

National Oceanic and Atmospheric Administration

Overview and Status of the Joint Venture Partnerships Program (JVP) Broad Agency Announcement (BAA)

Determining Natural Language Processing Capabilities with an Open Knowledge Mesh

Toward Realization and Exploitation of the Digital Twins Vision

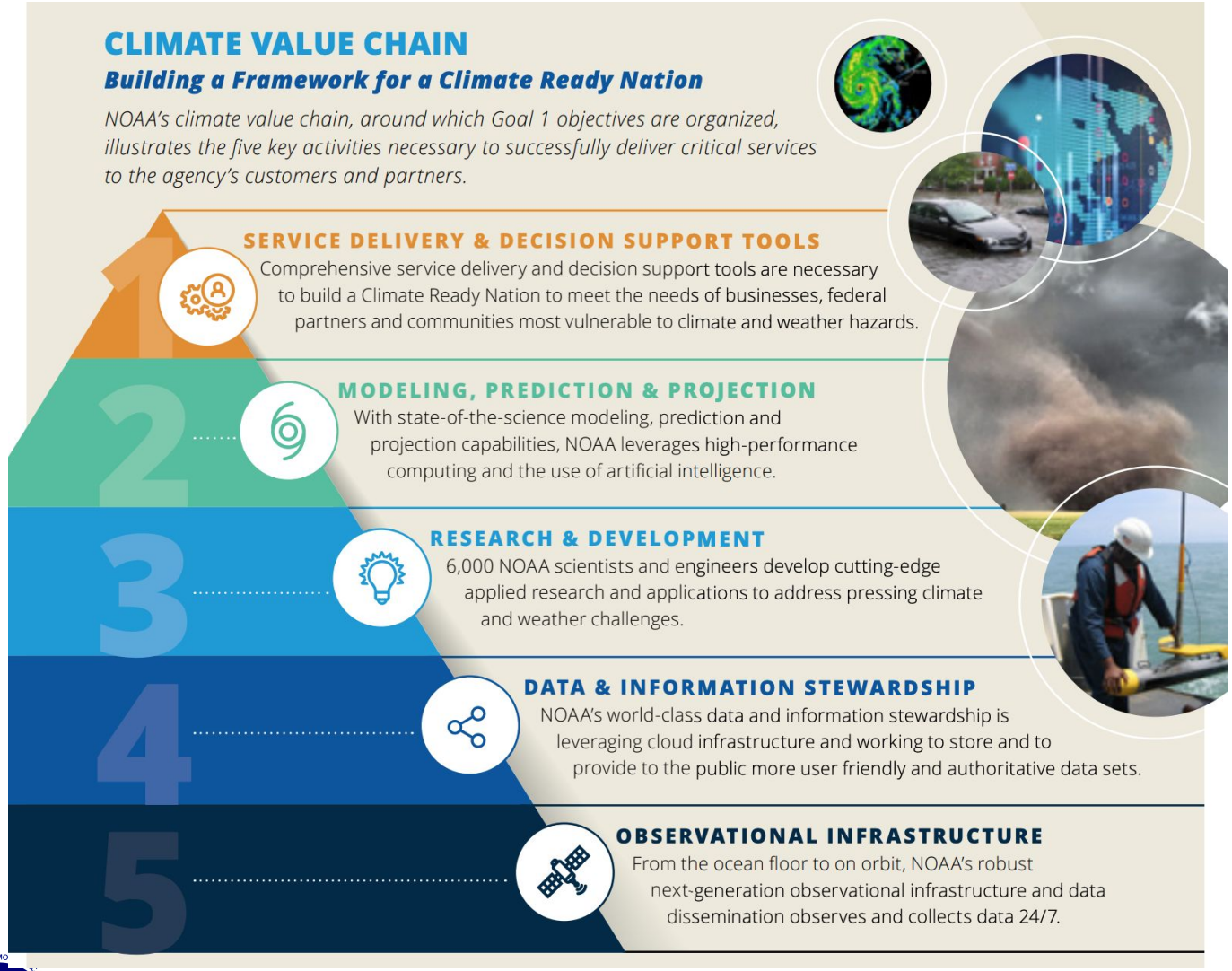
Prepared for: Session 15D - Special Look-Ahead Topics in Transition of Research to Operations (R2O)
As part of: AMS Annual Meeting 2025
Delivered on: 16 January 2025

Ryan Berkheimer[1], Beau Backus[2,3], Harshesh Patel[2], Ramesh Rangachar[2,4], Kevin Bartlett[2,5], Justin Dennison[2,3], Rob Redmon[1,6], Jennifer Webster[1], Tyler Christensen[7], Paul DiGiacomo[8]

[1]NOAA NESDIS National Centers for Environmental Information (NCEI), [2]NOAA NESDIS Systems Architecture and Engineering (SAE), [3]Johns Hopkins Applied Physics Laboratory, [4]The Aerospace Corporation, [5]Science and Technology Corporation (STC), [6]NOAA National Center for Artificial Intelligence, [7]NOAA NESDIS Office of Common Services (OCS), [8]NOAA NESDIS Center for Satellite Applications and Research (STAR)

Strategic Foundations for Digital Twins
NOAA Perspectives

Supporting and **Leveraging** NOAA **Strategic Plans** that Address **Grand Challenges**





NOAA Artificial Intelligence Strategy
Analytics for Next-Generation Earth Science

NOAA Cloud Strategy
Maximizing the Value of NOAA's Cloud Services

NOAA 'Omics Strategy
Strategic Application of Transformational Tools

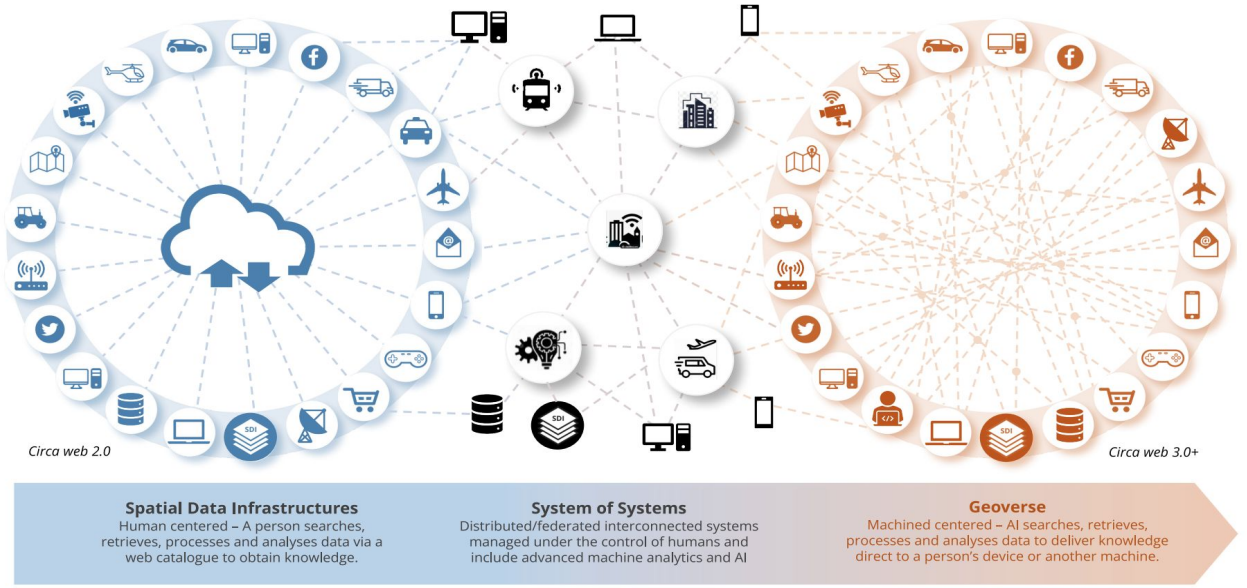
NOAA Data Strategy
Maximizing the Value of NOAA Data

NOAA Uncrewed Systems Strategy
Maximizing Value for Science-based Mission Support

NOAA Citizen Science Strategy
Applying the Power of the Crowd

Alignment with the Next Generation **Informatics Ecosystem**

- A **democratized** 'system of systems' powered in large part by **machine to machine** communication
  - Systems may be '**smart agents**'
- An **ecosystem** in which all users can both consume and contribute information
- A federated **trust-based network** for realizing **universally useful understanding**



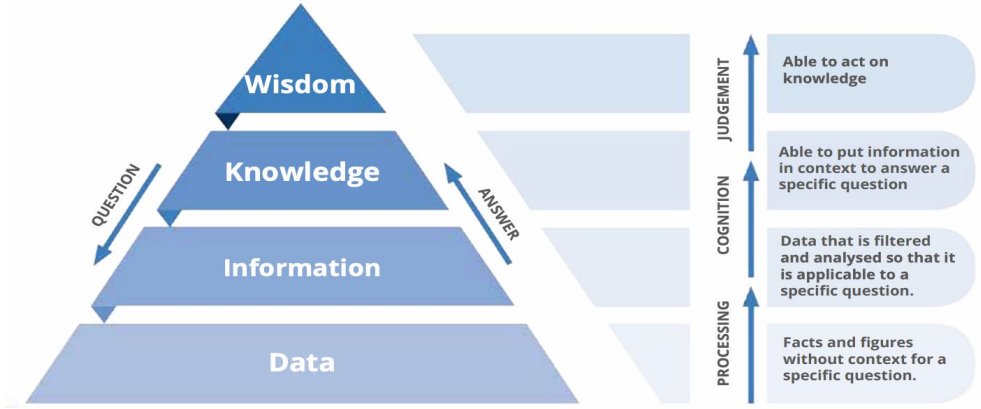


Figure 4. From data to information, knowledge and wisdom. Adapted from DIKW Model for knowledge management and data value extraction.

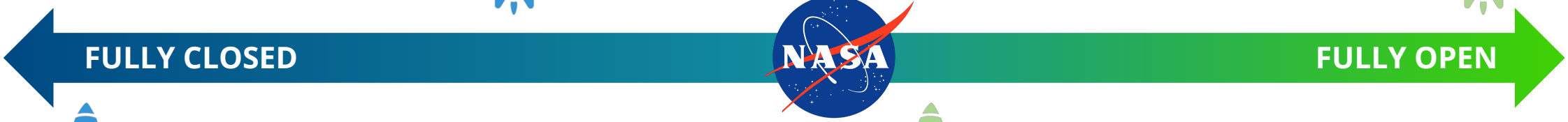
Strategic Pathway Elements	Spatial Data Infrastructures	Future Geospatial Information Ecosystem
Standards Governance and Policy	A commitment to assess, establish, and maintain a common standards framework	A commitment to adopt standard models for knowledge representation
Technology and Data Interoperability	Voluntary standards for data and technology	Anticipatory regulations, FAIR data and human-centered standards for knowledge creation
Compliance Testing and Certification	Regular assessments, training, government mandates, performance metrics, testing and certification	Regulatory sandboxes, proof of concepts and incentives for deploying and scaling 4IR technologies
Community of Practice	Sharing and leveraging data and technology standards	Inclusive and participatory roles in data vocabularies, digital twins, and other 4IR standards development

Figure 12. Standards Strategic Pathway – the Step Change

Delivering the **Open Science** Promise

Data access ($$)
Accessible Publications ($$)
Siloed systems
Limited communication
Proprietary Software
“Closed-Tent” culture

Free unlimited data access
Fully documented open software and algorithms
Fully linked data and publications
Open Access Journal publications
Fully Transparent processes
Reproducible across platforms
“Teaching” culture
Open science meetings



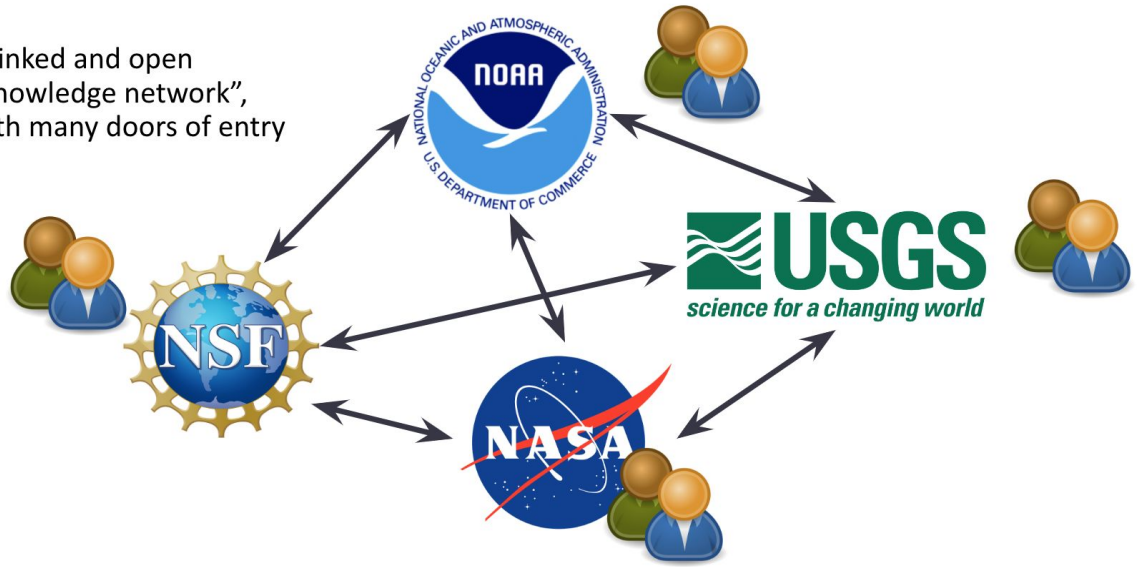
No public access data
No publications
No insight into processes
No reproducibility
“Black Box” culture

Free data access
Open software and algorithms
“Green” Journal publication
Documented processes
Reproducible in specific environments
“Open-Tent” culture

Source: C. Gentemann (2022, NOAA EDM Workshop presentation)

Achieving **Improved Outcomes** via Advanced **Partnerships**

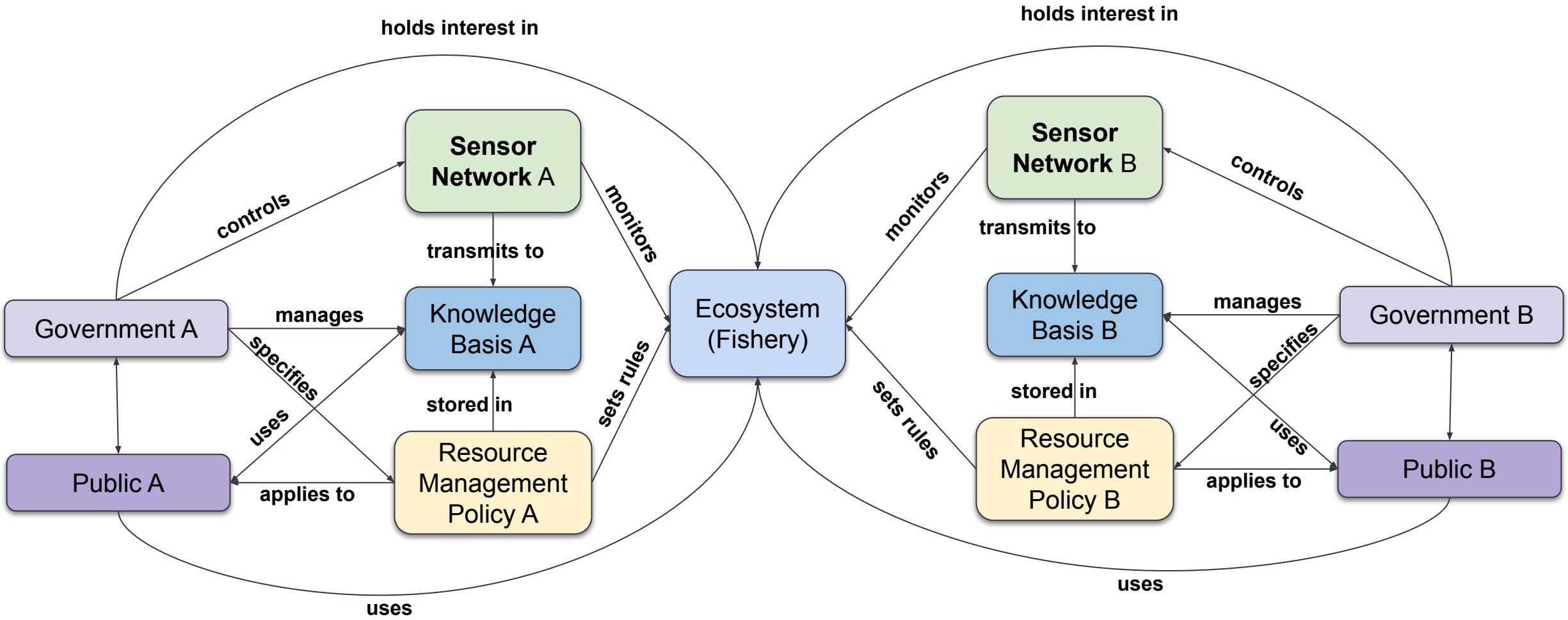
A linked and open
“knowledge network”,
with many doors of entry



