Video Linking and Intelligence from Non-Collaborative Sensors (Video LINCS) Proposers' Day

Dr. Reuven Meth | Program Manager | February 7, 2024

Intelligence Advanced Research Projects Activity

IARPA

Creating Advantage through Research and Technology

Video LINCS

Welcome to the Video LINCS Proposers' Day

Thank you for your interest in this program and participating in this event

To assure a clear broadcast stream, audio and video are disabled for meeting participants

Comments and questions can be submitted to the IARPA team via the WebEx chat tool submission or via index cards for in-person attendees

- Please direct questions to "All Panelists" in the chat if you are virtual

Questions submitted to the alias (dni-iarpa-videolincs-proposersday@iarpa.gov) prior to this meeting and during this presentation, and corresponding answers, may be posted in writing online

Disclaimers

- This presentation is provided solely for information and planning purposes
- The Proposers' Day does not constitute a formal solicitation for proposals or proposal abstracts
- Nothing said at Proposers' Day changes the requirements set forth in a BAA
- **The BAA language supersedes anything presented or said by IARPA at the Proposers' Day**
- This meeting is being recorded and will be posted for public viewing
- For those viewing the recording, email aliases and POCs may be dated. Please refer to IARPA.gov for updated information.

Proposers' Day Goals

1. Familiarize participants with IARPA's interest in the Video LINCS program and solicit questions and feedback
2. Foster discussion of complementary capabilities among potential program participants, i.e., TEAMING
   - Teaming information can be found at the following address: **https://www.iarpa.gov/index.php/research-programs/video-lincs**
   - An attendance list, with contact information of participants who approved of sharing will be distributed soon
   - The chat feature is enabled for participants to plan future discussions associated with teaming
   - Teaming interests, capability summaries, and lightning talk slides will be posted publicly on the IARPA Video LINCS webpage until the BAA submission period closes

Please ask questions and provide feedback, this is your chance to alter the course of events.
Please talk with others, find great team members.

Feedback and Questions

- Questions can be submitted <u>until 11:00am ET</u>
- There will be a break after the contracting presentation at 11:00am ET
- Responses to selected questions will be broadcast at 12:30pm ET, so please don't log out or close your WebEx connection
  - All programmatic and contractual questions will be captured but will not be answered in this session
- Feedback (but not questions) about the draft technical section may be submitted to the IARPA team email at dni-iarpa-videolincs-proposersday@iarpa.gov
  - A new alias will be established when the full BAA is released
- After this Proposers' Day, IARPA will review all the feedback received for a final BAA to be posted on SAM.gov and Grants.gov

Teaming

- Participants are encouraged to find partners and collaborators . . . someone might have a missing piece of your puzzle
- Lightning talks will take place following the Program presentations
- Collaboration and capability summaries will be accepted, with minimal review for appropriateness, and made available to the public
  - Teaming documents and summaries can be submitted until the BAA closes. Submit to dni-iarpa-videolincs-proposersday@iarpa.gov
  - If you would prefer your information not be shared (any recorded videos cannot be modified or removed) email dni-iarpa-videolincs-proposersday@iarpa.gov

Agenda

Time	Topic	Speaker
9:30 a.m. - 9:40 a.m.	Welcome, Logistics, Proposers' Day Goals	Reuven Meth, Program Manager
9:40 a.m. - 9:50 a.m.	IARPA Overview	Robert Rahmer, Director Office of Analysis Research, IARPA
9:50 a.m. - 10:40 a.m.	Video LINCS Program Overview	Reuven Meth
10:40 a.m. - 11:00 a.m.	Contracting Overview	IARPA Contracts
11:00 a.m. - 12:30 p.m.	Break (**Submit questions in Webex chat before 11:00 a.m.**)	
12:30 p.m. - 1:30 p.m.	Answers to Selected Technical Questions	Reuven Meth
1:30 p.m. - 1:45 p.m.	Break	
1:45 p.m. - 3:15 p.m. (est.)	Lightning Talks*	Potential Performers
3:15 p.m. - 4:30 p.m.	Informal Teaming Discussions*	In-Person Participants

*The Government will not attend these events

LIGHTNING TALKS AGENDA (1)

Time	Speaker	Institution	In person
1:45 p.m.	Lance Besaw	Applied Research Associates	YES
1:50 p.m.	Colin Adams	Arete	YES
1:55 p.m.	Naresh Cuntoor	Blue Halo	YES
2:00 p.m.	Rodolfo Valiente Romero	HRL Laboratories	YES
2:05 p.m.	Scott McCloskey	Kitware	YES
2:10 p.m.	Derek Walvoord	L3Harris	YES
2:15 p.m.	Xiaoming Liu	Michigan State	YES
2:20 p.m.	Meryl Spencer	Michigan Tech Research Institute	YES
2:25 p.m.	Nigel Mathes	MORSE Corp	YES

LIGHTNING TALKS AGENDA (2)

Time	Speaker	Institution	In person
2:30 p.m.	Kemal Davaslioglu	Nexcepta	YES
2:35 p.m.	Jia Deng	Princeton U.	YES
2:40 p.m.	Ashley Antonides	Two Six Technologies	YES
2:45 p.m.	David Brady	U. of Arizona	YES
2:50 p.m.	Tejas Gokhale	UMBC	YES
2:55 p.m.	Yu-Wing Tai	Dartmouth College	NO
3:00 p.m.	Jeff Benson	DejaVuAI	NO
3:05 p.m.	Bar Cohen	Reichman U.	NO
3:10 p.m.	Mauricio Pamplona Segundo	U. of South Florida	NO

