Regional Cabled Array (RCA)
Current Data QA/QC Activities and Priorities

UW Science Team

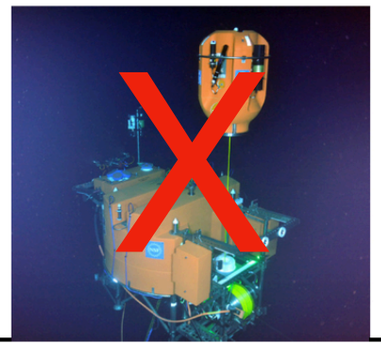
(Wendi Ruef, Mike Vardaro, Katie Bigham, Orest Kawka)

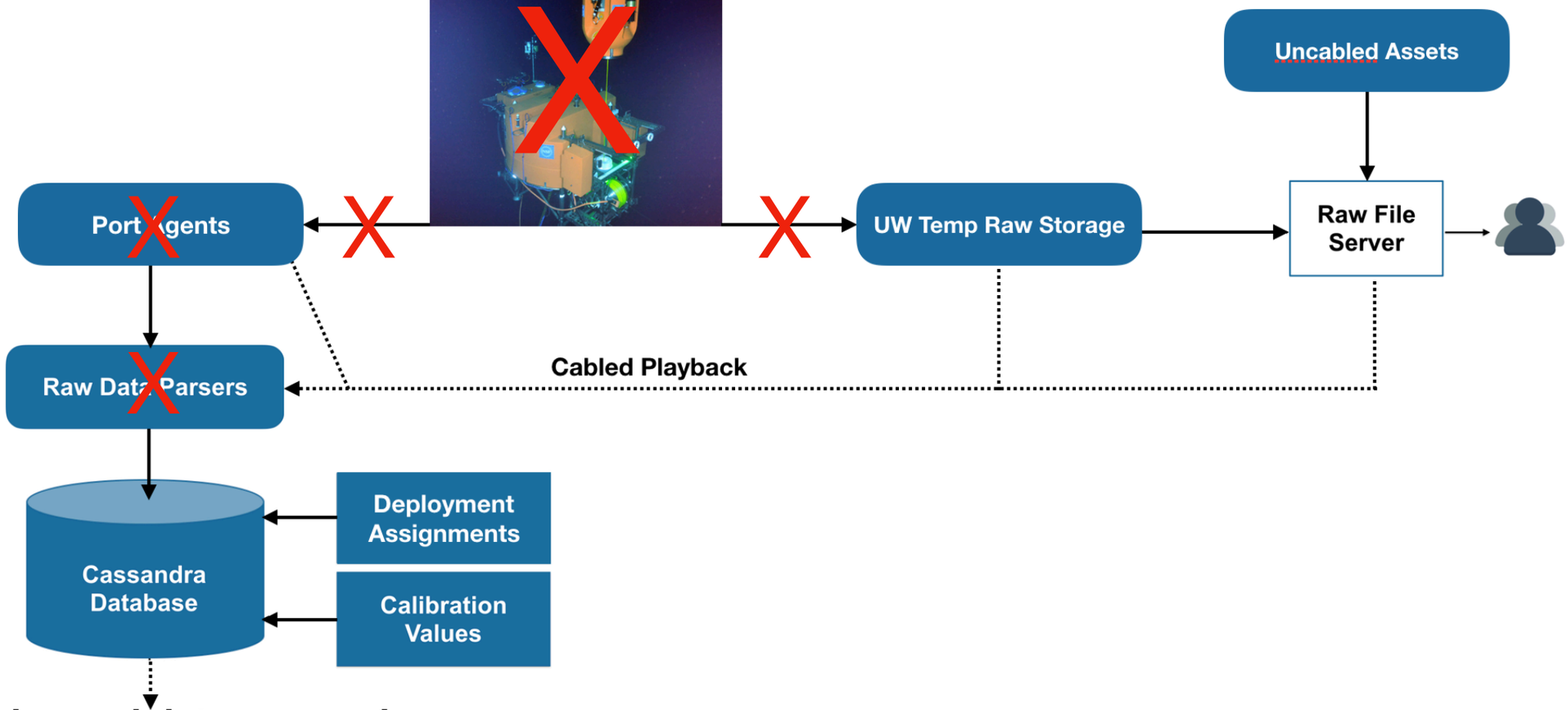
OOI Regional Cabled Array
School of Oceanography
University of Washington
Seattle, WA 98195

OOIFB and DDCI Joint Meeting, May 22, 2019

Data Delivery and Availability

- Causes of missing streamed data
  - Network outages
  - Port Agent/Parser errors
  - Cabled Array or instrument offline

