impact

What is ImPACT?

Ilya Baldin RENCI/UNC Chapel Hill

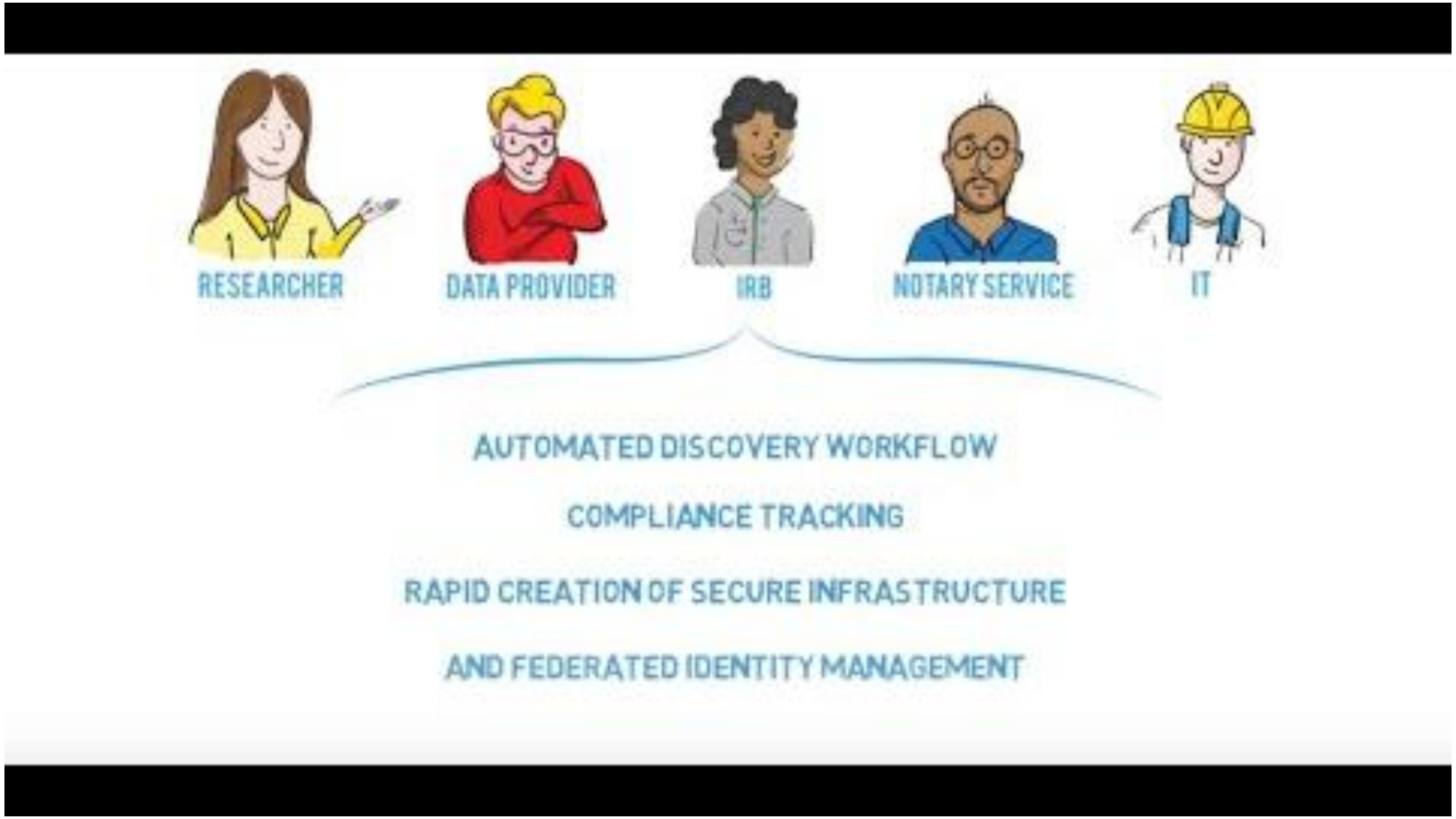
ibaldin@renci.org

This project is sponsored by the National Science Foundation grant OAC-1659367

Project Overview

- Challenges:
  - Social science and many other data-oriented disciplines depend on data belonging to multiple stakeholders
  - Governed by a variety of use policies
  - Multi-institutional research requires cooperative analysis
  - Need to satisfy the privacy concerns of the owners while producing interesting research outcomes by analyzing data
- Goal: to enable cooperative processing across the stakeholder-owned datasets, while respecting the privacy policies of the individual owners, and to provide a model for collaboration that could be readily used by other institutions.