Video LINCS Overview

Intelligence Advanced Research Projects Activity

IARPA

Creating Advantage through Research and Technology

Technical Slides Disclaimer

- All images, references, and articles are included as illustrative examples only
- ODNI and IARPA do not endorse any product or company referenced within
- Changes may occur before the final released BAA is released

Security video sensors abound …

- In today's world, video sensors are ubiquitous
  - Aerial, building-mounted, ground-based
  - Stationary, mobile
  - Indoor, outdoor
  - …

... but are not always effective

Capitol Police had cameras outside Pelosi home that weren't monitored at break-in: Sources

USCP saw police lights, rewound the video and saw the break-in, sources said.

By Rachel Scott, Jack Date, and Luke Barr
November 2, 2022, 4:21 PM

ABC news

attention span of a human watching monitors drops 50% in 20 minutes then precipitously thereafter ipvm.com

"59K hours of video in an active shooter forensic analysis" briefcam.com

"...20 minutes a human can watch video before their attention span falls apart... can't spend much more than 5 actually doing it"
Milestone Integration Platform Symposium

Effective-247-cctv-monitoring

IS EFFECTIVE 24/7 CCTV MONITORING REALLY POSSIBLE

The human brain has limited attention capabilities. A study found that after 20 minutes, CCTV operator watching a video scene will miss up to 95 percent of all activity.

Much video, even when taken for security objectives, is not viewed

Video Sensors are Typically Non-Collaborative

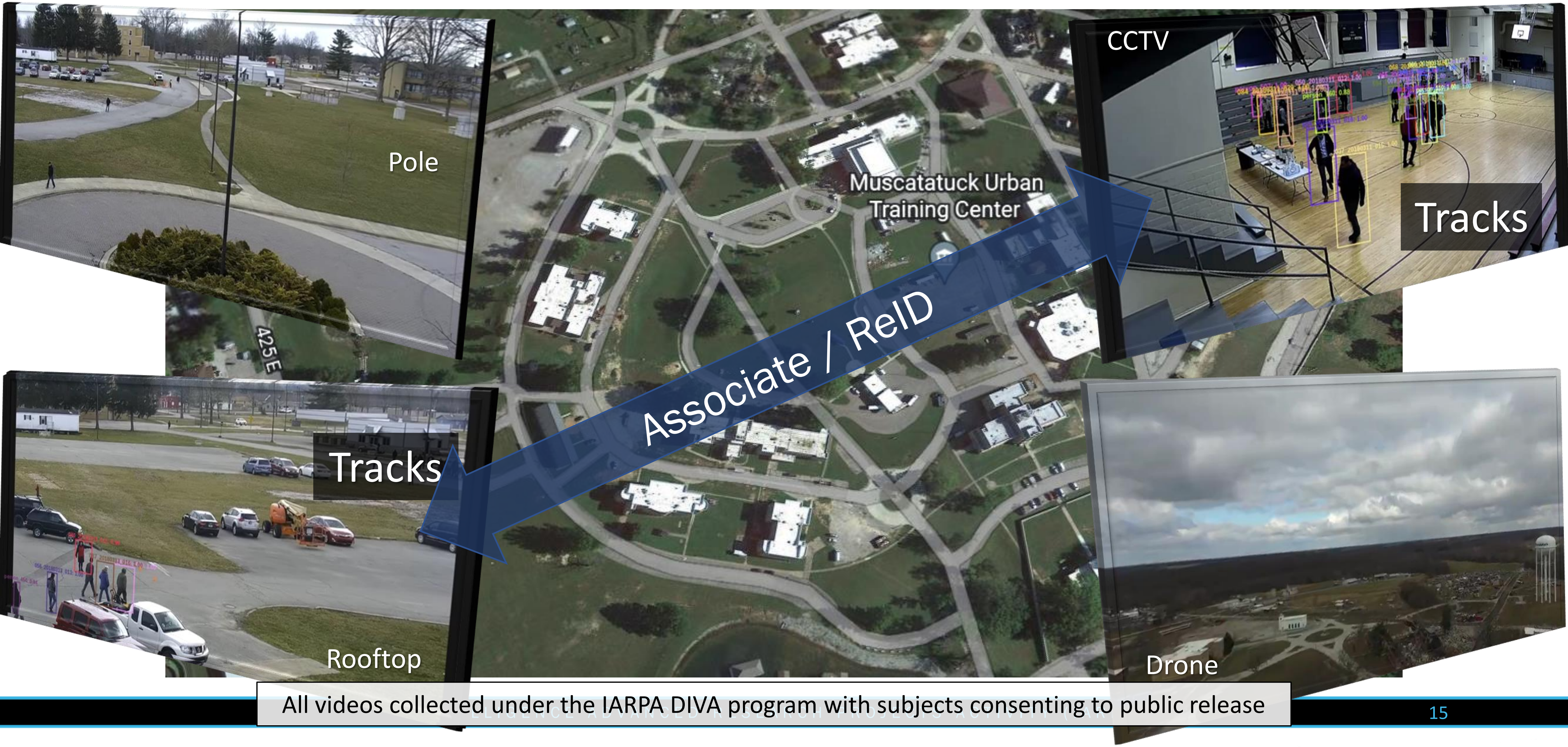
- Footage is available from many disparate video sensors that are largely independent and do not collaborate with each other
  - Collaborative video sensors
    - Cameras work together towards a specific goal
    - Content of one camera's imagery is used to cue another camera
    - Collaborative video technologies previously developed, e.g. DARPA VSAM, VIVID
  - Non-collaborative video sensors
    - Cameras are independent and do not work together
    - Typical scenario for many video sensor collections

