

OARC support Overview HPC/REDCap/CRDW

Vlad Kholodovych, PhD

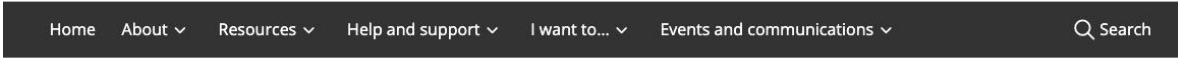
Director of Research Support

Office of Advanced Research Computing (OARC)

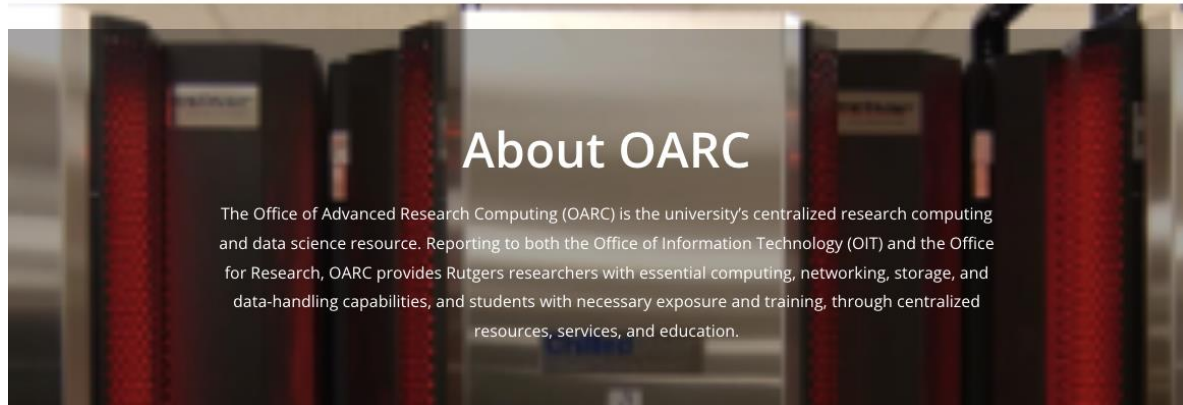
Rutgers University

vlad.khol@rutgers.edu

High Performance Computing (HPC)



OFFICE OF ADVANCED RESEARCH COMPUTING / ABOUT OARC

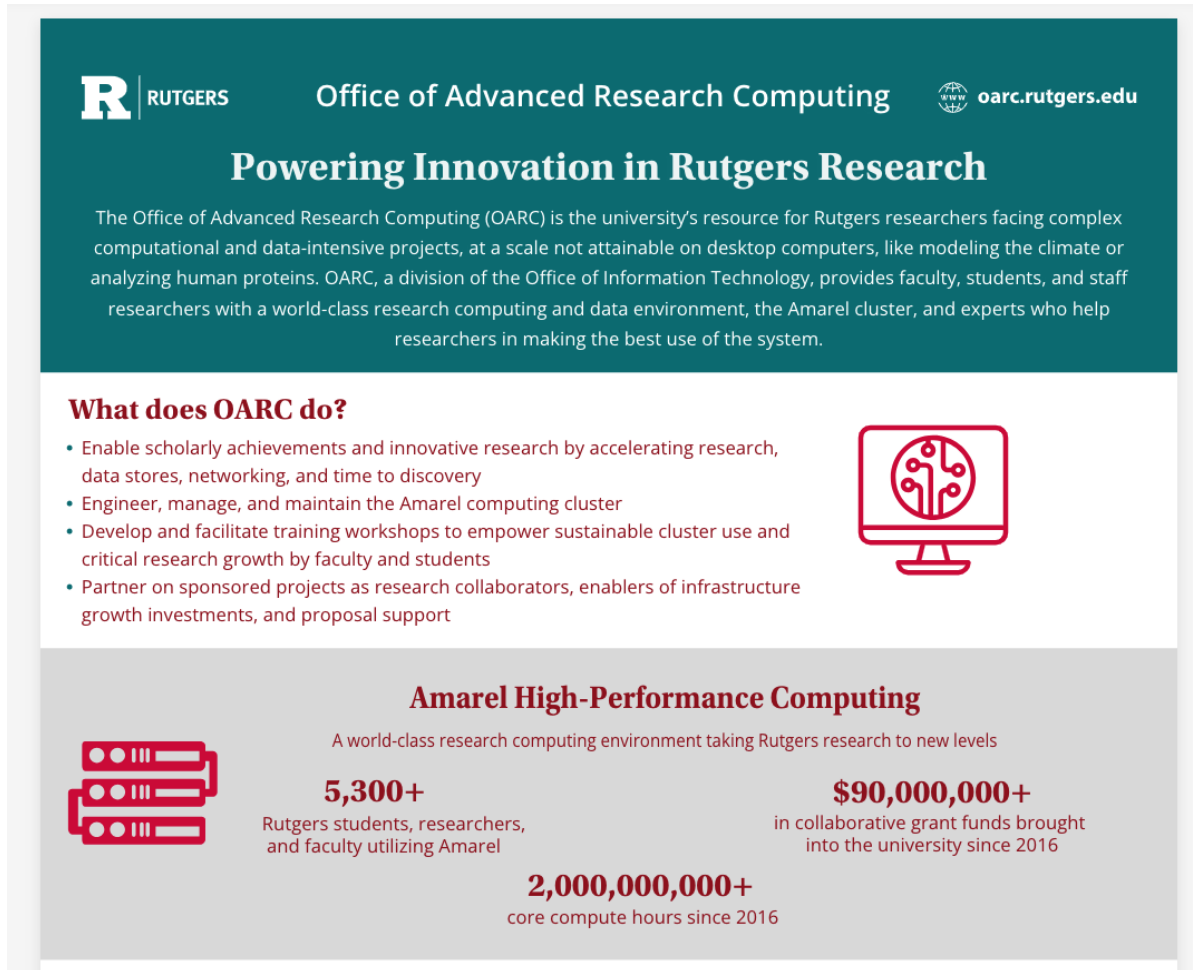


<https://oarc.rutgers.edu>

In addition to these core services, OARC manages critical resources like the Clinical Research Data Warehouse (CRDW), REDCap for secure data collection and management, and the Research Data Storage (RDS) system, ensuring comprehensive support for the university's research community.

The Office of Advanced Research Computing (OARC) is the university's centralized resource for research computing and data science. As part of the Office of Information Technology (OIT), OARC provides Rutgers researchers with essential computing, networking, storage, and data-handling capabilities, and offers students valuable exposure and training through centralized resources, services, and education.

Amarel – A Community Owned HPC cluster (updated March 2024)



The infographic features a teal header with the Rutgers logo and 'Office of Advanced Research Computing' text. Below is a white section with a red icon of a computer monitor displaying a circuit. The bottom section is grey with red icons of server racks and large numerical statistics.