-More coverage in **use cases**
-Providing more **holistic viewpoints**
-**Facility specialization** and **sharing** of **capabilities**
-Ability to **capture, track,** and **integrate decisions** and their **outcomes**, enabling **improved wisdom** over time (tracking **multidimensional uncertainty-reduction feedback** cycles - a digital twin requirement)



Supporting an **Operational** Ecosystem of **Twins**

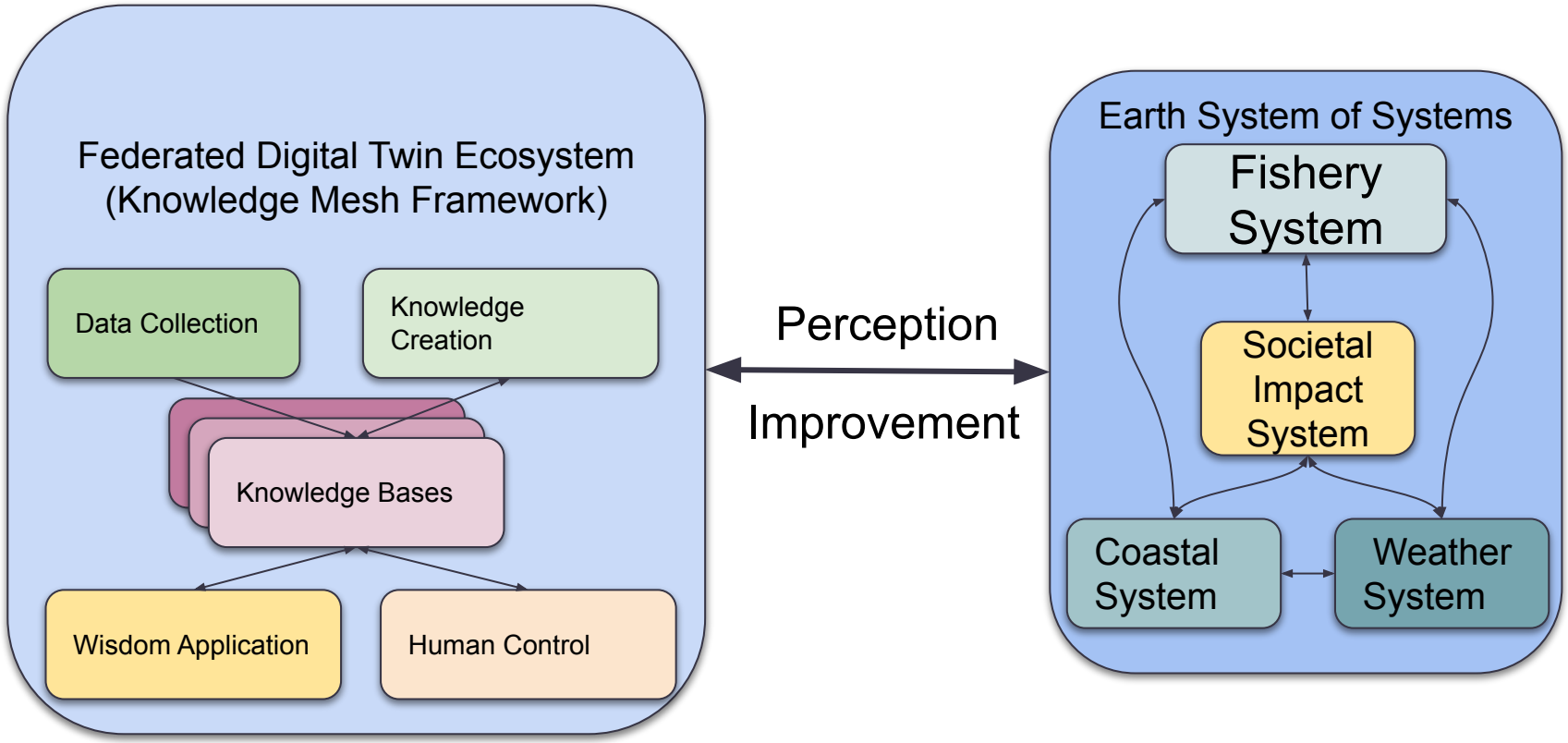


Tactical Activities for Realization

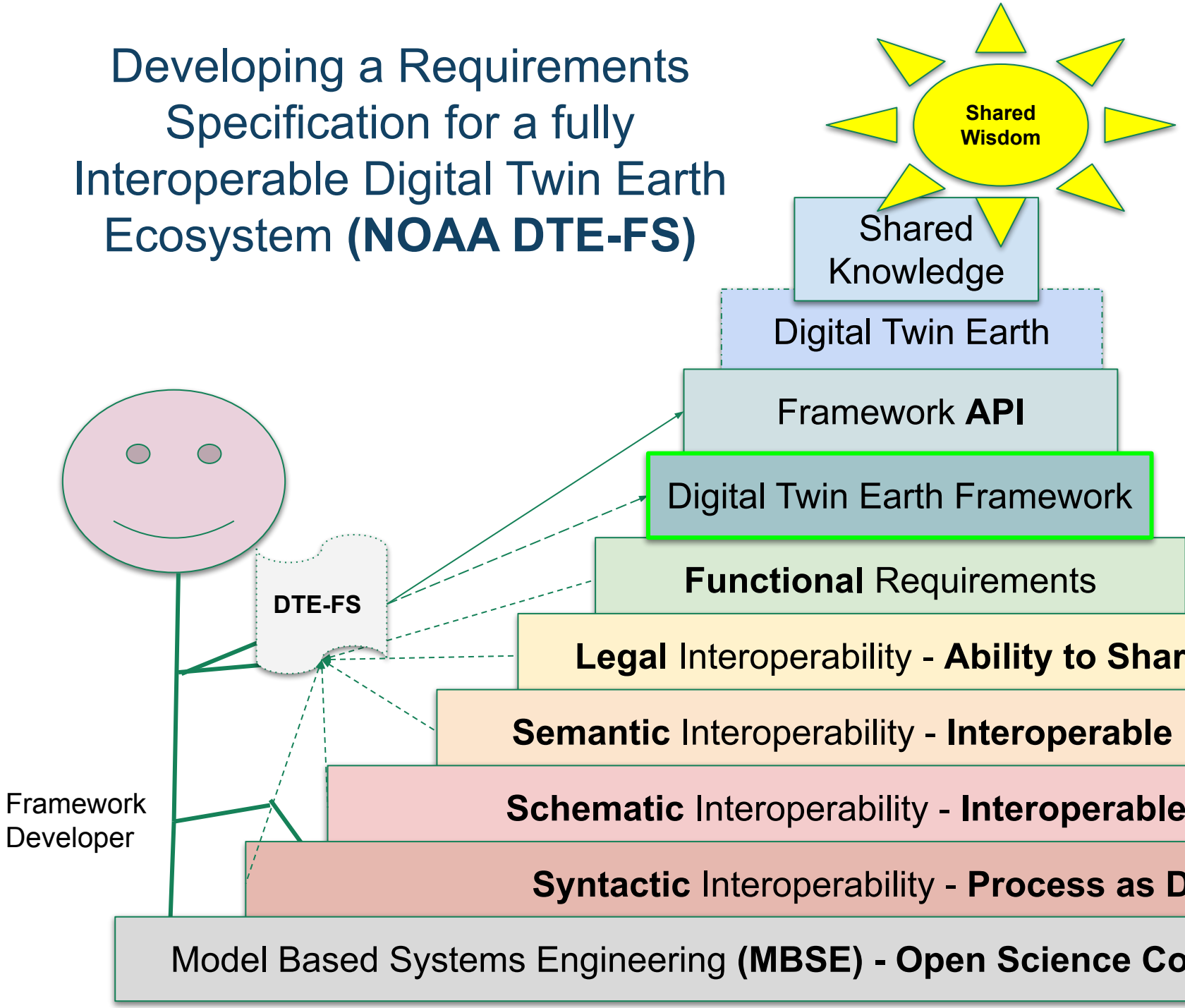
Defining Earth Digital Twins
and **Mapping** Essential Features & Capabilities

- **Data Collection**
  - Heterogeneous Physical Sensor Networks
    - e.g. Fishing Vessel Tracking
  - Pan-Domain Knowledge Contribution
  - Internet-of-Things, Big Data
- **Knowledge Creation**
  - Data Integration
  - Environmental Modeling
  - Human Impact Modeling
  - Response Modeling
- **Wisdom Application**
  - Shared Knowledge Mesh
  - Automated Alert System
  - Automated Sensor Control
- **Human Control**
  - Decision making Dashboard
  - Human Intervention
  - Knowledge Consumption
- **Knowledge Bases**
  - Historical Data
  - Ecological Knowledge
  - Adjacent Inferred Knowledge
  - Socioeconomic Knowledge
  - Regulatory Policy Knowledge
- **Improvement of the Earth System**
  - Sustainable Resource Management
  - Ecosystem Preservation
  - Collaborative Governance

Something like… *"a virtual, dynamic, up-to-date representation of a multi-user earth system of systems (ecosystem) that enables making and tracing decisions about the earth and their outcomes through the use of process techniques including simulation, machine learning, and reasoning."*



Developing a Requirements Specification for a fully Interoperable Digital Twin Earth Ecosystem **(NOAA DTE-FS)**

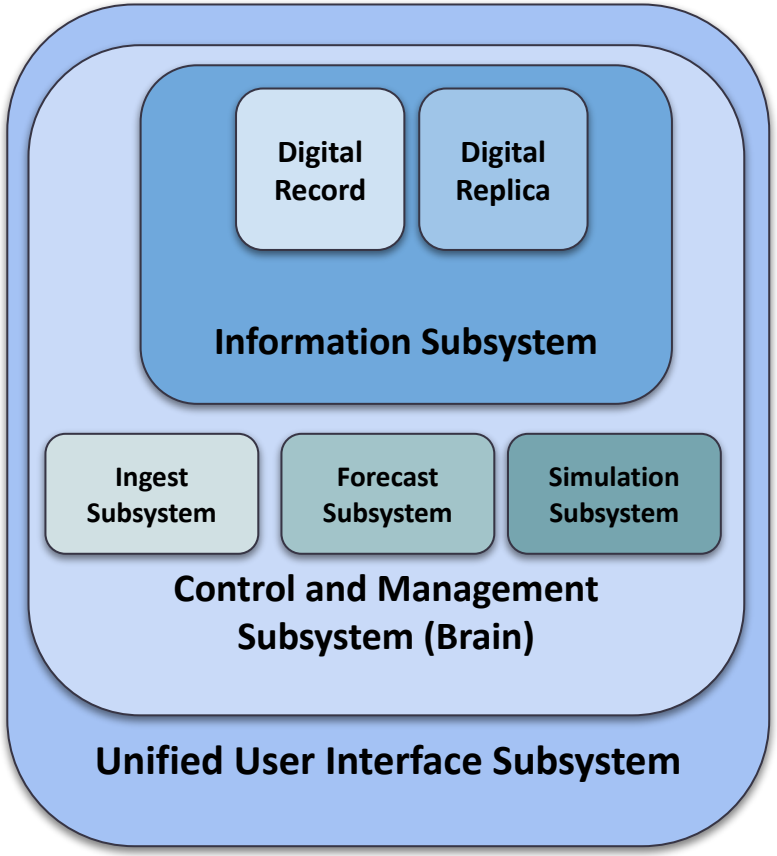


Notable Characteristics and Enablements

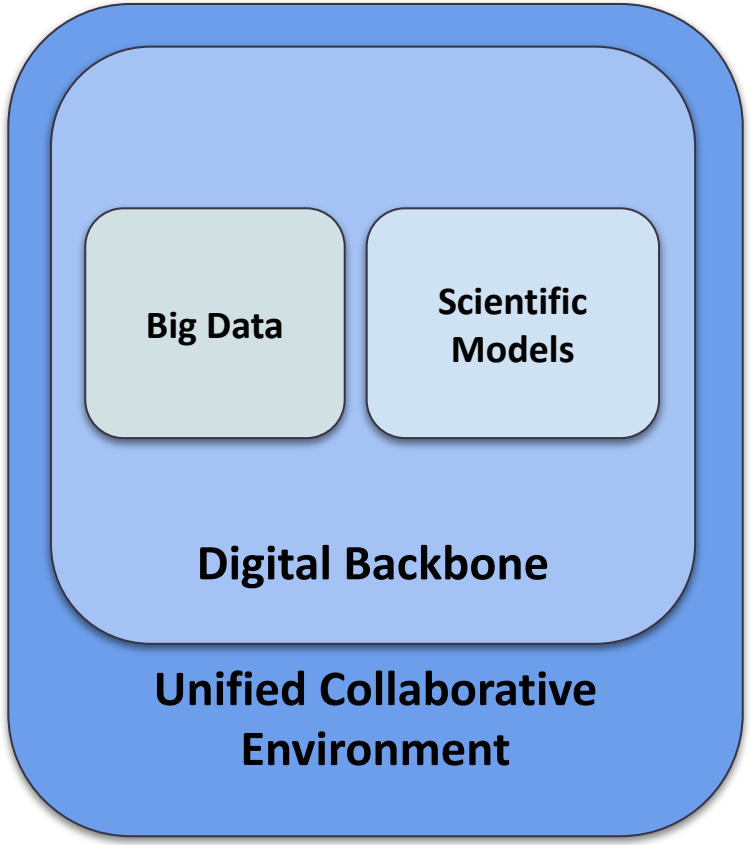
- **Digital Twin Foundation Model**
  - 'build me a twin that..'
- **Interoperable Digital Twins**
  - 'how does this change affect…'
- **Full Enterprise Knowledge**
  - 'give me a capability that…'
- **Democratization of Archive**
  - 'add my API/model…'
- **Use Driven Discovery**
  - 'what else is relevant to…'
- **Control and Query**
  - 'reconfigure this sensor network…'
- **Orchestrated Auto-evolution of capability**
  - 'plan me a sensor fleet based on…'

Empirical Studies of Practical and Emerging DT **Ecosystem Components**

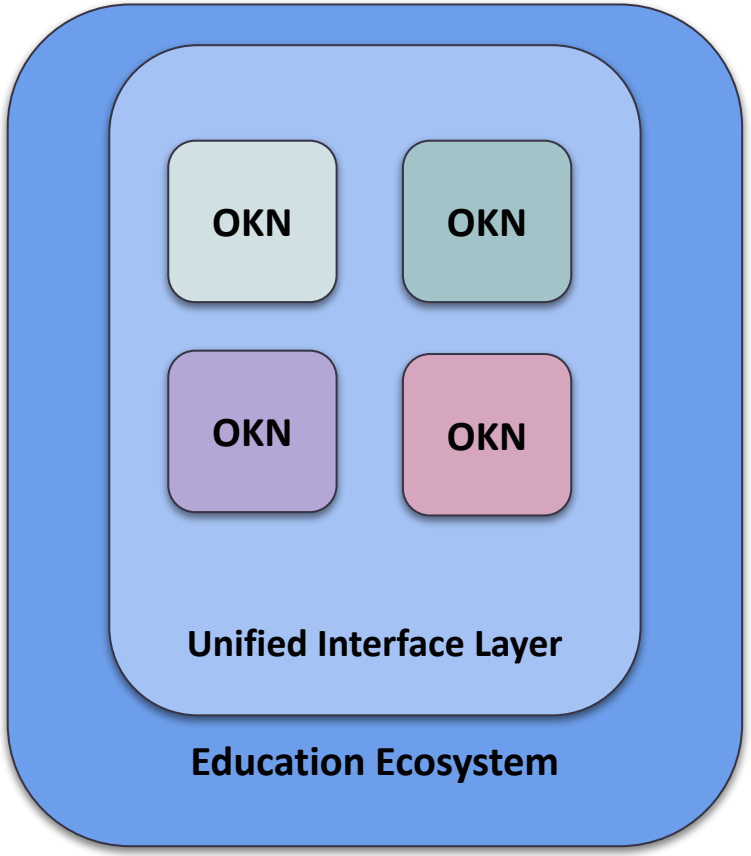
Studying features, capabilities, characteristics, & relationships of interoperable **Agents**, **Fabrics**, and **Meshes**



Earth System Digital Twin (ESDT)



