Accountability for AI Enabled Systems used in Critical Decision-making

Andy Lacher, NASA Langley Research Center

September 2024

AI4SE & SE4AI
RESEARCH AND APPLICATION WORKSHOP

SEPTEMBER 17-18, 2024 | Arlington, VA

Premise

For concepts that leverage increasingly autonomous systems (including those that are AI-Enabled) to be viable, there needs to be clarity regarding accountability for decisions and outcomes

- **Challenges associated with clearly defining roles and responsibilities for increasingly automated processes**
  - **Responsibility, authority, and accountability**
- **Stakeholders concerned with accountability**

Authority, Responsibility, and Accountability

✅ ACCOUNTABILITY

The obligation to answer for an action taken by a responsible entity

✅ AUTHORITY

The power to give orders and/or make decisions.

✅ RESPONSIBILITY [1]

The obligation or duty to carry forward an assigned task to its successful conclusion; this is closely coupled to the authority.

- **Today: automation has <u>limited</u> authority and responsibility**
  - **Revokable by the human**
- **Future: <u>Significantly greater</u> responsibility and authority to make decisions and act, <u>delegated</u> by humans**

Accountability can only reside with a human or human-based organizational entity

1: AC 377, Technical Report #3, "Regulatory Barriers to Autonomy in Aviation", 2022 ASTM, https://www.astm.org/tr3-eb.html

7 Principles for Trustworthy AI

Emphasis on Requirements
- Mechanisms in place to ensure accountability for AI systems and their **outcomes**

https://www.aepd.es/sites/default/files/2019-12/ai-ethics-guidelines.pdf

Accountability for AI System in Aviation

- "AI may have a degree of control authority over specific flight functions but is not accountable for anything"
- Talks about "Aberrant Behavior" and the need to account

Roadmap for Artificial Intelligence Safety Assurance

25 July 2024
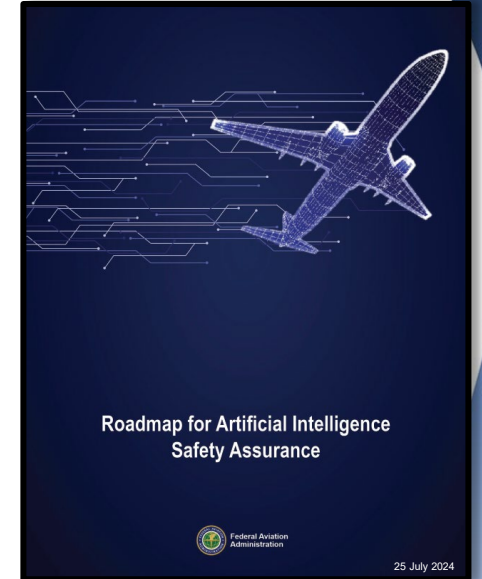
https://www.faa.gov/aircraft/air_cert/step/roadmap_for_AI_safety_assurance

Shifting Decision-Making

Human-centric decision-making → Accountability Relatively Straight Forward
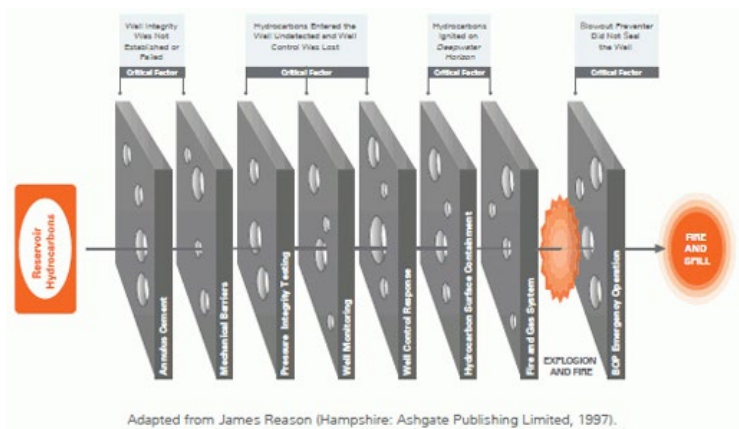
Shift to automated decision-making → Clarity on Accountability Required

Designer?
Maintainer?
Operator?
Data Provider?

