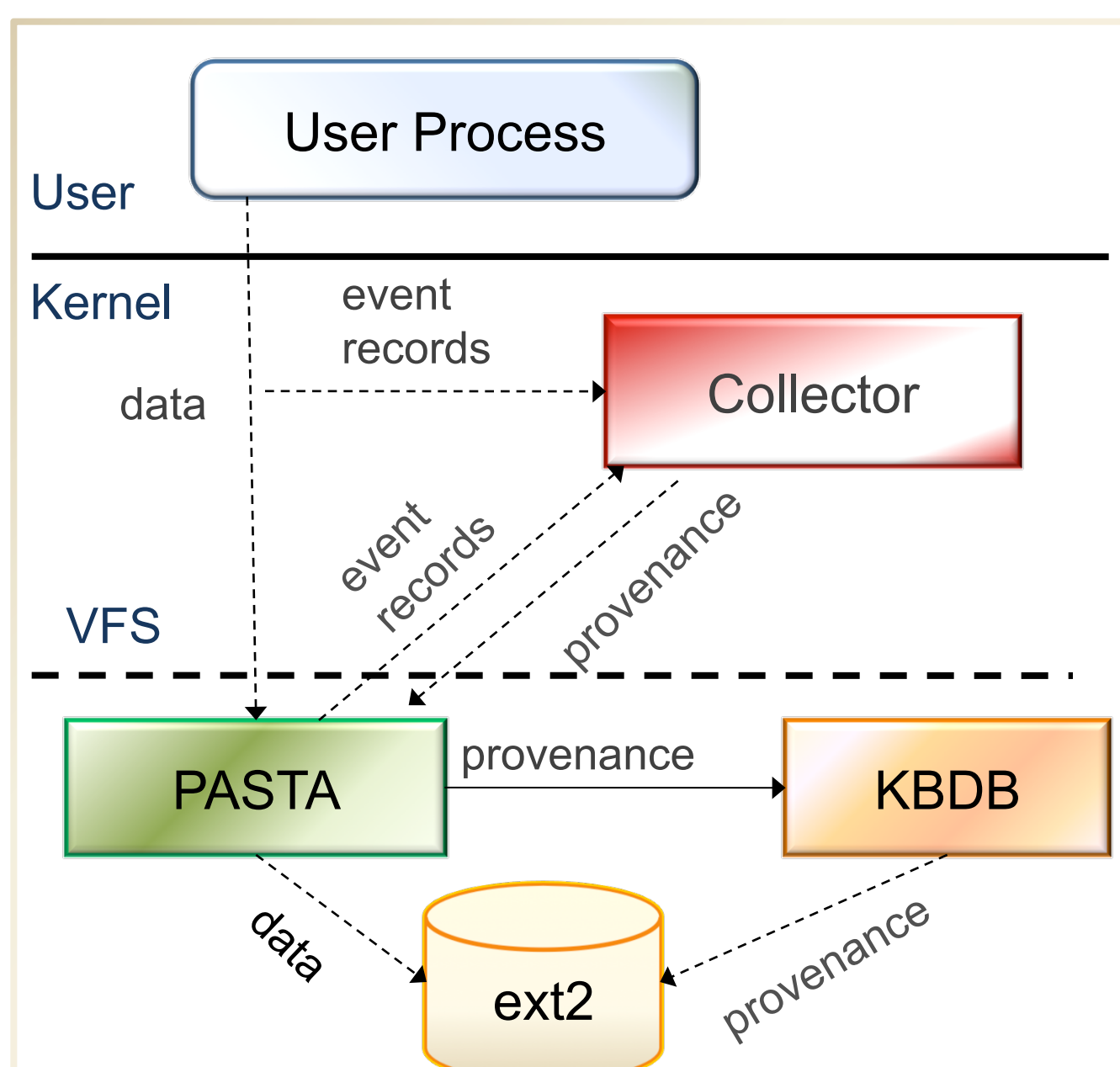


FiPS – A File Provenance System

Salmin Sultana, Elisa Bertino

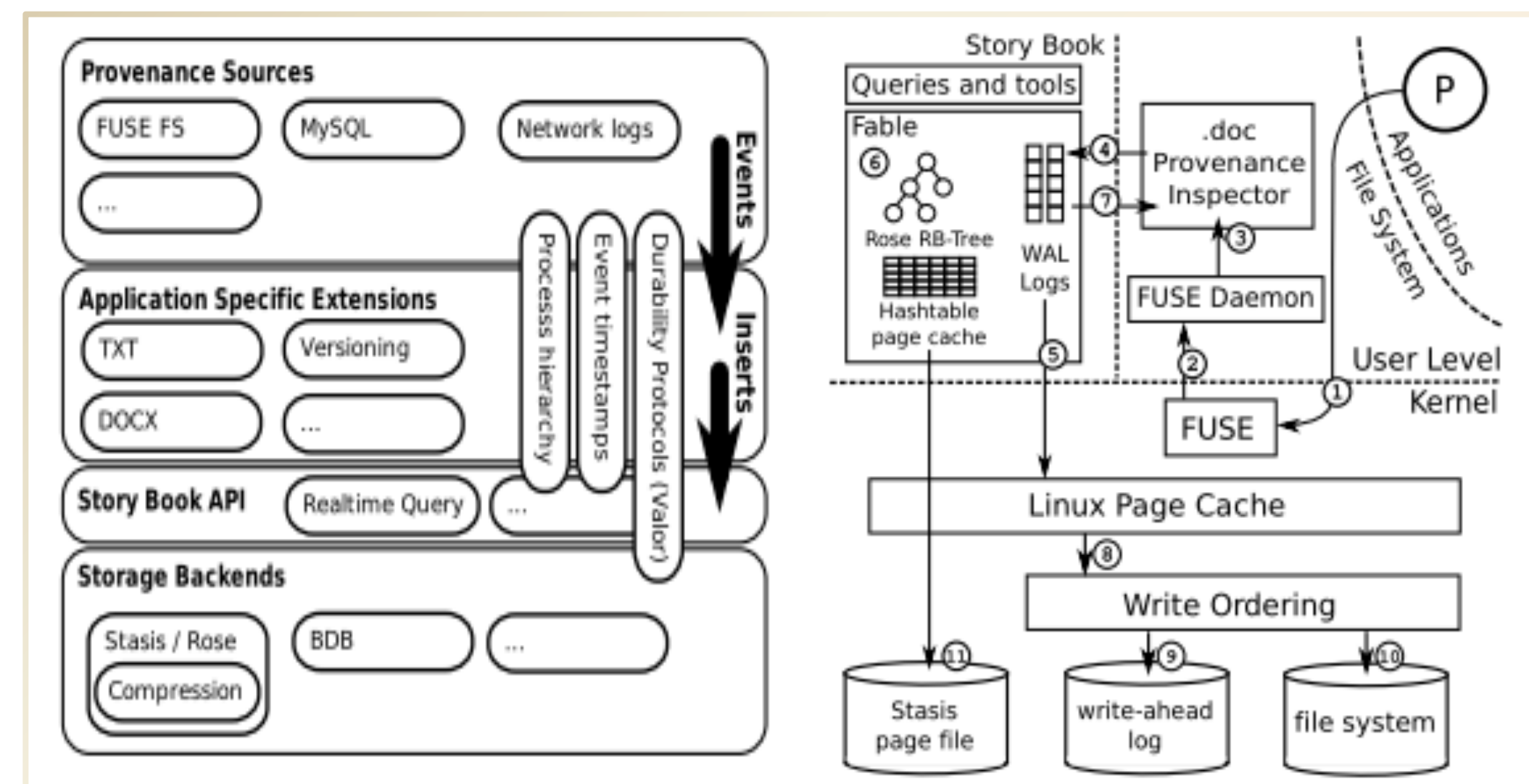
Motivation

PASS



- ❑ Captures unnecessary Provenance Information
- ❑ Cannot trace memory-mapped I/O
- ❑ NFS server-side operations cannot be tracked

Story Book



- ❑ Slow due to FUSE
- ❑ Security Concerns due to user space capturing

Design Objectives

Portability

- Provenance for any file system without significant change to the OS or file system.