RUTGERS Office of Advanced Research Computing oarc.rutgers.edu

Powering Innovation in Rutgers Research

The Office of Advanced Research Computing (OARC) is the university's resource for Rutgers researchers facing complex computational and data-intensive projects, at a scale not attainable on desktop computers, like modeling the climate or analyzing human proteins. OARC, a division of the Office of Information Technology, provides faculty, students, and staff researchers with a world-class research computing and data environment, the Amarel cluster, and experts who help researchers in making the best use of the system.

What does OARC do?

- Enable scholarly achievements and innovative research by accelerating research, data stores, networking, and time to discovery
- Engineer, manage, and maintain the Amarel computing cluster
- Develop and facilitate training workshops to empower sustainable cluster use and critical research growth by faculty and students
- Partner on sponsored projects as research collaborators, enablers of infrastructure growth investments, and proposal support

Amarel High-Performance Computing

A world-class research computing environment taking Rutgers research to new levels

5,300+ Rutgers students, researchers, and faculty utilizing Amarel	\$90,000,000+ in collaborative grant funds brought into the university since 2016
2,000,000,000+ core compute hours since 2016	

- Officially designated in summer 2016
- Named for Dr. Saul Amarel, Rutgers pioneer in AI
- Available to any Rutgers investigator or student with research computing projects
- Equipment in Piscataway, Newark, and Camden
- 612 Compute nodes
 - 408 in Hill (Piscataway)
 - 132 in Englehard (Newark)
 - 72 in BSB (Camden)
- Total of ~21,000 Intel Xeon cores
- 5300+ Rutgers students, researchers, and faculty work on Amarel
- A total over 2 billion core hours since 2016 (updated March 2024)



OARC expertise in Research Support

Areas of expertise

- Machine learning methods
- Computer architecture, optimization and performance
- Big data and deep learning
- Image processing and visualization
- Design of biomaterials
- Drug design
- Genomics
- Bioinformatics
- Statistical analysis
- Protein folding

General knowledge

- Molecular biology
- Clinical research
- Data processing
- Wet lab routines
- Data analysis

What is different?

- We speak the same language with researchers
- We understand computational demands, help to run and install software programs
- We can suggest different approach to tackle the problem



Certificate program for Rutgers Health students in Clinical Informatics and Data Science is offered by the New Jersey Alliance for Clinical and Translational Science at no cost to students. This program was developed with the assistance of Branimir Ljubic and Travis Williams, Senior Scientists at OARC.

Certificate in Clinical Informatics and Data Science: 3 Digital Badges



[Register Here](#)

[For More Information](#)

Description of the Clinical Informatics and Data Science Certificate

Course Instructors

Branimir Ljubic, Ph.D., M.D. – Dr. Ljubic will be the primary instructor for the core technical elements (computer science methods, machine learning) of the program. He is widely published on the use of machine learning and other computer methods in clinical applications. He holds a Ph.D. in Computer and Information Science from Temple University and an M.D. from Medical School, University of Belgrade. He is the Senior Scientist for Clinical, Medical Informatics and Advanced Research Computing in the Rutgers Office of Advanced Research Computing (OARC). The OARC is the university's centralized resource for research computing, including machine learning and clinical informatics. As part of the Office of Information Technology (OIT), OARC provides Rutgers researchers with essential computing, networking, storage, and data-handling capabilities, and offers students valuable exposure, training, and education in advanced research computing (HPC computing), clinical and medical informatics, machine learning methods, data analysis, and programming languages. In addition to these core services, OARC manages critical resources like the Clinical Research Data Warehouse (CRDW), REDCap for secure data collection and management, and the Research Data Storage (RDS) system, ensuring comprehensive support for the university's research community.

REDCap




































- REDCap stands for Research Electronic Data Capture.
- It's a secure, web-based application designed to support data capture for research studies. REDCap provides tools for building and managing online surveys and databases, allowing researchers to collect and manage data efficiently and securely. It's commonly used in academic research, clinical trials, and other types of research studies where data collection and management are crucial.
- REDCap was created in 2004 at Vanderbilt University.

<https://www.project-redcap.org/>

REDCap 2023

There are currently 5 known supported instances of REDCap at Rutgers University. These include the newly established central OIT instance, as well as instances under the management of The Cancer Institute of NJ, Rutgers School of Dental Medicine, New Jersey Medical School, and Robert Wood Johnson Medical School. Each of these instances are maintained and supported independently and vary in implementation, scope, and availability. The table below provides a high-level summary of the environments based on some of the key parameters. Details for each instance are provided in the next section.

LEGEND	
	Present
	Partially Present
	Not Present

Parameter	OIT - Central Instance	CINJ	RSDM	NJMS	RWJMS
Environments	Prod, Dev/Test	Prod, Test, Dev	Prod, Dev/Test	Prod	Prod, Staging, Test, Dev
Instances	2	7	2	1	4
Version	12.4.12	13.8.1	13.1.32	13.10.0	13.10.0
Users	~220	~250	~75	~4000	~2100
Projects	~210	~300	~30	~2600	~1800
Authentication Methodology	LDAP (Shibboleth) + Table	Ldap + Table	Ldap + Table	Ldap + Table	Ldap + Table
High Availability (Single Datacenter)					
Fault Tolerant (Two Data Centers)					
Disaster Recovery/Technology Continuity Planning					
HIPAA Compliant					
Infrastructure Support					
Application Support					

Enterprise REDCap Oversight Committee led by Erin Bauzyk, Associate Vice President - Health Sciences IT

April – October 2023 Report and Recommendations

During the review process, the committee reviewed the existing REDCap infrastructure, services, support, governance, and the community of users. Fifteen recommendations across these areas were identified (see below). **It is important to note that all the recommendations are dependent upon the first, the creation of a dedicated support team to manage the REDCap environment.**

1. Dedicated Support Team
2. Extended Infrastructure
3. Change Management & Stakeholder Communication Processes
4. Compilation of Regulatory Compliance Documentation
5. Disaster Recovery and Technology Continuity Planning
6. Review and Implement 3rd Party Integrations
7. Leveraging Cloud Resources
8. Project Archival and Data Retention Strategy
9. Clearly Documented Processes
10. Tiered Support Model
11. Templated Language to Support Researchers with Funding
12. Streamlined Ticketing & Customer Communication Methodology
13. Enhanced Website & Training Offerings
14. Establishment of a Shared Governance Structure
15. Establishment of a Community of Practice

Approved New Positions
REDCap Admin - 2
REDCap Analyst – 2



Rutgers University Biostatistics & Epidemiology
Services (RUBIES) from School of Public Health

REDCap current state (2024)

There are currently 5 known supported instances of REDCap at Rutgers University. These include the newly established central OIT instance, as well as instances under the management of The Cancer Institute of NJ, Rutgers School of Dental Medicine, New Jersey Medical School, and Robert Wood Johnson Medical School. Each of these instances are maintained and supported independently and vary in implementation, scope, and availability. The table below provides a high-level summary of the environments based on some of the key parameters. Details for each instance are provided in the next section.

