Beyond the Instrument: How to Manage and Master Your Core Facility Data

Aarya Vaikakkara Chithran, Ph.D.

Research Scientist
Microscopy and Histopathology CoLab
Seattle Children's Research Institute

Seattle Children's®

Overview of Seattle Children's Research Institute

Core Services

Shared Research Resources (SRR)

- ABSL-3 Core Facility
- Behavioral Phenotyping Core
- Flow Cytometry Core
- Genomics and Spatial Biology CoLab
- Microscopy and Histopathology CoLab
- Preclinical Imaging Core

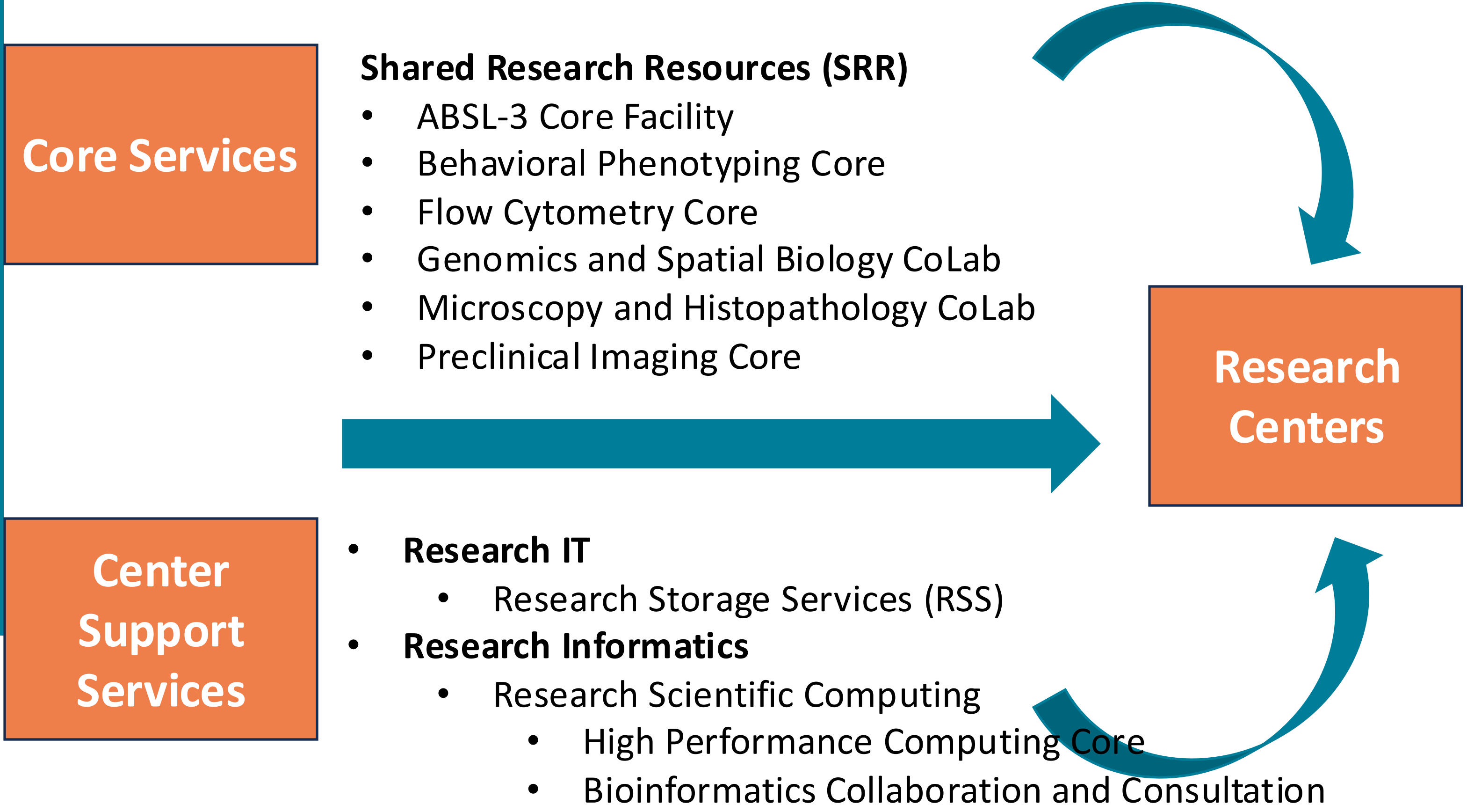
Research Centers

Center Support Services

- **Research IT**
  - Research Storage Services (RSS)
- **Research Informatics**
  - Research Scientific Computing
    - High Performance Computing Core
    - Bioinformatics Collaboration and Consultation

Seattle Children's

Overview of Seattle Children's Research Institute



Core Services

Shared Research Resources (SRR)
- ABSL-3 Core Facility
- Behavioral Phenotyping Core
