

**Chief Research Information Officer  
Strategic Review**

*Working with biomedical researchers across the institution  
to change the world from The Ohio State University*

**Timothy R. Huerta, PhD, MS**

Chief Research Information Officer (CRIO)

The Ohio State University Wexner Medical Center and College of Medicine

**Professor**

College of Medicine Department of Family Medicine

College of Medicine Department of Biomedical Informatics

College of Public Health Division of Health Services Management and Policy

## Table of Contents

---

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b><i>Executive Summary</i></b> .....   | <b>3</b>  |
| <b>2</b> | <b><i>Introduction and Principles</i></b> .....                               | <b>4</b>  |
| 2.1      | <b>Mission</b> .....  | <b>4</b>  |
| 2.2      | <b>Vision</b> .....   | <b>4</b>  |
| 2.3      | <b>Values</b> .....   | <b>5</b>  |
| <b>3</b> | <b><i>The Research IT Action Plan</i></b> .....                               | <b>5</b>  |
| 3.1      | <b>Leverage existing research infrastructure</b> .....                        | <b>6</b>  |
| 3.1.1    | Facilitate the use of Epic as a platform to support discovery .....           | 6         |
| 3.1.2    | Secure high-performance computing to support basic sciences .....             | 7         |
| 3.1.3    | Improve collaborative research capacity through resource identification ..... | 8         |
| 3.2      | <b>Develop New Research Infrastructure</b> .....                              | <b>10</b> |
| 3.2.1    | PARTNER: The PARTNERship for Enhancing Research .....                         | 11        |
| 3.2.2    | PROMPT: The Patient Reported Outcome Measurement Preference Tool .....        | 11        |
| 3.2.3    | SQUIRE: Seeking Quality Improvement for Research Evaluation (SQUIRE) .....    | 12        |
| 3.2.4    | Advance data as a strategic resource .....                                    | 14        |
| 3.3      | <b>Engage the research community</b> .....                                    | <b>19</b> |
| 3.3.1    | Ensure IT is an accelerator and multiplier, not a barrier or inhibitor .....  | 19        |
| 3.3.2    | Enhance engagement with the research community .....                          | 24        |
| 3.3.3    | Partnerships .....  | 24        |
| 3.4      | <b>Restructure Research Information Technology Infrastructure</b> .....       | <b>26</b> |
| 3.4.1    | Research Informatics: Electronic Data Capture Core .....                      | 26        |
| 3.4.2    | Research Informatics: DataCore .....  | 27        |
| 3.4.3    | Research Software Development and Operations (Dev/Ops) .....                  | 27        |
| 3.4.4    | Data Science .....  | 28        |
| 3.4.5    | Research Infrastructure Development .....                                     | 28        |
| 3.4.6    | Revised Organization Chart .....  | 29        |
| 3.4.7    | Budget .....  | 30        |
| <b>4</b> | <b><i>Conclusion</i></b> .....  | <b>31</b> |

## 1 Executive Summary

*Working with biomedical researchers across the institution  
to change the world from The Ohio State University*

This two-year plan is designed to equip researchers at The Ohio State University to accelerate biomedical discovery with local, regional and global relevance. This plan articulates four interdependent areas of focus that, when deployed in concert, will increase the university's collective ability to engage in research.

1. **Leverage Existing Research Infrastructure** to facilitate rapid discovery.
2. **Develop new Research Infrastructure** to advance new areas of research.
3. **Engage the Research Community** to utilize that infrastructure effectively.
4. **Restructure Research Infrastructure** to more effectively meet the needs of basic, clinical, and translational biomedical research.

When appropriately deployed, this plan will transform how research is supported and advanced by our infrastructure, increasing its quality, scale, flexibility and responsiveness.

## 2 Introduction and Principles

*Working with biomedical researchers across the institution  
to change the world from The Ohio State University*

This strategic plan outlines the vision and plan for building the information technology infrastructure required to facilitate biomedical research with local, regional and global relevance. As one of the largest universities in the United States, Ohio State is uniquely positioned to leverage its research and clinical excellence toward improving the health of both the populations we serve directly in central Ohio and those we serve by extension through groundbreaking research.

Our work requires addressing increasingly complex problems via innovative methodologies. This in turn requires innovative and reliable technology infrastructure. This plan outlines a series of investments to the university's research information technology (RIT) infrastructure to facilitate our research and translational aims. It details the initial investments and capacities required to accelerate discovery across our biomedical mission. Our ability to meet current and future research and translational challenges will be enhanced, or inhibited, by the effectiveness of our technology platforms, processes and services.

Just as any organizational endeavor is influenced by its leadership, a clearly articulated vision and mission serve as touchstones for goal-setting and the values we choose to hold ourselves to indicate the means by which we seek to accomplish them. Staff across our unit developed these principles to guide the efforts outlined in this document.

### 2.1 Mission

The Ohio State University Health Sciences campus provides a research technology infrastructure to support discovery as both an experimental and experiential laboratory to advance biomedical discovery. Working in concert with researchers across our campus, our mission is to put technology in the effective service of conducting research at Ohio State to save and improve lives.

### 2.2 Vision

Our intention is no less than to transform how our research is supported and advanced by our infrastructure, increasing its quality, scale, flexibility and responsiveness. We aim to transform the relationship between IT and the research community, acting as a trusted partner to both. Finally, we will work to advance the operational and translational efficacy of our research through improving the *practical* efficacy of our IT infrastructure. All of our initiatives advance the medical center's strategic plan:

- Become a national leader in biomedical breakthroughs and translating research into health care solutions.
- Significantly increase extramural funding for health sciences research.
- Recruit and retain the nation's best scientists and physicianscientists.
- Increase the quantity and quality of interdisciplinary biomedical research space and infrastructure.
- Strengthen basic and translational research programs.
- Build strategic partnerships to enhance the research portfolio.
- Focus on Precision Medicine Research to pinpoint the molecular underpinnings of disease and spur the discovery of novel therapies.

## 2.3 Values

To deliver on this vision, we commit to the following guiding principles:

| Our Shared Value | We commit to:  | We will:   |
|------------------|--|--|
| Transparency     | being open and honest in our interactions with the research community and with IT  | work to identify and clearly communicate how our efforts create benefits, as well as costs, toward empowering researchers to make informed choices                       |
| Accountability   | working diligently to maximize benefits, minimize costs, and respond to feedback   | focus our efforts to prioritize effectiveness in supporting discovery  |
| Engagement       | embracing a robust and researcher-centric communication structure and strategy   | actively engage the research community and IT, and we will embrace platforms and processes that improve shared understanding to advance research                         |
| Capacity         | working to developing and deploying infrastructure to support discovery, including through collaboration and partnership | reach across the university and beyond to identify and leverage opportunities to expand and improve the efficiency and effectiveness of technology services for research |

## 3 The Research IT Action Plan

Establishing the Department of Research Information Technology under the leadership of the Chief Research Information Officer (CRIO) affords new ways of thinking and strategic decision making to our research technology infrastructure. It creates a vehicle for making choices to increase our efficacy, efficiency, and competitive advantage. The CRIO's role, specifically, is to:

- identify the Ohio State's Health Sciences research community's current and emerging needs;
- assess the state of the art and trends in information technology resources and practices; and,
- advance the capabilities and strategic orientations required to meet the needs of health research information technology capacity, in ways that are adaptive and cognizant of environmental challenges.

Success in these endeavors requires leveraging our existing competencies as well as identifying and exploiting new opportunities. Leveraging current technologies to facilitate discovery allows us to make the most of Ohio State's significant resources, such as a deep understanding of the Epic platform that can and should be used to facilitate clinical discovery. Similarly, science frequently involves exploring new domains and testing new hypotheses, and in turn creates unprecedented infrastructure needs. New technologies can offer new structures, processes, and outcomes that create the potential for changes that could not be easily achieved within our current technological framework. The decision whether to use our internal resources

or look outside our walls is a strategic choice that influences our collective research capacity and ability to lead.

In consultation with the College of Medicine and the Medical Center, the CRIO has four operational foci for the next two years:

1. **Leverage Existing Research Infrastructure** to facilitate rapid discovery.
2. **Develop new Research Infrastructure** to advance new areas of research.
3. **Engage the Research Community** to utilize that infrastructure effectively.
4. **Restructure Research Infrastructure** to more effectively meet the needs of basic, clinical, and translational biomedical research.

### 3.1 Leverage existing research infrastructure

Ohio State has, for years, invested in research infrastructure that positioned it as a national leader in research. Frequently, a scientist's ability to execute their research vision is dependent on their access to enabling technologies. However, time spent by researchers building new infrastructure often duplicates existing resources, does not represent an investment in future capacity, and fails to produce sustainable or scalable results. Leveraging our existing infrastructure will be done in three ways:

1. Facilitating the use of Epic to support discovery.
2. Securing High-Performance Computing to support the basic sciences.
3. Improving Collaborative Research Capacity through resource identification

#### 3.1.1 Facilitate the use of Epic as a platform to support discovery

Ohio State has had to choose tools to support both our clinical and research mission. Epic is our organization's single largest IT investment. In addition to its central and critical role in delivering care, it enables information flow and communication, facilitates clinical trial management, and shapes how we engage in clinical discovery. Epic has in turn relied on Ohio State as a partner in the development of its technology since its adoption at the medical center. We have made great investments in Epic components that support and facilitate research, but must consider where our current practices are rooted in legacy technologies or have otherwise been superseded by better enabling technologies – many of which have been developed in concert by Epic and their partners, such as:

- *Cosmos*: a database that can facilitate discovery from the shared experience of participating populations.
- *Caboodle*: a simplified data model to facilitate discovery.
- *Cogito*: an analytics platform that uses learning tools such as machine learning

To be pioneers, rather than simply users, requires forethought and planning as we leverage these new tools. This includes investments such as complementing tools with researcher training. It also means handling governance issues, including those around data sharing, to expand secure access to non-clinicians, developing protocols to enable honest brokers to quickly deliver data via the latest tools, and ensuring that researchers understand how Epic's structure shapes the data it produces.

