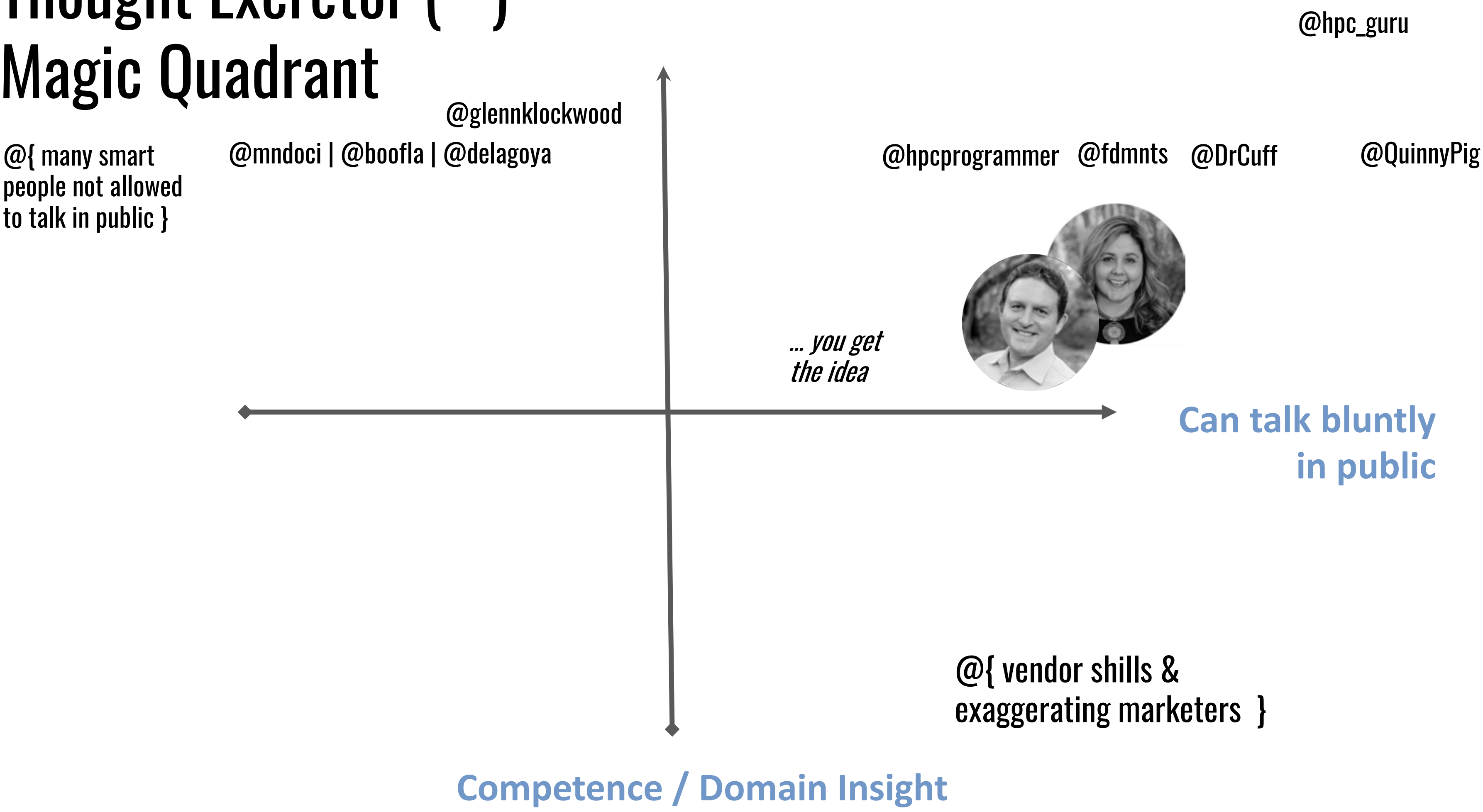
What to expect from this talk…

- Chris Dagdigian – gave this talk for 10 years
  - Retired from Trends last year
  - Focus has always been on blunt, honest and practical lessons learned from real world project
  - As consultants we get to see how different groups of smart people tackle similar challenges
  - We can (generally) speak in public about what we've experienced and learned - no filter, nothing to sell and no Marketing BS

So, who the heck are we?

- Laura Boykin Okalebo – Scientist, Senior TED Fellow, Senior Consultant at BioTeam
  - Passionate about equality and empowerment in science, solving REALLY hard problems
  - Practical applications of Trends in the Field
- I’m Ari – BioTeam’s CEO
  - I’m a scientist, technologist, been with BioTeam for 12 years
- Goal is to be informative and hopefully entertaining – I’ll do trends, Laura will talk about when trends fail and what to do about it

Thought Excretor (TM) Magic Quadrant



Biomedical Research Grand Challenges

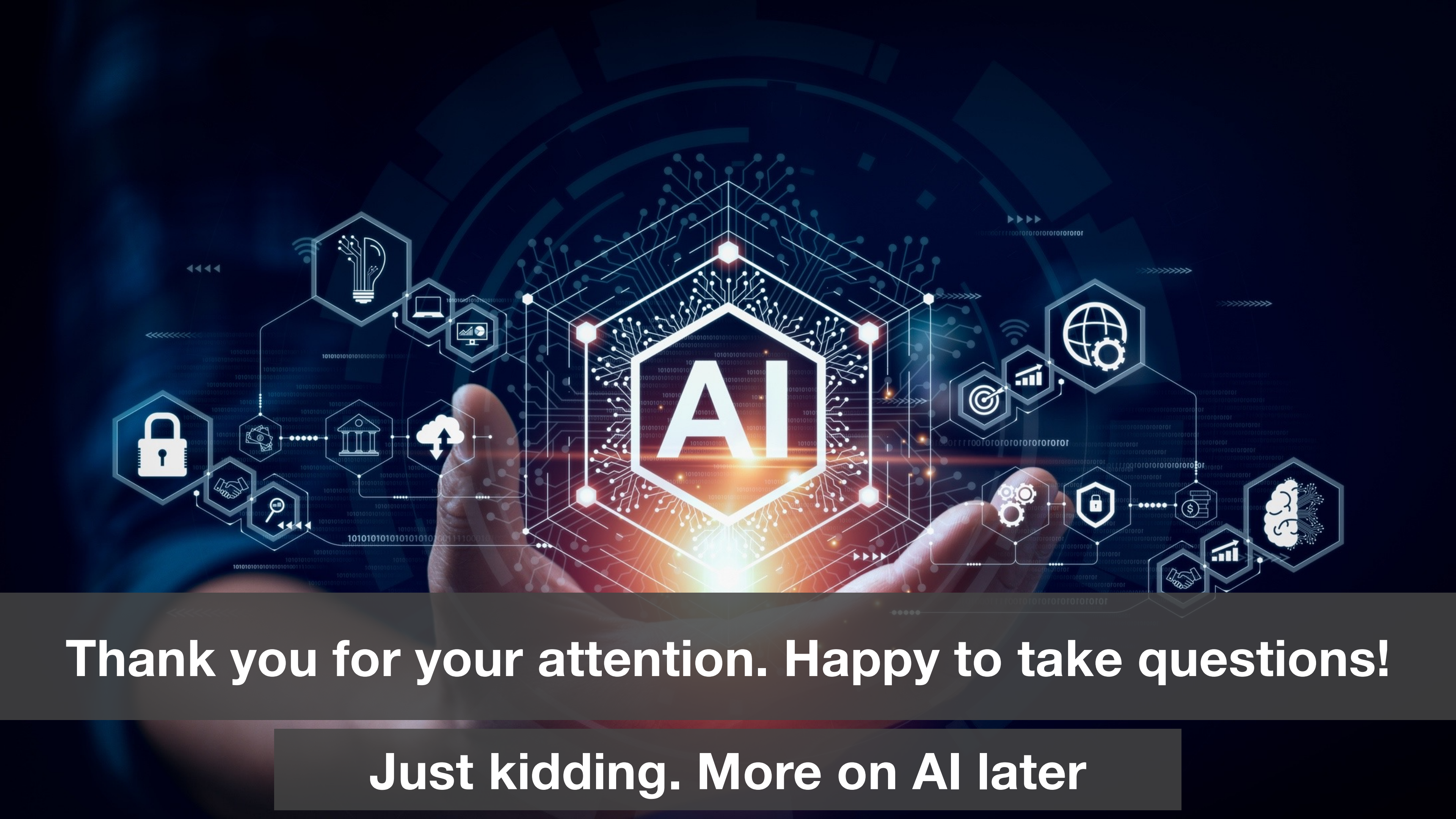
Context for this workshop, driven by what we’re trying to help the community do better

Grand Challenges

Current state	IT support for science inconsistent, not aligned with science	Fractured data ecosystem - minimal standards/FAIR	Advanced analytics are built by data scientists for experts
Effect on science	Infrastructure is a barrier, not an asset, scientists do it themselves	Enormous productivity sink just getting data ready for analysis - 80% of time	Data scientists become a bottleneck, lab scientists can't do analysis on their own
Grand challenge	Create advanced infrastructure for science - abstracted from users	Unify Biomedical Data Ecosystem	Make advanced analytics accessible to the long-tail of scientists

Science goes faster!

