

How we're transforming the practice of learning from incidents in a 12,000-person organization

DevOps Enterprise Summit

David Leigh

IBM CIO | Resilience Engineering





David Leigh

Distinguished Engineer, Resilience Engineering

Location: Chapel Hill, NC

IBM Tenure: 26 years

Key Skills:

- DevOps
- Coping with Complexity
- Resilience Engineering
- Learning from Incidents
- Systems Thinking

About me

- 26 years at IBM
- Many projects driving adoption of modern practices and tools
- Started studying Resilience and Learning from Incidents in 2013
- Helped deploy and scale Slack, GitHub, and many other tools to all of IBM
 - Many opportunities to learn from incidents!
- Representing IBM in the SNAFU Catchers Consortium for Resilient Internet-Facing Business IT
- Now focused on engineering resilience for CIO

Two perspectives of system safety

Traditional view:

- Our systems are inherently safe
- Accidents are caused by identifiable failures or malfunctions in specific technical or human components of the system.
- Safety is measured by the absence of failures.
- One of the best ways to improve safety is via enforcement of processes and rules.

New view:

- Our systems are inherently *unsafe* because they are complex and constantly changing.
- It is *people* who are keeping these systems working via continual adaptation to gaps, challenges, and surprises.
- Safety is the presence of *adaptive capacity*.
- One of the best ways to improve safety is to look at how people adapt in the face of incidents and near misses.
- How do they coordinate activities, make decisions, acquire information, assess risks, etc.

Root Cause vs. A Resilience Engineering approach to Learning from Incidents

Root Cause

Based on linear cause & effect models

Focus on fixing: Find the part that failed

Human variability is a problem control

Improve the system by addressing the failed part and constraining adaptability

Goal is to prevent recurrence

Learning from Incidents

Based on Systems Thinking

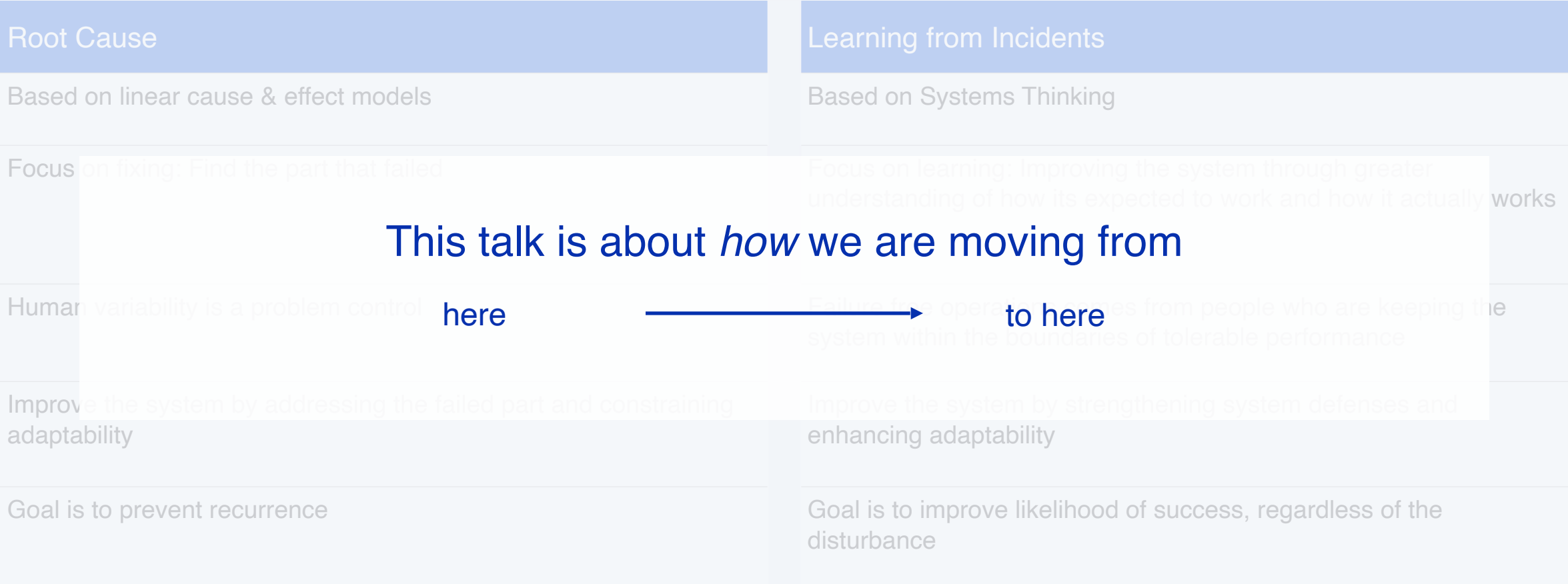
Focus on learning: Improving the system through greater understanding of how its expected to work and how it actually works

