IAEA Technical Meeting EVT2300917 on
Deployment of Artificial Intelligence Solutions for the Nuclear Power Industry:
Considerations and Guidance18-21 March 2024
U.S. Nuclear Regulatory Commission Headquarters, Rockville, MD, USA

Some issues in the Assurability of safety-critical digital systems
Part 1 – Assurance and AI

Doug Eskins
Senior Computer Engineer
U.S. Nuclear Regulatory Commission
Office of Nuclear Regulatory Research

The views expressed herein are those of the author and do not represent an official position of the U.S. NRC.

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

Assurance

- A claim (about **X**) is supported by sound, valid evidence (under the assumptions and conditions identified in **Y**).
- **X** could be a system design or an O&M process.
- **Y** is a set of conditions and assumptions under which the claim holds.
- Assurance is sometimes referenced to a CAE triplet (claim, arguments, evidence)

Artificial Intelligence

A machine-based system that can go beyond defined results and scenarios and has the **ability to emulate human-like** perception, cognition, planning, learning, communication, or physical action (NRC AI Strategic Plan).

Note: Each human-like capability is referenced to some (domain-specific) application.

AI & Assurance

- How can AI be assured?
- How can AI be used for assurance?

Assuring AI

- What are the bounds of application?
  - In nuclear: safety or non-safety, design or O&M?
- Is assurance comparable between humans and AI?
- How will the CAE needed to assure an application differ for AI?
  - Ex) Can non-interference with a safety function be assured?

AI for Assurance

- Can AI facilitate the CAE needed for assurance?
  - Data collection, processing, and analysis to support Evidence generation
  - System modelling to support Argument construction and validation
  - System and domain analysis to ensure a necessary and sufficient set of Claims to support assurance.

Assuring AI for Nuclear Cybersecurity Applications

- Ongoing NRC research exploring the use of AI to characterize nuclear cybersecurity states.
- Issues encountered relevant to assurance of cybersecurity classification models:
  - Data artifacts & joint IT/OT data
  - Model performance measures & coverage of plant states
- Answers can be very application dependent

IAEA Technical Meeting EVT2300917 on
Deployment of Artificial Intelligence Solutions for the Nuclear Power Industry:
Considerations and Guidance18-21 March 2024
U.S. Nuclear Regulatory Commission Headquarters, Rockville, MD, USA