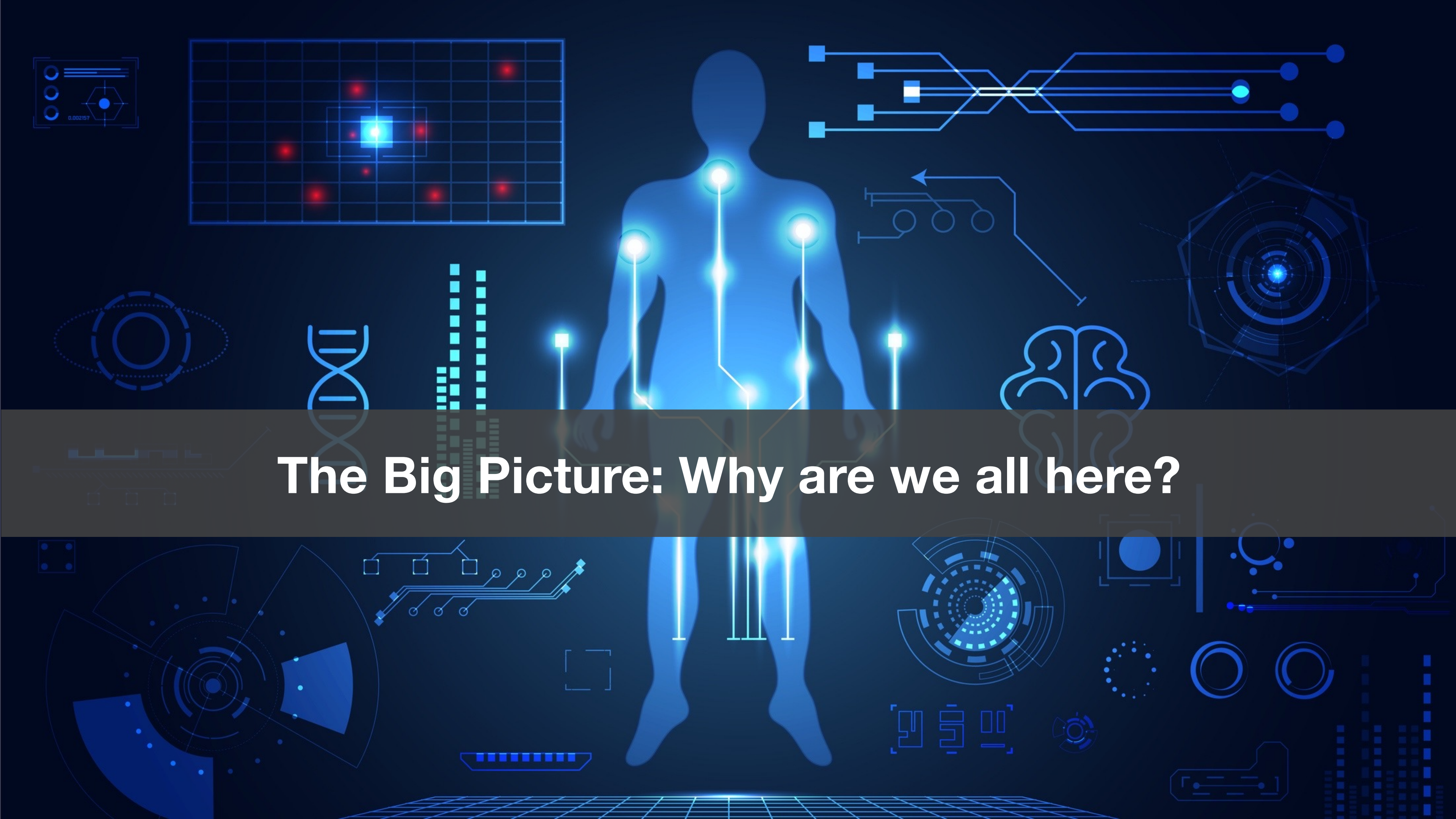
Let’s jump into the Trends



Thank you for your attention. Happy to take questions!

Just kidding. More on AI later

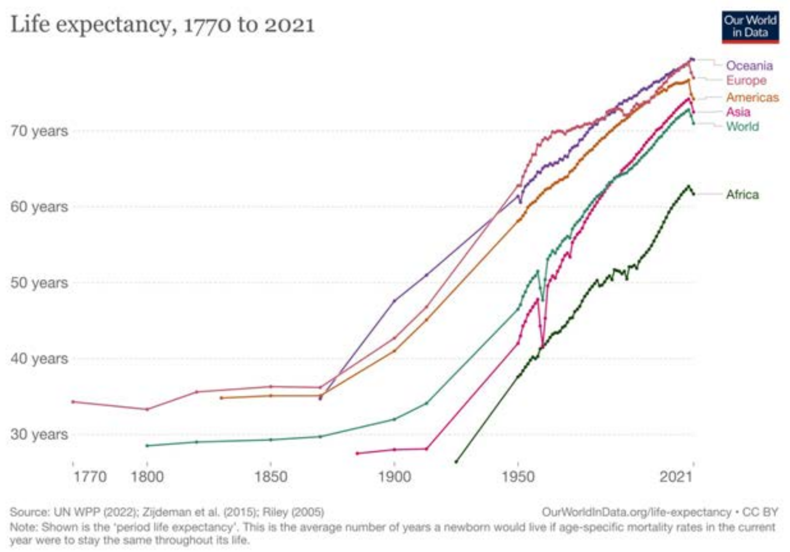
The Big Picture: Why are we all here?



Scientific Discovery Drives Modern Medicine

Science: *the systematic study of the structure and behavior of the physical and natural world through observation and experiment*

- Human lifespan hit a maximum in 2019 (76.7 years US)
- First major inflection in 1865 – Germ Theory, antiseptics in surgery, washing hands (doubled lifespans)
- Vaccinations, epidemiology, anesthesia, antibiotics are others
- All driven by basic life sciences research and applied to human health



Long Lifespan and Healthspan Now Out of Sync

- Healthspan – the percentage of life that one is considered healthy
- "Healthy" means different things to different people
- Average age of health decline in US is 63yo vs. 76.7-year lifespans
- Live nearly 20% of our lives unhealthy – with lower quality of life
- Resulted in soaring healthcare costs and increased burden on medical and economic systems as the population ages poorly

Disease	Deaths per year	Age of 1st occurrence
Heart Disease	610,000	65
Lung cancer	158,060	60
Chronic obstructive pulmonary disease (COPD)	147,101	45
Stroke	140,000	65
Lower respiratory infections	131,800	75
Alzheimer's disease	93,541	65
Type 2 diabetes	69,071	54
Colorectal cancers	50,260	70
Breast cancer	40,000	62
Prostate cancer	25,000	66

Table 1: Top ten causes of death in the US and their average or median age of first occurrence.

Medical Science Needs a New Focus

- Historically, medicine has focused on reducing risk of death as an outcome
- Medical science has advanced to the point where that focus isn't as relevant anymore
- Need new breakthroughs and renewed focus on being healthy for longer, vs being alive longer (likely linked)
- Modern diagnostic and computational methods have already started driving towards this reality

Drive Towards Precision Medicine is Data Intensive

- Move away from the “one-size-fits-all” approach to health care delivery and to instead tailor treatment and prevention strategies to people's unique characteristics, including environment, lifestyle, and biology
- Medical decisions, practices, interventions and/or products being tailored to the individual patient based on their predicted response or risk of disease
- Goal of increased Healthspan
- Foundational data initiatives like All of Us, CANDLE, TCGA, InsightRX, and many more
- Involves a large amount of data collection, modeling, and classification to pull off

Infrastructure in Life Sciences and Healthcare: Scoping the problem

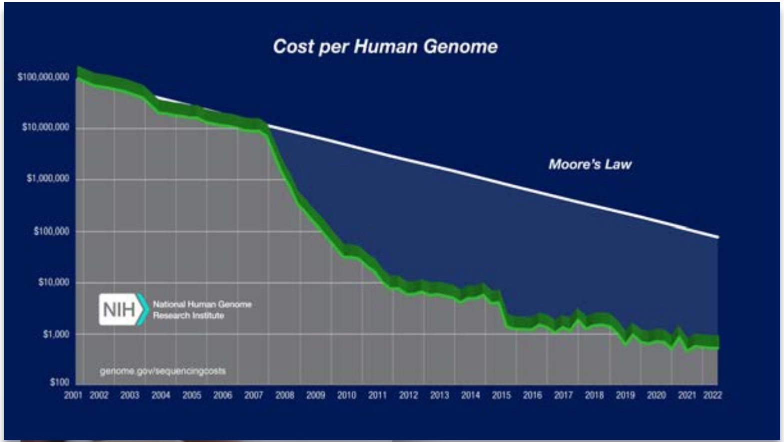
Data Generation: All-time high
Estimated 120 ZB of total data collected to date