What Is ImPACT?



AUTOMATED DISCOVERY WORKFLOW

COMPLIANCE TRACKING

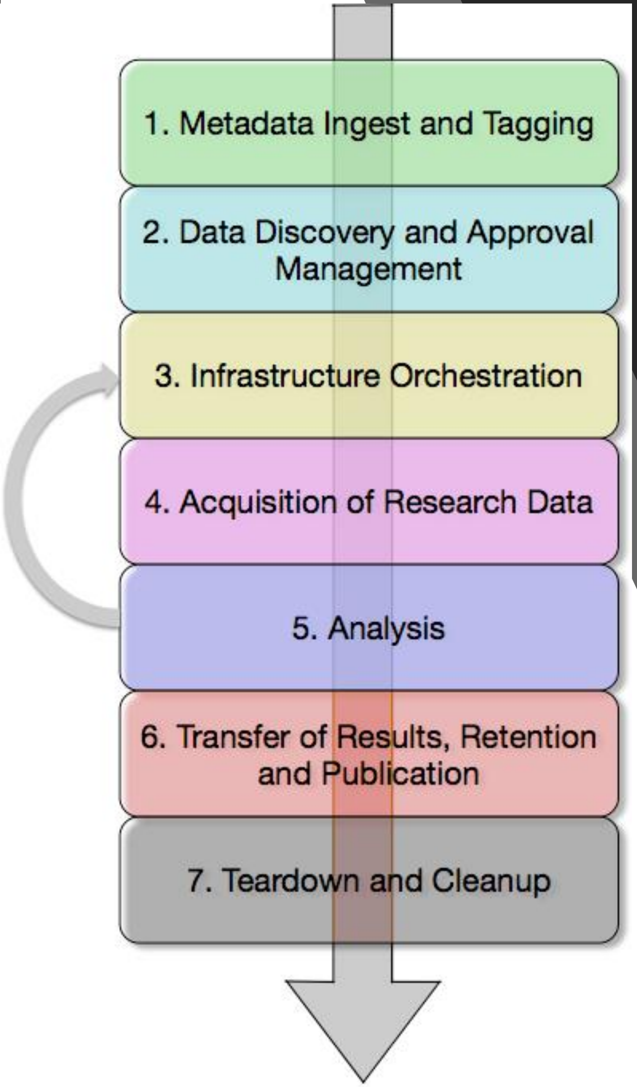
RAPID CREATION OF SECURE INFRASTRUCTURE

AND FEDERATED IDENTITY MANAGEMENT

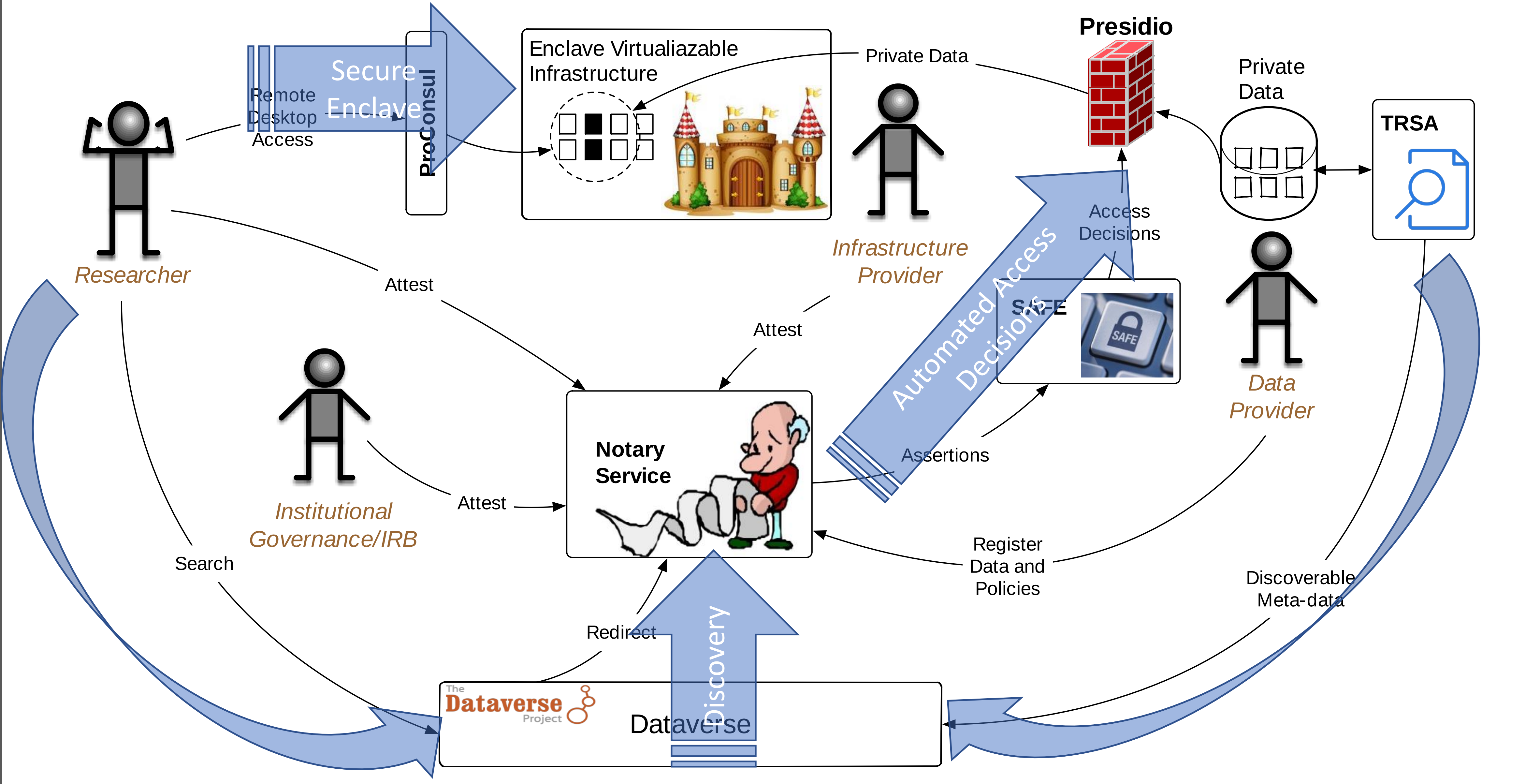
Why this architecture

- Two models by which the community evolves appear viable
  - “Aggregator model” – large institutions act as aggregators and gatekeepers of important datasets
  - “Distributed model” – multiple data owners and gatekeepers
- Aggregator model has significant value
  - We should still assume there will always be multiple owners of data with different access policies
- Need a way to
  - Make private data discoverable
  - Specify policies that control access to data
  - Automate to the extent possible the process of satisfying those policies
  - Automate and provide verification of data access based on those policies
  - Factor out policy storage and policy control so that distributed operation becomes possible
  - Leverage inter-institutional authentication and authorization mechanisms

The Approach


1. Metadata Ingest and Tagging
2. Data Discovery and Approval Management
3. Infrastructure Orchestration
4. Acquisition of Research Data
5. Analysis
6. Transfer of Results, Retention and Publication
7. Teardown and Cleanup

- Provide a suite of solutions that are designed to work <u>together,</u> but can also be <u>leveraged independently</u>
- Solutions:
  a) Make data discoverable in Dataverse, without compromising its privacy
  b) Create <u>repeatable</u> secure infrastructure for performing collaborative analyses on privacy restricted data
  c) Automate decisions about data access and link to DUA (Data Use Agreement) approval process



Secure Enclave

Remote Desktop Access

ProConsul

Enclave Virtualiazable Infrastructure

Private Data

Presidio

Private Data

TRSA

Researcher

Infrastructure Provider

Access Decisions

Automated Access Decisions

SAFE

Attest

Attest

Notary Service

Institutional Governance/IRB

Attest

Assertions

Data Provider

Register Data and Policies

Search

Discoverable Meta-data

Redirect

Discovery

Dataverse

TRSA (Trusted Remote Storage Agent)