Considerations

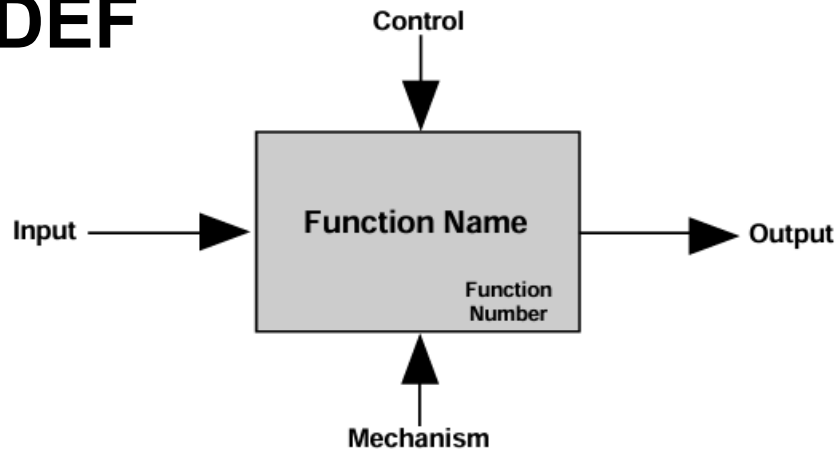- ## Root Cause Analysis / Causal Analysis

Deepwater Horizon Accident Investigation Report, September 8, 2010
https://www.sec.gov/Archives/edgar/data/313807/000119312510216268/dex993.htm

Sequencing "Chain"

