Web Image Processing Pipeline (WIPP) and Polus Services

1Nathan Hotaling & 2Peter Bajcsy
1National Center for Advancing Translational Sciences (NCATS) NIH
2Information Technology Laboratory (ITL) NIST
Foundational Problem

How can a biologist harness petabytes of image data to derive quantitative measurements and gain meaningful insights?
NIST Measurement: It is not just a number!

• Trusted
• Traceable
• Repeatable
• Searchable
• Immutable
• Persistent
• Verifiable by humans
• Accessible to multiple parties

Challenge: Incorporate all measurement attributes
Institutional Challenges

Software Features for

- Users
- Developers
- System Administrators
<table>
<thead>
<tr>
<th>Ingestion + OMERO/OpenBIS + CryoSpark + Python</th>
<th>Automated Data Ingestion</th>
<th>Fast Data Previews</th>
<th>GUI-Workflows and Analysis</th>
<th>Hierarchical Catalog GUI</th>
<th>Collaborative Sharing &amp; Coding</th>
<th>Advanced Automated Analysis</th>
<th>Simple Metadata Association</th>
<th>Easily Extensible</th>
<th>Native Broad Support</th>
<th>Scales Easily &amp; Efficiently</th>
<th>Built-in Traceability and Reproducibility</th>
<th>Secure</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ingestion + OMERO/OpenBIS + IMOD + MATLAB</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Ingestion + Columbus + Palantir/Spotfire/etc.</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Ingestion + Midas + CMake + VTK + ITK + Paraviewer</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Polus/WIPP 3.0</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Ingestion +</td>
<td>Open Source</td>
<td>Robust Code Documentation and Installation</td>
<td>Access to All Feature and Resource APIs</td>
<td>Well Documented APIs</td>
<td>Long Term Legacy Support APIs</td>
<td>Easy to Customize GUI</td>
<td>Simple Multi-Lingual Plugins Support</td>
<td>Open Plugin Repository</td>
<td>Easy &amp; Automated Porting of Plugins</td>
<td>Native Broad Microscope Support</td>
<td>Collaborative Sharing/Coding</td>
<td>Scales Easily &amp; Efficiently</td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>-------------</td>
<td>------------------------------------------</td>
<td>------------------------------------------</td>
<td>----------------------</td>
<td>-------------------------------</td>
<td>----------------------</td>
<td>-------------------------------------</td>
<td>-----------------------</td>
<td>-----------------------------------</td>
<td>-------------------------------</td>
<td>-----------------------------</td>
<td>-----------------------------</td>
<td></td>
</tr>
<tr>
<td>OMERO/OpenBIS</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>CryoSpark + Python</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Ingestion +</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>OMERO/OpenBIS</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>IMOD + MATLAB</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Ingestion +</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Columbus +</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Palantir/Spotfire/etc</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Ingestion + Midas +</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>CMake + VTK + ITK + Paraviewer</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Polus/WIPP 3.0</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>
### Data Independent Admin. Desired Features

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ingestion + OMEROS/OpenBIS + CryoSpark + Python</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ingestion + OMEROS/OpenBIS + IMOD + MATLAB</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ingestion + Columbus + Palantir/Spotfire/etc.</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ingestion + Midas + CMake + VTK + ITK + Paraviewer</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polus/WIPP 3.0</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Link to Sheet**
Technical Approach

The computational and institutional challenges cannot be addressed without the involvement of communities of stakeholders

- Open Source Web Image Processing Pipeline (WIPP)
- Cloud service called Polus
High-Level Perspective on WIPP

Automated Processing:
- Flat field correction
- Noise filtering
- Stitching
- Image pyramid building

Interactive Analyses:
- Tabular and Image Annotations
- Visual Verification and Validation
- Sub-setting and Measurements

Pyramid Representation of One Large Field of View

Upload
Workflow Construction

Visualization
Computation

Cloud Computing
Servers

Many Small Image Fields of View
Algorithmic Plugins
Interactive Interfaces

Servers
Workflow Construction
Visualization
Computation

Upload
Workflow Construction

Cloud Computing
Servers

Pyramid Representation of One Large Field of View
WIPP: Client - Server System - Image Processing + Analysis
WIPP 3.0 Client-Server Architecture

Focused on Functionalities for Users, Developers, and System Administrators

- **Web Browser**: Client app
- **Database**: MongoDB 3.6
- **E.g. Java Spring Boot 2.x**
- **File System**: Jobs
- **Frontend UI**: Angular 6+
- **Job Management**: Argo Workflow
- **Job Status Notifications**: HTTPS
- **Kubernetes**: TCP

**Diagram**: Shows the interactions between these components.
Containerization and Orchestration Technologies

- Docker
  
  “Docker containers wrap up software and its dependencies into a standardized unit for software development that includes everything it needs to run: code, runtime, system tools and libraries.”

