FloCon 2022

18th Annual Open Forum for Large-Scale Data Analytics

Using Data to Defend

What Do We Mean by a Science of Security?

Dr Scott Mongeau

scott@sark7.com

Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA 15213

FloCon presentations: Cybersecurity Data Science (CSDS)

1. **2019** CSDS Field
   interview insights
2. **2020** CSDS Best Practices
   research results
3. **2021** CSDS Trends
   emerging methods
4. **2022** Security Science
   future directions

Cybersecurity Data Science

I. Foundations

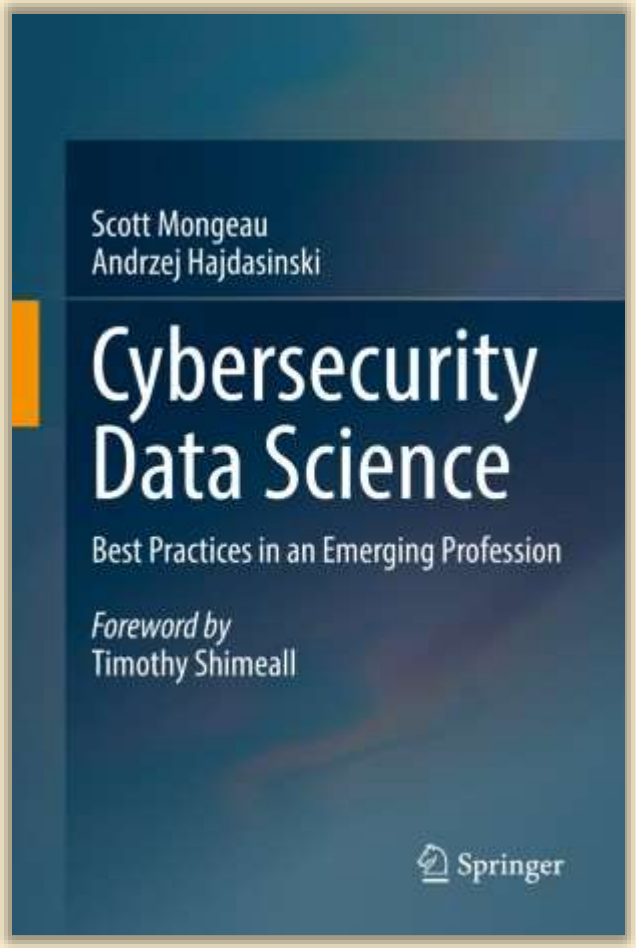
- What is CSDS?
- Status as a profession?

II. Interviews

50 CSDS practitioners

III. Best practices

- Data management
- Scientific methods
- Cross-domain collaboration



Cybersecurity Data Science: Springer 2021

Cybersecurity ✓

Data ✓

Science ?

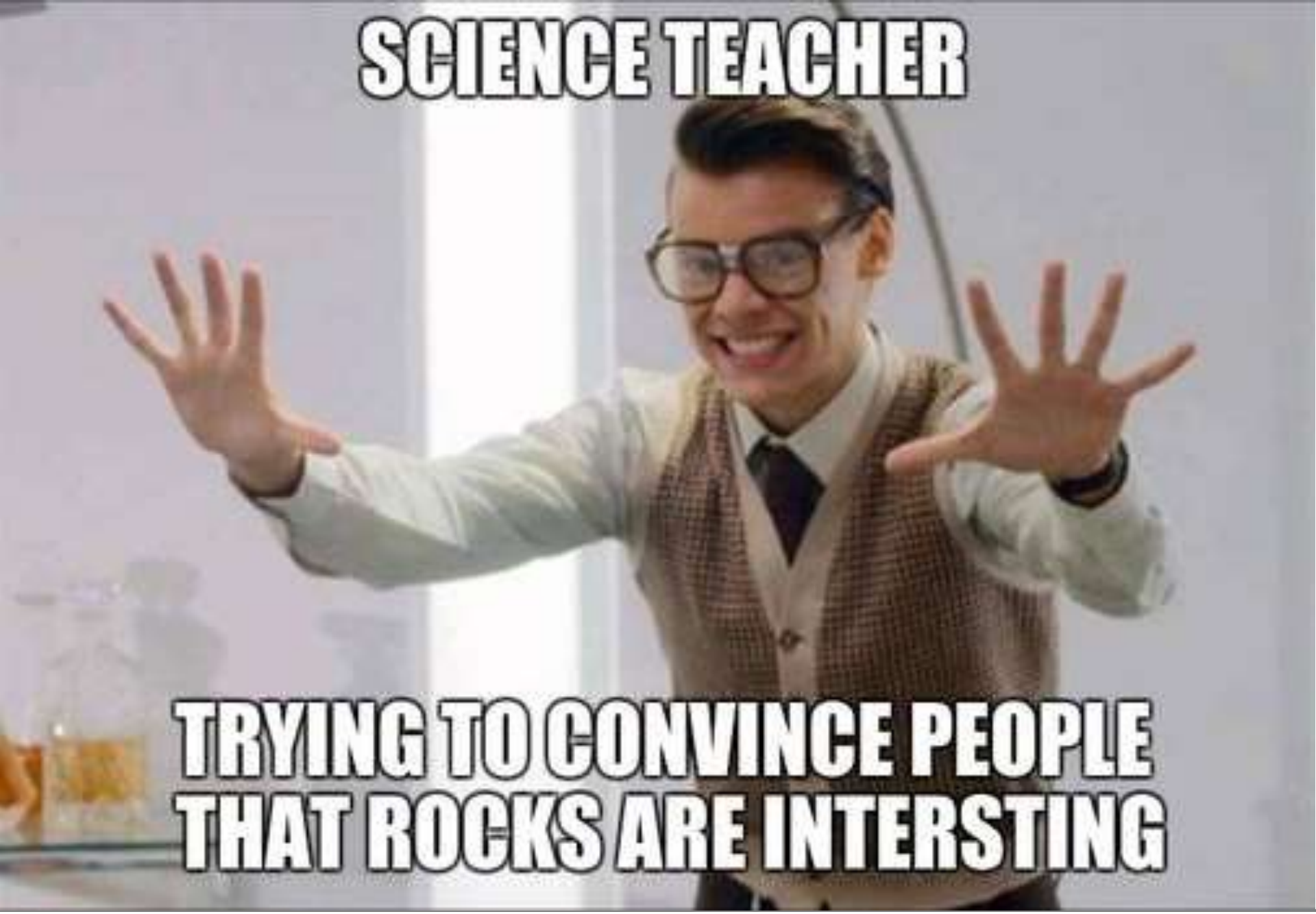
Data Science *offers both solutions and challenges*

Engineering ✓

Analysis ✓

Theory ?