Failure free operations comes from people who are keeping the system within the boundaries of tolerable performance

Improve the system by strengthening system defenses and enhancing adaptability

Goal is to improve likelihood of success, regardless of the disturbance

Root Cause vs. A Resilience Engineering approach to Learning from Incidents



Key Ingredients

- Leadership appetite for a new approach
- Credibility with practitioners
- Readiness to seize an opportunity
- Engagement with owners of existing processes



IBM CIO

Enabling a large, complex global enterprise



300,000+

IBMers
worldwide



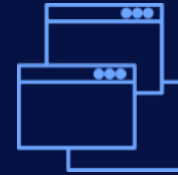
170

countries
where IBM does
business



\$55B
























annual revenue across consulting,
software, and infrastructure
businesses



5,000+

business applications

IBM's CIO Portfolio

	Manufacturing supply chain		Compensation and benefits		389,000 workstations		6 cloud regions
	Software delivery operations		Tax and treasury		167,000 mobile devices		31 data centers
	Consumption-based pricing		Sales incentives		4.7 billion+ mail messages		20,000 servers
	Global contract license management		Building and data center operations		13+ enterprise tools		56 petabytes storage
	Workforce management services		Contract lifecycle management		736,000 IT support tickets		
	Multi-vendor support		Global logistics		82 million searches		10 network points of presence (PoPs)

Previous state of learning from incidents (pre-2022)

- Weekly IT Operations Meeting with CIO Leaders, featuring RCA reviews of “major incidents”
- RCAs focused on technical system, how it failed, and how failure can be prevented
- Generally, if an incident was not characterized as “major”, no organized post-incident learning activity happened

- In 2015, I was part of a team of newcomers to the organization
- We had some experience in studying our incidents differently
- We used Dekker's *Field Guide to Human Error* and the *Etsy Postmortem Facilitation Guide* as inspirations
- We did this completely under the radar of senior leadership
- We slowly gained some reputation among peer teams and were eventually invited to help them study a few incidents



In the summer of 2021, shortly after our new CIO came in, a very big incident occurred.

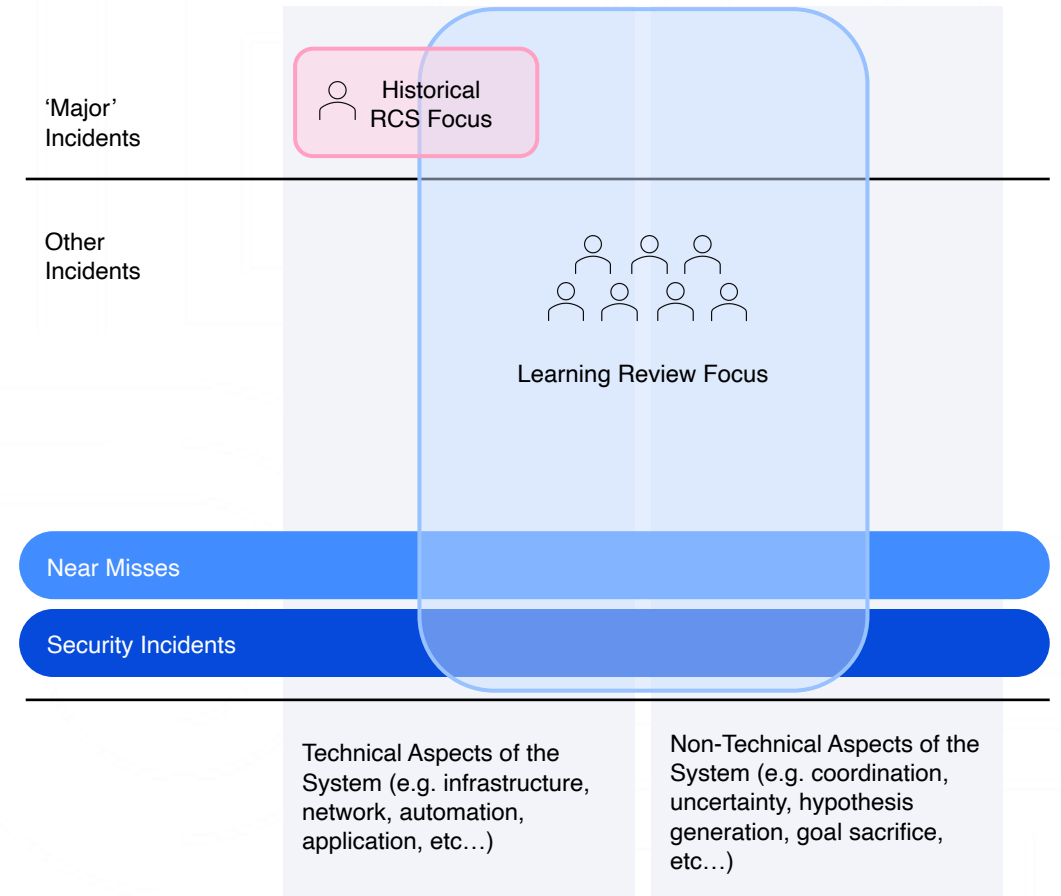
A big incident can provide a motivation to do things differently.

