GEN3 DATA COMMONS

Security and Compliance

Gen3 Community Forum
July 15, 2026

THE UNIVERSITY OF CHICAGO CENTER FOR TRANSLATIONAL DATA SCIENCE

OHSU

OCC

Australian BioCommons

Waipapa Taumata Rau University of Auckland

The Agenda

- **Navigating compliance frameworks in the US - FedRAMP moderate as a use case** - Clint Malson, University of Chicago
- **Advancing Childhood Cancer Research Securely: Deploying and Fortifying the NL-4C Gen3 Portal** - Jasper van Dalum, Princess Máxima Center for Pediatric Oncology
- **Compliance-as-code in the modern platform** - Colin Griffin, Krumware
- **Lessons on ISO27001 for Data Commons** - Nakul Deshmukh, Australian BioCommons

Community Forum Lecture

Securing the Gen3 Commons

5 Years of FedRAMP, False Positives, and Arguing with Auditors

Building secure, scalable open-source solutions and sustaining robust baselines under FedRAMP.

GEN3 DATA COMMONS

FedRAMP

The Compliance Journey: Paradox & Perseverance

GEN3 DATA COMMONS

The Gen3 Paradox

CTDS operate in a weird, high-friction space. We are a global, cloud-native SaaS provider, but we are also a scientific research center. The mission is public discovery; the federal mandate is total lockdown.

The 5-Year Survivor Badge

We've sustained FedRAMP Moderate for five years now. We didn't buy our way out of it with multi-million dollar proprietary tools. We built it on a massive stack of open-source and cloud-native tooling.

Shedding the Inertia

The initial uplift is brutal (hardening images, applying STIGs, defining FIPS). But once baked into the pipeline, the inertia sheds and becomes a standard operational rhythm.

FedRAMP: Hardening the Community Baseline

Community Security Uplift

When we solve a FedRAMP requirement, whether it's baking FIPS-validated libraries into public images or selecting AL2023 for its track record of backports, the entire community inherits that baseline or additional security control. We often take the compliance friction first so you get the enterprise path more easily—many of these controls also map out to ISO 27001, CMMC, and others.

The Baseline Reality

Our federal mandate actively improves Gen3's default security posture, as does our drive to ensure security beyond just security for audit's sake; we care deeply about building secure, scalable solutions that go beyond just the federal mandate. We push security left in the pipeline, so the open-source code you pull is continuously becoming more secure and resilient.

The Audit Reality Check

GEN3 DATA COMMONS

The Appendix Q Grind

You have to "walk the diagram." Every connection, line, and function on that Data Flow Diagram must map 1:1 to a record in your Appendix Q Cryptographic Modules Table. Starting with threat modeling and understanding hop interactions makes this simpler.

The Secret to Record-Time Audits

Achieved through compliance as code, a baked-in security culture, and the ability to quickly defend the mission goal and explain why things are the way they are.

Defending the Mission: How to Argue with Auditors

GEN3 DATA COMMONS

Never Blindly Say "No"

Checklist compliance can kill the science. Protect access while satisfying the mandate.

Threat Modeling > Checklists

How we defend keeping GraphQL introspection turned on and maintaining accessible data streams.

Creative Compliance

Rely on deep code context and layered security-in-depth to defend your architecture.

Meeting Engineers Where They Live

GEN3 DATA COMMONS

Dumb Scanners Destroy Velocity

Let's talk about Amazon Linux 2023. AL2023 backports patches constantly. If your scanner is just blindly looking for package names, your dashboards will claim thousands of vulnerabilities that aren't real.

Killing Alert Fatigue

We trained our scanners to look deeper at packages, not just names. True security requires deep context—knowing if a vulnerable function is *actually* executed in your environment.

Intelligent Routing

Don't throw raw scanner data over the fence. We auto-generate context-rich tickets and route them to the exact Jira/GitHub queues where the owners of that code actually live.

Stop Duplicating the Pain

GEN3 DATA COMMONS

Easing the Compliance Burden

Those of you moving forward with FISMA, FedRAMP, or other controls should know that we are actively working on a way to get a list of false positives from us, so you can hand those to auditors directly.

Maintain the Mainline Posture

It remains crucial to ensure your scanners understand the context of the OS. We want to prevent you from having to fork from us—taking our latest patches and releases is the safest way to keep your environment up to date.

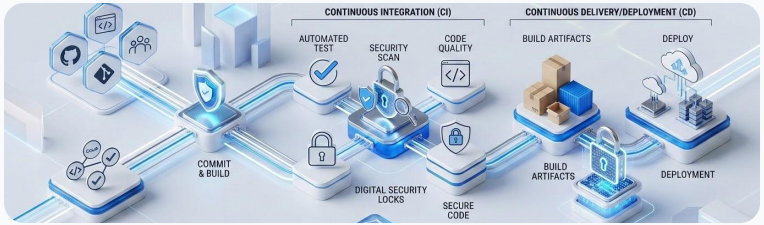
Looking to FedRAMP 20x

GEN3 DATA COMMONS

Risk Paradigm

The New PAIN Index

FedRAMP 20x and CR26 are changing the game. The priority is shifting from flat CVSS scores to active risk and continuous validation.



Pipeline Integration

Setting the Stage

Because we've baked all of this into our pipelines, we are ready for it. This crucial automation integrates compliance as code directly into all aspects of your development workflow.

Advancing Childhood Cancer Research Securely: Deploying and Fortifying the NL-4C Gen3 Portal