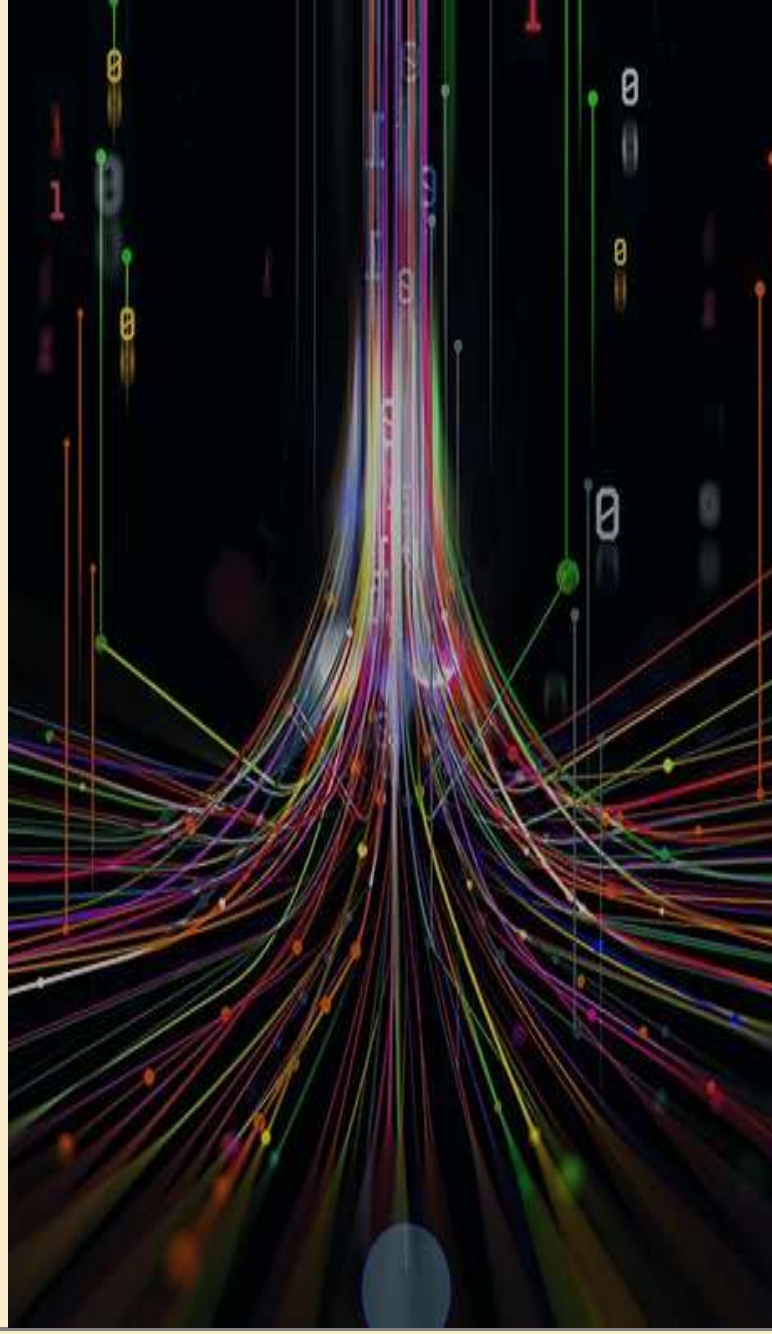
Data Science *offers both solutions and challenges*



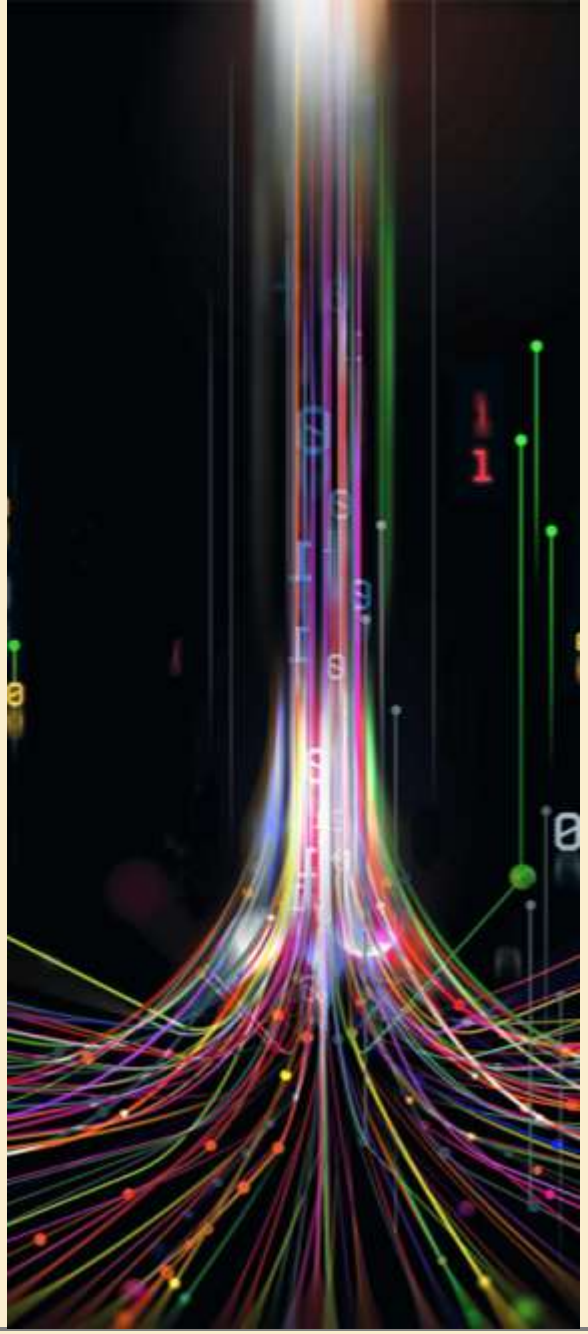
Sciency… things

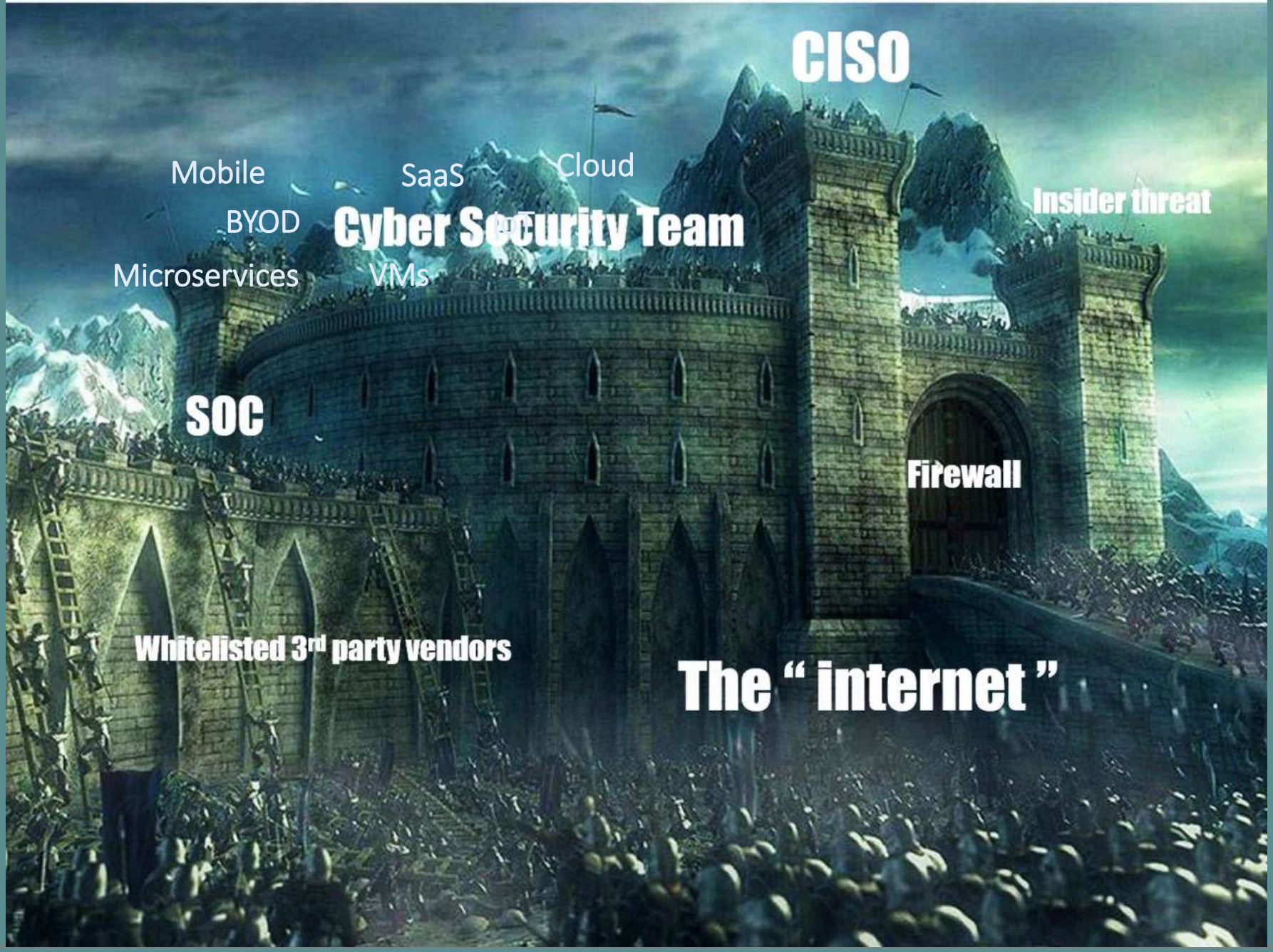
Agenda

1. Science of Security?
2. Engineering versus Science
3. Science as a Process
4. Operational Security Science
5. Where do we go from here?