Subsequently, during retrospective discussions about IT Operations meetings and RCAs, I found the opportunity to propose a new approach and this proposal was accepted.

The CIO Learning from Incidents Monthly Meeting

“Incidents are *unplanned investments*.” (John Allspaw)

We want to improve our return on these investments by discussing a small number of incidents in depth with CIO Executives, STSMs, and others.



The “oh shit” moment

- *Can I do the analysis necessary to tell a story about a recent incident that will resonate with a large group of business and technical leaders?*
- *Can I present the case in a way that will invite constructive discussion in the meeting?*
- *Will these meetings be satisfying to those who are looking for something to be labeled “root cause”*
- *Can I do this every month?*



Program Goals



Repair mental models

Recalibrate the mental models of team members to be more up-to-date with the actual state of the system



Identify Hazards

Improve awareness of what vulnerabilities, fragilities, and surprises can show up in our systems.



Recognize Patterns

Improve understanding of the themes and patterns to look out for in designing and managing similar systems.



Understand Challenges

Improve understanding of the difficulties people face as they work to repair systems and how we can make this work less difficult.



Enhance Culture

Sustain and grow our psychological safety, so that individuals talk freely about the mistakes that they make, help others avoid similar mistakes, and help to design systems that reduce the likelihood of mistakes.

CIO LFI Meeting Objectives

By the end of this meeting, we hope that you...



...understand the story of the incident and the themes we can observe.



...recognize the benefits of the improvement opportunities.



...learn something new about the technical and/or social aspects of the systems involved.



...recognize differences between this approach and the RCA approach.



...find the material and the discussion interesting and engaging. Ask questions.



...are looking forward to next month's meeting.