LEGEND	
✓	Present
⊙	Partially Present
✗	Not Present

Parameter	OIT - Central Instance	CINJ	RSDM	NJMS	RWJMS
Environments	Prod, Dev/Test	Prod, Test, Dev	Prod, Dev/Test	Prod	Prod, Staging, Test, Dev
Instances	2	7			4
Version	12.4.12	13.8			13.10.0
Users	~220	~250			~2100
Projects	~210	~300	~30	~2600	~1800
Authentication Methodology	LDAP (Shibboleth) + Table	Ldap + Table		Table	Ldap + Table
High Availability (Single Datacenter)	✓	✗			✗
Fault Tolerant (Two Data Centers)	✗	✗			✗
Disaster Recovery/Technology Continuity Planning	✗	⊙	✓	✗	⊙
HIPAA Compliant	✓	✓	✓	✓	✓
Infrastructure Support	✓	✓	✓	✓	✓
Application Support	✗	⊙	✗	⊙	✗

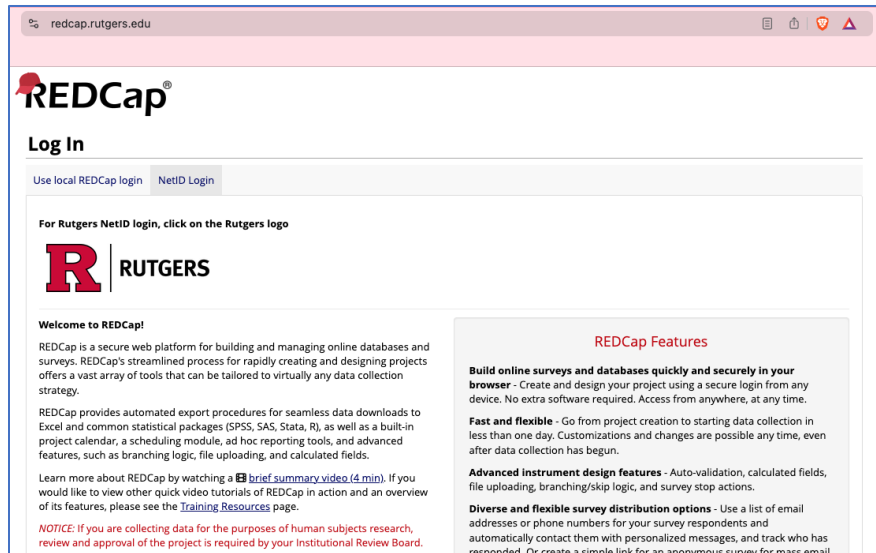
User base growth

2022	59
2023	222
2024	573
Total:	854

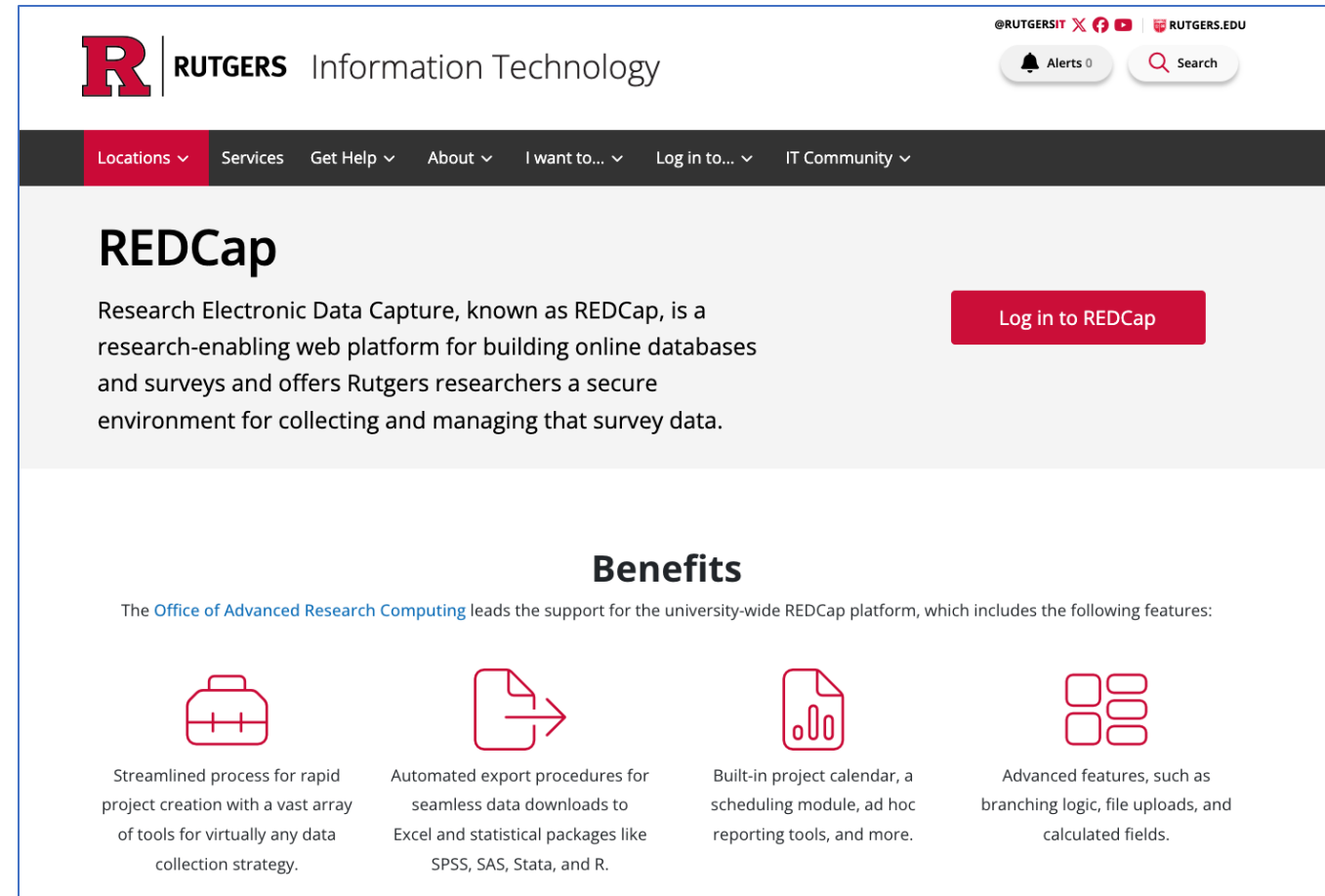
Projects growth

2022	88
2023	158
2024	210
Total:	446

An updated and optimized website with an improved user experience (UX) interface



<https://redcap.rutgers.edu>



<https://it.rutgers.edu/redcap>

Clinical Research Data Warehouse



@RUTGERSIT X f y RUTGERS.EDU



- Locations ▾
- Services
- Get Help ▾
- About ▾
- I want to... ▾
- Log in to... ▾
- IT Community ▾
- CRDW Quick Links ▾

Clinical and Research Data Warehouse (CRDW)

A secure, HIPAA-compliant central repository that includes diverse healthcare data and essential analytical tools for gaining insights into medical conditions, outcomes, and treatments.

crdw.rutgers.edu

What is CRDW?