15/07/2026

Jasper van Dalum

j.b.vandalum@prinsesmaximacentrum.nl

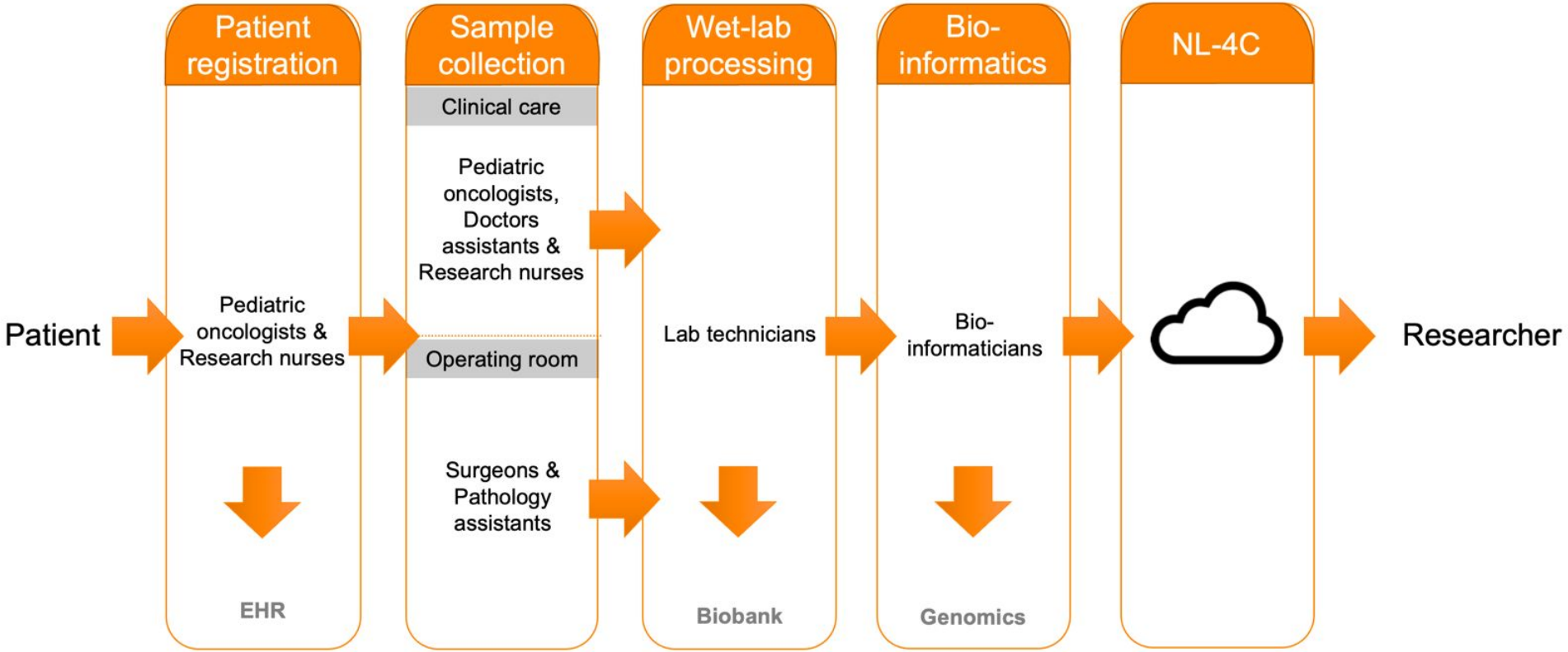
Childhood cancer

- Childhood cancer survival rate has increased the last decades to around 75-80% today
- Still major cause of disease-related child death in high-income countries

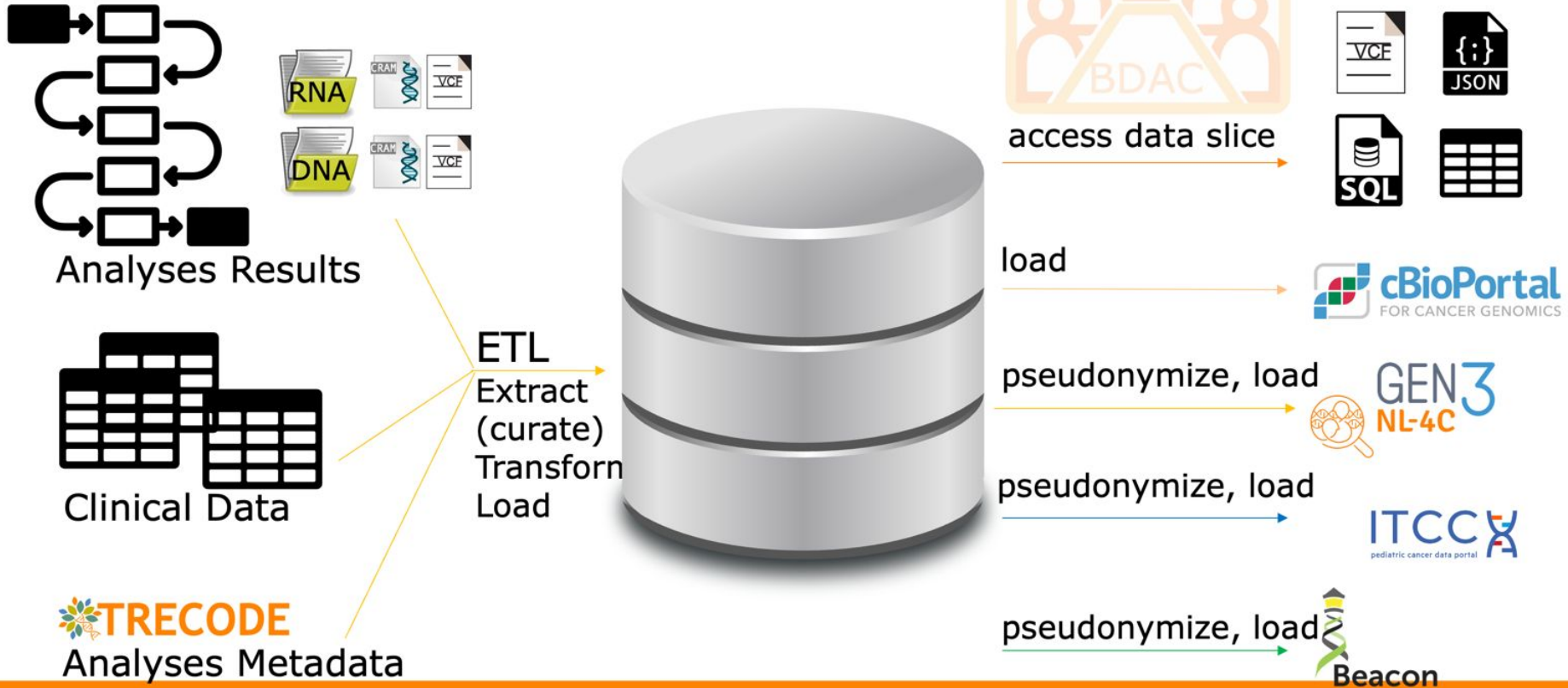
Addressing childhood cancer

- Childhood cancer is rare compared to adult cancer
- Mission: Curing every child with cancer, with an optimal quality of life
- Consorted efforts are needed to increase survival rates and quality of life
- The Princess Máxima Center comprises the first single national center and largest pediatric oncology institute in Europe
- Unique position to contribute to childhood cancer research world-wide