- Kubernetes
  
  “Kubernetes is an open-source system for automating deployment, scaling, and management of containerized applications. Production-Grade Container Orchestration.”
Container Workflow Technologies

- Argo Workflows

“Argo Workflows is an open source container-native workflow engine for orchestrating parallel jobs on Kubernetes.”
Algorithmic Plugins

Plugin Components

- **JSON Manifest**
  - Location of docker file
  - Inputs/Outputs
  - User interface definition

- **Docker Container**
  - Algorithms
  - Libraries
  - Abstraction Middleware

https://hub.docker.com/
Plugin Component: Docker Image

WIPP plugin as a Docker image available on DockerHub

Docker image definition (Dockerfile)

```
FROM wipp/pyramid-building:1.0.0
LABEL maintainer="National Institute of Standards and Technology"
ARG EXEC_DIR="/opt/executables"
ARG DATA_DIR="/data"

# Create folders
RUN mkdir -p $(EXEC_DIR) \n    && mkdir -p $(DATA_DIR)/inputs \n    && mkdir $(DATA_DIR)/outputs

# Install java 8 jdk
RUN apt-get update \n    && apt-get install --yes openjdk-8-jdk \n    && update-alternatives --set java /usr/lib/jvm/java-8-openjdk-amd64/jre/bin/java

# Copy wipp-pyramid-plugin JAR
COPY target/wipp-pyramid-plugin+.jar $(EXEC_DIR)/wipp-pyramid-plugin.jar

# Set working directory
WORKDIR $(EXEC_DIR)

# Default command. Additional arguments are provided through the command line
ENTRYPOINT ["/usr/bin/java", "-jar", "wipp-pyramid-plugin.jar"]
```
Dynamic Plugin UI: JSON Manifest

Information about the plugin:
- name, version, description
- Docker image to use

Inputs description
- name, type, description

```
{
  "name": "WIPP Pyramid plugin",
  "version": "0.6.1",
  "title": "WIPP Pyramid building",
  "description": "Pyramid building using NIST accelerated C++ algorithm",
  "creator": "National Institute of Standards and Technology",
  "containerId": "wipp/wipp-pyramid-plugin:0.6.1",
  "inputs": [
    {
      "name": "inputImages",
      "type": "collection",
      "options": {
        "format": "tilediff"
      },
      "description": "Input Images",
      "required": true
    },
    {
      "name": "inputStitchingVector",
      "type": "stitchingVector",
      "description": "Input Stitching Vector",
      "required": true
    },
    {
      "name": "blending",
      "type": "enum",
      "options": {
        "values": [
          "overlay",
          "max"
        ],
      },
      "description": "Blending method when assembling tiles",
      "required": false
    }
  ],
  "outputs": [
    {
      "name": "output",
      "type": "pyramid",
      "description": "Output pyramid"
    }
  ]
}
```

Outputs description
- name, type, description

WIPP will use this manifest to generate the job configuration form.
Workflow Creation from Algorithmic Plugins

- Interactive workflow creation and visualization
- Configure workflow tasks
- Chain job outputs
Workflow Monitoring and Traceability

- Monitor Execution in Argo
- U-Net training
- U-Net inference
- Monitor Convergence in TensorBoard
- Traceable Parameters

Monitor Execution in Argo

TensorBoard
Interoperable Algorithmic Plugins

Pros of Container Plugins

• Multi-lingual
  • C++, Python, Java, TensorFlow, etc.

• Containerized
  • Reproducibility of results
  • Compatibility of library dependencies
  • Simplicity of deployment
  • Modularity of analyses

• OSS or Paid

• Interoperable between NIH & NIST
Polus Services: Monitoring, Execution, and Data Storage - Managed Infrastructure!

- User
- Ingestion Daemon
- Network Attached Storage
- Image Acquisition
- Resource and Application Monitoring
- Analysis and Workflow Execution
- Data and Metadata Storage
- Elasticsearch
- Database
- Ingestion Daemon
- Network Attached Storage
- Image Acquisition
- Workflow Plugins
- Resource and Application Monitoring
- Analysis and Workflow Execution
- Data and Metadata Storage

Cloud Services:
- Amazon Web Services
- Google Cloud Platform
- Microsoft Azure
Plugins and Interaction
Algorithmic Plugins

- Image stitching (MIST algorithm)
- Image thresholding (ImageJ/Fiji multiple algorithms)
- Pyramid building (accelerated algorithm with low RAM requirements)
- Universal Notebook Execution (Scalably Execute Jupyter Notebooks as plugins)

Many Overlapping Microscope Fields of View

Image Stitching and Pyramid Building

Dynamic Workflow Builder

Stitched Interactive Image Pyramids
Interactive Data-Driven Visual Discovery

Illustration of WDZT annotation tool

Illustration of linking features and ROIs

Interactive whole sample visualization/manipulation
Prototyping Using Jupyter Notebook Interactive Programming

Configurable in-browser Polyglot IDE located where images are stored

Collaborative Notebooks using GitHub GUI or terminal

Tight Integration with WIPP Jupyterlab Extension
Prototyping Using Jupyter Notebook Interactive Programming

Configurable in-browser Polyglot IDE located where images are stored

Collaborative Notebooks using GitHub GUI or terminal

Tight Integration with WIPP Jupyterlab Extension
Scalable Interactive Visualization of Tabular Data