- Makes private data discoverable via Dataverse
- Owned by data provider
- Harvests and sends only the metadata to selected Dataverse instances
- Uses Dataverse API

ProConsul (Protected Data Enclave)

Provides federated, web-based login to enclave VMs

Granular access control

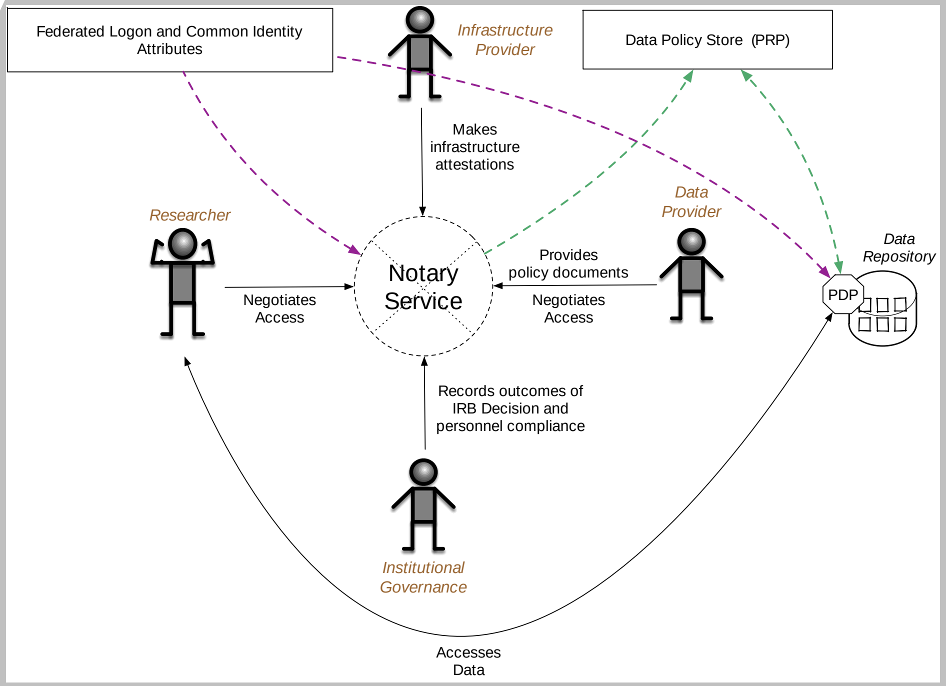
OS-independent (guest and host)

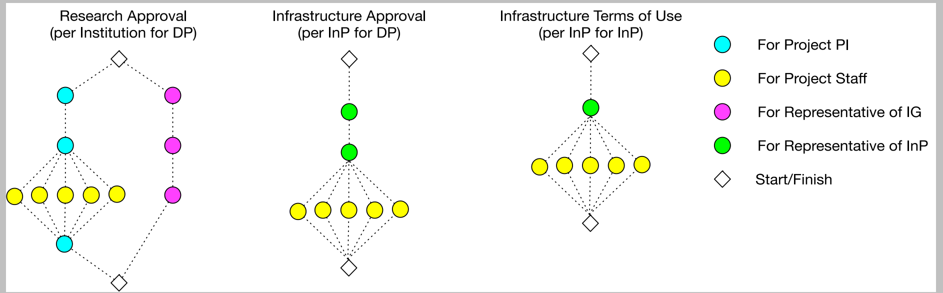
Presidio (Policy Enforcement)

- Guards access to private data hosted by data provider
- Supports complex access policies
- Validates that assertions made by various stakeholders (researchers, IRB, etc) satisfy access policies in automated fashion
- Browser- and API-based access

