ExtremeXP

Erik Pahor

April 2025

Abstract

This paper presents the design and current implementation status of ExtremeXP, a knowledge graph framework supporting reproducible experimentation across five distinct domains. The system addresses critical challenges in workflow contextualization, entity relationship mapping, and cross-domain data integration. Completed work establishes formal entity models for public protection (ADS), cyber threat analysis (i2CAT), activity prediction (MobyX), industrial monitoring (IDEKO), and flood forecasting (CS) use cases. Remaining challenges focus on dataset contextualization and dynamic intent propagation.

1 Introduction

Modern experimental systems face increasing complexity in managing heterogeneous workflows while maintaining reproducibility. The ExtremeXP framework addresses this through domain-specific knowledge graphs that capture entities, relationships, and operational semantics. Building upon prior work in workflow management systems and semantic modeling, this implementation spans five critical domains requiring distinct but interoperable knowledge representations.

2 Design Methodology

2.1 Domain Analysis

Each use case presents unique requirements for experimental reproducibility:

The ADS (Public Protection and Disaster Response) domain emphasizes access-controlled experiment regeneration, where security policies must govern both data and workflow usage. i2CAT's cyber threat classification requires specialized handling of imbalanced datasets through synthetic data generation. MobyX combines automated machine learning with human validation loops for activity prediction, while IDEKO focuses on real-time anomaly detection in industrial settings. The CS (flood prediction) use case integrates geospatial modeling with emergency response planning.

2.2 Entity Modeling

The core knowledge graph structure centers on six fundamental entity types that maintain consistency across domains while allowing specialization:

- **Users** represent authenticated actors with domain-specific roles (e.g., PPDR officers in ADS, industrial engineers in IDEKO)
- **Datasets** incorporate provenance metadata using NFT-based tagging developed by ICCS
- **Workflows** implement either predefined or dynamically generated task sequences

Domain extensions build upon this foundation. For instance, ADS introduces *AccessControlPolicy* entities that govern experiment visibility, while MobyX adds *GroundTruth* entities for human validation of automated predictions.

3 Remaining Challenges

Two critical limitations must be addressed to achieve full operational capability:

3.1 Dataset Contextualization

The Matic integration requires careful mapping of raw datasets to entity attributes.

3.2 Intent Propagation

The current implementation lacks a mechanism for propagating user intent across workflows. This is particularly important in domains like ADS, where security policies must adapt dynamically to user actions.

4 Conclusion

ExtremeXP's knowledge graph approach demonstrates viability for multi-domain experimental systems. Current progress establishes a robust foundation for entity management and workflow execution, while identified challenges focus future development efforts. Successful completion of dataset contextualization and intent propagation will enable the framework's deployment across all five target domains.