**Roadmap for Representing Information and Analysis in Cyber Domains**

Aim: establish a widely accepted standard for representing and exchanging digital forensic information that extends to cyber investigation, incident response, and counter terrorism.

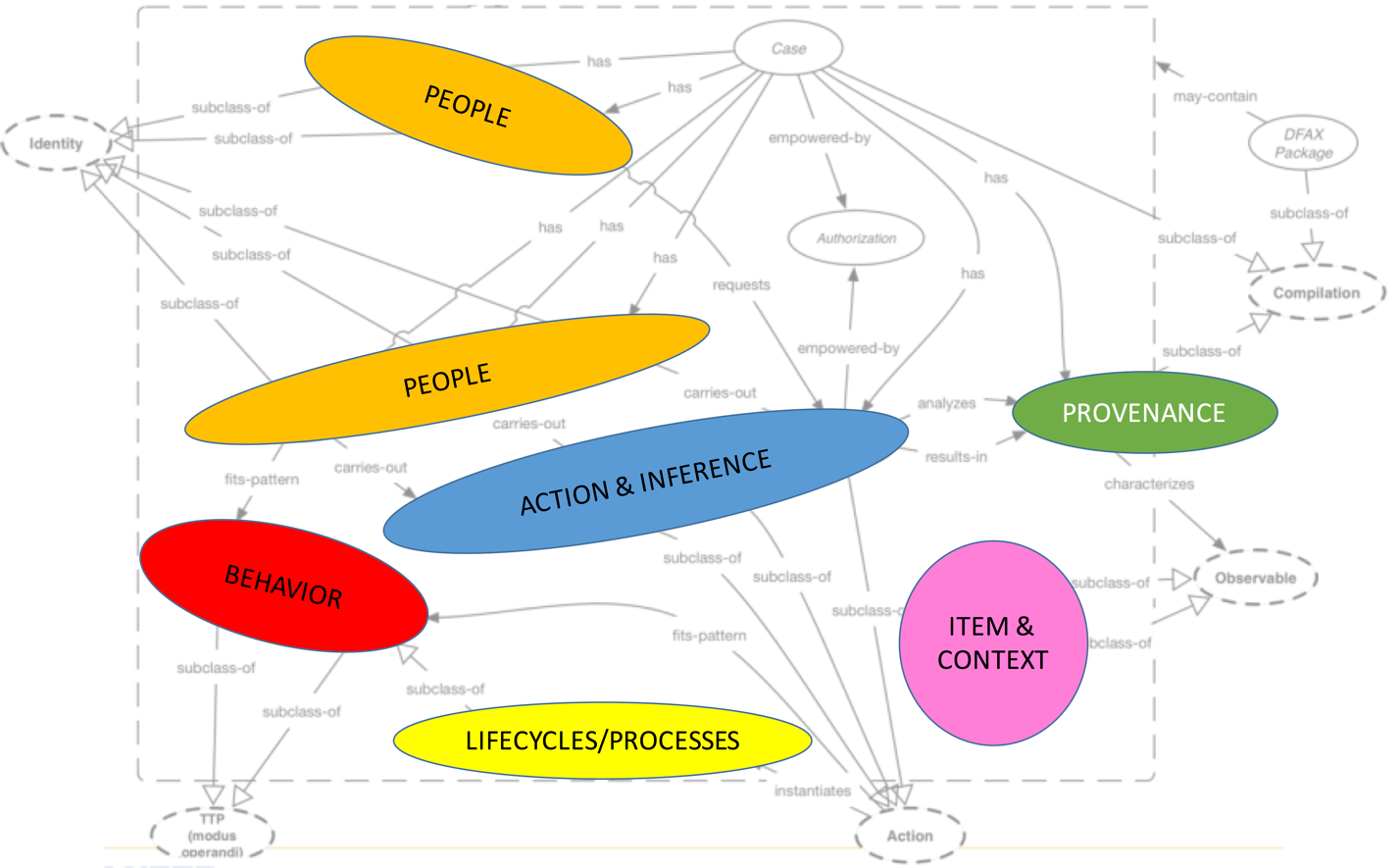
**Vocabulary**:

* Item: Physical or digital objects or entities, such as a mobile device, a file extracted from a device, an email address extracted from a file.
* Facet: Characteristics of an item. For example, a file can have multiple facets (picture, JPG, thumbnail).
* Action: Some observed activity caused by a person or system, which may be associated with item(s). For instance, execution of malware can result in various files, Registry entries, logs, etc. Actions can be an adjunction (object added to the crime scene, matter transferred, malware stored on a computer), a suppression (object stolen, file deleted), a transformation (misplaced/displaced object in a crime scene, file moved), a change of state (modified access rights of a file).
* Provenance: Where an item originated or was found, as well as how an item was handled after it was found.
* Grouping: An assortment of items that are logically grouped together. For instance, a MESSAGES are logically grouped into a THREAD.
* Relationship: An association or link between uco objects.

**Types of information to represent**:

* Case & Evidential Items
  + Case details
  + Items, Actions, and Relationships
  + Tool details
  + Output / results
* Provenance\_Records & Forensic\_Actions
  + Where an evidential item originated or was found, as well as how an item was handled after it was found (who did what to evidence where, when, and how?)
* Human Actors (Subject, Victim, Examiner, Investigator, Attorney, etc.)
* Forensic Processes (Lifecycle)
* Offender & Victim Processes (Lifecycle)
* Offender & Victim Behaviors (Action Patterns)
* Inferences (human reasoning or automated analytics)





**Requirement 1**: Represent multi-faceted extracted items and inter-relationships

Proposal:

1. Define base FILE item, and then define facet extensions for various file types (picture, audio, video, database, MS office, PDF, LNK, Jumplist).
2. Define base MESSAGE item, and then define facet extensions for various message types (sms, email, chat).
3. Define relationships between item, e.g., SQLite “mail.db” FILE contains a “Yahoo! Mail” MESSAGE and a blob FILE attached to the MESSAGE.
4. Represent other facets such as language using facets.

**Requirement 2**: Represent tools and system environment

Proposal:

1. Represent a specific version/configuration of tool or system as a distinct entity, with its own unique identifier.
2. Capture provenance, i.e., each instance of a tool being run and the specific system used to perform the operation.

**Requirement 3**: Have logical Groupings/Assemblages of extracted items

Proposal:

1. Define logical groups, including:
   * The Actor grouping/assemblage logically groups information about a person, including name, aliases, email addresses, age, etc.
   * The Thread grouping/assemblage logically groups message that are part of the same conversation
   * The Narrative grouping/assemblage logically groups information that represent investigation results or outcomes, including a timeline or link diagram.

**Requirement 4**: Represent Actions associated with extracted items, including time and location dimensions.

Proposal:

1. Define Action to apply to any item.

**Requirement 5**: Maintain provenance details for original, extracted, and processed items

Proposal:

1. Define Provenance\_Record, including who did what, where, when, how, and sometimes why.

**Requirement 6**: Represent Forensic\_Actions, both automated and human

Proposal:

1. Define Forensic\_Action, with input either from original evidential item or Provenance\_Record. Output must have an associated Provenance\_Record.
2. Distinguish between Forensic\_Actions that are purely fact-based observations and Forensic Actions that use human reasoning (abductive, deductive, inductive) whether human or automated to reach analytic assertions. For the output of human reasoning, assign a confidence such as a likelihood ratio or a scale (low, medium, high confidence).

**Requirement 7**: Represent Actions of offender(s), victim(s), and system

Proposal:

1. Define Actions associated with item, such as Offender sent a “Yahoo! Mail” MESSAGE to Victim.

**Requirement 8**: Represent behaviors of offender(s), victim(s), and system

Proposal:

1. Define Action Patterns to capture behaviors.

**Requirement 9**: Represent forensic process lifecycle

Proposal:

1. Define Action\_Lifecycle that can be used to represent forensic process lifecycle (e.g., evidence processing, evidence analysis).

**Requirement 10**: Represent offender process lifecycle

Proposal:

1. Define Action\_Lifecycle that can be used to represent offender process lifecycle (e.g., intruder breaking into a network, sexual predator grooming victim.

**Requirement 11**: Represent victim process lifecycle

Proposal:

1. Define Action\_Lifecycle that can be used to represent victim lifecycle.

**Requirement 12:** Represent what language(s) are within digital files/messages

**Requirement 13:** Represent data markings (classification, handling, etc) on content

Open questions:

* How to handle versioning of content?