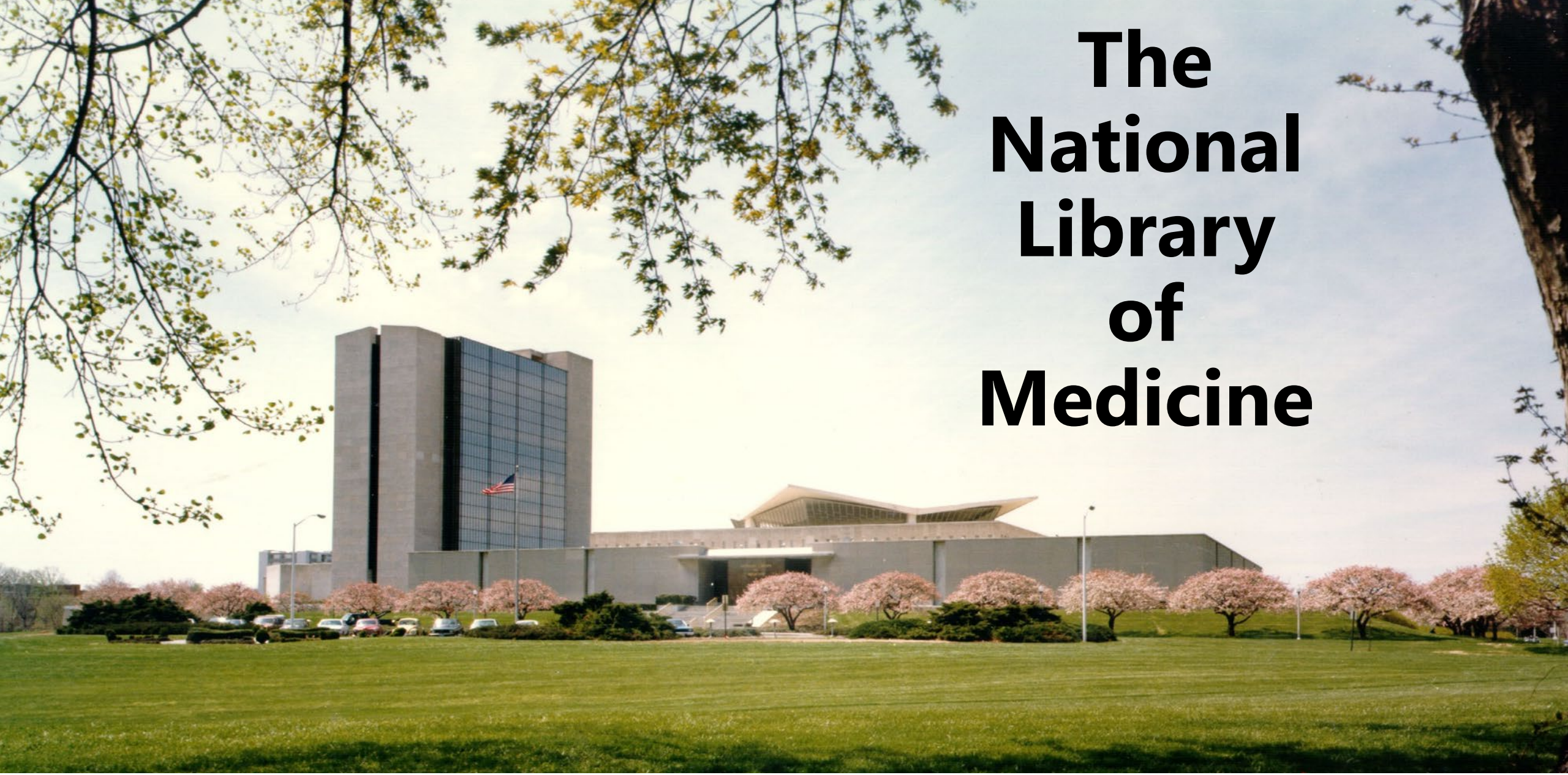
The *Digital NIH* strategy and a glimpse at ChatGPT

Patricia Flatley Brennan, RN, PhD
Director, National Library of Medicine
Co-Chair, Enterprise IT Council
Presented to the NINR Council May, 2023



The National Library of Medicine

NIH National Library of Medicine

Serving Science and Society Since 1836

Research Enterprise for Biomedical Informatics & the World's Largest Biomedical Library

1836-1968: A Collection of Books and Journals

With roots in the office of the U.S. Army Surgeon General, Congressional authorization moved NLM to the Public Health Service. In 1965, NLM opened its doors on the NIH Campus in Bethesda, MD.

