



OpenMFC Evaluation Infrastructure

Lukas Diduch^{*+}

Chocot Baptiste⁺, Ghorbanian Bajgiran Ilia⁺,
Haiying Guan⁺ and Yooyoung Lee⁺

^{*}(CTR) Dakota Consulting, Inc.

⁺ Multimodal Information Group (MIG), Information Access Division (IAD), NIST

OpenMFC2022 Workshop : Day 1, Tuesday, Dec. 6, 2022



Disclaimer

- Certain commercial equipment, instruments, software, or materials are identified in this article in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by NIST, nor is it intended to imply that the equipment, instruments, software or materials are necessarily the best available for the purpose.
- The views, opinions and/or findings expressed are those of the author and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.
- All images, graphs, and charts are original works created for DARPA MediFor as well as other NIST Programs.

Acknowledgement

- NIST contributors
 - Peter Fontana
 - Timothee Kheyrkhah
 - Jesse G. Zhang
 - Edmond Golden
 - Noah Dove

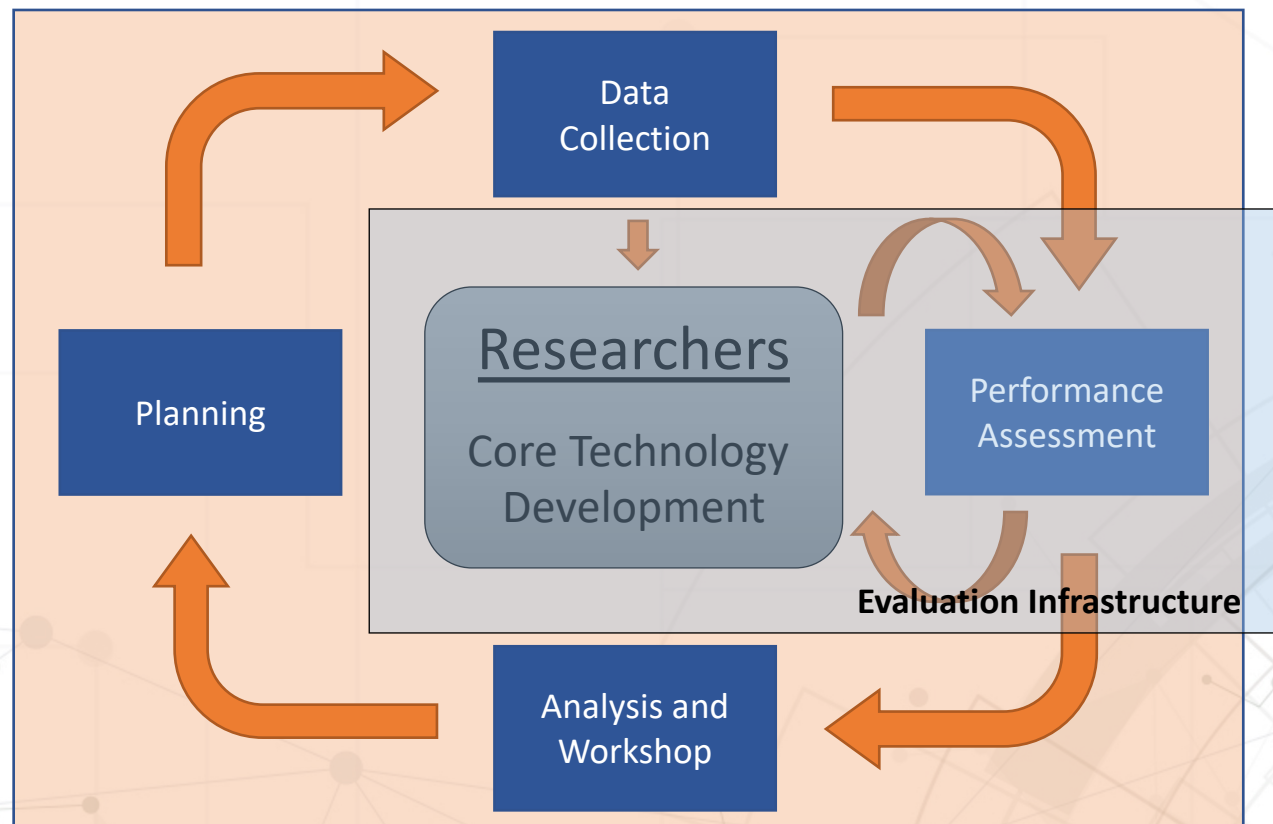
Talk Overview

- Evaluation Driven Research Cycle
- OpenMFC Evaluation Infrastructure Components
 - Overview
 - Public-Facing Infrastructure
 - Website
 - Leaderboards
 - Internal Infrastructure
 - Indus Framework
 - Scoring Code