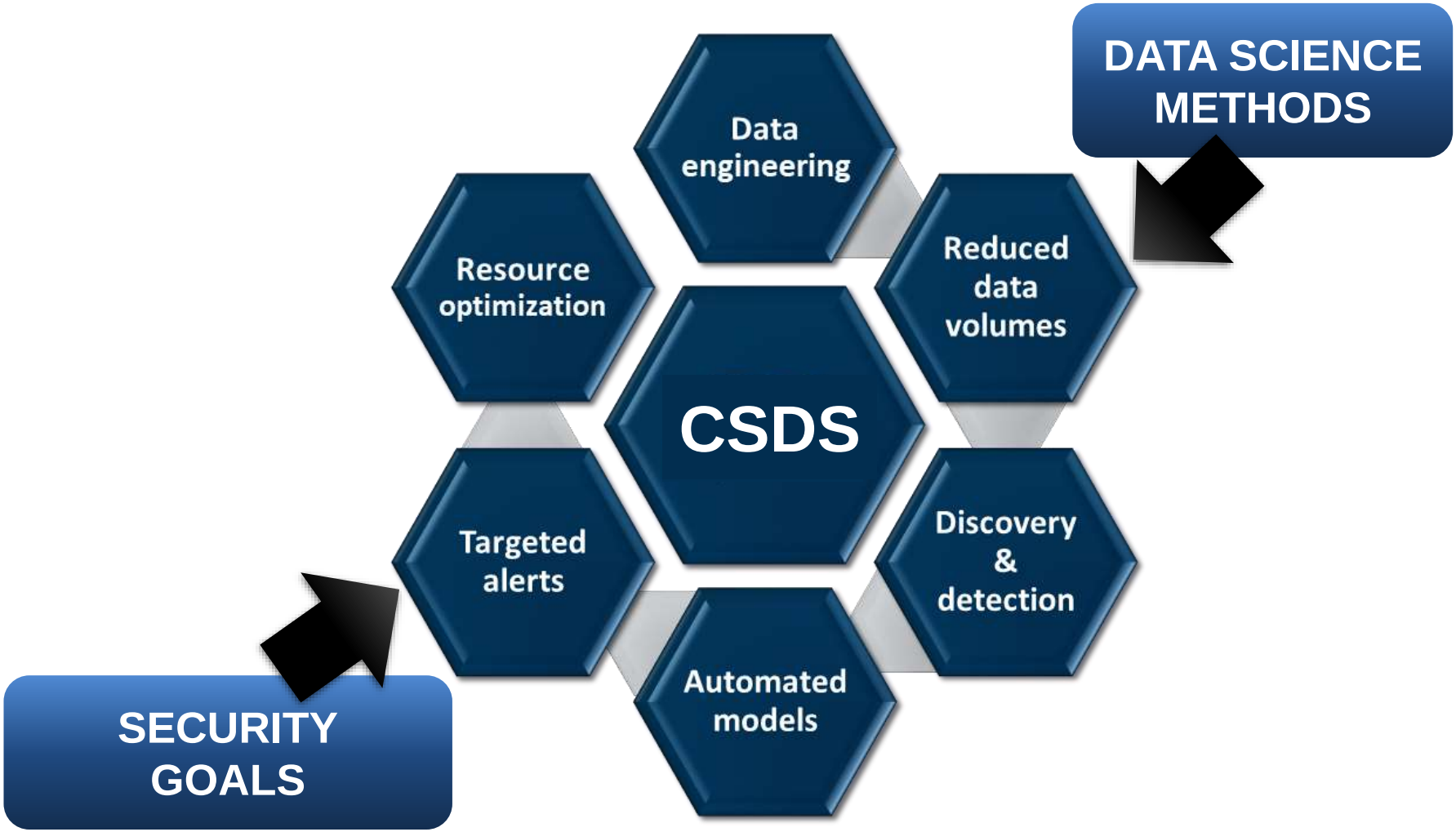


1. Science of Security



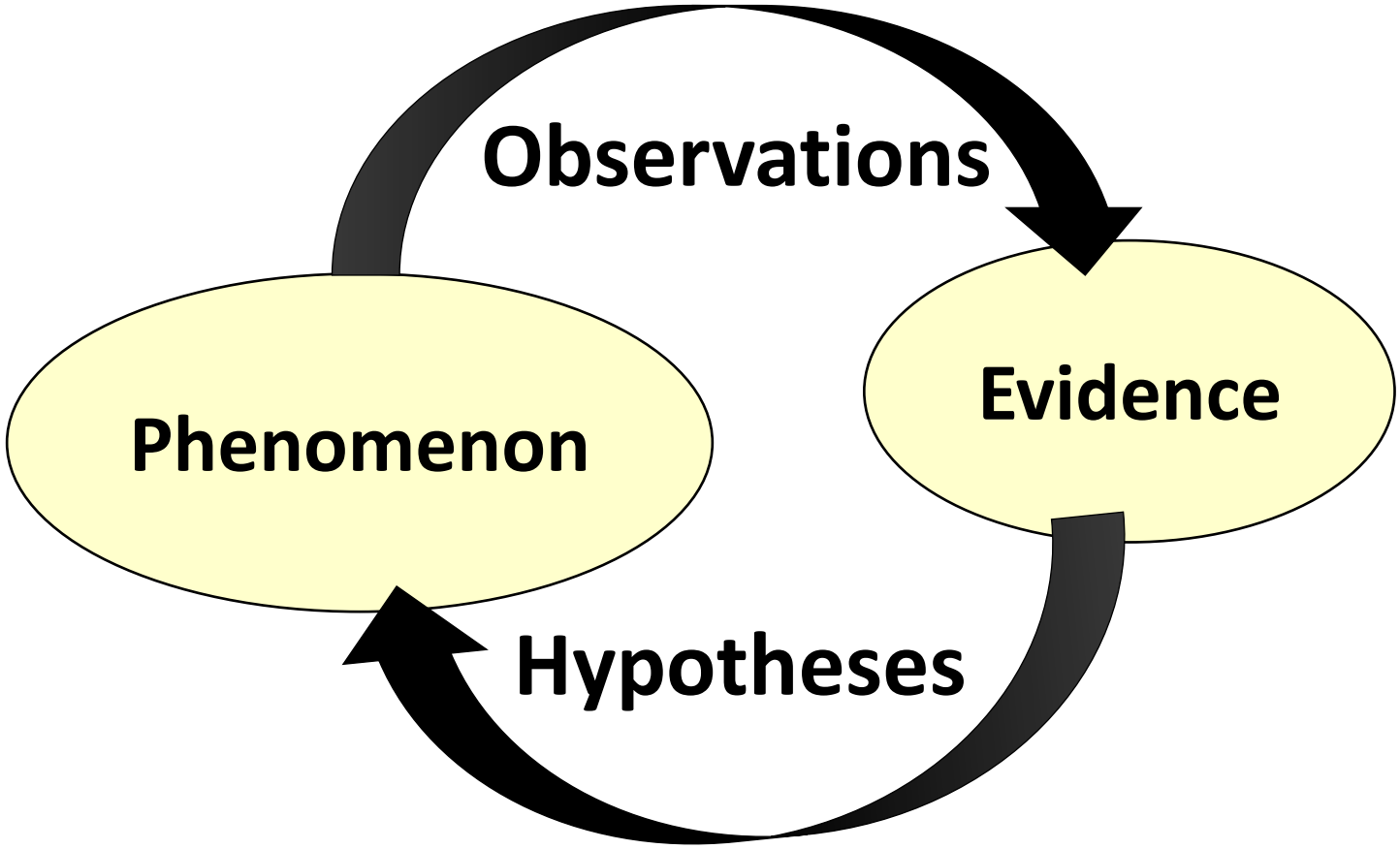

CISO
Mobile
SaaS
Cloud
BYOD
Cyber Security Team
IoT
Insider threat
Microservices
VMs
SOC
Firewall
Whitelisted 3rd party vendors
The " internet "