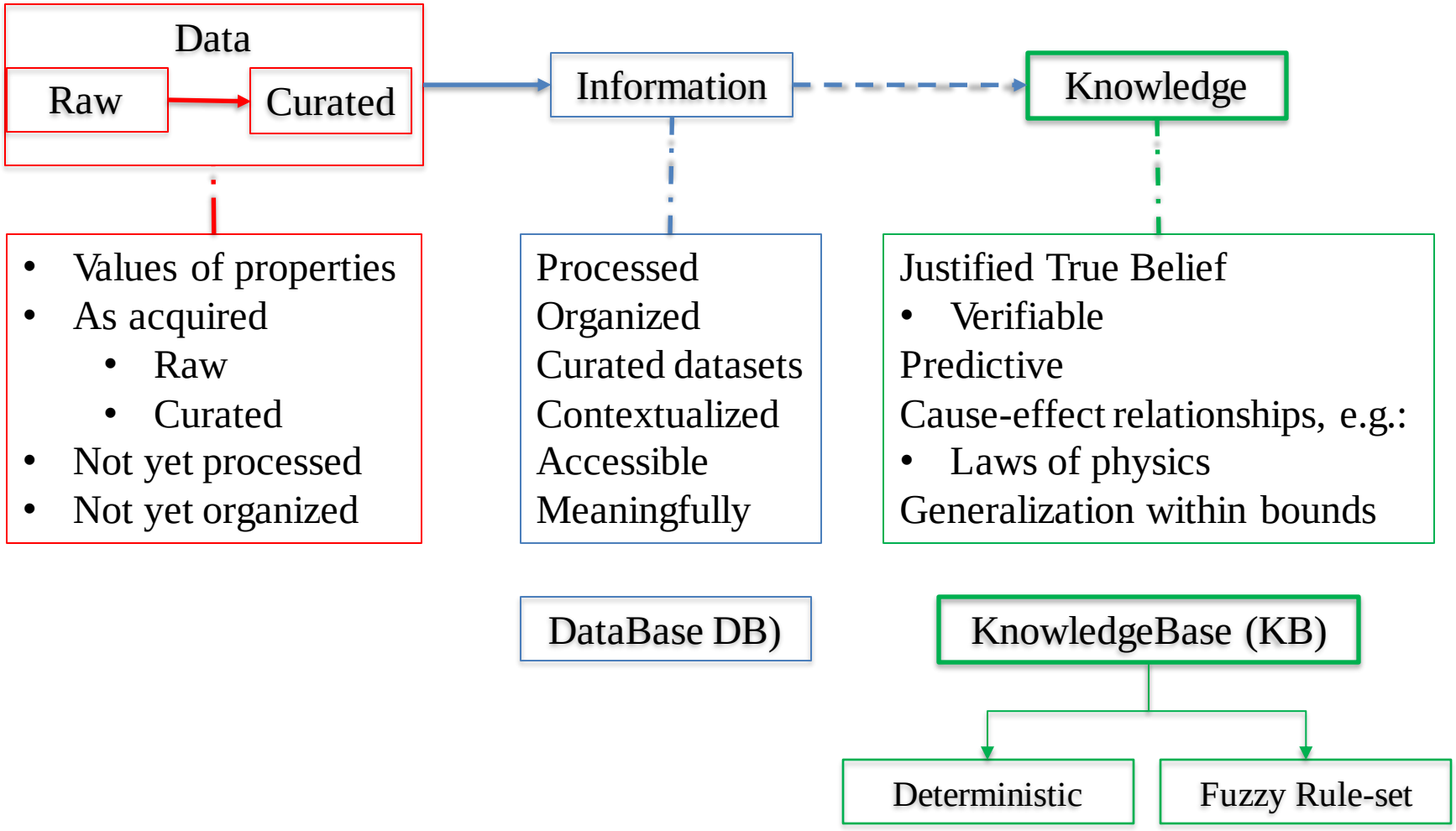
Some issues in the Assurability of safety-critical digital systems Part 2 – Knowledge Engineering is on the back burner

Sushil Birla
Senior Technical Advisor
U.S. Nuclear Regulatory Commission
Office of Nuclear Regulatory Research