The Rutgers Clinical and Research Data Warehouse (CRDW) serves as a central repository, uniting diverse healthcare data sources including Electronic Medical Records (EMRs), Clinical Trial Management System (CTMS), Tumor Registries, Biospecimen Repositories, and cutting-edge medical imaging and genomics. [Learn more about CRDW.](#)

JUMP TO:

[ABOUT CRDW](#)

[GENERAL FAQs](#)

[BENEFITS](#)

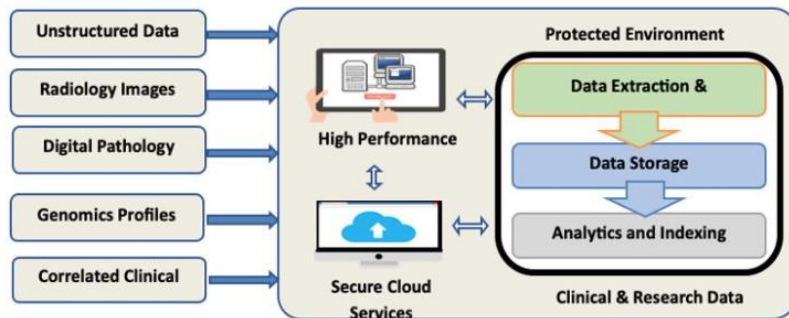
[CORE SERVICES](#)

[FEE SCHEDULE](#)

[CERTIFICATE PROGRAM](#)

Statement from Michele Norin, Senior Vice President and Chief Information Officer at Rutgers

“As a direct outcome of the partnership between Rutgers Health and RWJBarnabas, this has truly been a collaborative effort involving many individuals from across our organizations. Very specifically, we recognize the innovation, effort, creativity, and leadership invested by the team at Rutgers Cancer Institute of New Jersey (CINJ) in developing the **first-generation multimodal clinical and research data warehouse**, automated ETL (extract, transform, load) interface, and mining tools that became the foundation for the enterprise-level CRDW environment. Because of the efforts by all involved, the CRDW is expected to become an invaluable resource to our academic healthcare enterprise.”



Data from the original CINJ legacy data warehouse was initially procured starting in 2012. The CINJ warehouse was among the first multimodal oncology based Clinical Data Warehouses to be successfully developed and implemented.

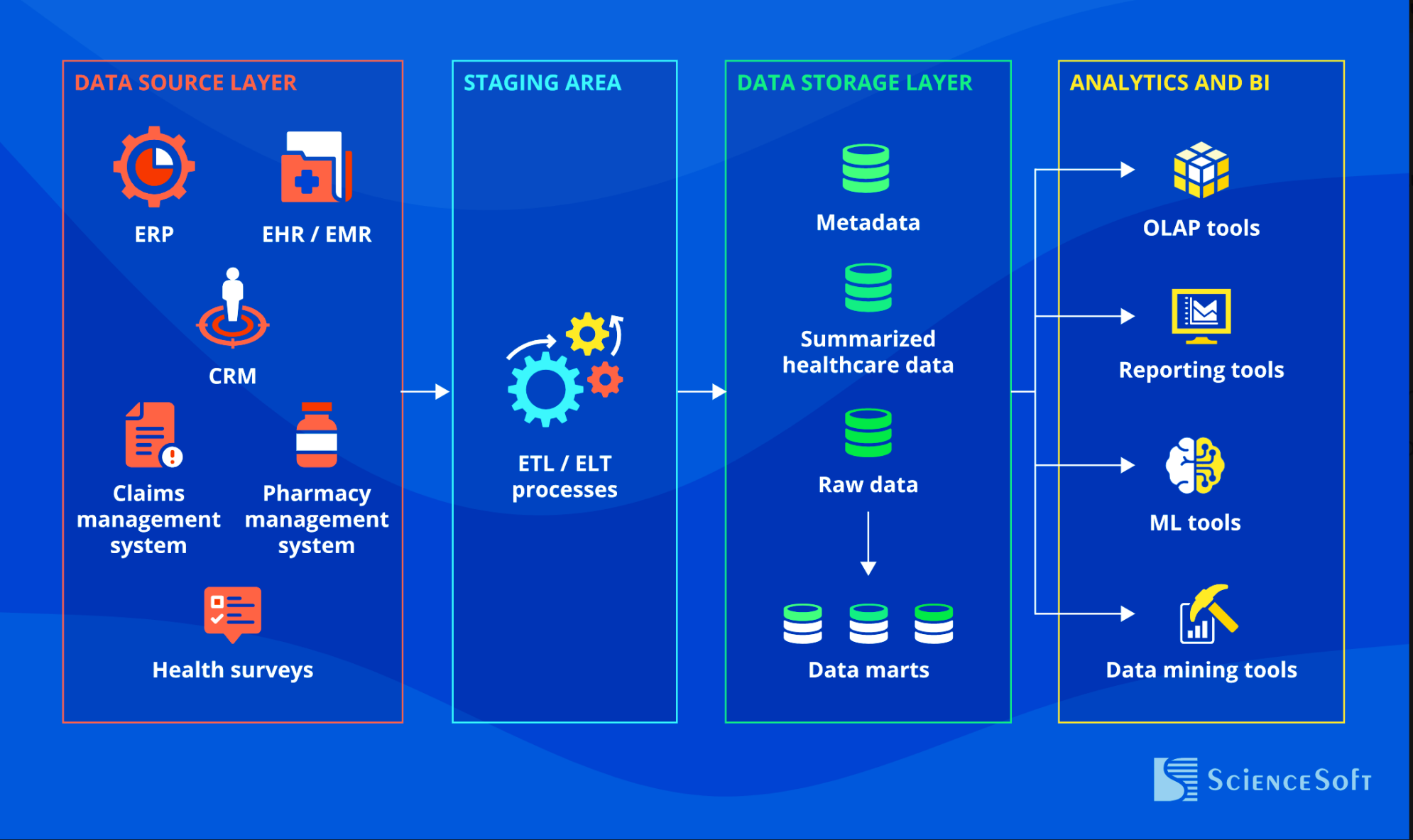


New Jersey Alliance for Clinical and Translational Science



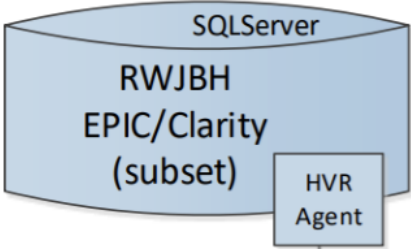
Data from RWJBH transferred to the CRDW includes data from RWJBH hospitals or facilities which is entered into the **Epic electronic health record system (EHR)**. These data were transferred starting in 2021 and included three years of data. Subsequent transfers of data from RWJBH facilities as they went live with Epic occurred in a similar manner (i.e. transmission of three years of data) and will occur as such until the RWJBH Epic roll-out is complete (projected in late 2024.) Additionally, data from all of the facilities live on Epic will be refreshed, going forward, at regular intervals depending on the type of data.