Motivation

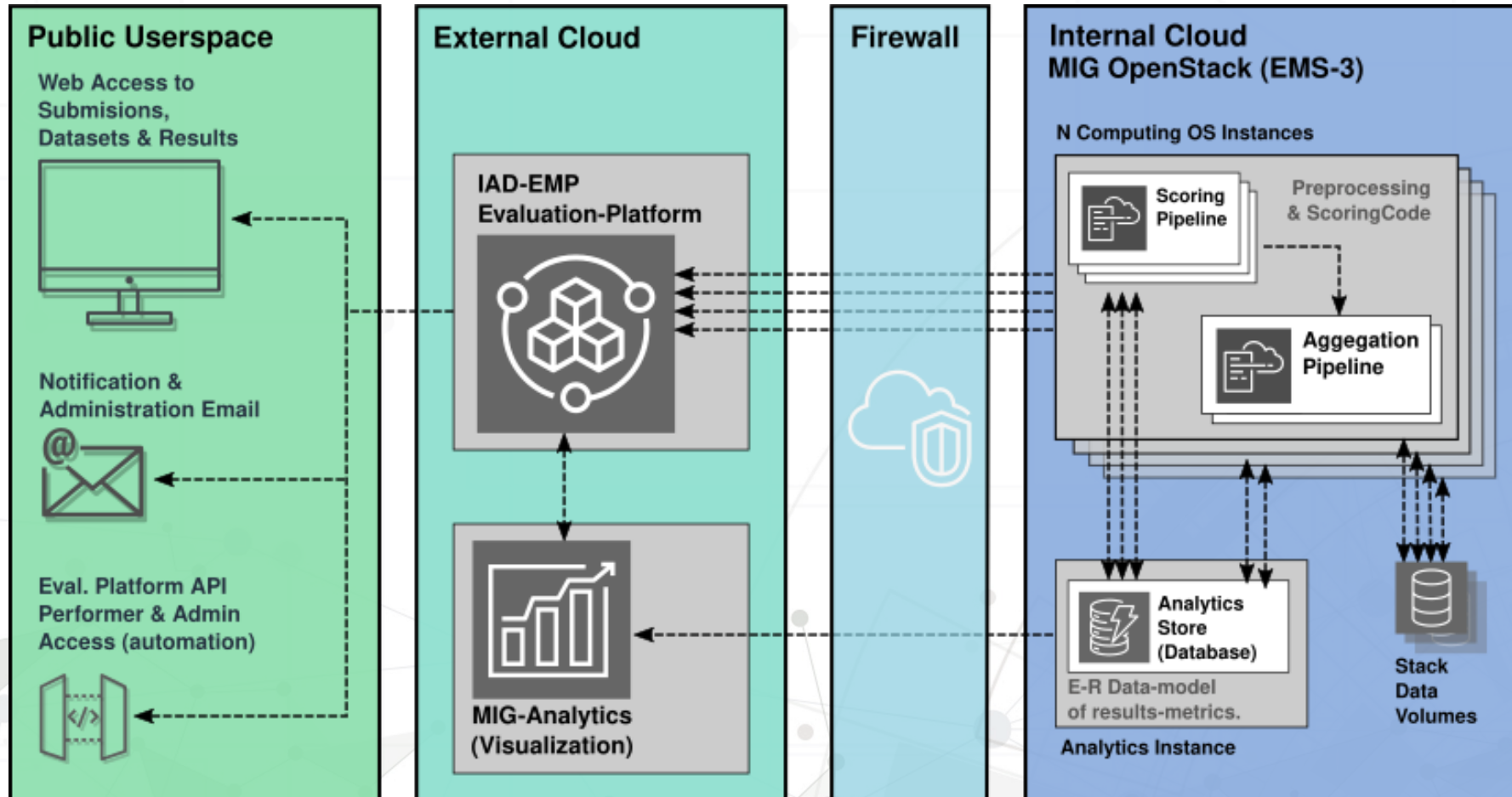
- Evaluations across various domains are conducted at MIG each year
- Conditions and requirements vary across domains:
 - Tracks/Challenges/Phases, Sites/Teams, Licensing, Transfer-Modes
 - Constraints: Data release, Timing, Reporting, Leaderboard
- Need for a tailored but rapidly adaptable Evaluation Management Platform (EMP) while addressing core requirements such as availability, security and usability
- Platform must play along with internal scoring microservices as well as internal and external cloud services