23 ZB last year – 10GB/person/day

Laboratory and Health Diagnostics Innovation Rate

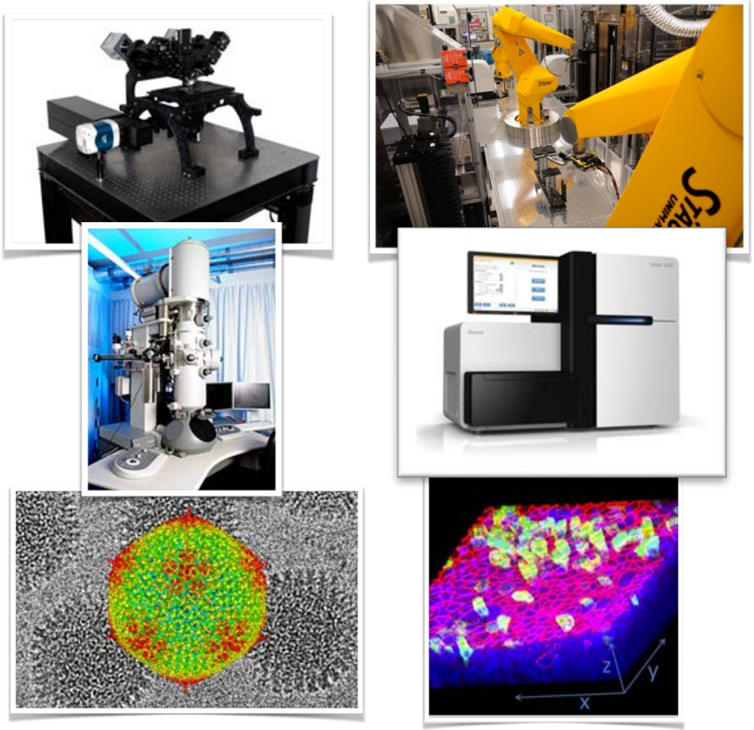
Rate of data generation in life sciences far outpaces Moore's Law

- Laboratory technology innovating on the order of a year or less
- HPC and IT tends to move on the order many years – equipment lifecycles are 3-6 years
- Cloud has enabled more rapid cycling, but at increased cost and complexity
- Health diagnostics, sensor data, making sense of clinical trials, pushes amount and complexity of data much higher



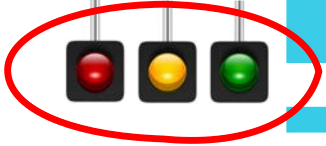
Laboratory Data Generation at an All-time High

- Sequencers were just the beginning
- Life sciences and Healthcare among the top data generators of all the sciences
- Most labs have equipment that can generate 10s TB of data/week
- Bioinformatics/Data Science have taken over biomedical analytics space (and AI)
- Computational sophistication required to compete
- Clinical informatics becoming a major health diagnostics tool




High Performance Computing is a laboratory tool, not an enterprise IT service!


23
2024
TRENDS



Nomenclature

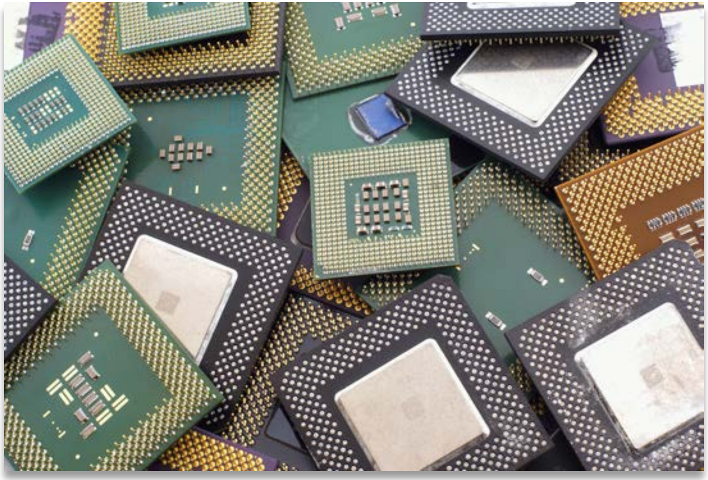
For each trend, I'll show a stoplight showing its effect on science

- Red means a more challenging or negative trend – makes science harder
- Yellow means a sustaining trend – keeping science moving as it is now
- Green means an innovating trend – helps science go faster

Trends: Infrastructure for Biomedical Research

*Trend: Computing is hard again

- The HPC infrastructure and services landscape has started to diverge away from the “figured out” landscape in 2019
- There are way more options and far more specializations in hardware, configurations, places to use compute, services to use it with
- This re-differentiation is causing a lot of problems for science – there aren’t uniform platforms for people to do research



The pandemic changed the game

- Everyone went home into isolation
- Labs shut down – researchers started remotely analyzing data
- On-premises HPC and compute systems ramped up usage a lot – users that wouldn't use it that often were now clamoring for access and hours on systems
- On-premises infrastructure became hard to manage remotely, overloaded VPNs, not prepared for remote science
- Local IT could barely maintain the status quo (still)

Infrastructure got scarce

- It’s genuinely harder to get hardware now
  - Global supply chain issues
  - Explosion of AI
  - Everything is more expensive (inflation)
    - $100,000 H100s???
- Change in investment priorities from the pandemic
  - Everyone is paying for buildings that aren’t really being used
  - Orgs are questioning capital expenditures
  - Scientists got started on deep analytics when labs closed – demand is UP

…Plus, a declining market in our space

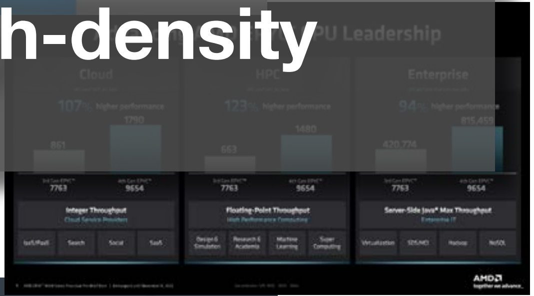
- Our government couldn't fund itself – set off a series of downward spirals
  - Grants and new investments delayed with NIH
  - Contractors not getting enough work – laying people off
- Pharma downsizing in an inflated and stressed drug market
  - Stressed by underutilized office space and properties
  - Drug pipelines are slower, early development sluggish, approvals even slower
- VC and PE dried up 2021-2023, seemed to be turning around, but maybe not: "Biotech winter"


BUT – There’s a resurgence of on-prem HPC planning!
Shows promise that things might start moving again

CPU Trends

Intel isn't the only game in town anymore, and is starting to lose their foot hold in the HPC space – choice

- Intel Sapphire Rapids (Xeon Scalable Processors
  - Up to 60 cores and 120 threads/PCIe 5
- AMD Milan/Genoa
  - Up to 96 cores and 192 threads/PCIe 5
- ARM Neoverse V3
  - Up to 128 cores/PCIe 5
  - Fugaku HPC (Japan, 4th on top-500) based on Fujitsu A64FX (7,630,848 cores, 0.5 Eflops Armv8.2-A)
- **Moral of the story**: Ultra high-density compute – 150kW/rack, 768 cores per quad chassis, ~11,000 cores in a 56U rack!



High-density compute = high-density discovery

GPUs Trends

- Literally **INSANE** demand for GPUs (NVIDIA)
- People paying $100,000 for a single H100 (see Dan Stanzione's comment of gold bars)!
- AI-crazed fever pitch has led to this trend
- **REALITY CHECK:** GPUs are not magic, they make 30% of scientific workloads go faster (deep learning is one of them), the rest CANNOT USE THEM!
- They are not fast CPUs, they are highly parallel vector processors made for graphics
- If you're only planning on GPU capacity, you're alienating 70% of your research

GPUs Trends

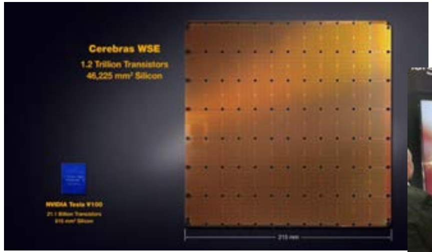
- The world is bigger than NVIDIA
  - Hopper (H100) very fast, very HOT, very expensive, 350W, HBM very interesting
  - A100/V100 still work great
  - NVIDIA DGX Superpods are the marketing goal

Need a balanced approach: balance cost, usability for science, and performance

- AMD Instinct MI300X
  - Excels in both data retrieval speeds and storage cap... 400W
  - Also, no CUDA – need to use ROCm (though ZLUDA looks promising)
- Intel GPU Max 1550
  - Interesting, harder to use than the others, lagging in production, possibly useful performance in simulations

Other co-processors

- Cerebras – wafer-scale “AI”
- AWS Inferentia, SAGEMaker, Google TPU, etc.
- Voltron Data and Next Silicon Accelerators
  - Thinking differently about how to process high-intensity algorithms
- SpiNNaker from SpiNCloud
  - Very flexible hardware architecture – can approximate quantum (more on that later) – built to work like the human brain
- FPGAs are still a thing



VOLTRON DATA

NEXTSILICON

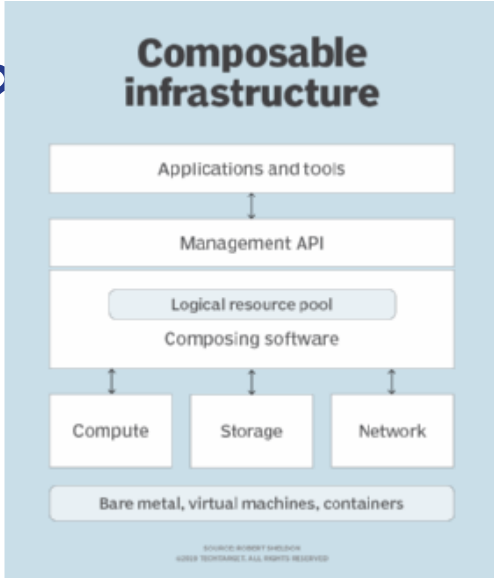
SpiNNcloud

Trends for HPC interconnect/networking

- Interconnect is still fairly standard, except that NVIDIA bought Mellanox and tend to come as a “package deal” to get discounts
  - Omnipath (Cornelis) is still a thing and may resurge due to the NVIDIA thing
  - Rockport Networks is interesting as a RoCE solution - scaling issues
- Composable architectures and CXL may change how HPC is designed – Liqid and GigaIO are the main players here
  - Memverge is also very interesting

Composable Infrastructure?