European Digital Twins of the Ocean Ecosystem (DTO)

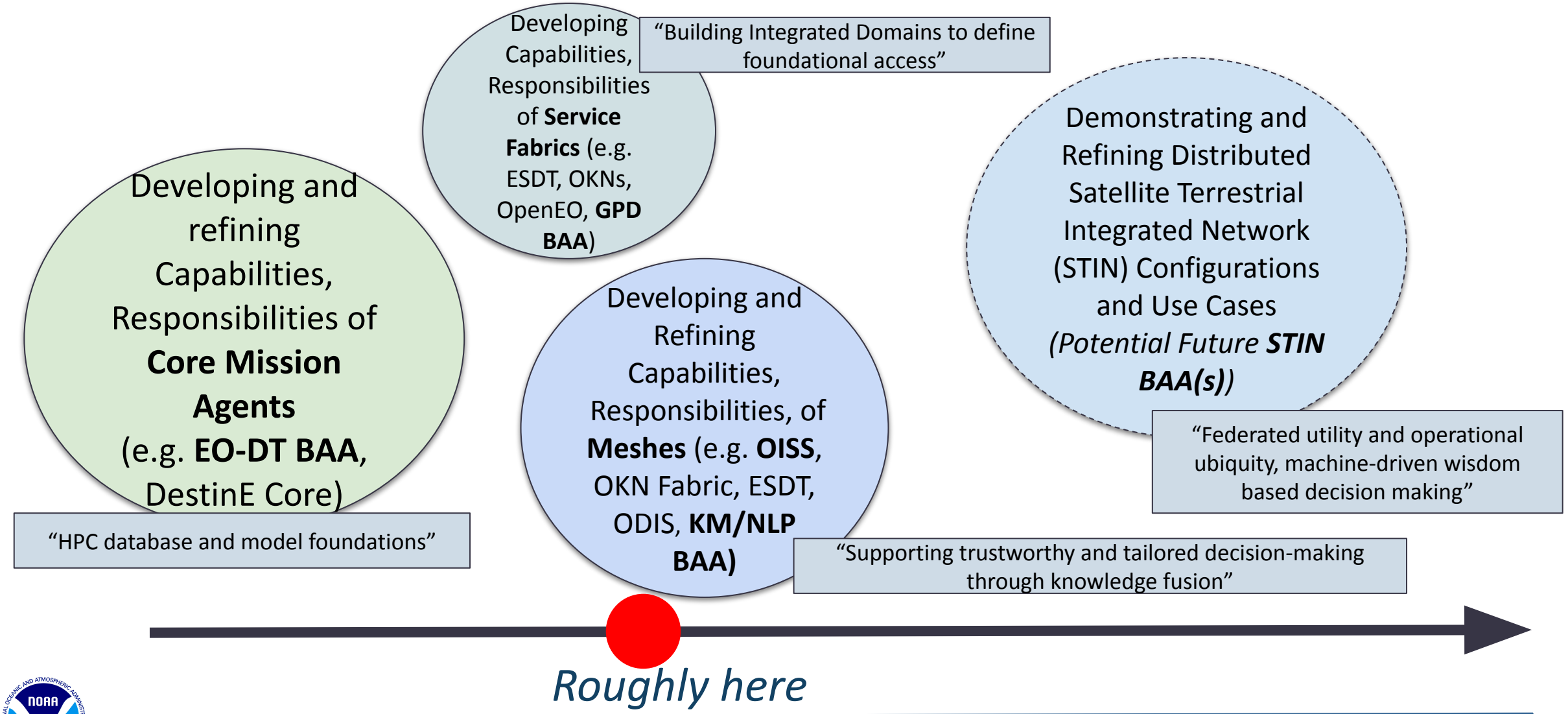


Prototype Open Knowledge Network (Proto-OKN)

Relating and Nurturing an **Open, Comprehensive,** and **Inclusive Community of Practice**

- **Standards and Best Practices Development**
  - CEOS Interoperability Framework (CEOS Members)
  - OGC Climate and Disaster Resilience Pilots (CDRP) (International OGC Members)
  - IEEE Digital Twins of the Earth Study and Standards Groups (Industry, DITTO, DestinE, NOAA, NASA, etc.)
  - NSF Academies DT Foundational Research Gaps Fast Track Action Committee (Cross Domain)
  - Spatiotemporal Thinking, Computing, Applications (large uni cohort setting 5-10 year research agenda)
- **Technical Tools and Capabilities**
  - The NESDIS Common Cloud Framework Open Information Stewardship Service (NCCF OISS) operational knowledge mesh/interoperability framework
  - Uncrewed Systems (UxS) Data Factory
  - National Weather Service Data Fabric
- **Collaborative Studies, Testbeds, and Pilots**
  - NASA Earth Systems Digital Twins (ESDT), and the Coastal Zone Digital Twin Pilot (NASA/NOAA/CNES/NSF)
  - AI For Public Good (NOAA/ECMWF/DestinE)
  - NSF Proto-Open Knowledge Network (NASA, NOAA, USGS, NIH, NIJ, NREL, USDA)
  - EO-DT, Knowledge Mesh/NLP NOAA Broad Agency Announcements (Industry, Academia, NOAA, NASA)
  - Iliad Cohort Exploration - DTO, Ifremer, NOAA

Pathfinding Future Directions through targeted efforts including **JVP BAAs**



Establishing an Agile R2O(2R) Development Pattern

- Using shared semantics via a (reference model based) interoperability framework allows creation of truly portable resources, as the **same process can run on any implementation**
  - Process as semantically interoperable data - it's just (open) config files, and processes contain full semantic representation
  - Easy to set up data-driven promotion tiers, with fast read-eval-print-loop (REPL) based development
- Using shared and fully open standards in protocol and vocabulary (w3c, ogc, esip, etc.) means that the 'source of truth' (the full **digital record**) is both maximally portable, maximally transformable to other formats (like virtualizarr, umm-g/c, parquet, etc.) for specific uses (multi-world **digital replica**), and maximally accessible (multitude of specification based tools can retrieve and operate on the record-held data/metadata directly, using a single, static, API specification)
- Tailored implementation of a computational reference architecture (**control and management subsystem**) means that process management is universally defined, stable, and easily controlled
  - Platform specific implementations tailored to need profile, e.g. **Open Information Stewardship Service (OISS)** serving as the **Stewardship Service of the NESDIS Common Cloud Framework (NCCF)**
- Complete process provenance, together with explicit process definition, enables **a single (unified user) interface to full enterprise knowledge**
  - Full results of what now, what next (forecast), what if (simulation) explorations are produced at runtime and immediately available for discovery (**instant refresh digital record**)

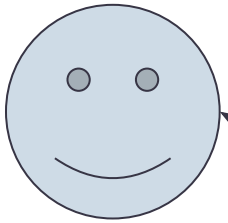
Knowledge Mesh Key Concepts

Knowledge Mesh (KM) in Technical Context

- A KM includes the full content (the knowledge itself), but it is common to shorthand/couple KM with the architecture that builds it
  - A KM is a 'special' type of **knowledge graph**
  - A KM represents the collective contribution from users - 'data mesh' alignment
- The framework that produces a KM is conformant to the DTE-FS
  - Supports **communications between an ecosystem of digital twin technologies** (leveraging linked data to orchestrate agents, fabrics, meshes) **as well as** construction and operation of **fully independent digital twins**
    - *Understanding - Think about how different domains depend on each others activities, implicitly or explicitly, through relationship - environment, ecosystem, land use, geophysical, medical, built environment, etc*
    - *Understanding - Think about the particulars of how we combine different 'resolutions' of information domains - atomic to terrestrial to cosmic*
- Use of the mesh architecture ensures all resources in the mesh are intrinsically interoperable, and that **the collective system represents a 'digital twin earth' that naturally conforms to the (need-based) expressions of its users**
- **The NCCF Stewardship Service (OISS) KM is also conformant to the NASA ESDT Architecture Framework**
  - This might apply generally due to relationships between DTE-FS and the ESDT-AF

Illustrating by Example

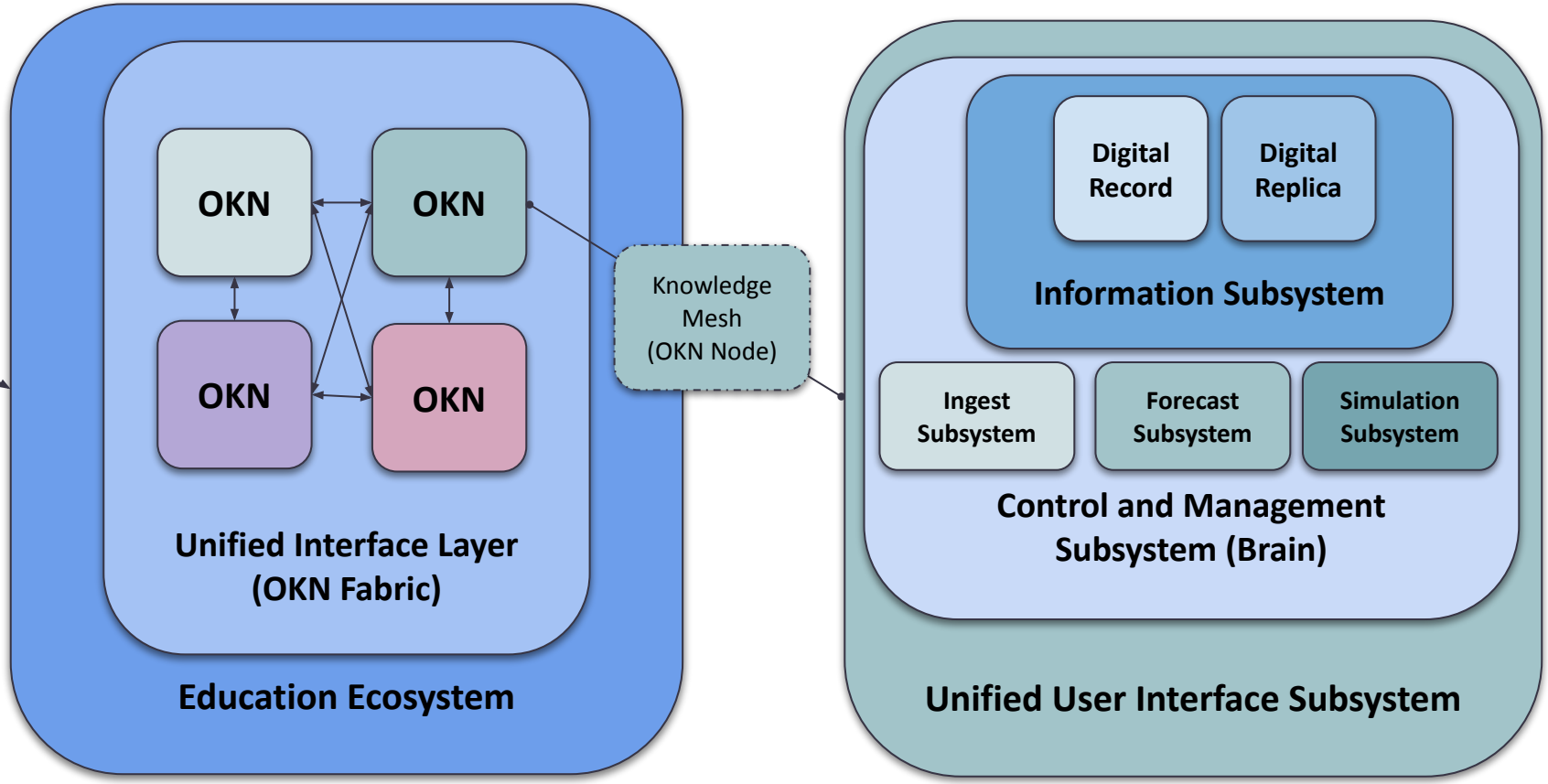
Exposing the Coastal Zone Digital Twin (CZ-DT) (a NASA ESDT) as an NSF OKN Node



"What if a 100 year flooding event occurred in the Chesapeake Bay - what direct seasonal impact would it have on bird migrations?"

"What impact will this next storm have on fishing this year?"

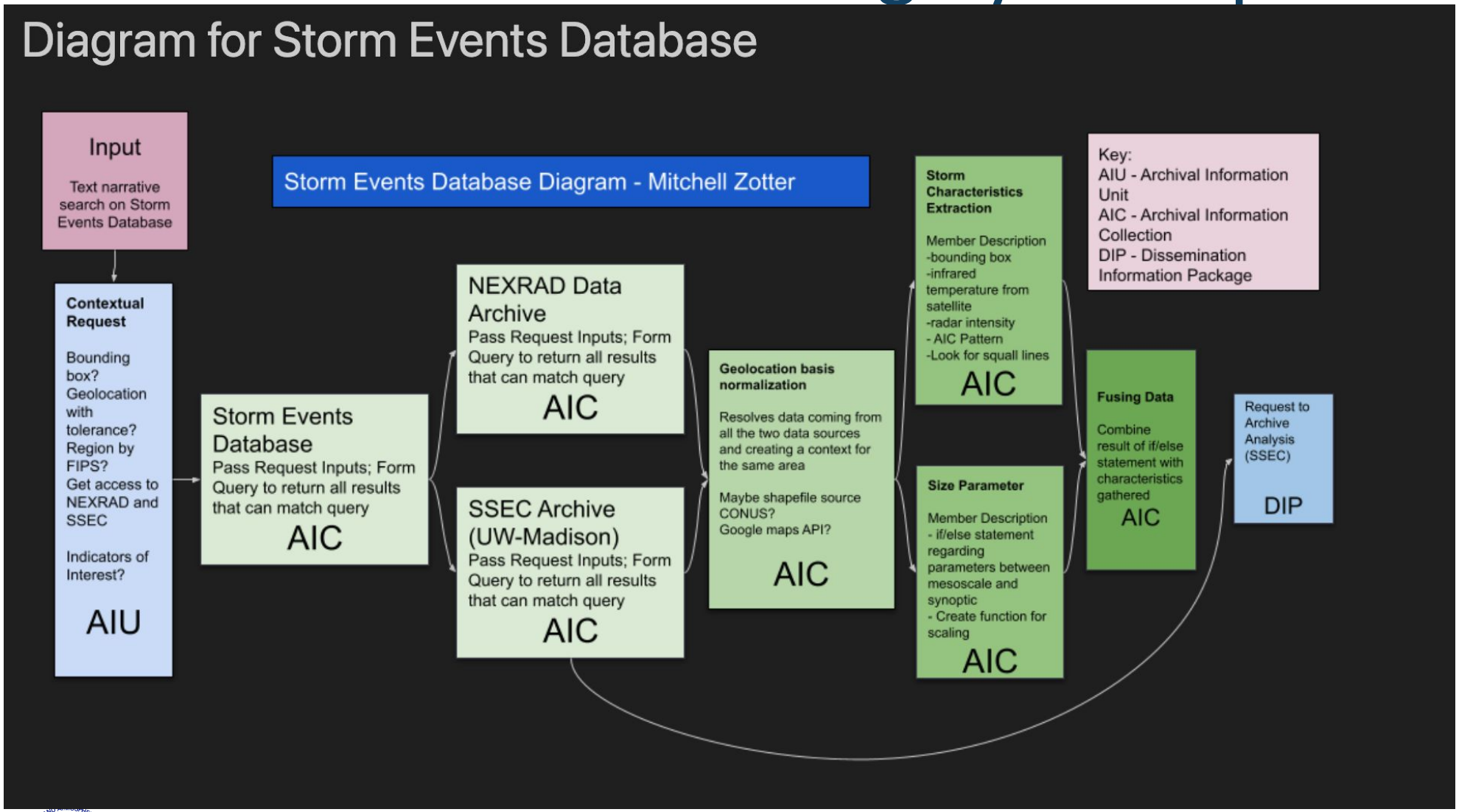
"Please run me an ensemble of scenarios to identify some of the areas of concern this year."



