

# Metrics 2023

## **TOWARDS DEVELOPING A MODEL FOR ASSESSING THE PROGRESS, SUCCESS, AND IMPACTS OF CYBERINFRASTRUCTURE (CI) PROJECTS**

- Ritu Arora, Wayne State University
- Sukrit Sondhi, Macmillan Learning

# CI Project Metrics: Challenges and Opportunities

## Current State

- ▶ Billions in CI investment by federal funding agencies
- ▶ Several frameworks have evolved - DORA, REF, Snowball, STAR, etc. for measuring research impact of investments but no standard framework for CI project assessment
- ▶ For measuring the impact of the projects, it is important to define well-thought, project-specific metrics and monitor how the metrics evolve during the projects' lifecycles

## Challenges

- ▶ Diverse, interdisciplinary, fast-evolving nature of projects and their project-specific metrics make it challenging to identify patterns and trends for assessing the overall impact of the assessments at the funding program level
- ▶ No common semantics or yardstick
- ▶ Lack of well-defined processes for metrics capture and trustworthiness
- ▶ Calibrating a metrics model and producing benchmarks can be time-consuming for individual teams because

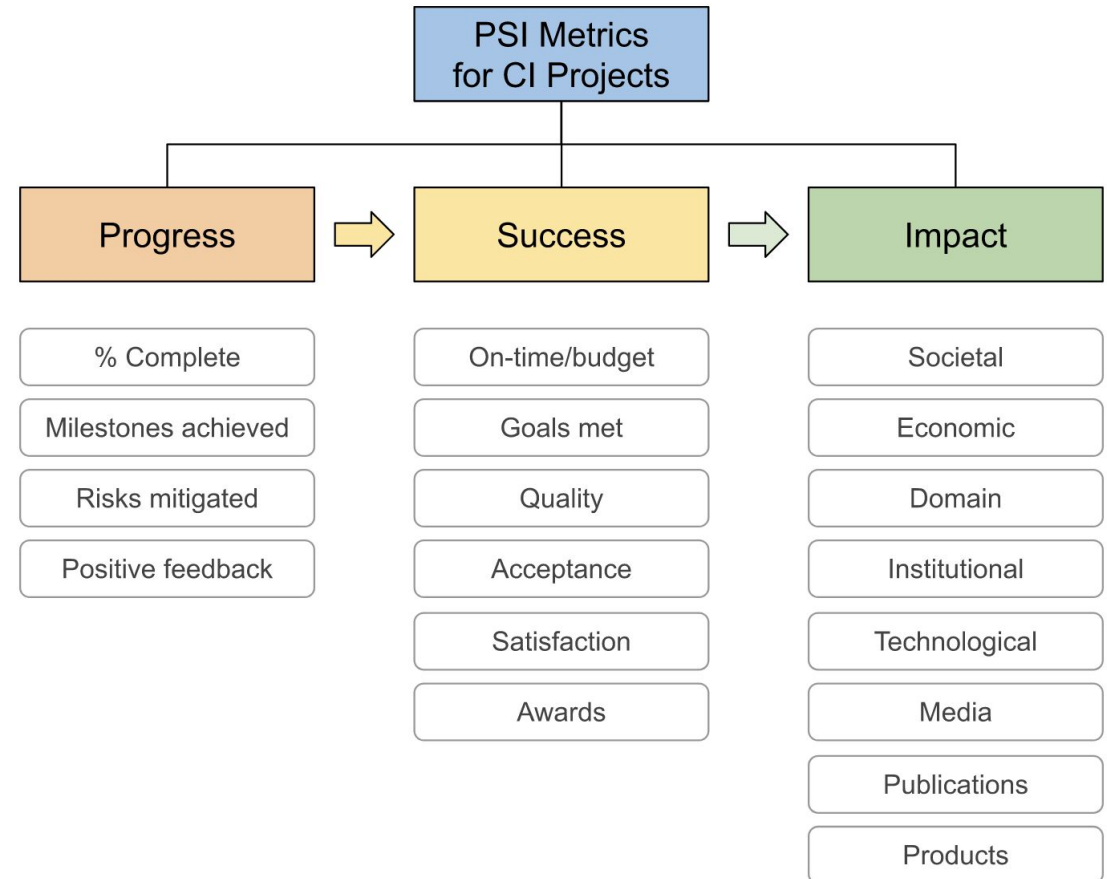
## Opportunities

- ▶ Defining and adopting a standardized set of metrics and processes in a community-driven manner while being sensitive to institution sizes, project types, and disciplines
- ▶ Leveraging advances in CI provisioning, management & instrumentation
- ▶ Analyzing large datasets of metrics
- ▶ Leveraging automation and intelligence
- ▶ Better managing and rewarding performance - of institutions, projects and individuals