LFI Outcomes



Began in January 2022

85-100

Participants attending each monthly meeting

+270

People attending one or more meetings

+100

Unique views of the meeting recordings

+300

Unique views of the reports



Feedback from this meeting series has been extremely positive

Result

Practice drills for collaboration during Slack outages

Result

Major project to consolidate our IT inventory

Result

Leaders asking for incidents to be included

Result

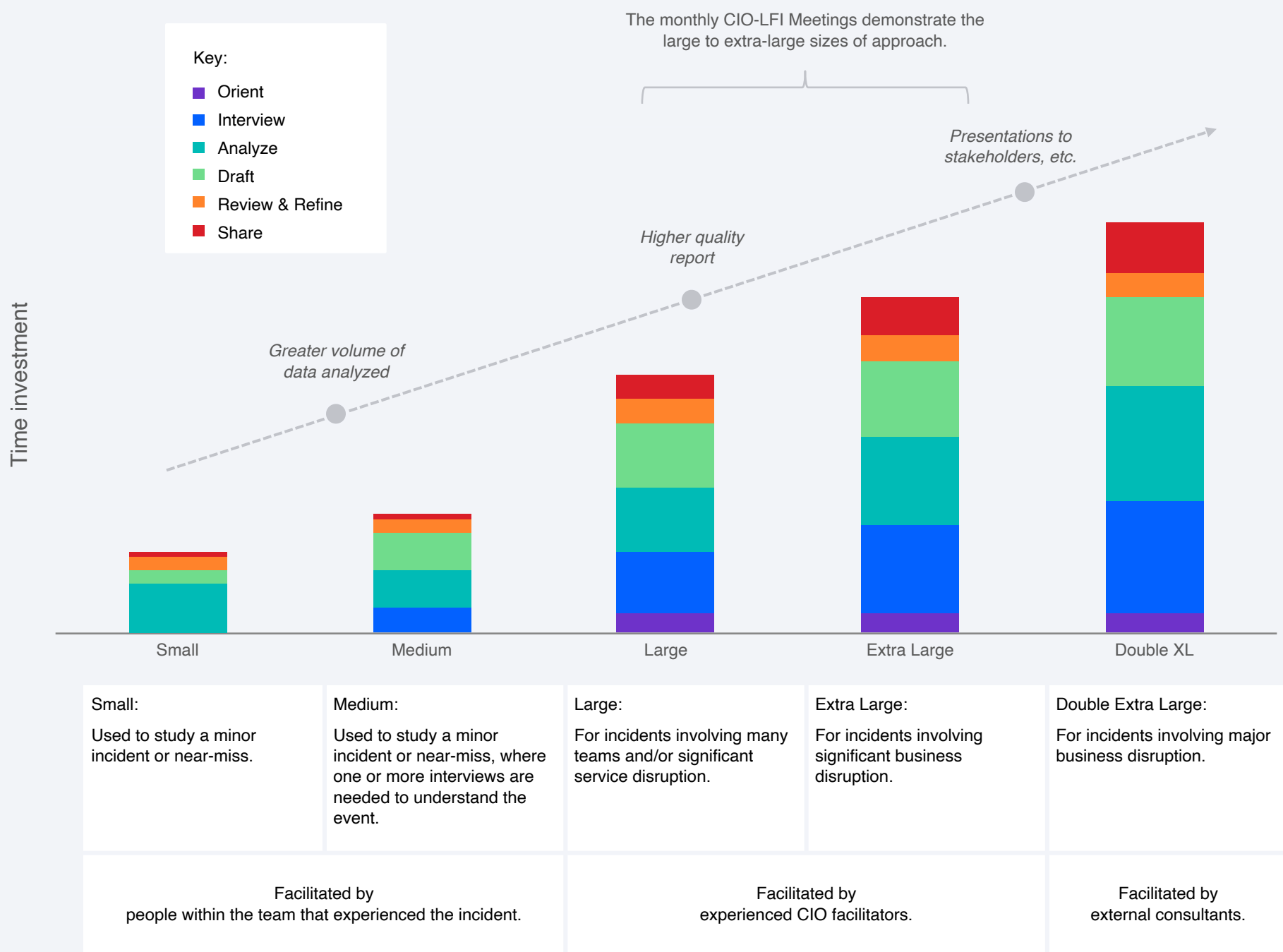
Scaling impact through inspiration

T-shirt Sizes of Learning from Incidents

LFI Process Phases:

- Orient:** If the facilitator is from outside the team, they need a high-level understanding of what happened, who was involved, what data sources are available, etc.
- Interview:** 1-1, or small group discussions to understand individuals' experiences and perspectives.
- Analyze:** Build a timeline, synthesize data, identify patterns, themes, and improvement opportunities
- Draft:** Write an initial report draft
- Review & Refine:** Discuss the draft with those who were involved in the incident. Includes discussion about improvement opportunities which is often done separately.
- Share:** Make the report available, present to stakeholders and others as appropriate.

CIO LFI



Our Personas



Carlos
The Guiding Coach

“I have experience with ITSM and ITIL, and I’m curious about this new approach to learning from incidents.”

- Ambitious
- Unorganized
- Dependable



Sharice
The Protective Leader

“My team is my priority. We recently had a small incident with a service disruption, but we responded and reacted quickly.”

- Protective
- Apprehensive
- Defensive



Parth
The Newbie Facilitator

“I’ve been asked to look into an incident that recently occurred. I’m not sure where to get started.”