Prototype Open Knowledge Network
(Proto-OKN)

Coastal Zone Digital Twin
(Earth System Digital Twin)

Illustrating by Example

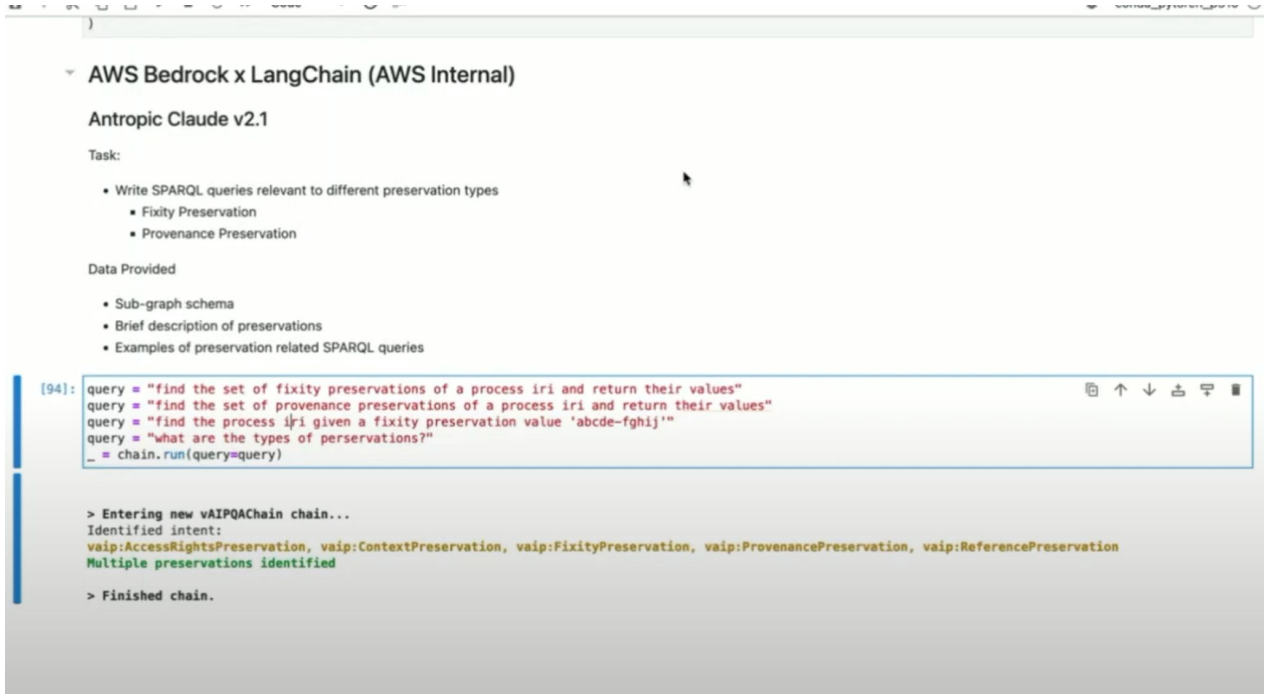


Data Fusion Process

"Build an online process to augment event details of storm events database results with satellite data from SSEC and NEXRAD data."

Leveraging the KM via a unified language Interface

- By its layered interoperable nature all data in the KM may be exposed through a single unified interface
  - We can use the interface to discover, compose, modify, and execute processes
  - As use increases, the information known to the KM grows, creating a holistic foundation model for the digital twin ecosystem
    - Machines can eventually leverage this to take over many duties - model selection, process composition, process execution, etc.
  - In addition to leveraging advanced AI/ML models and actors within processes (e.g. AI agents), the KM serves as an agent itself
    - E.g. - we can put advanced chat interface over the mesh, creating a conversational interface of control to the mesh and all of its contents
- For OISS (NCCF Stewardship Service), this interface and all of its contents is being called the ‘OISS marketplace’
  - Essentially - in the Agent/Fabric/Mesh view, where OISS is seen as an agent, the marketplace is a single-endpoint fabric
- **We can call other agents and fabrics by calling the KM fabric!**



KM NLP Study and Status

Study goals, among other things:

- **AI Safety**
  - Ensuring trust through provenance and uncertainty management
  - Maintaining mission control and authority
  - Providing controls for model evaluation, validation, and evolution
  - Studying advanced access, consumption, and presentation of results
- **Answering Advanced Questions for Decision Support leveraging NLP-based control of Complex 'what now, what next, what if' questions**
  - "How will this upcoming hurricane affect bird migrations in Florida this year?"
  - "Can you help me develop my fishing plan for this season?"
  - "Help me plan for potential solar E&M disasters for insurance"
- **Federated Interoperability in Processing**
  - Using integrated data and processes from across NOAA, the federal government, etc
  - Testing and driving standards adoption

NOAA to Study New Technologies for Enhancing Environmental Intelligence

May 10, 2024

NOAA announced the release of a Broad Agency Announcement (BAA) to review the capability and feasibility of deploying advanced technologies to expand the discovery, access, and use of environmental intelligence for the purposes of social environmental equity and enhancing weather and climate resilience.

NOAA is undertaking this study after recognizing the need to improve and expand use, reproducibility, and trust in its systems and information to inform weather and climate decision-making. This BAA represents, in part, a follow-on to the Earth Observation Digital Twin BAA released in September 2022.

NOAA NCCF Open Knowledge Mesh

NOAA Works with Industry to Determine Natural Language Processing Capabilities with the NCCF Open Knowledge Mesh

September 19, 2024

On Sept. 13, NOAA awarded Knowledge Mesh/Natural Language Processing (KM/NLP) contracts with a cumulative value of $5.4 million to four companies:

- Booz Allen Hamilton, of McLean, Va., for $1,084,718.69, with a 12 month period of performance;
- Element 84 Inc., of Alexandria, Va., for $1,195,618, with a 24 month period of performance;
- Noblis Inc., of Reston, Va., for $1,203,452, with a 15 month period of performance; and
- Orion Space Solutions, an Arcfield Company, of Louisville, Colo., for $2,036,930, with a 24 month period of performance.

These awards are to examine, deliver, and iteratively demonstrate working models, features, and capabilities related to the development and management of an enterprise information processing system, backed by a knowledge mesh of process contextualized, semantically enabled facts that is capable of dynamically leveraging Application Programming Interface (APIs). Further, using the knowledge mesh capability for the purpose of answering highly contextual queries from a number of personas representing a diverse swath of NOAA users. In part, this Broad Agency Announcement (BAA) represents an idealized follow-on to the Earth Observation Digital Twin (EO-DT) BAA.

These contracts will explore the development of techniques for future applications, but they are not an operational mission or associated with an operational mission. The scope of work supported by this BAA is exploratory in nature.

NOAA completed the award on Sept. 13, following a competitive solicitation that resulted in 27 proposal submissions, and a thorough evaluation of the solutions received.

The contracts are managed by the Joint Venture Partnerships program, which is a program of the Office of Systems Architecture and Engineering, NOAA's National Environmental Satellite, Data, and Information Service.

For additional information, please see: https://www.nesdis.noaa.gov/news/noaa-study-new-technologies-enhancing-environmental-intelligence

- Utilize a knowledge mesh of process contextualized, semantically enabled facts
- Dynamically leverage Application Programming Interface (APIs)
- Answer highly contextual queries from a number of personas representing a diverse swath of NOAA users
- Follow-on to the Earth Observation Digital Twin (EO-DT) BAA

ESTO Earth Science Technology Office

IS

Highlighted Goals and Opportunities

- Study different domains **independently** using the **shared KM reference model**
  - **Diverse persona driven views of regional concerns and their effects**
  - Apply **unique solutions that implement their own agents, fabrics, and meshes**
- Demonstrate intrinsic **interoperability** of the studies
  - Leverage **common use of OISS** (and the OISS Marketplace Fabric) to demonstrate native interoperability and **relationship between study solutions**
- Demonstrate **interoperability** between **studies and operations**
  - Leverage **OISS** to provide **intrinsic links between operational upstream and downstream products and the study efforts**
- Demonstrate interoperability between **studies and other efforts**
  - Leverage **OISS** to demonstrate connection between studies and **CZ-DT, OKN integration, DestinE integrations, EO-DT integrations**
- **Develop database for digital twin foundation model - future studies could look at exploiting this!**
- Provide solution pathways for high profile (national/international) gaps, like **uncertainty combination, reduction, communication**
- Provide a baseline to design follow-on studies - moving beyond query into topics including coupled systems control, auto-evolutionary capabilities (leveraging foundation model), and overall demonstration of earth system improvements, validated via uncertainty/error function tracing

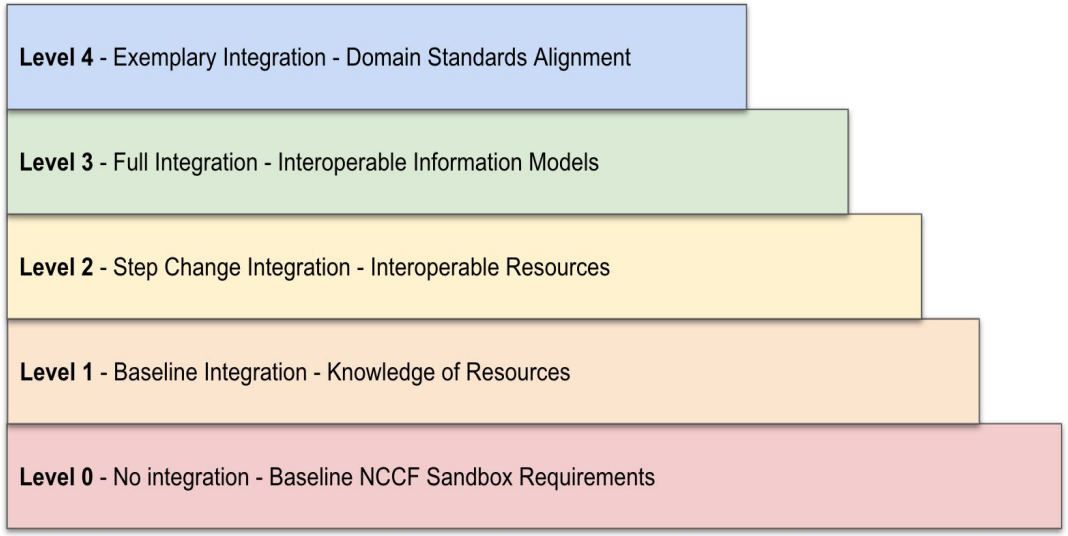
Backup/Supportive

Enabling **Iterative Transformation** at NOAA

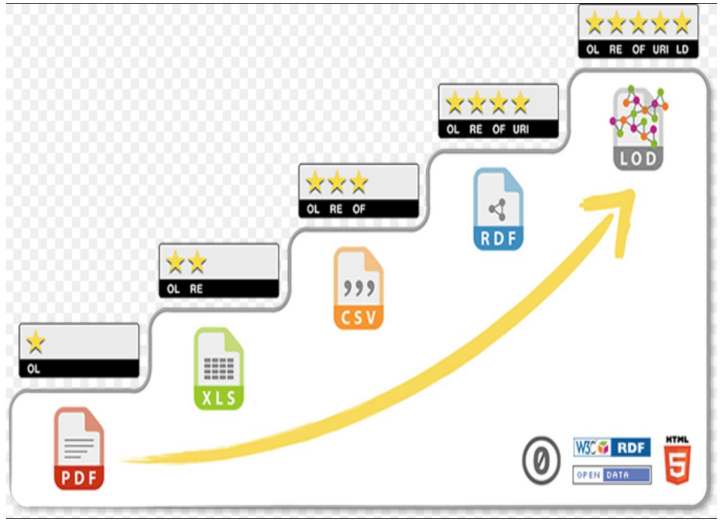
Working toward a sustained cooperative continuous improvement model