Changing our organizational approach to clinical data will be require a substantial effort requiring engaging a significant range of stakeholders. It will require that we establish best practices to streamline interventional research that leverages our electronic health record (EHR) capabilities, and doing so in a manner that is always respectful to all stakeholders – patients,

physicians, nurses, IT staff, non-clinician researchers, students, the medical center and the university. Balancing the need to be efficient with the need to be respectful requires committed governance that complements existing structures – for example, IRB authorization does not in and of itself guarantee access to data, but multiple layers of approvals and opaque authorization processes can inhibit important efforts toward discovery.

Moving forward, the College of Medicine will identify the research pain points, bureaucratic challenges and dysfunctional systems and support thoughtful, research-aware policies throughout the organization. These will be emergent but are core to the success of the CRIO in effectively facilitating meaningful research across the enterprise.

### 3.1.2 Secure high-performance computing to support basic sciences

Fast, well-managed, secure and accessible research computing resources are another key prerequisite for advancing the research mission. Increases in datasets' size and complexity create incentives to capitalize on several opportunities and best practices in analytic computing:

1. Virtual environments leverage economies of scale, are more flexible and resilient than desktop or laptop hardware, and reduce the total hardware resources needed to support multiple projects and research teams.
2. Multithreaded execution allows accelerating analysis via parallel processing. The price-performance curve for computer processors is not linear, and statistical analysis software is increasingly designed to take advantage of parallelization.
3. Growing data creates an increased demand for storage. Genomic, administrative and telemetric datasets are measured in terabytes, outstripping the ability for laptops and desktops to conduct effective analysis. Managed storage arrays, such as the medical center's distributed file system (DFS), hold data far more efficiently.
4. Fast computers and efficient storage must be complemented by fast connectivity. Office and laboratory networks are up to a hundred times slower than those in the medical center's datacenters. Computing environments that put computation and storage together in higher-bandwidth network environments improve efficiency by reducing bottlenecks.
5. Finally, virtualizing research computing reduces the likelihood of information breaches as the data never leaves the data center – the researcher's desktop, laptop, or tablet simply provides the screen and keyboard.

Leveraging these opportunities yields faster, safer and more robust research computing. The **Computational High-Performance System (CHiPS)** is a collaborative effort between the Office of Research and Wexner Medical Center IT (WMC-IT) to invest resources in scalable, sustainable, high-performance computing. The CHiPS infrastructure is under development to facilitate the efficient repurposing of legacy infrastructure supporting the medical center to create computational clusters to facilitate research requiring high-performance computing. Currently, we are testing a proof-of-concept computational cluster with the capacity of about 175 concurrent processors.

The College of Medicine has made an initial investment in high performance research storage capacity through the purchase of an 800 TB Hewlett Packard storage device. Termed the **Storage Array for Large-Scale Analytics (SALSA)**, this investment provides a large-scale *working* storage solution that is compliant with personal health information (PHI) requirements. It provides the fast, massive storage required for efficient data manipulation.

Finally, in a proof-of-concept effort, we have instantiated a graphics processing unit-based

system that is PHI compliant. The **Graphics processing Unit Analytics Compute resource (GUAC)** will allow Ohio State to safely engage in artificial intelligence and machine learning using PHI. These resources provide a test bed for large-scale analytics similar to those available at the Ohio Supercomputer Center (OSC), which is not yet available for analysis of data containing PHI. While we are simultaneously building a relationship with OSC to enable HIPAA-compliant computation, providing this resource locally will facilitate pilot testing and serve as a springboard for sophisticated analyses complemented by “high touch” support.

These three components will work together to:

- Reduce the effort and project resources required to engage in high performance computing;
- Increase data security by consolidating research data and computation on a centralized platform and keeping it off users’ devices; and,
- Improve support by making the platform accessible to IT staff 24/7/365 even if researchers are distributed over the globe,

CHiPS, SALSA and GUAC will support basic science research across the organization, including work in precision medicine, genomics, genetics, simulation and artificial intelligence.

### 3.1.3 Improve collaborative research capacity through resource identification.

The value of Research IT sometimes lies more in solving common procedural and organizational problems, often tedious ones, than in providing cutting-edge tools. Resource discovery and inventory are a great illustration. Simple tools can be built or otherwise leveraged to illuminate, track and analyze our research capacity. Two examples of this approach are a tool for resource identification and management (Eagle-i), and a tool to track research trainees in support of education and training grants (Training Tracker). Research IT will continue to identify ways to leverage its resources to improve our collective collaborative research capacity.

#### *3.1.3.1 Training Tracker*

National Institutes of Health (NIH) Institutional Training Grant applications (T32 and T15) require a significant investment of time and effort. In response to an inquiry by the Ohio State Department of Surgery, and subsequent to Dr. Huerta becoming the CRIO, Dr. Huerta led the development of an online tool to manage trainee data related to T32 projects. We presented this to the Department of Surgery and have been working on the development of the prototype.

More recently, the Ohio State University Health Sciences Libraries has provided resources (0.5 FTE) to support ongoing development. We have expanded the stakeholder group to include the Center for Clinical and Translational Science (CCTS), the Biomedical Sciences Graduate Program; and the new Center for Cancer Mentoring, Education, Leadership and Oncology-Related Training (CAMELOT) program created by The Ohio State University Comprehensive Cancer Center – Arthur G. James Cancer Hospital and Richard J. Solove Research Institute. Through the support of the Health Sciences Library we have outlined a plan to improve our ability to efficiently meet current and future training programs’ reporting requirements. The tool will have the flexibility to accommodate a variety of reporting requirements; will store data in a way that is agnostic to report formats; and can facilitate use of the data for planning, evaluation and reporting.

#### *3.1.3.2 Eagle-i*

The Ohio State University Office of Research formed the Facilities and Infrastructure Working

Group (FIWG) to identify and evaluate platforms. FIWG reached out to Dr. Huerta before he assumed the CRIO role and asked how we might build a catalog of research resources through the development of new software. Dr. Huerta noted that other software already existed and that a limited pilot of existing software would allow us to test if, and how, a resource catalog might be leveraged to increase organizational research capacity and collaboration. Subsequently, FIWG accepted this recommendation and has proposed to the Office of Research to pilot the adoption of Eagle-i<sup>1</sup>, an open-source application developed by Harvard University for cataloging and searching research resources of all kinds.

College of Medicine Research IT has offered to lead the effort as a member of the FIWG, using the College as a pilot site to:

- Develop a combined needs assessment and resource identification survey to identify what we have and what current faculty indicate we need. The CRIO will deploy this to faculty and staff throughout the College of Medicine.
- Ensure the basic functionality and suitability of Ohio State's current Eagle-i instance<sup>2</sup> with regard to the instance's own resources and configuration.
- Catalog resource data obtained via the survey and other reviews of resources into the Ohio State instance of Eagle-i.
- Develop workflows to facilitate resource discovery via the creation of an engagement survey to match potential new faculty with collaborative opportunities leveraging our current resources.
- Report the outcome of the pilot and makes recommendations for the value of a university-wide implementation.

This assessment will allow us to evaluate how such a system can be used to support research. The information gathered should provide a method to analyze resource gaps and resource strengths within the college as a model for how it can be used across the university.

---

<sup>1</sup> <https://harvard.eagle-i.net/>

<sup>2</sup> <http://eagle-i.rf.ohio-state.edu>

### 3.2 Develop New Research Infrastructure

Translational Clinical and Population Health Research sits at the juncture of *what we know* in clinical science and *what we do* in clinical care (known as “T3 research”), and how clinical care results in improved population health (“T4 research”) as a component of public health. The College of Medicine Research Strategic Plan notes that four of the eight areas of focus – health policy, health services, population health management and behavioral health – sit firmly at or are closely related to this T3 and T4 intersection.

The national focus on T3 and T4 research, particularly the study of the comparative effectiveness of different clinical strategies on patient-centered outcomes, can be traced to research demonstrating substantial variation in the performance of specific medical procedures. External to Ohio State, significant investment has been made to support discovery in T3 research. The Patient Protection and Affordable Care Act, enacted by Congress in 2010, raised the profile of comparative effectiveness research through the establishment of the Patient-Centered Outcomes Research Institute (PCORI), an independent, nonprofit organization whose mandate was recently extended through 2030. PCORI represents one of several funding agencies – NIH, NSF, CDC, DoD, and AHRQ – whose funding announcements weave an underlying narrative that patients are the focus of care and their experiences and perceptions are at least as important to health as clinical endpoints.