- Automatically associating data across non-collaborative sensors presents significant technical challenges

Video LINCS Objectives

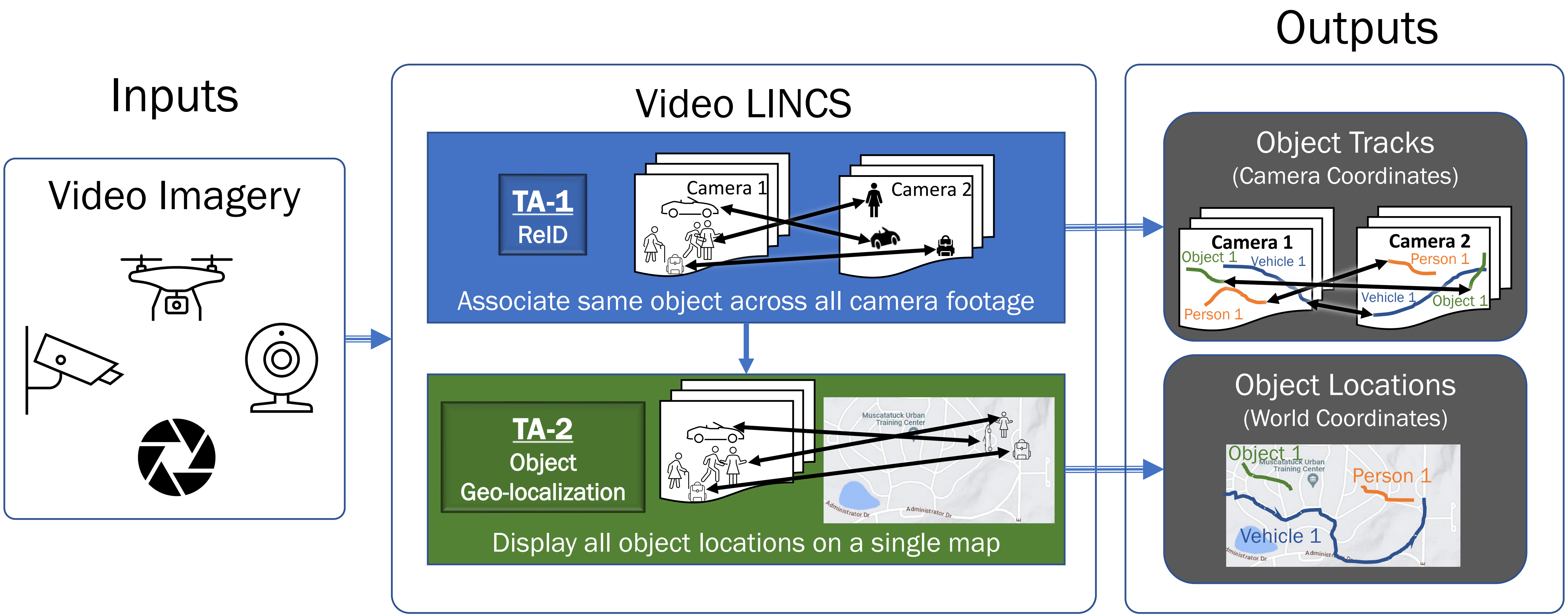
- Re-identification (ReID) - Autonomously associate all instances of the same object across all video footage in a collection
  - Across space
  - Across time
  - Across sensor types and collection geometries
- Autonomously transform object locations to a common reference frame
- Provide a gestalt view of content, with the ability to rapidly
  - View objects motion patterns in a unified world reference frame
  - Understand motion patterns across space and time
  - Analyze patterns for anomalies and threats

Concept



All videos collected under the IARPA DIVA program with subjects consenting to public release

Video LINCS Overview



Video LINCS will provide geolocated, time-stamped, tracks of objects across open-world, video footage

Why is Video LINCS so difficult?