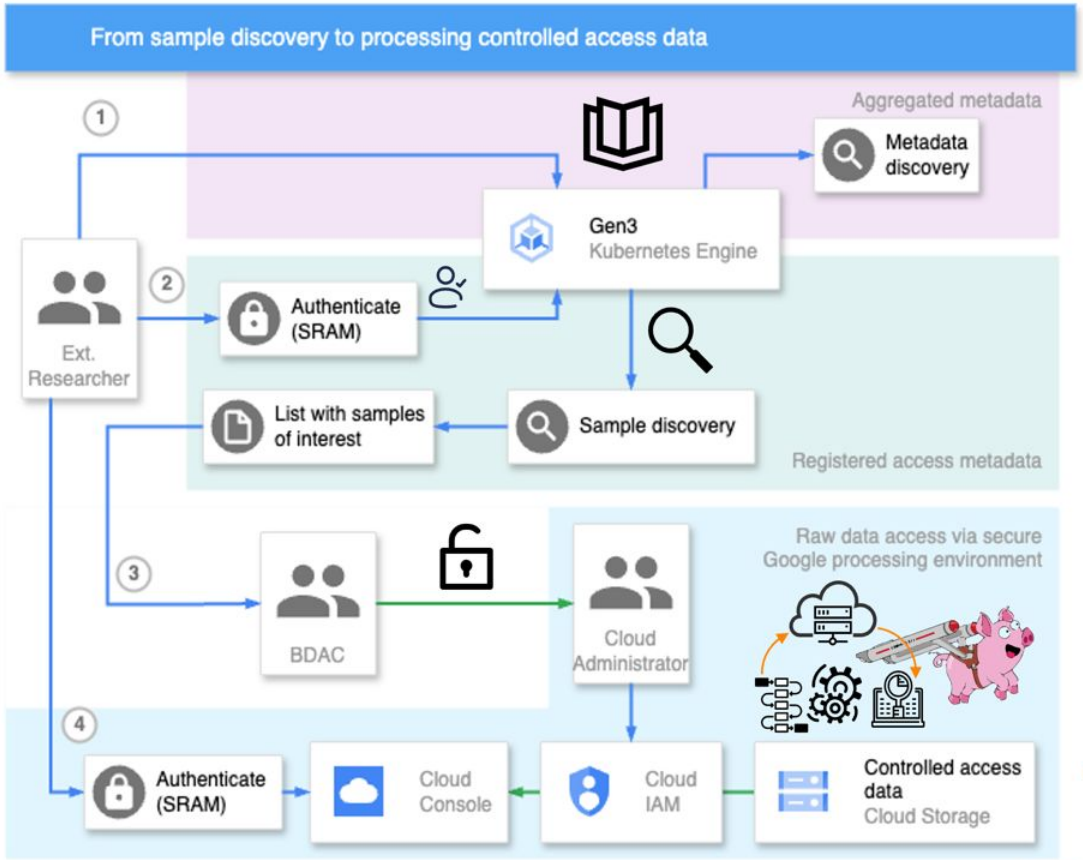
Dutch Comprehensive Childhood Cancer Commons

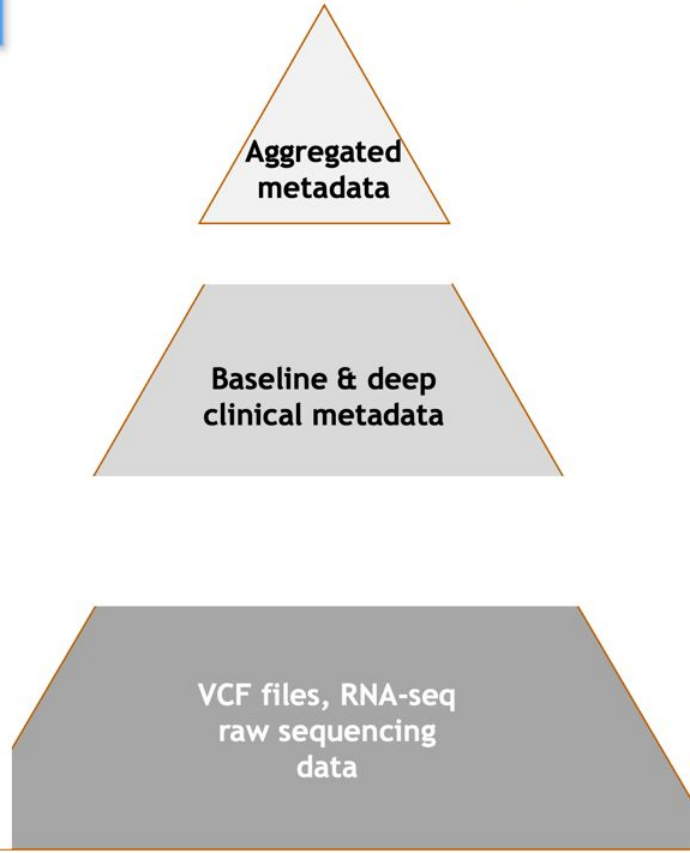


Architecture for data sharing



Data discovery through the Gen3 dataportal





Compliance: Navigating the EU/Dutch Landscape

- Laws & standards: General Data Protection Regulation (GDPR), ISO27001/NEN7510
- Patient/parent consent: broad consent that allows data to be used for research purposes
- Data Privacy Impact Assessment (DPIA): Mapping privacy risks of data processing operations
- Information security risk analysis

Biobank information letter with consent form for parent(s)/guardian(s)

Storage and use of body material and data collected at the Prinses Máxima Centrum for future scientific research

Official title: "Biobank of the Prinses Máxima Centrum: establishing a biobank for extensive characterization of pediatric malignancies".

Introduction

Dear parent(s)/guardian(s),

Your child comes to the Prinses Máxima Centrum for Pediatric Oncology because there is a suspicion of a form of cancer or some other disease that is treated at our centre. It requires, for example, treatment with chemotherapy or a stem cell transplant.

The purpose of this information letter is to explain why we are asking for your permission to collect, store and use body material and (medical) data of your child.

The letter explains what it will mean for your child if you grant permission for this and what the advantages and disadvantages are. Please read the information carefully and then decide whether or not to consent. Consent is granted by completing and signing the consent form. This form is provided in the Enclosure.

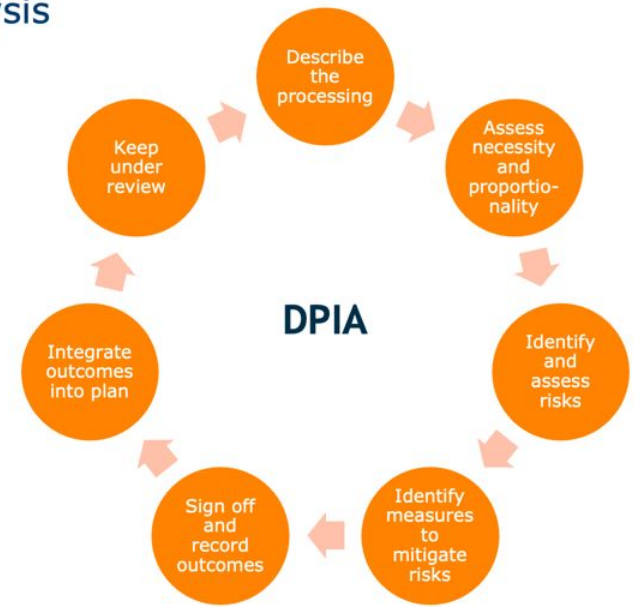
Ask your questions

You can base your decision on the information provided in this information letter. We also recommend that:

- You put your questions to your child's treating doctor or study nurse;
- Discuss this with your partner, family or friends;
- Put your questions to the independent expert, Dr. W. Kollen.

