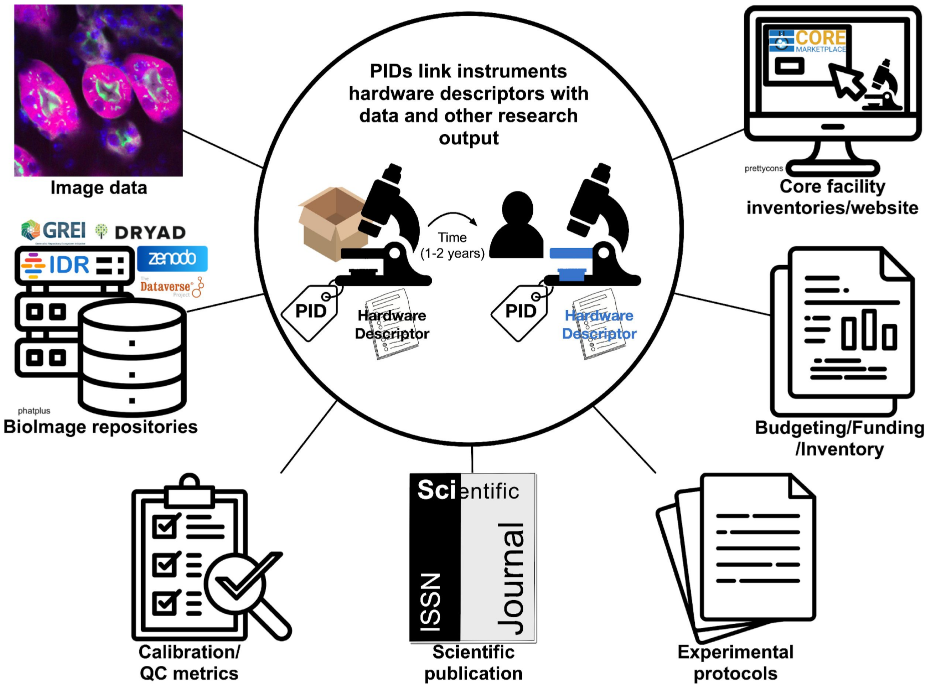
NSF Cyberinfrastructure for Sustained Scientific Innovation (CSSI)

Imaging-PHD: Empowering data reuse and reproducibility through microscopy-community-defined Persistent Hardware Descriptors

PIDs form a cornerstone of research



Metadata is Scattered

- Critical imaging metadata is stored across disparate locations (files, devices, documents).
- This fragmentation hinders reproducibility, reuse, and proper attribution.