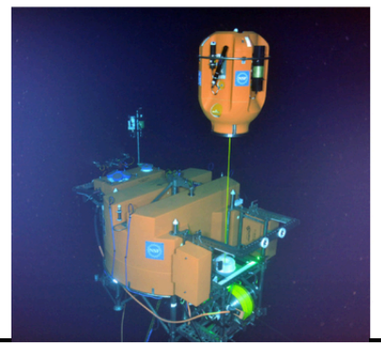


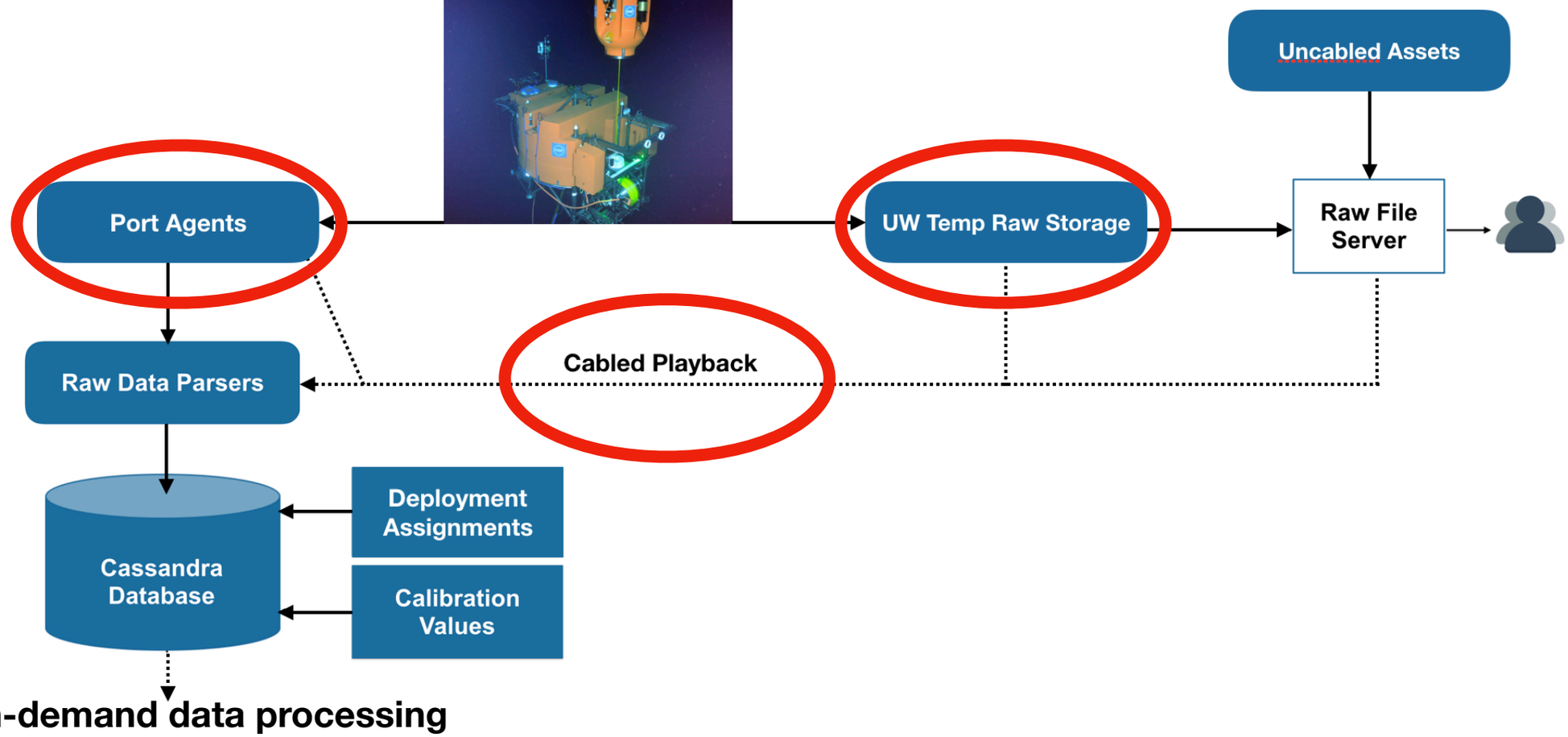


On-demand data processing

Data Delivery and Availability

- Data availability, backfilling missing streamed data
  - Accurate Cassandra/UFrame data availability stats require comparison of both raw data archives
  - Large-scale cabled array playback from Port Agent logs still in process





On-demand data processing

RCA Historical Critical Metadata Check

Phase 1 — Oct 2018 - Jan 2019

- Evaluation of scope and workflow development
- 2i-HITL comparison between GitHub and vendor cals
- Flagged files corrected and re-verified

Phase 2 — Feb - May 2019

- Refinement of workflow and documentation SOPs
- Scripted check of all digital cals
- 2i-HITL check of all pdf cals
- GitHub history crawler

Phase 3 — June - Sept 2019

- Final critical metadata checks
- Annotations and user outreach
- Select sensor deep dives
- Tier One/IRIS metadata
- Deployment image database
- Engineering platforms in GitHub

RCA Instrument and Critical Metadata QA/QC Workflow (Roles & Responsibilities)

Instrument in Need of Servicing

Prepare for Shipping
Create MAF
Enter Purchase Request
Log in Refurb Sheet/MAF
Create Service Request Form

Ship to Vendor

Vendor Service

Receive from Vendor

Inspect and Test
Close MAF
Complete Instrument Test
Update Refurb/MAF log

No

Pass?

Yes

Prepare for Integration
Create cal files:Github
Deployment Assignment
Update Master Tracking List