# CI Project Metrics: Defining a General Set of Metrics

Three top-level metrics, intended to be broadly applicable to CI projects and programs, based on the project *lifecycle*:

- ▶ Progress: tasks or activities for achieving the project goals
- ▶ Success: The required output for making the projects successful - typically completion metrics
- ▶ Impact: What difference did the project make in different ways - post-completion metrics

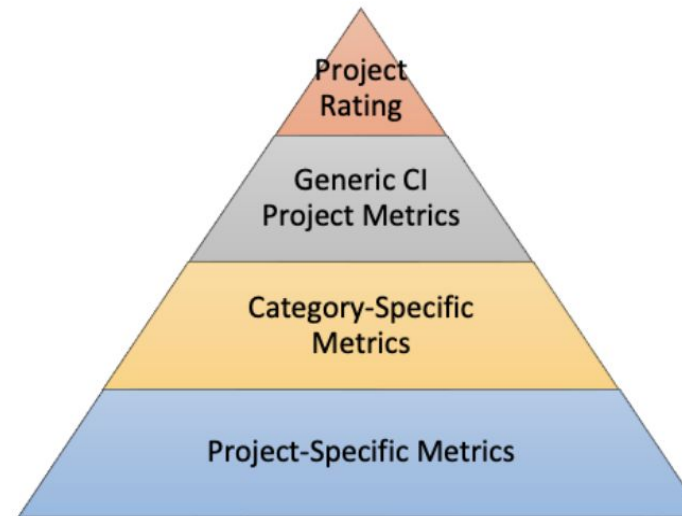


# CI Project Metrics: Classification of Projects

1. Hardware or Instruments
  - 1.1. Computing
    - 1.1.1. Cloud Computing
    - 1.1.2. Distributed Computing (Web Servers)
    - 1.1.3. Data-Intensive Computing
    - 1.1.4. High Performance Computing (HPC)
    - 1.1.5. IoT and Edge Computing
    - 1.1.6. Quantum Computing
    - 1.1.7. Volunteer Computing
    - 1.1.8. High-Throughput Computing
    - 1.1.9. Reconfigurable Computing (FPGAs)
  - 1.2. Data Storage
  - 1.3. Network and Communication Equipment
  - 1.4. Remote Sensing
2. Software
  - 2.1. Application Software
  - 2.2. CI/CD and Process Automation
  - 2.3. Cloud Computing Tools, Frameworks, and Environments
  - 2.4. Code Optimization and Modernization Tool
  - 2.5. Code Compression Tools
  - 2.6. Cybersecurity
  - 2.7. Content Management Tools
  - 2.8. CRM tools
  - 2.9. Data Management
    - 2.9.1. Data Archival and Preservation
    - 2.9.2. Data Protection and Recovery
    - 2.9.3. Data Privacy
    - 2.9.4. Data Warehousing and Data Lake Platforms
    - 2.9.5. Database and Data Processing
      - 2.9.5.1. Distributed Databases
    - 2.9.6. Semantic Web
  - 2.10. Decision-Support System, Expert System, Knowledgebase
  - 2.11. Embedded Software
  - 2.12. Fault-tolerance
  - 2.13. Filesystems and Parallel I/O
  - 2.14. Generative programming tools and frameworks
  - 2.15. High-level interfaces, libraries, compilers, and runtime systems for parallel programming
  - 2.16. HPC Science Gateways (HPC in the Cloud)
  - 2.17. Large-scale HPC applications (tuning, optimization, and implementation on HPC resources)
  - 2.18. Image and Video Processing
  - 2.19. Learning Management Systems
  - 2.20. Measurement and Monitoring
  - 2.21. Mobile Applications
  - 2.22. Networking
  - 2.23. Pattern Recognition
  - 2.24. Programming Languages, Programming Environments, and Runtime Systems
  - 2.25. Quantum Computing toolkits
  - 2.26. Simulation Platforms
  - 2.27. Software Configuration Management
  - 2.28. Software Libraries
  - 2.29. Software Verification and Validation
  - 2.30. Supporting Software and Middleware for HPC environments
  - 2.31. System and Network Management Software
  - 2.32. Tools and techniques for Code Modernization
  - 2.33. Tools and techniques for Memory and Power Optimization
  - 2.34. Tools for Profiling, Debugging, and Parallelizing Applications
  - 2.35. Tools for Supporting Collaborative and Virtual Environments, Virtual Assistance
  - 2.36. Tools for Supporting Volunteer Computing and High-Throughput Computing
  - 2.37. Tools and Environments for Scientific Visualization
  - 2.38. Virtualization and Containerization Software
  - 2.39. Web Portals, Web Services, Middleware, and Web-accessible Products
  - 2.40. Workflow Management
3. Data
  - 3.1. Structured Databases
  - 3.2. Unstructured Data Collections
  - 3.3. Semi-Structured Data
4. Processes
  - 4.1. Software Engineering
  - 4.2. Cybersecurity
  - 4.3. Project Management
  - 4.4. Automation
5. People
  - 5.1. Broadening Participation
  - 5.2. Management of Resources
  - 5.3. Research Facilitation Services
  - 5.4. Workforce Development
    - 5.4.1. Training
    - 5.4.2. Education
6. Combination of the above
7. Other