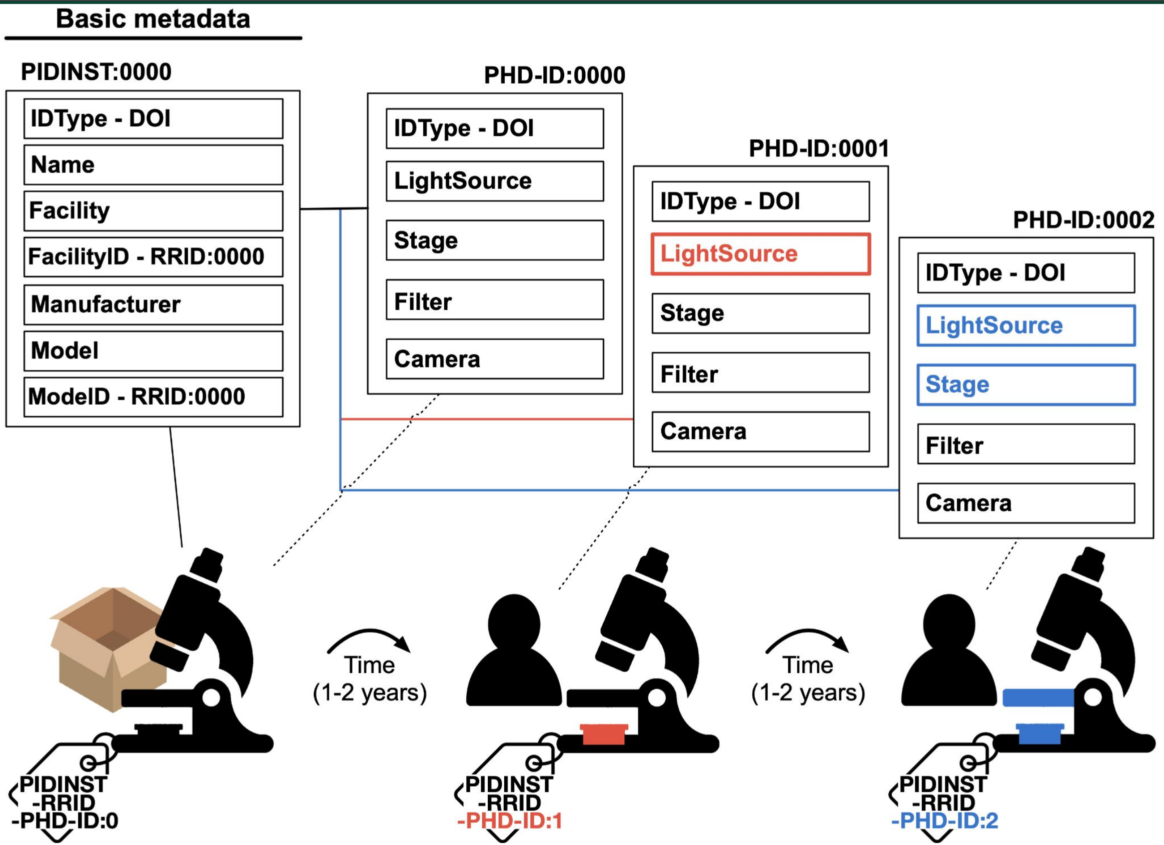
What Are Persistent Identifiers (PIDs)?

- Unique, long-lasting digital codes assigned to entities like instruments, datasets, and people.
- Maintained by trusted authorities and resolve to stable landing pages with rich metadata.

Why PIDs Are Essential

- **Linking**: Connect hardware, datasets, publications, and contributors.
- **Preservation**: Maintain long-term accessibility and integrity of research assets.

Solution - Persistent Hardware Descriptors enable reuse and reproducibility



Design Goals

These foundational goals guided the structure and priorities of the Imaging-PHD platform.

1. **Creation of Persistent Hardware Descriptors (PHDs)**: Fault-tolerant, distributed, and low-cost metadata records that provide citable, persistent descriptions of complex and evolving instrument hardware configurations. **This fosters reproducibility and reusability of data.**

2. **Automation and Standardization**: Development of vendor-friendly and community-driven frameworks to automate metadata capture directly from instrument manufacturers. **This minimizes manual entry and errors.**

3. **Community Engagement and Outreach**: Educating and empowering researchers, technical staff, and manufacturers to adopt and contribute to standardized hardware metadata practices. **This raises awareness and ability in the workforce and community.**

4. **Extensibility and Interoperability**: Designing solutions that are extensible beyond microscopy to other scientific domains, supporting FAIR (Findable, Accessible, Interoperable, Reusable) data principles. **This ensures that PHDs can be used for different instrument types, disciplines and domains**

Methods and Approaches - Building and Scaling the Imaging-PHD Ecosystem

🏛️ **Institutional Deployment**	🌐 **Federated Infrastructure**	🤝 **Community Integration**
Flagship installation at UMass Chan Medical School, featuring durable PHD storage, searchable metadata dashboard (Elasticsearch), and smart visualization tools.	Future deployments will connect via a federated network to support distributed, collaborative metadata management across institutions.	Strategic partnerships with BioImaging North America, QUAREP-LiMi, ABRF, OME to define standards, align incentives, and ensure adoption.