Integration

Deployment/Recovery

Initialization Checkout
First raw file

CA Google Drive

Check APL Refurb/ Assignment Sheet
Once a Week

If New

Record on Refurb Sheet

Metadata Entered on Github
Scripted when possible

ICDA Verification Protocol
2 sets of eyes

No

No

Yes

Pull Request

Merge to Github
Travis CI Check

Yes

Release Tag Created
Once a Week

Metadata Ingestion

Cyber Infrastructure Asset Management

No

Deployment Verification
Github = First Raw File
Github = OOINet

Yes

Verified

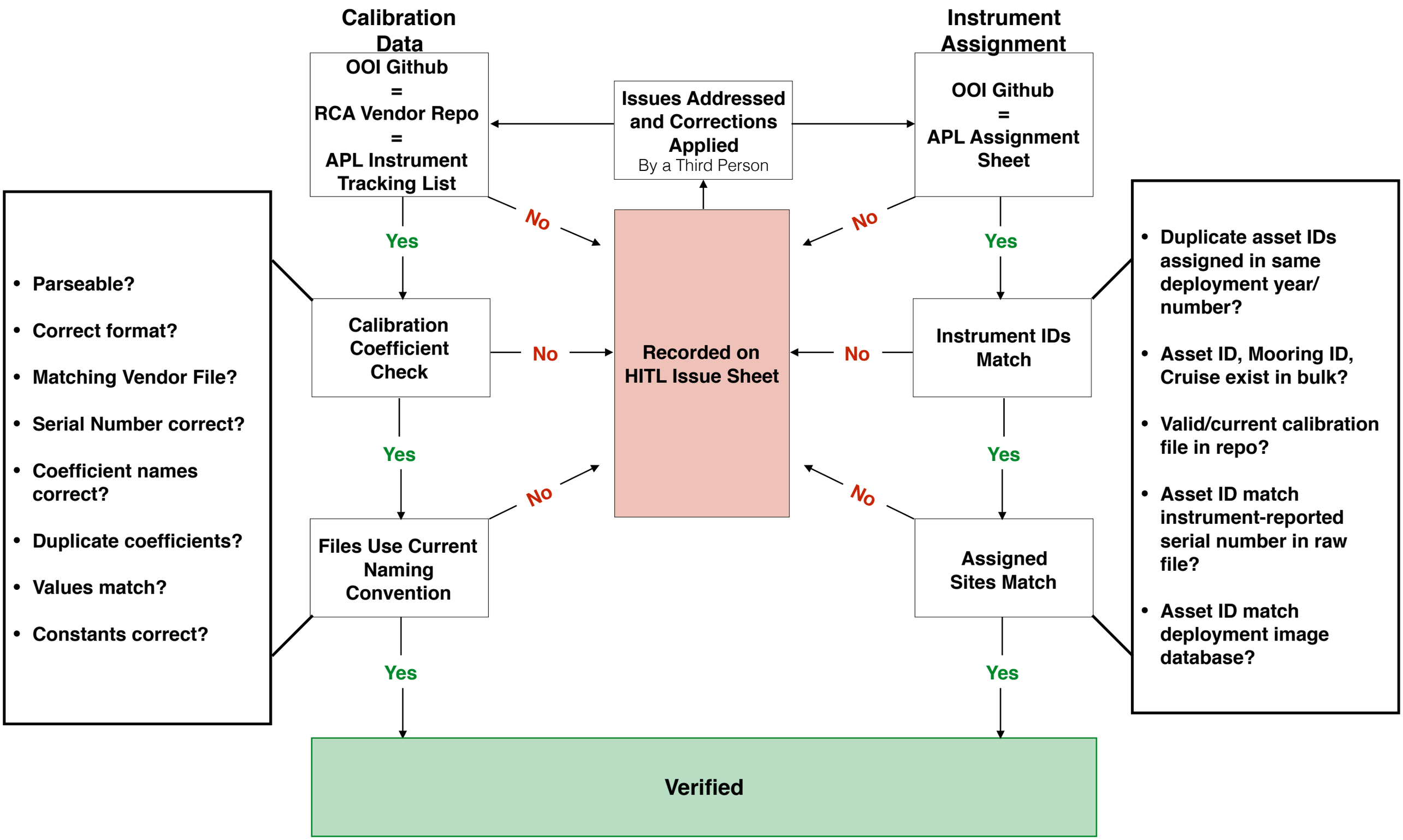
Legend

- RCA Data Team
- RCA Engineering Team
- RCA Equipment Manager
- Vendor
- OOI CI

Possible sources of error

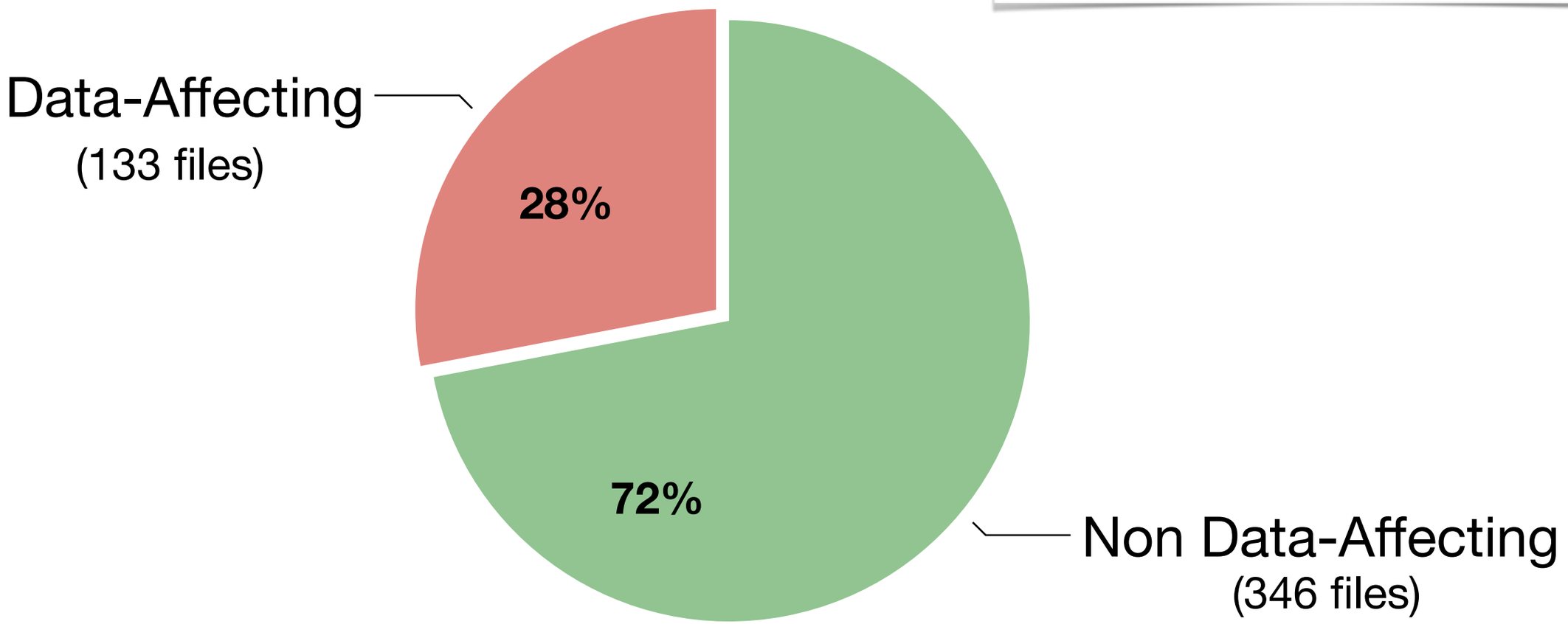
Verification touchpoint

RCA Instrument Calibration Data and Assignment (ICDA) Verification Workflow



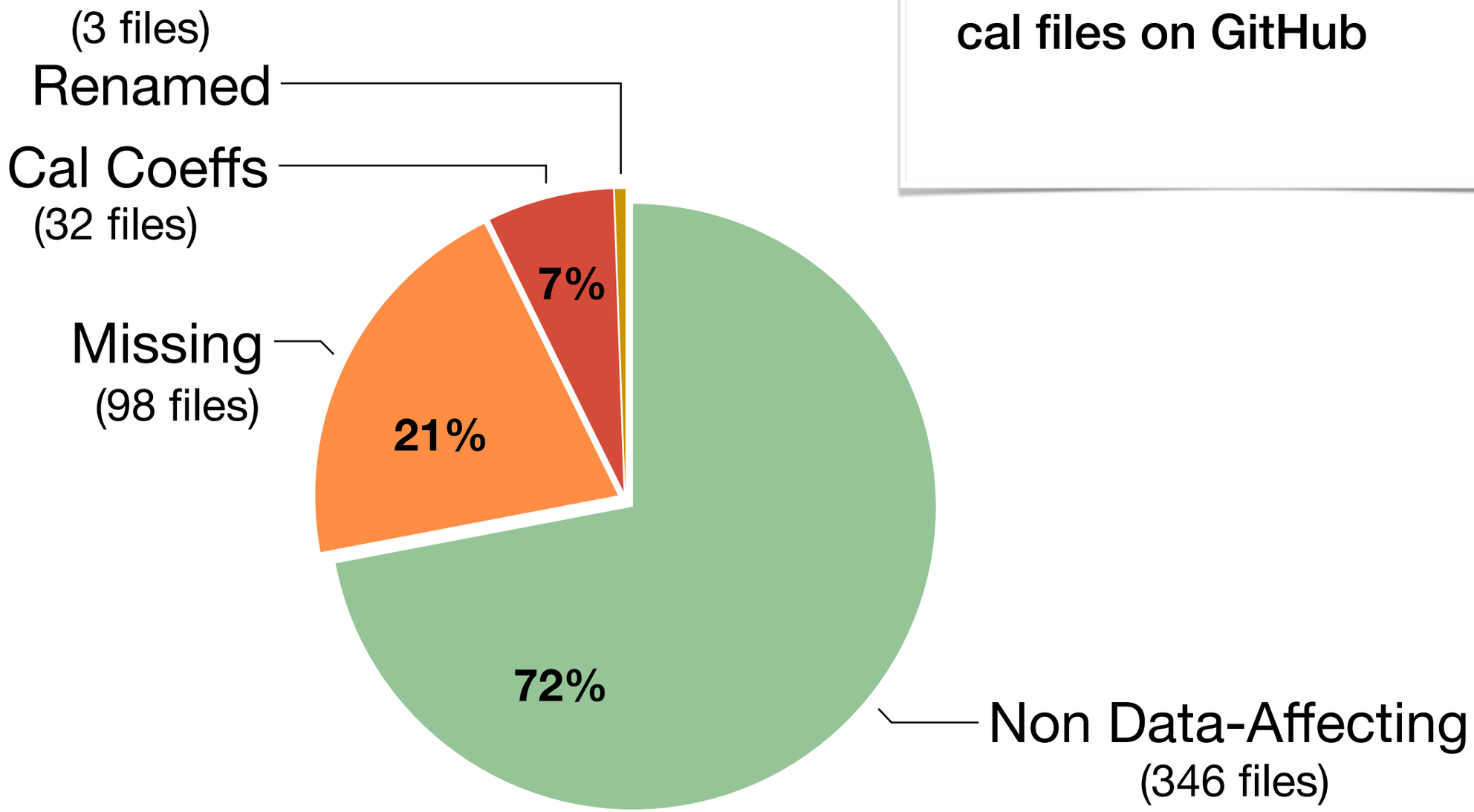