- Flow Cytometry Core
- Genomics and Spatial Biology CoLab
- Microscopy and Histopathology CoLab
- Preclinical Imaging Core

Research Centers

Center Support Services

- **Research IT**
  - Research Storage Services (RSS)
- **Research Informatics**
  - Research Scientific Computing
    - High Performance Computing Core
    - Bioinformatics Collaboration and Consultation

Data Workflow in SRR

Core Services

Shared Research Resources (SRR)

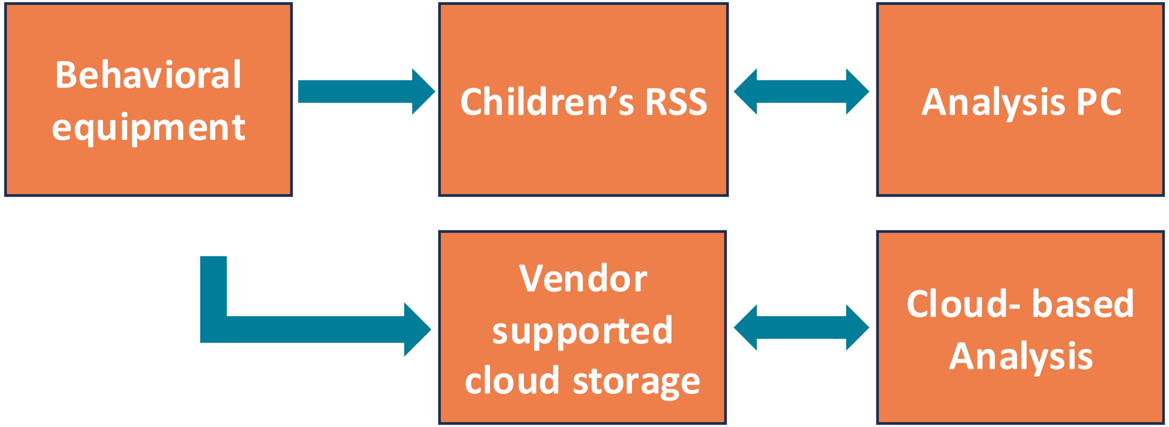
- ABSL-3 Core Facility
- Behavioral Phenotyping Core
- Flow Cytometry Core
- Genomics and Spatial Biology CoLab
- Microscopy and Histopathology CoLab
- Preclinical Imaging Core

Seattle Children's®

Data Workflow in SRR

Behavioral Phenotyping Core (BPC)

- Data type: Behavioral videos (few hours to > 24 hours)
- Data size: ~2 TB/ day
- Data storage: AWS cloud (integrated with Allentown Inc.) → Children's RSS
- Data Analysis: Machine-Learning; High-end PC with a dedicated graphics card