- ## IDEF
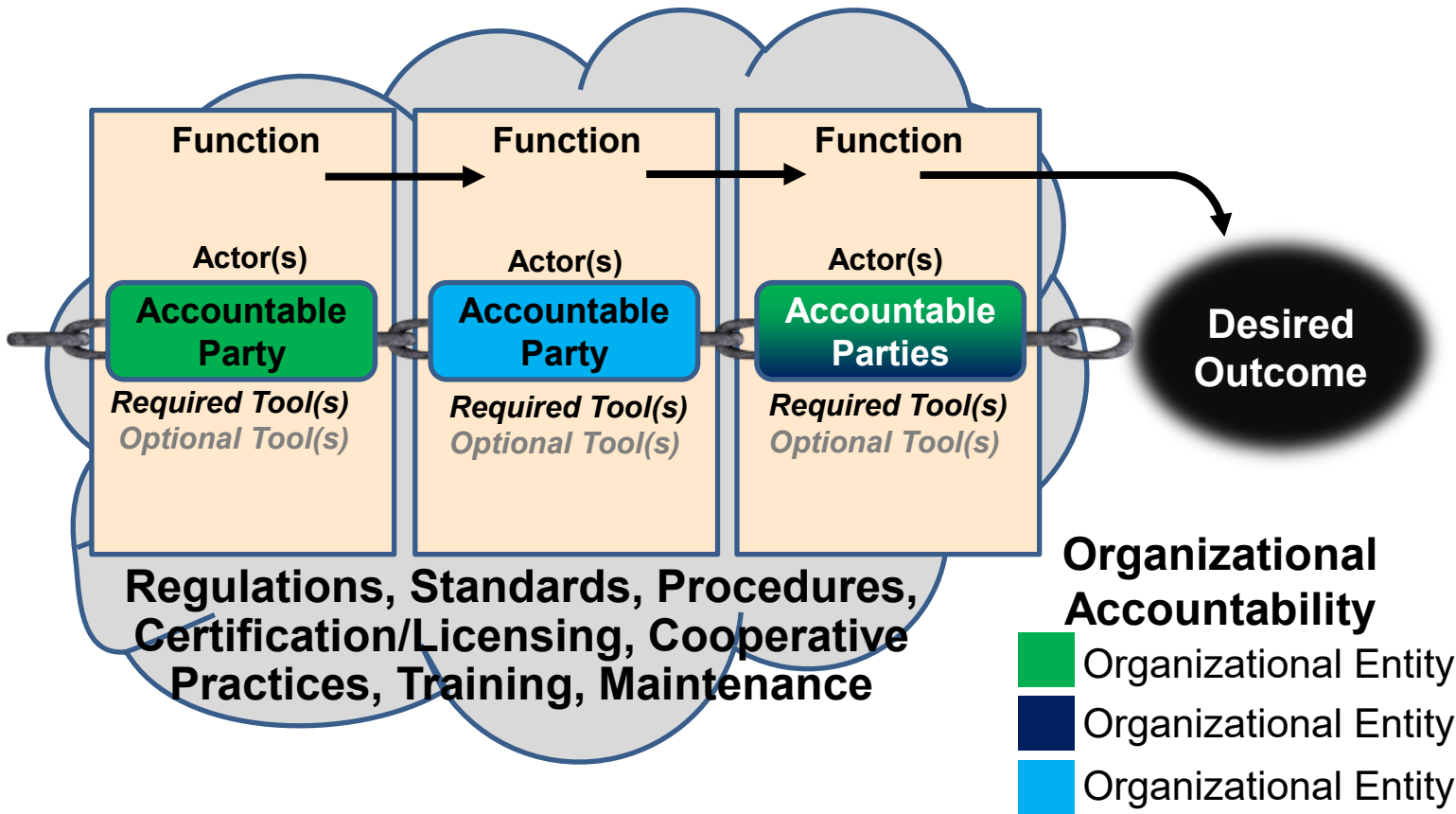

Defense Acquisition University Press, "Systems Engineering Fundamentals",
Defense Acquisition University Press, Fort Belvoir, Virginia, January 2001.
https://acqnotes.com/wp-content/uploads/2017/07/DAU-Systems-Engineering-Fundamentals.pdf

Building Block Format

Function

Actor(s)

Accountable Party

Required Tool(s)
Optional Tool(s)

Our New Model – Accountability Chain

Function → Function → Function

Actor(s) — Accountable Party (green)
Actor(s) — Accountable Party (blue)
Actor(s) — Accountable Parties

Required Tool(s)
Optional Tool(s)

Desired Outcome

Regulations, Standards, Procedures, Certification/Licensing, Cooperative Practices, Training, Maintenance

Organizational Accountability

- Organizational Entity (green)
- Organizational Entity (dark blue)
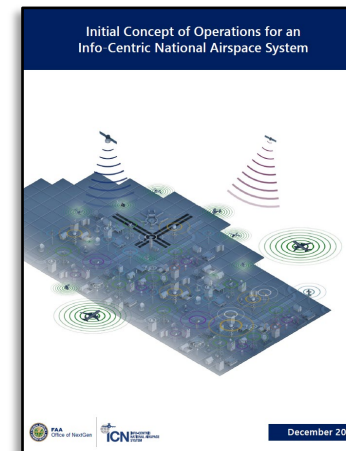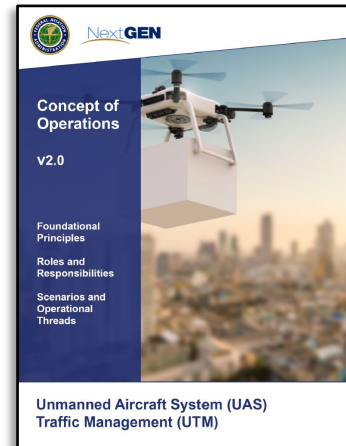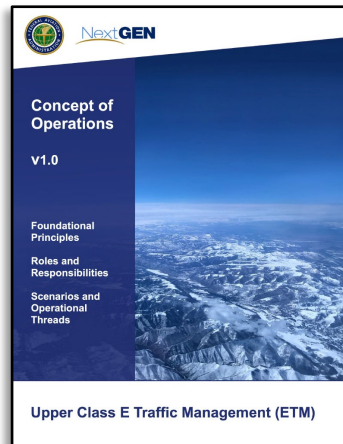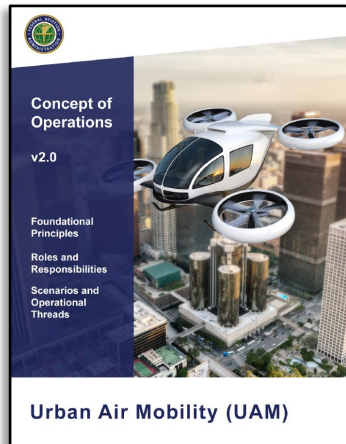- Organizational Entity (light blue)

- **Desired Outcome:** The specific result the functional process is attempting to realize.
- **Functions:** Specific activities that need to be performed to contribute to the desired outcome.
- **Actors:** The individuals or <u>systems</u> that have the authority (power) and responsibility (obligation) to perform a function.
- **Tools:** Capabilities, devices, and systems leveraged by actors to perform the function.
- **Accountable Parties:** While often actors are also the accountable entities, in some cases an <u>organizational entity</u> that manages a system may be the accountable entity.