Calibration Verification Status

- 479 total files modified
- 28% of modifications were data-affecting

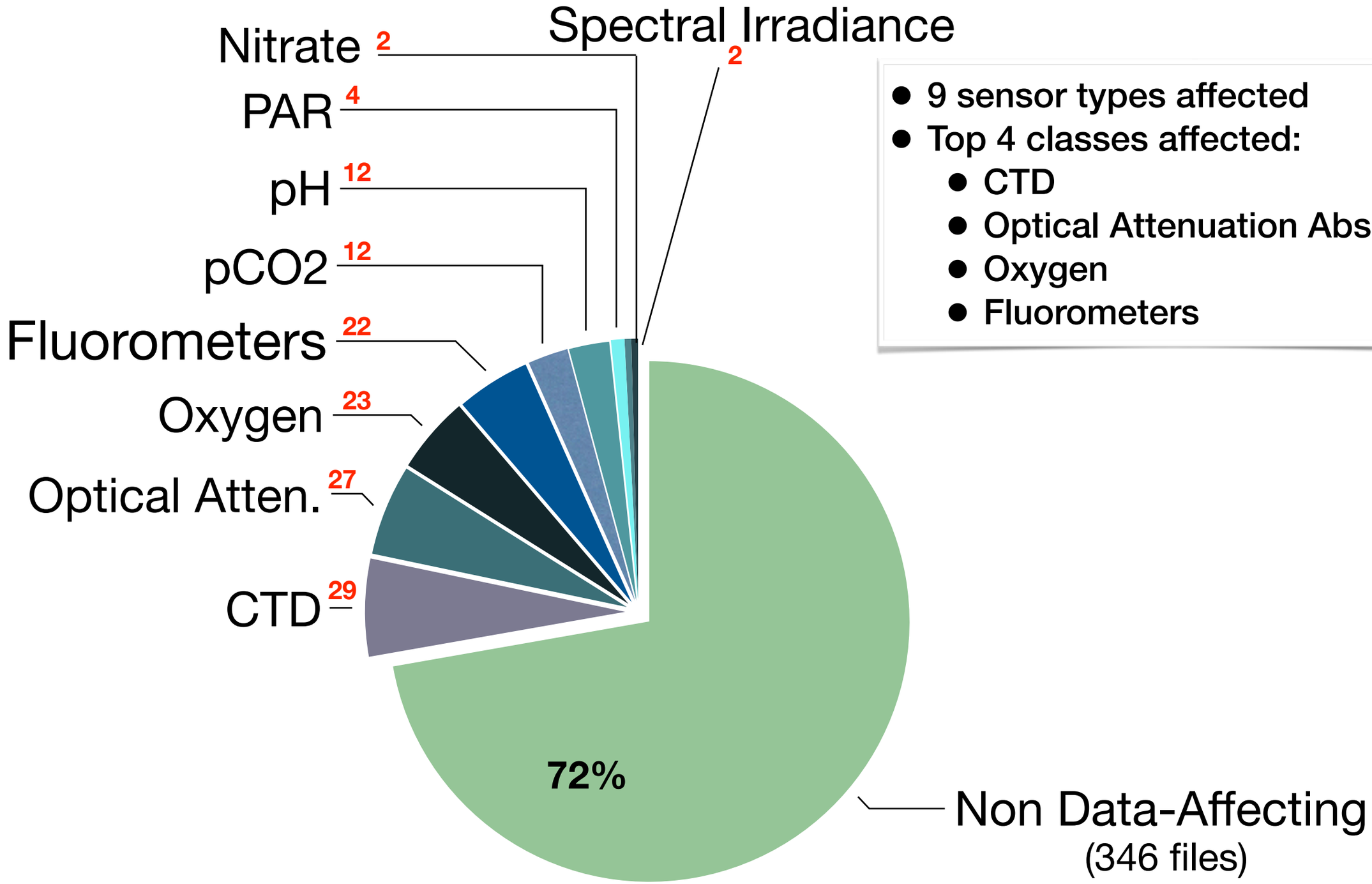


Calibration Verification Status

- Most data-affecting modifications were missing cal files on GitHub



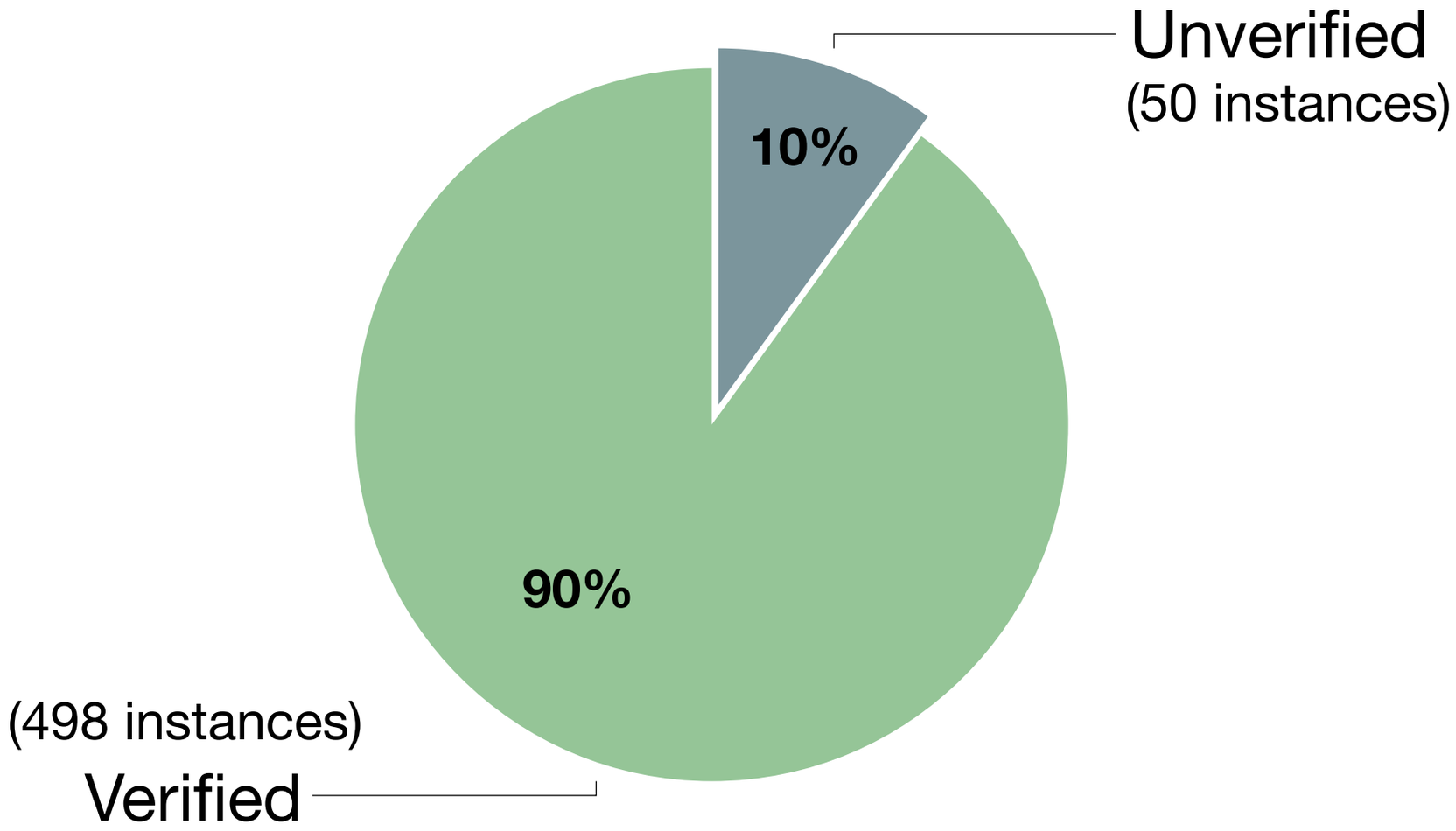
Calibration Verification Status



- 9 sensor types affected
- Top 4 classes affected:
  - CTD
  - Optical Attenuation Absorb.
  - Oxygen
  - Fluorometers