1968-2000: Foundation of a Modern Library

Expansion and impact in the Information Age.

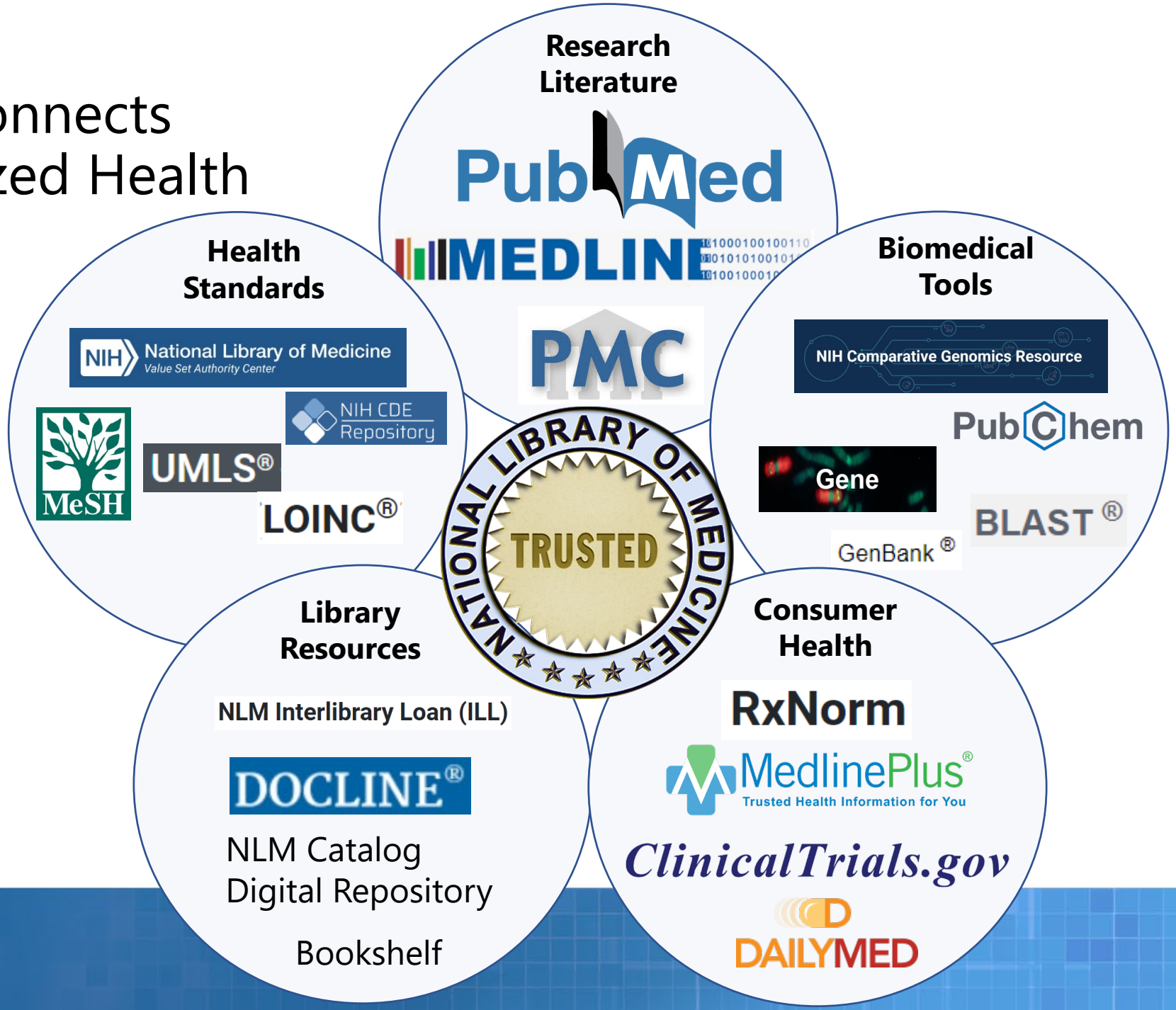
- Lister Hill National Center for Biomedical Communications
- National Center for Biotechnology Information

2000-2036: The 21st Century Library

NLM leads innovative research to accelerate NIH's mission and reach scientists and society with trusted health information.