Clinical Data Warehouse General Structure



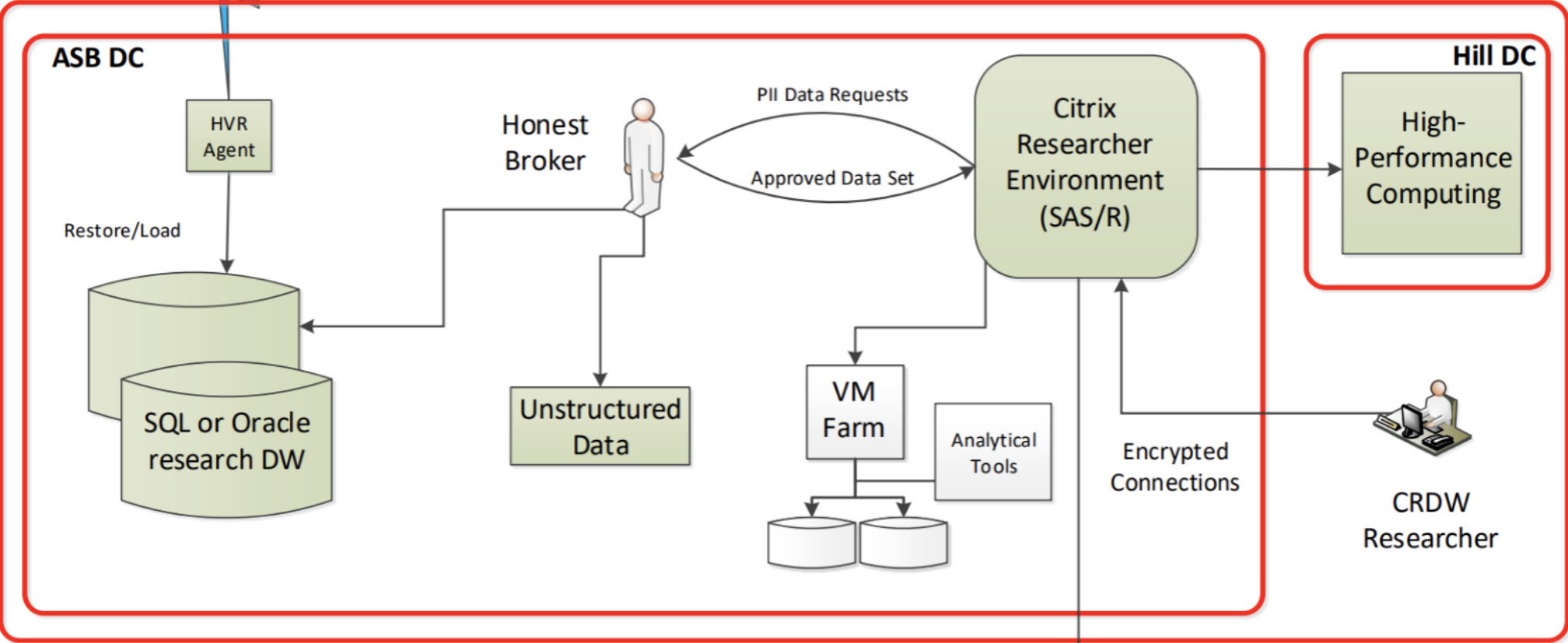
GLOSSARY:
CRM – Customer relationship Management
ERP – Enterprise resource planning
EHR/EMR - electronic Health/Medical Records
ELT – Extract, Load, Transform
OLAP – Online Analytical Processing
CDM – Common Data Model (Clinical Data Mart)

RWJBH / Rutgers – Clinical Research Data Warehouse Architecture



VPN Tunnel

Rutgers Clinical Research Data Warehouse (CRDW)



Reports & Analytics (non PII)
De-identified data



DATA in CRDW

Currently (October 2024) CRDW holds the following data:

1. RWJBH EPIC Clarity database: (2018- current)

Approved Data Types:

These are the data elements that Data Provider will identify in its EPIC record and consistent with the description of same in the EPIC Clarity relational database.

o Access

- Schedule Management • Appointments • Referrals • ADT Events • MyChart • Release of Information • Deficiency tracking • Coding • Identity • Scanning • Kiosk Usage • Registration • Insurance Verification • Bed Planning • Environmental Services

o Clinical

- Patient encounters • Episodes • Diagnosis / Chief complaint / Problem list • Orders and results • Medications • Events • Registries • Alerts • Flowsheets • Research study records

o Revenue Cycle

- Coverages • RVU's

o Data Dictionary

- User Profiles • Procedures • Payors/Plans • Facility Structure • DRG • Visit Types • Report Groupings • System Audit Logs • System Rules

Exclusions:

Data and Sensitive Health Data excludes, and Data Providers will not provide, the following "Psychotherapy notes" (as that term is defined by 45 CFR § 164.501) and specific financial data (such as financial transactions, reimbursement, fee schedules, claims, patient accounts, account receivables, denials, work queues, and/or follow-up).

2. **Additional data** can be accessed through our collaboration with **Rutgers Cancer Institute (CINJ) and Institute for Health (IFH)** (<https://ifhcore.rutgers.edu/data>)

ACCESS REQUIREMENTS

Typically, access is granted to authorized researchers, clinicians, and stakeholders associated with Rutgers University and RWJBH, as well as their partner institutions. However, approval from the CRDW Data Governance Council (DGC) is necessary.

CRDW is a restricted service, implying that researchers or research groups must first get approval from the Institutional Review Board (IRB) for their research projects before applying for CRDW access. We cannot grant access to projects lacking IRB-approved protocols.

SERVICES

CRDW provides two types of data services:

- **Data Access:** Researchers can request access to data already stored in the CRDW warehouse.
- **Data Tenancy:** This is when a researcher needs to transfer their own data into the protected CRDW environment. We currently prioritize projects with clinical and medical data for tenancy requests.

In addition to data access and hosting services, CRDW offers research support ranging from basic analytical consultations and initial assessment of the available clinical data to comprehensive examination of data to identify trends and patterns, data mining and predictive modeling.

All current information about the service can be found on our website, crdw.rutgers.edu

Get started with CRDW

Data access form

Need access to data hosted in CRDW?
Submit your request via the following
form.

[Data request](#)

Tenant request form

Want to host your own data from a
project in CRDW? Complete the tenant
request form.

[Tenant request](#)

Consultation request form

Need a consultation with CRDW staff
scientists? Submit your request through
the following form.