Open Information Stewardship Service (OISS)
Integration Maturity Scale



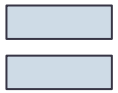
5 Star Linked Open Data



Domain of Authority

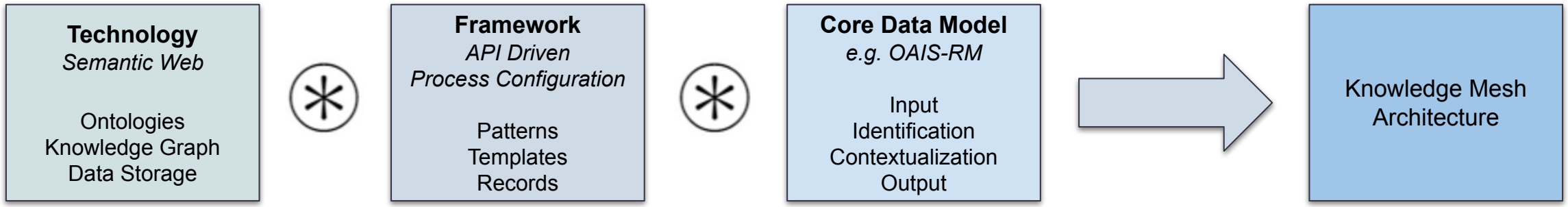


Recursive Stepwise Integration



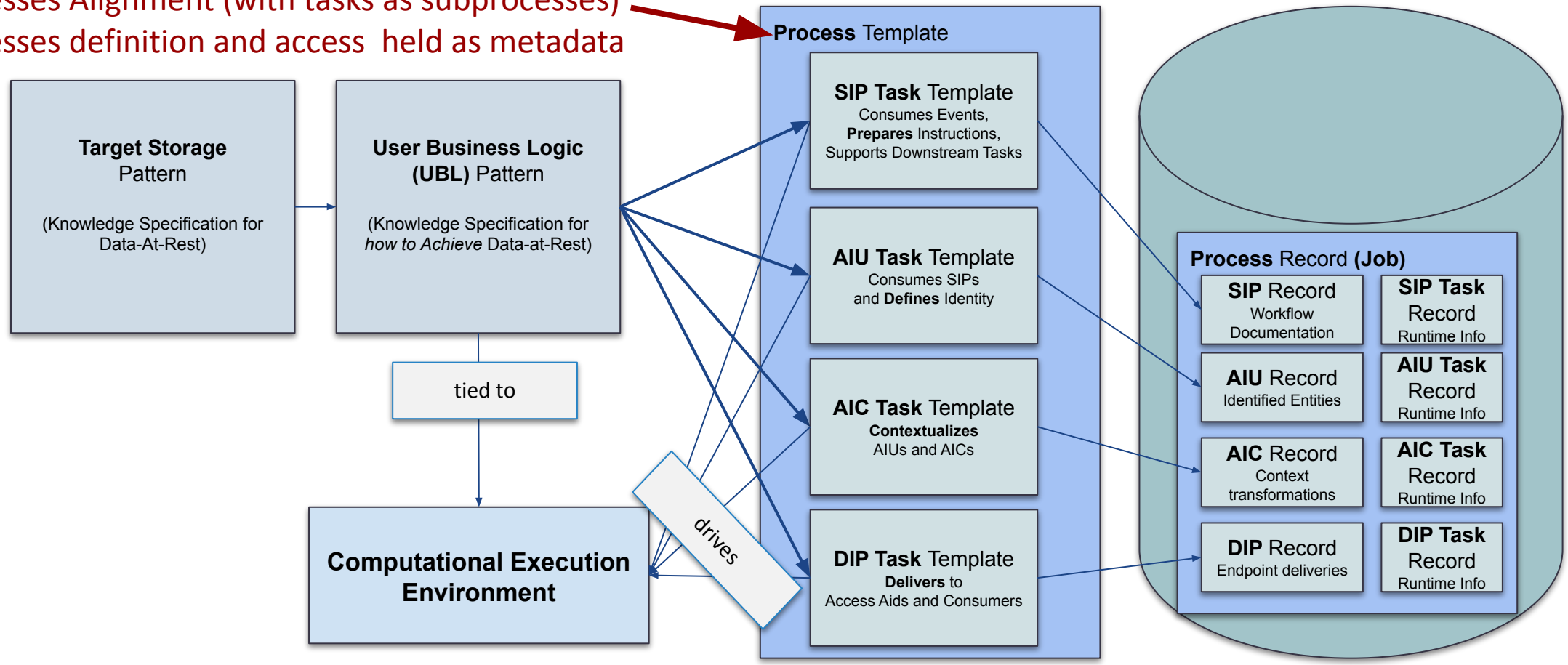
Iterative Increases to Openness

Framework Concept Convolution

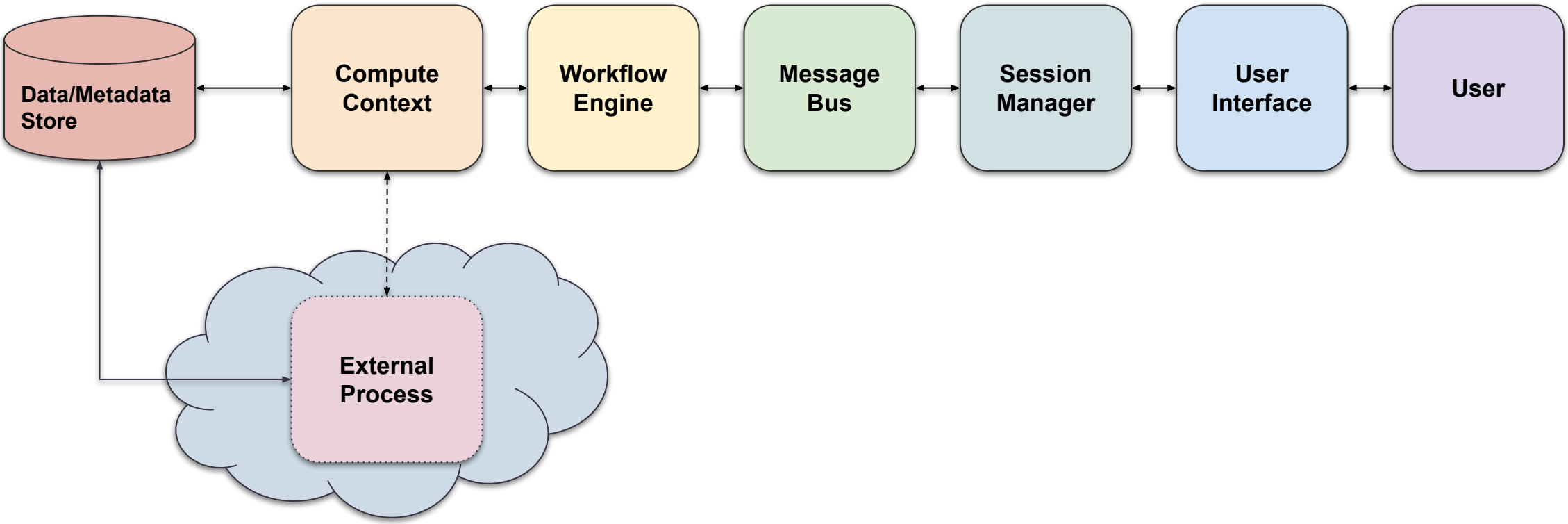


Semantic Topology

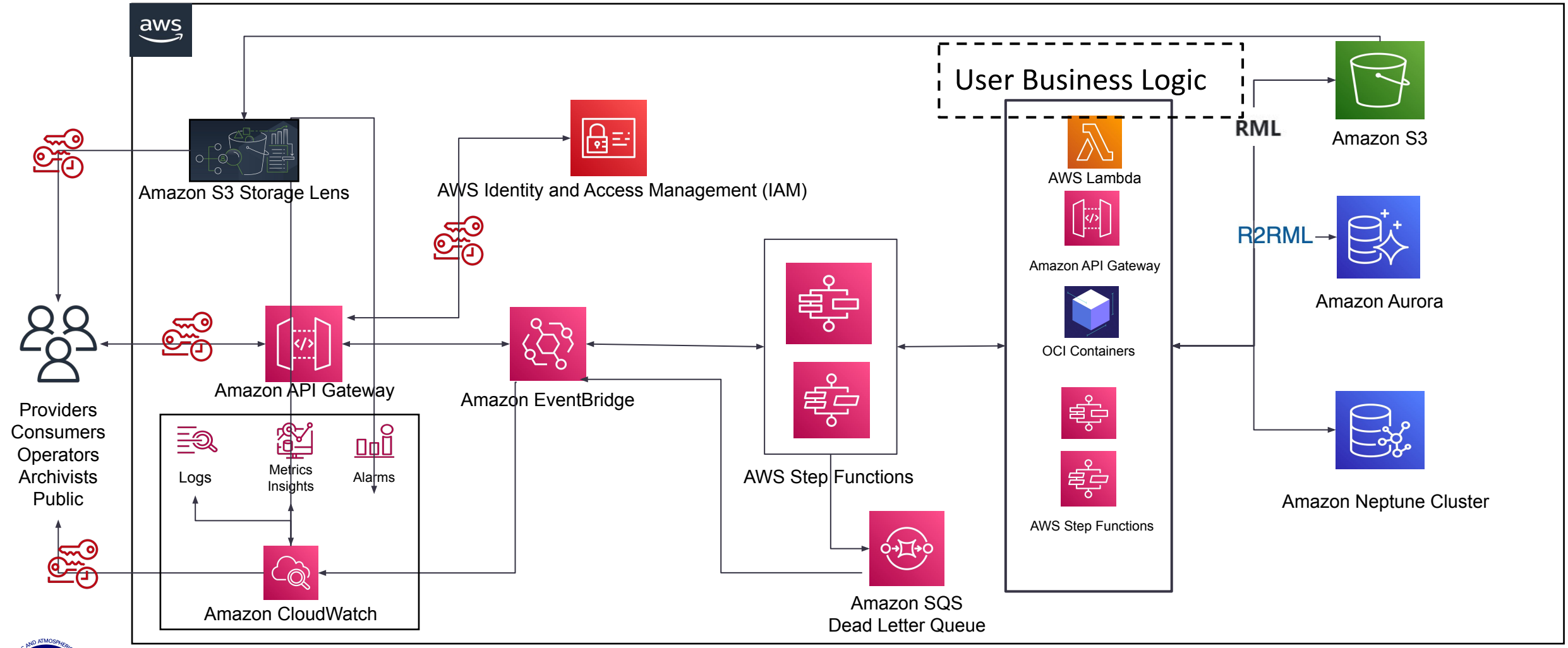
OGC Processes Alignment (with tasks as subprocesses)
OGC Processes definition and access held as metadata



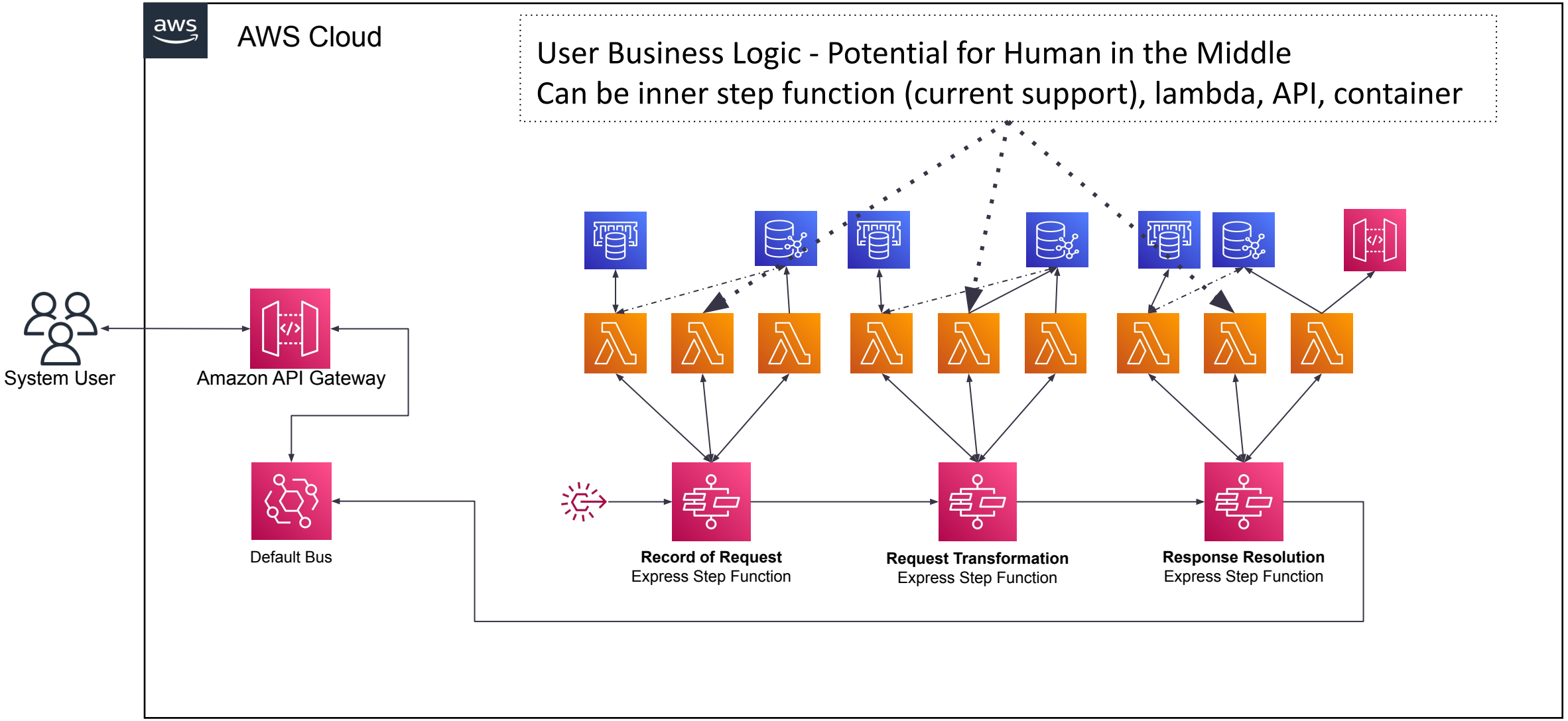
Evolved System Model for Access and Archive (Knowledge Graph) Framework (2023) Agile Reference Architecture



1.1.0 Release Configuration - API-Driven Bidirectional Web Use

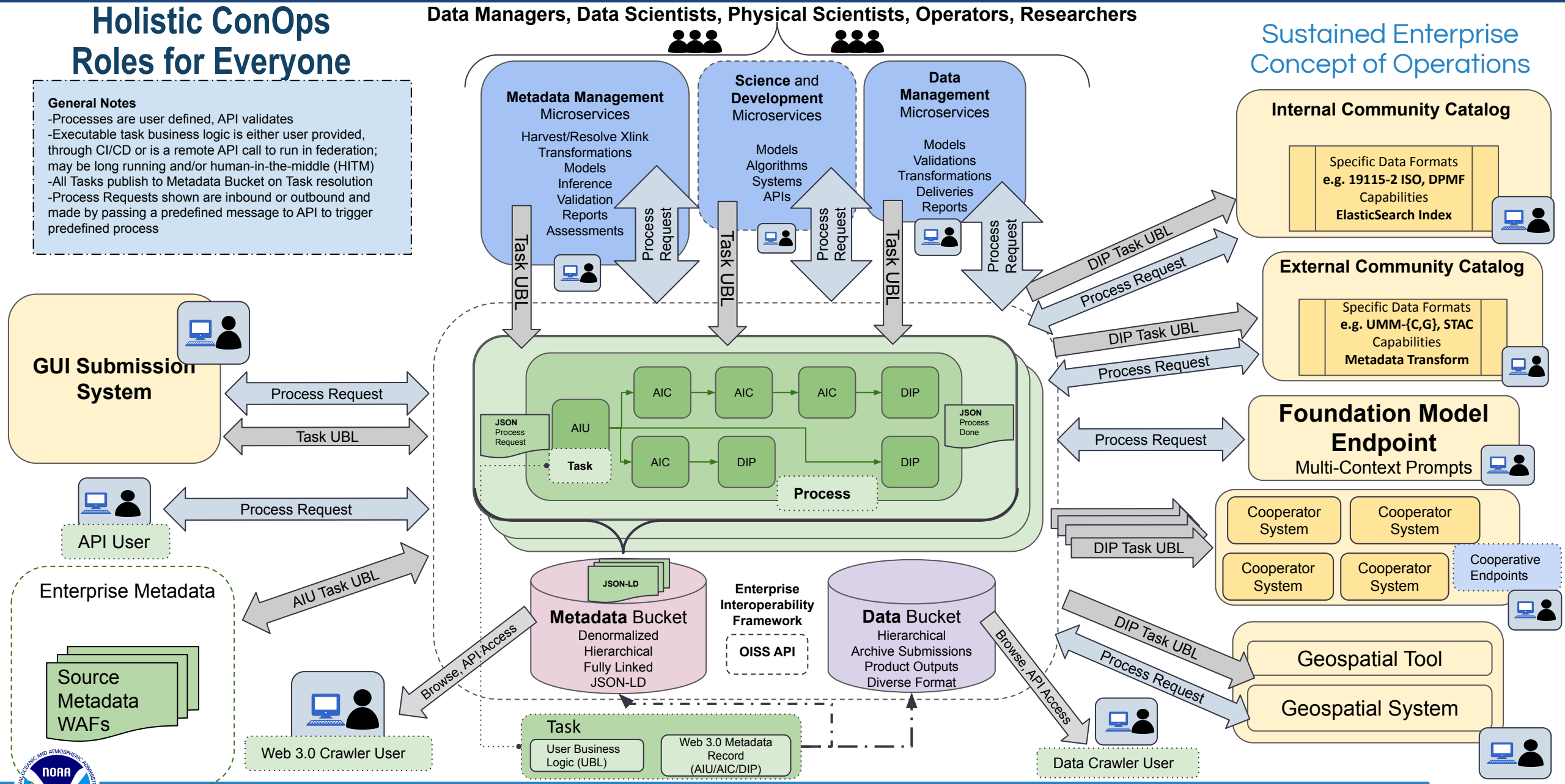


Simple (Non-Branching) Workflow Illustration

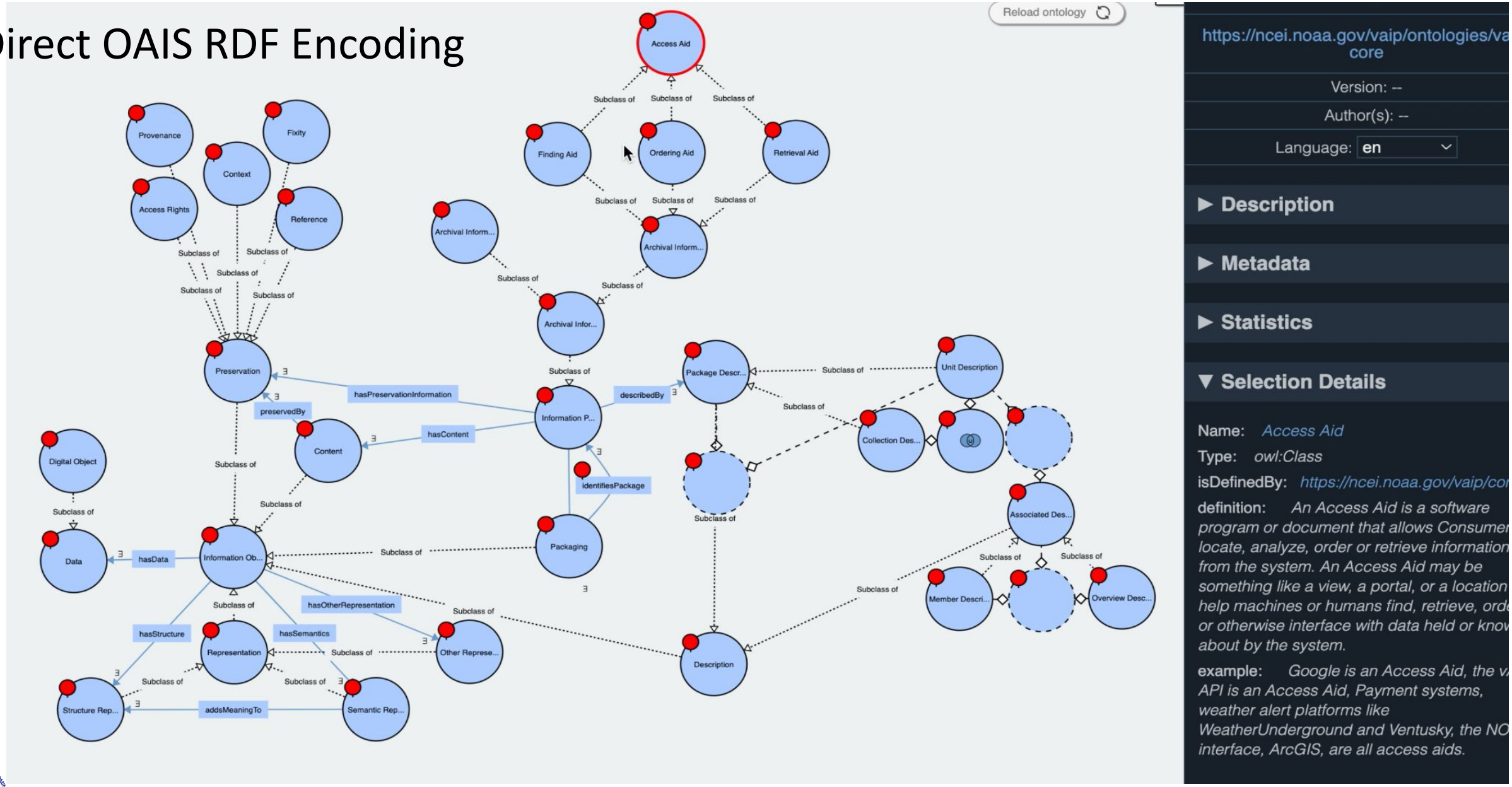


Holistic ConOps Roles for Everyone

General Notes
-Processes are user defined, API validates
-Executable task business logic is either user provided, through CI/CD or is a remote API call to run in federation; may be long running and/or human-in-the-middle (HITM)
-All Tasks publish to Metadata Bucket on Task resolution
-Process Requests shown are inbound or outbound and made by passing a predefined message to API to trigger predefined process

