### Introduction

The modern biomedical research paradigm recognizes that data is a valuable resource that can be reanalyzed in secondary contexts to allow for synthesis into new knowledge.

Researchers at the University of Michigan Medical School utilize data from over 300 enterprise repositories that often need to be integrated with data from hundreds of public repositories and a growing wealth of omics scale data.

Currently researchers are required to integrate these resources on a point-to-point basis within the scope of a single project.

Although centralized data warehousing can facilitate data access and interoperability, the modern economy of information requires a federated data liquidity framework that can scale with the exponential growth of big data and evolve with changing analytical platforms.

#### Solution

We have designed the MedBus Service-Oriented Architecture to increase the efficiency of deploying high-value research data web services while reducing the institutional risk associated with access to protected health information.

MedBus implements a FAIR data delivery strategy in which data are:

Findable by exposing web service functionality in Swagger and defining query parameters in our enterprise information management glossary.

Accessible by following a provisioning model that is protected by an integrated institutional directory that is transitioning to an OAuth implementation.

Interoperable by allowing data to be consumed and utilized across multiple platforms including Excel, R, SAS, Oracle, Hadoop and others.

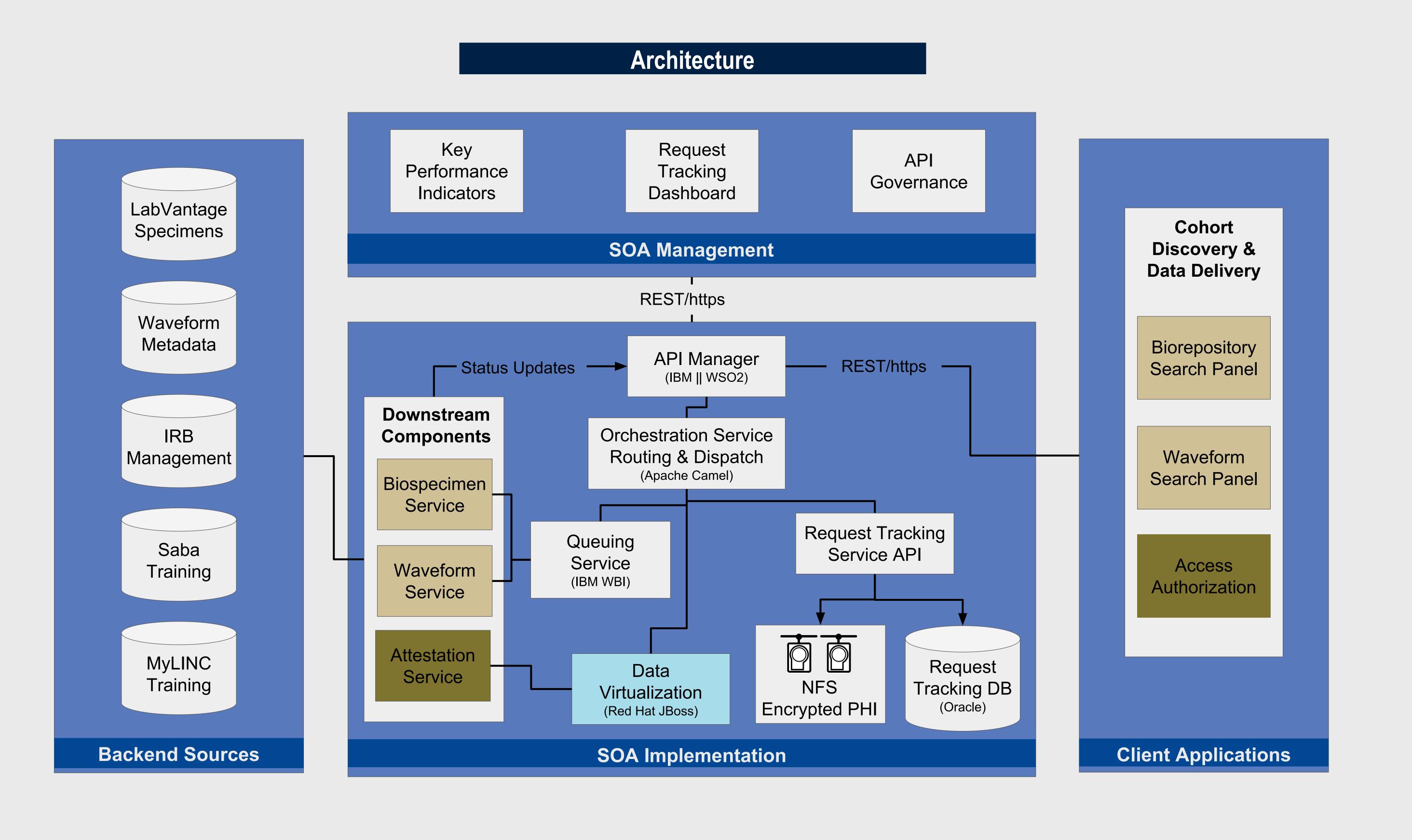
Reusable by providing basic units of function that can be uniformly accessed as services that are usable within multiple application and analysis platforms.

# MedBus Applies arises to describe the reservoir of the state of the s

# A service-oriented architecture for enabling the research data economy

Amitava Shee, BE<sup>1</sup>, James Estill, PhD<sup>1</sup>, John Brussolo, MBA<sup>1</sup>

1. University of Michigan, Medical School Information Services, Ann Arbor, MI



#### **Key Capabilities**

Industrializes data integration in a cheap and fast manner that leverages open source components.

Extensible data sourcing from small scale Excel documents to big data platforms.

FedEx request & delivery model supports long running queries lasting multiple days.