As a research university with a robust clinical enterprise, we have the opportunity to leverage our clinical journey as one that is centered around health outcomes that matter to patients, and decisions should reflect the priorities of those who receive care. Within that context, the role of consumer- and patient-focused health technology has opened a new frontier where behaviors are influenced by engagement. Patient engagement has been called the next “blockbuster drug” because of its potential to improve health outcomes and the possibility for generating significant health care savings. With an increasing emphasis on engaging patients as partners in, and often drivers of, their health and health care, patient engagement is seen as a necessary component in achieving the quadruple aim of improved experience of care, improved health of populations, lower per capita health care costs and less clinician burnout. Similarly, the rise of EHRs has allowed for the collection of data that has been previously unavailable. As such, data-driven experiential learning has become a mandate in clinical care.

Leveraging opportunities to learn and apply lessons to improve the delivery of care using experiences throughout the health care system is essential in efforts to transform health care delivery. One coherent conceptual approach advanced nationally has been the Learning Healthcare System (LHS), an organizational approach to health services delivery where efforts to improve efficiency and effectiveness are grounded in experience with the populations served. An LHS seeks to optimize the structure and process of care delivery toward improving the organization’s ability to get the *right treatment* to the *right person* at the *right time*. While these efforts have been greatly supported by increased access to data made available by EHRs, developing a system in which feedback informs practice and research drives improvement remains challenging. The CRIO proposes a three-pronged approach to the development of a robust research infrastructure to support discovery.

First and foremost, we propose an Ohio State-developed, research- focused platform called the **Basic Research Using Technology to Understand the Science of patient engagement (BRUTUS)** platform to support clinical discovery in patient-reported outcomes, patient-generated data, patient preferences and clinical engagement. BRUTUS comprises three components – PARNTER, PROMPT and SQUIRE – and is conceived as the technology

environment necessary to support an LHS. Second, given the reliance of research on high-quality data, the university must establish an approach for leveraging data as a strategic resource to enable efficient access to data resources to support research. Finally, there is a collective need to develop competencies in technologies that we currently do not have. Chief among them, due to NIH's requirements (NOT-OD-19-150), is our capacity to support Fast Healthcare Interoperability Resources (FHIR)-based projects and tools. This will require investment to ensure our clinical research is properly supported as it advances.

### 3.2.1 PARTNER: The PARTNERship for Enhancing Research

**PARTNER** is an IRB-approved protocol analogous to Total Cancer Care (TCC) in Ohio State's Comprehensive Cancer Center – James Cancer Hospital and Solove Research Institute, but with a focus on the general medical population. PARTNER asks patients to join Ohio State in discovery and provides a framework for broad engagement with the patient population to allow data collection to function more efficiently and effectively. Notably, studies that leverage PARTNER can share data collected on the platform via pre-approved shared protocols. Results can support both exploratory and confirmatory research, improving our collective understanding of the patient experience.

The PARTNER protocol includes:

- Broad consent for the use of clinical data for research (excluding Cancer).
- A protocol that supports biobanking of NIH-identified rare diseases not otherwise banked at Ohio State.
- Shared data across studies to allow patient data collected in one study to be used in other studies.

PARTNER supports positioning Ohio State as a research hub and data coordinating center focused on patient engagement in concert with other universities, similar to what we accomplished with the ORIEN protocol. Current partners included via the NSF IUCRC process, discussed later in this plan, are Indiana University Health, Penn State University, the University of Florida, the University of Alabama at Birmingham, the Medical University of South Carolina, and the University of California, San Francisco.

### 3.2.2 PROMPT: The Patient Reported Outcome Measurement Preference Tool

PROMPT is a research platform focused on providing the patient engagement infrastructure to enable discovery focused on patient outcomes. A patient-reported outcome, or PRO, is “any report of the status of a patient’s health condition that comes directly from the patient, without interpretation of the patient’s response by a clinician or anyone else.”<sup>3</sup> PROs assist both healthcare providers and patients in making informed decisions about patient care; they also help in clinical trials to establish intervention efficacy.

A robust PRO collection and engagement approach can provide benefits to care and research that would be highly valuable to the medical center; however, current PRO approaches suffer from a number of issues:

1. They fail to engage patients using the forms of communication with which they are most

---

<sup>3</sup> <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-reported-outcome-measures-use-medical-product-development-support-labeling-claims>

- comfortable.
2. Data collection doesn't adapt to changes in a patient's health.
  3. Collected data is not provided to patients and providers in a manner that informs decision making.

These issues cause low patient engagement and submission to PRO systems, to the detriment of both research and clinical care.

**PROMPT** aims to engage patients using the tools and technologies that are most comfortable for them. Its design addresses the limitations with current PRO approaches by employing features to increase engagement and impact:

1. PROMPT is grounded in shared decision making; the patient and provider agree on the PROs that are of greatest importance in the patient's care.
2. The PROMPT system offers patients options in how they communicate their experiences; options include phone, text, email, web portal, and mobile app.
3. PRO collection will include intelligent patient engagement scheduling to reduce response burden.
4. PROMPT will synthesize PRO data for health care providers to provide clinical decision support that recommends evidence-based courses of action.
5. PROMPT integrates PRO data into the medical record, making it available for use in both clinical decision making and research.

PROMPT proposes to improve population health by creating a robust feedback loop between patients and their health care team, making clinically relevant data available in (near) real time to inform medical decisions regarding a patient's care plan. In so doing, PROMPT creates the opportunity for discovery – specifically related to question of how health care organizations appropriately include this type of data in the care delivery process. PROMPT will make a significant new line of research possible at Ohio State that is grounded in comparative effectiveness research to identify what works best, for whom, and under what circumstances.

### 3.2.3 SQUIRE: Seeking Quality Improvement for Research Evaluation (SQUIRE)

Health interventions in the real world involve complex configurations of persons and organizations. There are currently no standard tools for data collection that address this complexity. **SQUIRE** will provide an ecosystem of tools to effectively gather data from program participants, community members, community-based organizations, and local and state authorities to increase understanding of how public health interventions impact the communities they serve. The SQUIRE tools facilitate discovery in partnership with the community, organizations, across the region and as an NIH-funded Data Coordinating Center (DCC).

Current efforts include developing:

- *Participant-facing tools* that collect data via mobile apps, text messaging and interactive voice response tools, aligning with the same person-centric values as PROMPT.
- *Organization-facing tools* that help organizational research partners collect and share data, as well as gain insights, about how they engage with the populations they serve. These tools enable organizations to record and analyze data about the individuals they serve in cases where the organization does not already have a system in place (for example, the Ohio Equity Institute study's portal for data collection):
  - data management tools to facilitate organization sharing of data such as those



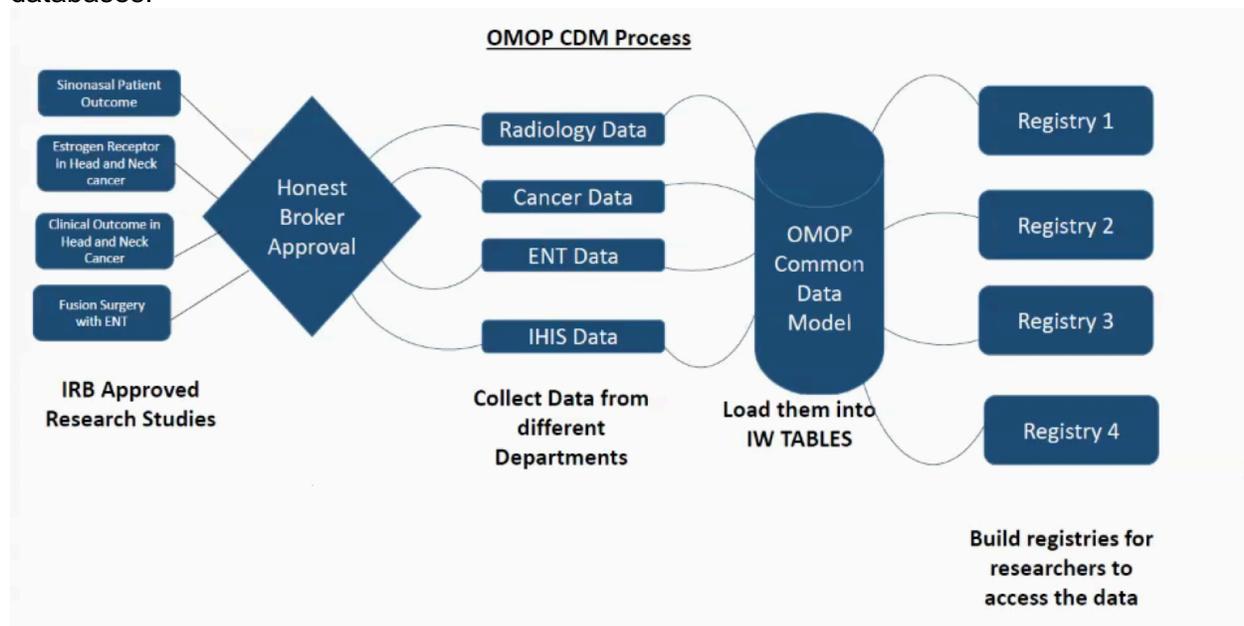
### 3.2.4 Advance data as a strategic resource

Discovery support across the clinical sciences is facilitated through four new resources:

1. SCARLET for All
2. LifeScale (Primary) DataCore
3. Secondary DataCore
4. FHIR for Research

#### 3.2.4.1 SCARLET for All

The Ohio State University Center for Clinical Translational Science, in concert with programmers in the Department of Biomedical Informatics, developed a research registry platform named the Scalable Analytics Registry for Rapid Learning and Translational Science (SCARLET). SCARLET is a secure, web-based application that leverages REDCap and Integrated Health Information Systems (IHIS) for data capture, merging research data sourced from the former with clinical data sourced from the latter. The registry pipeline software takes specific data points from the REDCap database and extracts IHIS data from the Information Warehouse; it then integrates this data into a standard Observational Medical Outcomes Partnership (OMOP) ontology. SCARLET's front end involves a query portal that has the potential to be used and further developed to provide search queries across a wider variety of databases.