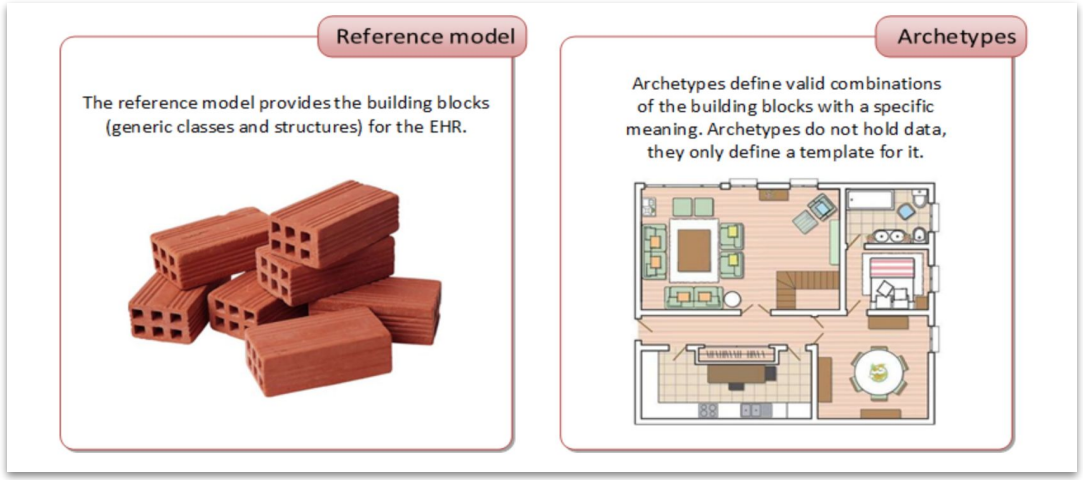


Direct OAIS RDF Encoding

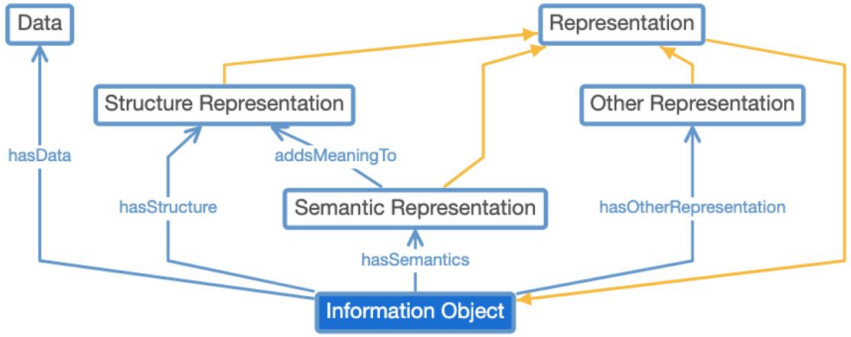


Supporting Flexible Multidimensional Semantic Interoperability

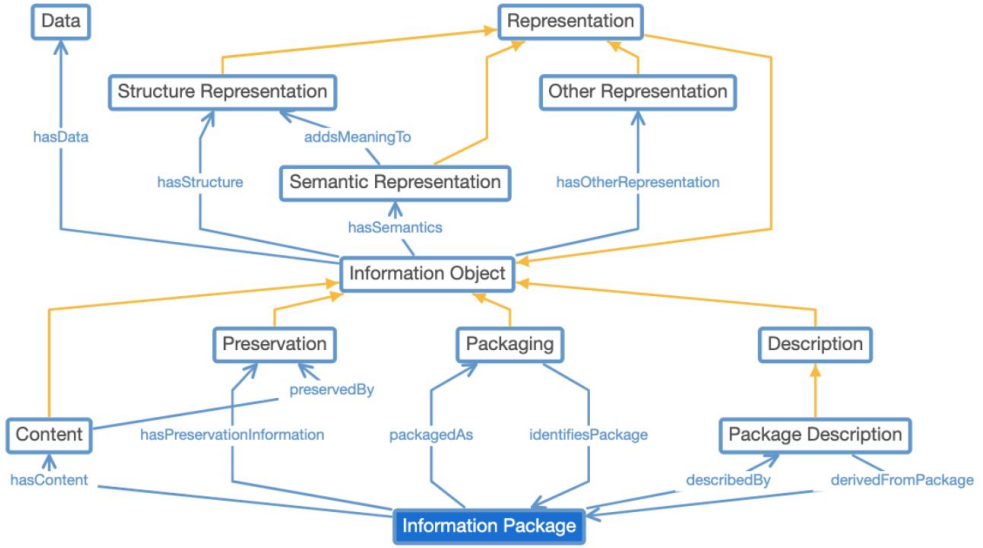
- Process-as-data alone cannot solve interoperability
- Semantic interoperability within a shared reference model is required to enable automated terminology resolution, value transformation, and evolutionary governance
- Critically, Semantic interoperability enables **higher order legal interoperability automation**
- Critically, Semantic interoperability enables decision definition and capture, enabling automated feedback and introduces observability to system improvement



Archetypes: Constraint-based Domain Models for Future-proof Information Systems ISO 13606



A basic foundational building block - recursive and generic



A complete generic model for DTE-FS data

SIP/AIU/AIC/DIP Metadata Documents

Intrinsic Links

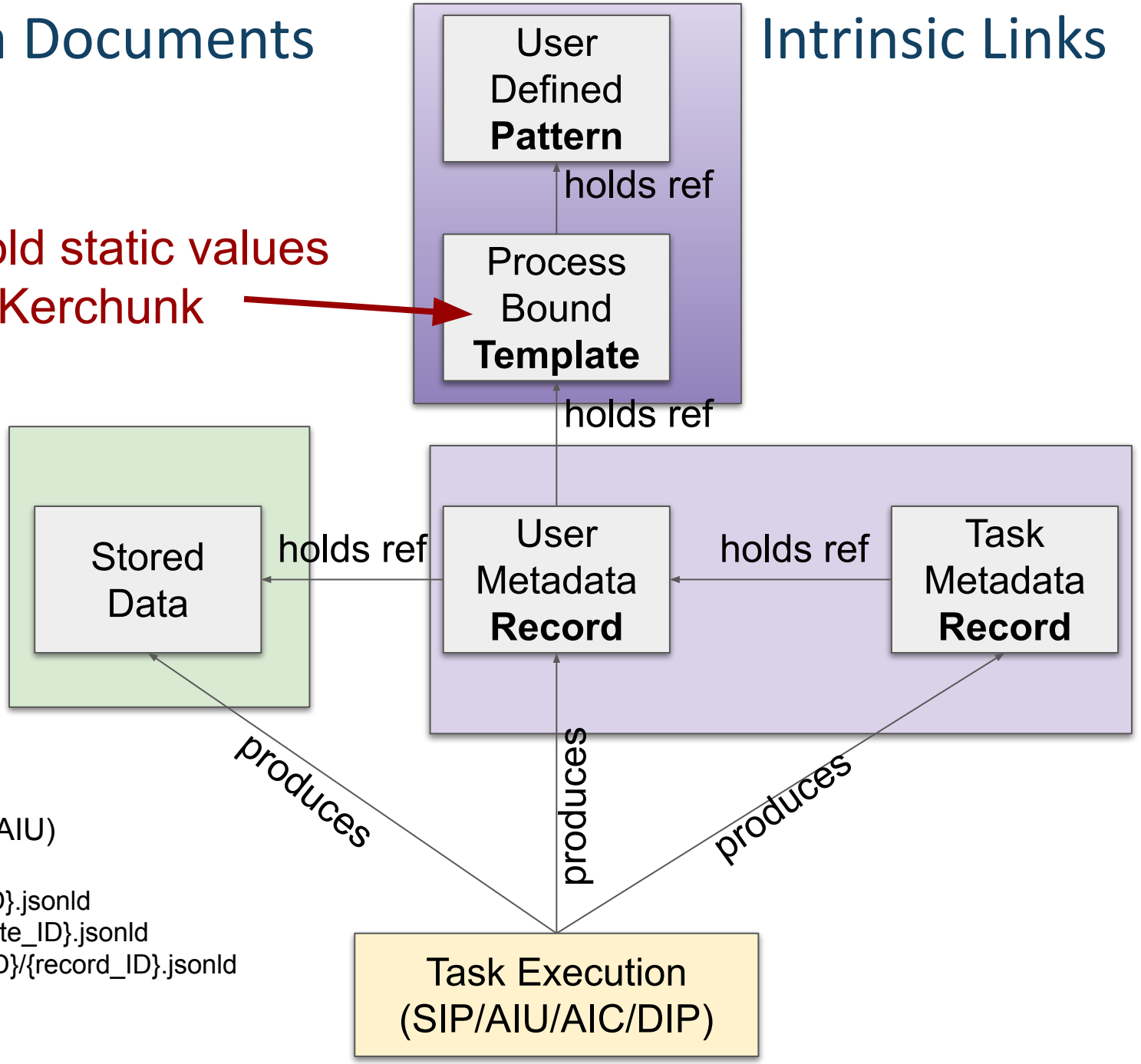
S3 Archive/Access Bucket

S3 Metadata Bucket

S3 Metadata Bucket/Neptune/Cache

Step Function

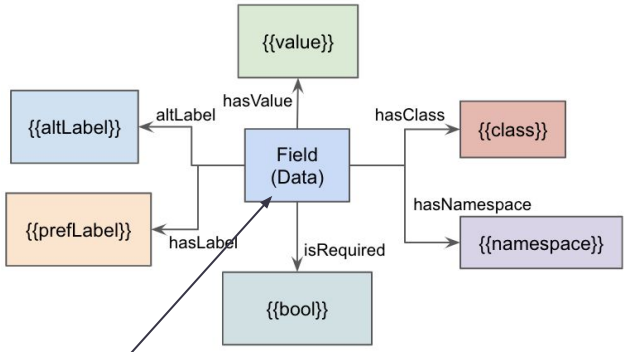
Can hold static values
Think: Kerchunk

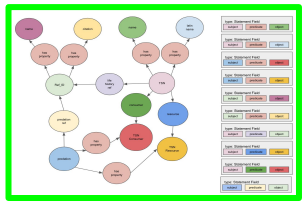


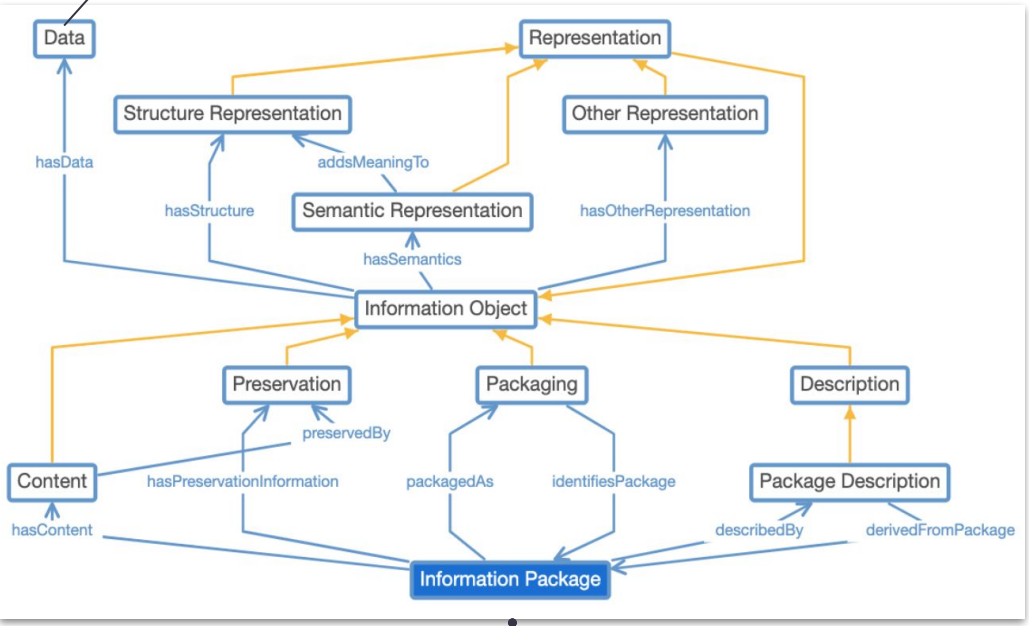
Metadata Bucket Hierarchy (example layout for AIU)

(https) s3://patterns/storage/aiu/{pattern_ID}/{pattern_ID}.jsonld
(https) s3://templates/storage/aiu/{template_ID}/{template_ID}.jsonld
(https) s3://records/storage/aiu/{template_ID}/{record_ID}/{record_ID}.jsonld

OISS Motifs







AIU
(Entities)

"granule" "weather event" "task" "sensor platform" "species" "hurricane" "collection record" "process" "cruise" "estuary" "wetland study" "data manager" "analysis" "UBL"

AIC
(Representations)

"PII Identified" "Data Quality Score" "Weather Advisory" "Lightning density" "DSRR" "Format Transformation" "Climate Impacts" "Extracted Entities" "Food Web" "Trust Assessment" "Hurricane Path" "Hazard Alert" "User Sentiment" "Similar Datasets" "Similar Orders" "Crop Yield Prediction" "Security Score"

DIP
(Access Disseminations)

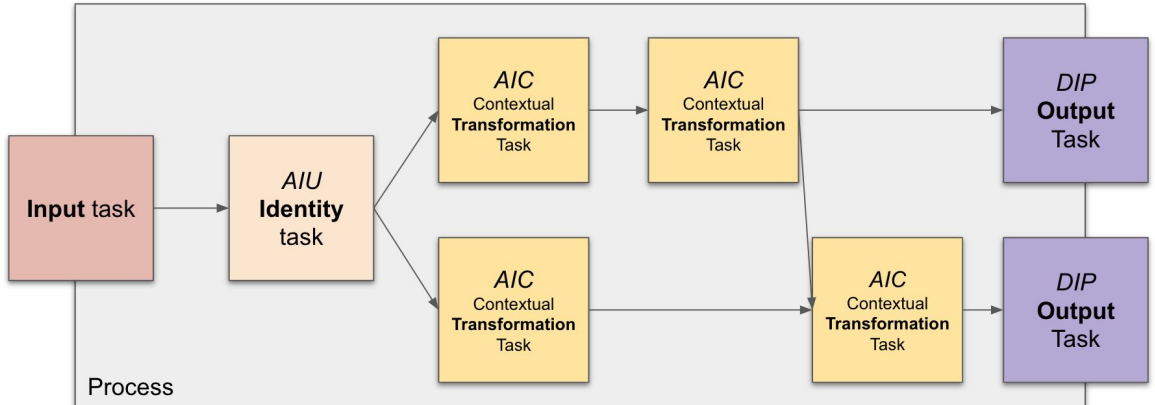
"Publish to OIH/ODIS" "Publish to data.gov" "Broadcast Text Message" "Notify Internet of Water" "Publish to One Stop" "Return API Response" "Send Email Alert" "Publish to DPMF" "Publish to CMR" "Feed foundation model"