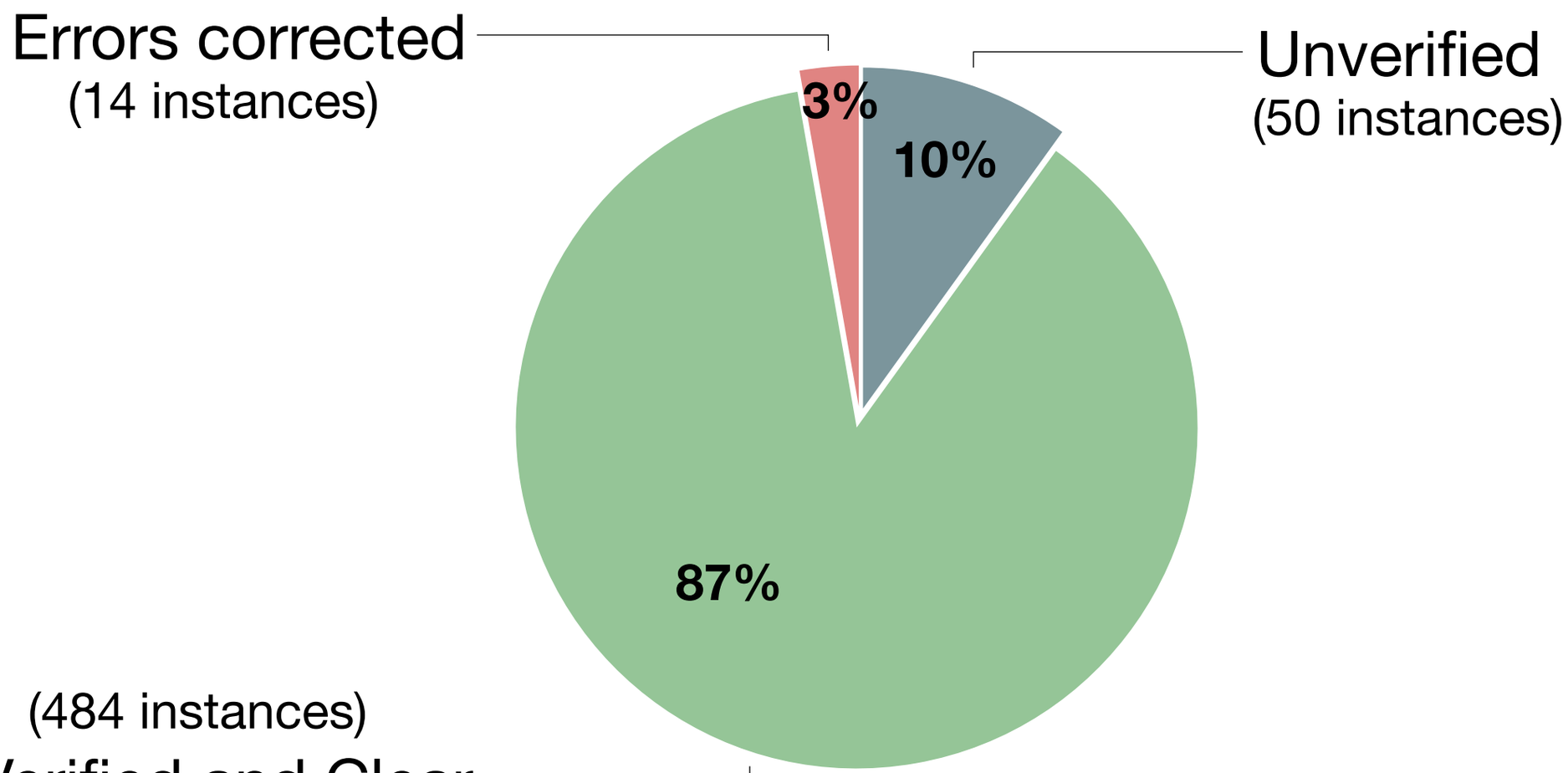
Deployment Assignment Verification Status