Conceptual Exploration into Shifts in Conflict Management

eXtensible Traffic Management (XTM)

Urban Air Mobility (UAM)

Upper Class E Traffic Management (ETM)

Unmanned Aircraft System (UAS) Traffic Management (UTM)

Initial Concept of Operations for an Info-Centric National Airspace System

CHARTING AVIATION'S FUTURE: OPERATIONS IN AN INFO-CENTRIC NATIONAL AIRSPACE SYSTEM

. . .

Automated decision-making
Shared information
Operator responsibility for conflict management
Cooperative operating practices

NASA is Exploring a New Digitally-Enabled Cooperative Operating Mode

Visual Flight

Pilot

Flight by Reference to Visual Cues

- Pilot is the separator (see and avoid)
- Limited to no intent sharing
- No conflict management technology required
- Cooperative through human interpretation & judgment
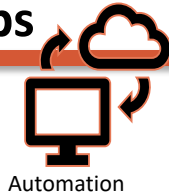- Limited to no dependence on traffic management

Human Centric

Digitally-Enabled Cooperative Ops
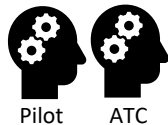
Decision-Maker

Operator   Automation

Operating Environment

Flight by Reference to Digital Capabilities

- Operator is the separator
- Required intent sharing
- Technology-dependent separation function
- Encoded cooperative practices
- Collaborative with traffic management