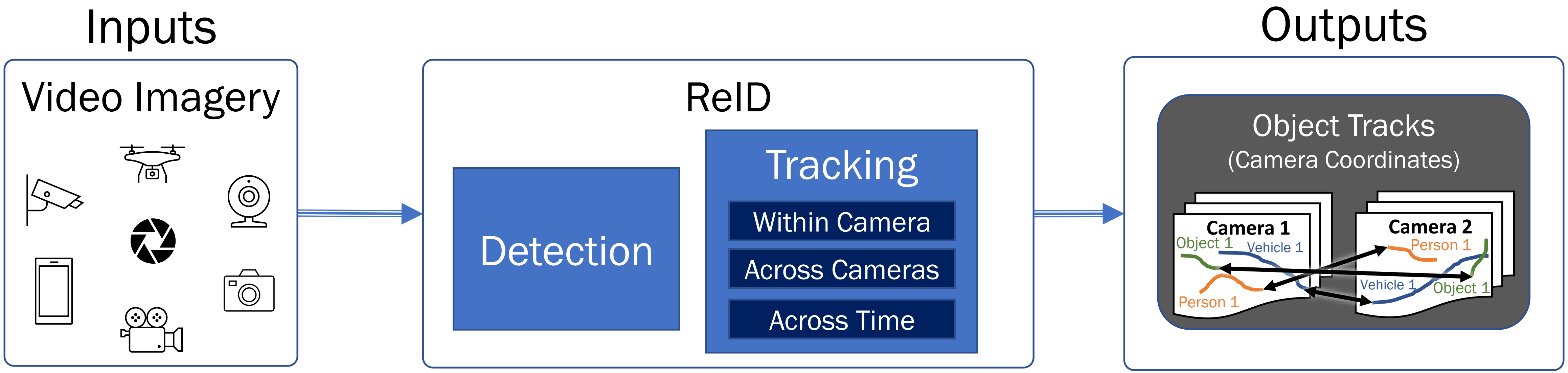
- Successful autonomous video reID requires research innovations in
  - Detection – Finding all objects across scale, aspect, density/crowding, obscuration, sensors
  - Matching – Associating all instances of the same object – within and across sensors, collection geometries, appearance variations
  - Geo-localization – Automatically transforming object positions to a geo-coordinate system, often using inexact or unavailable sensor location & pose (general vicinity typically known)
- Neither query[1] nor gallery[2] are available in advance (open-world)
- Sensor/video/object characteristics are not known *a priori* and can vary significantly across a collection

[1] Query: Object instance being looked for
[2] Gallery: Set/library of objects being matched to

Video LINCS will provide

- ReID for objects (open-world)
  - Across different video sensor types, without prior availability of annotated data from those sensors
  - Across time, robust to changes in external appearance
  - On raw video footage, without *a priori* detections of objects
  - Application to people, vehicles and generic objects in footage
- Complete tracks (not merely retrieval), generated entirely by automated process (including queries and galleries)
- Unified presentation in a common coordinate frame (geo-localization)

Video LINCS will research and implement novel technologies to perform autonomous, open-world, geo-localized reID

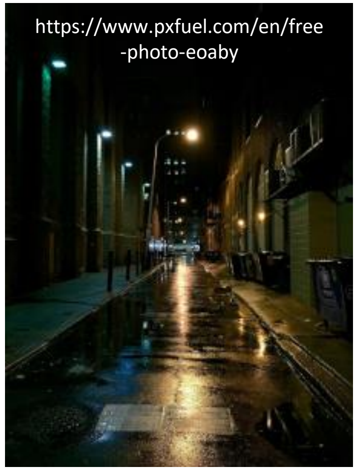


- TA1 Innovations
  - Detect all objects - Across scale, orientation, pose, crowding, occlusion, noise
  - Track all objects
    - Within camera - Retain object in presence of confusers, occlusions, diss/re- appearance, jitter
    - Across cameras - Varying locations, geometries, conditions (resolution, viewpoint, aspect, lighting, sensor)
    - Across time - Varying attributes (e.g. clothing, accessories, face masks, imaging conditions)
- Outputs – {video ID (vID), frame ID (fID), object ID (oID), row (r), column (c), Height (H), Width (W)}

ReID Challenges

- Different video sources, conditions, collection geometries
  - Ground-based, building/roof/pole mounted and aerial
  - Stationary and mobile
  - Outdoor and indoor
  - Day and Night
  - Varying weather
  - Temporally proximate and distant
  - Noise/artifacts/corruption


https://www.pxfuel.com/en/free-photo-empyn


https://www.pxfuel.com/en/free-photo-eoaby

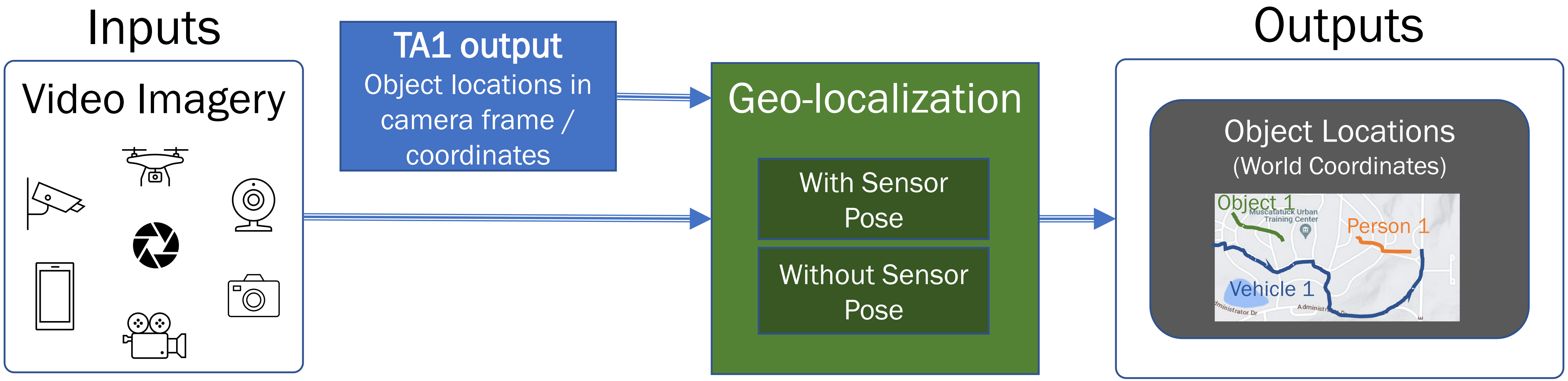
- Arbitrary collections with no *a priori* galleries of objects
- No prior representative data from sensors that collected the video
- Faces, license plates predominantly unavailable
- End-to-end solution (video input, reID output)