- 548 deployment instances on GitHub
- Verification 90% complete



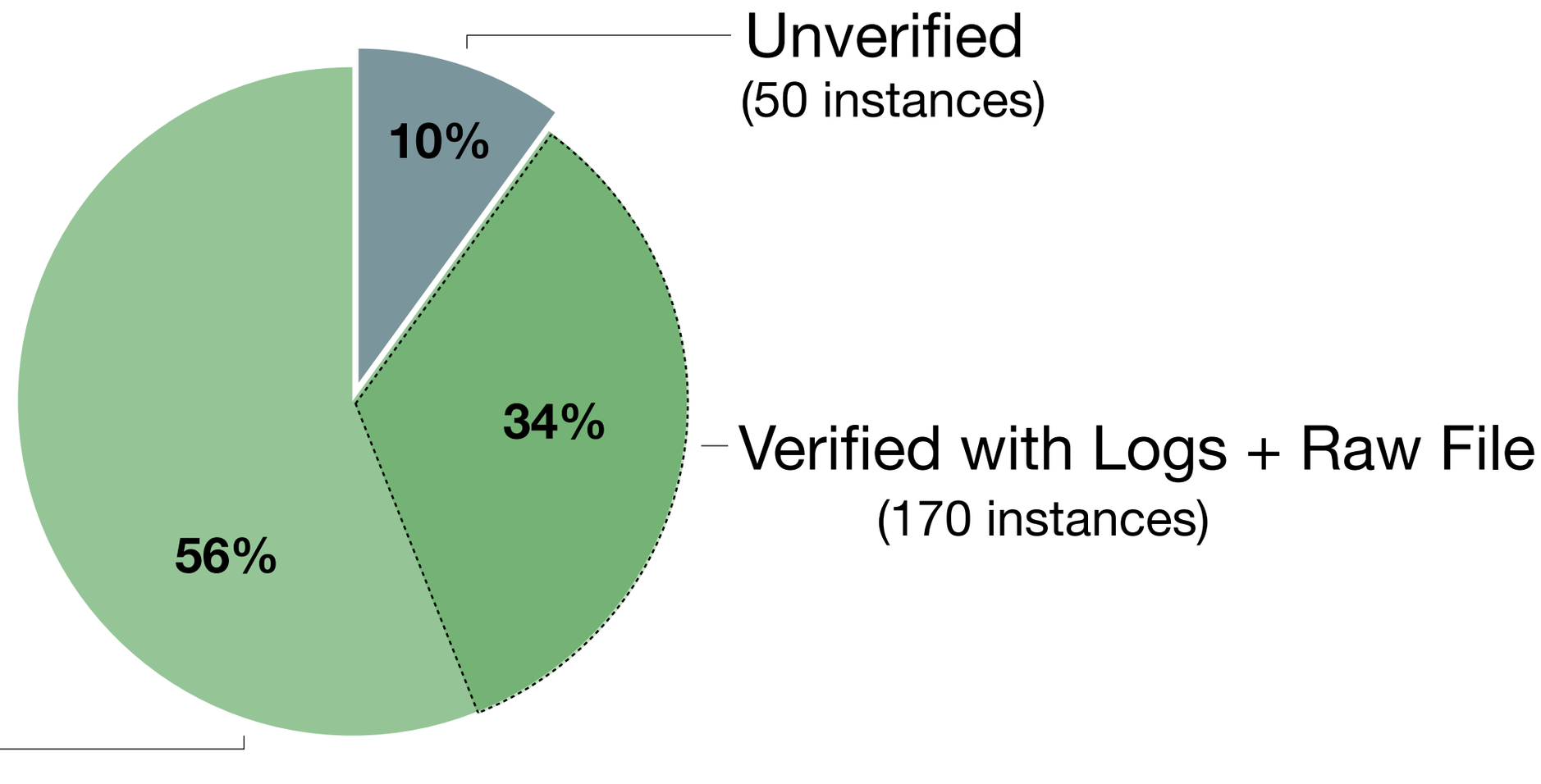
Deployment Assignment Verification Status

- 2.8% of verified deployment instances required correction (14 errors)



Deployment Assignment Verification Status

- Additional checks with sensor raw files complete on 34% of verified deployment instances



Sample File Reports

RefDes	changeDate	file	changeType	dateRangeStart	dateRangeEnd
CE02SHBP-LJ01D-06-CTDBPN106	11/29/18 13:37	ATOSU-69827-00003__20131207.csv	Missing file added	9/10/14 15:43	8/1/15 0:00
CE04OSBP-LJ01C-06-CTDBPO108	5/10/19 14:25	ATOSU-69828-00002__20131207.csv	calibration coefficients were modified	8/3/15 22:00	7/21/16 0:00
CE04OSBP-LJ01C-06-CTDBPO108	10/19/18 15:20	ATOSU-69828-00001__20171109.csv	Missing file added	6/25/18 0:00	10/19/18 15:20
RS01SBPS-SF01A-2A-CTDPFA102	10/19/18 15:20	ATAPL-66662-00002__20140528.csv	Missing file added	9/29/14 4:43	7/6/15 0:00
RS01SBPS-SF01A-2A-CTDPFA102	10/30/18 13:21	ATAPL-66662-00008__20150412.csv	Missing file added	7/7/15 20:47	7/17/16 0:00
RS01SBPS-SF01A-2A-CTDPFA102	10/19/18 15:20	ATAPL-66662-00002__20160303.csv	Missing file added	7/18/16 0:19	7/28/17 3:00
RS01SBPS-PC01A-4A-CTDPFA103	10/19/18 15:20	ATAPL-66662-00001__20140605.csv	Missing file added	9/29/14 1:53	7/6/15 0:00
RS01SBPS-PC01A-4A-CTDPFA103	10/30/18 13:21	ATAPL-66662-00009__20150419.csv	Missing file added	7/7/15 16:12	7/17/16 0:00
RS01SBPS-PC01A-4A-CTDPFA103	10/19/18 15:20	ATAPL-66662-00001__20160227.csv	Missing file added	7/18/16 0:19	7/28/17 3:00
CE04OSPS-SF01B-2A-CTDPFA107	10/30/18 13:21	ATOSU-66662-00007__20140214.csv	calibration coefficients were deleted	8/30/14 21:51	8/2/15 0:00
CE04OSPS-SF01B-2A-CTDPFA107	11/15/18 15:21	ATOSU-66662-00013__20150412.csv	calibration coefficients were modified	8/3/15 2:21	7/18/16 0:00

referenceDesignator	deployYear	deploymentSheetAssetID	rawFileVerification	ITM_Verification	calFileVerification	HITLnotes
RS03AXPS-SF03A-3D-SPKIRA301	2016	ATAPL-58341-00002	MATCH	MATCH	VALID_FILE	
RS03AXPS-SF03A-3D-SPKIRA301	2014	ATAPL-58341-00002	MATCH	MATCH	VALID_FILE	
RS01SBPS-SF01A-3D-SPKIRA101	2018	ATAPL-58341-00002	MISMATCH: raw: 267:ATAPL-58341-00004	MATCH	VALID_FILE	
RS01SBPS-SF01A-3D-SPKIRA101	2016	ATAPL-58341-00003	MATCH	MATCH	VALID_FILE	
RS01SBPS-SF01A-3D-SPKIRA101	2014	ATAPL-58341-00003	MATCH	MATCH	VALID_FILE	
CE04OSPS-SF01B-3D-SPKIRA102	2018	ATAPL-58341-00003	MATCH	MATCH	VALID_FILE	
RS01SBPS-SF01A-3D-SPKIRA101	2018	ATAPL-58341-00004	MATCH	MATCH	VALID_FILE	