- Composability means you can assemble a high-performance “node” virtually by meshing hardware from many machines into one using PCIe-switching
- Reduces need for exotic heterogenous HPC designs
- GigaIO, Liqid are the major players – open standard CXL



GIGAIO

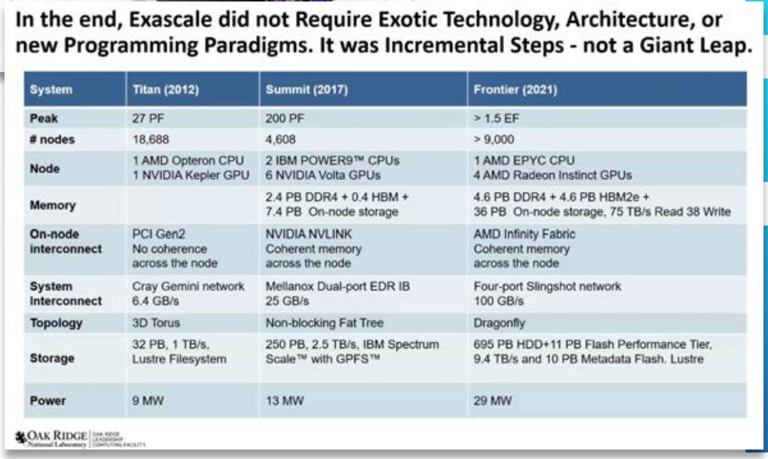
LIQID

Trends in Storage

- Majority of organizations go for volume and cost, over pure performance
- Cost/space/power is always a limiting factor
- Buying Petabytes of anything is very expensive
- Without data value and data lifecycles, have to plan for infinite storage – Cloud gets very expensive at 1PB+
  - Hammerspace is an interesting play in the data management space
- Most common in life sciences on-prem HPC: Spectrum Scale (GPFS)
  - Lustre doesn't handle the small files well (though promises from DDN to improve it this year)
  - VAST and Weka are the main players, with Pure making an impact

Power Trends

- AI industry: 85-134 Terawatts/year by 2027
  - That would power the Netherlands for a year!
  - This is not sustainable – too power intensive
  - Not to mention (actual) global warming from heat production
- Need other methods for cooling or power efficiency
  - Immersion cooling
  - Direct liquid cooling
  - Keep using CPUs, not everything has to be a GPU



In the end, Exascale did not Require Exotic Technology, Architecture, or new Programming Paradigms. It was Incremental Steps - not a Giant Leap.

System	Titan (2012)	Summit (2017)	Frontier (2021)
Peak	27 PF	200 PF	> 1.5 EF
# nodes	18,688	4,608	> 9,000
Node	1 AMD Opteron CPU 1 NVIDIA Kepler GPU	2 IBM POWER9™ CPUs 6 NVIDIA Volta GPUs	1 AMD EPYC CPU 4 AMD Radeon Instinct GPUs
Memory		2.4 PB DDR4 + 0.4 HBM + 7.4 PB On-node storage	4.6 PB DDR4 + 4.6 PB HBM2e + 36 PB On-node storage, 75 TB/s Read 38 Write
On-node interconnect	PCI Gen2 No coherence across the node	NVIDIA NVLINK Coherent memory across the node	AMD Infinity Fabric Coherent memory across the node
System Interconnect	Cray Gemini network 6.4 GB/s	Mellanox Dual-port EDR IB 25 GB/s	Four-port Slingshot network 100 GB/s
Topology	3D Torus	Non-blocking Fat Tree	Dragonfly
Storage	32 PB, 1 TB/s, Lustre Filesystem	250 PB, 2.5 TB/s, IBM Spectrum Scale™ with GPFS™	695 PB HDD+11 PB Flash Performance Tier, 9.4 TB/s and 10 PB Metadata Flash. Lustre
Power	9 MW	13 MW	29 MW

OAK RIDGE National Laboratory

Trends in Cloud

Cloud providers launched a BRILLIANT marketing campaign: convinced decision makers to go cloud.

People Tend to Think in Absolutes

- A solution can only be this or that
- If cloud, then ALL cloud
- Abandon things that work well, because this will be better!
- Reasons:
  - Everyone's doing it
  - It's cool so people will invest in it, support it
  - Uninformed Leadership likes to talk about it
- Reality:
  - Very nuanced situation, there are no absolutes, and it isn't simple

Aggressive Cloud Migration Programs were Started

- Most science organizations started planning for cloud-first or all-cloud transitions away from local infrastructure
- Cloud providers gave deep discounts and a lot of direct support during the planning and the migrations to cloud
- Stopped planning on local infrastructure: storage, HPC, even networking
- All future planning was for cloud-based architectures
- We've seen this movie before…

A Look Back: The Last Wave of Cloud Migrations

- 2008 – 2014: we helped several orgs migrate to the cloud and close their datacenters (AWS)
- The draw: cheap, easier to manage, endless supply of compute power, less staff needed
  - Better shared access for external collaborations
  - Better access to public datasets
- The reality:
  - 10-50x the cost of operating datacenters
  - Cloud couldn't replace all local infrastructure
  - Required specialized skillsets in IT, harder to use
  - Didn't meet scientists' requirements
- The Result: Cloud sobriety – massive pullback

Ah, what short memories we have

Now it's Happening Again…

- …for slightly different reasons this time:
  - There are more cloud providers, way more sophisticated, clouds are largely designed to handle enterprise needs now
  - Competition has forced huge innovation in cloud services
    - There are aspects of cloud that you can't reproduce locally
    - Deep learning applications and specialized hardware/services are huge draws
  - There's so much data now, that storing it locally is non-trivial
  - Promise of it being somehow cheaper to operate
- Flexibility of cloud architectures combined with supply chain issues for local hardware made moving that direction more attractive

Also: Cloud business model locks your data/operations into their platforms

- The most successful ransom scheme in history
- Free to put data in, charged to get it out
- Save money by using proprietary (and very useful) services like:
  - Serverless
  - Bulk operations services
  - Specific analytics services
- Locks you into using their services – can't reproduce them yourself (e.g. – Parler)

Relevant Questions: To Cloud or not to Cloud?

- Can you do real HPC in the cloud?
  - Yes, now you can, but depends on what kind you need
- Can you create a secure environment?
  - Absolutely, but it’s yours to mess up. The models are different than on-prem
- Can scientists use it out of the box?
  - Absolutely not: requires set up of services, user interfaces, platforms
  - Even the sophisticated users need to architect their environments
  - Some things are there out of the box, but they’re basic
- Can we use it for storage?
  - Definitely: but storage in the cloud is complex, hard to price, and easy to overdo
  - Once your data is in the cloud, it costs money to get it out again

Advantages of the Cloud

- Large capacity, automatic upgrades, reliability
- Virtual orchestration and serverless technologies
  - Can architect very sophisticated environments more easily
- Containerization and portability of workloads
- Share data more easily, and in more sophisticated ways
- Shared standards in architecture, great way to prototype
- Availability of accelerators, like GPUs
  - Though, they aren't so available, really
- Specialized services that can speed up deployments

So, Why not All Cloud?

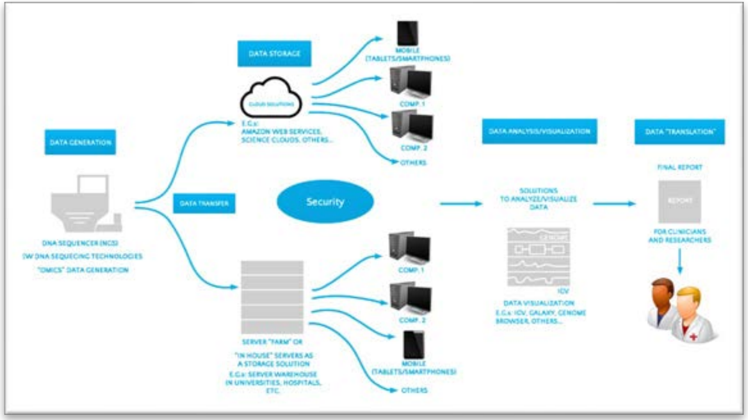
- Fundamental mismatch between cloud business model and long-term research goals
- These are private, for-profit companies, not public utilities. They offer services that they can change at any point (and do!)
- Shortest full scientific study, 5-10 years. Some 100s of years
- Clouds don't even disclose time period if they shut down to get your data back – this should concern you
- Hybrid computing model preserves data ownership – source of truth kept locally
  - Copy data to the cloud for analysis (ingress – no charge)
  - Only copy results of analyses back – delete source data (minimize data egress fees)
  - Still not solved!!

Trends in Networking

You have to move the data!