Framing CSDS



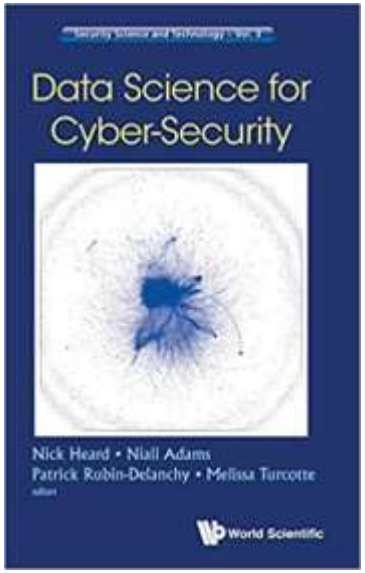
Limits of engineering-based prescriptions

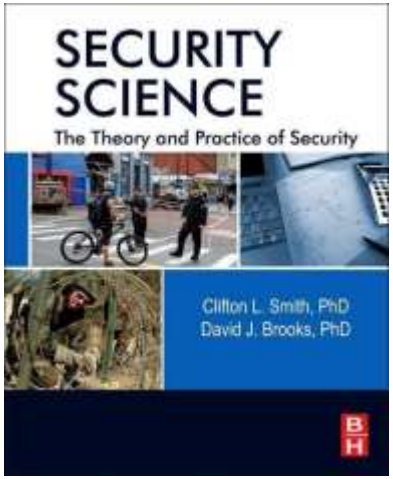
Inference Cycle

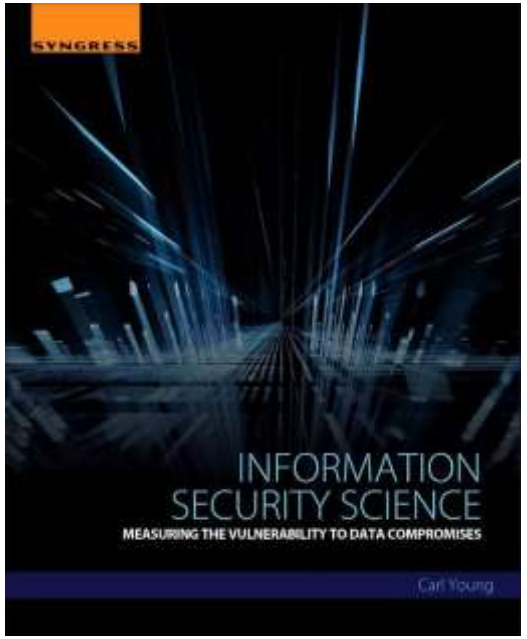


Science in security practice

SCIENCE OF SECURITY





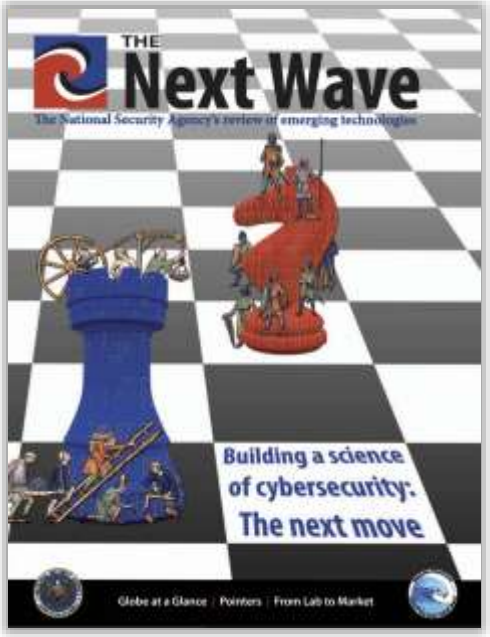


Science in security practice

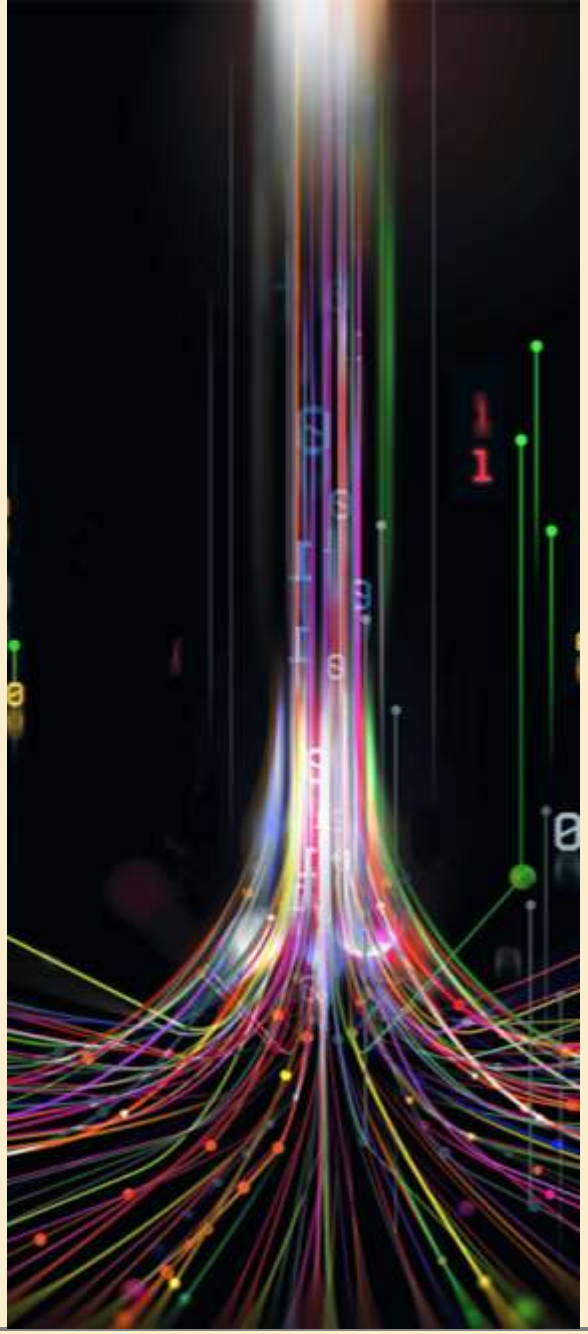
https://www.nsa.gov/Research/The-Next-Wave/

Vol. 19, No. 2 2012

- Amla et al Toward a secure and trustworthy cyberspace
- Landwehr Cybersecurity: from engineering to science
- Longstaff Barriers to achieving a science of cybersecurity
- Maxion Making experiments dependable
- Meushaw, Landwehr NSA initiatives in cybersecurity science
- Pavlovic On bugs and elephants: mining for science of security
- Schneider Blueprint for a science of cybersecurity
- Shostack The evolution of information security