Methods and Approaches - Supporting Scientific Rigor, Reproducibility and Reuse across domains

Challenges	Imaging-PHD Approach
Limited reproducibility due to inconsistent metadata	Captures detailed, machine-actionable hardware descriptors persistently linked to datasets
Manual metadata entry is error-prone and burdensome	Automates metadata capture from instruments, minimizing human input
Metadata standards don't translate across domains	Designed to be extensible beyond microscopy — supports FAIR principles across sciences
Technical metadata is often disconnected from publications	Links samples, datasets, and hardware metadata to enrich scientific outputs and credit technical staff

Deliverables

Next-Generation Metadata (NGM) Framework	A vendor-friendly metadata schema and protocol enabling automatic transfer of detailed hardware configuration data from microscopy instrument manufacturers to researchers.	NGM ensures metadata consistency, machine-aåctionability, adaptability and extensibility to different instrument types, disciplines and domains
Micro-Meta Platform Frontend (MMPF)	A user-centric Graphical User Interface (GUI) designed for authoring, reviewing, and editing NGM-encoded hardware descriptors.	The interface will promote community adoption and metadata accuracy.
Micro-Meta Platform Backend (MMPB)	A scalable service infrastructure that manages PIDs creation, metadata deposition, and interlinking of hardware descriptors with data and publications and post publication exploration, citation and data annotation. MMPB will incorporate exploratory Large Language Model (LLM) techniques to extract metadata from unstructured instrument technical documents, enhancing metadata completeness and quality.	The backend platform facilitates the management of the entire life-cycle of PHDs, from metadata capture, PID registration, publication and exploration
Coordinated Testing, Validation, Outreach and Dissemination	Actively engaging research scientists, core-facility staff, and data librarians to obtain early feedback during software development to ensure usability; implementing targeted education and training on FAIR practices, persistent identifiers (PIDs), and metadata registries at venues and organizations specifically selected to maximize impact; and creating technical documentation for microscopy hardware to support future community learning.	A coordinated strategy to promote adoption, training and education will be key to the broader impact of Imaging-PHD.

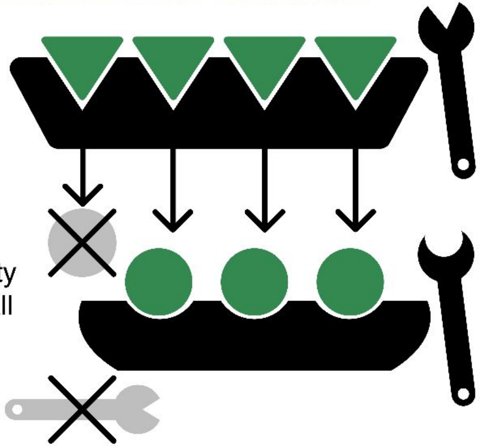
Deliverable - Composable Metadata as a Framework for Extensibility, Adaptability and Interoperability

Metadata components are modular, reusable, and designed to fit together across platforms and tools.

Current Situation: non-reusable schemas

A. Custom (e.g. vendor) schemas optimize for specific metadata, tools, and containers, but aren't generally accessible by community tools.

B. Translation into community schemas cannot represent all metadata, and vendor tools can no longer be used.



Objective 1: Next-generation metadata

C. With a composable framework, metadata need not be translated, but can be produced in a common container by users and instruments.