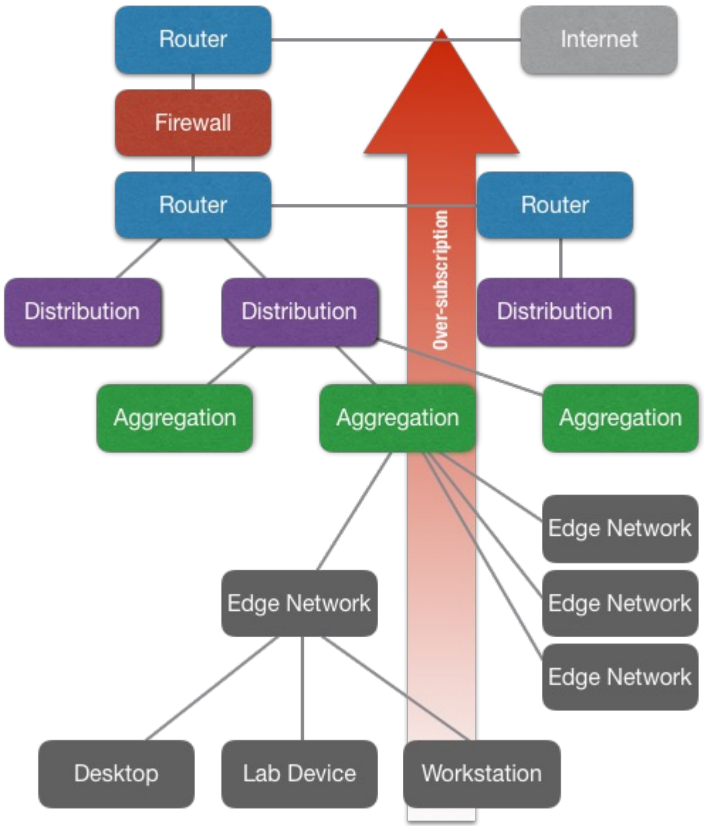
A lot of talk about moving compute to data, but you always have to move it at least once

- Laboratory equipment only able to store last few experiments
- Not designed for any analysis past initial data reduction
- Have to move it to compute storage for processing
- Many labs can generate 1PB/year or more – Non-trivial network requirements
- Data is valuable, needs to be protected, replicated, or backed up
- Data sharing is a standard requirement



Enterprise Networks

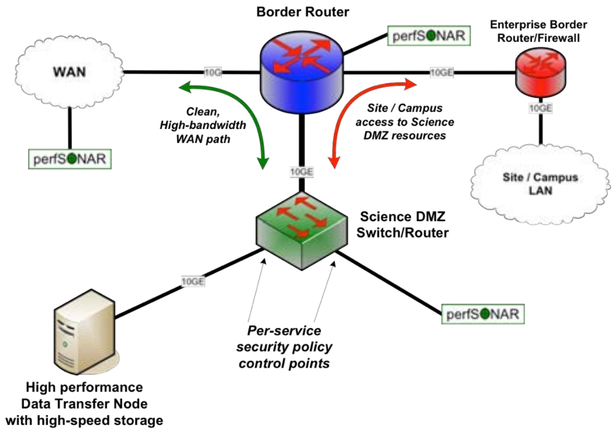
- Viewed as a cost to be controlled
- IT in general has flat budget, tight fiscal control
- Networks optimized for web/email traffic, not for large sustained data transfers
- Results in highly under- provisioned and oversubscribed network deployments
- Expertise split and siloed - networking org isolated
- Really need at least 10Gb from local data storage to Cloud for data intensive science
- Security designed for risk, not science



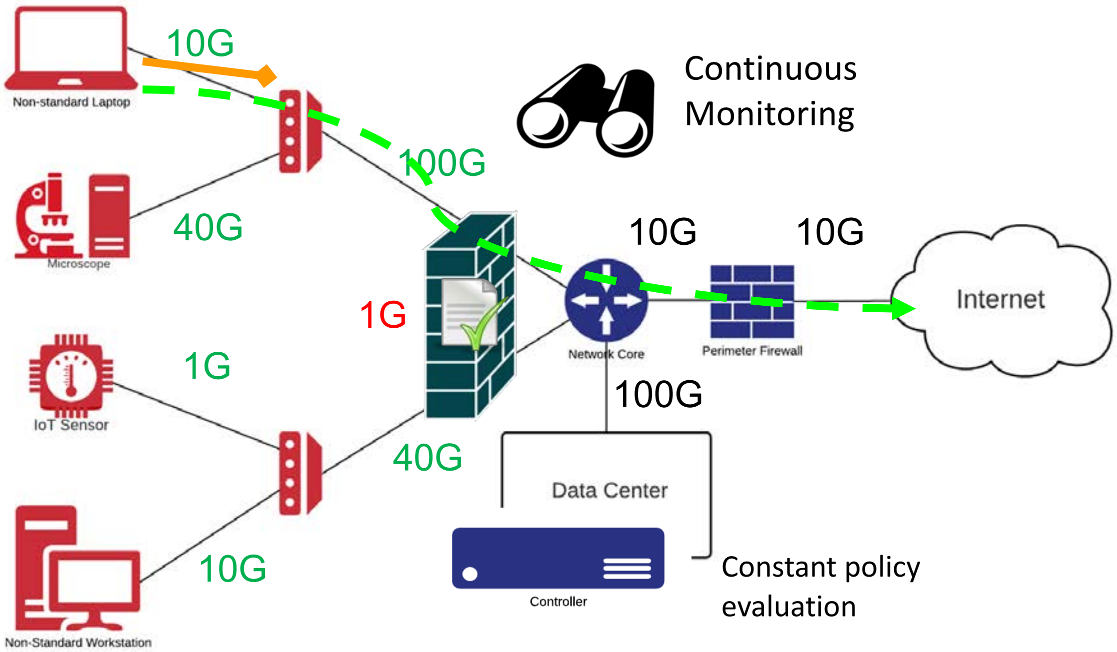
Most common high-speed network

Network: Science DMZ (perimeter network)

- Very fast "low-friction" network links and paths with security policy and enforcement specific to scientific workflows
- "ScienceDMZ" developed by DOE – ESNet connects National Labs
- Central premise:
  - Legacy firewall, network and security methods architected for "many small data flows" use cases
  - Not built to handle smaller #s of massive data flows
  - Also very hard to deploy 'traditional' security gear on faster networks
- More details, background & documents at fasterdata.es.net/science-dmz/

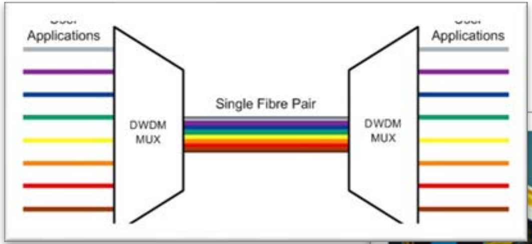


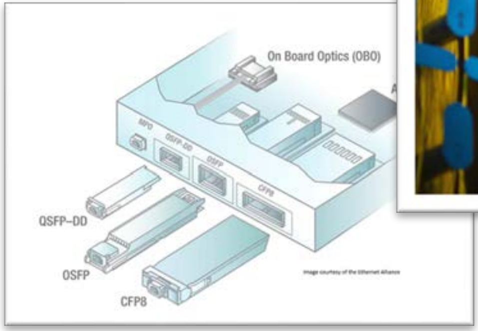
Zero-trust/microsegmentation



Networking: key for modern biomedical science

- Modern science requires 100Gb data speeds (1TB transfer in 1.5min*)
- Next-gen filesystems require 100Gb networking to function
- 400Gb current standard
- 600-800Gb Optical Transport Networks are out now
- 1 Tb networks in early release
- If you aren’t thinking of high-speed science for your network, you are hampering the science of your org





Trends in Data

What is FAIR?

- The four parts of the acronym describe the basic requirements for making scientific data available and useable across projects
- Why FAIR though?
  - **Obvious**: data should be FAIR—science and collaboration require fundamental standards
  - **Not obvious philosophy**: to align the life science and biomedical communities' data to convert ALL data into knowledge to improve understanding and treatment of disease—make all data analysis-ready
  - **Basis for creating a unified scientific data ecosystem**
  - Bold, ambitious, honorable, and an extremely difficult endeavor


How are we doing with FAIR so far?
Not great—limited pockets of excellence

What's blocking FAIR?

- Short answer: people, not technology
- Long answer—it's complicated:
  1. It's really hard to do, extremely complex—lots of historical data—but it can be done
  2. Lack of unified data standards in the field
  3. Lots of distraction, lack of sophistication, lack of tools and training, lack of accessibility of solutions
  4. No incentive to make the effort—all incentives are individual (i.e., NIH grants), this is a community effort (some EU efforts towards open data)
  5. It pays to do your own thing—your funding gets diluted if you join with groups of collaborators
  6. No clear vision, leadership, or directives from funding organizations to get there

Lack of unified data standards

- Biomedicine and life sciences research lack unified data standards
  - Very difficult to combine data from multiple sources for deeper analysis
- Clinical information largely contained in electronic health records (EHR) systems
  - Not designed for analytics, designed for compliance
  - Different systems store and make data available differently—far from FAIR
  - Deidentifying data is difficult and loses key information, while improving access
- Historical data locked in publications
- Each domain and each project has its own format/standards

Global Alliance for Genomics & Health

Collaborate. Innovate. Accelerate.

HL7 FHIR

OHDSI

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

MIAME and ArrayExpress

- a standard for microarray data annotation and a database to store it

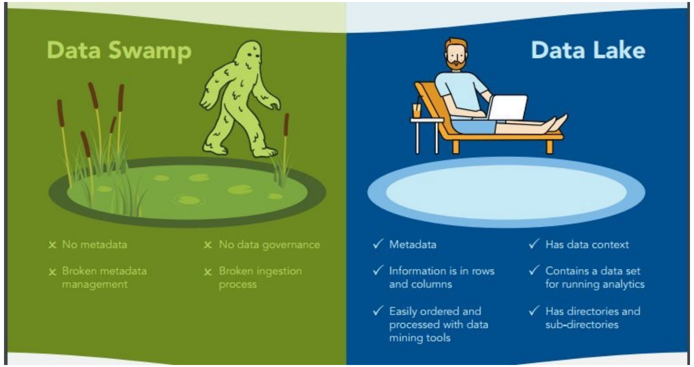
THE NATIONAL CENTER FOR BIOMEDICAL ONTOLOGY

RDA

RESEARCH DATA ALLIANCE

Data platforms—they're everywhere!