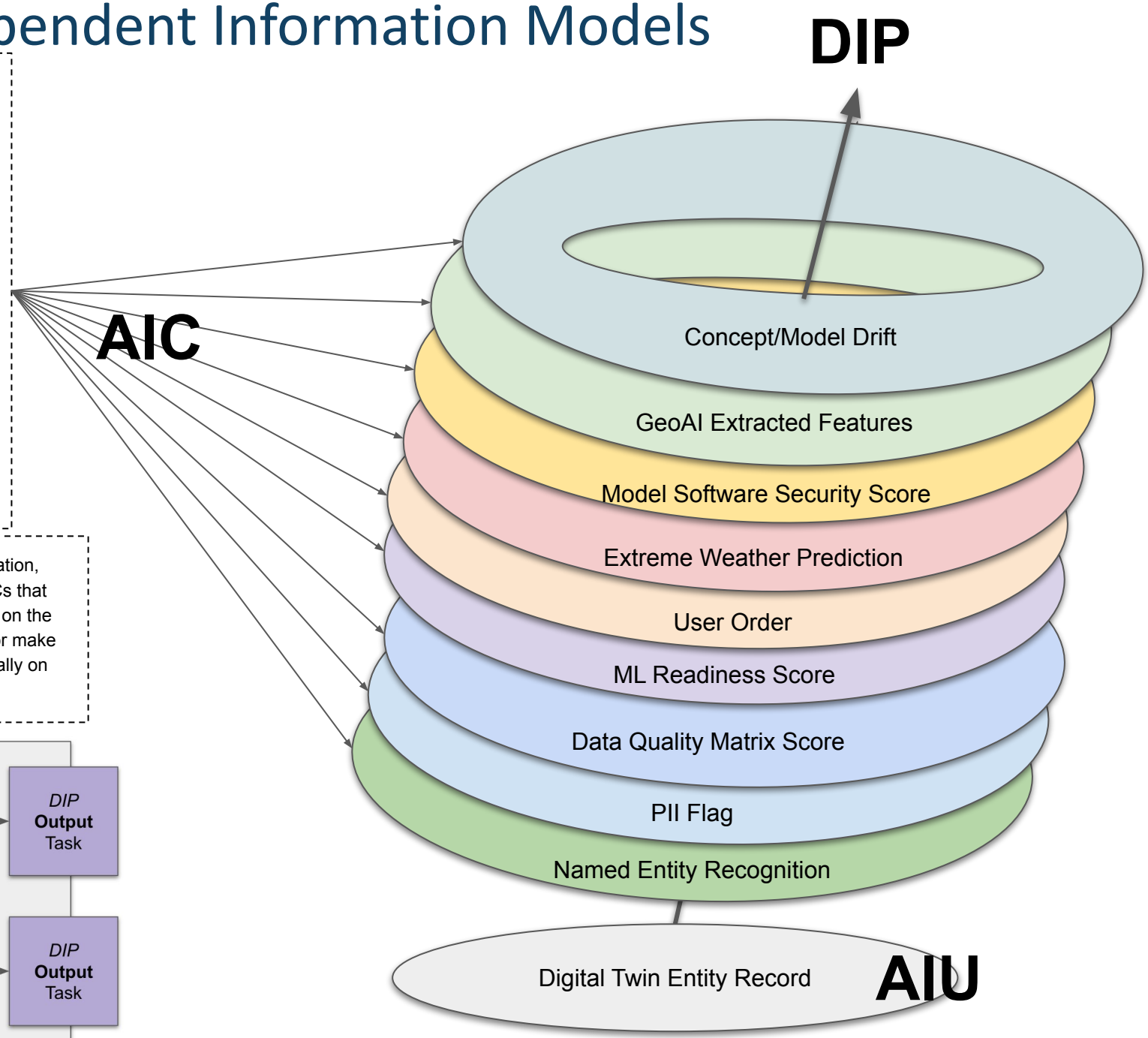
Pattern Flavors (*SHACL Validated*)

Linked Independent Information Models

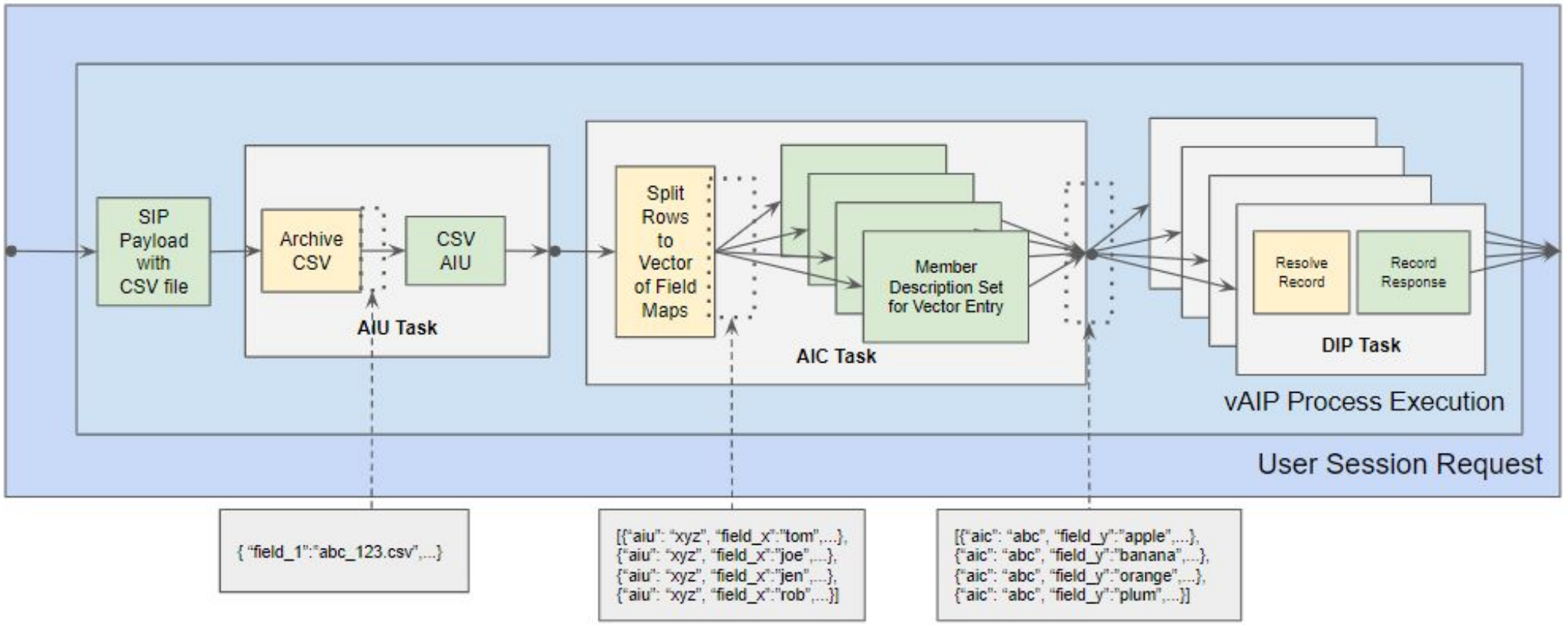
- n-ary References pointers, data, or mixed
- FAIR
  - Findable
  - Accessible
  - Interoperable
  - Reusable
- CARE
  - Fine-grain access rights
  - Legal Interoperability management
- Independent spaces
- API Discovery/Document Browsing
- Native Identity Stitching
- Loose explicit UBL and Information Model Coupling
- Materialized queries/views
- Overloaded class/namespace pairs

The **AIC** provides controlled access to the 'full graph' of semantically encoded data, information, and knowledge, allowing users authorized by associated Access Rights policy to define AICs that implement Patterns using one or more of any preexisting, well-defined ontologies available on the semantic web (e.g. SSN, SOSO, Schema.org, SensorML, Dublin Core, GCMD, etc.), and/or make use of newly minted, user defined ontologies that are constructed and published automatically on AIC Pattern deployment for system-wide discovery and reuse.



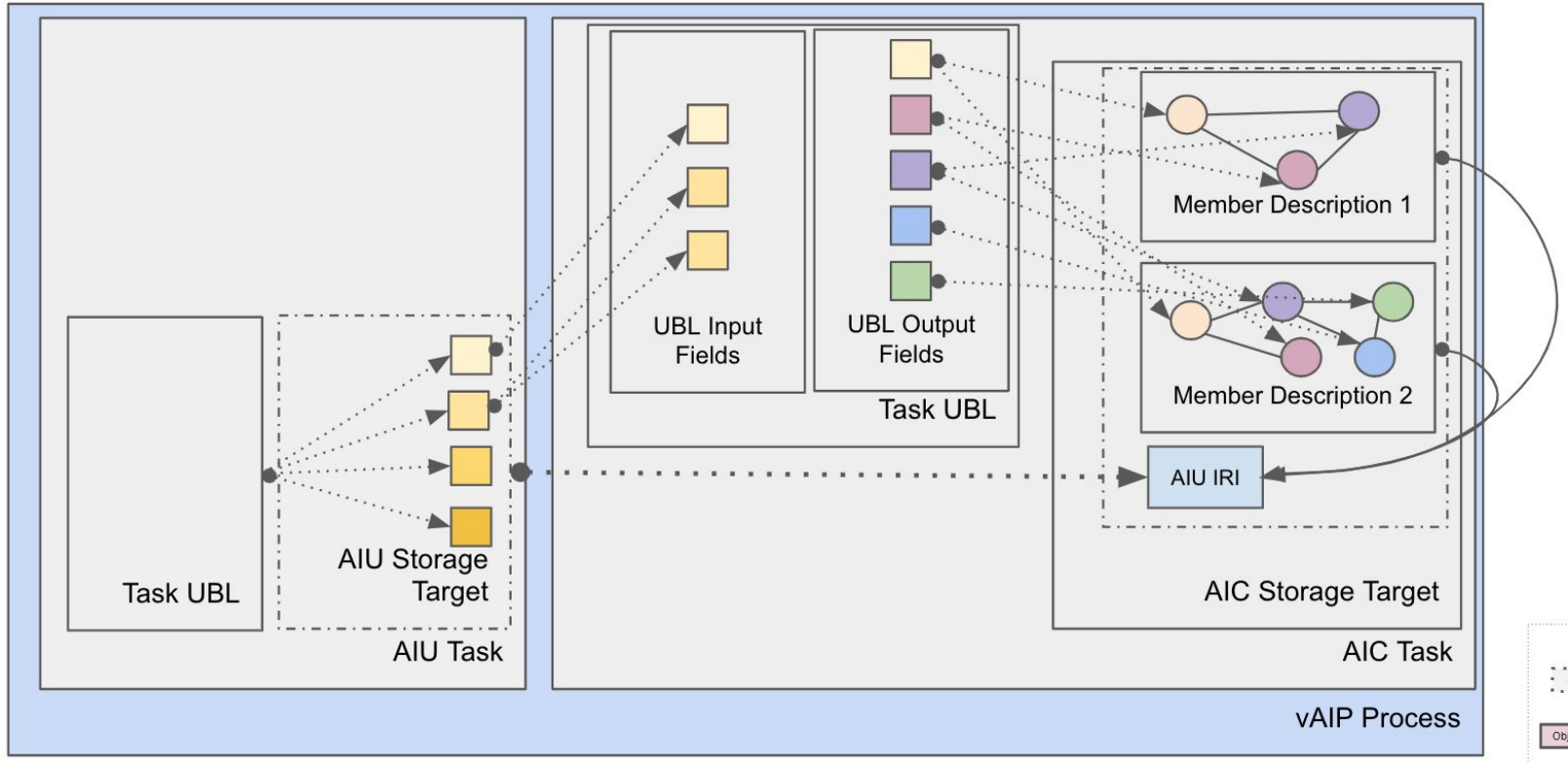


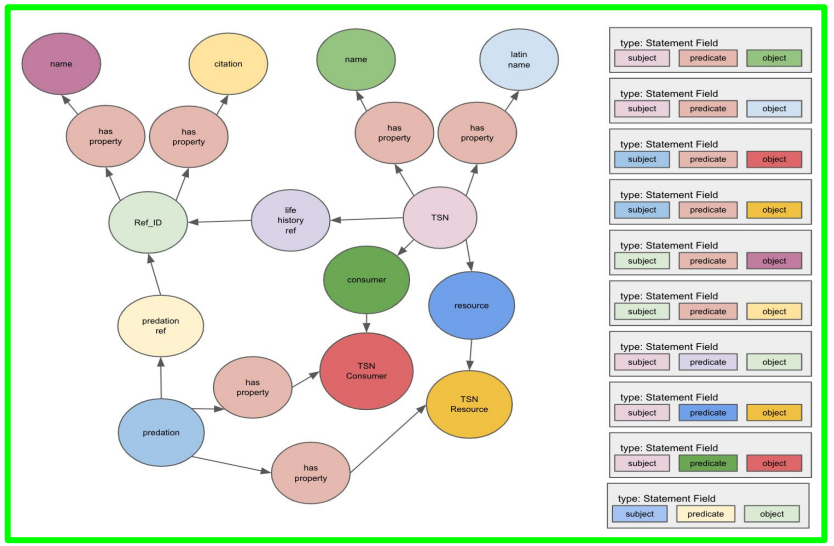
How Data Flows through through an OISS Process

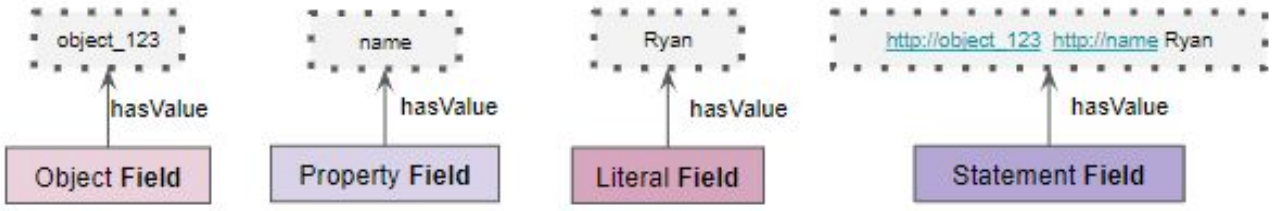


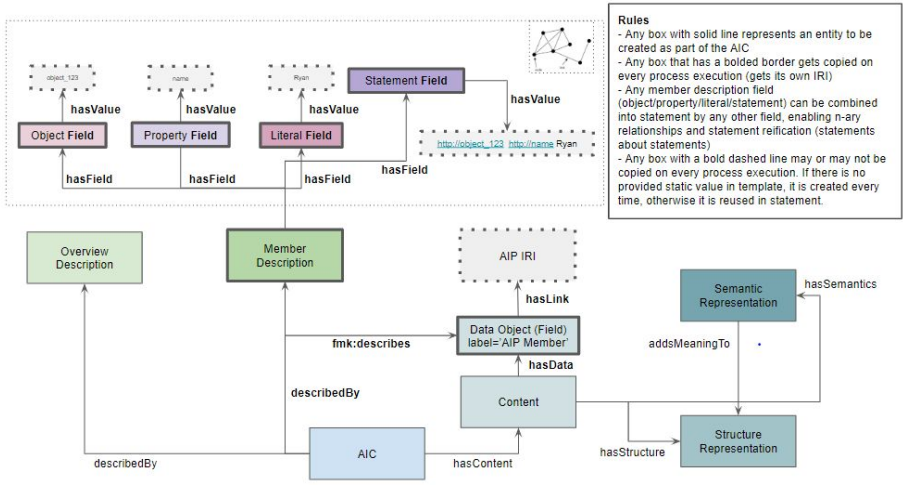
Enormous Amounts of Trusted, High Quality, Fully Contextual, and Regularized Data

Links; texts; documents; input/output parameters for functions; video; audio; biological; ecological; etc.