Research registries could be requested through the CCTS's service request system, called CoRR, and a standard registry deployment with a standard set of data elements corresponding to the OMOP data could be provided for a base cost of \$10,000. Customization beyond the standard framework, such as capturing data elements beyond the standard data set offered or enabling non-standardized data elements captured through REDCap to be queried through the query portal, will be charged at the standard department's rate of \$110/hour. There also was an associated hosting fee, with change requests/revisions will incur effort charges at the \$110/hour rate.

Research IT will extend SCARLET to "SCARLET for All" by bringing this technology to all eligible research studies by June 2020 at no cost to the researcher. Studies are eligible if their IRB and consent documents allow for full access to the medical record. Starting in July, we will

begin working with the IRB to identify active studies, reach out to the principal investigators and engage them to determine whether access is appropriate. This will reduce the effort required to secure research data for studies and facilitate statistical analysis.

#### 3.2.4.2 *The LifeScale DataCore*

The use of EHR systems has progressed from infrequent, to best practice, to regulatory requirement, and the growth of EHR data has fortunately been accompanied by an increase in their accessibility and usability. Transparency regarding a patient's past medical history supports optimal treatment, reduces avoidable errors and improves outcomes. However, while the current definition of a patient's history is understood to reflect the treatment by physicians, the research is clear – health is determined by a range of biological, genetic, behavioral, social, economic and environmental factors. Yet despite the fact that data regarding all of this information is available, Ohio State, like many other organizations, has neither sought to leverage its collective knowledge nor has it leveraged its strategic advantage to advance discovery. The status quo contains a number of limitations:

- Data about the individual is segregated by professional domain; for example, medical data is not linked to optometry, dental, or community databases.
- Data is segregated by organizational boundaries; for instance, information across the life span of the individual is often siloed because it is held by others.
- Data about the experience of the individual is often not linked; community exposure and behavioral data is often unavailable to researchers who do not have the capacity to gather it.

To address these gaps, we will undertake a three-step process to develop a health sciences-wide resource to create a 360-degree view of the patient:

1. In collaboration with Nationwide Children's Hospital, we will link individuals' clinical records across organizational boundaries to provide an improved view of the experience of individuals from birth and through their life.
2. In collaboration with the health professions schools across the university, we propose to link data across disciplines: dentistry, optometry, pharmacy, public health and, potentially, veterinary medicine.
3. Develop enabling policies, tools and training to support the effective use of these new resources in both identified and de-identified data contexts, specifically focused on two challenges: broad consent and supporting research using machine learning and natural language processing.

With more health sciences colleges on one campus than any other university in the United States, Ohio State is uniquely positioned to discover the keys to making our communities and our world healthier, happier and more productive. As more organizations shift toward capitation and shared risk payment models, a 360-degree view of the patient will help identify care gaps and address high-risk situations. In turn, this supports proactive rather than reactive care, supporting reductions in costs and readmissions simultaneous with higher quality of care and better quality of life.

This approach aligns with efforts to support research that “advances whole-person, 360-degree care especially those with multiple chronic conditions and/or socioeconomic disadvantage.”<sup>2</sup> A greater understanding of the whole person can orient the care team to consider all domains of a person's life when assessing and addressing needs and can facilitate the implementation of evidence-based approaches to identify, understand and overcome barriers to the adoption,

adaptation, integration, scale-up and sustainability of evidence-based interventions, tools, policies and guidelines.

#### 3.2.4.3 *The Secondary DataCore (SDC)*

The SDC is a shared resource for Ohio State researchers that stores and synthesizes large-scale clinical datasets on an easy to use analytic platform to facilitate outcomes research. The SDC aims to reduce the costs associated with data licensing, the time and effort associated with data acquisition, and both the effort and risk of error involved in processing and preparing data for analysis. The SDC streamlines research on secondary datasets to allow researchers to focus on discovery.

The SDC houses large-scale clinical datasets including but not limited to:

- Truven MarketScan
- The Healthcare Cost and Utilization Project (HCUP)
- Centers for Medicare & Medicaid Services (CMS) Claims

The SDC's principal contribution is its shared data model. Using ontology matching, the SDC uses common data elements across datasets to facilitate cross-dataset discovery. This allows researchers to ask their analytic question of many datasets via a single operation rather than having to obtain and prepare each dataset they wish to analyze. The SDC also streamlines reassessing and extending prior findings as new data become available, not only to advance science but providing useful information on trends to health behavior and promotion efforts, public policy, and health services organizations. Additionally, the SDC will advance discovery by rapidly testing localized data against national norms, automating and systematizing what to date has been a fragmented system of discovery.

The SDC seeks to take a data science-driven, systematic, systems-level approach to knowledge discovery with the following five specific aims:

- 1) Construct bundles – which contain the information necessary to assess prior findings associated with the prior research – into a coherent *Research System*;
- 2) Identify when new data become available, flag potential opportunities to explore, evaluate and extend prior research, and notify researchers of same (the *Surveillance System*);
- 3) Develop a system that allows researchers to upload instances of local data for comparison with bundles to allow individuals to explore local data through the lenses of prior research and national data (the *Discovery System*);
- 4) Provide a set of common standardized coding and operationalizations to reduce errors and help researchers make informed and efficient choices;
- 5) Leverage allocated collaborative resources to support both the development of a community of users through outreach and training and broaden the tool's value and reach among trainees and researchers.

The SDC will bring secondary data analysis into the realm of data science at scale, enabling continuous, systematic analysis and interpretation of both health-related data and the evidence those data inform, applicable to the planning, implementation and evaluation of public health practice. With only 3% of published studies using more than one year of data, the SDC represents a novel opportunity for a robust, replicable and open approach to discovery that could

be adopted by journals, researchers and research teams, and training programs that can improve our capacity to engage in high-quality research.

The SDC seeks to transform our approach to the use of secondary data, especially in cases where these data are publicly available and are gathered in series. The SDC is designed to be scalable and generalizable such that it can include other datasets, leveraging the data we already gather to help us understand behavioral phenomena in new ways. Social determinants contribute an estimated 70% to the causes of disease, yet informatics tools to help us better understand those factors remain a largely unexplored frontier and addresses a central mission of NIH and the National Library of Medicine to promote research-driven informatics technology across the development lifecycle to address priority needs in research. Once implemented, the SDC will:

- Support and guide every step of the data acquisition workflow, guiding the access process with regard to cost, rules and requirements surrounding use;
- Present clear listings of available data and their corresponding metadata and data dictionaries, allowing the user to browse and search data holdings in both targeted and general, exploratory ways;
- Allow researchers to use the analysis software of their preference to investigate the data by mating the SDC with high performance virtual computing environments;
- Allow uploading analysis program code to the SDC to be executed against all appropriate datasets specified by the user;
- Automatically update users' working datasets as new data become available;
- Contact SDC users with "push" updates as new data become available.

Taken together, the DataCore helps streamline, simplify, and automate scholarship and discovery:

- All work with the data is covered by a master exempt IRB, meaning researchers will not need to fill out a unique IRB for every SDC project they work on;
- All data undergo cleaning operations to address known errors, reducing the effort and inconsistency resulting from individual investigators and teams performing their own data cleaning actions;
- All data are presented to users in one or more harmonized formats following common ontologies, allowing researchers to focus on discovery rather than converting the data to a consistent format;
- Metadata and documentation are presented to users to help identify commonalities and opportunities to conduct analyses on different patient populations;
- All of the rules, restrictions and costs for datasets are clearly spelled out in a single location, saving researchers from having to hunt down or inquire specifically about every dataset. This also reduces the cost for the researcher by using the data license purchased by the SDC rather than necessitating the individual purchase of the data.

Research IT will support the administration of research using the SDC by providing a list of restrictions and rules in each dataset's use agreement, the cost of access, and a blanket IRB, using the same accessible documentation platforms described elsewhere in this plan.

The SDC is hosted in a single database structured to allow both intra- and inter-dataset analysis, with data standardized across multiple years or waves of the same data source. The SDC also contains metadata that explain what questions are being asked by each data source, what the

responses to those questions mean, and “codebook”-style univariate descriptive statistics that give the user an initial glimpse at completeness and distributional characteristics without the user having to load the data or perform any analysis.