ReID Under Temporary Attributes Changes

Temporally distant

- Temporary attributes change (e.g. clothing, hats, glasses)
- Same person has significant external appearance variations
- Phase 2 objective

All images from D. Davila et al, "MEVID: Multi-view Extended Videos with Identities for Video Person Re-Identification", Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), January 2023, and may also be found in the public MEVA dataset at mevadata.org



- TA2 Innovations
  - With sensor pose – Refine mapping to provide accurate coordinates impervious to sensor pose errors
  - Without sensor pose – Utilize auxiliary data (e.g. maps, building layout) to determine object positions in unified coordinate system
- Outputs – Object time-stamped geo-locations {object ID (oID), timestamp (UTC), longitude (long), latitude (lat), altitude (alt)} (e.g. KML / Google Earth compatible)

Geo-localization Challenges

- Transform from image coordinates (different for each camera) to world coordinates
  - Provide unified view for all objects independent of camera that collected imagery
  - Geo-locate with noisy or unavailable sensor pose (general location known)
  - Estimation of dynamic/mover location for non-planar objects
- Autonomous backprojection (camera to world coordinates)
  - No person-in-the loop (e.g. for homography computation)
  - Deal with non-planar content
- Differing levels of camera metadata and reference imagery
- Geolocalization to be evaluated
  - as standalone module
  - for end-to-end reID system – errors in camera frame detection location may propagate

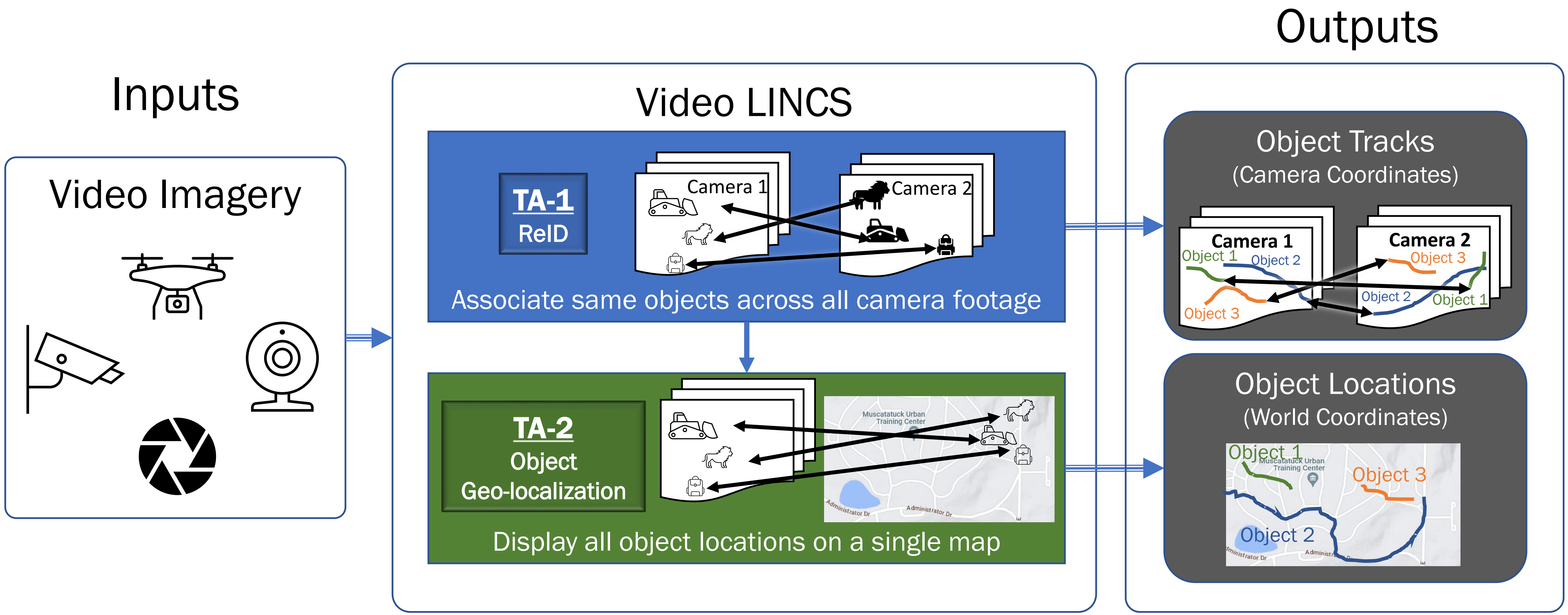
Additional Challenges

- Autonomous operation
  - Open-world
  - Not reliant on faces (people) / license-plates (vehicles)
  - Exhaustive (not merely retrieval)
- ReID under partial/significant occlusion
- ReID for objects in groups / close proximity
- Non-collaborative sensors
  - Autonomous at test time, without prior information or training data from video sensors that obtained the imagery
  - General operation to work with data from new, unseen video sensors
  - Work across collection conditions (resolution, aspect, viewpoint, lighting, quality, ...)
- Generic objects – no prior information on object to be re-identified