Long-Term Vision:

Knowledge Hub that Connects Everyone with Customized Health Information



National Institutes of Health

Digital NIH: Innovation, Technology, and Computation for the Future of NIH

Patricia Flatley Brennan, RN, PhD
Director, National Library of Medicine

Dennis Papula
Acting CIO

May 2023

National Institutes of Health

"We are seeing a new approach to science in which ***technologies have become an inseparable component of the scientific process****."*

– NIH Researcher

The **role of technology in science** is changing.

Achieving the promise of technologies requires **new funding and governance models**.

NIH REQUIRES NEW WAYS OF DOING BUSINESS BECAUSE…

Expectations of technology that is needed and desired are shifting.

National Institutes of Health

What is *Digital NIH*?

An **adaptive governance** model that aligns IC-specific technology investments with trans-NIH investments and its mission.

The strategy reflects contributions of over 400 NIH staff members and 19 leading sector organizations.

National Institutes of Health

Digital NIH will champion a **new, more enterprise-savvy approach to technology governance** and **decision-making** that reflects holistic, integrated planning across NIH.

The *Digital NIH* strategy aims to…

Provide rapid computational processing, flexible data storage and retrieval, and advanced analytical functions to support scientific endeavors.

Enable all ICs to be at the forefront of rapid changes in science and technology.

Maximize operational efficiency of scientific processes and technologies.

Adopt innovative frameworks that deliver faster, more automated, and more cost-effective digital technologies.

Digital NIH is a **framework to define high-priority capabilities** and **manage NIH technology investments** across functional areas common to all ICs to achieve its target outcomes.

The *Digital NIH* framework applies across…

EXTRAMURAL RESEARCH MANAGEMENT

INTRAMURAL CLINICAL & BASIC RESEARCH

ADMINISTRATION AND MANAGEMENT

CROSS-CUTTING CAPABILITIES

NIH must also look at enabling capabilities and solutions that cut across functional areas.

Extramural Research programs today are larger in scale, scope, and complexity than ever before.

National Institutes of Health

EXTRAMURAL RESEARCH MANAGEMENT CAPABILITIES…

- Enhanced eRA to conduct comprehensive and holistic portfolio views
- Technologies to help implement critical policies
- Artificial intelligence and machine learning algorithms incorporated in alert systems
- Digital tools for NIH's stewardship responsibilities

…SHAPE FUTURE IMPACT

- Cutting-edge integrated digital platforms for portfolio management
- Reduced researcher and staff burden
- Streamlined planning and tracking across the grant lifecycle

The pace and scope of NIH's progress in improving health are made possible by decades of first-class, NIH-funded **Intramural Basic and Clinical Research**.

National Institutes of Health

INTRAMURAL RESEARCH CAPABILITIES…

- Collaboration portal for researchers to connect on related data sets
- IRP-wide licensure for commonly used scientific software
- Suite of tools to support clinical trial management
- Increased computing power for large complex data sets

…SHAPE FUTURE IMPACT

- Researchers can access holistic information on NIH's research portfolio
- Researchers can design, conduct, and manage studies easily and securely
- Patient-reported needs are represented at the core of NIH's clinical research

The future of **Administration and Management** at NIH will be enabled through common platforms that can be tailored to specific IC needs.

National Institutes of Health

ADMINISTRATION AND MANAGEMENT CAPABILITIES…

- Management platforms that can be tailored to IC's unique needs
- Analytic, reporting and visualization tools
- Automation and digitization of administrative workflows
- Learning systems and tools

…SHAPE FUTURE IMPACT

- Optimized and efficient business processes for common administrative actions
- Reduced manual administrative workflows and processes

Cross-cutting Capabilities set standards for system interoperability and build core, modern technical solutions for all of NIH to enhance NIH's capacity to leverage new types of technology.

CROSS-CUTTING CAPABILITIES…

- Common architecture with well-defined standards to enable integration
- Innovative, cutting-edge storage, analytics, and computational infrastructure
- Increased technically competent workforce
- Technology to support anywhere, anytime workplace of the future
- Risk-based, embedded cybersecurity protections

…SHAPE FUTURE IMPACT

- Established standards for system interoperability
- Modernized technical solutions for all NIH ICs

Digital NIH provides an enterprise-savvy framework to **governing technology investments and data storage** at the IC and enterprise levels in an **accountable** way.