githubFile	instrument	ITM_check	googleDrive_check	Oct208Clear	HITLstatus	HITLnotes	fileParse	serialNumber	duplicateCoeff	vendorMatch
ATAPL-70111-00001__20140519.csv	FLCDRA	MATCH	MATCH	NO	NA		SUCCESS_TYPE1	MATCH_SENSORBULK	NONE	['COMPARED']
ATAPL-70111-00001__20151218.csv	FLCDRA	MATCH	MATCH	NO	NA		SUCCESS_TYPE1	MATCH_SENSORBULK	NONE	['COMPARED']
ATAPL-70111-00001__20170425.csv	FLCDRA	MATCH	MATCH	YES	NA		SUCCESS_TYPE1	MATCH_SENSORBULK	NONE	['COMPARED']
ATAPL-70111-00001__20171027.csv	FLCDRA	MATCH	MATCH	NO	NA		SUCCESS_TYPE1	MATCH_SENSORBULK	NONE	['COMPARED']
ATAPL-70111-00002__20131213.csv	FLCDRA	MATCH	MATCH	YES	NA		SUCCESS_TYPE1	MATCH_SENSORBULK	NONE	['COMPARED']
ATAPL-70111-00002__20151218.csv	FLCDRA	MATCH	MATCH	YES	NA		SUCCESS_TYPE1	MATCH_SENSORBULK	NONE	['COMPARED']
ATAPL-70111-00002__20180913.csv	FLCDRA	MATCH	MATCH	YES	NA		SUCCESS_TYPE1	MATCH_SENSORBULK	NONE	['COMPARED']
ATAPL-70111-00003__20140130.csv	FLCDRA	MATCH	MATCH	YES	NA		SUCCESS_TYPE1	MATCH_SENSORBULK	NONE	['COMPARED']
ATAPL-70111-00003__20160331.csv	FLCDRA	MATCH	MATCH	YES	NA		SUCCESS_TYPE1	MATCH_SENSORBULK	NONE	['COMPARED']
ATAPL-70111-00003__20161222.csv	FLCDRA	MATCH	MATCH	NO	NA		SUCCESS_TYPE1	MATCH_SENSORBULK	NONE	['COMPARED']
ATAPL-70111-00003__20171027.csv	FLCDRA	MATCH	MATCH	NO	NA		SUCCESS_TYPE1	MATCH_SENSORBULK	NONE	['COMPARED']
ATAPL-70111-00004__20141021.csv	FLCDRA	MATCH	MATCH	YES	NA		SUCCESS_TYPE1	MATCH_SENSORBULK	NONE	['COMPARED']
ATAPL-70111-00004__20170131.csv	FLCDRA	MATCH	MATCH	NO	NA		SUCCESS_TYPE1	MATCH_SENSORBULK	NONE	['COMPARED']
ATAPL-70111-00004__20171101.csv	FLCDRA	MATCH	MATCH	YES	NA		SUCCESS_TYPE1	MATCH_SENSORBULK	NONE	['COMPARED']
ATAPL-70111-00005__20141021.csv	FLCDRA	MATCH	MATCH	NO	NA		SUCCESS_TYPE1	MATCH_SENSORBULK	NONE	['COMPARED']
ATAPL-70111-00005__20170131.csv	FLCDRA	MATCH	MATCH	NO	NA		SUCCESS_TYPE1	MATCH_SENSORBULK	NONE	['COMPARED']

- Need to reconcile GitHub merge date with CI ingestion date
- CSV format is easily adaptable to provide GUI for user accessibility

Critical Metadata Check Lessons Learned

- **Missing files accounted for majority of calibration errors**
  - ✦ *Curated repo of vendor calibration documentation*
  - ✦ *Verification flag for deployment instances without valid or current calibration files*

- **Scripted entry does not eliminate all errors**
  - Inconsistencies in coefficient resolution and file types
    - ✦ *Verify highest resolution available used for coefficient entry*
  - Typos and erroneous coefficients not always obvious
    - ✦ *Validate vendor files for consistency in dates, headers, ranges of coefficients*

- **Need for consistent gold standards**
  - ✦ *pre/post deployment image database cross-referenced with deployment logs*
  - ✦ *Mining sensor information in raw file during QCT testing and post-deployment*

Ongoing Data Team activities

- **Created Shipboard Discrete Summary template**
  - Formed foundation for adoption of common format working with MIOs to adapt it for cross-program use
  - Summaries being added to Alfresco as they are completed

- **User and Community Outreach**
  - Redmine and Helpdesk tickets
  - Education outreach to undergraduates at UW
    - ✦ 3 volunteer research computing students working on data projects
    - ✦ 3 undergraduate lab assistants
  - Uploading and maintaining documentation on Alfresco for user accessibility

- **Asset management upload and verification for Visions 2019 deployments**

- **Data Management Working Groups**
  - QARTOD, Redmine, Communications, Analysis of Alternatives

Milestones

Phase 1 — Oct 2018 - Jan 2019

- Evaluation of scope and workflow development
- 2i-HITL comparison between GitHub and vendor cals
- Flagged files corrected and re-verified

Phase 2 — Feb - May 2019

- Refinement of workflow and documentation SOPs
- Scripted check of all digital cals
- 2i-HITL check of all pdf cals
- GitHub history crawler

Phase 3 — June - Sept 2019

- Final critical metadata checks
- Annotations and user outreach
- Select sensor deep dives
- Tier One/IRIS metadata
- Deployment image database
- Engineering platforms on GitHub

Milestones

- **Implementation of revised workflows for Visions 2019**
  - Asset management for recovery and deployment of 105 instruments
  - Deployment image database

- **Pursuing RCA-initiated Cabled Playback**
  - Deep profiler data
  - Backfilling data gaps

- **Sensor Deep Dives**
  - Broadband Hydrophones
  - Oxygen Optodes
  - pCO2 calibration ranges

- Final critical metadata checks
- Annotations and user outreach
- Select sensor deep dives
- Tier One/IRIS metadata
- Deployment image database
- Engineering platforms on GitHub