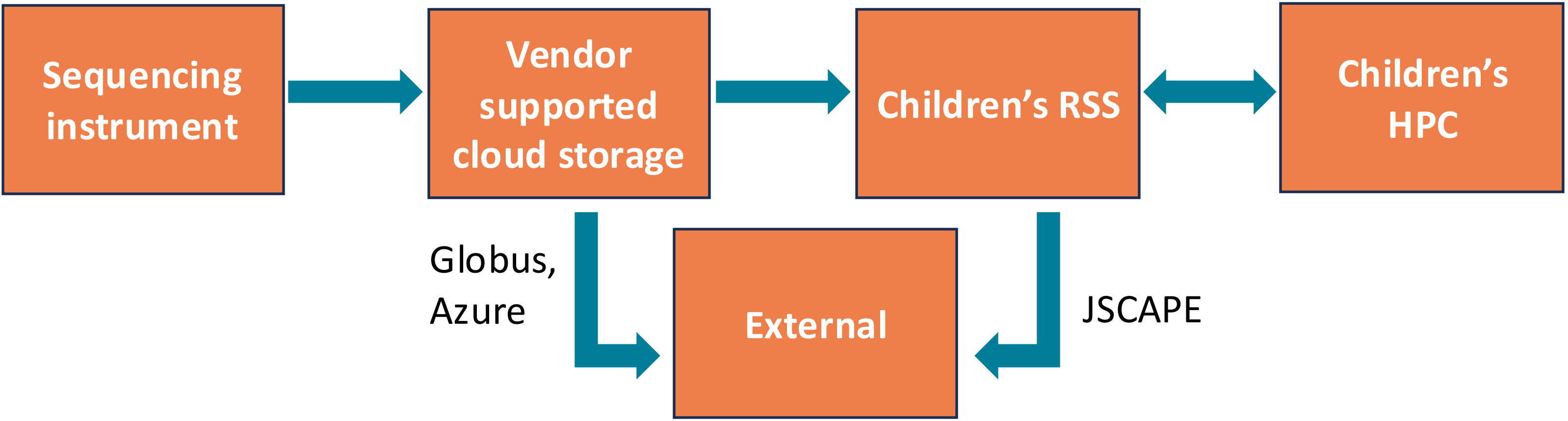


Seattle Children's

Data Workflow in SRR

Genomics and Spatial Biology CoLab (GSB)

- Data type: Sequencing data from Illumina/ PacBio instruments
- Data size: ~10 GB- 1 TB per file
- Data storage: Instrument cloud → Children's RSS
- Data Analysis: Children's HPC
- External Transfer: Globus Connect, JSCAPE, Azure



Seattle Children's

Data Workflow in SRR Microscopy

Data Generation > Data Processing > Data Analysis …..Data Storage + Transfer

Seattle Children's®

Data Workflow in SRR Microscopy

Data Generation > Data Processing > Data Analysis …..Data Storage + Transfer

Instrument	Avg. Raw data	Usage
Zeiss LSM 900 (Airyscan) Confocal	1- 80 GB	3-4 users per day; 8- 10 hours per day
Zeiss LSM 980 (Airyscan) Confocal	1- 80 GB	3-4 users per day; 8- 10 hours per day
Zeiss Axioscan 7 slide scanner	2- 20 GB	20-100 slides per week
LifeCanvas SmartSPIM light-sheet	1 TB	moderate- heavy usage
ImageXpress Nano High-Content Imager		Low-moderate usage
CSU W1 SoRa Spinning disk Confocal		moderate- heavy usage

Seattle Children's

Data Workflow in SRR Microscopy

Why use HPC for microscopy data?