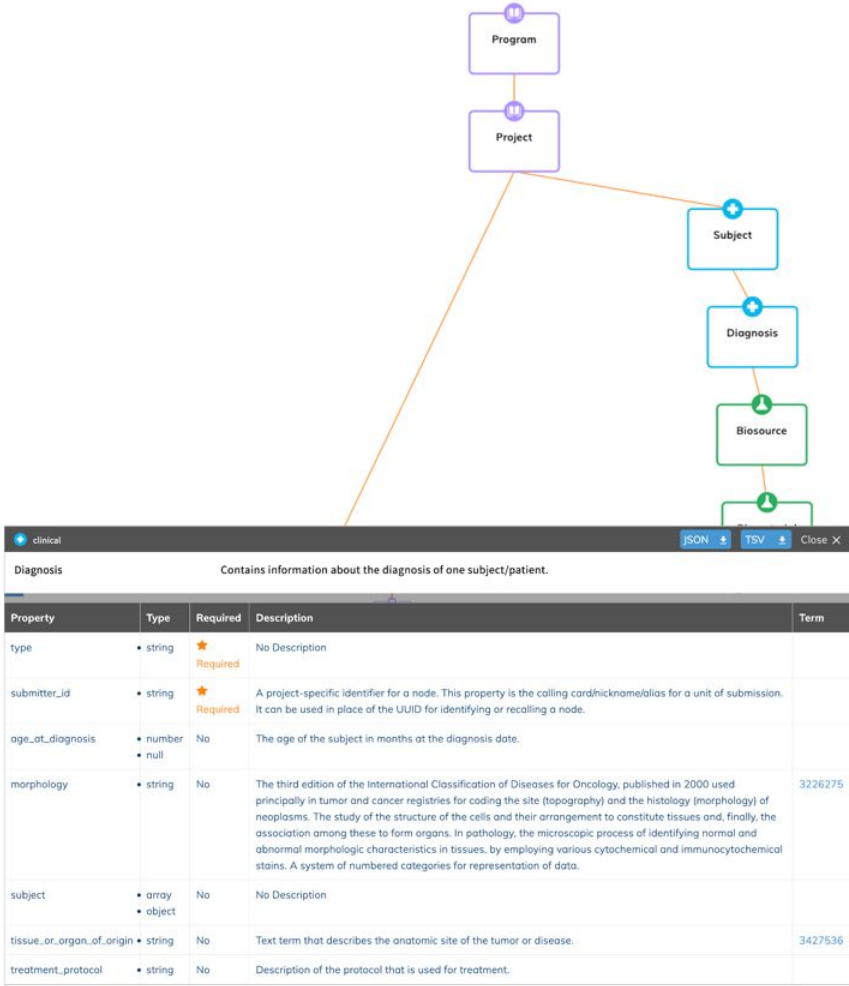
The contact information is provided in Enclosure A.

Further information can be found at www.rijksoverheid.nl/mensenonderzoek.



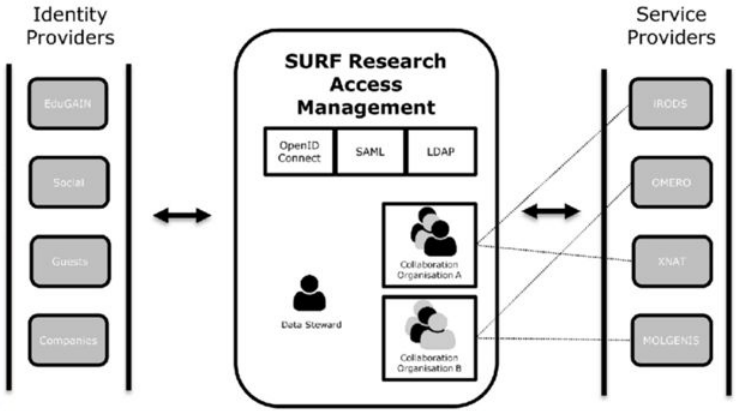
Most important findings DPIA

- Paediatric cancer is rare -> additional measures should be taken to protect patient privacy
  - Data generalisation
  - Date shifting & relative time transformation
  - Data suppression
- Do not re-use identifiers, reduce risk of profiling
- Consent is key, make sure we actively inform our client council, and publish transparent information on our website about (international) collaboration
- Have a strict process for access to deep clinical metadata & raw data



Most important findings information security risk analysis

- Keep systems up-to-date, prevent exploitable vulnerabilities
  - CI/CD pipeline to easily perform upgrades
  - Enabled Cloud Armor, block security exploits via threat signatures
  - Automated security scans via the web security scanner
- Data should be correct, only data for consented patients should be shared
  - Make use of an automated ETL process to provision the portal with data
- End-user accounts may be insecure
  - Make use of secure (federated) identities



More information about the NL-4C project

NL-4C website

NL-4C data portal (Gen3)

NL-4Cbioportal

Biobank

Data avaialble via the cloud to speed up research

Big Data Core

Acknowledgements

Kemmeren Group:

Patrick Kemmeren
Alex Janse
Alex Staritsky
Anastasía Spinou
Devin van Valkengoed
Eleonora Paganin
Emma Bernsen
Emmy Wesdorp
Eugène Verwiel
Femke Albertsboer
Fleur Wallis
Hinri Kerstens
Inês Henriques
Jack Jiang
Jasper van Dalum
Jayne Hehir-Kwa
Jet Zoon
Karina Borja
Karlijn Cammel
Marcel Santoso
Matthijs Pon
Nienke Biesot
Peter van Rijn
Rodrigo Jara
Roula Farag
Thimo Harmsen
Victoria Cruz