Generic Object ReID

- ReID capability to be developed for generic objects, where there is no knowledge of the specific object type to be re-identified at training time
- Generalized modeling, beyond class specific approaches
- Re-identified tracks of all foreground objects in collection
  - No prior information on what is an object
  - No prior information on foreground vs. background
  - Generic object reID capability should work for all objects the system is able to reidentify, without size limits (large or small)

Generic Object ReID



ReID for all foreground objects in video without prior knowledge of object types present in foreground

Relevant Expertise

- Artificial Intelligence
- Computer vision, including object detection, tracking, person/vehicle/object modeling, generic vision learning
- Deep learning
- Image and video geo-localization
- Geometric camera projections and inverse projections
- Machine learning
- Modeling and simulation
- Open set classification
- Re-identification
- Soft biometrics
- Software engineering
- Software integration
- Systems integration
- Truthed video data collection, annotation (truthing in both anonymized identities and geo-location), including potential HSR
- Vehicle fingerprinting
- Video data generation (including simulation, generative modeling)

Out of Scope

- Research of non-visible image/video modalities (e.g. audio, emissive infrared)
- Development of hardware or sensors
- Research on face recognition. Existing face recognition technologies may be employed, but there should be no expectation that recognizable faces will be present in the video.
- Research on license plate recognition. Existing license plate recognition technologies may be employed, but there should be no expectation that recognizable license plates will be present in the video.
- Video from space-based platforms
- Methods that require a human in the loop

Video LINCS Phases

Phase 1
Feasible

TA1 - ReID
- People
- Temporally proximate (people - same clothing)
- Within collection type
- Accurate time-stamps

TA2 - Geo-localization
- Camera pose provided (location & orientation)

18 Months

Phase 2
Robust

TA1 - ReID
- People, vehicles, generic objects
- Temporally distanced (people - attribute changes)
- Across collection types
- Time-stamp errors

TA2 - Geo-localization
- Camera location provided (no orientation)

18 Months

Phase 3
Generalize

TA1 - ReID
- Generic objects
- Temporally distanced (vehicles)
- Across broader collection types

TA2 - Geo-localization
- Camera general vicinity provided

12 Months

Metrics

- TA1: ReID
  - ReID accuracy – F_1 (measure of true positives with penalization for mis-matches and misses) in camera coordinate system
  - Processing time speedup – Autonomous reID vs manual reID (manual reference performed by T&E)
- TA2: Geo-localization
  - Ground plane accuracy, i.e. position error in meters

$$||(\text{long, lat, alt})_{Sys} - (\text{long, lat, alt})_{GT}||_2$$

  - Average over all object appearances
- TA1 & TA2: Runtime
  - Runtime on test server hardware relative to test video length/runtime (to ensure sufficient capacity for all submissions)

Metrics (TA1: ReID)

- F_1 [1]

$$F_1 = \frac{TP_{Sys}}{(TP_{GT} + (TP + FP)_{Sys}) / 2} = \frac{2TP_{Sys}}{2TP_{Sys} + FP_{Sys} + FN_{Sys}}$$

TP = True Positive, FP = False Positive, FN = False Negative

GT = Ground Truth, Sys = System

- Ratio of correctly identified detections over the average number of ground-truth and system detections
- Encapsulates true positives (correct re-identification), false positives (mismatches) and false negatives (misses) in single metric
- Computed for every frame where object appears

Measures how often the correct object is correctly re-identified, with even penalties for false positives and false negatives

[1] E. Ristani, F. Solera, R. Zou, R. Cucchiara and C. Tomasi, "Performance Measures and a Data Set for Multi-Target, Multi-Camera Tracking," *Proceedings of the European Conference on Computer Vision*, Amsterdam, the Netherlands, 2016.

Metric Milestones

Objective	Conditions		Phase 1	Phase 2	Phase 3
TA1: ReID accuracy: (F_1)	Person	Temporally proximate	0.6	0.75	0.9
		Temporally distanced	N/A	0.6	0.75
	Vehicle	Temporally proximate	N/A	0.6	0.75
		Temporally distanced	N/A	N/A	0.5
	Generic Object		N/A	N/A	0.5
TA1: ReID speedup	vs. manual annotation		10x	100x	1000x
TA2: Geo-localization (m) (accuracy in meters)	Full sensor pose provided (position and orientation)		10	5	2
	Sensor position provided (no orientation)		N/A	10	5
	General vicinity provided (w/in 100m, no orientation)		N/A	N/A	10
TA1 + TA2: Runtime (relative to video runtime)	On T&E servers (TBD)		5x	2x	1x

Test Set Conditions[1]