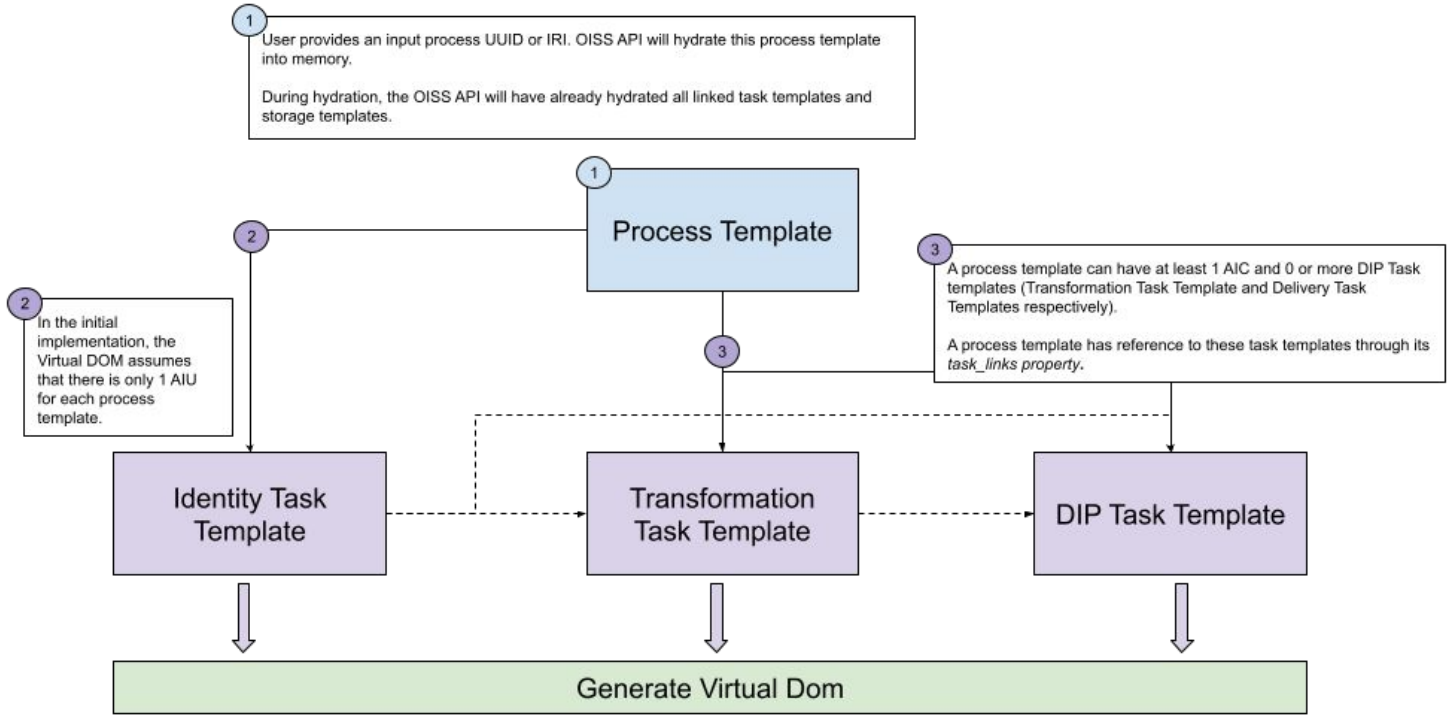




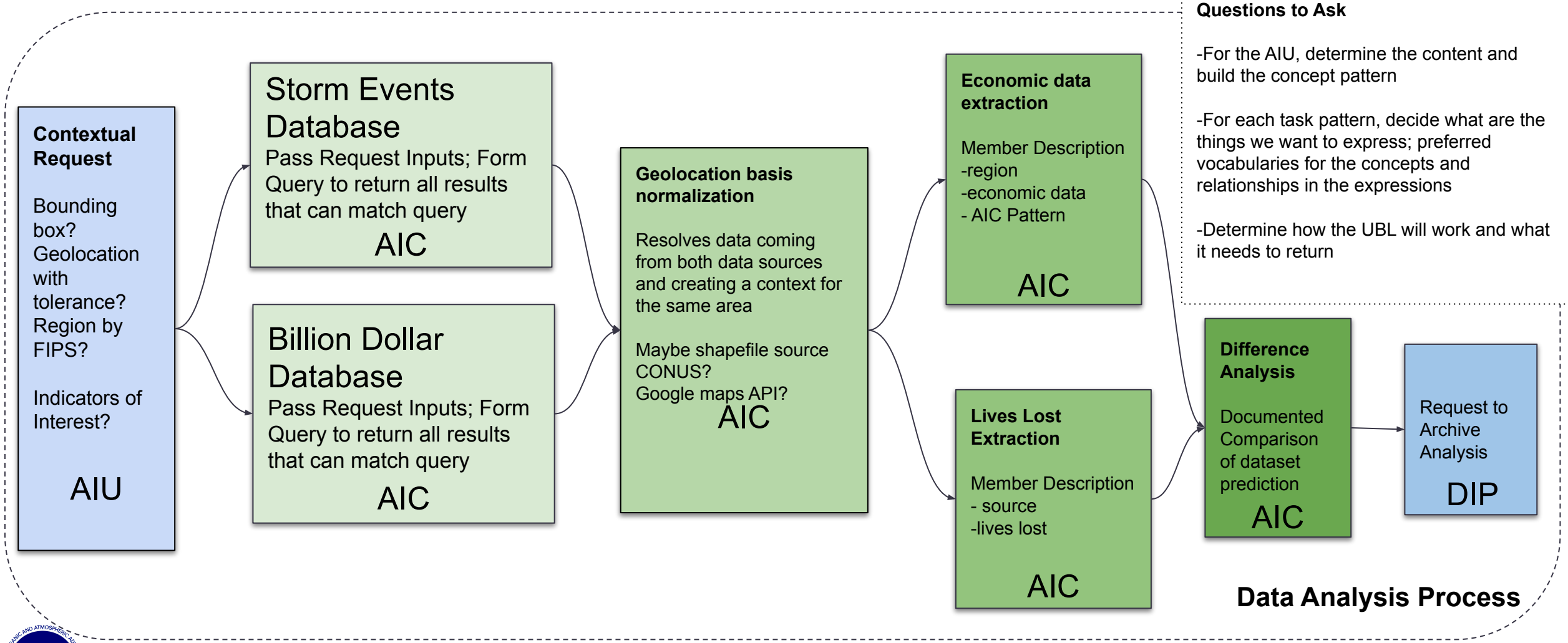




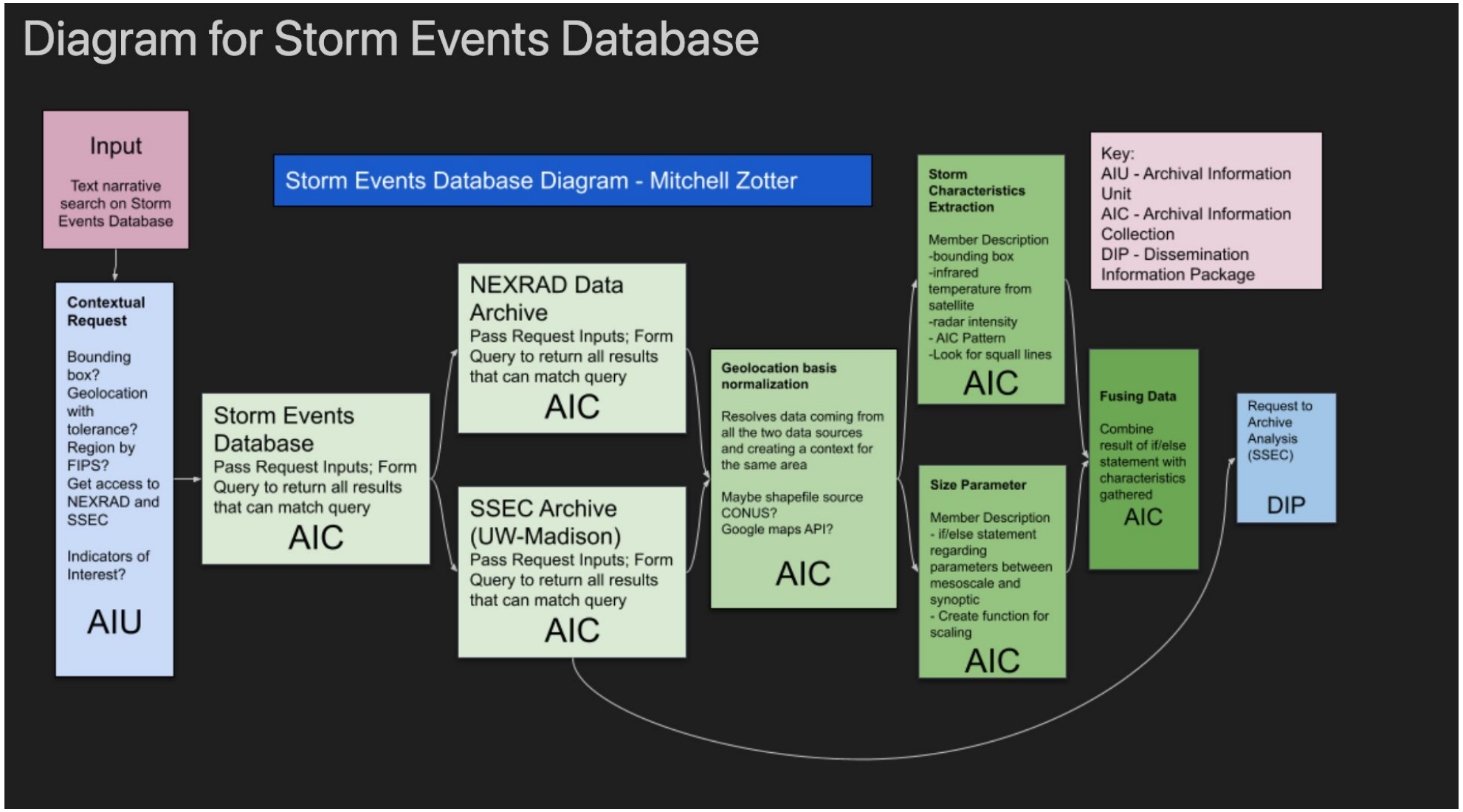
Near-Real-Time, Object Oriented, REPL-Style API-Driven Access to Fully Contextualized, User Defined, and Semantically Standardized Records



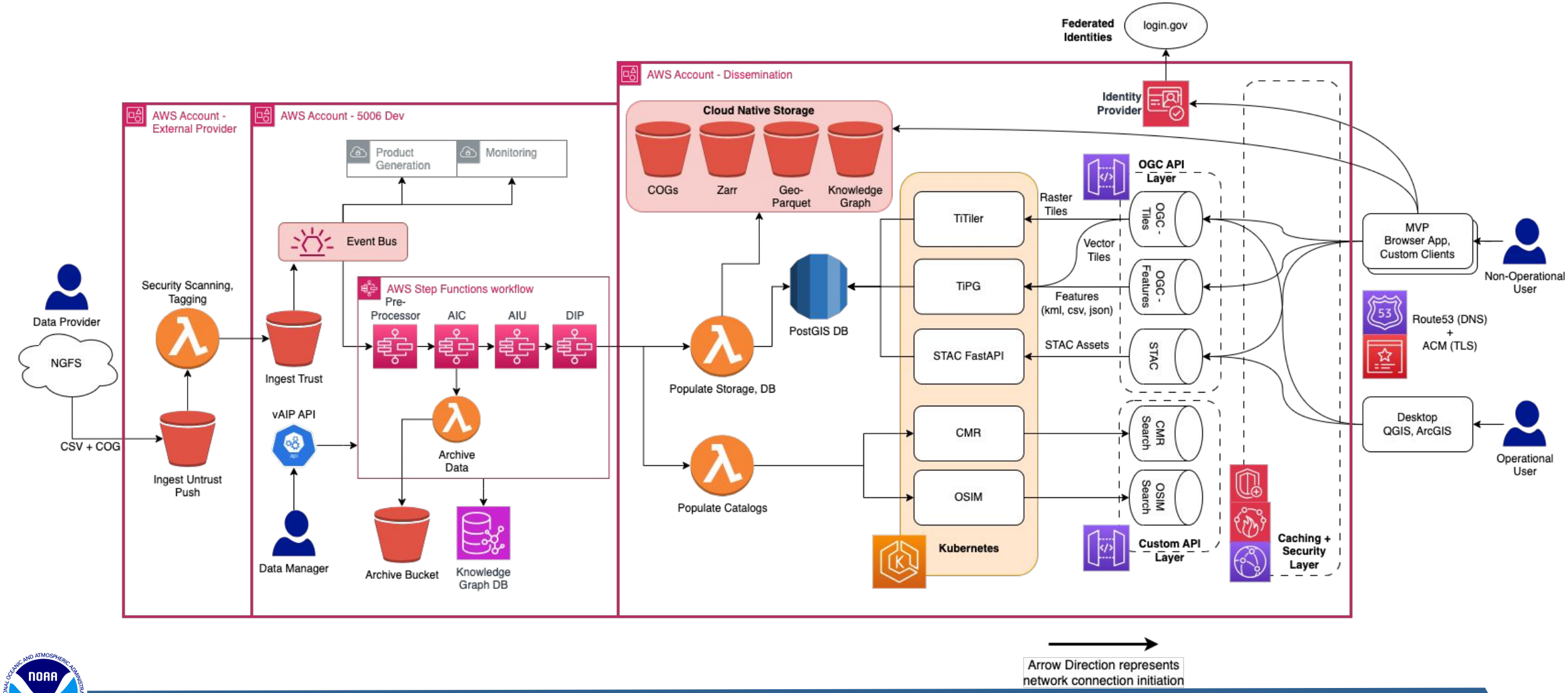
OAR-NCEI Societal Data Partnership



Storm Events Augmentation Project



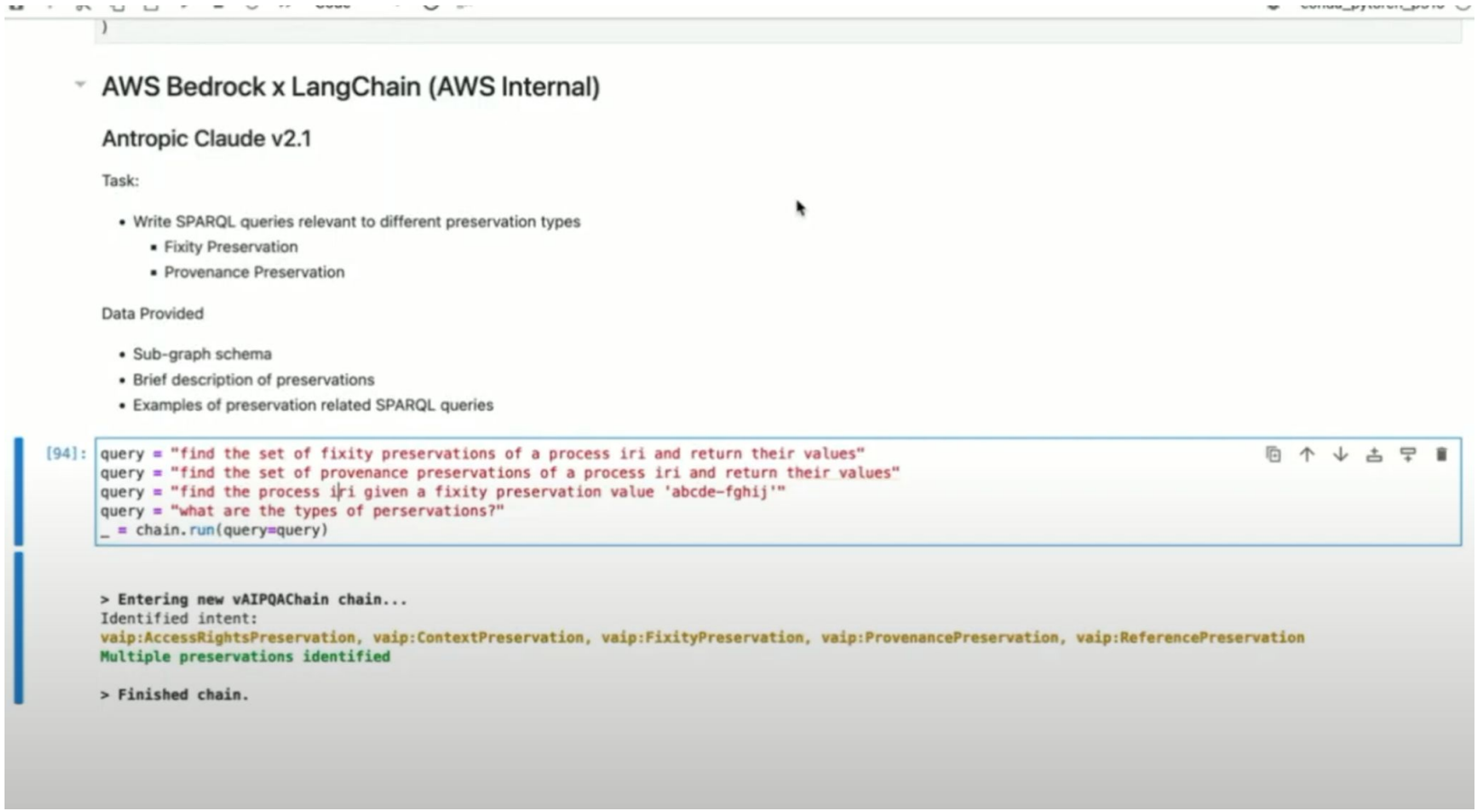
Multi-transform and Multi-access for Wildland Fire Storefront



What Does this Mean to the Enterprise?

- Working toward a new sustained model
- "Simple Made Easy" - Rich Hickey (2011)
  - The model is **simple (decomplected) - gives us maximum concurrency, sustainability**
  - Delivery of the model attempts to make it **easy - tools that are 'near to' users**
- Full provenance of the computational enterprise
  - Data flows, agents, etc
  - Self documenting(!) processes
    - Tracking evolution
    - We can ask the machine to create us new processes
- Shared reference model
  - No more independent stovepipes
  - Iterative improvement (leveling framework)
  - self-sustaining model of shared resources
    - optimizing to use cases, sharing general libraries, etc
  - **Enterprise Approach to Process; Automated, Standardized Access Rights Interface Capabilities**
  - **Agile Reference Architecture Enables best of both worlds - cost optimization and flexibility in deployment**
- Enabling advanced tools in the next-gen toolbox
  - Generative AI, Spiking neural networks
  - **Full provenance and automated access controls provide foundation for AI Safety!!**
- Enabling new users; user driven discovery; federated interoperability
  - Fully concurrent system provides AWS-bounded scale
- Enabling iterative standards alignment inter/intra domain
  - OGC APIs, ARCO references and Indexes in UBL, Template
  - Domain Ontologies that overload, blend, iterate, and include

Exploring Advanced Operational Capabilities



Enterprise Adoption via Education Ecosystem



Federating the Coastal Zone Digital Twin to the OKN