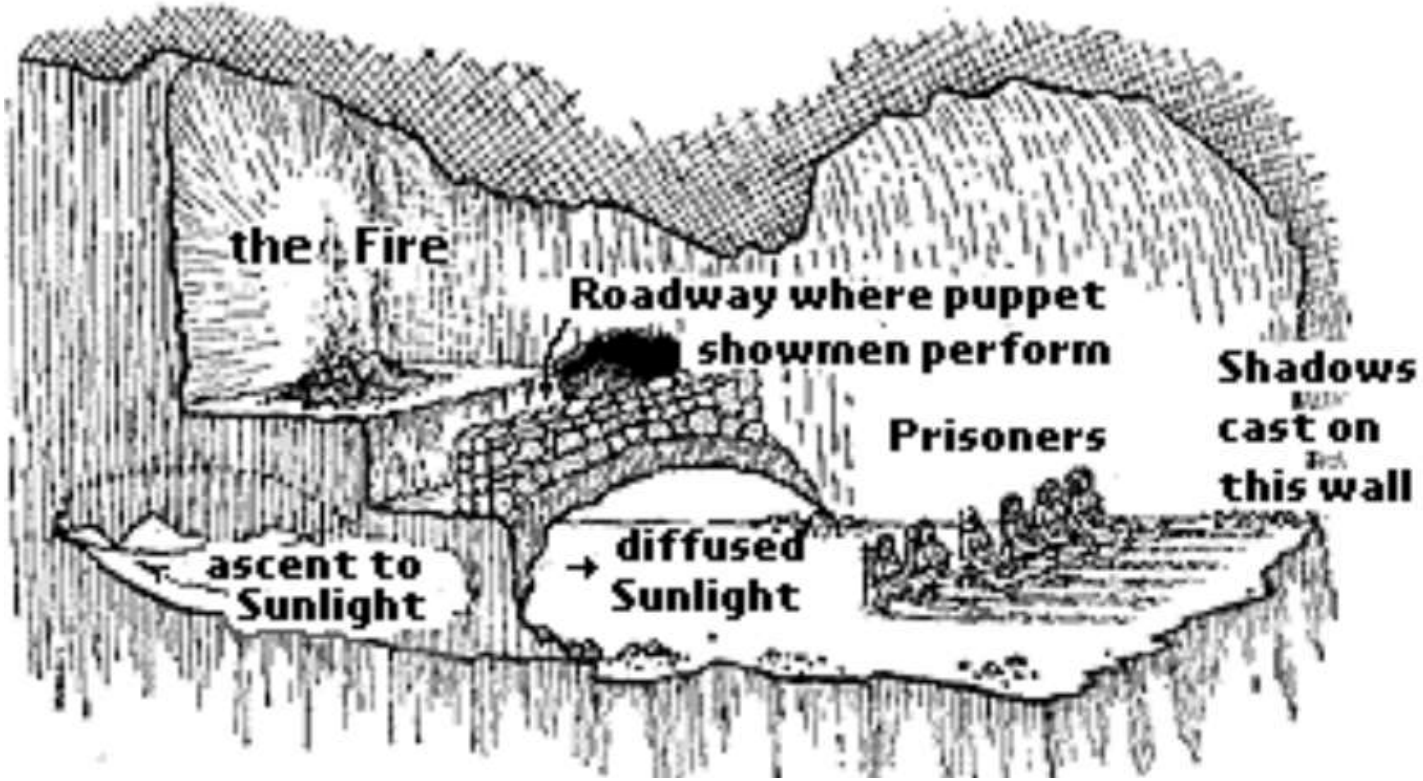
2. Engineering versus Science



Plato's Allegory of the Cave

Attribution https://www.philosophyzer.com/the-allegory-of-the-cave-by-plato-summary-and-meaning/

Plato's Allegory of the Cave



Simarro, Francisco Montero and Víctor López-Jaquero. "From Plato's Dualism to user Interface Development." *WEBIST* (2007).

Medical practice: clinical vs. research

Clinical practice

- Intuition based on experience
- Heuristics
- Processes and procedures
- Intervention on conditions
- Extrapolate from symptoms
- Applications of technology
- Rough consensus

Medical research

- Statistics
- Cumulative theory
- Protocols and rigor
- Address underlying causes
- Substantiate supposed causes

Clinical versus Actuarial Judgement

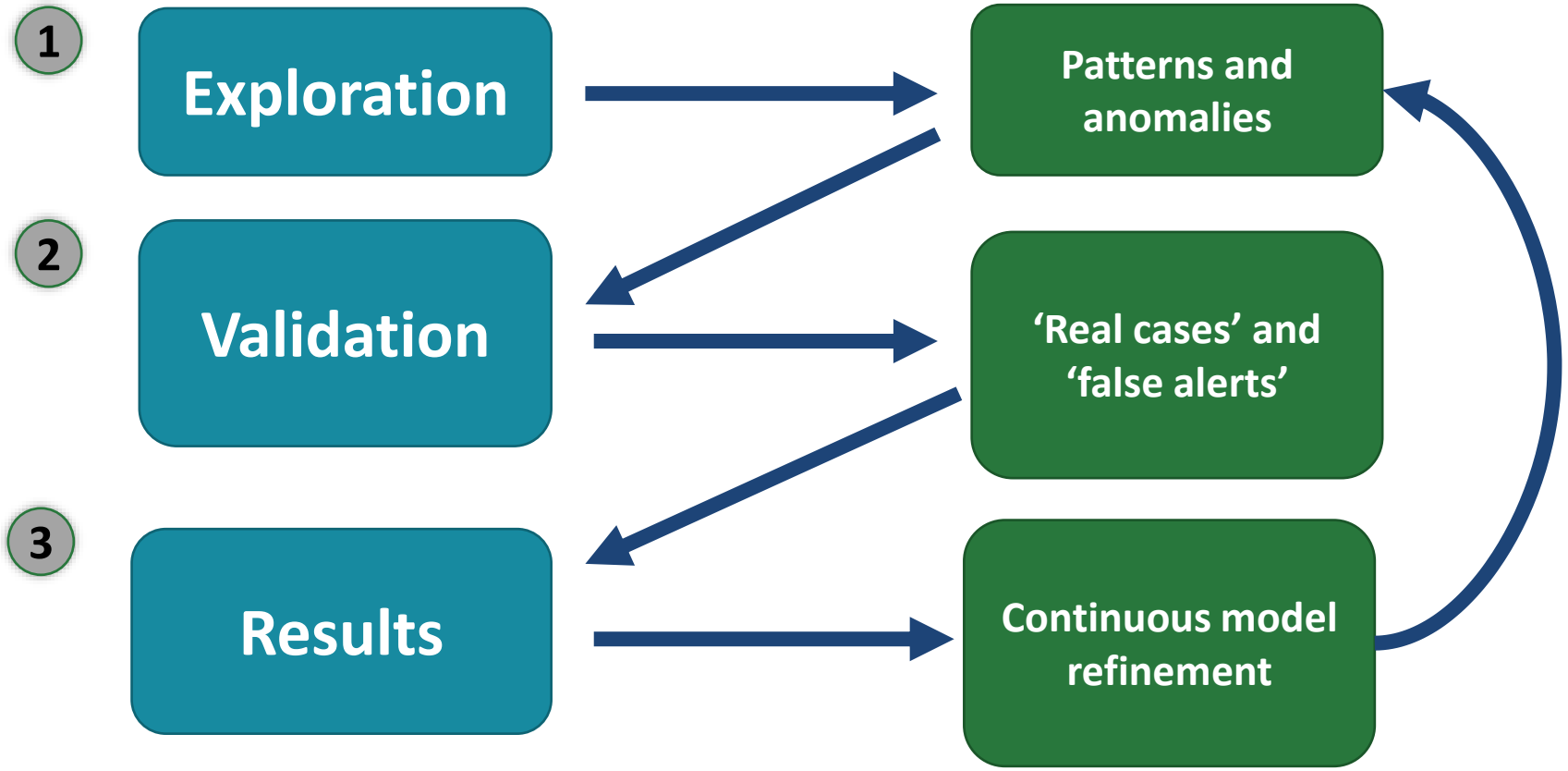
- **Clinical**: ad hoc field experience
- **Actuarial**: empirical statistical observations and testing