Notary Service

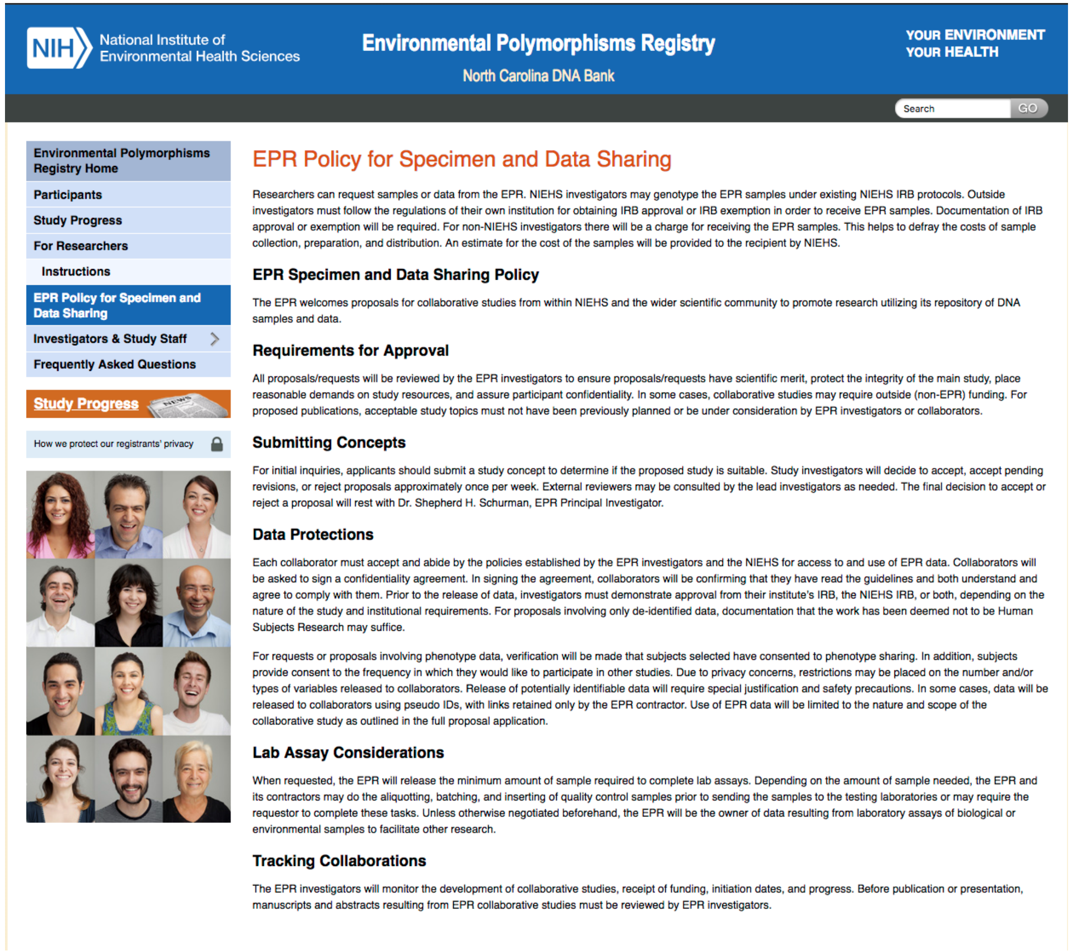
- Mediator of interactions between various stakeholders
- Digital trail of promises and attestations needed to satisfy data access policies (DUAs)
- Support for flexible document workflow and individual forms
- Extensive logging for statistical and audit purposes





Research Approval (per Institution for DP)
Infrastructure Approval (per InP for DP)
Infrastructure Terms of Use (per InP for InP)
● For Project PI
● For Project Staff
● For Representative of IG
● For Representative of InP
◇ Start/Finish

Policy conversion


NIH National Institute of Environmental Health Sciences
Environmental Polymorphisms Registry
North Carolina DNA Bank
YOUR ENVIRONMENT YOUR HEALTH
Search
GO
Environmental Polymorphisms Registry Home
Participants
Study Progress
For Researchers
Instructions
EPR Policy for Specimen and Data Sharing
Investigators & Study Staff
Frequently Asked Questions
Study Progress
How we protect our registrants' privacy

EPR Policy for Specimen and Data Sharing

Researchers can request samples or data from the EPR. NIEHS investigators may genotype the EPR samples under existing NIEHS IRB protocols. Outside investigators must follow the regulations of their own institution for obtaining IRB approval or IRB exemption in order to receive EPR samples. Documentation of IRB approval or exemption will be required. For non-NIEHS investigators there will be a charge for receiving the EPR samples. This helps to defray the costs of sample collection, preparation, and distribution. An estimate for the cost of the samples will be provided to the recipient by NIEHS.

