Florida Cancer Research (CARES) Network & PAC3R Platform Data Access and Sharing Policy

Version 0.2
October 19, 2024

Table of Contents

Introduction

This Data Sharing and Access Policy document represents a foundational framework for the Bankhead-Coley Florida Cancer Research (FL CARES) Network Infrastructure to delineate the rules and procedures that govern data sharing, access, and management in the Platform for Accelerating Collaborative Computational Cancer Research (PAC3R, pronounced “pacer”). These policies support a secure, transparent, and compliant data ecosystem following the guidance of FAIR Principles in collaboration among the Florida CARES Network’s institutions, with a dynamic digital landscape essential for the automatic discovery of datasets and tools.

Objectives

The goal is to drive collaboration and innovation while safeguarding the privacy and security within PAC3R.

The data access and sharing policy described in this document is based on a culture of transparency, fortifying security measures, and ensuring compliance with pertinent regulations,

as we increasingly rely on data to inform decision-making, drive innovation, and facilitate collaboration. The data shared, accessed, and managed within the Florida CARES Network organization or in external collaborations will:

- Promote responsible data management: Encourage the responsible use, access, and sharing of data to maintain its integrity, security, and accuracy.
- Ensure ethical handling of sensitive information: Establish best practice guidelines for handling sensitive, confidential, private data.
- Facilitate collaboration and information exchange: The Platform for Accelerating Collaborative Computational Cancer Research (PAC3R) provides a framework that enables seamless collaboration and information exchange, promoting innovation and knowledge sharing.
- Comply with relevant regulations and standards: Ensure alignment with applicable legal requirements, regulations, and standards, as well as data protection, privacy, and security.

Regular reviews and updates will ensure the access and classification level of access to data and tools according to the security measures, access controls, encryption, and audit trails implemented to the sensitivity level. We will implement training programs to ensure individual awareness of and adherence to data handling requirements within the PAC3R framework of policies.

Definitions of Abbreviations

FL CARES: Bankhead-Coley Florida Cancer Research Network Infrastructure.

PAC3R: Platform for Accelerating Collaborative Computational Cancer Research.
Pronounce “pacer”.

PII: Personally Identifiable Information.

API: Application Programming Interface.

RBAC: Role - Based Access Control.

ReBAC: Relationship-Based Access Control (ReBAC).

PAC3R Access Control

FL CARES will follow data security best practices for access control into PAC3R, through secured sharing of data and for their specific roles.

Relationship-Based Access Control (ReBAC)

Relationship-Based Access Control (ReBAC) manages access to PAC3R in a structured and efficient way. It controls the functionalities and data a user can access via the Application Programming Interface (API) based on their predefined Role.

Fine-Grained Authorization with SpiceDB

SpiceDB is an open-source, globally distributed database system designed for Fine-Grained Authorization (FGA), inspired by Google's Zanzibar. It is primarily a Relationship-Based Access Control (ReBAC) engine that models access as a graph of relationships between users and resources.

Core Concepts

- Objects: The protected resources (e.g., a document, a dataset, a user).
- Relationships: The connections between objects (e.g., "user:alice is a reader of dataset:cancer-data").
- Permissions: Rules defined in a schema that dictate what actions a subject can perform on an object by traversing the relationship graph (e.g., the `edit` permission is granted to anyone who is a writer of the document).

Role	Can Be	Responsibilities	Key Permissions
Administrator	Site administrator account.	Has overall oversight for user and dataset (endpoint) administration within PAC3R.	Assign/Revoke roles for any user. Assign/Revoke access to datasets (endpoints).
Member (Contributor/Collaborator)	Researcher or data analyst with an account.	Generates data, conducts research, analyzes data, derives insights, and generates reports.	Access based on specific functions (expertise). Analyze/Interpret data per endpoints. Access to specific datasets (endpoints/samples). Contributes to research projects.
Guest	Collaborator with an account.	Can access tools to create signatures and can download publicly available data.	Limited, often read-only, access to specific datasets and publicly shared research projects. Access to specific data/tools based on the nature of the collaboration.