- When data is unavailable, operational judgment defers to clinical instincts

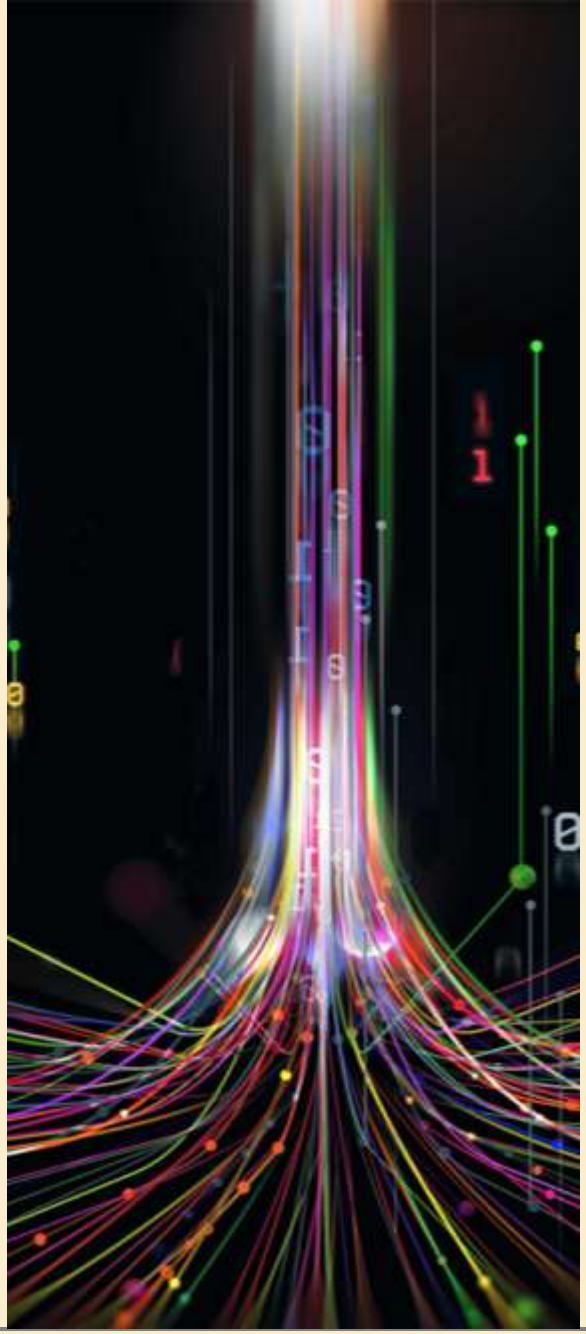
Dawes, R., Faust, D., & Meehl, P. (1989) [Clinical versus actuarial judgement](). Science 243:6

Research comparing these two approaches shows the actuarial method to be superior. Factors underlying the greater accuracy of actuarial methods, sources of resistance to the scientific findings, and the benefits of increased reliance on actuarial approaches are discussed.