		Phase 1	Phase 2	Phase 3
People	# Distinct truthed identities people	750	1,500	2,500
	% Small (< 150 pixels high)	≥ 50% all phases		
	% Medium (150 < pixels high < 250)	< 30% all phases		
	% Large (250 < pixels high)	< 20% all phases		
	# Additional confusers	1,000	5,000	10,000
	# Attribute changes per subject	N/A	≥ 5	≥ 5
	# Sensor collection types[2]	3	5	6
Vehicles	# Distinct truthed vehicles	N/A	1,000	2,000
Video sensor pose	Collection sensor location and pointing	Full pose provided	Location, no orientation	Approximate location within 100 m

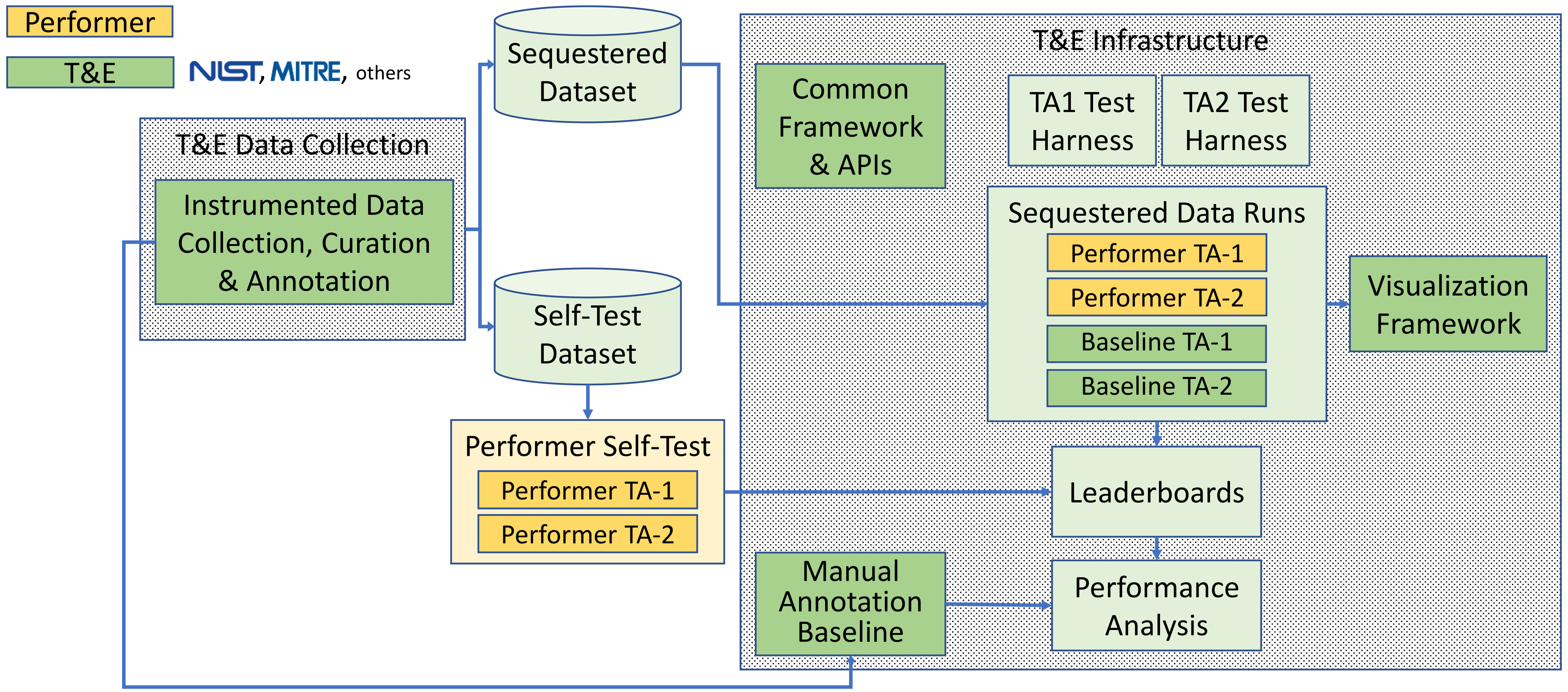
[1] Nominal, subject to update

[2] Sensor collection types to be provided at kickoff (nominally aerial, building, ground; stationary, mobile)