*Automation Intensive*

Instrument Flight

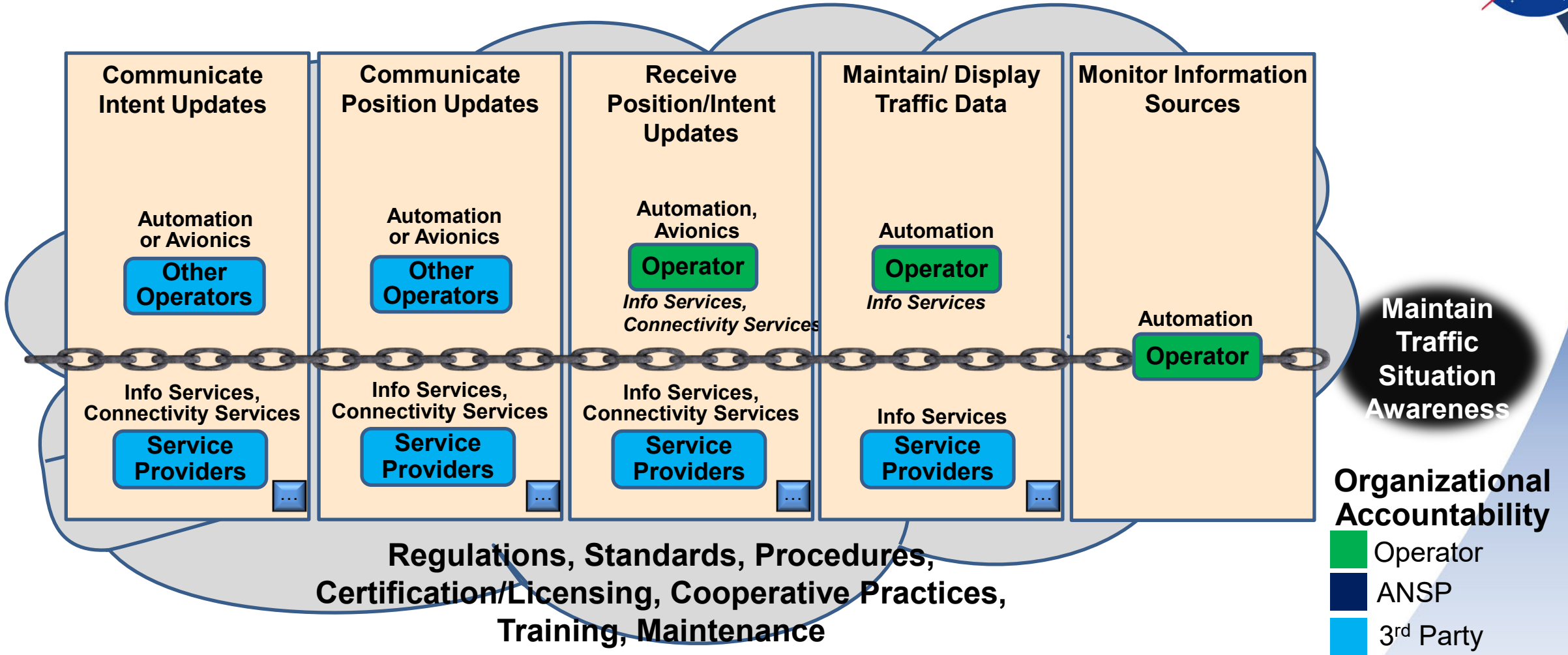Pilot   ATC

Flight by Reference to Instruments and ATC Services

- Air Traffic Control (ATC) is the separator
- Intent sharing with ATC, not with other IFRs
- Technology dependence (communications, navigation, surveillance)
- Cooperation through centralized control
- Integrated with traffic management

Human Centric

High-level Accountability Analysis for Separation Provision Capability
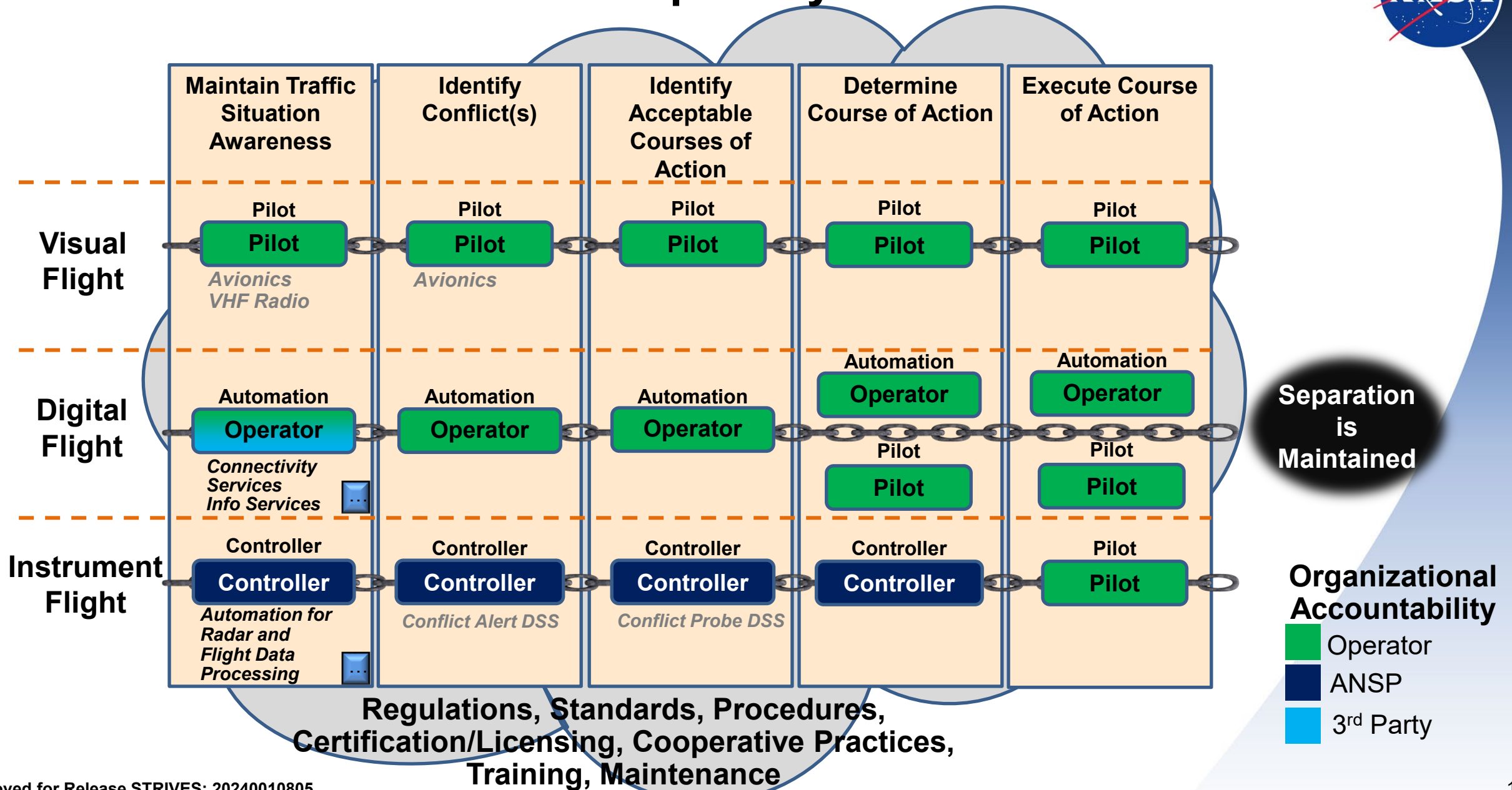
High-level Accountability Analysis for Maintaining Traffic Situation Awareness under Digital Operations