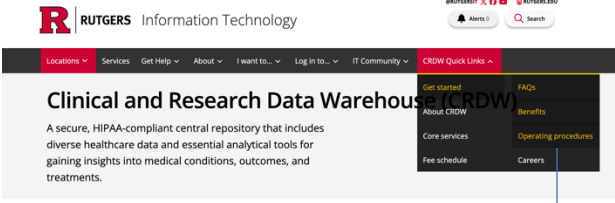
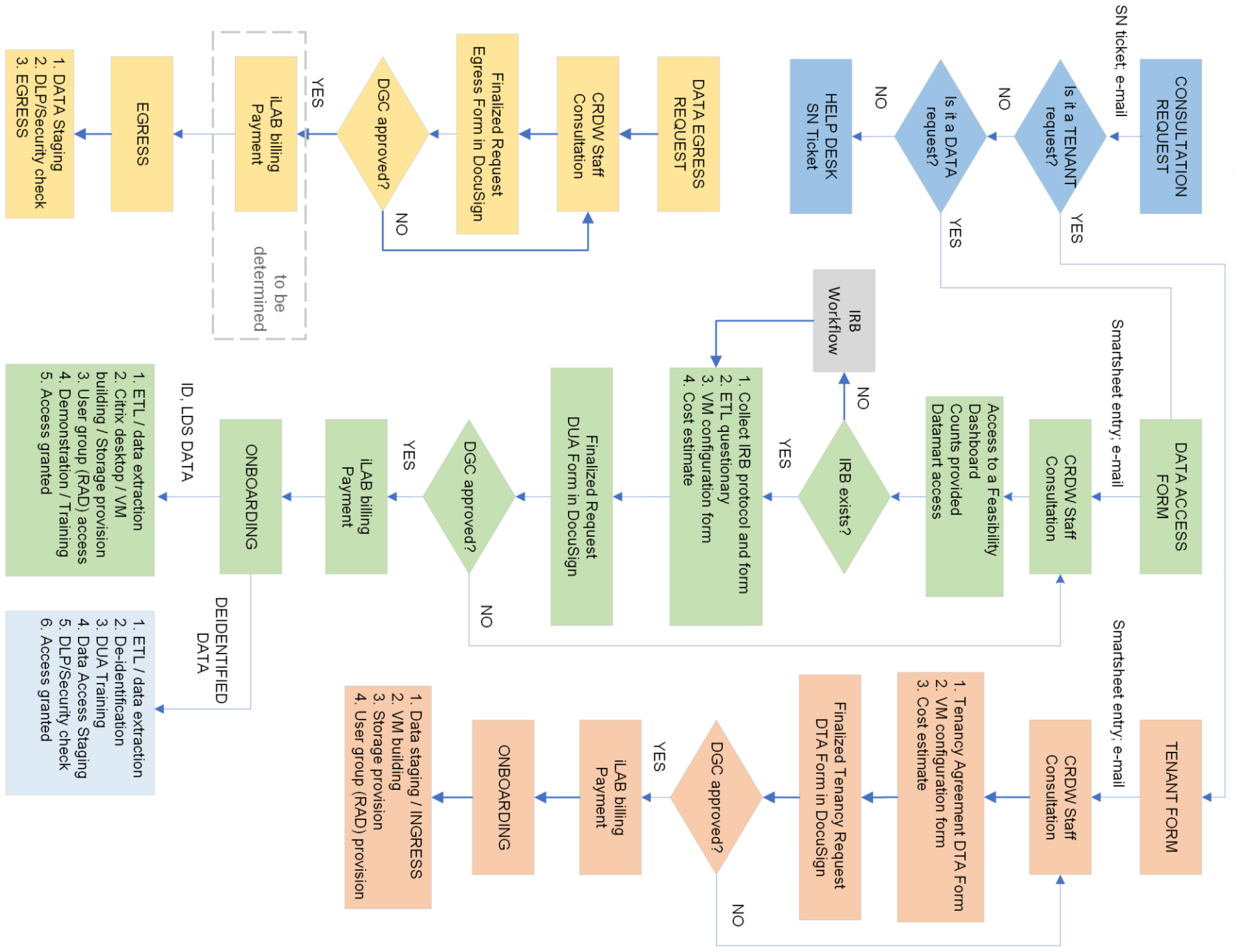
[Consultation request](#)

Contact us

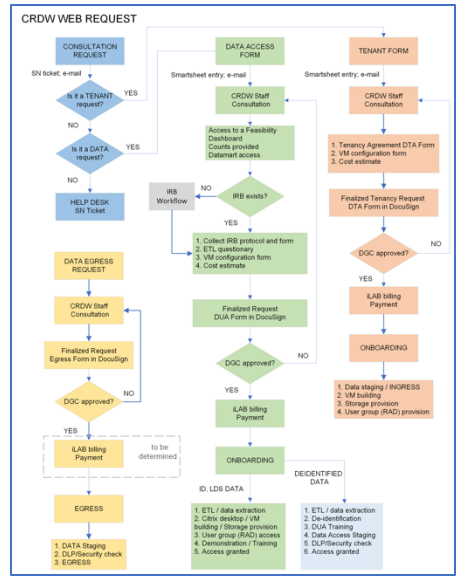
Submit any questions regarding CRDW services by sending an
email to crdw_requests@oarc.rutgers.edu.

Please visit: crdw.rutgers.edu

CRDW WEB REQUEST



What is CRDW?
The Rutgers Clinical and Research Data Warehouse (CRDW) serves as a central repository, uniting diverse healthcare data sources including Electronic Medical Records (EMRs), Clinical Trial Management System (CTMS), Tumor Registries, Biospecimen Repositories, and cutting edge medical imaging and genomics. [Learn more about CRDW.](#)



Fee* Schedule for the Rutgers CRDW services

*Prices subject to change

<https://crdw.rutgers.edu/fee-schedule>

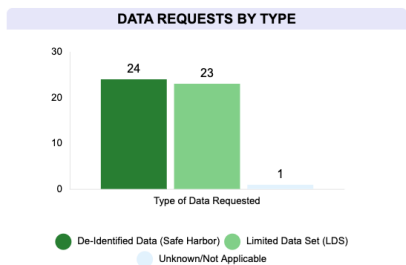
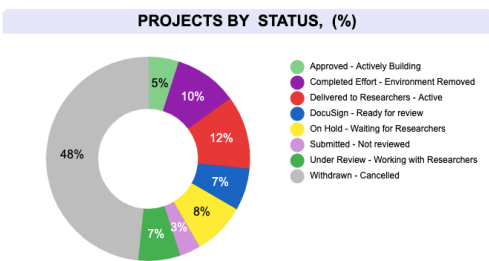
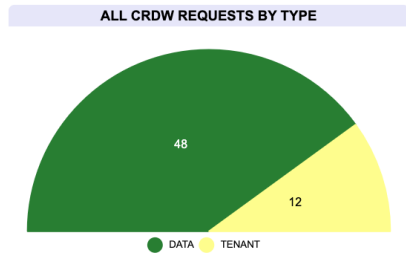
SKU	Description	Rutgers rate	External academic	Private sector	Unit
User support and consulting					
CRDW 1	Pre-consultation (Counts)	\$15	\$30	\$60	Per hour
CRDW 6	Data analytic services	\$90	\$180	\$360	Per hour
CRDW 7	Consultation	\$50	\$100	\$200	Per hour
Data extraction (ETL)					
CRDW 4	Data extraction (ETL)	\$45	\$90	\$180	Per hour
CRDW 5	Data transfer (ingestion/egress)	\$75	\$150	\$300	Per hour
Computation					
VM 1	Standard virtual machine	\$0.02	\$0.04	\$0.08	Per hour
BM 1	Bare metal CPU-only node	\$0.44	\$0.88	\$1.76	Per hour
Data storage					
CRDW 8	Encrypted object storage (Isilon)	\$6.25	\$12.50	\$25.00	1TB/month
CRDW 10	Encrypted block storage	\$16.70	\$33.40	\$66.80	1TB/month