The SDC will include as-received, cleaned and harmonized versions of the following datasets to facilitate outcomes research, empowering researchers to automatically investigate their question of many data sources and time periods:

- HCUP from the Agency for Healthcare Research and Quality: admission-level data about hospital admissions, readmissions, and emergency department use.
- CMS claims data: claims-level data about all claims filed through Medicare.
- American Hospital Association (AHA) Annual Survey (AHAAS) with Information Technology supplement (AHAIT): annualized survey about hospital demographics.
- Health Information National Trends Survey from the National Cancer Institute: person-level survey that asks people general thoughts on cancer and cancer-related topics.
- Patient-Centered Outcomes Research Institute (PCORI) PCORNET: patient EHR data along with PCORI studies in a unified data model.
- Epic Systems’ Cosmos: patient EHR data
- IBM’s Truven: a set of clinical datasets of which one is a patient EHR dataset

In addition, the SDC will empower users to automatically perform apples-to-apples comparisons with alternate dataset variables and automatically query the data commons once new data is available. When a user makes a request of the data commons, a query to provide the data will be procedurally generated to fulfill the request. The SDC can use the query as the basis of an automatic cross- reference operation to determine equivalent questions that could be asked in other SDC sources and provide those matches to the requestor, empowering them to make valid comparisons across the populations contained in all relevant matching SDC datasets.

Traditional analytic computing models require the researcher to adjust and re-execute their code to conduct analyses on updated data. The SDC moves beyond this model in two ways, leveraging server-side technologies to automatically update the requestor’s data with updated data once it’s available and to automatically execute the requestor’s stored analyses on the updated data. The SDC’s backend infrastructure will monitor data holdings and stored analyses, executing the latter on the former and notifying researchers once updated analyses are ready for review.

#### *3.2.4.4 FHIR for Research*

NIH has increasingly looked to standards-compliant approaches to bridge the discovery-implementation gap. Of particular importance is the use of Fast Healthcare Interoperability Resources (FHIR) to accelerate the use of clinical data for research. FHIR is a technology standard for communication across disparate health IT applications through an application programming interface (API), a kind of “lingua franca” for software.

Several Federal health agencies are promoting the use of FHIR in EHR systems. The 21st Century Cures Act requires that a health IT developer or entity "allow health information ... to be accessed, exchanged, and used without special effort through the use of application programming interfaces (APIs) ... including providing access to all data elements of a patient's electronic health record." To implement this provision, the Department of Health and Human Services, Office of the National Coordinator for Health Information Technology (ONC) has

proposed a new rule to support seamless and secure access, exchange and use of electronic health information that calls on the health care industry to adopt standardized APIs by using the FHIR standard to share patient data.

On July 30, 2019, NIH issued a notice ([NOT-OD-19-122](#)) to encourage NIH-funded investigators to explore the use of FHIR to capture, integrate and exchange clinical data for research purposes and to enhance capabilities to share research data. In addition, NIH issued a notice ([NOT-OD-19-127](#)) to small business communities that announces NIH's special interest in supporting applications that use FHIR in the development of health IT products and services.

Separately from the NIH, on September 9 2019, the Agency for Healthcare Research and Quality (AHRQ) also issued a notice ([NOT-HS-19-020](#)) encouraging the use of the FHIR standard. AHRQ believes that the FHIR standard is mature enough to warrant its use for health services research. AHRQ along with ONC is developing FHIR resources to support PRO collection, as well as clinical decision support.

The respective interoperability goals of the notices issued by ONC and CMS align with and facilitate many of the objectives asserted in the NIH Strategic Plan for Data Science, as well as NIH's long-term policy goals for data management and sharing. Additionally, in its 2017-2027 Strategic Plan, NLM proposes technical and scientific advances to ensure that research data are Findable, Accessible, Interoperable and Re-usable.

NIH is currently requesting input on how the FHIR standard could be integrated into NIH funding standards. NIH is exploring researchers' experiences using FHIR, the extent to which researchers plan or do not plan to use FHIR, what tools may be needed to effectively use FHIR, the need for research regarding standards development, and opportunities and challenges with using FHIR. Ohio State's own experience with NIDA has resulting in the head of NIH making requests for plans on how the HEALing Communities Study will deploy FHIR in its intervention. Building that knowledge will enable the organization to be on the leading edge and accelerate our research capacity.

### 3.3 Engage the research community

Ohio State can provide the most advanced technology in support of research, but if accessing that technology causes investigators to become frustrated, angry, or disheartened, they will begin to seek workarounds, subvert security, and search other less fruitful lines of inquiry. A key challenge to the effective use of research IT is ensuring the technology is usable and accessible.

#### 3.3.1 Ensure IT is an accelerator and multiplier, not a barrier or inhibitor

There are a number of reasons underlying why researchers struggle with information and the technology that supports it at Ohio State. Broadly speaking, the policies, procedures, documentation and supports created and deployed by IT are often poorly disseminated, poorly understood, do not reflect the spirit in which they were created, unnecessarily inhibit research, and/or fall out of sync with current practice. The absence of clear policies and procedures has led the institution to two key dysfunctions:

1. What can be achieved is too often a function of *whom you know*.
2. The answer to the question "Can we do this?" is too often a function of *whom you ask*.

Some broad issues in IT disproportionately affect research, and some research activities are disproportionately affected by certain issues. Moving forward, the Chief Research Information Officer represents and advocates for researchers in shaping how information and technology

within is leveraged within the College of Medicine, across the Health Sciences and in the broader academic context. In so doing, the strategic vision for changing the experience of the research community is grounded engagement. Specifically, the CRIO will focus on three areas:

1. **Improving Data Governance across the Research IT enterprise:** A top-tier biomedical research university requires that policies are clearly defined and are developed in a transparent fashion that invites and integrates feedback across the institution. We will work to improve the policies, procedures and communication surrounding research infrastructure, representing and advocating for a research-centric, and researcher-centric, perspective in matters of information and technology governance and decision making.
2. **Developing research-sensitive support for the use and access of RIT tools and technologies:** We will engage researchers through their full tenure spanning recruitment, onboarding, career growth, role transitions and offboarding. We will complement both current and newly created resources with documentation and training that is tailored for a research audience and responsive to its needs.
3. **Engagement with the Research Community:** Information and policy must be made accessible on platforms that are easy to access, simple to use and responsive to research community needs.

We will work with stakeholders across the university to adopt common platforms and processes for developing and disseminating policies, to engage the research community in policy development and review, and to emphasize consistency in the policies that affect research across the university.

#### 3.3.1.1 *Data Governance*

Data is an asset for the institution, and the governance for managing data that must balance two responsibilities:

1. The need to collect and secure information.
2. The responsibility to secure value from that information.

When we talk about health data, we must recognize we are stewards of that data; it can consist of patients' personal and health information as well as financial data. On the other hand, that clinical data is core to our research mission of improving the health of the populations we serve and those who benefit from the biomedical research we conduct. We also must recognize that our research mission is not limited to human data, and the rules that protect human subjects don't apply to non-humans. Our data governance approach must acknowledge and respect all the rules of the data that we manage, including requirements for data sharing, data stewardship and data management. When the rules of clinical data are applied without thought of the nature of the data itself, it can subvert our research mission. Leadership can help us navigate these issues, and policies and procedures can provide the map to researchers as they traverse these questions.

#### 3.3.1.2 *Coordinate governance through committees*

Committees are the most common locus of policy genesis, but the proliferation of committees, teams and workgroups created to address particular aspects of research infrastructure – while often necessary to administer those aspects – tends to contribute to fragmentation and work against a coherent vision and strategy across the enterprise. Without a coherent strategic approach, teams and individuals risk not only not working in harmony, but at cross purposes. We seek to help address the following challenges with the committee structures and processes

that affect research:

- committees are sometimes inconsistent with respect to adherence to their mission or charter (sometimes for good reason, as problems and opportunities are often emergent);
- committees do not follow common procedures for conduct and documentation (such as Robert's Rules of Order), making it difficult to identify when a committee made a decision; and
- committee membership turns over, and without durable documentation of committee charters, activities and changes, institutional memory is lost and effort is wasted.

These challenges yield preventable knowledge loss and wasted effort, and there are best practice solutions available. The CRIO will advocate for standards of "meta-governance" – that is, how governance conducts itself – to increase accountability, transparency, and effectiveness. This includes creating more consistency in how meetings are conducted and documented, both within teams/groups/committees and across them. CRIO staff will collaborate with bodies such as the Ohio State University Senate to identify robust platforms for meeting materials across committees to avoid reliance on oral tradition and other ephemeral or brittle systems, and to ensure that governance activities maintain accountability through transparency – such as via accessible documentation. The CRIO will work with upper leadership in IT, the College, the medical center, and the university to hold strategic planning sessions during which representatives come together to share insights and create a common vision to be instantiated by members as they represent the college and medical center.

#### *3.3.1.3 Link policy to procedure*

Many issues related to compliance have no process in place to ensure an appropriate endpoint. Policies define constraints and requirements, but without the corresponding elaboration of a procedure that explains how to conduct business consistent with a policy, the affected person is left to guess and speculate about how to do so –. This can lead to wasting valuable time and resources, slowing or even preventing potential discovery, and demoralizing researchers who rely on infrastructure services to guide and support them. Often rules or constraints are cited or employed that aren't in alignment with current policy. This is in part because the path for getting from vision to regulations is inconsistent or absent, and in part because the policy and information about it is not interpretable by researcher audiences. This problem is compounded by the need for researchers to engage with multiple services in the course of their work, encountering multiple layers of issues.