Implementing ***Digital NIH*** is a **multi-year journey that will iteratively explore solutions** to prioritized capabilities over the next several years.

Year 1

Establishing infrastructure, developing Capabilities Portfolio, and Roadmap, Beginning Capability Proof of Concept

Year 2

Collecting Lessons Learned consulting Capabilities Portfolio and Roadmap, Continuing subsequent Capability Proof of Concepts

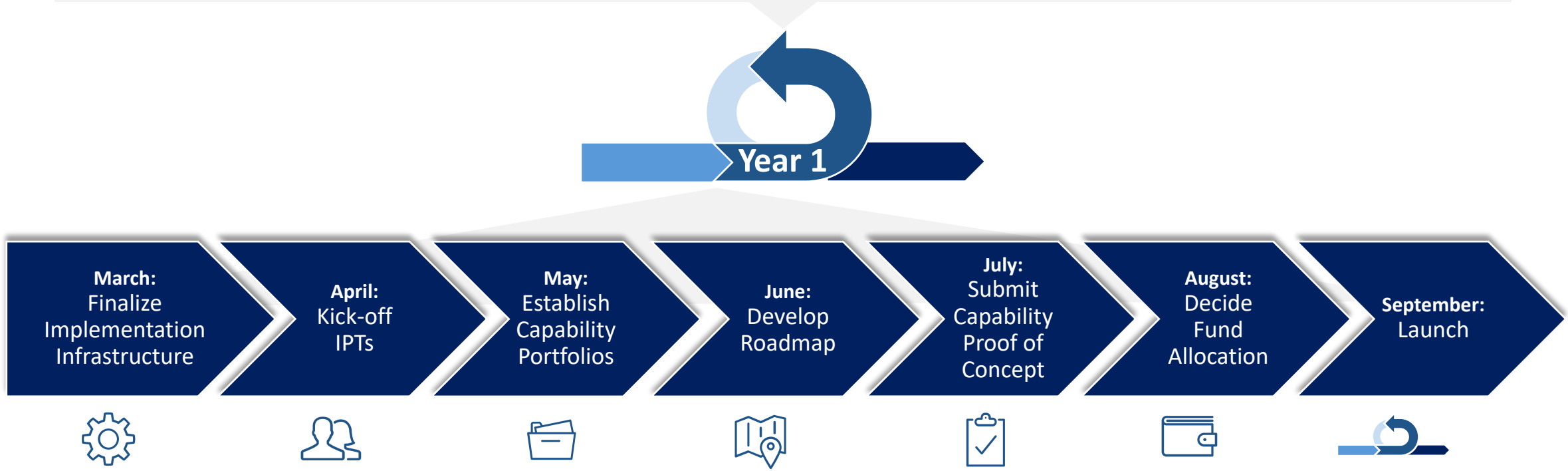
Year 3

Roadmap Refinement, Spinning off Full-Scale Projects, Integration into Operations

Year 4+

Integration into Operations, Revise and Vision next phase

To lay a foundation for the success of ***Digital NIH***, **several milestones and activities** are critical over the next six months.



By implementing *Digital NIH*, NIH can sustain a scientific enterp that is

driven by and empower through technology.

How will NINR bring the strategy to life across NIH?

Key to the success of the Digital NIH Strategy is the recognition that NIH requires new ways of doing business and, most importantly, new approaches to sustaining a scientific enterprise that is driven and empowered through technology.