CRDW Projects Dashboard

go.rutgers.edu/crdw-dashboard

CRDW REQUESTS BY STATUS

Primary	Working Status	Count
Total		60
Working Status		
Approved - Actively Building		3
Working Status		
Completed Effort - Environment Removed		6
Working Status		
Delivered to Researchers - Active		7
Working Status		
DocuSign - Ready for review		4
Working Status		
On Hold - Waiting for Researchers		5
Working Status		
Submitted - Not reviewed		2
Working Status		

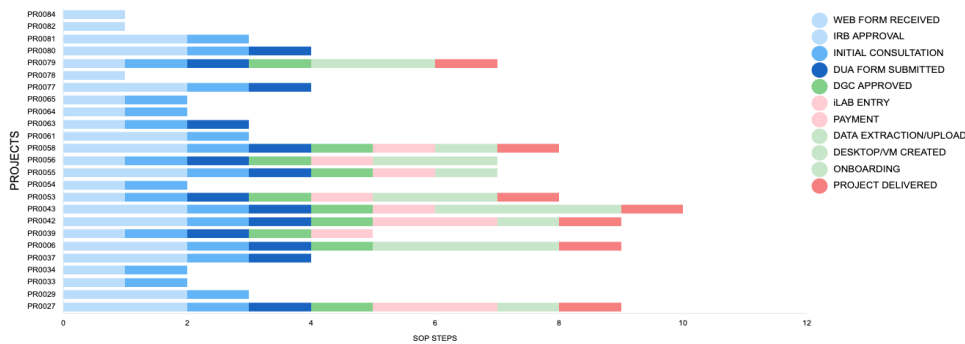


PROJECTS DETAILS (ACTIVE & PLANNED ONLY)

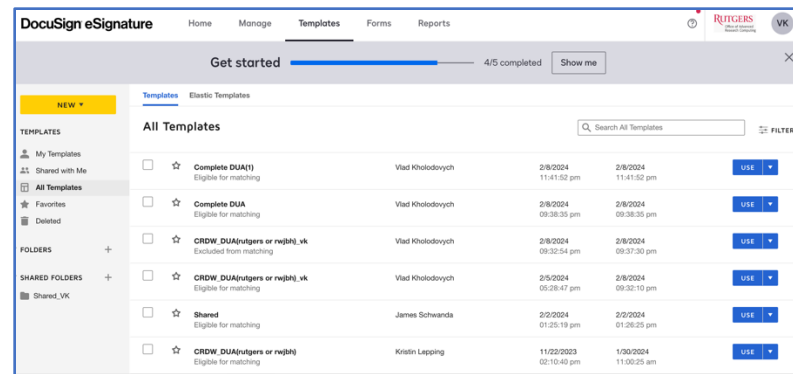
o - in progress o - attention needed o - waiting for approval o - not reviewed

Project ID	REQUEST TYPE	PI Name	Project Title	BH data	Working Status	Created	Institution Name	Assigned to
Total	Count	25						
DATA TYPE	Count	9						
PR0082	DATA	David Awad	The True Impact of Early Tocilizumab for Cytokine Release		Submitted - Not reviewed	09/18/24	Robert Wood Johnson New Brunswick	Mark Baker
PR0077	DATA	Branimir Ljubic	Computer Science Analytics and Modeling of Hypertension		DocuSign - Ready for review	08/21/24	OARC, Rutgers	Branimir Ljubic
PR0065	DATA	Noelle Foster	Core data for clinical informatics students - depression		Under Review - Working with R	07/30/24	NJACTS	Sailaja Balaramamaha

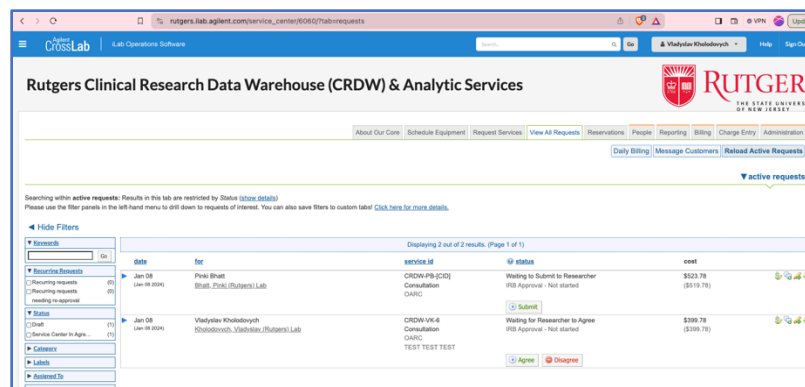
CRDW PROJECTS PROGRESS REPORT (ONLY ACTIVE)



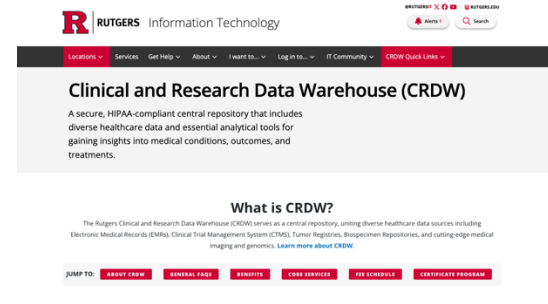
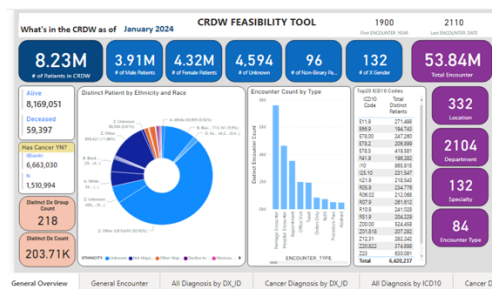
DocuSign optimized workflows and procedures