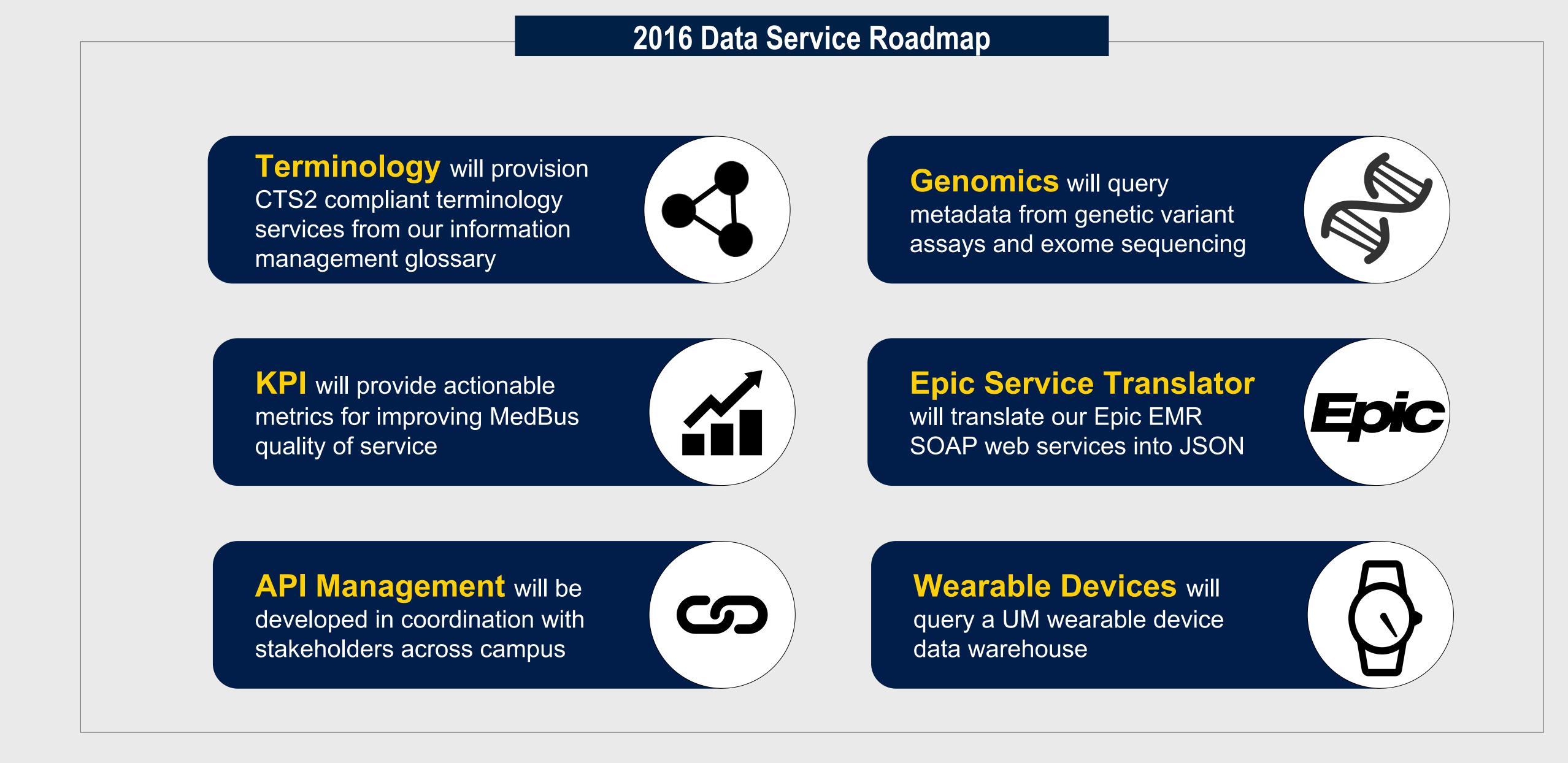
High availability at low cost using cheap virtual machine clusters with software managed failover.

HIPAA compliant with a centralized visibility and control of ePHI access across downstream queries.

Researcher friendly allows data consumption via R, SAS, Tableau, Excel, services or any ODBC/JDBC

Aligned to NIH open data strategy and the community cloud model.

#### Data Services in Use Training queries 2 web-based Waveform queries a 5GB training systems for 38,000 staff metadata store of over 6TB of members enrolled in 881,000 archived waveform recordings Biospecimen queries a central LIMS warehouse that indexes 164,000 specimens from Payload secures compressed ttestation unifies disparate query request and response training services into a single payloads by encryption with a transcript reporting service



## Summary of Conclusions

MedBus is a platform for enabling the new research data economy in which existing data assets can produce new knowledge that translates into tangible research value.

MedBus unblocks the technical, regulatory, and organizational bottlenecks that limit research data liquidity, allowing IT to focus on the value-added components of service delivery and researchers to focus on testing hypotheses.

Since a March 31 2015 soft launch, MedBus has responded to >300 requests from a cohort discovery application that integrates MedBus services with EMR data.

Red Hat JBoss Data Virtualization has improved MedBus by allowing us to deliver secure federated data services in weeks instead of the months required for custom coding efforts.

# 2016 Roadmap Highlights

- Extend data service delivery to more clients in addition to the current use for cohort discovery
- Collaborate in the development of a university-wide central API management and governance framework
- Deploy a SOA test harness that will accelerate regression testing and facilitate change management
- Extend open source collaboration to other units within the University of Michigan
- Extend the biospecimen service to include specimen data that are stored in Freezerworks

Send feedback and ideas to: MSIS-Research-Program@med.umich.edu

More Information at: <a href="http://medbus-umich.github.io">http://medbus-umich.github.io</a>