Efficiency

- Not too much overhead in terms of time and space

Granularity

- User preferences to customize provenance capture

Security

- Security and Access control to provenance storage

Query Management

- Quick response to queries

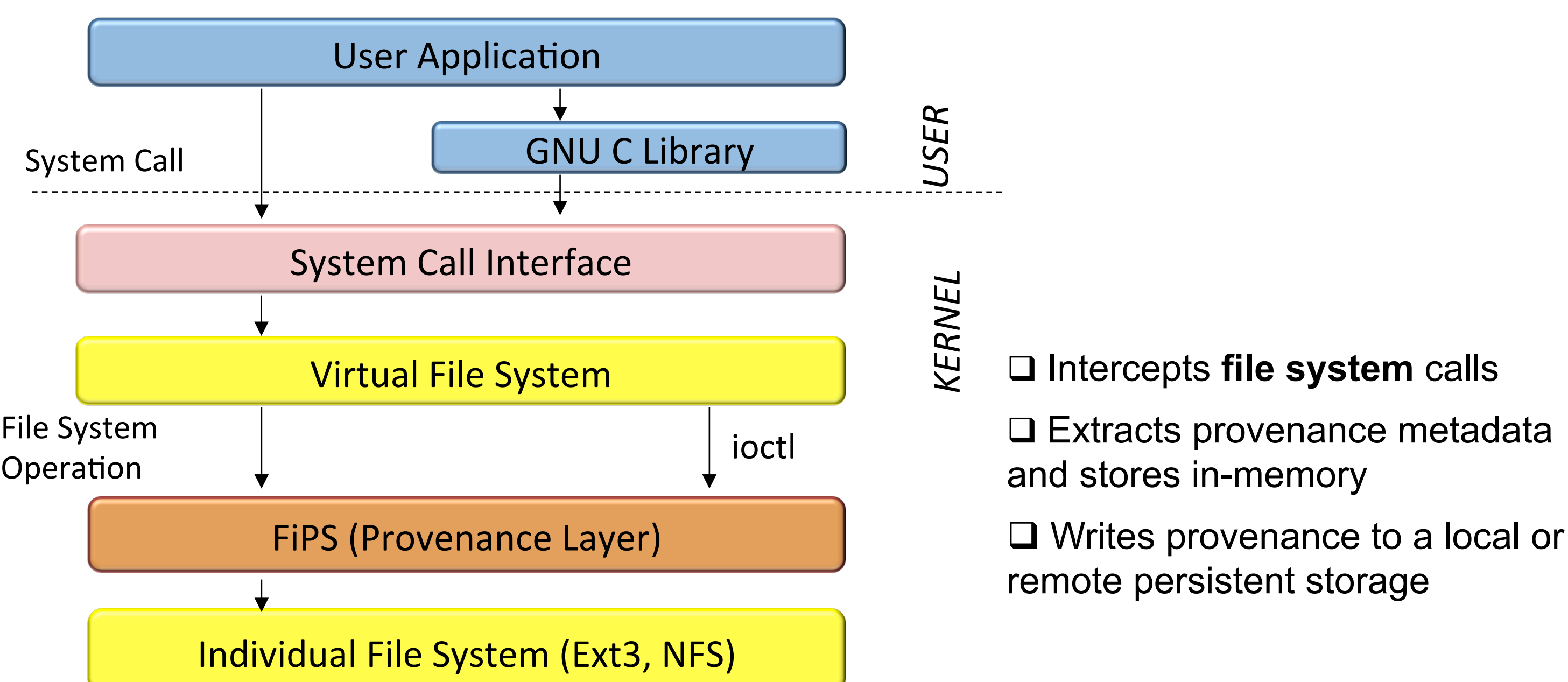
Provenance Model

Entities

- Process
- Operation
- Communication
- Lineage
- Environment
- Actor
- Access Control Policy
- Granularity Policy