EPR Specimen and Data Sharing Policy

The EPR welcomes proposals for collaborative studies from within NIEHS and the wider scientific community to promote research utilizing its repository of DNA samples and data.

Requirements for Approval

All proposals/requests will be reviewed by the EPR investigators to ensure proposals/requests have scientific merit, protect the integrity of the main study, place reasonable demands on study resources, and assure participant confidentiality. In some cases, collaborative studies may require outside (non-EPR) funding. For proposed publications, acceptable study topics must not have been previously planned or be under consideration by EPR investigators or collaborators.

Submitting Concepts

For initial inquiries, applicants should submit a study concept to determine if the proposed study is suitable. Study investigators will decide to accept, accept pending revisions, or reject proposals approximately once per week. External reviewers may be consulted by the lead investigators as needed. The final decision to accept or reject a proposal will rest with Dr. Shepherd H. Schurman, EPR Principal Investigator.

Data Protections

Each collaborator must accept and abide by the policies established by the EPR investigators and the NIEHS for access to and use of EPR data. Collaborators will be asked to sign a confidentiality agreement. In signing the agreement, collaborators will be confirming that they have read the guidelines and both understand and agree to comply with them. Prior to the release of data, investigators must demonstrate approval from their institute's IRB, the NIEHS IRB, or both, depending on the nature of the study and institutional requirements. For proposals involving only de-identified data, documentation that the work has been deemed not to be Human Subjects Research may suffice.

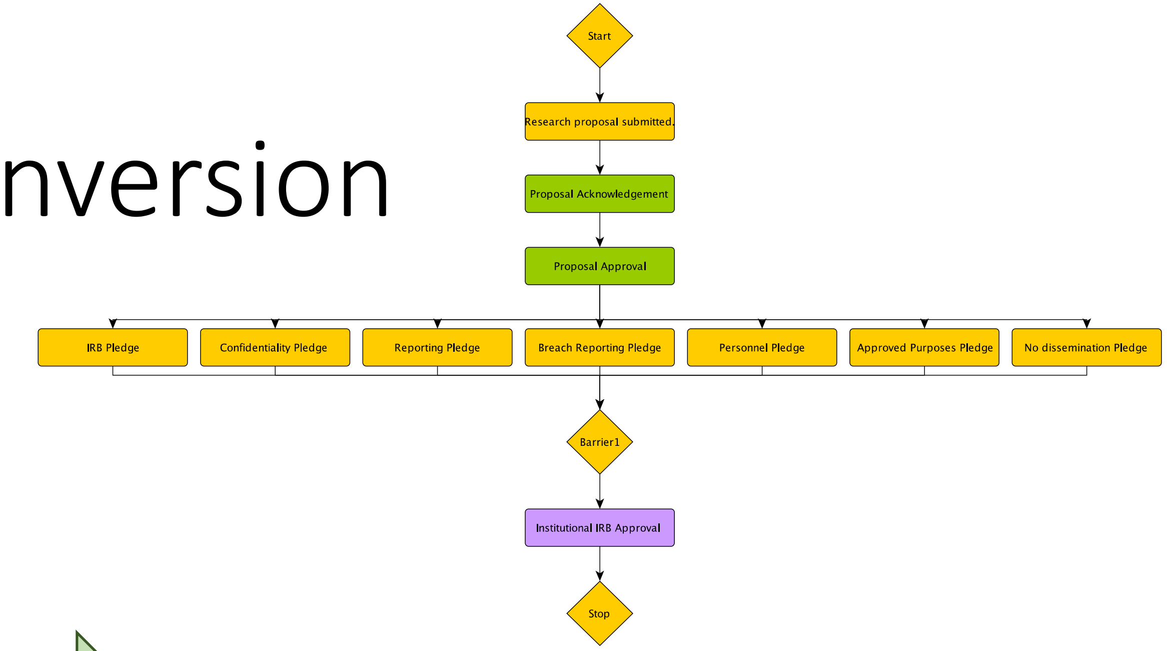
For requests or proposals involving phenotype data, verification will be made that subjects selected have consented to phenotype sharing. In addition, subjects provide consent to the frequency in which they would like to participate in other studies. Due to privacy concerns, restrictions may be placed on the number and/or types of variables released to collaborators. Release of potentially identifiable data will require special justification and safety precautions. In some cases, data will be released to collaborators using pseudo IDs, with links retained only by the EPR contractor. Use of EPR data will be limited to the nature and scope of the collaborative study as outlined in the full proposal application.