Communicate Intent Updates

Automation or Avionics

Other Operators

Info Services, Connectivity Services

Service Providers

Communicate Position Updates

Automation or Avionics

Other Operators

Info Services, Connectivity Services

Service Providers

Receive Position/Intent Updates

Automation, Avionics

Operator

Info Services, Connectivity Services

Info Services, Connectivity Services

Service Providers

Maintain/ Display Traffic Data

Automation

Operator

Info Services

Info Services

Service Providers

Monitor Information Sources

Automation

Operator

Maintain Traffic Situation Awareness

Regulations, Standards, Procedures, Certification/Licensing, Cooperative Practices, Training, Maintenance

Organizational Accountability

- Operator
- ANSP
- 3rd Party

Moving Forward

- **Is clarity associated with accountability a critical barrier to the viability of the use of highly automated system (including AI-enabled systems) in critical operations?**

- **Can the barrier be overcome?**

> The *Accountability Chain* methodology could be a means to analyze accountability for increasingly automated operating modes in which multiple parties have interdependent roles

Thank You

Andrew.R.Lacher@nasa.gov
David.Wing@nasa.gov

Backups

Detailed Framework for a Digitally Enabled Operating Mode

Info Services & Connectivity

Maintains a digital model of the operating environment for use by decision-making automation systems

Integrated Information Environment

Cooperative Practices

Governs the behavior of digital operations to ensure harmonized use of the airspace

Essential Elements

- Information Services & Connectivity
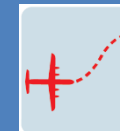- Shared Traffic & Intent Awareness
- Cooperative Practices
- Automated Conflict Management Capabilities

Shared Traffic & Intent Awareness

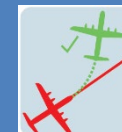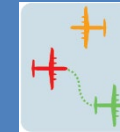Sharing of relevant traffic & intent information for use by automation in conflict management

Automated Capabilities

Automates the critical functions in tactical separation and strategic conflict management