# CI Project Metrics: Mapping General and Specific Metrics – MICI Model

| Level | Metrics Type               | Description                                   |
|-------|----------------------------|---|
| 1     | Project-Specific Metrics   | Metrics unique to each CI project             |
| 2     | Category-Specific Metrics  | Metrics common to a category of CI projects   |
| 3     | Generic CI Project Metrics | Top-level metrics, common to all CI projects  |
| 4     | Project Rating             | A numeric rating derived from Level 3 metrics |



| Project Name:        | XYZ                      |       |        |       |                          |        |       |                   |        |
|----------------------|--------------------------|-------|--------|-------|--------------------------|--------|-------|-------------------|--------|
| <Other Project Info> |                          |       |        |       |                          |        |       |                   |        |
| MICI Category:       | Software                 |       |        |       |                          |        |       |                   |        |
| Project Rating       |                          |       |        |       |                          |        |       |                   |        |
| Project Metric Group | Project-Specific Metric  | Value | Weight | Score | Category-Specific Metric | Weight | Score | Generic CI Metric | Weight |
| Software Development | Services created         |       |        |       | % Modules developed      | 25%    |       | Progress          |        |
| Software Development | Help pages/files created |       |        |       | % Documentation Done     | 5%     |       | Progress          |        |
| Data Analysis        | Data points analyzed     |       |        |       | Reports prepared         | 10%    |       | Progress          |        |
| Training & Outreach  | Trainings Conducted      |       |        |       | Trainings Conducted      | 15%    |       | Success           |        |
| Implementations      | Use cases covered        |       |        |       | % Requirements Met       | 30%    |       | Success           |        |
| Publications         | Number of publications   |       |        |       | Number of publications   | 10%    |       | Impact            |        |
| Citations            | Number of citations      |       |        |       | Number of citations      | 5%     |       | Impact            |        |
|                      |                          |       | 0%     | 0%    |                          | 100%   |       |                   |        |