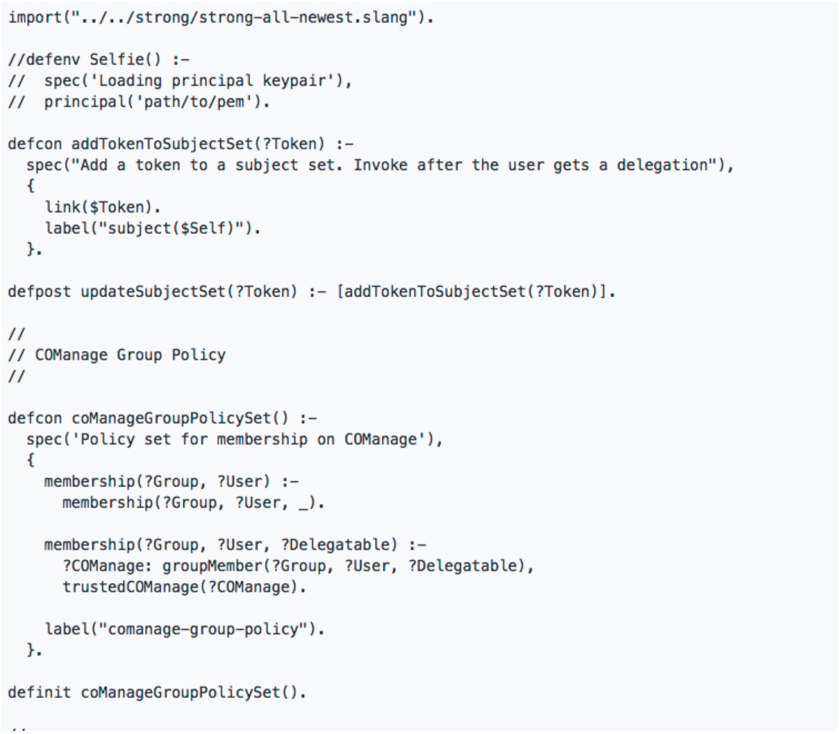
Lab Assay Considerations

When requested, the EPR will release the minimum amount of sample required to complete lab assays. Depending on the amount of sample needed, the EPR and its contractors may do the aliquotting, batching, and inserting of quality control samples prior to sending the samples to the testing laboratories or may require the requestor to complete these tasks. Unless otherwise negotiated beforehand, the EPR will be the owner of data resulting from laboratory assays of biological or environmental samples to facilitate other research.

Tracking Collaborations

The EPR investigators will monitor the development of collaborative studies, receipt of funding, initiation dates, and progress. Before publication or presentation, manuscripts and abstracts resulting from EPR collaborative studies must be reviewed by EPR investigators.