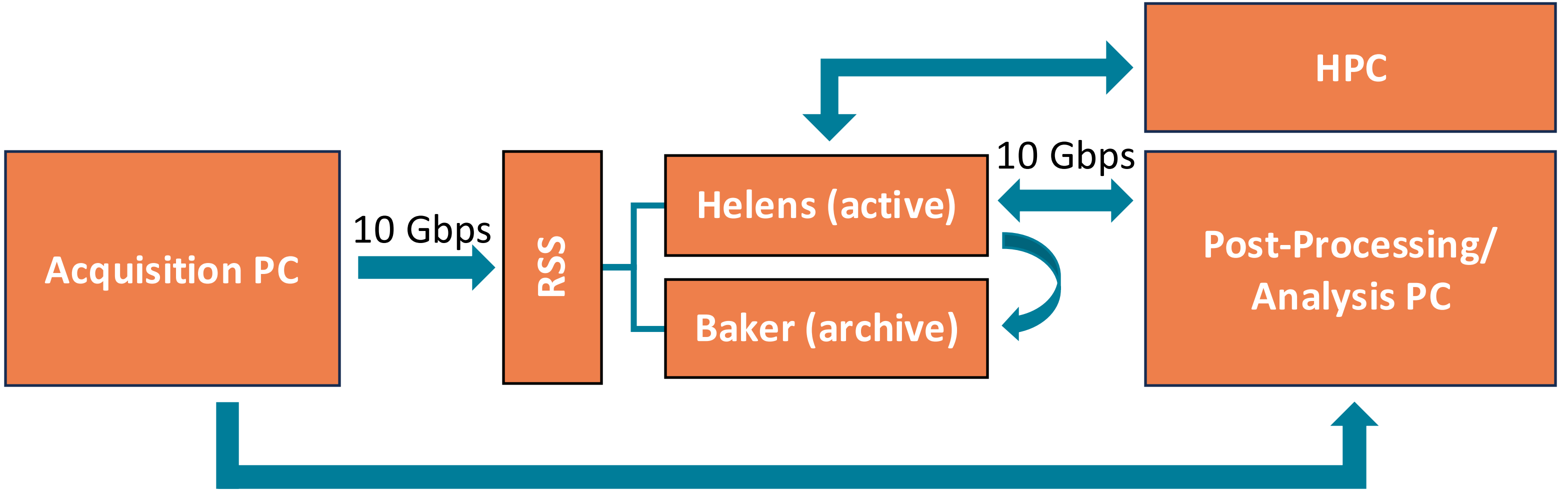
- Explosion in data volume and complexity (increasing > 5x in recent years)
- Growing demand for rapid analysis and sharing
- Need for scalable, reliable infrastructure

HPC can:

- Handle large-scale processing (TB-scale datasets)
- Accelerate image analysis workflows (segmentation, registration, AI)
- Faster turnaround for users
- Enable automated, reproducible workflows
- Better sharing and collaborative access