Design Attributes

Safety
- Safety Benefits
- Embedded Risk Mitigations

Integration
- All Airspace Classes
- Non-Disruptive to Incumbents

Performance
- Tailorable Separation
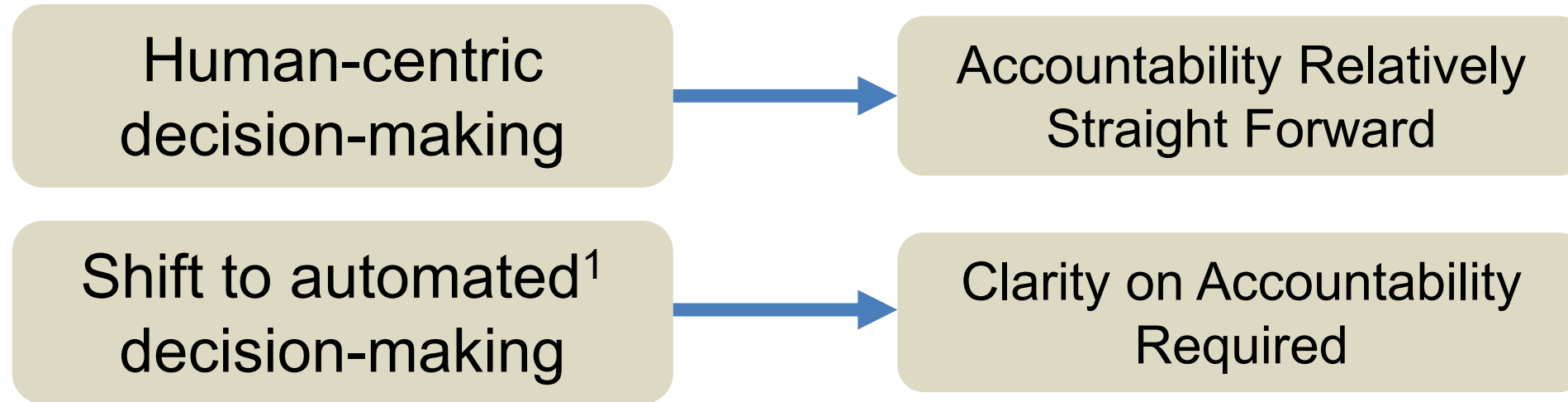- Aircraft Diversity

Value
- Community
- Individuals

ICAO Layers of Conflict Management

ICAO Conflict Management[1]	Visual Flight	Digital Flight	Instrument Flight (ATM Services)
Strategic Conflict Management *"Achieved through the airspace organization and management, demand and capacity balancing and traffic synchronization components; Aim to reduce the need to apply the second layer — separation provision."*	**Procedures, Practices, and Airspace Structure** • Traffic Pattern • Ordinal Altitudes • Routes, Flyways	**Traffic Flow Management & Airspace Structure** • Cooperative compliance (PC-4) • Self organization (PC-5) • Self limiting (PC-6)	**Traffic Flow Management & Airspace Structure** • Executed via ATC
Separation Provision *"Tactical process of keeping aircraft away from hazards by at least the appropriate separation minima; Only used when strategic conflict management cannot be used efficiently."*	**Remain Well Clear** Airspace user is the separator for its activity in respect of one or more hazards Visual separation, assisted by technology • CDTI • FLARM	**Self-Separation** Airspace user is the separator adhering to a defined minima in respect of one or more hazards • Operator self-separation (PC-1) • Cooperative conflict management (PC-2) • Adaptive pair-wise separation (PC-3)	**Separation Services** • Radar separation • Procedural separation
Collision Avoidance *"Must activate when the separation mode has been compromised; Not part of separation provision; Must be compatible with separation provision mode."*	**"See" and Avoid** • Pilot vision • CDTI – Traffic Alerting • TCAS (in some cases)	**"See" and Avoid** • Pilot vision (optional) • TCAS/DAA (in all cases)	**"See" and Avoid** • Pilot vision • TCAS/DAA (in most cases)

1: International Civil Aviation Organization, ICAO Doc 9854, *Global Air Traffic Management Operational Concept*, First Edition-2005.

Shifting Decision-Making – Digital Operations

Human-centric decision-making	→	Accountability Relatively Straight Forward
Shift to automated[1] decision-making	→	Clarity on Accountability Required

- **Dependence Upon Automation**

- **Expansion of Operator Role in Conflict Management**

- **Conflict Management with Instrument Flights**

- **Dependence Upon Information and Connectivity Services**

- **Dependence Upon Operational Intent**

1: Digitally enabled cooperative operations is highly dependent upon automated decision-making, but the issue extends to automated/autonomous flight in general

What is an Operating Mode?

Regulatory, procedural, and technical means for aircraft to *operate safely* within the airspace

aka "*flight rules*" as defined in regulations, policies, procedures, training materials, …

Routine Operating Modes	Specialized Operating Modes
Widespread use in most airspace classes *Routine* for pilots & controllers	*Limited* to certain operations and/or airspace incompatible with VFR & IFR *Routine* for pilots who use them; *Nonroutine* for controllers and other pilots
• Visual Flight Rules (VFR) • Instrument Flight Rules (IFR)	• State/Military Aircraft • Moored Balloons / Kites / Amateur Rockets / Unmanned Free Balloons • Ultralight Vehicles • Parachute Operations • Small Unmanned Aircraft Systems