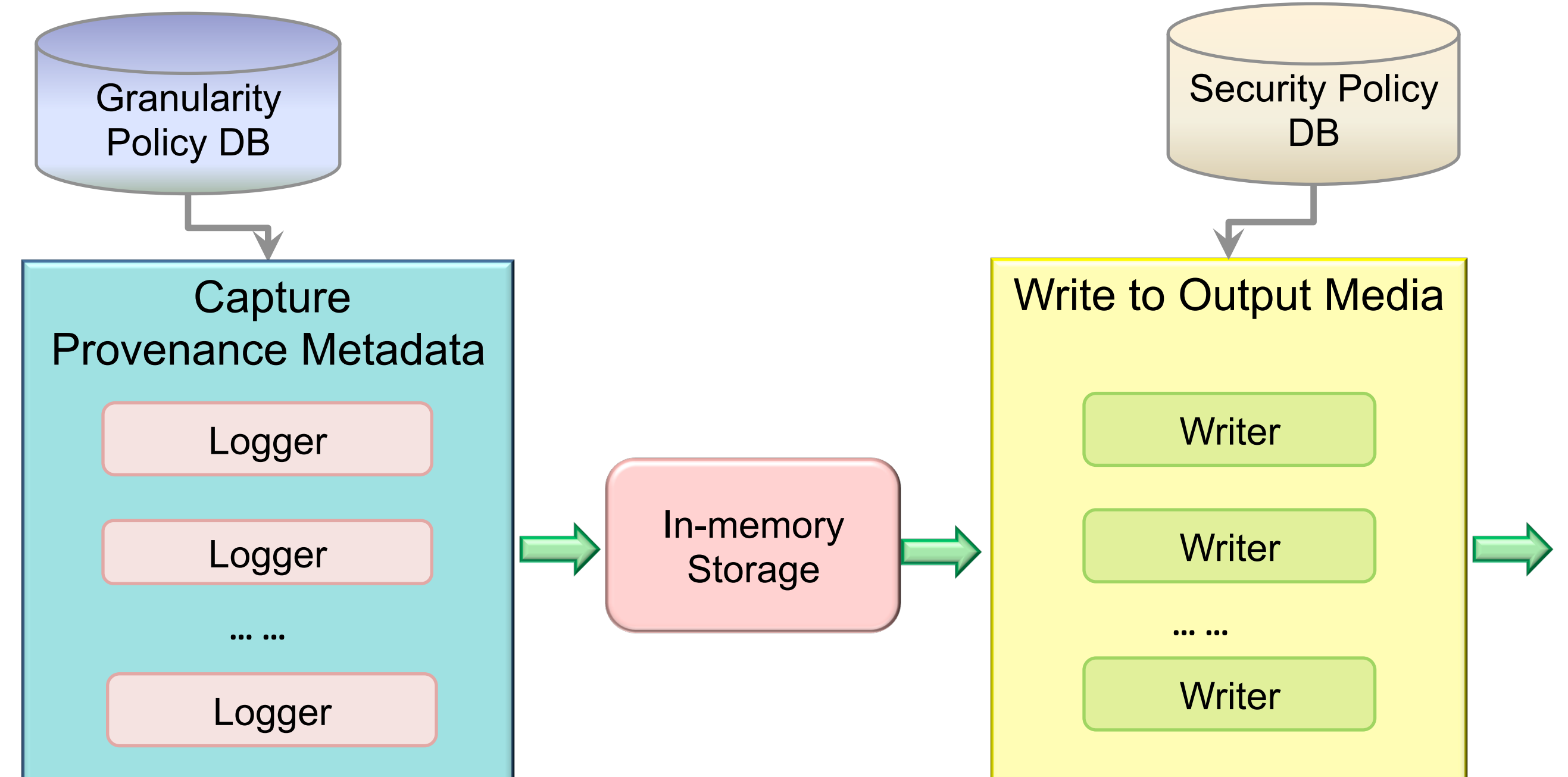
The provenance of a data object is the documented history of the actors, process, operations, inter-process / operation communications, environment, access control and other user preferences related to the manipulation of the object. The relationships between the entities form a provenance graph (DAG) for the data object.

Proposed Framework



- ❑ Intercepts **file system** calls
- ❑ Extracts provenance metadata and stores in-memory
- ❑ Writes provenance to a local or remote persistent storage

FiPS Layer



Implementation

Virtual File System	• Built on the stackable wrapper file system <i>Wrapfs</i>
Policy Databases	• In-kernel Berkeley DB (KBDB)
In-memory Storage	• Buffer or KBDB
Redundancy Elimination	• <i>Redactor</i> to compress and prune long term history for older files
Security	• Secure interface between the modules

Future Works

- Finish the implementation
- Incorporate with Networking File System (NFS)
- Extend FiPS to be placed at virtual machine monitors