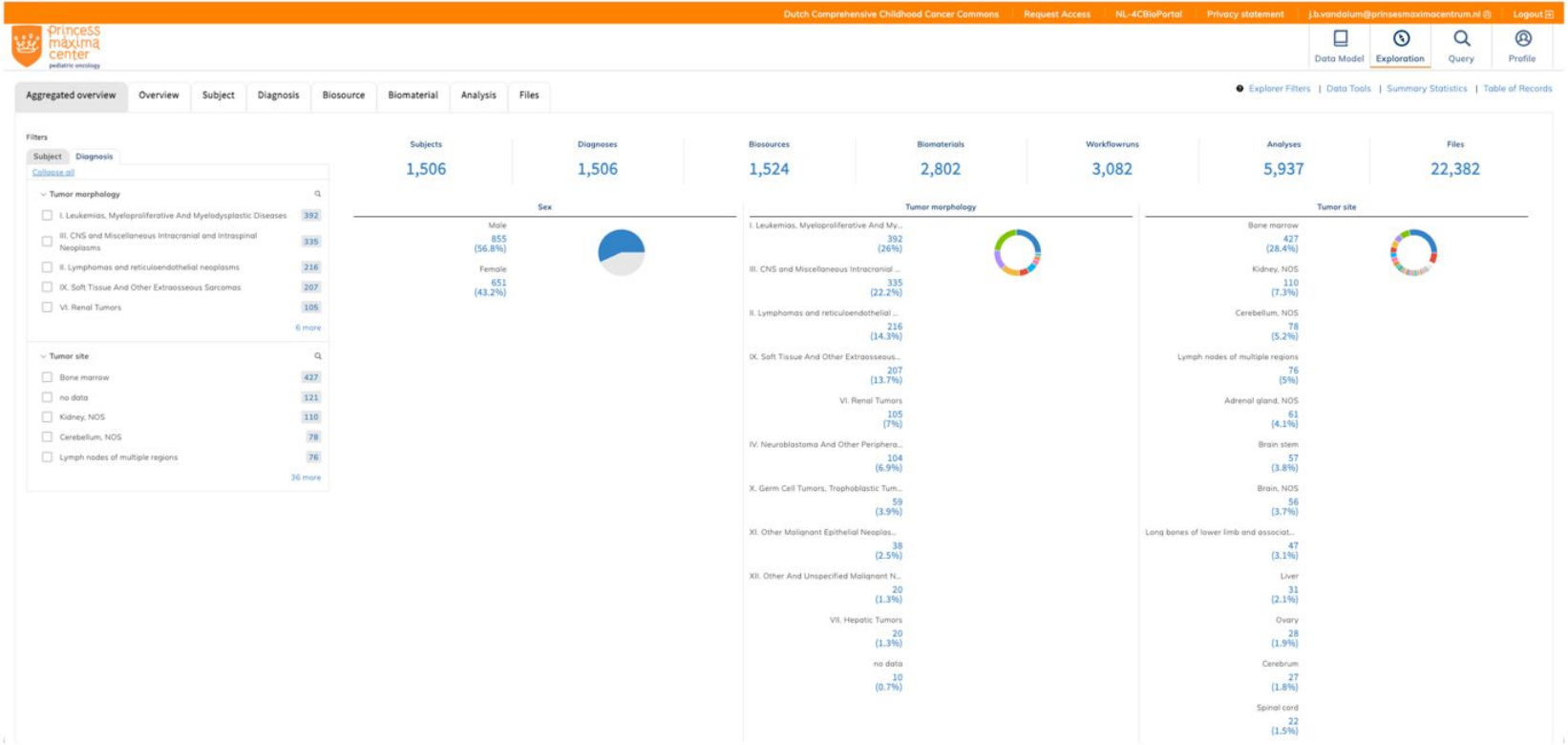
SE4Bio

Gen3 Community

Gen3 development team

The children and their parents

Questions


Dutch Comprehensive Childhood Cancer Commons
Request Access
NL-4CBioPortal
Privacy statement
j.b.vandalum@prinsesmaximacentrum.nl
Logout
Data Model
Exploration
Query
Profile
Aggregated overview
Overview
Subject
Diagnosis
Biosource
Biomaterial
Analysis
Files
Explorer Filters | Data Tools | Summary Statistics | Table of Records
Filters
Subject
Diagnosis
Collapse all
Tumor morphology
I. Leukemias, Myeloproliferative And Myelodysplastic Diseases 392
III. CNS and Miscellaneous Intracranial and Intraspinal Neoplasms 335
II. Lymphomas and reticuloendothelial neoplasms 216
IX. Soft Tissue And Other Extraosseous Sarcomas 207
VI. Renal Tumors 105
6 more
Tumor site
Bone marrow 427
no data 121
Kidney, NOS 110
Cerebellum, NOS 78
Lymph nodes of multiple regions 76
36 more
Subjects 1,506
Diagnoses 1,506
Biosources 1,524
Biomaterials 2,802
Workflowruns 3,082
Analyses 5,937
Files 22,382
Sex
Male 855 (56.8%)
Female 651 (43.2%)
Tumor morphology
I. Leukemias, Myeloproliferative And My... 392 (26%)
III. CNS and Miscellaneous Intracranial ... 335 (22.2%)
II. Lymphomas and reticuloendothelial ... 216 (14.3%)
IX. Soft Tissue And Other Extraosseous... 207 (13.7%)
VI. Renal Tumors 105 (7%)
IV. Neuroblastoma And Other Peripher... 104 (6.9%)
X. Germ Cell Tumors, Trophoblastic Tum... 59 (3.9%)
XI. Other Malignant Epithelial Neoplas... 38 (2.5%)
XII. Other And Unspecified Malignant N... 20 (1.3%)
VII. Hepatic Tumors 20 (1.3%)
no data 10 (0.7%)
Tumor site
Bone marrow 427 (28.4%)
Kidney, NOS 110 (7.3%)
Cerebellum, NOS 78 (5.2%)
Lymph nodes of multiple regions 76 (5%)
Adrenal gland, NOS 61 (4.1%)
Brain stem 57 (3.8%)
Brain, NOS 56 (3.7%)
Long bones of lower limb and associat... 47 (3.1%)
Liver 31 (2.1%)
Ovary 28 (1.9%)
Cerebrum 27 (1.8%)
Spinal cord 22 (1.5%)

Krumware

Compliance-as-Code for Modern Platforms

Making regulatory policy versioned, testable, and continuous

Colin Griffin
Krumware
Co-Chair CNCF Platform Engineering TCG

THE PROBLEM

Compliance can't keep pace with the platform

Overlapping obligations

- HIPAA / HITECH - PHI safeguards
- FDA 21 CFR Part 11 - electronic records & signatures
- GxP - validated systems for research
- IRB / data-use agreements per study
- GDPR for multi-site / international trials

! The gap

- Audit today is retrospective & manual - screenshots, spreadsheets, point-in-time attestations
- Cloud-native platforms change faster than manual audit cycles can track
- Grey areas - Compliance requirements are changing as technology evolves
- **Policy lives in documents; systems live in code - so they drift apart**

DEFINITION

What we mean by a "modern platform"

Not a product you buy - a curated internal capability your teams build on. The CNCF Platforms White Paper frames a platform as something that presents common capabilities and ***"consistent, compliant experiences"*** so teams deliver faster.