Evaluation-Driven Research



Evaluation Infrastructure Overview

Network Partitioning & Systems Architecture of the Evaluation system



Public-facing Infrastructure

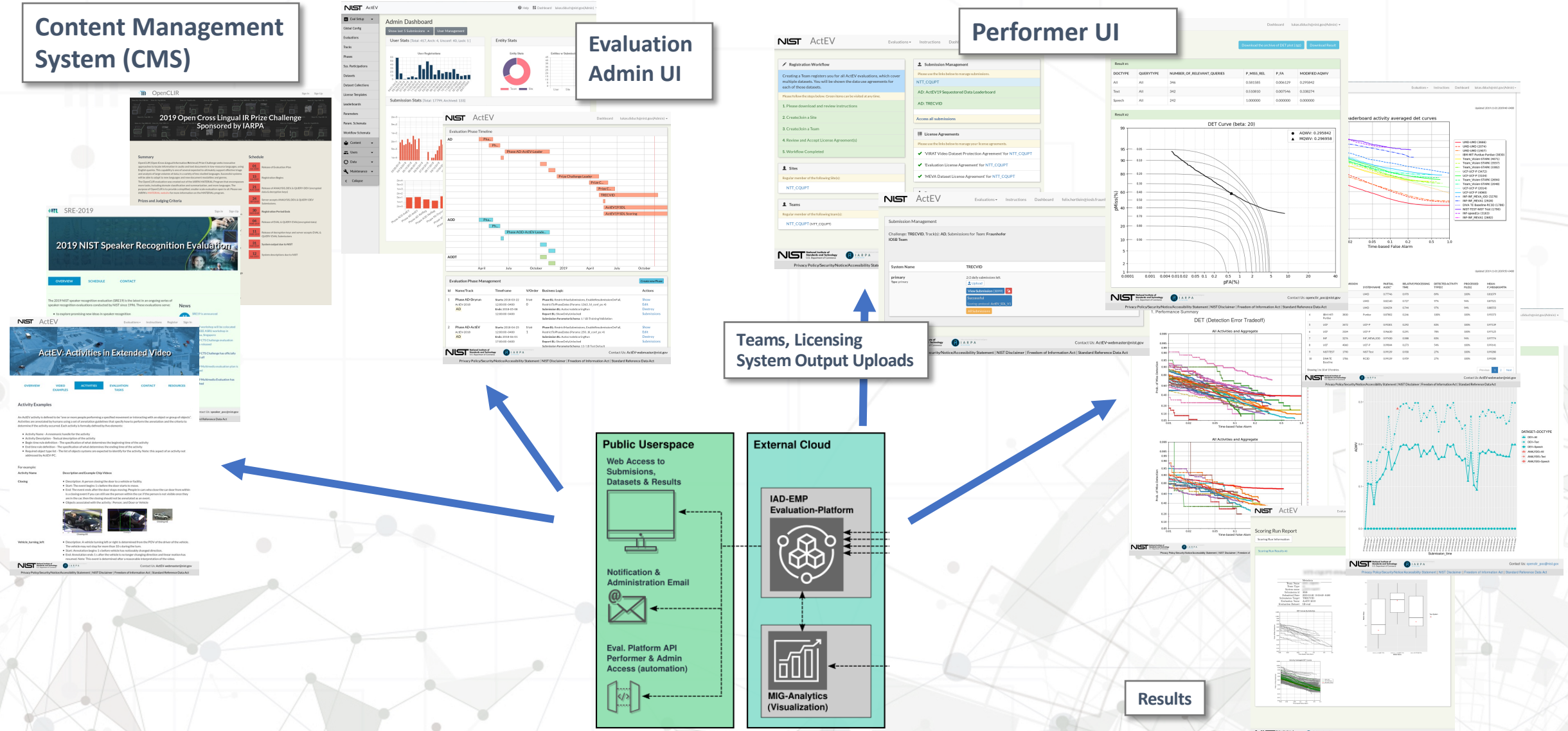
Content Management System (CMS)

Evaluation Admin UI

Performer UI

Teams, Licensing System Output Uploads

Results



Content and Access Management

The screenshot shows the NIST ActEV website with several challenge pages. The top page is 'ActEV: Activities in Extended Video'. Below it is 'OpenCLIR' for the '2019 Open Cross Lingual IR Prize Challenge'. The main focus is 'Open Media Forensics Challenge' (OpenMFC22), which includes sections for Summary, Prizes and Judging Criteria, Task, How to Participate, Questions?, Overview, and News. The News section lists several updates with dates: 01 AUG 22 (Stego data available), 26 JULY 22 (New OpenMFC data download webpage), 03 MAR 22 (OpenMFC2022 Eval Plan), 15 FEB 22 (OpenMFC2021 Workshop Talks and Slides Uploaded!), 7-9 OCT 21 (OpenMFC2021 Workshop), 3 NOV 21 (Sign up OpenMFC slack for workshop discussion), 19 OCT 21 (Evaluation schedule updated), 06 JUL 21 (User Guide for MFC Datasets), 23 APR 21 (New leaderboard launched), 01 MAR 21 (Dataset details updated), and 24 NOV 20 (Please sign up on OpenStack).

OpenMFC21

Email

Password

Password confirmation

(12 characters minimum with at least one special character, capital letter and one digit)

[Sign up](#)

[Log in](#)

[Didn't receive unlock instructions?](#)

[Didn't receive confirmation instructions?](#)



OpenMFC21

Email

Password

[Log in](#)

[Register new account](#)

[Forgot your password?](#)

[Didn't receive unlock instructions?](#)

[Didn't receive confirmation instructions?](#)

Performer Dashboard

General Participant Interface: Home Dashboard

1. Evaluation Workflow
2. Sites & Team Managements
3. Evaluation Tasks & Submission management
4. License agreements
5. Datasets Access

The screenshot shows the Performer Dashboard interface. It is divided into several sections:

- Registration Workflow:** A list of steps from 1 to 6. Step 3, 'Sign and upload license', is highlighted in green and has a blue circle with the number 1 next to it.
- Sites:** A list of sites. 'Academia Sinica' is listed with a blue 'Owner' tag. A blue circle with the number 2 is next to it.
- Teams:** A list of teams. 'UIIA (Academia Sinica)' is listed with a blue 'Owner' tag. A blue circle with the number 2 is next to it.
- Submission Management:** A section with a heading 'Submission Management' and a sub-heading 'UIIA'. It lists several submission categories: 'Image Manipulation Detection and Localization (MFC20)', 'Video Manipulation Detection (MFC20)', 'Video GAN Manipulation Detection (MFC20)', and 'Image GAN Manipulation Detection and Localization (MFC20)'. A blue circle with the number 3 is next to the first item.
- License Agreements:** A section with a heading 'License Agreements' and a sub-heading 'License Agreements'. It lists two license agreements: 'Open MFC2020 Evaluation Registration' for Academia Sinica and 'Open MFC2020 Data Use Agreement' for Academia Sinica. A blue circle with the number 4 is next to the first item.
- Datasets:** A section with a heading 'Datasets' and a sub-heading 'OpenMFC Datasets'. It lists 'OpenMFC Datasets'. A blue circle with the number 5 is next to it.