Data Workflow in SRR Microscopy

Data Generation > Data Processing > Data Analysis …..Data Storage + Transfer

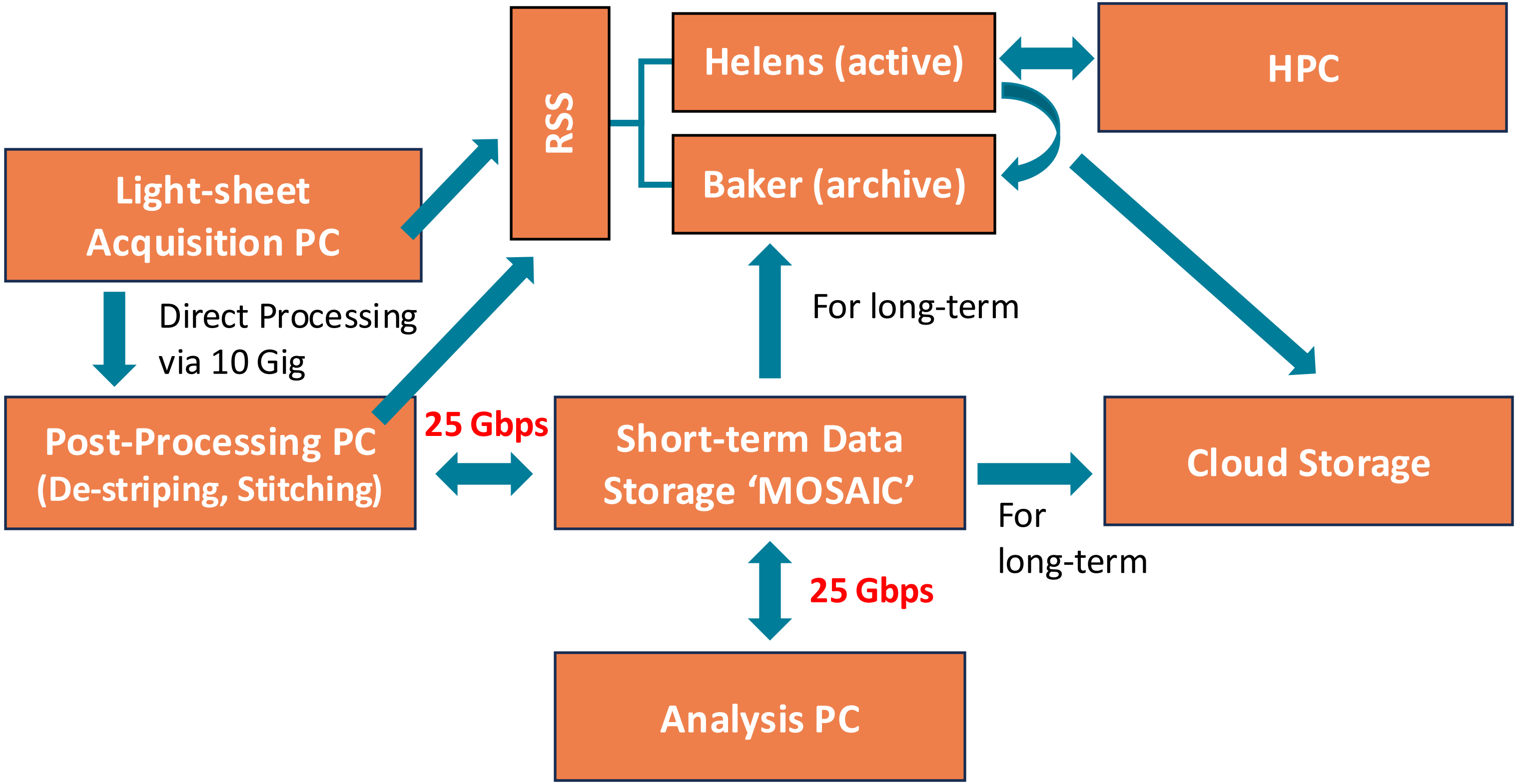


Direct Processing during
acquisition via 10 Gig connection

Seattle Children's

Data Workflow in SRR Microscopy

Data Generation > Data Processing > Data Analysis …..Data Storage + Transfer



Light-sheet Acquisition PC

Direct Processing via 10 Gig

Post-Processing PC (De-striping, Stitching)

25 Gbps

RSS

Helens (active)

Baker (archive)

HPC

For long-term

Short-term Data Storage 'MOSAIC'