✓ You already have one

Every org relies on an internal platform crafted for its own needs - even if it's just documentation on how to use third-party services.

? So the real question

Not whether you have a platform - but how mature and how governable it is. That's what the rest of this talk is about.

Two lenses on platform maturity

Top-down

Cloud Native Maturity Model

- 5 levels across 5 dimensions: Business Outcomes, People, Policy, Processes, Technology
- Connects technical goals to business outcomes
- **Speaks the language of leadership**

Bottom-up

Platform Engineering Maturity Model

- Concrete companion to the Platforms White Paper
- Lower levels = tactical; higher levels = strategic
- **Designed to complement the CNMM**

Both insist: **outcomes come from balancing people, process, and policy alongside technology.**

Maturity ≠ "climb to level 5"

Each level demands more funding and more people's time.

> Reaching the highest level ***should not be a goal in itself.***

- Target investment where the qualities actually benefit your context
- The models were built to reflect both fast-moving and highly regulated industries - healthcare included

START WHERE YOU ARE

You already own a good framework

CIS Controls v8 | NIST 800-53 / CSF | ISO 27001 | HITRUST CSF

- These already map to your obligations (HIPAA, Part 11) - they are not the problem
- **The problem is their form: Word docs in SharePoint, PDF binders, annual review**
  - Hard to audit, hard to maintain, hard to see gaps

{ } **The insight**

CIS v8 is already a schema:

18 families → 153 safeguards → 3 implementation groups.

That structure is begging to be version-controlled.

1

STEP 1

Define what a source of truth must guarantee

"Source of truth" is a set of properties, not a tool. Apply the same control-thinking you use for PHI to the policy documentation itself.

Version control

Every change → immutable, retrievable version

Peer review

Changes reviewed before becoming authoritative

Access control

Author vs. approve vs. read is enforced

Auditability

Who changed what, when, and why is reconstructable

Lifecycle / state

Explicit status: draft / active / deprecated

Traceability

Policy → control → evidence links hold

Review cadence

Staleness is visible and enforced

Single source

One authoritative location; no scattered copies

1

STEP 1 · APPLIED

Meeting the bar in Confluence, done right

Your existing wiki can be a legitimate source of truth - if the controls are deliberately configured.

- ✓ **Version control** → Page history & version diffs
- ✓ **Peer review** → Approval workflows / required reviewers
- ✓ **Access control** → Space & page permissions, restricted edit
- ✓ **Auditability** → Version history + mandatory change comments
- ✓ **Lifecycle** → Trackable statuses / page properties
- ✓ **Review cadence** → last_reviewed properties + expiry reporting

! **The catch:** out of the box a wiki is free-for-all editing with invisible staleness. The properties come from **governance, not the tool.**

2 NEXT EVOLUTION

Docs-as-code: properties enforced by the system

From governance discipline → to mechanism

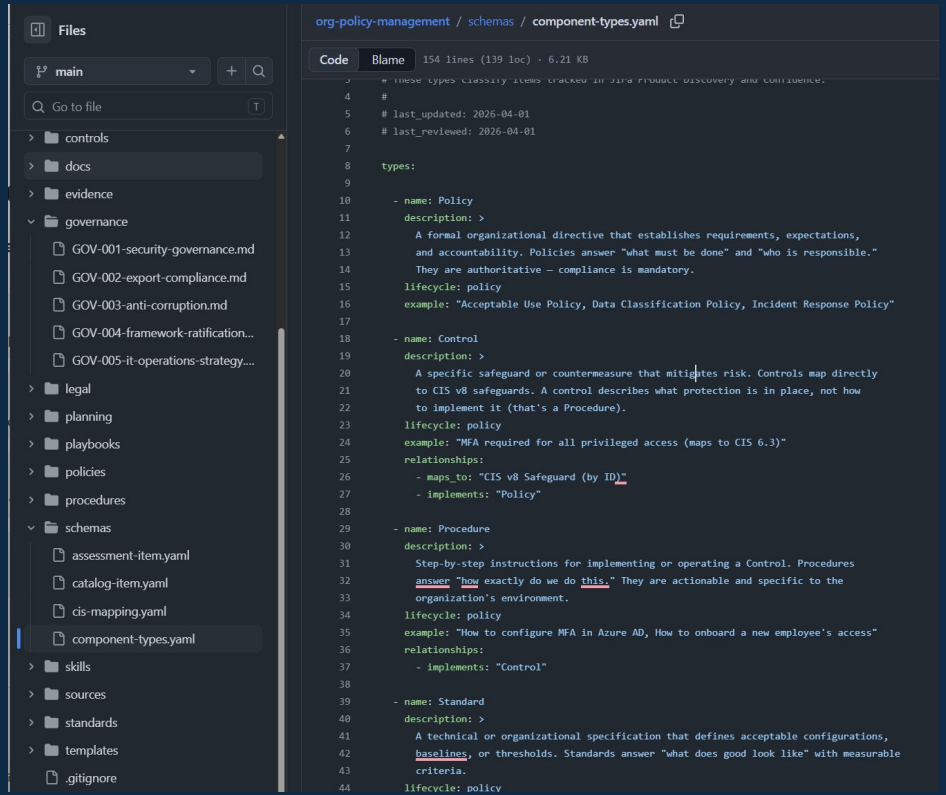
- Version / history → git commits, diff & blame (immutable)
- Peer review → pull requests with required approvers (can't merge without them)
- Lifecycle & metadata → structured frontmatter
- Traceability → machine-checkable links
- **Wiki doesn't go away: git = enforced source of truth, Confluence = published reading surface**