The NIST logo is visible at the bottom left of the dashboard, and the contact information 'Contact Us: mfc_poc@nist.gov' is at the bottom right.

Phases, Systems, Submission Management

- **Tracks** represent an evaluation task
- **Phases** represent stages of a Track across a time-period
- **Systems** represent different system instances/ implementations (e.g. training sets or system parameters etc.). Systems and Phases form a Matrix
- **Submissions:** System Output to be scored against sequestered test dataset

NIST ActEV

Evaluations ▾ Instructions Dashboard Participant ▾

Submission Management

Challenge: TRECVID,
Track(s): AD, Submissions for Team: NIST-TEST

[Add new system](#)
2 systems left.

System Name	TRECVID	TRECVID20	TRECVID21
Test System I Type: primary	Phase closed, Submissions Disabled. View Submission (3806) FAIL-scoring Scoring-protocol: ActEV_SDL_V1 All Submissions	Phase closed, Submissions Disabled. View Submission (22841) Original FN: 42_zipbomb.zip FAIL-uncompress Scoring-protocol: ActEV19_AD All Submissions	Phase closed, Submissions Disabled. View Submission (23863) Original FN: baselineACT_1_AD.tgz FAIL-generate_report Scoring-protocol: ActEV19_AD All Submissions
Test System II Type: primary	Phase closed, Submissions Disabled. View Submission (1741) FAIL-scoring Scoring-protocol: ActEV19_AD FAIL-scoring Scoring-protocol: ActEV18_AD All Submissions	Phase closed, Submissions Disabled. View Submission (19749) Original FN: systemnist.zip FAIL-scoring Scoring-protocol: ActEV_SDL_V2 All Submissions	Phase closed, Submissions Disabled. All Submissions

Submissions

- **UI Generation : Submission Form example**
- Customized evaluation parameters stored as Json-schemas
- Parsed & Validated in ruby to generate the HTML & JavaScript interface

Form Features:

- Form elements (text input, drop-down list, multiple selections, etc.)
- Conditional forms (mutual exclusions between form inputs)

New Submission

Select submission parameters if applicable. Use the 'Browse' button to select a file from your computer and upload it to the server by clicking on 'Submit'.

System: test
Evaluation: OpenMFC 2020
Track: Image Manipulation Detection and Localization
Phase: MFC20
Allowed file formats: .url

Dataset

OpenMFC20 Image

Input condition

Image only

Evaluation protocol

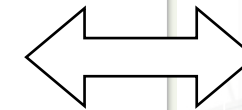
Detection
Localization

Specify Submission URL

Enter URL to fetch ressource from here.

Submit

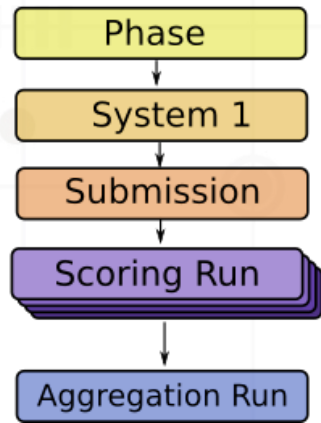
Back



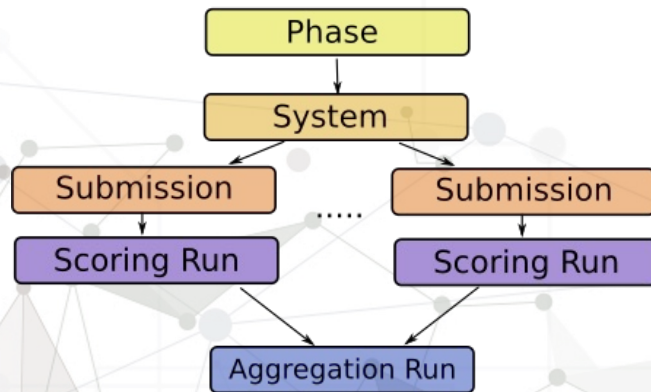
```
{  
  "type":"object",  
  "required":[  
    "Dataset",  
    "Evaluation Protocol",  
    "Input Condition"  
  ],  
  "properties":{  
    "Dataset":{  
      "type":"string",  
      "enum":"nil"  
    },  
    "Evaluation Protocol":{  
      "type":"array",  
      "items":{  
        "enum":[  
          "Detection",  
          "Localization"  
        ]  
      },  
    },  
    "Input Condition":{  
      "type":"string",  
      "enum":"nil"  
    }  
  }  
}
```