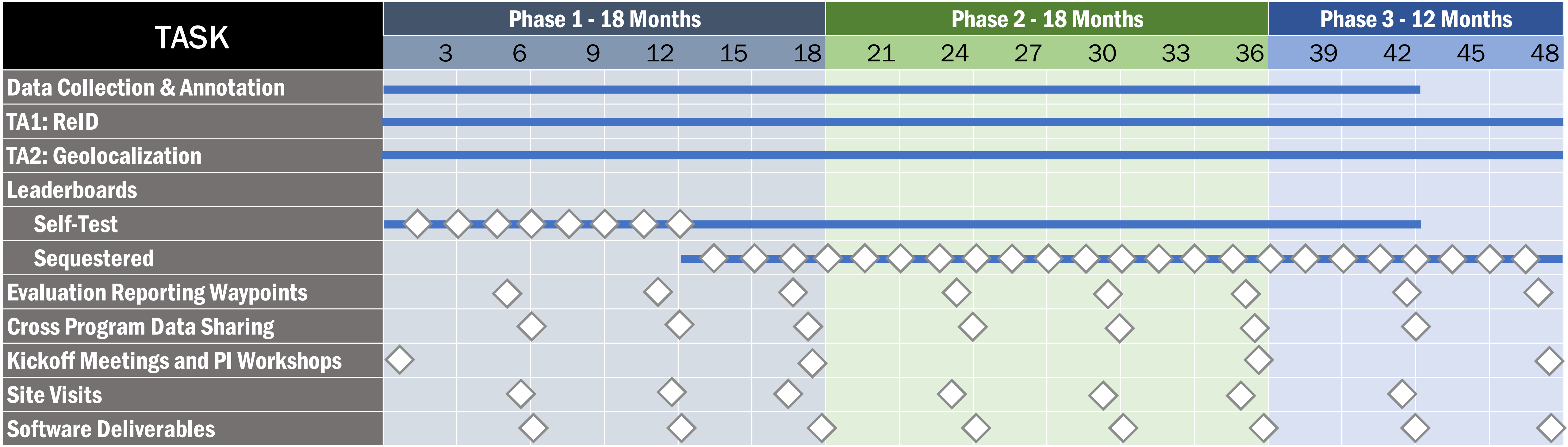
Data

- Privacy Protection Plan (PPP) required for any team that will use video data and/or annotations
- Institutional Review Board (IRB) approval required for any collections with potential Personally Identifiable Information (PII)
- All performer proposals that will utilize data for R&D (development and/or validation) must share data (video and annotations) program-wide every 6 months (after evaluation waypoint where data is used)
- T&E team will collect data for T&E purposes - There should be no expectation of Government collected datasets for R&D

T&E Structure



Schedule


TASK
Phase 1 - 18 Months
Phase 2 - 18 Months
Phase 3 - 12 Months
3
6
9
12
15
18
21
24
27
30
33
36
39
42
45
48
Data Collection & Annotation
TA1: ReID
TA2: Geolocalization
Leaderboards
Self-Test
Sequestered
Evaluation Reporting Waypoints
Cross Program Data Sharing
Kickoff Meetings and PI Workshops
Site Visits
Software Deliverables

Additional Information

Feedback, thoughts and comments:

- Video LINCS Team Alias: dni-iarpa-videolincs-proposersday@iarpa.gov

Additional information:

- Video LINCS website: www.iarpa.gov/index.php/research-programs/Video-LINCS

Video LINCS

Point of Contact

Dr. Reuven Meth

Program Manager

Office of the Director of National Intelligence

Intelligence Advanced Research Projects Activity (IARPA)

Email: dni-iarpa-videolincs-proposersday@iarpa.gov

Website: www.iarpa.gov/research-programs/Video-LINCS

Answers to Selected Questions

Agenda

Time	Topic	Speaker
9:30 a.m. - 9:40 a.m.	Welcome, Logistics, Proposers' Day Goals	Reuven Meth, Program Manager
9:40 a.m. - 9:50 a.m.	IARPA Overview	Robert Rahmer, Director Office of Analysis Research, IARPA
9:50 a.m. - 10:40 a.m.	Video LINCS Program Overview	Reuven Meth
10:40 a.m. - 11:00 a.m.	Contracting Overview	IARPA Contracts
11:00 a.m. - 12:30 p.m.	Break (**Submit questions in Webex chat before 11:00 a.m.**)	
12:30 p.m. - 1:30 p.m.	Answers to Selected Technical Questions	Reuven Meth
1:30 p.m. - 1:45 p.m.	Break	
1:45 p.m. - 3:15 p.m. (est.)	Lightning Talks*	Potential Performers
3:15 p.m. - 4:30 p.m.	Informal Teaming Discussions*	In-Person Participants

*The Government will not attend these events

LIGHTNING TALKS AGENDA (1)

Time	Speaker	Institution	In person
1:45 p.m.	Lance Besaw	Applied Research Associates	YES
1:50 p.m.	Colin Adams	Arete	YES
1:55 p.m.	Naresh Cuntoor	Blue Halo	YES
2:00 p.m.	Rodolfo Valiente Romero	HRL Laboratories	YES
2:05 p.m.	Scott McCloskey	Kitware	YES
2:10 p.m.	Derek Walvoord	L3Harris	YES
2:15 p.m.	Xiaoming Liu	Michigan State	YES
2:20 p.m.	Meryl Spencer	Michigan Tech Research Institute	YES
2:25 p.m.	Nigel Mathes	MORSE Corp	YES

LIGHTNING TALKS AGENDA (2)

Time	Speaker	Institution	In person
2:30 p.m.	Kemal Davaslioglu	Nexcepta	YES
2:35 p.m.	Jia Deng	Princeton U.	YES
2:40 p.m.	Ashley Antonides	Two Six Technologies	YES
2:45 p.m.	David Brady	U. of Arizona	YES
2:50 p.m.	Tejas Gokhale	UMBC	YES
2:55 p.m.	Yu-Wing Tai	Dartmouth College	NO
3:00 p.m.	Jeff Benson	DejaVuAI	NO
3:05 p.m.	Bar Cohen	Reichman U.	NO
3:10 p.m.	Mauricio Pamplona Segundo	U. of South Florida	NO