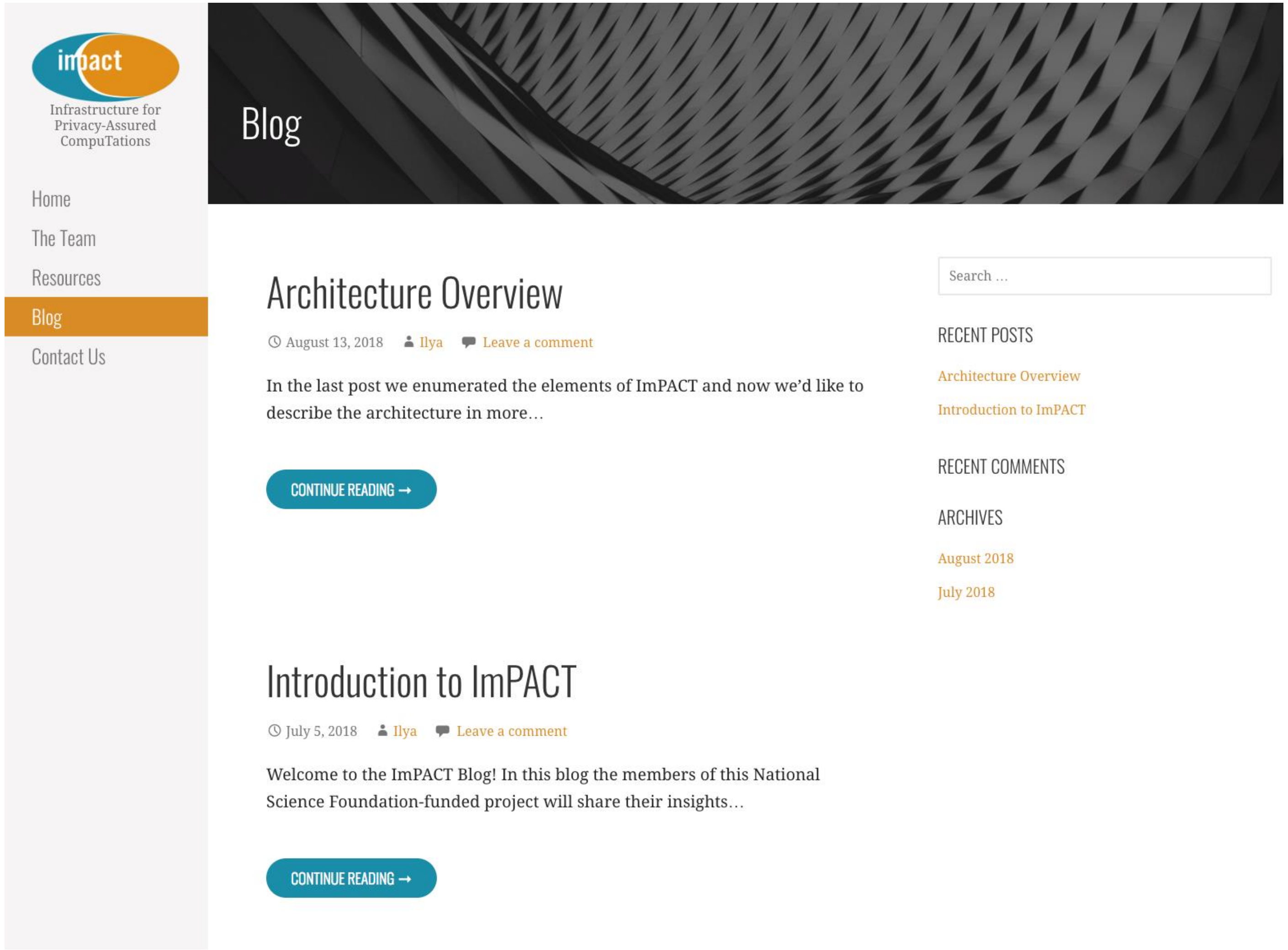




```
import("../../strong/strong-all-newest.slang").

//defenv Selfie() :-
//  spec('Loading principal keypair'),
//  principal('path/to/pem').

defcon addTokenToSubjectSet(?Token) :-
  spec("Add a token to a subject set. Invoke after the user gets a delegation"),
  {
    link($Token).
    label("subject($Self)").
  }.

defpost updateSubjectSet(?Token) :- [addTokenToSubjectSet(?Token)].

//
// COManage Group Policy
//

defcon coManageGroupPolicySet() :-
  spec('Policy set for membership on COManage'),
  {
    membership(?Group, ?User) :-
      membership(?Group, ?User, _).

    membership(?Group, ?User, ?Delegatable) :-
      ?COManage: groupMember(?Group, ?User, ?Delegatable),
      trustedCOManage(?COManage).

    label("comanage-group-policy").
  }.

definit coManageGroupPolicySet().

//
```

More Information

http://cyberimpact.us



Blog

Architecture Overview

August 13, 2018 · Ilya · Leave a comment

In the last post we enumerated the elements of ImPACT and now we'd like to describe the architecture in more…

Introduction to ImPACT

July 5, 2018 · Ilya · Leave a comment

Welcome to the ImPACT Blog! In this blog the members of this National Science Foundation-funded project will share their insights…

RECENT POSTS

Architecture Overview

Introduction to ImPACT

RECENT COMMENTS

ARCHIVES

August 2018

July 2018