We propose to address this issue in the following ways:

1. advocating in the case of any of new policies to always couple policy development with developing a corresponding procedure second, to identify.
2. identifying existing policies that lack clear associated procedure and supporting their development either by IT or by the CRIO.
3. Working to identify issues in both in existing policy and in new policy development that make compliance or procedural execution difficult or impossible.
4. integrating "user testing" strategies to evaluate the consequences of policy for the conduct of research and the effectiveness of procedures for maintaining compliance.

#### *3.3.1.4 Research-sensitive policies and procedures*

As Ohio State seeks to position itself as a center of research for the community, region and nation, our local policies and procedures must reconcile regulatory requirements with the needs

of research in general. This is often done through ensuring that rules and regulations have means through which exceptions can be made, and that these paths are well understood. They also require that IT community understand these policies and procedures, enabling our IT staff to empower and enable research. This includes writing policies and procedures in ways that are meaningful and intelligible by research audiences. Working together, we can improve the experience of researchers at Ohio State and accelerate research, for example by developing procedures and tools for resources such as Amazon Web Services, Box.com, the Ohio Supercomputer Center, and the State of Ohio Innovate Ohio Platform.

#### *3.3.1.5 Facilitate compliance*

When policies apply, providing tools and education can make it easier for researchers to do the right thing. For instance, the need for Data Management Plans, as required by funding agencies like the National Institutes of Health, can be made easier by providing best practice templates and training for research staff and faculty. Facilitating check out procedures that make transparent the needs of data ingress and egress can make future research easier to conduct. A robust system of this nature can serve to improve data loss and compliance.

One area of particular concern surrounds data archiving. Ohio State's Institutional Data Policy, Research Data Policy, and NIH policies on research data make it clear that while Principal Investigators are stewards of the research data they generate, the university is the data owner. As such, the university is required to provide the archival storage for projects. However, there is neither an established Ohio State or Ohio State Wexner Medical Center standard, nor an infrastructure, for the archiving of research data. This problem is particularly severe for health sciences data that have stricter privacy standards and regulatory controls, and tend to be large. As such, the archival requirement represents an unresourced and ungoverned need for the university, and an unfunded mandate for investigators.

The CRIO will partner with University Libraries and the Office of Research to identify sponsor and regulatory requirements for research archival data, survey platforms and resources already deployed at or available to the university. We will collaborate with those parties to document those requirements, draft guidelines for meeting them, and draft procedures to support investigators and research teams. We will take the same approach to these resources as for policy development, submitting them for review by the research community to ensure they are both usable and useful to their audiences.

The above policy areas have some gaps as well as contain some existing policies that either do not reflect their ostensible spirit or impose unnecessary and/or avoidable burdens on research and discovery. To pilot a policy development process that includes both the research community and IT as stakeholders and coauthors, the CRIO has drafted a policy on guest access that clarifies existing policy and codifies a process that to date has largely relied on oral tradition and has been inconsistent across the College of Medicine and medical center in ways that have distributed collaboration with non-medical center partners.

#### *3.3.1.6 Develop research-sensitive support for accessing and using tools and technologies*

Conversations to date with IT show a clear vision of a future that would be both very productive and wildly popular if it could be shared. We need a more consistent, institutionalized and audience-appropriate approach to engaging faculty and staff. The CRIO will serve as a bridge and translator between IT's vision and the research community to help communicate and contextualize both how IT's vision serves and advances research and how research can most effectively embrace and leverage that vision.

### *3.3.1.7 Align IT operations to support research practice*

A number of issues have been raised related to current IT support of research operations. An illustrative example is the disconnect between the clinical workstation and the research workstation, which often includes the need for what IT has, in the past, considered non-standard equipment. Chief among this has been the use of non-Windows computers, especially those that use the MacOS operating system (hereafter referred to as Mac[s]). Macs, across the research and clinical mission. However, such issues are seen in the way that IT focuses on the clinical mission without the concomitant flexibility necessary for research software. The result can be software that becomes difficult to upgrade and images that focus on clinical security at the expense of research flexibility. That has led to support of versions of tools that is context-dependent and idiosyncratic to the relationships involved or that require following steps and procedures that are opaque to customers and are not incorrectly built into the eServices workflow.

In providing solutions to problems there are tradeoffs between addressing symptoms and addressing root causes. There is a parallel tradeoff between addressing immediate and manifest research computing needs and searching for enterprise-optimal solutions. A framework is needed to structure the process of defining problems and needs on the one hand, and targeting and implementing solutions on the other, that creates collaboration between IT and the researcher to prioritize the type and scale of solution employed to ensure that immediate needs are met as well as reconciled with broader efforts to provide solutions at scale.

There have been IT decisions made that affect end users without engaging them as partners in testing, or without notifying users that potentially consequential changes are being made. First, there are members of the research community who would gladly serve as alpha and beta test users. Second, lack of information that changes are being made, or the provenance and potential implications of those changes, deprives researchers of the ability to engage IT to discuss mitigation, prevents users from identifying the origin of changes or preparing for their consequences, and results in “wild goose chases” where users struggle to identify the source or a new problem or disruption. Creating a framework and culture of early adoption, phased deployment, and shared governance on IT changes would yield substantial dividends for both IT and research in terms of trust minimized disruption.

Many of the solutions for these problems lie in increased end-user engagement. The current model of engagement imposes too much responsibility on users for IT decisions and request/idea preparation. IT should be more thoughtful in working with end users to identify issues; finding workable solutions; and then creating architectures that balance our support of the clinical, education and research missions of the organization. Such a process can serve to clarify requirements, including identifying required processes, and streamline processes to eliminate steps that do not create value. This will spare both IT and the research community unproductive discussion, searches, administrative ritual and frustration.

### *3.3.1.8 Develop and improve research-supporting documentation*

Just as obtaining resources, navigating procedures, or making sense of policies too often depends upon personal connections to the right individuals rather than to a robust infrastructure that creates equal opportunity to be productive, understanding how to access, use and further develop research IT resources is too often a function of access to specific individuals or on an oral tradition that is error-prone and can lead to “technical rituals” that fail to align with the most efficient way to get work done. In addition to being inappropriate for ensuring coherence, this

configuration makes IT and research vulnerable by risking the loss of essential knowledge if personnel depart, take substantial leave, or change roles.

### 3.3.2 Enhance engagement with the research community

Research infrastructure must be accessible to and usable by the research community; hidden resources are no better than nonexistent ones, and resources that require specialized technical knowledge to use are hardly better than none. All resources must be accompanied by comprehensive, clear, usable and researcher-focused documentation and user education that is kept current and delivered as much as possible on researchers' own terms. We will complement our infrastructure with appropriate documentation and training, as well as provide educational resources that bridge the gap between traditional IT communications and researchers needs, styles and professional rhythms.

#### 3.3.2.1 *Improve researcher communication*

Two existing platforms hold promise for structuring the development, dissemination and review of research infrastructure policies. First, there is a need to more actively engage researchers with messages tailored to the community. Messaging tends to be ad hoc, ephemeral and lacks ongoing engagement with the research community. Communications infrastructure for engaging user communities about the status, trajectory, strategy and vision for IT activities is insufficient or absent. To address these concerns, the CRIO has partnered with the medical center to hire a professional educator focused on research messaging and technology. As a corollary to Policies and Procedures and Communications, end-user training for tools is often lacking. Similar to the difficulties in gaining declarative knowledge (the "what") from research data and IT services and supports, there is a challenge to gaining procedural knowledge (the "how"). Employing a professional training developer to build, deliver and maintain training materials can allow the college to make its research infrastructure available to more people and with training that is aligned to pedagogic best practices.

Second, there is a need to bring these messages of change together under one common umbrella and to ensure research service providers, including IT, embrace a standard "one-stop shop" communication infrastructure for providing notifications. One model for this approach is the university's policy platform and policy development process, which are established and familiar tools for developing, maintaining and disseminating policy. These tools adopt and embrace a standard protocol for communicating research service activities that incorporates a proposal phase with review by user communities, eliciting feedback on proposals with the goal of providing improved policies and procedures.

The CRIO has obtained agreement from the custodians of the university's policy platform to share the platform's code, and is in the process of securing hosting. This approach will allow for RIT to both adopt and embrace a standard protocol for notifying users about changes comprising, at minimum, the following: (a) what change is coming, when, and what users can expect; (b) the change was performed and went as expected; (c) the change was performed and something unexpected happened that will affect users; (d) the change could not be performed and the new expectation is .

### 3.3.3 Partnerships

One purpose of the CRIO is to build and maintain relationships within the Health Science campus and across the university. Stakeholders of note include those within the university, such as Ohio State's Comprehensive Cancer Center – James Cancer Hospital and Solove Research Institute, the Translational Data Analytics Institute, and the Center for Clinical and Translational Science. These organizations have cross-cutting mandates that include biomedical science, and

to the extent that such efforts include Research IT, there is a need for a strong partner in advancing the collaborative research capacity of our medical mission.