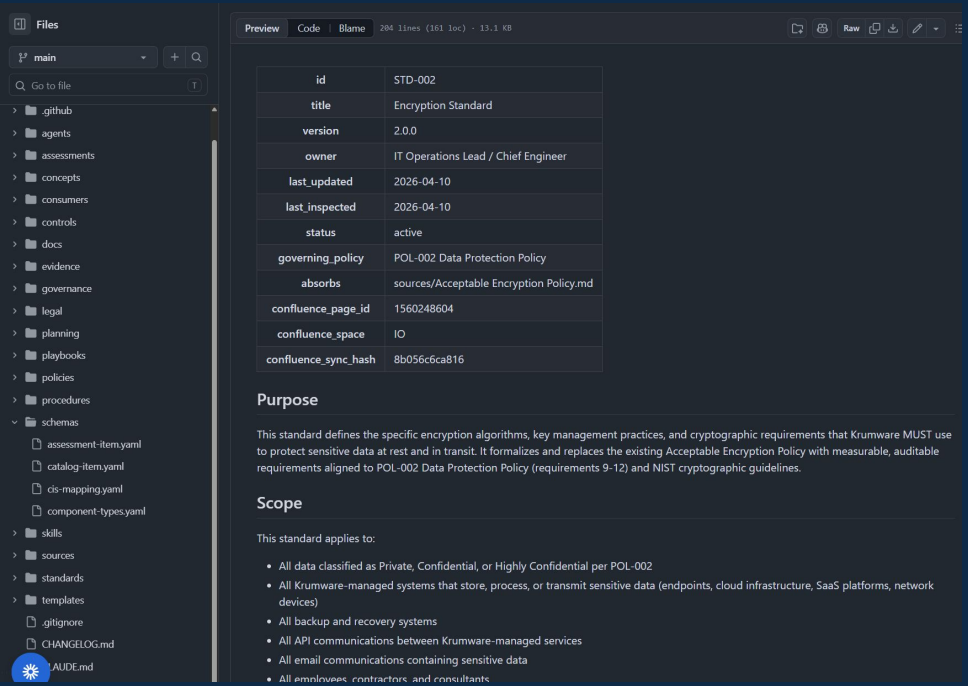
Case study - Krumware CIS v8 repo

- Every control = a tracked artifact with auditable frontmatter (id, version, owner, last_inspected, status, CIS mapping)
- Honest implementation status - partial coverage is stated, not hidden
- 67 gaps for higher tiers tracked as first-class artifacts
- confluence_page_id + sync hash keeps git ↔ wiki linked with drift detection

CIS safeguard → control → evidence, answerable by design

2 NEXT EVOLUTION





3

STEP 2+ · TO BE EXPANDED

From tracked docs to executable enforcement

To flesh out

- Policy engines / policy-as-code - (Kubernetes:) OPA/Rego, Kyverno, Cloud Custodian
- Guardrails at admission / deploy time (e.g. Kubewarden, Neuvector) vs. detective controls after the fact
- **Evidence generated as a byproduct of the pipeline - not a separate audit project**

AN EMERGING STANDARD · OPTIONAL TARGET

OpenSSF Gemara

GRC Engineering Model for Automated Risk Assessment

OpenSSF · sandbox

Seven categorical layers

7	Audit & Continuous Monitoring
6	Preventive & Remediative Enforcement
5	Intent & Behavior Evaluation
4	Sensitive Activities
3	Risk & Policy
2	Threats & Controls
1	Vectors & Guidance

■ Definition (1–3) ■ Action (4) ■ Measurement (5–7)

Why it matters for this journey

- **The standards-body version of the docs-as-code pipeline you just saw**
- Descends from the CNCF Automated Governance Maturity Model (Policy · Evaluation · Enforcement · Audit) - reinforces the CNCF thread
- Machine-readable schemas (Control Catalog, Policy, Risk, Audit Log) - a CIS-v8 catalog can target it
- Model is intentionally stable, so tooling can build on it safely
- **Adopt as a convergence target - align, don't reinvent**

Coverage for compliance, at the code level

Context as a Product - the docs-as-code discipline, pushed all the way down to the source file.

Bind code to the standard it implements

```
# @knowledge CXP-001@0.6.0
# Frontmatter follows the control spec.
def issue_session_token(uid):
    return sign(uid)
```

A one-line annotation = the code-level equivalent of citing a control.

Measured like test coverage

- **Breadth** - every file: covered / exempt / gap / unattributed
- **Depth** - inline refs score higher than blanket folder coverage
- **Audit** - LLM checks code content matches its claim (planned)

Why it closes the loop

- Gates in CI like coverage does - min_coverage and min_avg_depth fail the build below target
- --against origin/main shows the coverage delta on a PR
- **Completes the chain: policy → structured control → code that cites & satisfies it, measured continuously**
- Where it heads: the same bindings drive agents to generate code that reflects org standards - by construction, not after the fact

v0.1.0-alpha · emerging

One control, end to end

Example: "PHI must be encrypted at rest" (CIS 3.11)

1. **Regulation** HIPAA / Part 11 requirement, in prose
2. **Tracked control** CTL-CIS-3.11 artifact, versioned & owned
3. **Enforcement** policy rule / admission gate
4. **Evidence** auto-generated audit artifact

Before: manual quarterly screenshot | After: continuous enforcement + auto-evidence

WHERE DO YOU SIT?

Mapping to your maturity journey

0 **Docs in SharePoint** Fails most source-of-truth controls

1 **Governed source of truth** Confluence configured to the properties

2 **Docs-as-code** Properties enforced by the system · Krumware baseline

3 **Mapped to enforcement** IaC scanning, policy engines in CI/CD

4 **Code-level coverage** Code bound to controls; coverage & depth gated in CI · CxaaP

5 **Continuous & generative** Continuous evidence, standards-aware code generation

Pick the level that fits your risk & budget context - a governed wiki (Level 1) is already defensible.

BE HONEST · TO BE EXPANDED

Risks & honest caveats

! Still code

Policy-as-code needs review, testing, and ownership like any code

! False assurance

Encoded policy can be confidently wrong - a green check isn't proof

! Auditor education

Machine-generated evidence may need explaining to regulators

! Upkeep cost

Governed wiki depends on discipline; docs-as-code trades that for tooling upkeep

TAKEAWAYS

The through-line

- A modern platform already implies compliant experiences - build on that
- You already own a good framework - change its form, not its content
- Source of truth is a set of controls, not a tool - hold your system to the bar
- Docs-as-code is the next evolution: the same properties, enforced by the system
- Then treat policy as an executable dimension: control → enforcement → evidence
- Mature deliberately - invest where risk and value meet, not everywhere

Your next step: *define your source-of-truth controls, assess your current tool against them.*

REFERENCES

Sources & further reading

CNCF Cloud Native Maturity Model	https://maturitymodel.cncf.io (4.0 beta added AI, FinOps)
CNCF Platforms White Paper	https://cloudnativeplatforms.com/whitepapers/platforms/
CNCF Platform Engineering Maturity Model	https://cloudnativeplatforms.com/whitepapers/platform-eng-maturity-model/
Platform Eng. Maturity Model Assessment	https://cloudnativeplatforms.com/initiatives/platform-maturity-model-assessment/
CIS Critical Security Controls v8	https://cisecurity.org/controls/v8
Open Source Security Foundation (OpenSSF)	https://www.openssf.org/
OpenSSF's Gemara	https://gemara.openssf.org/