For long-term

Cloud Storage

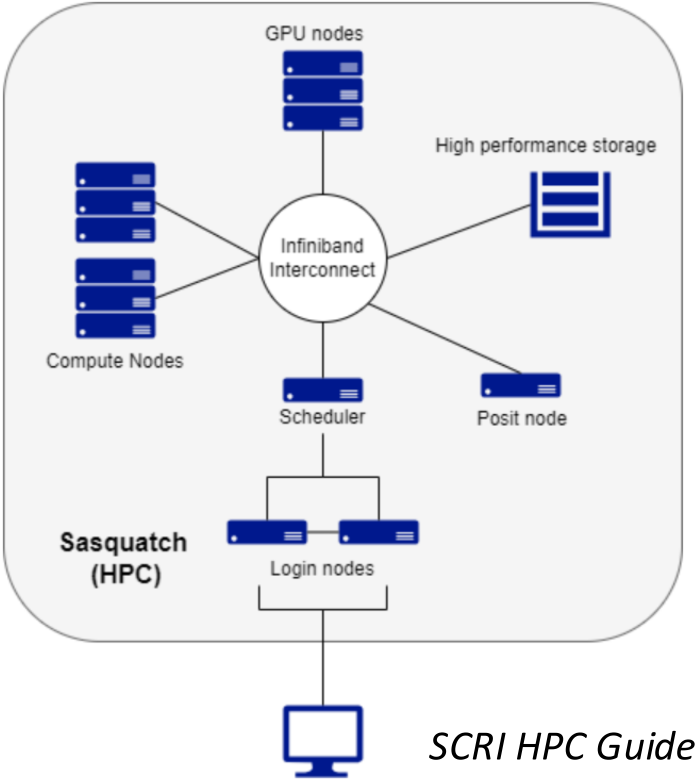
25 Gbps

Analysis PC

Seattle Children's

HPC Workflow

- Onboarding to request an account
- Orientation and Training
- **Connect to HPC** (login nodes)
  - Home directory (100 GB)
  - Association directory (1 TB increments)
- **Transfer data** from RSS to HPC (not queued)
- Use **Slurm** for scheduling (jobs are queued)
  - Determine the resources required
  - CPU, GPU, memory, time
- Job dispatched to one or more of the nodes
- Job completed- output written to the specified subfolder in association directory
- Transfer analyzed data to RSS



SCRI HPC Guide

Seattle Children's

Using HPC for Microscopy Data

Use Case 1: 3D/ 4D/ 5D Confocal analysis

- Large screen of >800 mutants
- High-content imaging of several 384-well plates
  - 10+ ROIs per well
  - Multichannel z-stacks; time series
- Repetitive pipeline (masking, filter, segmentation, quantification of 5+ parameters)
- Ideal for containerized batch analysis on HPC
- **1.3 million cell instances segmented per day on HPC**
- 30 experiments run parallel for analysis

Seattle Children's

Using HPC for Microscopy Data

Use Case 2: Light-sheet Imaging

- 1 TB per sample
- Destriping, stitching, tile correction, atlas registration
- Too slow and resource heavy on local machines

Pilot project

- Identify workflows common to all users (post-processing of light-sheet data)
  - Destriping using ‘Pystripe’
  - Stitching using ‘TeraStitcher’
- Deploy shared storage + processing containers
- Evaluate usage and impact → scale as needed

Seattle Children's

Integration Challenges

Data Transfer Bottlenecks

- Slow **data transfer** speeds from acquisition/ RSS to HPC storage creates workflow delays.

Workflow Compatibility

- Many microscopy tools are **Windows-only or GUI-based**, while HPC environments are typically **Linux and CLI-based**.
- Software requiring **proprietary licenses** may not be easily portable to HPC.

Usability Gap

- Most users are **more comfortable with GUIs**, but HPC access often requires **command-line skills**.
- Limited coding experience makes it hard for users to **optimize or build pipelines** on their own.

Progress & Lessons Learned

Progress made:

- **Direct Processing** Pipelines via 10 Gbps
- Established **25 Gbps** dedicated link for light-sheet data transfer
- Launched specialized **image analysis workshops** (Python scripts) for researchers
- Identified pilot projects for **HPC integration**

Lessons learned:

- **Core's Responsibility: Bridge the Gap.** We must translate researcher needs to IT requirements and vice versa.
- **The Partnership Model:** Make **IT your best friend**. Ongoing discussions with Research Support Services (RSS) and HPC are crucial for policy and infrastructure.
- **Actionable Start:** Create and deploy **basic, common pipelines** (e.g., deconvolution, batch processing) to immediately lower the barrier to entry.

Seattle Children's

“People. Ideas. Technology – in that order.”

-attributed to John Boyd

Seattle Children's®

HOSPITAL • RESEARCH • FOUNDATION

Hope. Care. Cure.