However, it should be noted that we have a number of external collaborators that would benefit from a leader-led engagement; among them are Nationwide Children's Hospital, the Ohio Supercomputer Center and regional health care providers. Ohio State's role in the community combined with the need to provide research leadership (for example around data sharing), creates an opportunity to the biomedical sciences to provide leadership consistent with the University's broader mission.

More broadly, to foster engagement external to the university with the goal of leveraging that relationship to advance research, the CRIO will work with other universities to establish an Industry-University Cooperative Research Center (IUCRC). The IUCRC Program was started in 1973 by the National Science Foundation with the goal of developing long-term partnerships among industry, academia and government. The National Science Foundation (NSF) invests in these partnerships to promote research programs of mutual interest, contribute to the nation's research infrastructure base, enhance the intellectual capacity of the engineering or science workforce through the integration of research and education, and facilitate technology transfer.

To meet national needs, multi-university IUCRCs are preferred to single-university IUCRCs because multi-university centers contribute to an increased research base as well as to increased interaction among center participants. The centers are catalyzed by an investment from NSF with primary support derived from the private and public sector. The NSF takes a supporting role in the development and evolution of the IUCRC, providing a framework for membership and operations as well as requirements derived from extensive center experience and evaluation.

The NSF has agreed to receive and application from a national collaboration based on existing engagements and relationships entitled The Center for Population Health Analytics & Technology (CPHAT) which seeks to develop and advance technologies, knowledge and analytic methods that improve population health and wellbeing. The proposed university partners – Indiana University, the University of California San Francisco, Ohio State, the University of Florida, and the Medical University of South Carolina – do not have any formalized partnerships; however, the lead investigators and key supporting faculty have longstanding (more than a decade) and ongoing research projects that include collaborations on multiple papers and grants, and shared work on multiple technical advisory panels. In particular, the goal of CPHAT is to focus on:

- Social determinants of health, which include the conditions in which people live, work and age, along with individual behaviors. Collectively, these factors are estimated to drive 60% of health care utilization and are composed of sociodemographic factors such as poverty, transportation barriers, food insecurity and physical inactivity. Examples of proposed research include aggregate and individual level measurement of social determinants and informatics tools to typify our understandings of communities and enhance individual engagement in their health as well as connect individuals and organizations to address social needs.
- Population health analytics, which transforms health information technology, tools and analytic methodologies from the traditional individual patient setting to identifying and making decisions about risks, needs and management of populations. For example, current risk models predict outcomes such as death, care costs or utilization and are focused on care decisions made by clinicians. While important, these outcomes are

distal from the basic goals of population health. Instead, prediction of needed services or the risk because of upstream factors are more relevant to population health goals. To support population health, new information sources and methods of presenting risks must be put in front of health system leaders and shared with partner organizations to foster action.

- Addiction, which is widespread and encompasses opioids, tobacco, alcohol and illicit substances. Tobacco use is one of the leading causes of preventable illness and death. Notably, opioid addiction complicates care delivery and treatment adherence. The addiction epidemic is often driven by social factors and policies outside of health care organizations' direct influence. As such, effective interventions require the combination of unique data, differing guidelines and novel interventions such as decision support tools that engage patients in self-management as well as link them with addiction to certified counselors or leveraging consumer shopping and pharmacy data to assess and monitor risk.

### 3.4 Restructure Research Information Technology Infrastructure

Accomplishing these ambitious goals is not possible with current staffing. Prior to September 2019, research information infrastructure was managed by the Department of Biomedical Informatics. Subsequently, the university selected its new Chief Research Information Officer and charged him with restructuring Research IT to more effectively facilitate discovery across the biomedical sciences. As a consequence, staff were moved into a new organizational unit – College of Medicine Research IT – which is headed by the Chief Research Information Officer. In January, 2020, Research IT was organized to facilitate the execution of this strategic plan.

As the new organization chart shows, Research IT has broad service lines: Research Informatics Data Services, Research Informatics Software Engineering, Data Science and Engineering, and Research Infrastructure Development spanning Research User Experience and Design and Research Infrastructure Development and Engineering. These service lines dovetail with each other, a study or project is likely to draw on multiple of them, and the success of the broader efforts described in this strategic plan is dependent upon their coordinated contributions.

Research Informatics Data Services involves the use of technologies, structures and processes to leverage information, created through both research and practice, in the discovery and management of new knowledge relating to health and disease. Individual investigators or project teams often lack the funds or personnel to provide expert support, and resources such as datasets are often more efficiently procured and managed via a central service. We characterize research informatics as spanning two core areas: support for primary data collection (Electronic Data Collection Core, or EDCCore), and support for the analysis of secondary data (DataCore), with each core embodying specialized skills to deliver high quality research resources.

#### 3.4.1 Research Informatics Data Services: Electronic Data Capture Core

The EDCCore supports and provides consultation on a repertoire of electronic tools to help ensure robust, valid and reliable data collection for both clinical and nonclinical studies. These tools include REDCap, Qualtrics and EHR-to-EDC, and they are used to support project data collections ranging from simple surveys to longitudinal data capture, quality improvement and workflow tracking.

The use of REDCap and other EDC tools has grown rapidly across the Ohio State Wexner Medical Center and has spread to projects in other colleges. The majority of support comes from the CCTS, but current support resources are insufficient and are lower than for similar-

sized colleges. and is generally relegated to treatment as infrastructure. We have 1.5 FTE in support of REDCap at Ohio State. Our peer institutions invest **between three and six FTE** in EDC support. The resultant impact of this has been significant:

- Our relative lack of support results in a greater reliance on self-serve approaches where the clinical faculty are required to manage the process to support their research, and that has resulted in studies losing data.
- We currently don't allow plug-ins to REDCap because we don't have sufficient technical support. As a result, we have to work with the university to stand up versions of REDCap to support research. Further, we do not have the IT infrastructure necessary to support forked versions of REDCap. For example, the HEALing Communities Study is a multi- institutional project requiring a different (newer) version of REDCap than the medical center's production version, but it also requires a static version of the software.
- Our version of REDCap is not REDCap National Institute of Standards and Technology/Federal Information Security Management Act environment certification compliant.
- Our deployment of REDCap Sync is not as robust as it should be.
- The medical center REDCap account provisioning process is cumbersome and inefficient.
- The availability of multiple EDC tools creates the need for triage/concierge services to identify the appropriate tool for a given need. For example, Qualtrics has no current shared support, and expertise is concentrated in departments/centers.
- We currently update REDCap semiannually. We should be updating at least quarterly.

The new structure of the EDCCore will fall under the leadership of an EDC Leader who will coordinate EDC services. This approach will allow for broader strategic changes to support researchers, including:

- Containerizing REDCap which would facilitate supporting multiple versions and multiple configurations.
- Developing tools to support self-service such as a "random walk" form-filling bot that could be used to test REDCap instruments.
- The support of an EDC users' group to provide a forum through which service bottlenecks could be resolved, in part by improving training for "powerusers."
- A defined certification process for REDCap Admins across the campus that could involve initial training combined with periodic recertification.

### 3.4.2 Research Informatics Data Services: DataCore

The DataCore team will lead the development of the SDC, provide convenient access to secondary data from the medical center as well as other sources such as national databases, and provide the metadata and data management services needed to ensure data are accessible and usable for discovery. This core serves as a centralized resource for data extraction, data linkage, metadata management, and harmonization and standardization to common data models. The DataCore also will manage the licensing of and access to datasets acquired for use across the college. Additionally, the DataCore supports cohort identification technologies such as i2b2 and SHRINE as well as research support tools like Scarlet for All; it also will likely support implementations of Epic Cosmos.

### 3.4.3 Research Informatics Software Engineering (RISE)

Research includes human interaction and engagement activities that require the creation of new

or customized software tools. This in turn requires specialized expertise that is generally beyond the scope of any single project, and individual project teams risk duplicating work across the institution and forfeiting economies of scale. Further, software tools require support from “cradle to grave,” including deployment, ongoing support and maintenance. Research IT needs to support such effort through a cadre of software developers who will help researchers identify requirements and specifications; develop software tools; facilitate deployment, spread, and maintenance; and work with vendors on behalf of the research community. By providing a central, “at large” development resource, we can leverage prior work, provide pre-award feasibility and needs analyses, and exploit opportunities outside the context of an individual project.

However, maintaining this resource in house comes with an associated risk. Specifically, software development needs, like research itself, occur in bursts. As a result, developers were tasked with identifying products that could benefit from their efforts. Moving forward, unfunded time will be earmarked for the development of tools to support the research community for the common good. A number of the projects listed here will have build requirements that will leverage moments when there is excess capacity. For example, there are a number of tools that are currently not supported by IT and rely on researchers to manage. An example of this is the Genome Browser. Current implementations at Ohio State were made possible by an assistant professor who was told by IT that supporting applications was not its responsibility. To prevent wasted effort and missed opportunities, a process is needed to ensure ongoing support of tools that were developed in the context of a study but continue provide value to the community.