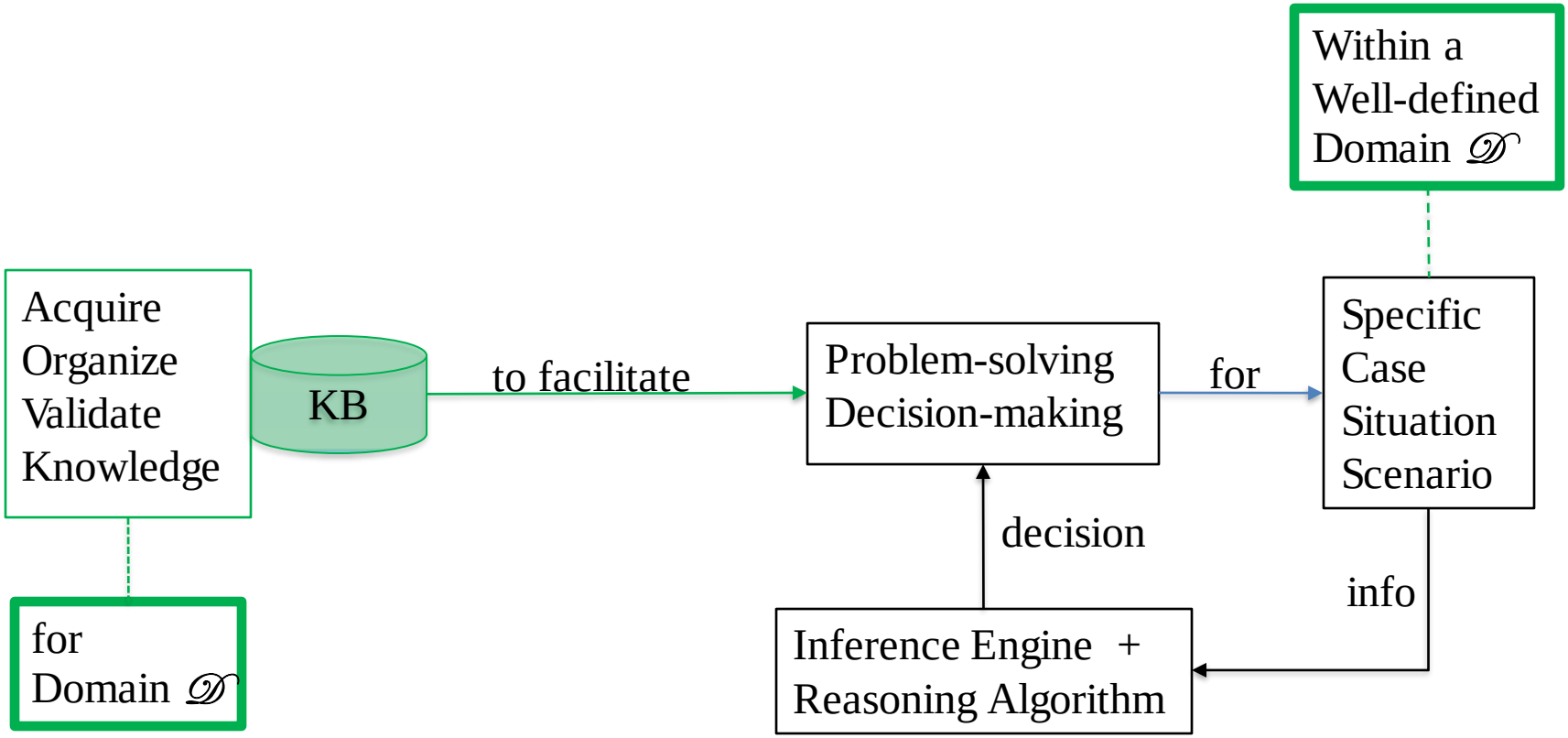
The views expressed herein are those of the author and do not represent an official position of the U.S. NRC.

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

Distinguish between data, information & knowledge



Knowledge Engineering (KE)



Knowledge Representation (KR)

KR: the field of artificial intelligence (AI)
dedicated to representing knowledge about the world
in a form that can be mechanized to solve complex tasks.

Means of KR — example: Ontology

a set of concepts and categories in a subject area or domain that shows their properties and the relations between them

KR formalisms – characteristics of interest:

- Expressivity
- Tractability
- Comprehensiblity
- Usability; Learnability

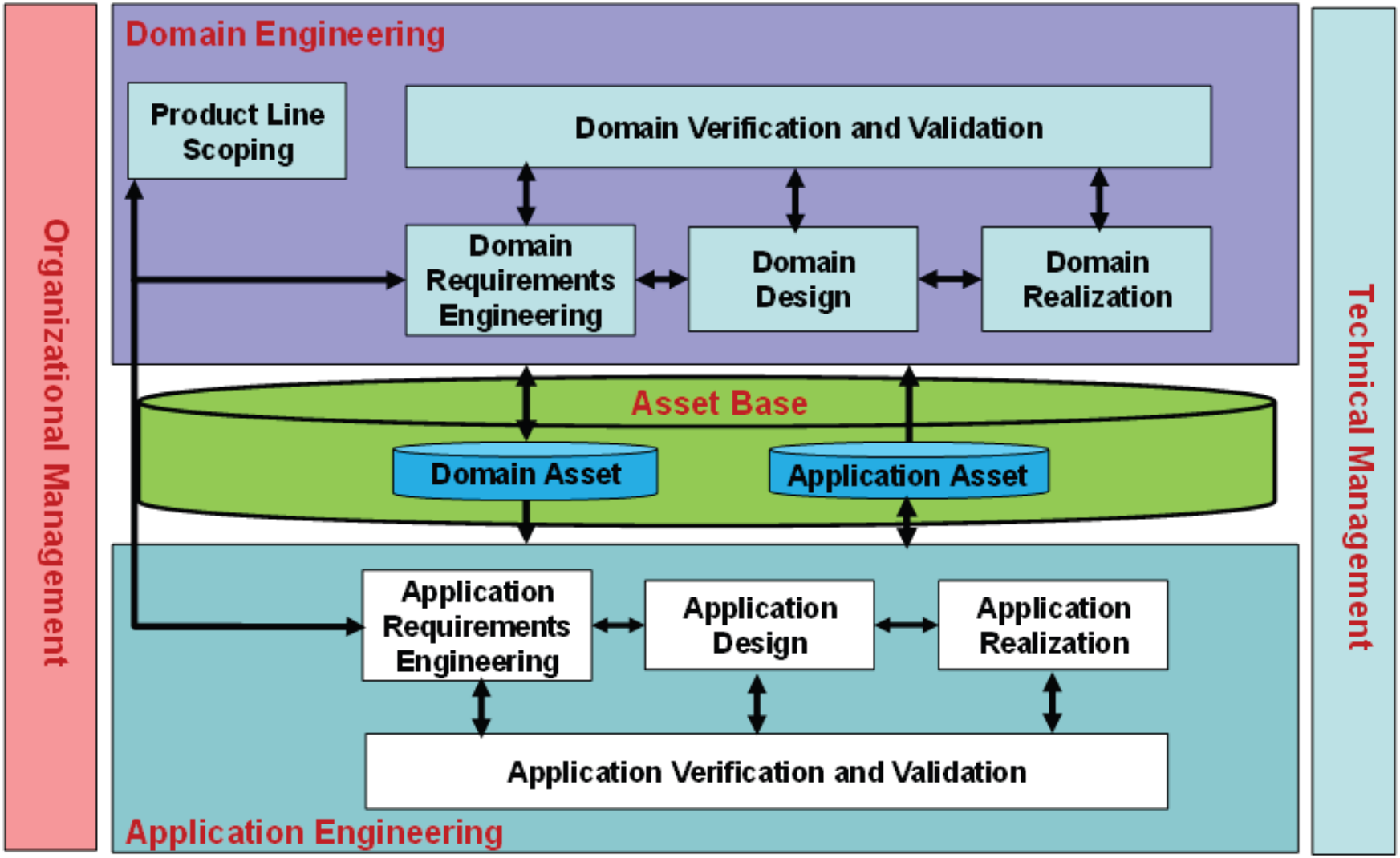
Domain Engineering

Development, evolution, and sustenance of domain specific knowledge and artifacts to support the development and evolution of particular systems in the domain.

Includes engineering of:

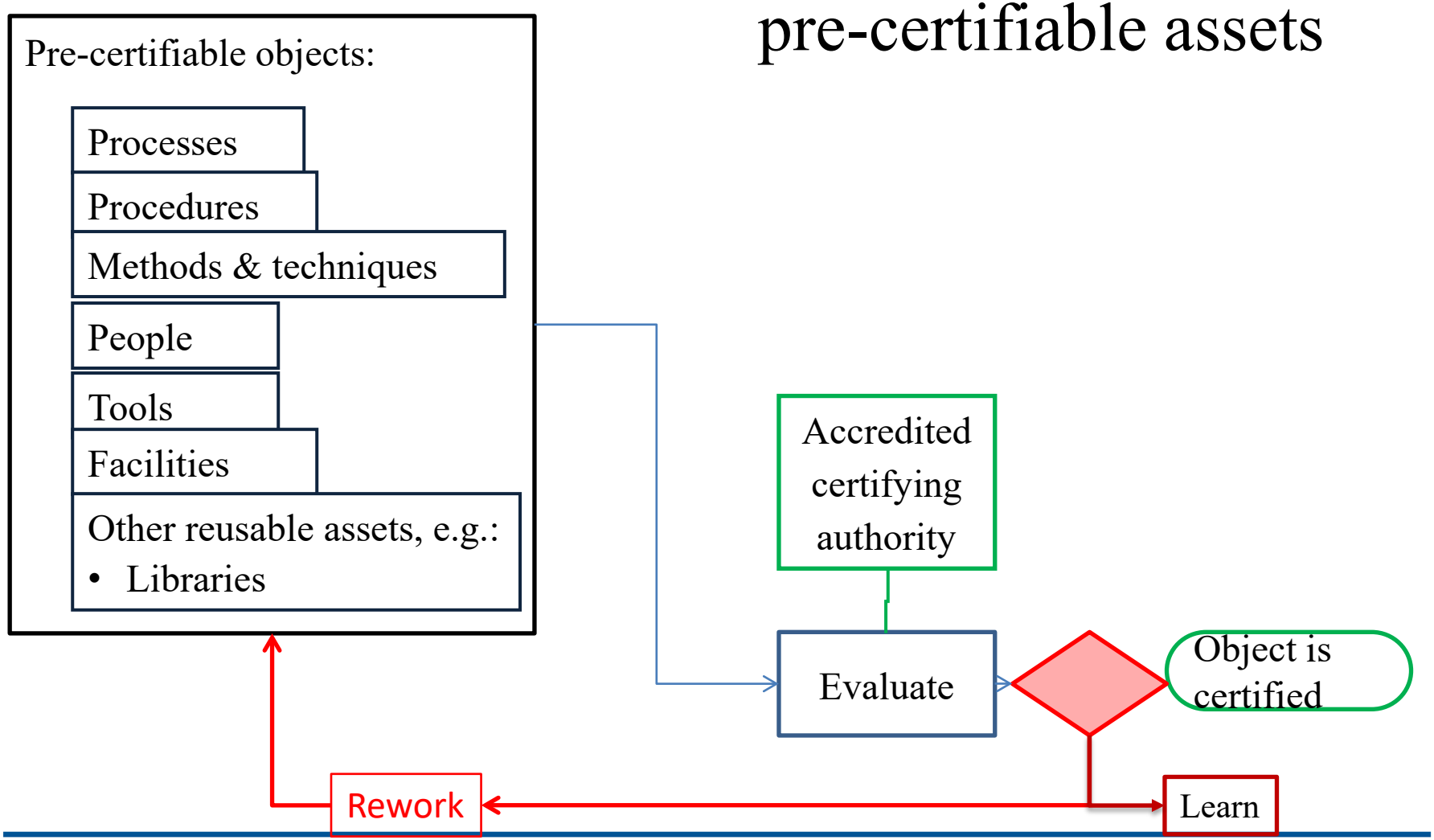
- Domain models
- Components
- Methods
- Tools
- Asset management (possibly)