- Another buzz-loaded space:
  - Data Lakes, Oceans, Fogs, Swamps, Islands, Universe, Warehouse, Commons, Ecosystem, Mesh, Fabric…
- Truth: just building more silos of excellence—and not accessible to long tail of scientists (laboratory)
- Need a diversity of tools and approaches, but they aren't cross-compatible
- Lack of data standards and interoperability lose the power of the data in collaboration



GEN3 DATA COMMONS

DCPPC DATA COMMONS PILOT PHASE CONSORTIUM

Terra

Palantir

Data value in life sciences: key for digital transformation

- In general—no understanding of scientific data value
  - Investment—how much is data worth, what did it cost to generate
  - Scientific value—will this data ever be reused, how long do we keep it, what's the impact of each unit of data?
- Without this understanding, everything is high value so we just keep everything
- Infinitely expand storage—hoarding
- Can't prioritize investments in data that are aligned with strategy
- FAIR also doesn't contemplate data quality

The hype is increasingly ridiculous – but why the interest?


Transitioned from Information to Analytics age:
gain value from collected data
AI put pressure on field to innovate again

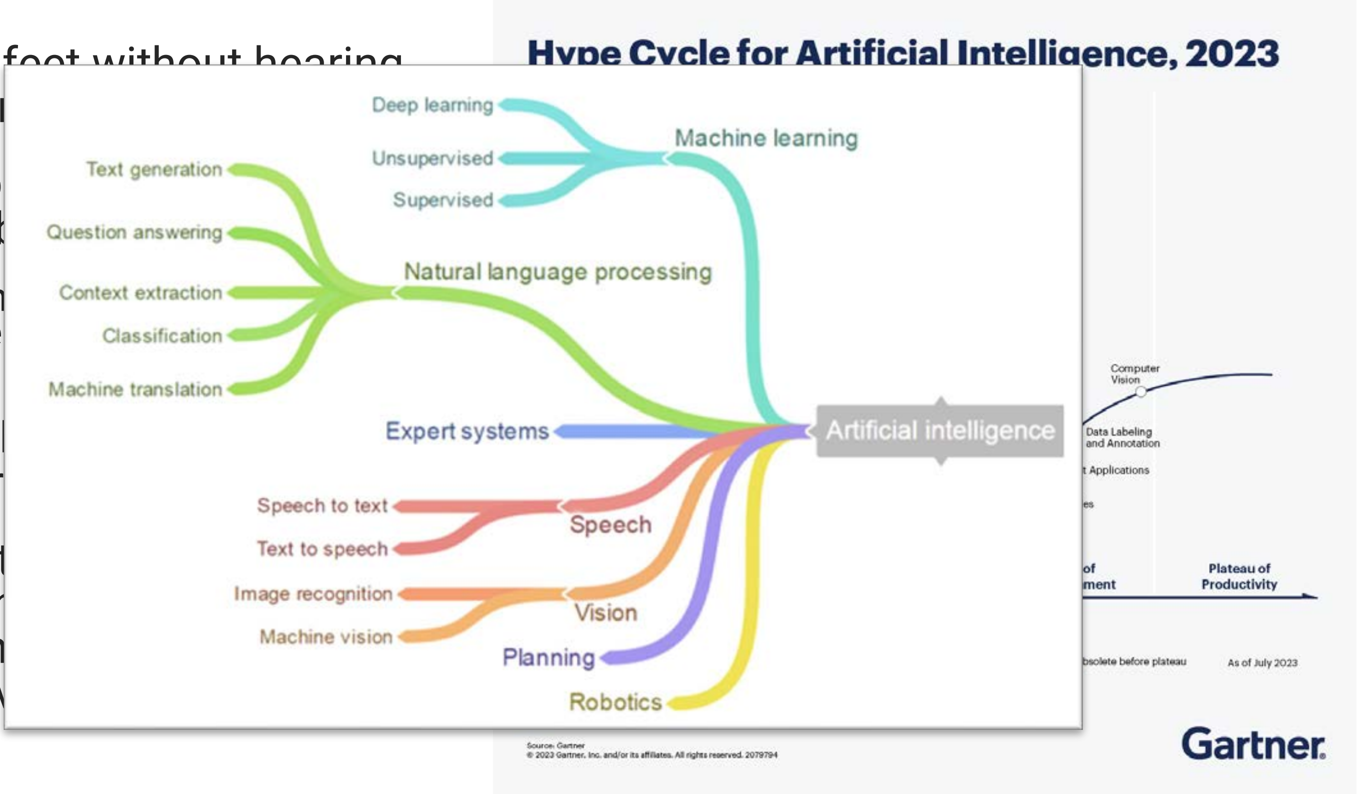
The AI Slide

Because no talk is complete without an AI Slide

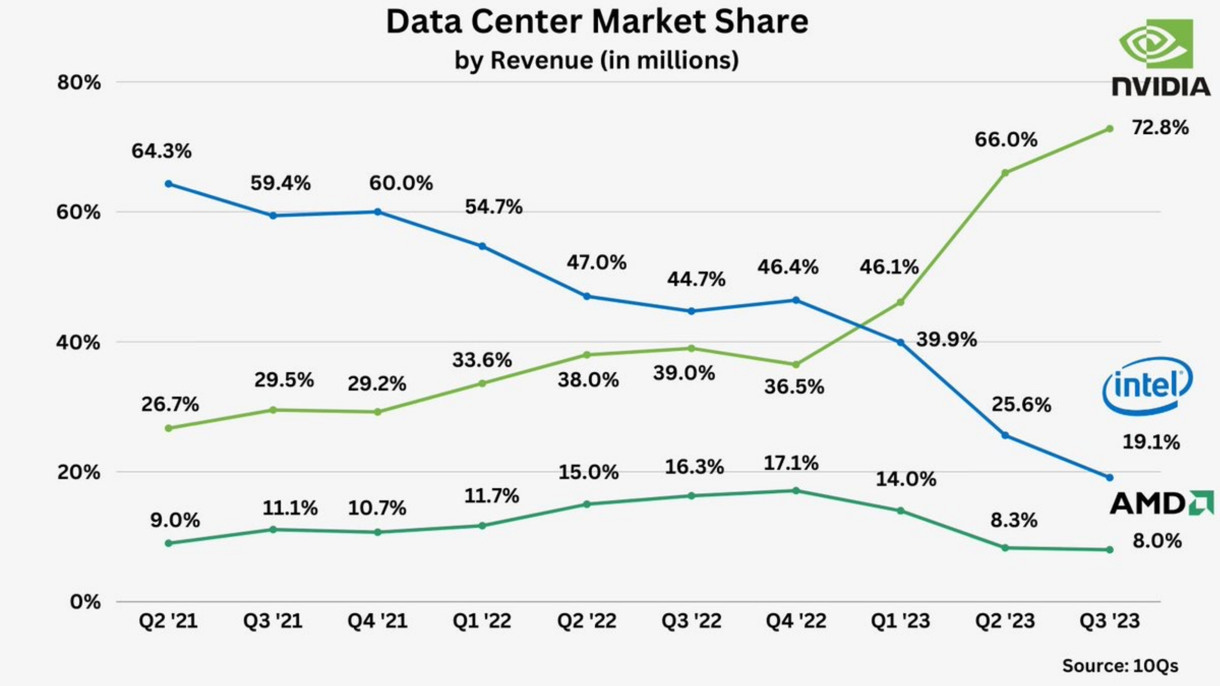
- Let's be honest [illegible] reminiscent of Big Data
- Most people don't really understand what AI/ML is, most think it means this:
- All organizations are telling us they need data to be AI-ready.

AI: Super useful, but way over-hyped

- Can't go two feet without hearing AI/ML being r
- AI as a comp has many sub
  - Deep Learnin learning type problems
- Especially in away from pr
- AI: everyone t don't really kn everyone thin doing it, so ev doing it



What's Driving AI Hype?



Trends Recovery Slide

- AI!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!! (Also: AI!!!!!!)
- Still a big push to cloud (#3 of Top500 0.5EF – Azure)
- Cloud push slowing down, on-prem planning is back: GPUs!!!
- We have had the same data problems for 15 years, they just got bigger and more complex – MORE STORAGE (for AI?!?)
- New silos have popped up that make a difference locally

Meta:

- IT is disconnected from science mission
- Science is disconnected from supporting infrastructure and not incentivized to change
- Data science is caught in the middle, being a bottleneck
- We're on the verge of making meaningful change, because of/in-spite of AI/ML