Submissions, Scoring Runs, Aggregation Runs

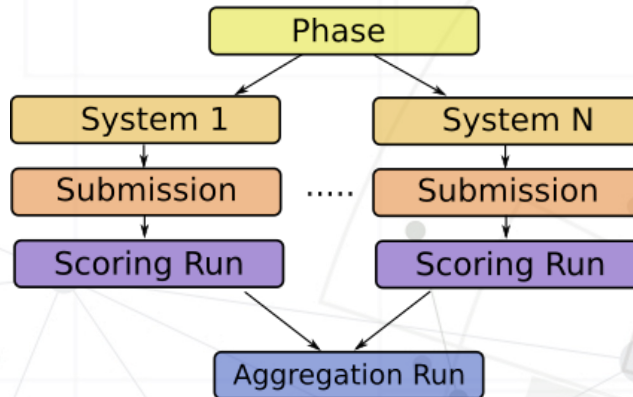
Across Scoring-Runs
(one System, one Submission)



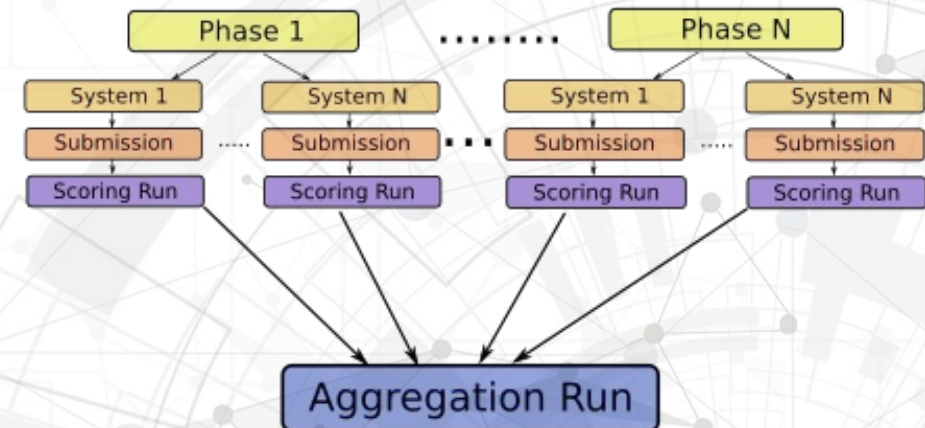
Across Submissions
(one System)



Across Submissions
(multiple Systems)



Across Phases
(multiple Systems,
multiple Phases)



- System Output is considered a **Submission**
- Multiple **Scoring Runs** can be run against a submission
- **Aggregation Runs** can be run against a set of submissions or scoring-runs, or aggregation-runs and merge them together for comparison

Scoring-Run Report

Scoring Run Report

Show Scoring Run Information

Results

QUERY | TRR | SYS_RESPONSE | AUC | EER | FAR_STOP | AUC@FAR | CDR@FAR | CI_LEVEL | AUC_CI_LOWER | AUC_CI_UPPER | AUC_CI_LOWER@FAR | AUC_CI_UPPER@FAR | CDR_CI_LOWER@FAR | CDR_CI_UPPER@FAR

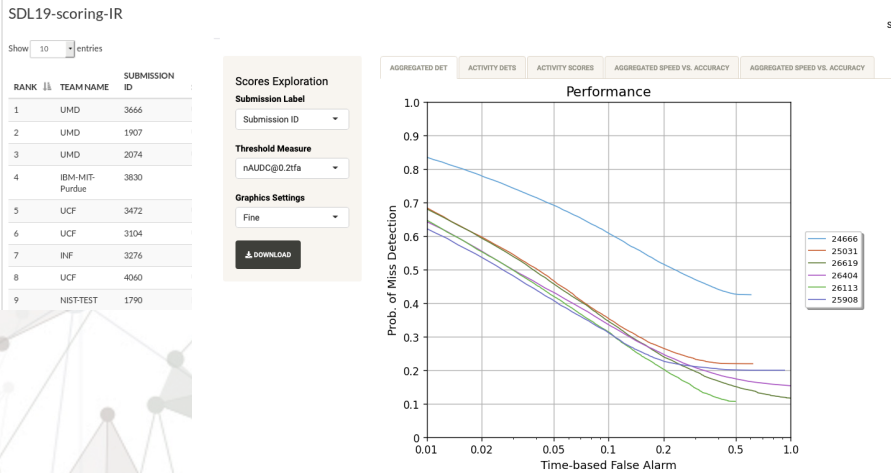
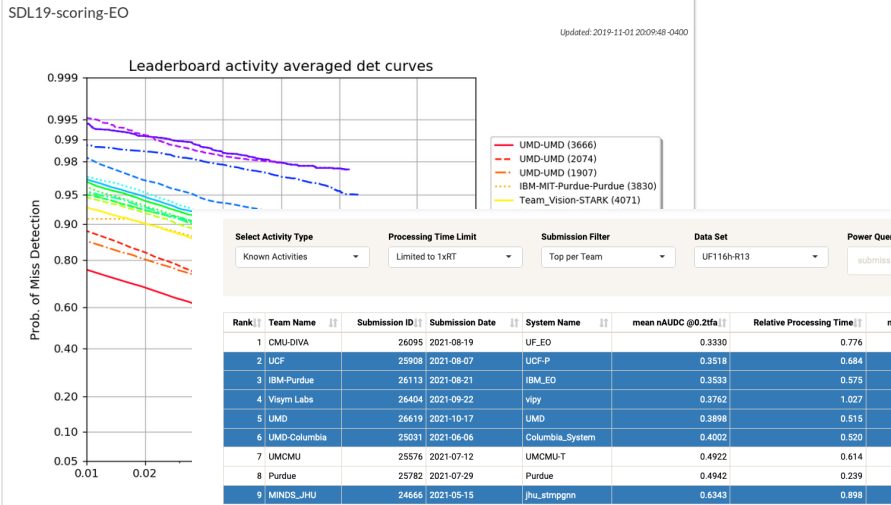
TaskID == ['manipulation'] | 1.0 | all | 0.616186 | 0.396983 | 0.05 | 0 | 0.071351 | 0.9 | 0.609714 | 0.622467 | 0 | 0 | 0.069902 | 0.072862