D. Metadata remains accessible by all schema-aware tools, and unknown terms are not lost.

E. Schema types

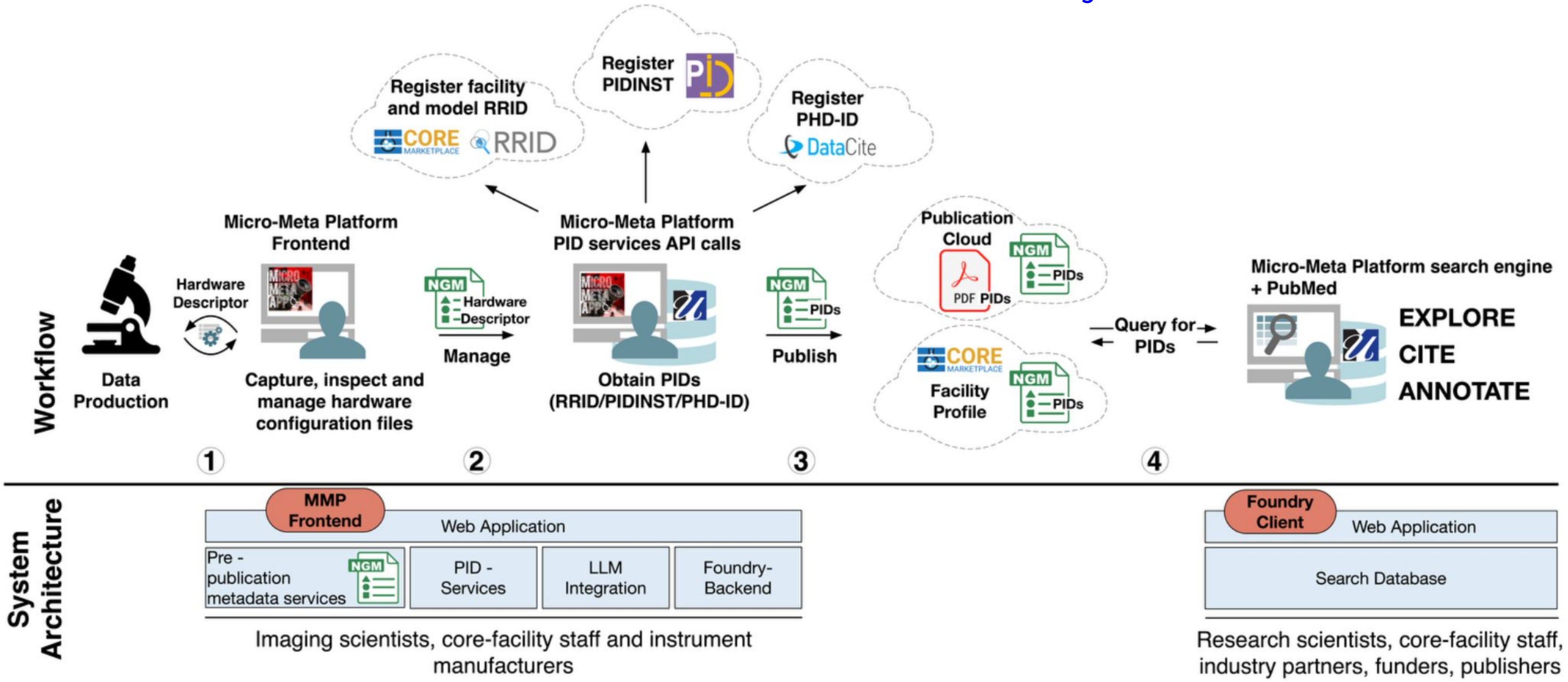
Custom schemas capture terms that have not yet been agreed upon. In the future they can extend community (circle) and common (square) schemas.

Community schemas like OME, NBO-Q, and REMBI seek consensus on terms, provide reusable base classes, and enable open-source tools.

Core schemas like DataCite provide the most general building blocks and even more re-use.

Deliverable - Micro-Meta Platform Workflow and Architecture

Capturing, registering, and reusing persistent hardware metadata across the research lifecycle.



Intellectual Merits - How this Work Advances Science

Instrument → Metadata → Persistent ID → Publication → Reuse

Merit Area	Imaging-PHD Contribution
Democratizing Scientific Infrastructure	Makes detailed, standardized instrument metadata openly available via persistent identifiers. Enables smaller labs and under-resourced institutions to build on others' hardware setups.
Ensuring Long-Term Research Quality	PHDs improve instrument lifecycle tracking and quality assurance, supporting replicability and reducing hidden sources of error.
Enabling Workforce Development	Empowers core facility staff and early-career researchers with tools and training for metadata best practices.
Catalyzing New Discovery	Harmonized metadata enables data pooling and cross-study reanalysis, accelerating multi-institutional discoveries.

“Our technical design promotes long-term reuse, community adoption, and the scalability of reproducible research.”

Broader Impacts - Empowering Researchers, Core Facilities, and Industry Partners

Who Benefits	**How Imaging-PHD Helps**
Researchers	Gain access to well-documented instrument configurations, improving reproducibility and enabling meta-analyses
Educators & Trainees	Use PHD-linked metadata for real-world training on instrumentation and FAIR practices
Core Facility Staff	Receive credit for their contributions via PID linkage and gain tools for quality tracking
Under-resourced Institutions	Can reuse metadata and design experiments modeled after advanced facilities
Software Developers & Industry Partners	Integrate their tools with standardized APIs, enhancing interoperability and adoption

“Broader impact is embedded in every layer of our platform—tools, people, and community.”