Moving forward, we must…

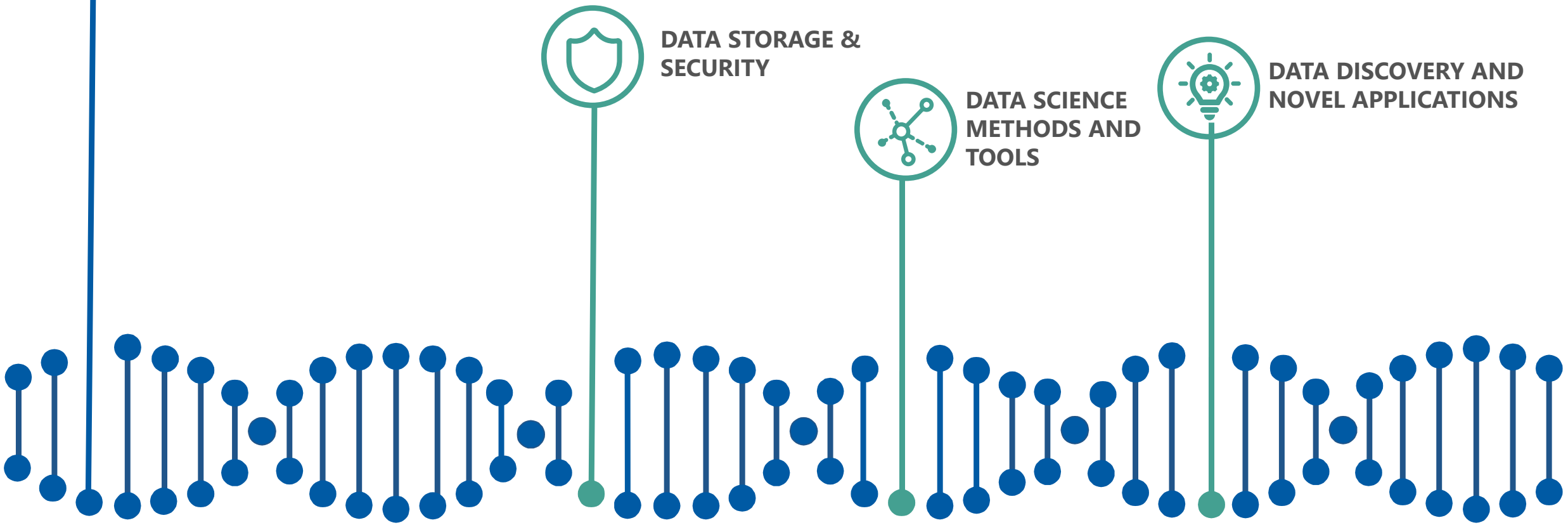
1. Treat **technology as a mission-critical resource**
2. Apply **new holistic and collaborative planning** to give priority to **innovative shared solutions**
3. Uphold the IC's ability to meet **unique IC** technology needs

This requires a **culture change** that will not be easy but **is critical to success moving forward.**

Examples at NINR could include:

- **Evaluate** existing technology solutions before we pursue new acquisition or development
- **Serve** as a Center of Excellence where we have technology or tools that can benefit other ICs and the enterprise
- **Work** with Implementation Planning Teams to identify and stand-up capabilities over the next five years
- **Enable** powerful research through our support of common architecture, standards, data interoperability, AI, machine learning, and process automation

Digital NIH will provide the light-weight infrastructure needed to **modernize the NIH biomedical data-resource ecosystem** through the ***2023 Strategic Plan for Data Science***



National Institutes of Health

NIH 2023
Strategic Plan for Data Science

NIH 2023-2028 Strategic Plan for Data Science

National Institutes of Health

GOAL 1

Capabilities to Sustain the NIH Data Management and Sharing Policy

GOAL 2

Programs to Enhance Human Derived Data for Research

GOAL 3

New Opportunities in Software, Computational Methods, and Artificial Intelligence

GOAL 4

Support for a Federated Biomedical Research Data Infrastructure

GOAL 5

Strengthen a Broader Community of Data Science

Cross-Cutting Themes Strategic Plan for Data Science

National Institutes of Health

Support capabilities that will develop and adopt common services, tools, workflows, and standards

Increase data discovery and broaden the use of clinical and healthcare data, while preserving participants' rights

Leverage resources, standards and capabilities from industry and other agencies

Integrate ethics, policy, health equity, and transparency in the development of data science methods, and tools

Engage researchers and communities in data science training across biomedical and behavioral disciplines

National Institutes of Health

Questions?

For questions on *Digital NIH*, Reach out to the *Digital NIH* mailbox (digitalnih@od.nih.gov) or Patti Brennan (patti.brennan@nih.gov)

For questions on 2023 Data Science Strategy, reach out to Susan Gregurick, Associate Director for Data Science (susan.gregurick@nih.gov).

National Institutes of Health

Generative AI in the Scientific Research Landscape

Based on a presentation by Dianne Babski, NLM
Associate Director for Library Operations

Hype or Hope??