Thanks for listening!
Now Laura…

April 17, 2024

Trends from the Portable Genomics Trenches

Laura Boykin Okalebo, PhD
Senior Scientific Consultant
TED Senior Fellow

bioteam.net

BioTeam
Accelerate Science

How we are using DNA tech to help farmers fight crop diseases

TED 2019

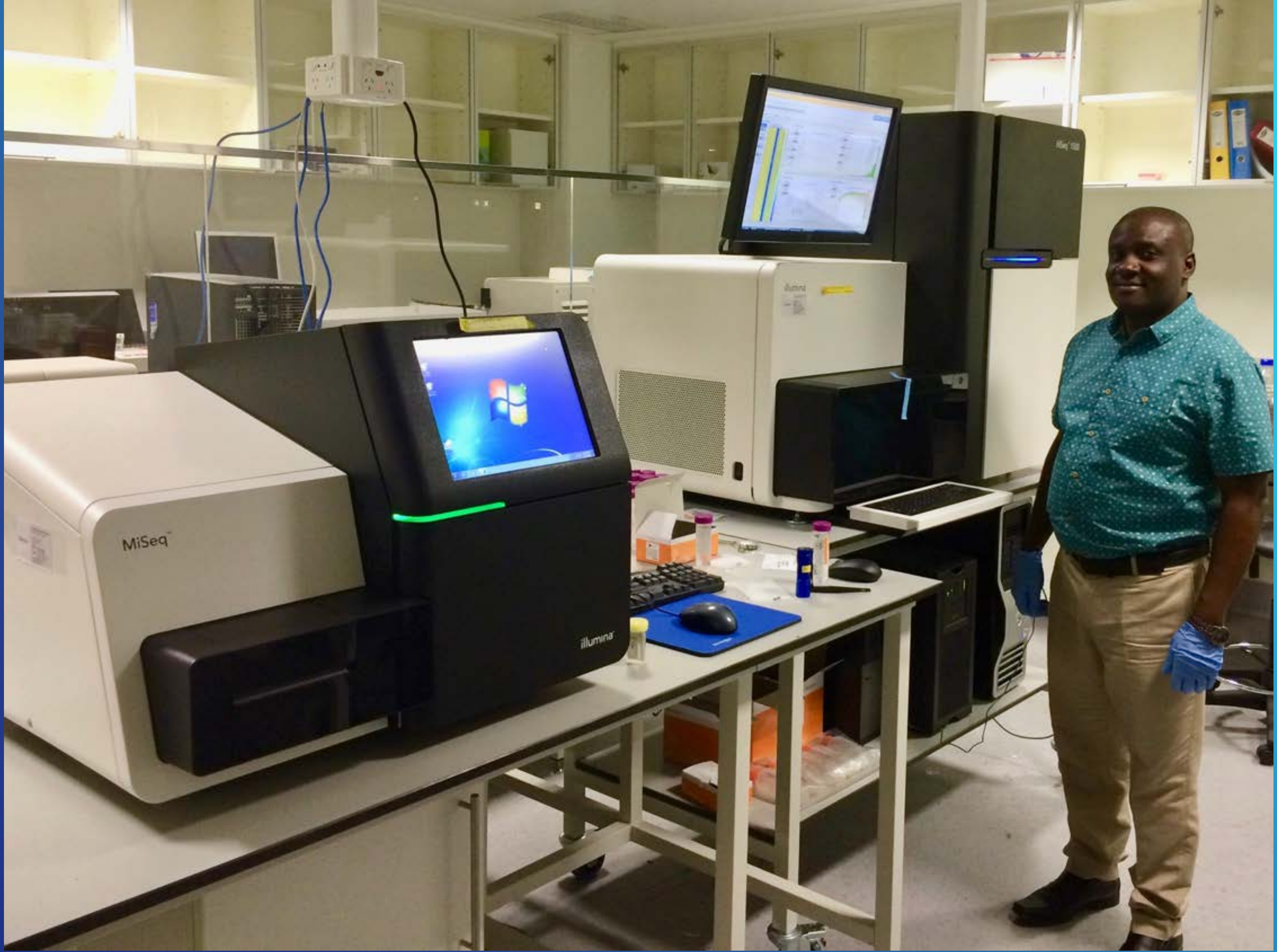
Courtesy: TED/Ryan lash

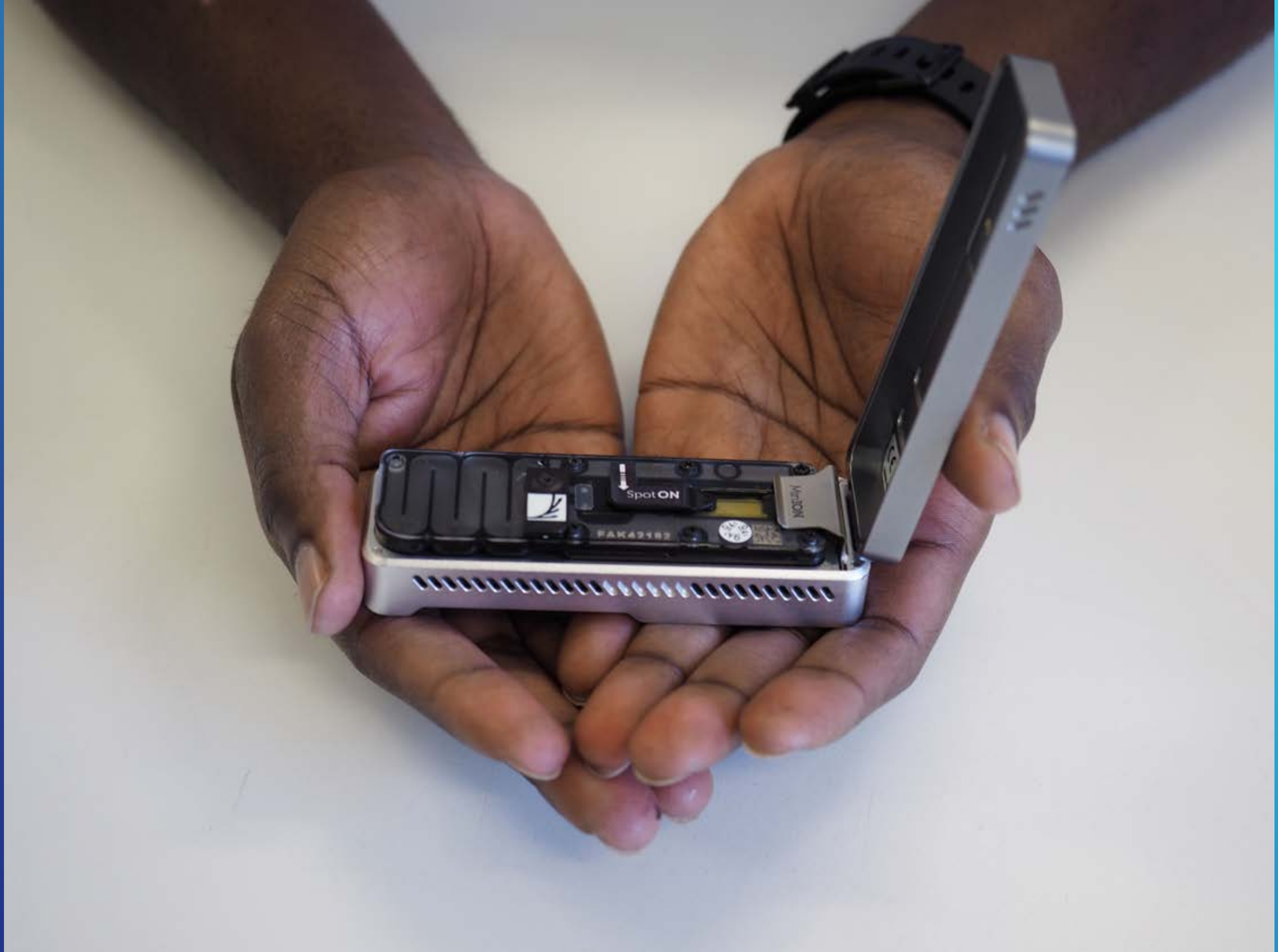
What are portable genomics?

Diagnostics results in seconds, minutes, hours

Portable means different things to different people

Portable- space station, ocean, Ebola, Zika, amazon


MiSeq
illumina

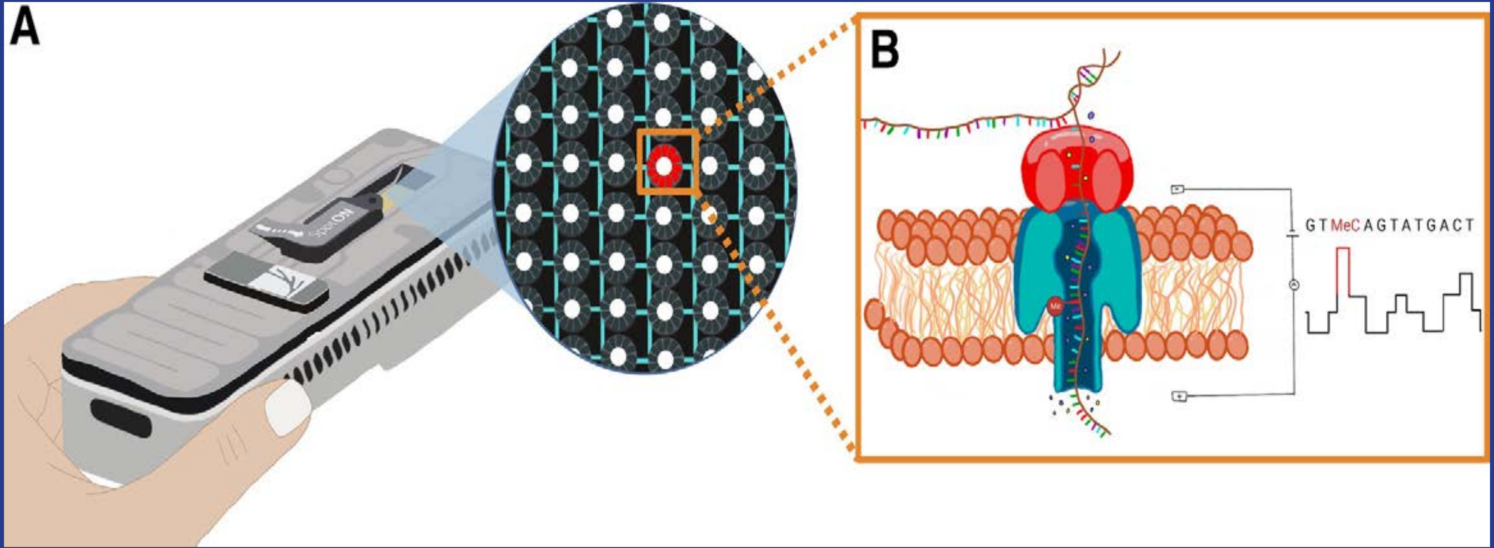


Courtesy: Monica A. Kehoe

What is portable DNA sequencing?