Australian BioCommons

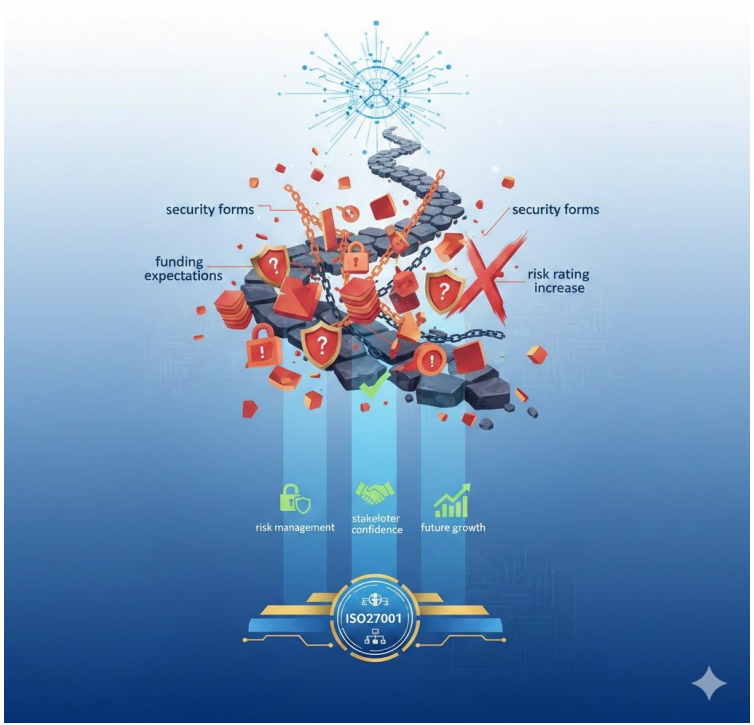
Managing Security

Prepared by Nakul Deshmukh
Contact: nakul@biocommons.org.au

Security Uplift

As BioCommons continues to mature as into its role as **a nationally significant research infrastructure provider**, some key challenges are surfacing at an increasing rate.

- increasing number of security forms from third-parties
- increase in expectations of security for funding cases
- increase in risk rating from host institutions
- increase in regulatory scrutiny
- increase in attack complexity and frequency



^ gemini images are funny

Australian BioCommons

Biocommons operates as a UoM-led project but involves multiple partner institutions sharing operational responsibilities.

ISO 27001 forces clarity on questions that would otherwise remain ambiguous: who **owns what controls**, how are **responsibilities** distributed, what's the **governance** structure.

Australian BioCommons

Actively Managing Risks:

- Sensitive Data Disclosure
- Account Compromise
- Public Service Unavailability
- Ransomware or Destructive Malware
- Exploit of Unpatched Vulnerabilities
- Denial-of-Service or Abuse
- Supply Chain Compromise
- Insecure Development
- Inadequate Security Awareness
- Governance Failure
- Inadequate Incident Response or Monitoring

What is the scope?

Australian
BioCommons

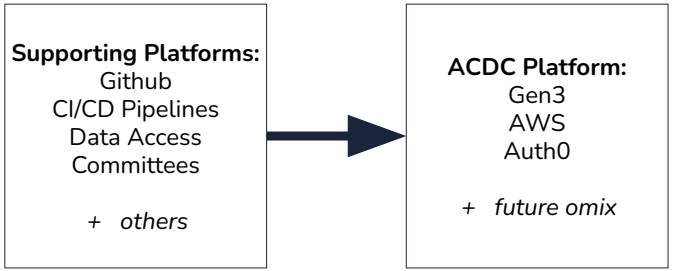
Main Scope

ACDC Platform:
Gen3
AWS
Auth0

\+ *future omix*

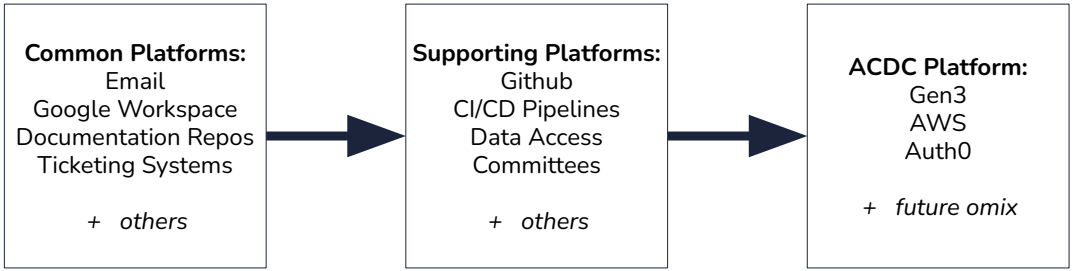
Australian BioCommons

Supporting Scope

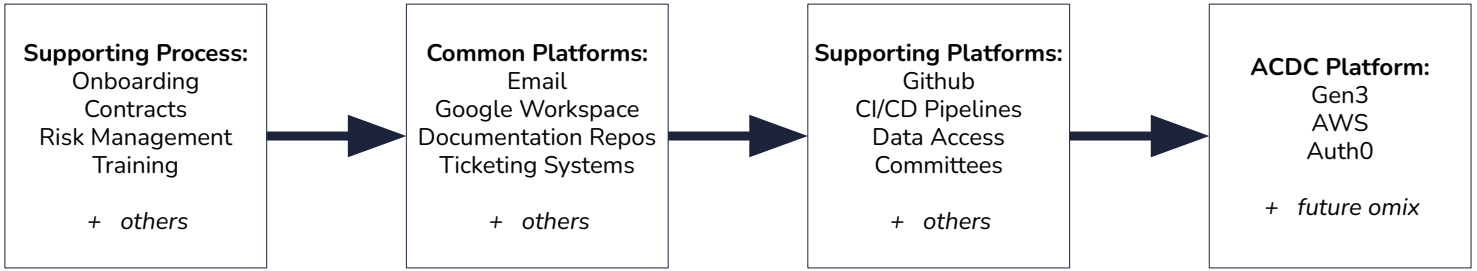


Australian BioCommons

Common Scope



Process Scope



Actively Managing Risks:

- Sensitive Data Disclosure
- Account Compromise
- Public Service Unavailability
- Ransomware or Destructive Malware
- Exploit of Unpatched Vulnerabilities
- Denial-of-Service or Abuse
- Supply Chain Compromise
- Insecure Development
- Inadequate Security Awareness
- Governance Failure
- Inadequate Incident Response or Monitoring

Team

- Jeff Christiansen
- Jonathan Arthur
- Steven Manos
- Conrad Leonard
- Samitha Amarapathy
- Nakul Deshmukh

Assistive Processes:

- Fortnightly Sprints
- Risk Elicitation Exercises
- Control Mapping
- AI Assistant (Rosie)*
- University of Melbourne endorsement
  - Architecture reviews
  - Partner support
  - Tooling
- University of Chicago advisory
- CI-ISAC advisory

Nakul has finished talking. (allegedly)

Australian BioCommons