National Institutes of Health

ChatGPT Heralds an Intellectual Revolution

Generative artificial intelligence presents a philosophical and practical challenge on a scale not experienced since the start of the Enlightenment.

By Henry Kissinger, Eric Schmidt and Daniel Huttenlocher
Feb. 24, 2023 2:17 pm ET

GPT-4 Offers Human-Level Performance, Hallucinations, and Better Bing Results

How ChatGPT Has, and Will Continue to, Transform Scientific Research

AI Breakthrough: ChatGPT can almost pass US Medical Licensing Exam, study finds

Disinformation Researchers Raise Alarms About A.I. Chatbots

Researchers used ChatGPT to produce clean, convincing text that repeated conspiracy theories and misleading narratives.

GPT Defined

(GPT = Generative Pretrained Transformer)

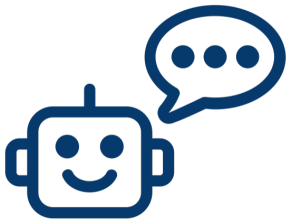
- A **Chatbot** is a tool "[d]riven by [artificial intelligence], automated rules, natural-language processing (NLP), and machine learning (ML)…[to] ***process data to deliver responses to requests of all kinds***." Artificial intelligence (AI) "broadly refers to the idea of computers that can learn and make decisions in a ***human-like way***." https://wame.org/

- **Generative AI *creates something completely new* -** data, images, sounds, or other types of information - similar or related to the training data they were trained on.

- **Large Language Model (LLM)** can ***recognize, summarize, translate, predict and generate*** text and other content based on knowledge gained from massive datasets. A **multimodal LLM** means it can respond to both text and images.

What makes Generative AI different than other AI?

Conversational language



Easier for people to communicate with machines

Creative outputs

Generates novel content in response to prompts (writing, art, music, etc.)

Highly versatile

Useful for a variety of applications and use cases

Know Before You Go: Content entered in chatbots, including personal Information, are used to improve model performance, unless users opt out. Examples: Privacy Policy and Terms of Use

Generative AI *Capabilities* in Scientific Landscape

- Analyze large and complex datasets more efficiently.
- Identify patterns and trends that may not be visible to the human eye.
- Generate new ideas and hypotheses, potentially leading to new discoveries.
- Assist in designing experiments and optimizing experimental parameters.
- Facilitate the development of new drugs and materials.

Generative AI ***Limitations*** in Scientific Landscape

1. **Accuracy and reliability:** particularly if the training data used to develop the system is biased, incomplete or not factoring developments and complexities unique to science. This can lead to errors in scientific research and undermine the credibility of scientific publications, especially when large language models “hallucinate”.
2. **Reproducibility:** particularly if the underlying algorithms are complex and not well-understood. This can make it difficult for other researchers to verify or build upon the results generated by the AI system.
3. **Authorship and attribution:** whether generative AI systems can be considered authors of scientific publications and whether they should be given credit for their contributions to the research.
4. **Ethical considerations:** particularly with regard to issues such as data privacy, bias, and accountability. Hypotheses or ideas may not be ethical to pursue.
5. **Intellectual property:** particularly if the system is trained on copyrighted or proprietary data. Copyright stays with machine.
6. **Cost and accessibility:** particularly for researchers and institutions with limited resources. This can lead to disparities in scientific publishing and hinder scientific progress.
7. **Interpretability:** particularly for results generated by deep learning or generative models difficult to validate by other methods.

This list was generated in a ChatGPT dialogue

Generative AI as a author of scientific papers? The Guidance says NO!

National Institutes of Health

ICMJE
INTERNATIONAL COMMITTEE of MEDICAL JOURNAL EDITORS

ICMJE Authorship Criteria:

1. Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND
2. Drafting the work or revising it critically for important intellectual content; AND
3. Final approval of the version to be published; AND
4. Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

WAME
world association of medical editors

WAME Recommendations:

1. Chatbots cannot be authors.
2. Authors should be transparent when chatbots are used and provide information about how they were used.
3. Authors are responsible for the work performed by a chatbot in their paper (including the accuracy of what is presented, and the absence of plagiarism) and for appropriate attribution of all sources (including for material produced by the chatbot).
4. Editors need appropriate tools to help them detect content generated or altered by AI and these tools must be available regardless of their ability to pay.

@NLM_NIH
@NLMdirector

patti.brennan@nih.gov