#### 3.4.4 Data Science and Engineering

Data Science in the context of Research IT comprises four key services: the creation of secure high performance computing resources for conducting sophisticated data analyses on big data requiring appropriate protection; data engineering to construct analysis-ready datasets; data visualization and dashboarding; and the development and dissemination of computing practices for secure reproducible research computing. This includes establishing research computing environments that – through the use of scalable hardware platforms, virtualization and containerization, and improved governance – simultaneously enhance the performance, regulatory compliance, usability and reproducibility of research computing. We also are developing programs to increase data availability and accessibility (e.g. interfaces such as SCARLET), as well as collect new forms of data such as patient-reported outcomes. Data engineering includes standardizing and harmonizing datasets for the Secondary DataCore. Visualization and reporting tools include the development required to make tools such as SQUIRE effective, and to create the data architectures that complement systems such as PROMPT. Development of reproducible analytic workflow templates and supporting datasets from onboarding through close-out will increase the efficiency and effectiveness of individual projects as well as support both the College’s and IT’s more holistic views for needs assessment and data governance. Our investments in infrastructure and governance will also help position Ohio State as a stronger leader in multisite research, such in serving as a Data Coordinating Center. These include maintenance of RShiny and Tableau servers as well as engaging the research community on the use of data analytics platforms. We expect to build partnerships with the new Chief Analytic Officer Dr. Susan White as well as the Department of Biomedical Informatics to increase the availability of resources to faculty, staff and students in support of research.

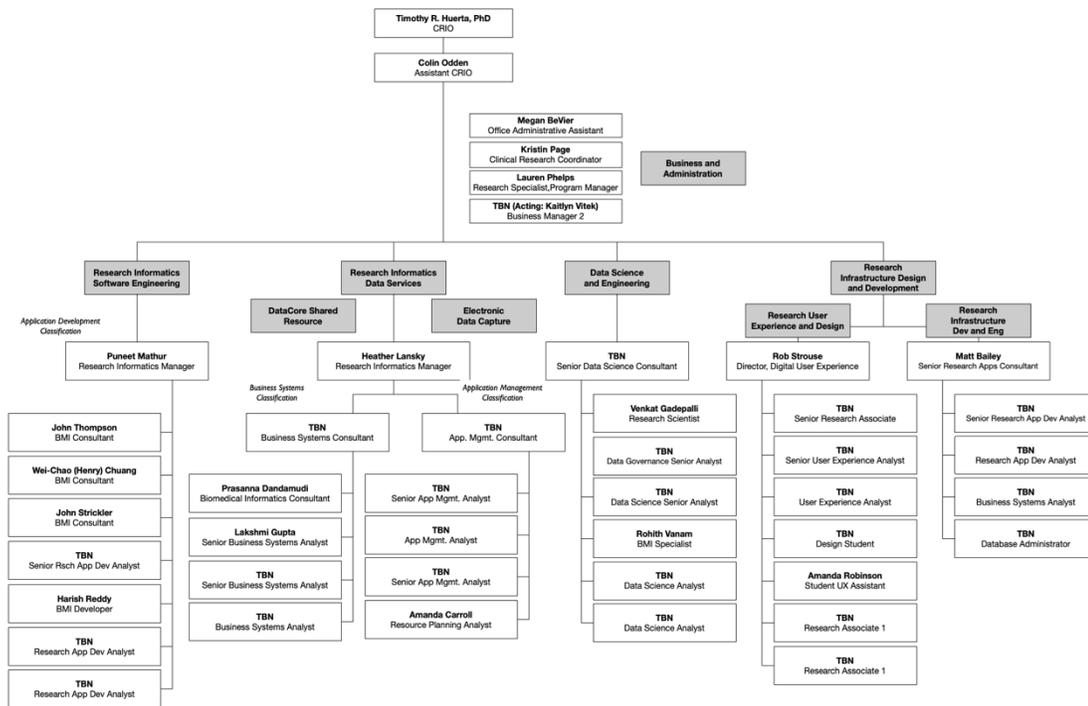
#### 3.4.5 Research Infrastructure Design and Development

Initiatives such as PARTNER, PROMPT and SQUIRE belong to a class of general challenges in research infrastructure: the problems are pervasive, general solutions may be identifiable but

cannot be achieved in the context of a particular study or research program, and general solutions cannot fit within an individual study's aims. Bespoke tools of the kind developed by the RISE team are tailored for specific studies or applications, and as such are complementary to building broader research infrastructure but do not replace it. A different approach is required to produce frameworks that are both sufficiently general, yet flexible and customizable, to be useful to the research community at large. The Research User Experience and Design and Research Infrastructure Development and Engineering groups

Again, as NIH begins to require that clinical projects include FHIR as a part of the study to foster implementation, we will be required to have these resources available for the research community. As a result, Ohio State will need to maintain these core capacities and will need to grow them as we move forward.

### 3.4.6 Revised Organization Chart



This organization chart outlines the proposed revised reporting structure for Research Information Technology. As illustrated, the new Department is organized around four major areas and this document includes employees current as of 4/1/2020, hires planned both to meet current demand to the needs consequent to the strategic directions explained in this plan, and others awaiting transactions.

### 3.4.7 Budget

| CRIO Budget  | COM                   |                     |                       | TOTAL                 |
|--|-----------------------|---------------------|-----------------------|-----------------------|
|  | Effort                | Earnings            | General Funds         |                       |
| Administration Total                               | \$27,384.69           | \$0.00              | \$534,198.67          | <b>\$561,583.37</b>   |
| Wexner Medical Center Information Technology Total | \$0.00                | \$0.00              | \$448,841.62          | <b>\$448,841.62</b>   |
| Research Dev/Ops Total                             | \$476,644.18          | \$233,644.25        | \$213,920.89          | <b>\$924,209.32</b>   |
| Research Informatics Total                         | \$290,686.46          | \$287,004.12        | \$263,441.78          | <b>\$841,132.36</b>   |
| DataCore Total                                     | \$109,516.74          | \$182,527.91        | \$73,011.16           | <b>\$365,055.82</b>   |
| EDC Total  | \$104,476.21          | \$104,476.21        | \$139,301.62          | <b>\$348,254.04</b>   |
| Data Science Total                                 | \$322,899.74          | \$82,068.75         | \$133,197.75          | <b>\$538,166.23</b>   |
| Research Infrastructure Total                      | \$659,399.30          | \$163,164.57        | \$276,434.97          | <b>\$1,098,998.83</b> |
| User Experience Total                              | \$343,186.10          | \$86,602.17         | \$142,188.57          | <b>\$571,976.83</b>   |
| Infrastructure Development Total                   | \$316,213.20          | \$76,562.40         | \$134,246.40          | <b>\$527,022.00</b>   |
|  | <b>\$1,777,014.36</b> | <b>\$765,881.69</b> | <b>\$1,870,035.68</b> | <b>\$4,412,931.73</b> |

#### 3.4.7.1 Changes in departmental operations in FY20-21

- Dr. Timothy R. Huerta began his tenure as CRIO on September 1, 2019
- The Department of Research Information Technology was established on November 1, 2019
- The Department requested to establish an earnings operations to charge for Research Informatics, Software Development and Operations, Data Science, and Infrastructure. We estimate that FY21 to be approximately \$1.4M.
- Secured lease agreement with off-site space at 530 Spring Street. Lease agreement begins April 1, 2020 and is in effect until April 30, 2025. Monthly rental amount is \$18,004 (annual \$216,048) with utilities billed based on utilization. These costs are not currently part of the CRIO budget as they are covered by the College of Medicine.
- Purchased data sets during FY20 to accommodate needs of existing and recently recruited CoM faculty. (multiple PR's / \$19,285)
- Purchase order generated to DuetHealth, Inc. to build the Patient Reported Outcome Measurement Prioritization Tool (PROMPT). This will enable patients and their families to report patient outcomes regarding their health experience outside of a clinical setting. (PR5508787 / \$660,000).
- Computers and laptops were purchased during FY20 to accommodate new employees as well as refresh existing units. (multiple PR's / (\$34,569)

#### 3.4.7.2 Faculty/Staff

- Staffing changes included in the budget (new hires, retirements and/or promotions)
- Staff transitioned from the Department of Biomedical Informatics (BMI) effective November 1, 2019.
- Hired Colin Odden as Assistant CRIO in early second half of FY20. (\$115,000 + benefits)
- Two(2) general admin/research positions are posted; Office Admin Assistant and Sr. Research Associate. Both will support the general CRIO operation. (\$121,440 + benefits)
- One(1) Research Application Developer Analyst, within the Research Development/Operation Core, is posted. (\$89k + benefits)
- One (1) Data Governance Sr. Analyst, within the Data Science Core, is posted. (\$97,500 + benefits)
- Four (4) positions are posted within the Research Infrastructure Core. Two of the positions are Research Associate 1's, one is a Sr. Research Associate, and the other is a Sr. User Experience Analyst. (\$266,250 + benefits)
- Nineteen (19) other positions are needed based on this Strategic Plan

#### 4 Conclusion

At the beginning of this document, we note the following guiding principle:

*Working with biomedical researchers across the institution  
to change the world from The Ohio State University*

This framing is one that was brought to frame the CRIO's original hiring. It was held in contrast to the idea of simply improving the experiences of Central Ohioans. The potential for the University to act as an agency of change is the central defining frame for what must be advanced in the next two years that would and could be felt by faculty day-to-day. **Guided by this Strategic Plan, the Chief Information Research Officer and the Department of Research Information Technology within the Office of Research in the College of Medicine are committed to building the infrastructure that will improve the likelihood of research funding.** As we move forward working with faculty, administrators, leaders and staff, our goal is to establish a robust infrastructure on which discovery can be built.