Functional iLab portal for CRDW Core facilities

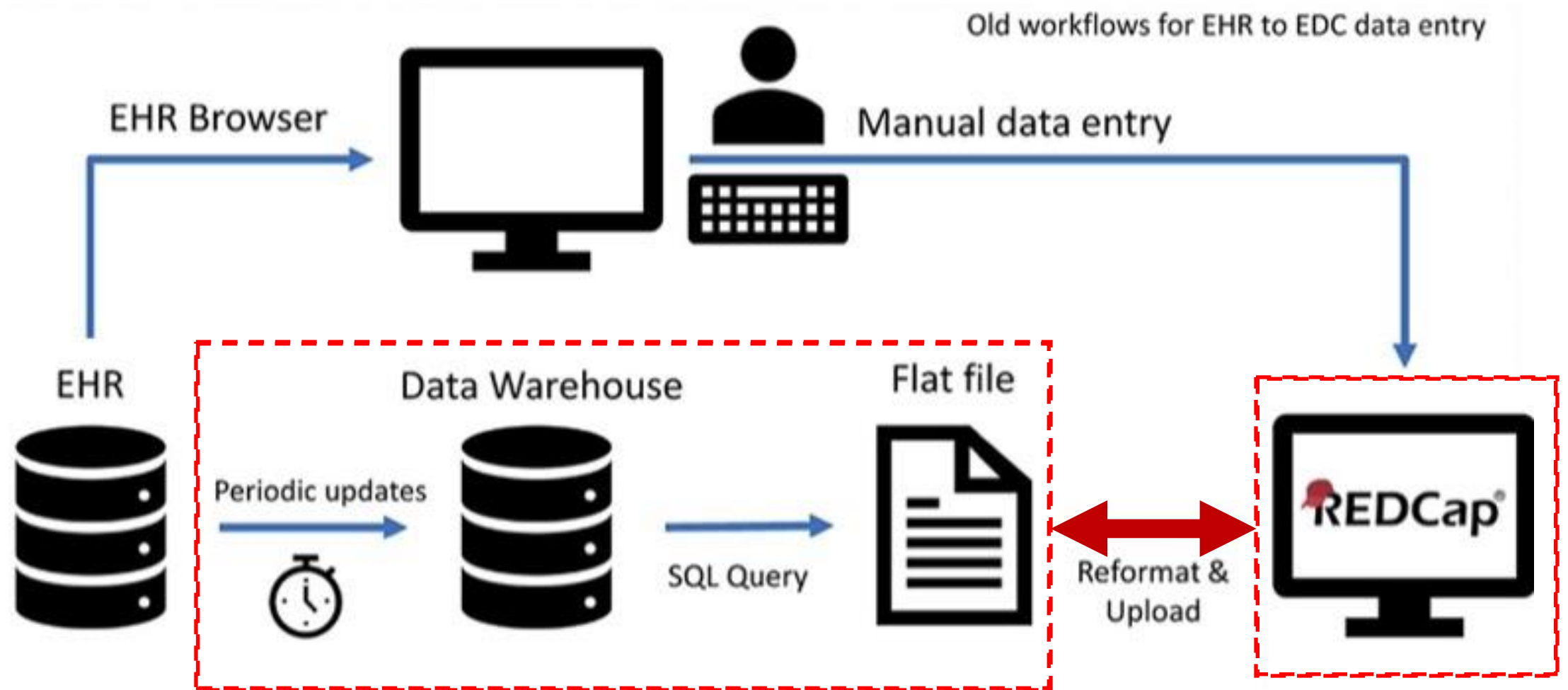


Updated web portal and feasibility dashboards

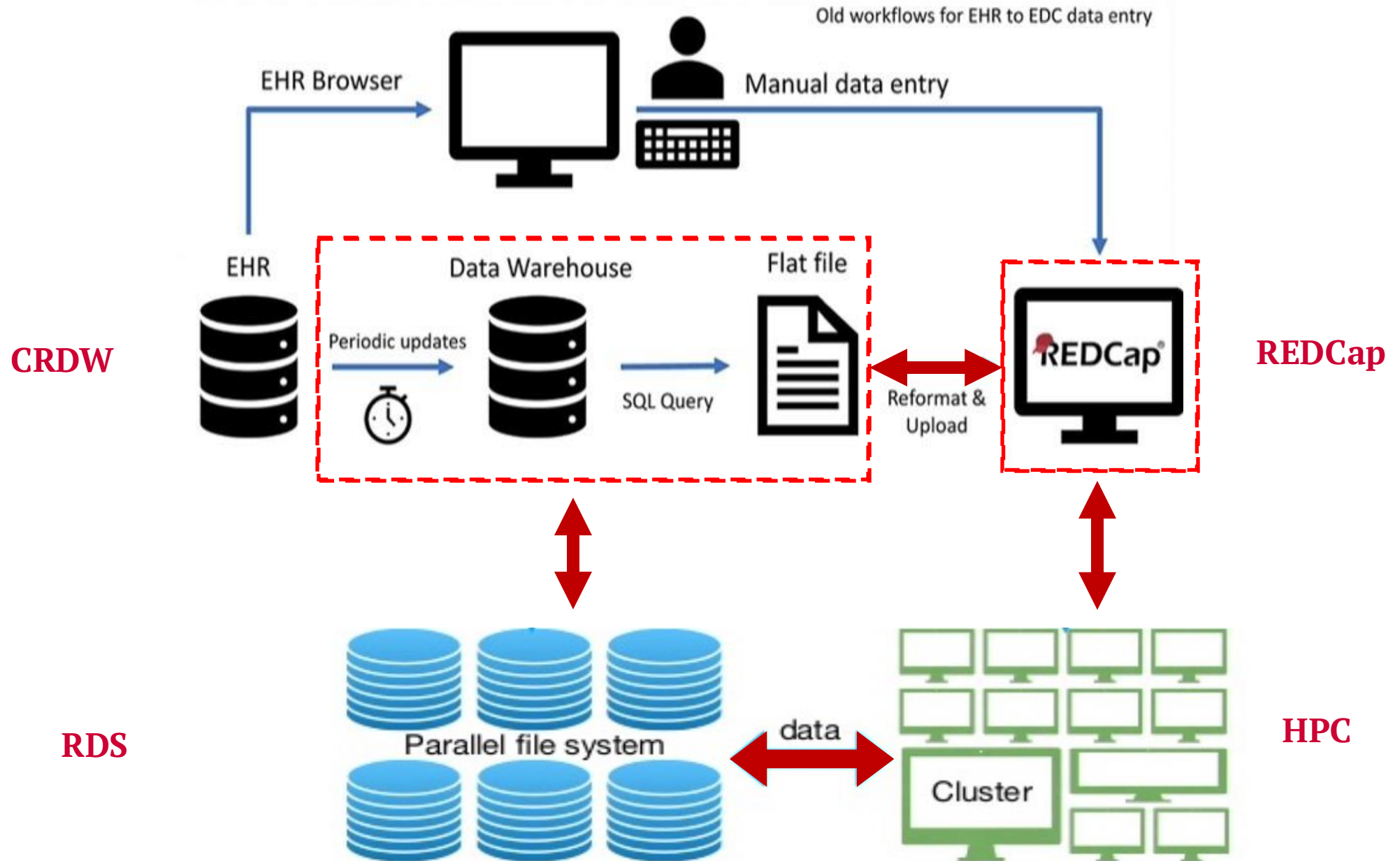


Since the official launch of the CRDW in January 2024, CRDW team processed and delivered 20 projects to researchers, a remarkable increase from only two projects delivered over the previous four years.

FUTURE WORK



SEAMLESS WORKFLOW



THANK YOU

**We look forward to having you
as our clients and partners at OARC**

oarc.rutgers.edu

REDCap: it.rutgers.edu/redcap

CRDW: crdw.rutgers.edu