- Overwhelmed
- Confused
- Logical



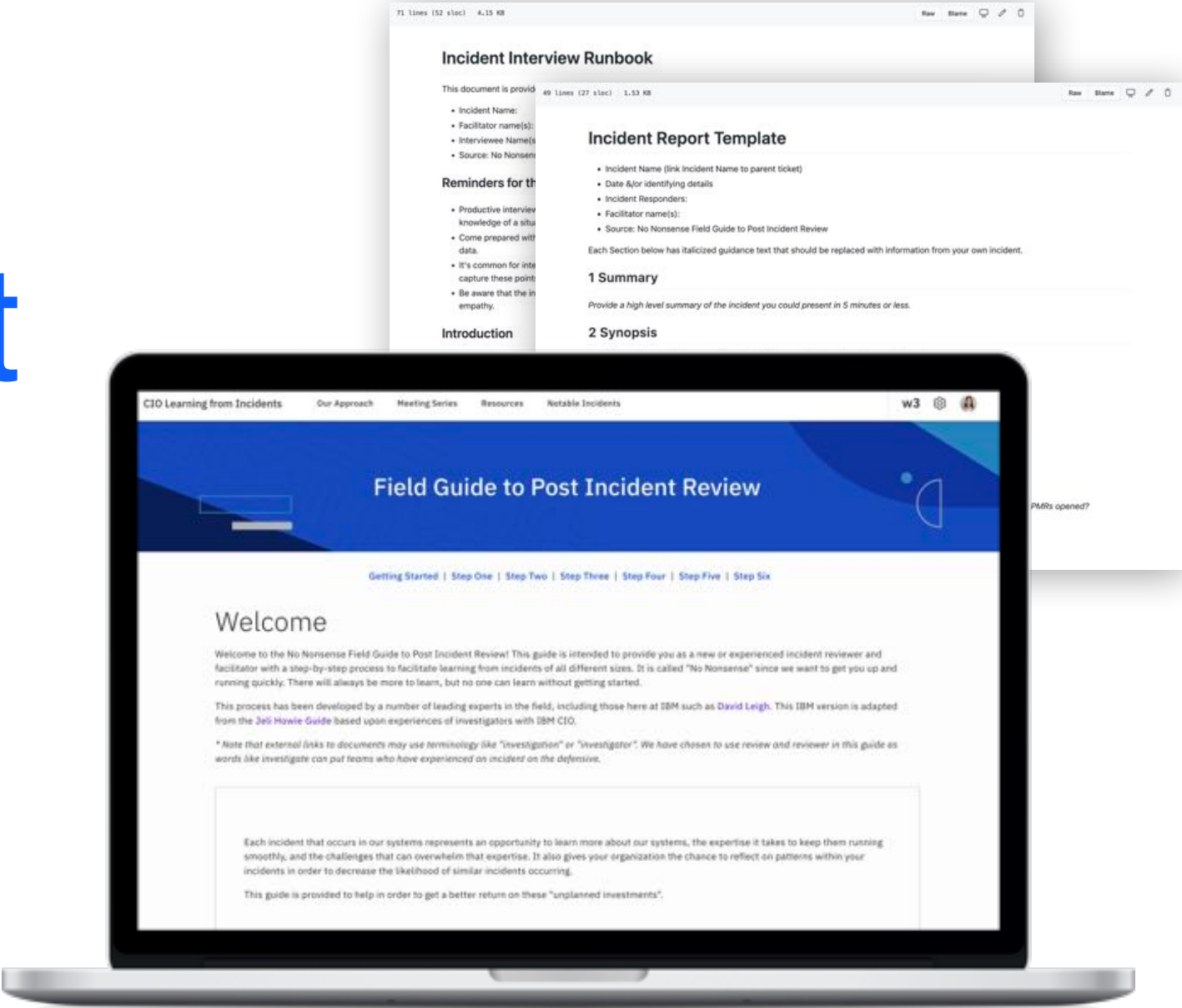
Gina
The Intrigued Learner

“I want to hear interesting events that have occurred and learn how to apply to my team.”

- Curious
- Unchallenged
- Inquisitive

Design Led

Our Enablement Materials



Here are some
problems that
remain:

How to help teams to grow their own
facilitation capacity.

How to make it easy for teams to
share incident reports in a central
library.

How to socialize the contents of the
incident library.

References

How Complex Systems Fail

Richard Cook, M.D.



Dr. Richard Cook

Four Concepts for Resilience Engineering and the Implications for the Future of Resilience Engineering

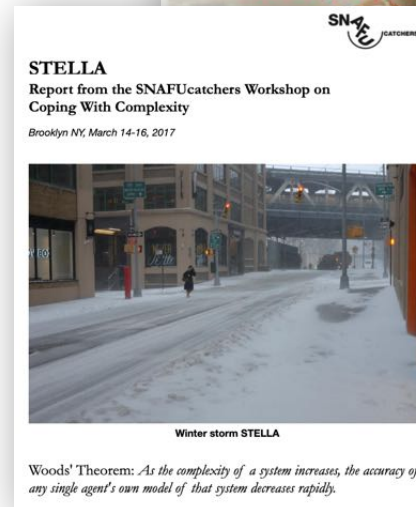
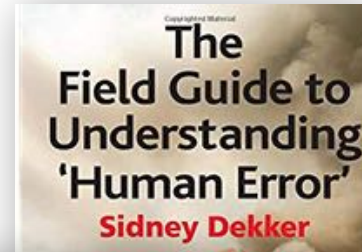
Dr. David Woods

Common Ground and Coordination in Joint Activity

Gary Klein, Paul J. Feltovich, Jeffrey M. Bradshaw, David D. Woods

Managing the Hidden Costs of Coordination

Laura M.D. Maguire



Dr. Laura Maguire



Dr. David Woods



John Allspaw



Dr. Sidney Dekker

Thank you