Dimensions

Levels

Processes

# CI Project Metrics: Tools and Enablers

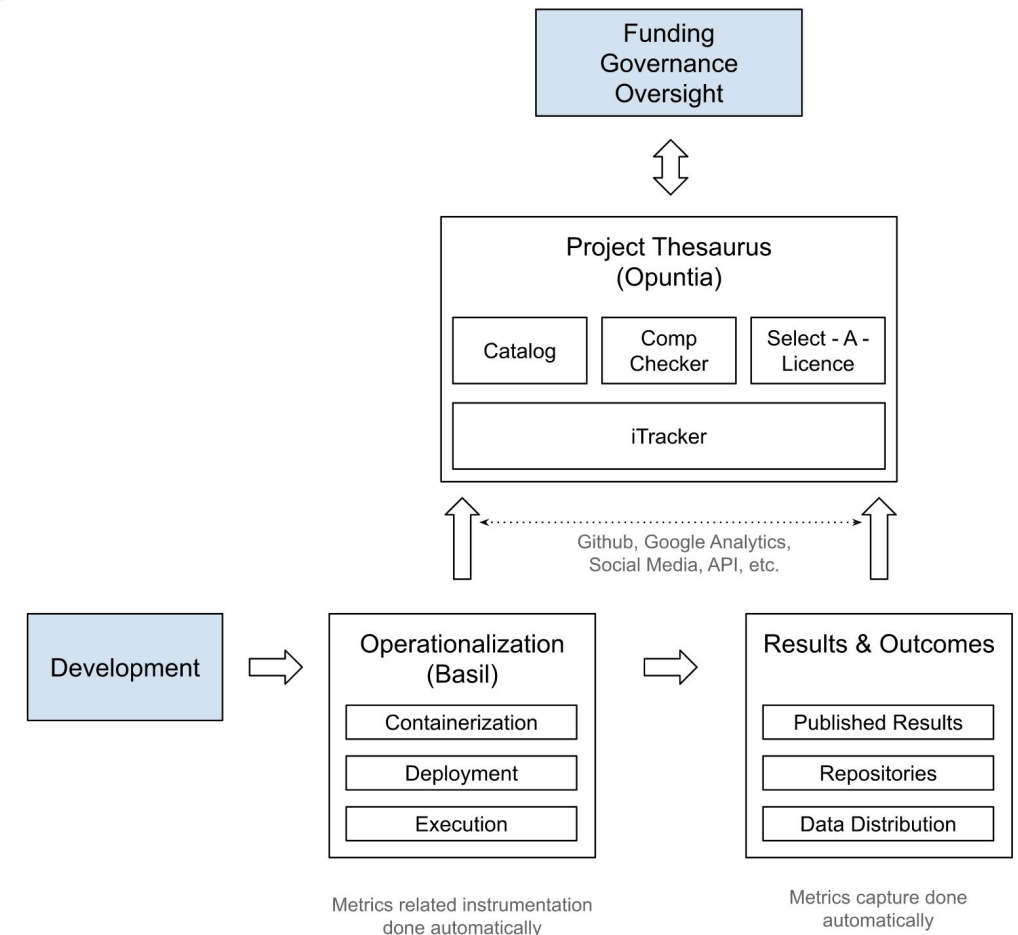
- ▶ **Opuntia:** Provides prototype cyberinfrastructure for facilitating the assessment, discovery, dissemination, and reuse of publicly accessible software and data products; selected subcomponents:
  - a. CI Project Taxonomy: Classification scheme for CI projects
  - b. iTracker: Tracks, gathers, and aggregates user-defined/project-specific metrics
  - c. CompChecker: for checking the compatibility of the products and determining if they could legally and technically interoperate
  - d. Select-A-License tool
  - e. Catalog of NSF funded projects - include information of projects depending upon other projects

<https://opuntia.online>



# CI Project Metrics: A Case Study

- ▶ A software-centric CI project can benefit from a metrics framework and enabling tools
- ▶ Basil is a project for automating many containerization, deployment, and management tasks, and also feeds operational metrics to Opuntia and to other repositories, databases and portals
- ▶ Opuntia acts as a *Project Thesaurus*, maintaining a catalog of projects organized by the CI Taxonomy, with licence and compatibility checks
- ▶ Funding, governing and other stakeholders can get detailed, aggregated (and potentially correlated) metrics from Opuntia



# Conclusion

- ▶ CI project metrics are difficult to capture and represent in a standardized manner
- ▶ The **MICI Model** provides a standardized, but flexible and customizable, metrics framework for CI projects
- ▶ The case study presented illustrates real-world use, leveraging a publicly available toolset, developed as part of the **Opuntia** and **Basil** projects
- ▶ For **ease of adoption** and **sustainability**, guidelines, associated process and a governance model are also discussed
- ▶ **Future work:**
  - ▶ Evolving the MICI Model
  - ▶ Providing a simple, spreadsheet-based tool
  - ▶ Enhancing and integrating the Opuntia and Basil toolset
  - ▶ Inviting collaborations and deeper engagement key communities



# **Towards Developing a Model for Assessing the Progress, Success, and Impacts of CI Projects**

**Thank You!**