Few minutes after teams make a submission they can see scores / error-logs associated for their submission.

The collage displays several key components of the scoring system interface:

- Resources monitoring report:** Three line graphs showing CPU usage, Free GPU Memory, and Free Memory over time.
- All Results:** A table listing submission results with columns for DOCTYPE, QUERYTYPE, NUMBER OF RELEVANT QUERIES, P_MISS_REL, P_FA, and MK.
- DET Curve (beta: 20):** A plot of AOVY (y-axis, 0.0 to 0.4) versus pFA(%) (x-axis, 0.001 to 10). It shows multiple curves for different submissions and a highlighted curve for the current submission.
- Scoring Run Report:** A log of system commands and their outputs, including file paths, directory listings, and execution results.

Leaderboards



OVERVIEW TASKS DATA SCHEDULE LEADERBOARD RULES RESOURCES CONTACT

IMD ISMD VMD **IDD** VDD STEGD

Some systems on the leaderboard are built with training data (with reference ground-truth information known), we will report training data and testing data performance separately in the NIST evaluation report. We would not recommend direct comparisons.

Image Deepfakes Detection (IDD) Updated: 2022-11-23 09:18:03 -0500

Previous 1 2 Next

RANK	SUBMISSION ID	SUBMISSION DATE	TEAM NAME	SYSTEM NAME	AUC	CDR@0.05FAR	AVERAGE OPTIMAL MCC
1	90	2021-07-10 09:56:14	UIIA	test	0.689716	0.207018	
2	93	2021-07-30 18:13:11	UIIA	test	0.683956	0.187125	
3	75	2021-06-14 04:12:30	UBMDFL_IJMD	dry-run	0.554261	0.012865	
4	52	2021-06-01 15:41:20	UBMDFL_IJMD	dry-run	0.547125	0.009357	
5	82	2021-06-23 09:21:26	UIIA	pitchat	0.500033	0.051077	
6	36	2021-05-06 07:32:03	UIIA	test	0.5	0.05	
7	91	2021-07-21 15:53:47	UIIA	test	0.478445	0.009357	
8	86	2021-07-07 06:59:01	UIIA	test	0.41193	0.00117	
9	87	2021-07-07 07:04:14	UIIA	test	0.403957	0.004678	
10	89	2021-07-08 04:20:47	UIIA	test	0.398674	0.003509	

Showing 1 to 10 of 16 entries

Download CSV Updated:

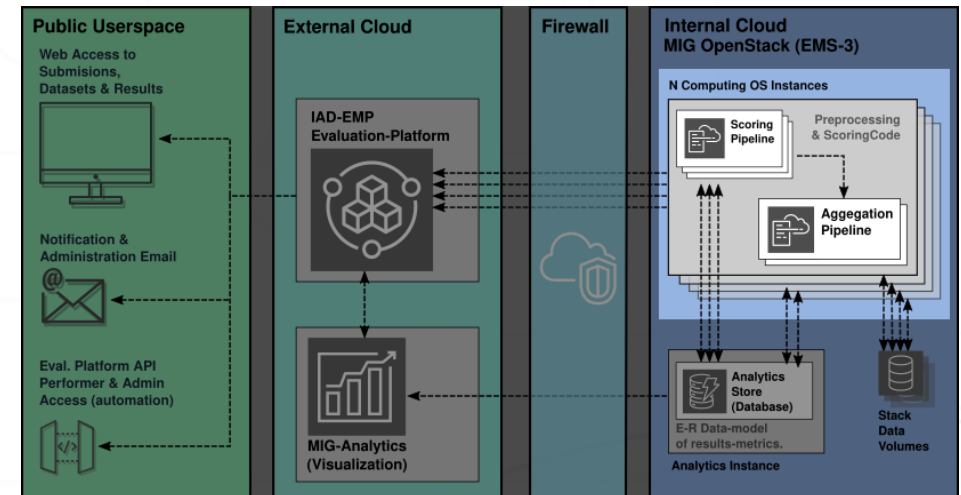
Aggregated ROC Curves

- Manipulation Detection (MD)
 - Image_MD (IMD)
 - ImageSplice_MD (ISMD)
 - Video_MD (VMD)
- Deepfake Detection (DD)
 - Image_DD (IDD)
 - Video_DD (VDD)
- Steganography Image Detection (StegD)

