Gen3 Webinar
NCI DCFS
Quality Assurance Processes
Monday, August 24, 2020
12:30 PM - 1:30 PM (CST)
NCI DCFS Quality Assurance Processes

Jiaqi Liu
Marcelo Costa
Andre Paredes

Center for Translational Data Science,
University of Chicago
Because Framework Services serve controlled access data on our platforms, we want to ensure we provide reliable and scalable solutions to our users. This presentation comprises an overview of the processes we follow before a release to ensure that DCFS provides a reliable, scalable and secure way to access data.
Introduction

- Gen3 is an open source solution for operating data commons.
- NCI Data Commons Framework Services (DCFS) runs an instance of Gen3 Framework Services.
- Gen3 is built and operated by the Center for Translational Data Science at the University of Chicago
The Agenda

● Intro to QA
  ○ Testing Pyramid / Observability

● How Code is Released
  ○ Our Release Process / Monthly releases

● How Data is Released
  ○ Data Indexing / GDC Data Replication

● Testing
  ○ Automation
  ○ Observability
  ○ Independent Infrastructure Tests
  ○ Load testing
  ○ Long Running Tests

● Conclusion
  ○ Managing Quality
Test Pyramid and Observability

- Manual Tests
- GUI Tests
- Acceptance Tests (API Layer)
- Unit Tests/Component Tests

- Logs and Events
- Tracing
- Metrics
Our Release Process

- Feature Freeze every 2nd Friday
- Code freeze every 4th Friday
  - The Release is published and PROD PRs are produced.

<table>
<thead>
<tr>
<th>Month A</th>
<th>Month B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Release Management version N</td>
<td>Cut testing branch version N</td>
</tr>
<tr>
<td>Testing version N</td>
<td>Publish Release N</td>
</tr>
<tr>
<td></td>
<td>Release Management version N+1</td>
</tr>
<tr>
<td></td>
<td>Cut testing branch version N+1</td>
</tr>
<tr>
<td></td>
<td>....</td>
</tr>
</tbody>
</table>

- https://github.com/uc-cdis/cdis-manifest/tree/master/releases/2020
How Data is Released

NCI CRDC Data Releases

- 08/03/2020: GDC Data Release v25.0 available through DCFS. See more at GDC.
- 07/22/2020: PDC CPTAC3-Discovery CPTAC LUAD Discovery Study - Acetylome available though DCFS. See more at PDC.
- 07/22/2020: PDC CPTAC3-Discovery CPTAC LUAD Discovery Study - CompRef Acetylome available though DCFS. See more at PDC.
- 07/21/2020: ICDC GLIOMA01 available through DCFS. See more at ICDC.
- 06/08/2020: GDC Data Release v24.0 available through DCFS. See more at GDC.
- 06/02/2020: CPTAC CCRCC Discovery Study DIA Proteome available through DCFS. See more at PDC.
- 06/02/2020: PDC CPTAC HNSCC Discovery Study available through DCFS. See more at PDC.
- 06/02/2020: PDC ICPC Academia Sinica LUAD-100 available through DCFS. See more at PDC.
- 05/15/2020: PDC ICPC Early Onset Gastric Cancer dataset available through DCFS. See more at PDC.
- 05/08/2020: CDS LCCC 1108 available through DCFS. See more at LGCC 1108.
- 05/05/2020: GDC Data Release v23.0 available through DCFS. See more at GDC.
- 04/08/2020: CDS PKTC available through DCFS. See more at PKTC.
Testing: Automation

- Continuous Integration

Pre-commit hooks:
- Security analysis
- Formatting/Style

Gen3 QA tests
> 300 Integration Tests

PR Automation:
- Run Integration Tests
- Builds Docker image
- Unit Tests
- Static Analysis

CI Environment 1

CI Environment 2
Test Reports

https://dcf.gen3.org/software-releases

Example Negative Tests
Testing: Observability
Tests that require Complex / Extensive infrastructure provisioning
Not suitable for PR checks

- Bucket manifest generation
- GDC Data replication
Testing: Load Testing

- Benchmarking
- Performance bottleneck investigations
  - HAR (HTTP Archive) testing
  - PreSigned URL Transactions

QA or Staging Environment
Testing: Long Running Tests

- **Hundreds of thousands of instances:**
  - Clinical metadata subjects (Sheepdog)
  - Data lake studies (Metadata Service)
Ensuring Quality in Production

- **Continuous Monitoring**
  - Healthchecks
  - Monitor for Kubernetes Pod Health

- **Alerting**
  - NCI DCFS has alerts for if services are unavailable

- **Logging**
  - Logs for triaging bugs and issues

- **Rollbacks/Hotpatches**
  - If an issue is identified in production, we can rollback to a previous working version of the code
  - Bug fixes and hotpatches go through the QA process before they are released to production
Learn More / Contact Us

- [GitHub](https://github.com/uc-cdis)
- [Gen3.org](https://Gen3.org)
- Gen3 Community on Slack
- [dcf-support@datacommons.io](mailto:dcf-support@datacommons.io)
- [ctds.uchicago.edu](https://ctds.uchicago.edu)
Questions?