- A tool to help reduce time to diagnosis
- Made by Oxford Nanopore Technologies
- **$1000** starter pack vs $1 million USD for the alternative
- Utilizes Nanopores (proteins) to gather genomic data
- Easy (ish) to use
- Results immediate so the user can take action

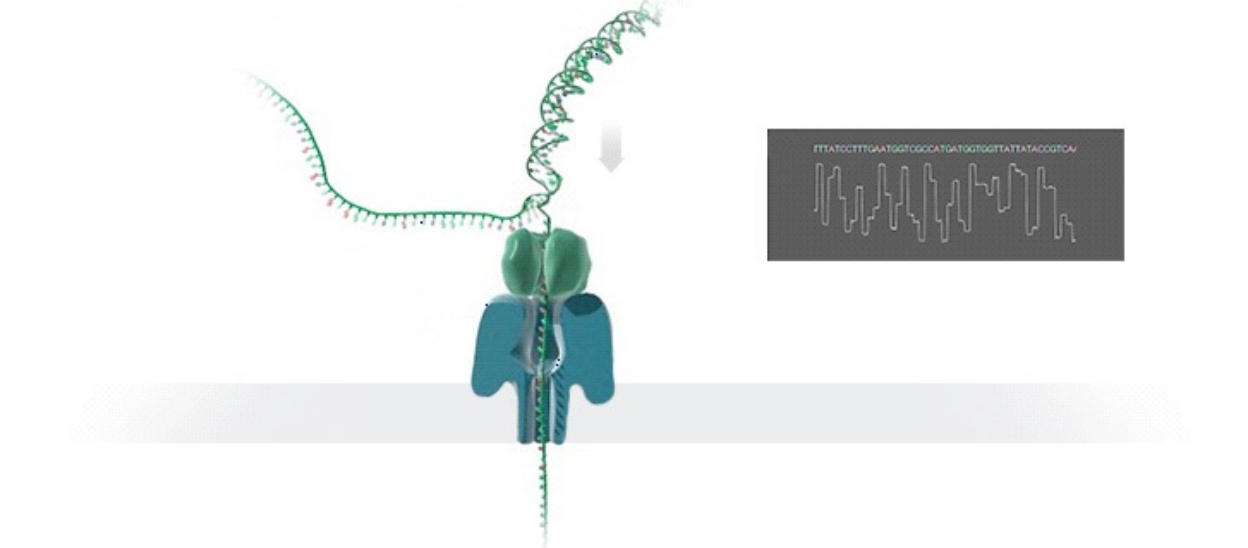
Oxford Nanopore Sequencing



Molecular Systems Biology, Volume: 19, Issue: 8, First published: 16 June 2023, DOI: (10.15252/msb.202311686)

How does this DNA sequencer work?

- A strand of DNA passes through a nanopore. The change in current is detected as the A, T, C and G pass through in different combinations.
- ~400-450 bp per second



From lab to portable genomics lab

9 months later…

Asha and her group have enough cassava to feed the entire village!

Portable genomics is NOT just the sequencer!

- Sample
- Extraction
- Sequencing
- Data Analyses
- Data Storage
- FAIR

Sample

Plant,
animal
microbes & viruses
unknowns

Extraction

Microgem
Biomeme
Claremont BioSolutions
Qiagen Dneasy Kit
Mini PCR
Bento lab

Sequencing

Oxford Nanopore Technology

Data Analyses

Locally:
Base Calling
Building custom databases

Cloud:
EPI2ME
CZ ID's Impact
MemVerge

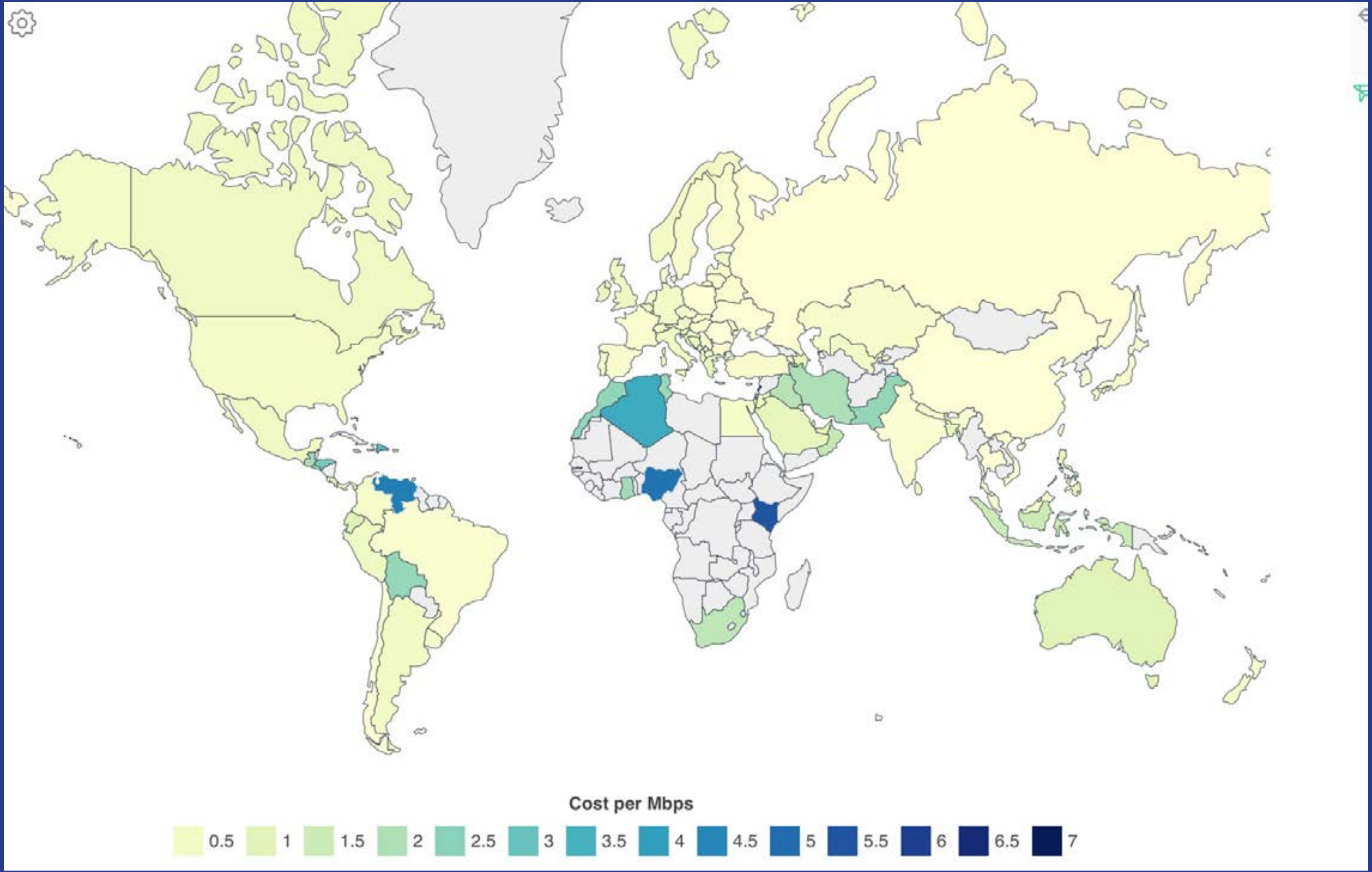
Data Storage

Hard drives
Local repositories (lab computers)
Global repositories
Cloud

Ensuring FAIR

Sensitizing and engaging local communities
Adopting metadata standards

Cloud struggle- internet is expensive



Source: World Population review

Data struggles

- Need for customized pathogen databases
- Need for portable compute
- Need for on prem HPC resources
- Need for cloud compute workflows once network is established
- Raw data is reanalyzed frequently
- Current DB resources are geographically limited
- Development of FAIR


AFRICAN

Build communities- FAIR

Tanzania

Democratic Republic of the Congo

Sierra Leone

Kenya

Trends from the portable genomics trenches

- Offline Nanopore analyses pipelines are needed
- Power is limited
- Internet is limited
- Very few studies are equitably designed
- Care needs to be taken with portable genomics or “labs in a suitcase”
- Flying in with reagents doesn’t help
- Reagents are not available OR they are too expensive in LRC
- Local governments need to be engaged early on to help with importing
- Data is power- it needs to be stored and processed locally- FAIR.

My challenge to you, to us- be an equity designer

- Work locally with communities if you can
- Build inclusive platforms
- AI ethics and more AI ethics
- Please remember power and internet are a privilege
- What are you doing in your space to work on being more globally inclusive?
- Who is speaking at your events?
- We have the chance to get this right, but it will take all of us.

- Let's put the "global" into

  "Building a **Global** Network for Precision Medicine"

BioTeam
Accelerate Science

Thank You
Q&A

Booth #404