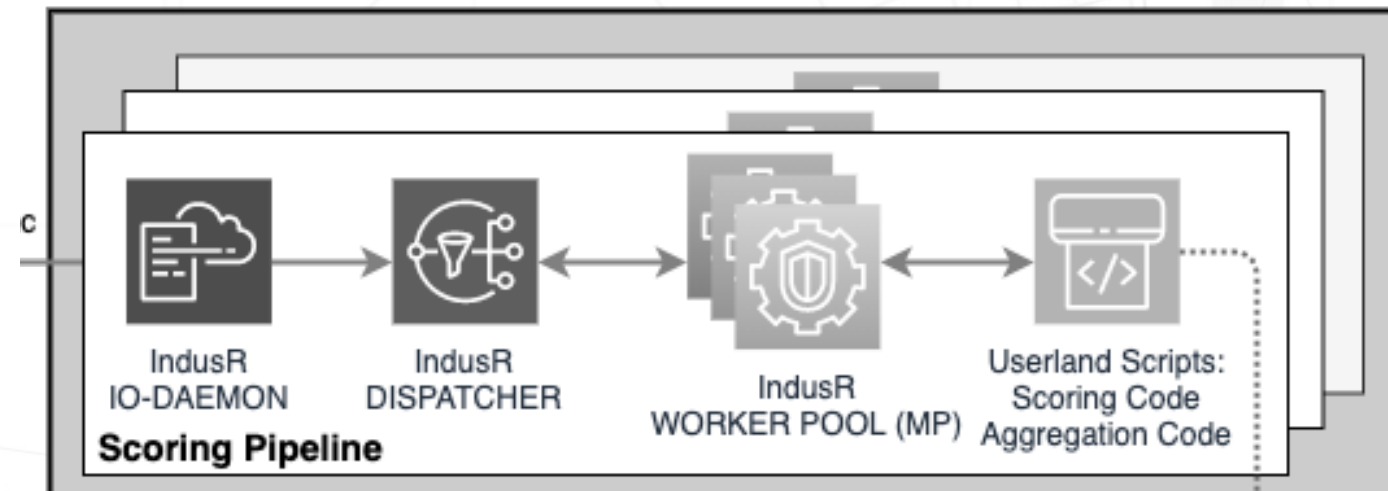
IndusR Backend

Software overview

- IndusR is a *ruby command line tool* for running pipeline jobs in concurrent and distributed environments
- Each pipeline is defined by the user's **environment, config file** and associated set of **scripts**
- The config file is used to fully describe the pipeline **parameters**, which includes server setup, sequence of steps with their respective scripts and associated hooks and hook parameters