Reference model

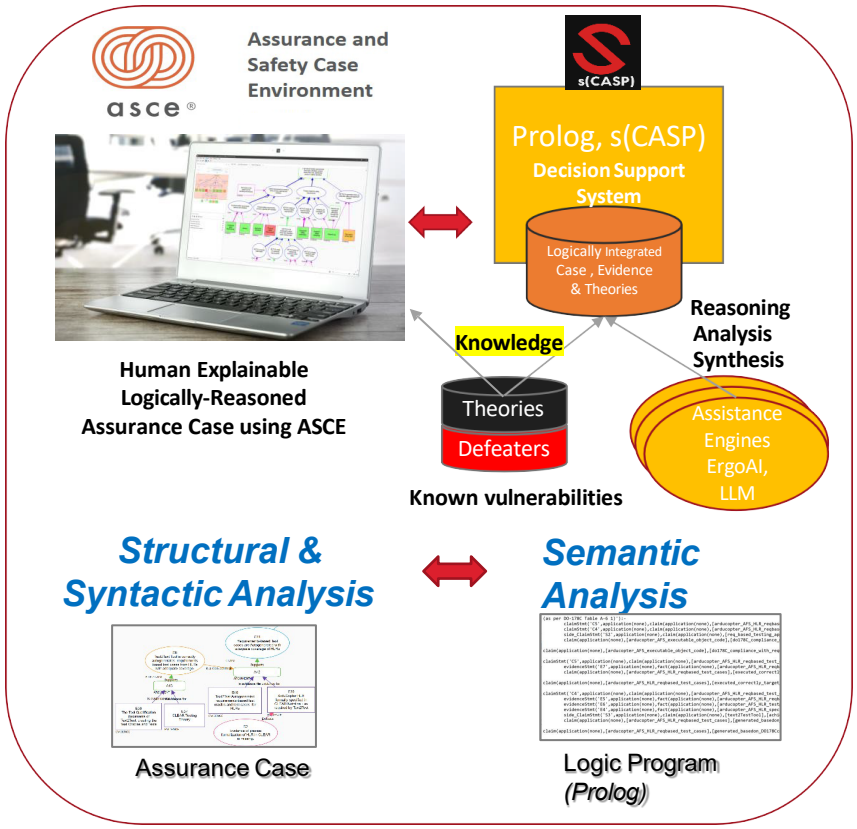


Source: **ISO/IEC 26550:2015(E)**

Domain engineered pre-certifiable assets

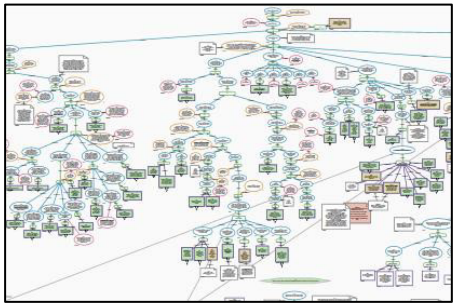


CLARISSA TOOLS



Step 1

Assurance Case to Logic Program



ASCE

```
claimStmt('C28',application('C28'),claim(application('C28'),
    evidenceStmt('E36',application('C28'),fact(application('
    evidenceStmt('E33',application('C28'),fact(application('
    evidenceStmt('E35',application('C28'),fact(application('
    evidenceStmt('E37',application('C28'),fact(application('
    evidenceStmt('E34',application('C28'),fact(application('
    evidenceStmt('E38',application('C28'),fact(application('
    claim(application('C28'),[tim,tim_driver_license,tim_car
    theory('C14',application('C28'),claim(application('C28'),

claim(application('C28'),[tim,tim_driver_license,tim_car,saf

evidenceStmt('E36',application('C28'),fact(application('C28'
    fact(application('C28'),[tim_driver_license],[valid],[us

fact(application('C28'),[tim_driver_license],[valid],[us]).

evidenceStmt('E33',application('C28'),fact(application('C28'
    fact(application('C28'),[tim,tim],[testimonials_of_good_
```

s(CASP)

Query:

?- claimStmt(C28,application(C28),claim(application(
[tim,tim_driver_license,tim_car,safe_driving_factor_list],[
theory. Tim is a safe driver in the US).

Answer:

yes

Model:

{ claimStmt(C28,application(C28),claim(application(C28)
[tim,tim_driver_license,tim_car,safe_driving_factor_list],[
theory. Tim is a safe driver in the US),

Semantic Reasoning

Step 2

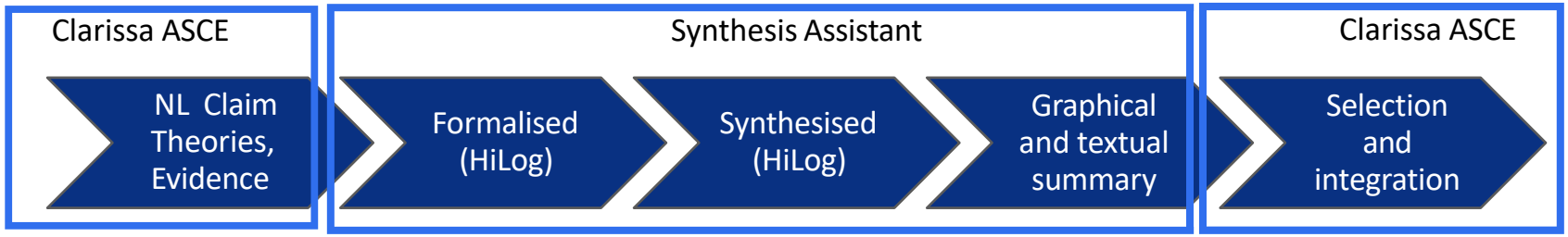
Source: CLARISSA team presentation at ARCOS meeting, Niskayuna NY March 12, 2024

DISTRIBUTION C. Distribution authorized to U.S. Government Agencies and their contractors; administrative or operational use; 10/19/2021. Other requests for this document shall be referred to DARPA I2O.