- DeepZoom of all pair-wise feature scatter plots
- Interactive data sub-setting across dimensions, zoomed and global views
- Familiar dynamic group selection tools with a history dendrogram and plot view
Scalable Interactive Visualization of Tabular Data

- DeepZoom of all pair-wise feature scatter plots
- Interactive data sub-setting across dimensions, zoomed and global views
- Familiar dynamic group selection tools with a history dendrogram and plot view
Examples of Data Sharing and Dissemination

- **Provenance data**: hyperlinked computations with all metadata about data and execution configurations
- **Large size image data**: pyramid representation with Deep Zoom enabled browsing of images
- **Image-based measurements**: hyperlinked thumbnails and numerical values to persistent image ROIs
Dissemination (WIPP 3.0 Beta version)

- Source code is in GitHub repositories
  - WIPP (main repo with documentation and deployment instructions): https://github.com/usnistgov/WIPP
  - WIPP-backend (Java Spring REST API): https://github.com/usnistgov/WIPP-backend
- Docker containers are in Docker Hub
- We provide single node installation instructions for all platforms
  - Tested on Linux 18.04, Windows 10 Enterprise, and Mac OS10
- WIPP 3.0 installation includes
  - Image processing and AI semantic segmentation plugins
  - Integrated Jupyter notebook
    - Prototyping in 13 programming languages
  - Interactive scatter plots
  - Visualization and sub-setting of millions of image-based measurements.
**Credit**

<table>
<thead>
<tr>
<th>NIST</th>
<th>NIH</th>
<th>NIST BIO/MAT</th>
<th>External Collaborator</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Mary Brady</td>
<td>• Sam Michael</td>
<td>• Anne Plant</td>
<td>• Marcin Kociolek</td>
</tr>
<tr>
<td>• Peter Bajcsy</td>
<td>• Anton Simeonov</td>
<td>• Kiran Bhadriraju</td>
<td>• Piotr Szczyński</td>
</tr>
<tr>
<td>• Joe Chalfoun</td>
<td>• Robert Hohman</td>
<td>• Michael Halter</td>
<td>• Carol Parent</td>
</tr>
<tr>
<td>• Mike Majurski</td>
<td>• Beth Fischer</td>
<td>• John Elliott</td>
<td>• Christina Stuelten</td>
</tr>
<tr>
<td>• Mylene Simon</td>
<td>• Marc Ferrer</td>
<td>• Jeff Stinson</td>
<td>• Michael Weiger</td>
</tr>
<tr>
<td>• Adele Peskin</td>
<td>• Sheldon Miller</td>
<td>• Ed Kwee</td>
<td>• Wolfgang Loser</td>
</tr>
<tr>
<td>• Petre Manescu</td>
<td>• Brian Brooks</td>
<td>• Carl Simon</td>
<td>• Amitabh Varshney</td>
</tr>
<tr>
<td>• Soweon Yoon</td>
<td>• Kapil Bharti</td>
<td>• Stephen Florczyk</td>
<td>• Daniel Hoeppner</td>
</tr>
<tr>
<td>• Mohamed Ouladi</td>
<td>• Tiziana Cogliati</td>
<td>• Nick Schaub</td>
<td>• Ron McKay</td>
</tr>
<tr>
<td>• Sarala Padi</td>
<td>• Arvydas Maminishkis</td>
<td>• Sumona Sarkar</td>
<td>• Jana Kosecka</td>
</tr>
<tr>
<td>• Antoine Gerardin</td>
<td>• Bryan Hansen</td>
<td>• John Henry Scott</td>
<td>• Suhas Sharma</td>
</tr>
<tr>
<td>• Philippe Dessauw</td>
<td>• Yuchi Chen</td>
<td>• Nicholas Ritchie</td>
<td>• Rafat Sarosh</td>
</tr>
<tr>
<td>• Walid Keyrouz</td>
<td>• Kelli Wilson</td>
<td>• Keana Scott</td>
<td>• Gary Mays</td>
</tr>
<tr>
<td>• Derek Juban</td>
<td>• Nina Mezu</td>
<td>• Charlie Camp</td>
<td>• Dan Gareau</td>
</tr>
<tr>
<td>• Antonio Cardone</td>
<td>• Reid Simon</td>
<td>• Young Lee</td>
<td></td>
</tr>
<tr>
<td>• Tim Blattner</td>
<td>• Busola Grillo</td>
<td>• Marcus Cicerone</td>
<td></td>
</tr>
</tbody>
</table>

(1) Technical University of Lodz, Poland  
(2) NIH National Cancer Institute  
(3) University of Maryland at College Park  
(4) The Lieber Institute for Brain Development  
(5) George Mason University  
(6) Axle Informatics  
(7) Rockefeller University

Thank you to all who aren’t specifically called out here!  
The goal is huge and so is the team that got us here.
Disclaimer

Commercial products are identified in this document in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by NIST, nor is it intended to imply that the products identified are necessarily the best available for the purpose.