OpenStack Instance



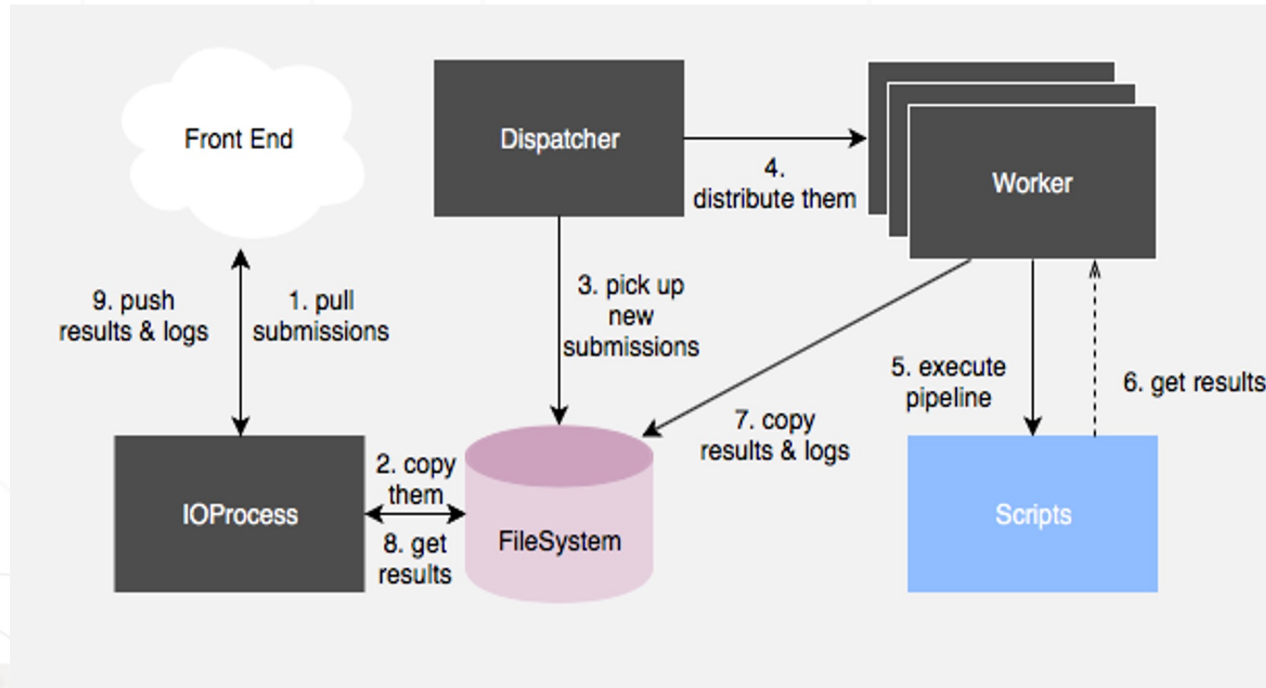
IndusR: Configuration Driven

Pipeline Config File :

```
evalid: MFCBackend
redis: "redis://localhost:6379"
dispatch_type: scoring_run
eval_root_dir: !ENV INDUS_BE_EVAL_ROOT_DIR
io_interval: 45 # seconds
pipeline:
  setup: step_setup.sh
  download: step_download.sh
  uncompress: step_uncompress.sh
  validate: step_validate.sh
  detscore: step_detscore.sh
  locscore: step_locscore.sh
  updatedb: step_updatedb.sh
hooks: [...]
[paths...]
```

- **Additional Scoring Pipeline Hooks :**
- **IO Hooks**
 - RSYNC/SSH: push, pull
 - REST API: push, pull + complex query for pull
- **POST step success/fail Hooks**
 - bind a script to execute based on step condition
- **Configurable Log-scrubbing**
 - Define scripts filtering out sensitive information and excessive detail

IndusR Components



- **I/O process:** automatically retrieve and upload submissions/scoring- and aggregation-runs/status/results between WebUI and Scoring Cluster either via SSL or REST API
- **Dispatcher process:** queue incoming runs
- **Worker processes:** Execute single step of a step-sequence expressed as a scoring- or aggregation pipeline running against the performers submission. Workers can be bound to individual steps
- **Redis:** Key-value store to synchronize application state across all distributed/parallel processes
- **Scripts:** Userland scripts written by evaluators

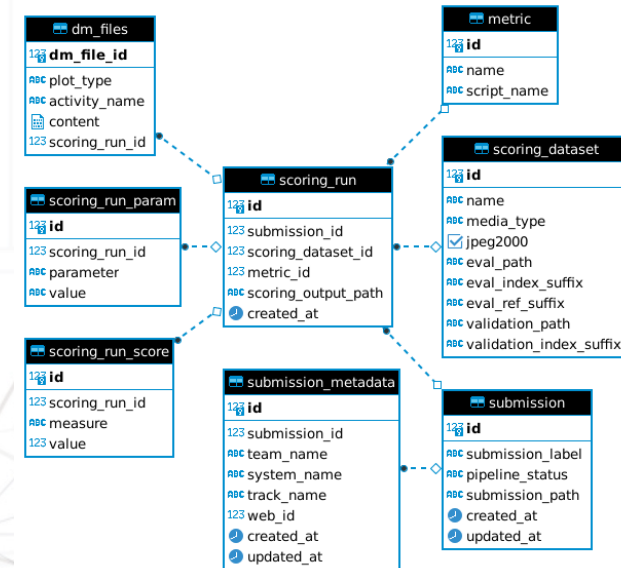
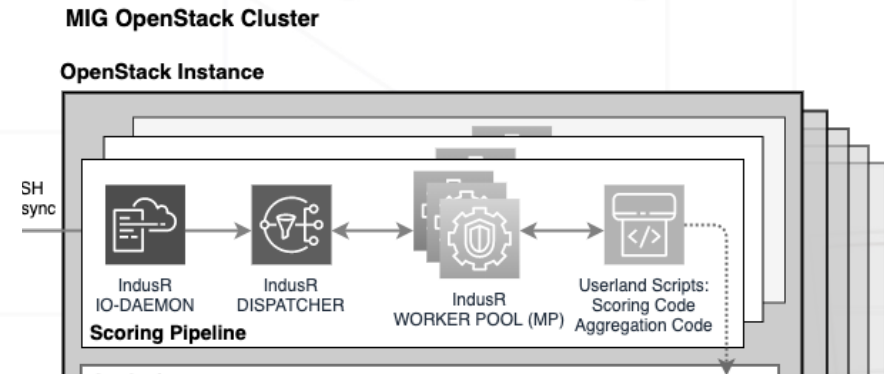
OpenMFC Implementation

Scoring Pipeline Implementation:

- One scoring pipeline handles all eval Tracks
- One aggregation pipeline handles all Leaderboards

Configuration Driven mainly through Database:

- Database stores
 - **Track configuration**
 - Track scores
- Leaderboards can use DB for aggregation or as a source for complex analysis online and offline
- Advantages:
 - Adding new tasks and datasets quickly to the evaluation logic is straight-forward
 - Database w/ Scores can be hosted anywhere
- Disadvantages
 - Advanced setup



Conclusion

- As an **independent Scoring Entity** we are providing
 - Fair **comparison of systems performance** across community peers
 - **Standardized metrics** against **sequestered Dataset(s)**
- **Our Evaluation Infrastructure** is
 - Providing all **essential evaluation resources** online
 - Facilitating and **managing scoring process** within eval-constraints
 - Providing **tracking of progress** across different modalities
 - Providing scoring computation resources (for open evaluations)
- Leaderboard based **evaluation cycle** enables **rapid R&D**

Questions?

OpenMFC team: mfc_poc@nist.gov



Thank You!