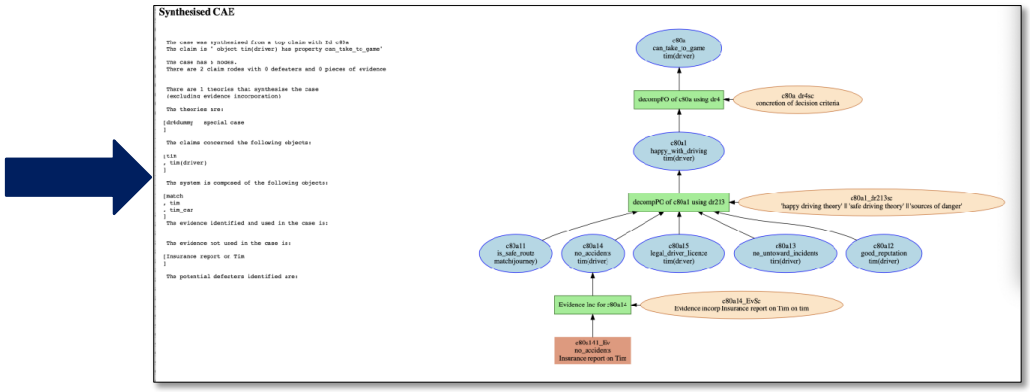
U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

ASSURANCE CASE SYNTHESIS

Synthesis Assistant is a research tool designed to synthesize claims, arguments and evidence structures from a root or top-level claim.



- Given:
  - Top-level claim (defined in ErgoAI or node imported from an ASCE file)
  - Definition of the system structure
  - Possible defeaters
  - Theories used to develop the case
  - Evidences for the case



Source: CLARISSA team presentation at ARCOS meeting, Niskayuna NY March 12, 2024

DISTRIBUTION C. Distribution authorized to U.S. Government Agencies and their contractors; administrative or operational use; 10/19/2021. Other requests for this document shall be referred to DARPA, I2O.

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

Supporting information

ISO/IEC 26550 family of standards

ISO/IEC 26550:2015(E)
Software and systems engineering — Reference model for product line engineering and management

ISO/IEC 26551:2016(E)
Tools and methods for product line requirements engineering

ISO/IEC 26552:2019(E)
Tools and methods for product line architecture design

ISO/IEC 26553:2018(E)
Processes and capabilities of methods and tools for domain realization and application realization

ISO/IEC 26554:2018(E)
Methods and tools for domain testing and application testing

ISO/IEC 26555:2015
Tools and methods for technical management

ISO/IEC 26556:2018(E)
Tools and methods for organizational management

ISO/IEC 26550 family of standards

ISO/IEC 26557:2016(E)
Methods and tools for variability mechanisms

ISO/IEC 26558:2017(E)
Methods and tools for variability modeling

ISO/IEC 26559:2017(E)
Methods and tools for variability traceability

ISO/IEC 26560:2019(E)
Methods and tools for product management

ISO/IEC 26561:2019(E)
Methods and tools for technical probe

ISO/IEC 26562:2019(E)
Processes and capabilities of methods and tools for transition management

ISO/IEC 26563:2022(E)
Processes and capabilities of methods and tools for configuration management of assets

ISO/IEC 26564: 2022(E)
Methods and tools for product line measurement

ISO/IEC 26550 family of standards

ISO/IEC 26850:2021(E)
Methods and tools for the feature-based approach to software and systems product line engineering

ISO/IEC 26565 to ISO/IEC 26599: To be developed