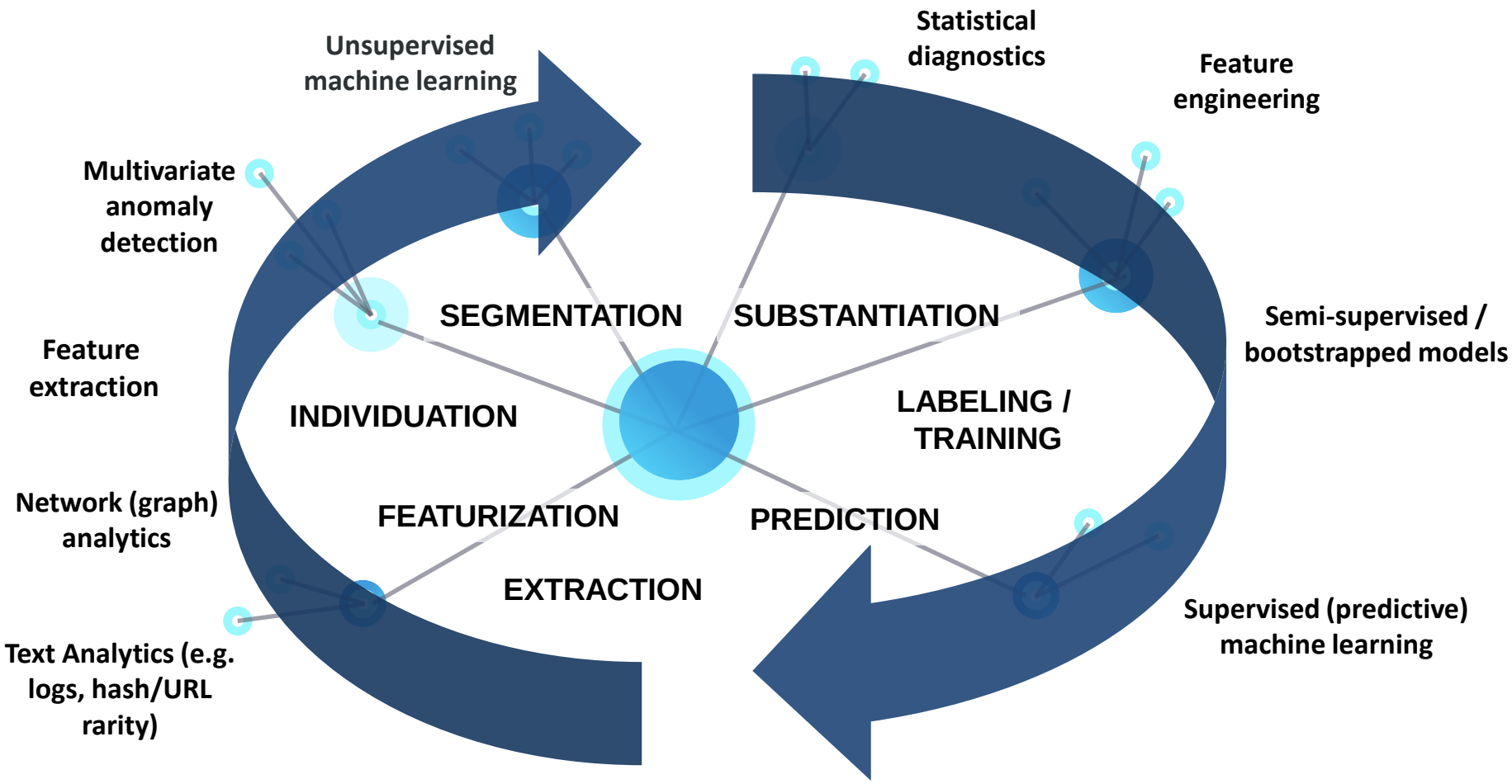
Continuous Improvement Process



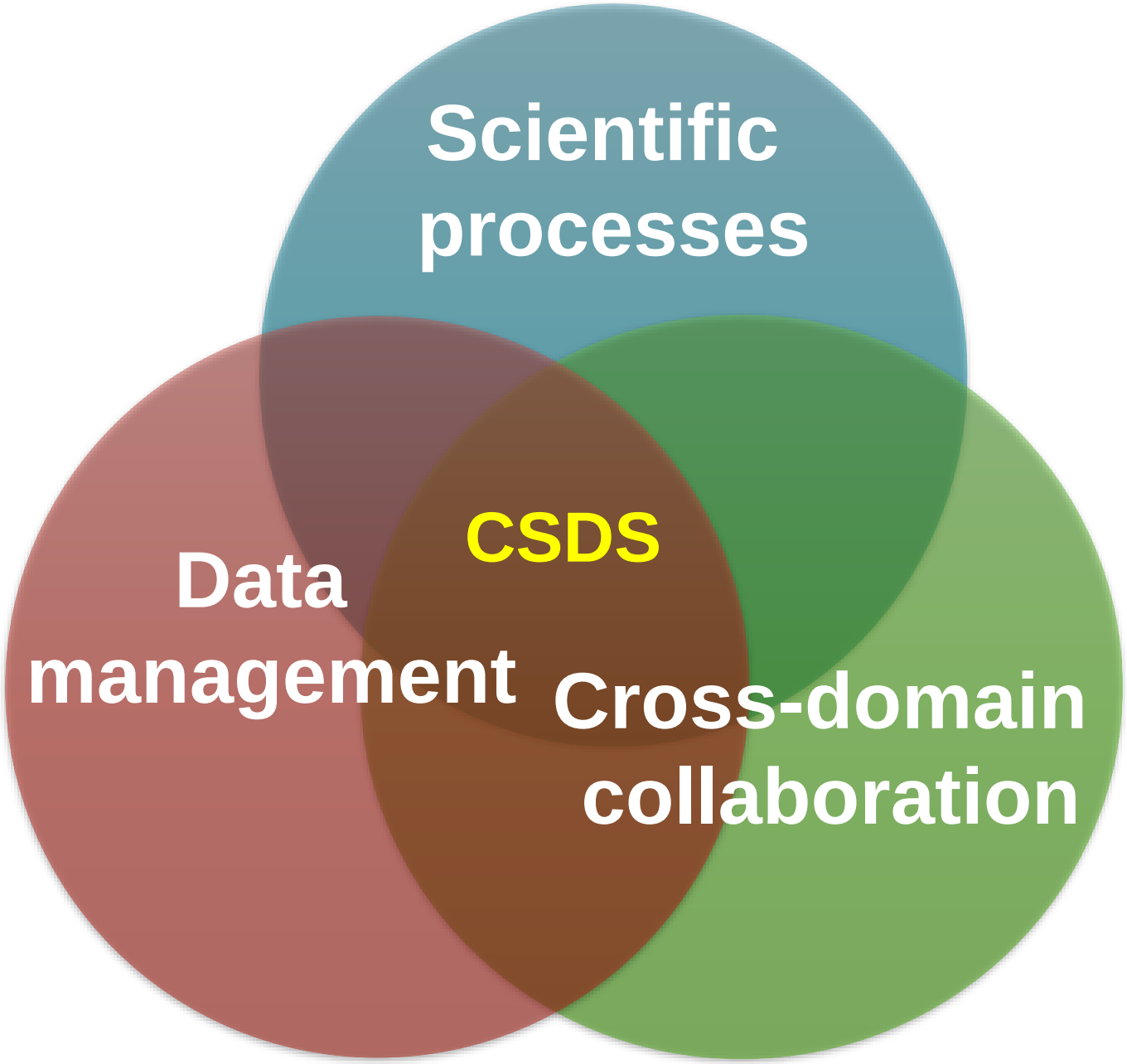
3. Science as a Process



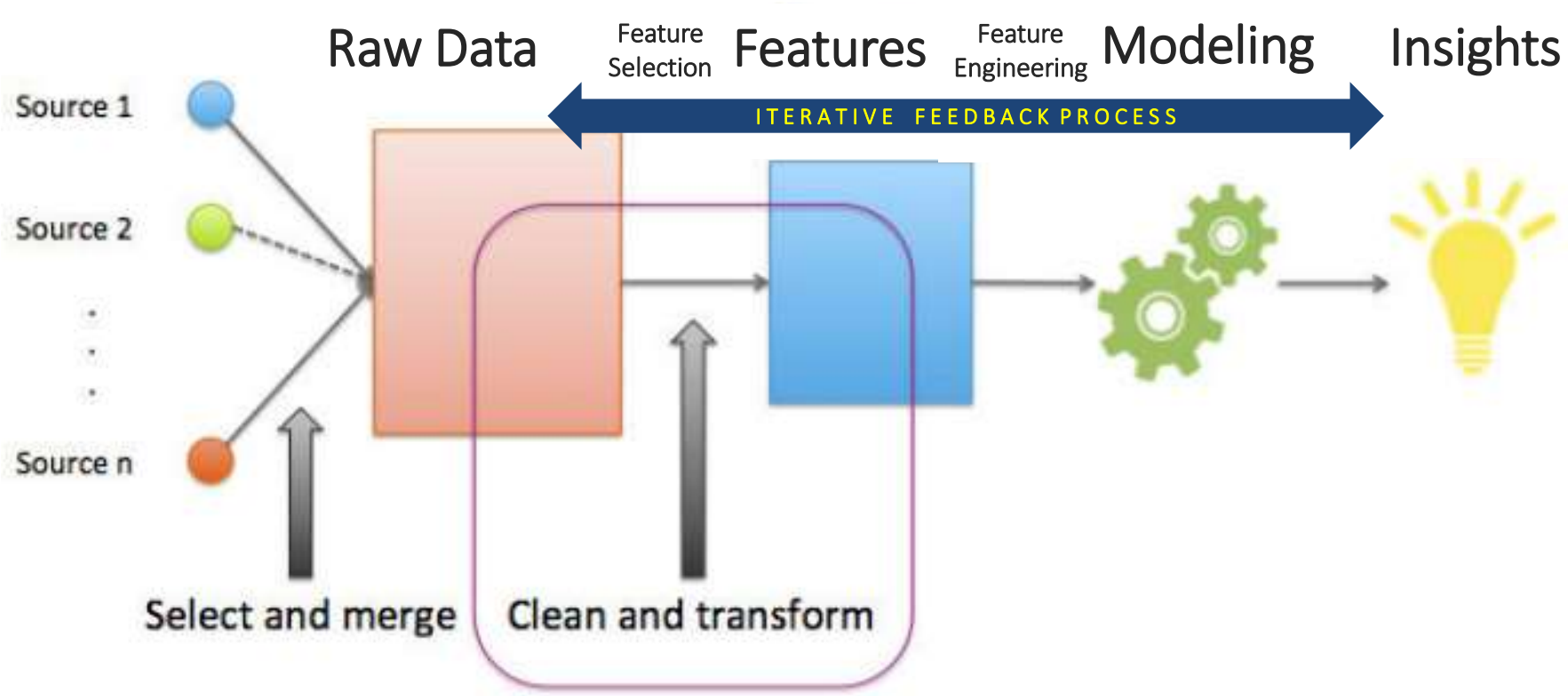
CSDS Process: Linking Methods



We are already doing science, albeit in bits and pieces…

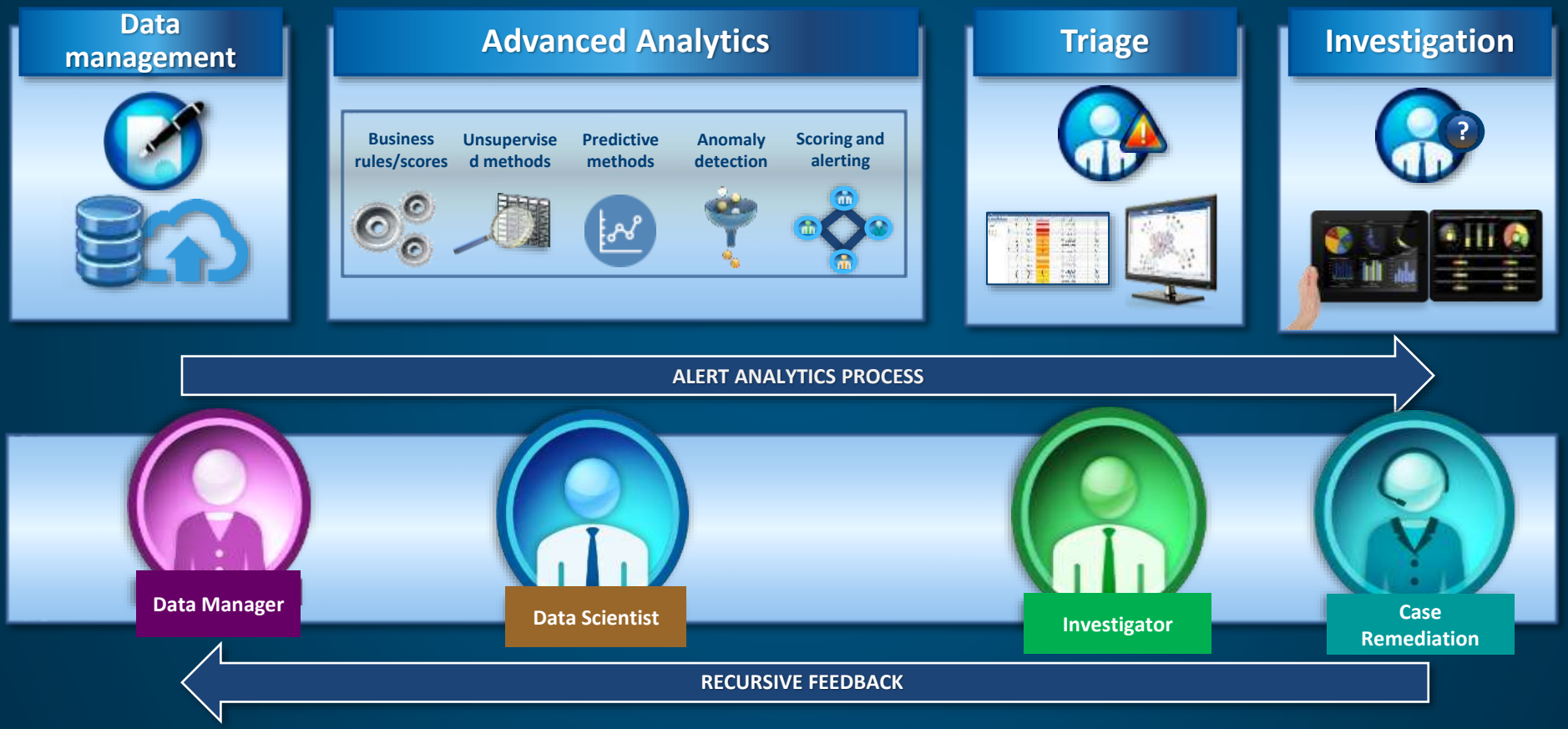

Scientific processes