Data Protection Classification

FL CARES data classification for PAC3R is a critical aspect of ensuring responsible handling and protection of sensitive information for next-generation cancer sequencing and research. The classification categorizes the data based on sensitivity, criticality, and specificity of research impact. Cancer research discovery fosters collaboration and innovation toward safeguarding the privacy and security of patient data and genomic information. Regular reviews and updates shall be conducted to adapt these guidelines to changes in regulation, technological advances, and the evolution of cancer research. The data classification process takes into account the type of information, and the data lifecycle, from collection and storage to sharing and disposal.

Public data

Publicly accessible data that do not contain confidential, sensitive, or PII.

Research data governed by a consortium or research group

Data intended for internal use within PAC3R, which may include non-sensitive research data and aggregated statistics.

Confidential data governed by institutional policies

Critical data that require protection and access restrictions, governed by institutional agreements. For example: Patient data involving severe medical conditions, proprietary research findings, and any data subject to legal or confidentiality agreements.

Regulated data governed by state of federal law

Sensitive information such as personally identifiable information (PII) or information that could be used to identify individuals. For example: Patient medical records, genomic data, and any data that could lead to the identification of individuals.
Highly sensitive data that require strict access controls and additional security measures. For example: genetic information, clinical information, experimental results, and any data that pose a high risk if accessed by unauthorized individuals.

In a ReBAC model, access is granted by establishing a relationship (e.g., *admin*, *owner*, *reader*) between a Subject (User/Group) and a Resource (System/Dataset).

- Only Subjects with the admin relationship to the system configuration resource can modify it.
- Only Subjects with the administrator relationship to the PAC3R system can grant the *initial access relationship* to other users.

Data Sensitivity and Relationship-Based Access Control (ReBAC)

ReBAC is integrated with data classification by using the sensitivity level as an attribute or a type of resource. Access is granted only if the user has a valid relationship to the data and the permissions defined in the schema meet the security requirements for that sensitivity level.

Data Ownership

Datasets are unambiguously associated with a Subject who has the owner relationship to the data. This relationship grants complete control.

- The researcher who uploaded the data, or their Institution, is assigned the owner relationship to the dataset.

Data Segmentation

Datasets are segregated based on sensitivity and research project requirements. A user's access depends on their relationship to the data segment (e.g., "user:alice is a member of project:lung-cancer-study" which grants them read permission on the associated sensitive-dataset:x).

Selective Data Sharing

This is the act of creating a specific relationship (e.g., viewer, editor) between a designated individual or group and a specific dataset, while withholding a relationship for other data.

Collaborative Data Sharing

Sharing data is achieved by creating temporary or project-scoped relationships with external parties (researchers, partners, or other organizations).

- This allows researchers to create relationships (e.g., collaborator) to share specific datasets or findings within approved limits.
- Access must be requested and is granted by establishing a time-bound relationship to specific datasets for collaborative purposes.

Federated Data Sharing

A model where data are distributed across different locations or organizations. Access is granted in a coordinated and controlled manner, typically by evaluating a user's relationships across the federated infrastructure. The authorization engine checks the user's relationship to the data's organization, project, and security domain to compute final access rights.

Training

Training will be provided for users on the roles and associated responsibilities to ensure the understanding of the access scope.

Audit Trail

The FL CARES Network team and PAC3R development team will maintain an audit trail per role assignment, and will monitor compliance.

Review and Revision

The implementation of RBAC in PAC3R streamlines access management, ensures that individuals have access to only the necessary content for their roles, and enhances security and overall administration.

Incident Response

Regular reviews of this Data Access and Sharing Policy will incorporate feedback, address emerging challenges, and ensure alignment with evolving regulatory requirements. Clearly defined reporting mechanisms will adhere to the relevant obligations.

Document Version Control

Version	Sign Off	Comments
Version 0.0.10 March 28, 2024	C.Chung	Incorporated edits and comments.
Version 0.1 December 17, 2024	S. Schurer C. Obregon M. Pilarczyk D. Vidovic L. Rupprecht M. Sinclair V. Stathias C. Chung	Review of roles
Version 0.2 October 19, 2024	S. Schurer C. Obregon M. Pilarczyk L. Rupprecht M. Sinclair V. Stathias F. Sotolongo C. Chung	Review ReBAC