CSDS
Data management
Cross-domain collaboration

Data Management: Analysis + Features

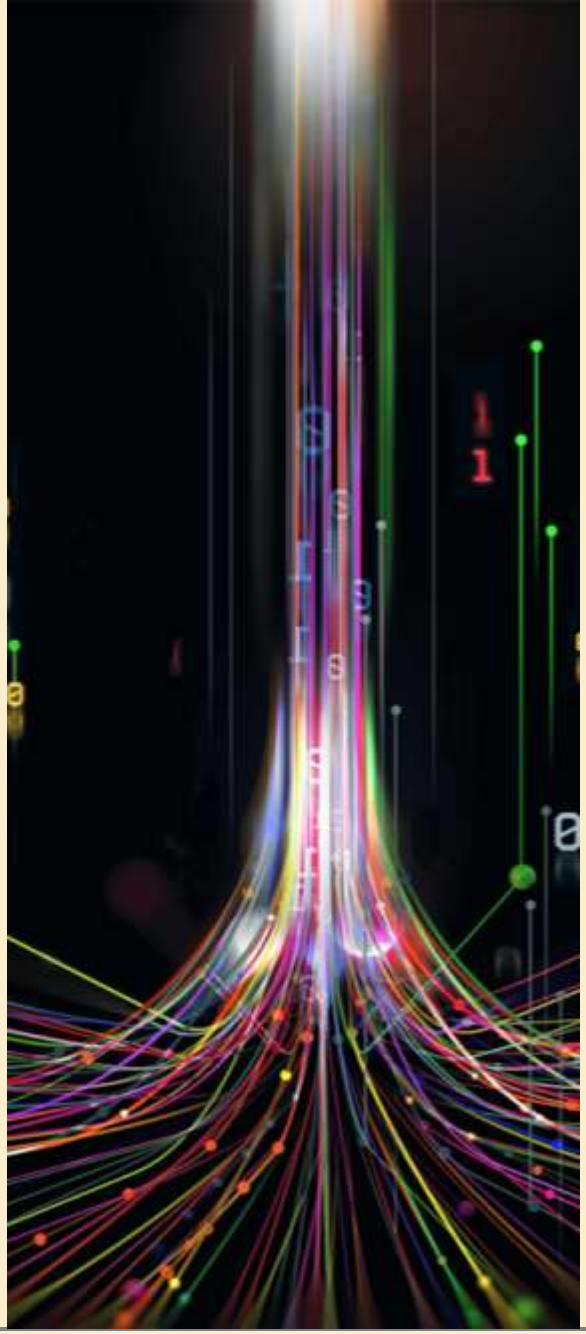


SOURCE: Alice Zheng, Amanda Casari. 2016. Feature Engineering for Machine Learning Models. O'Reilly Media.

CSDS: High-Level Functional Process

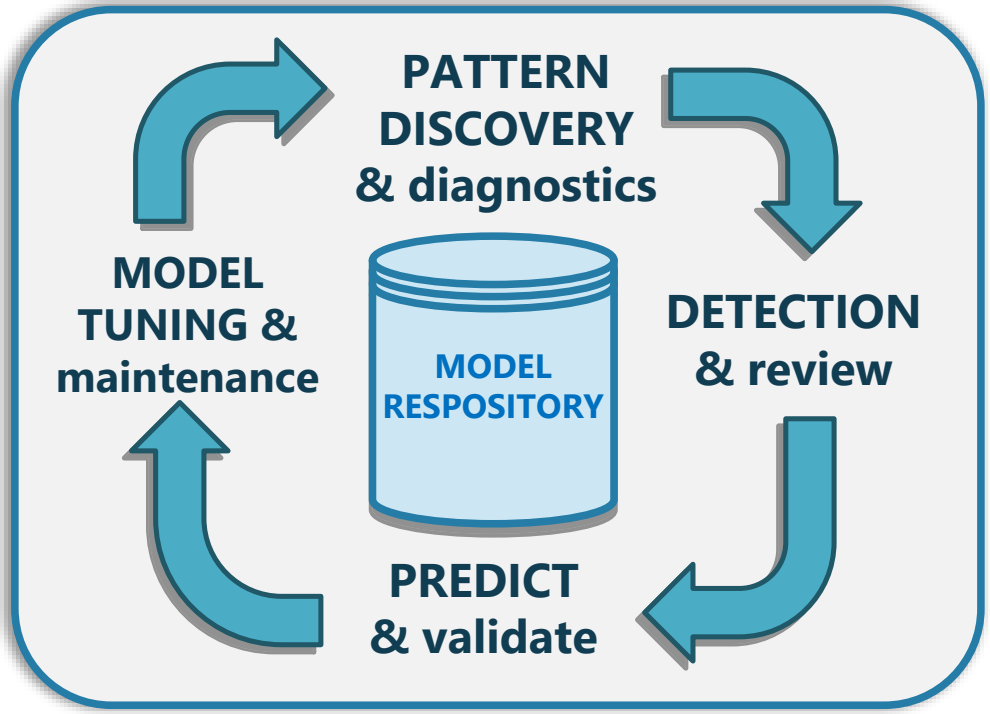


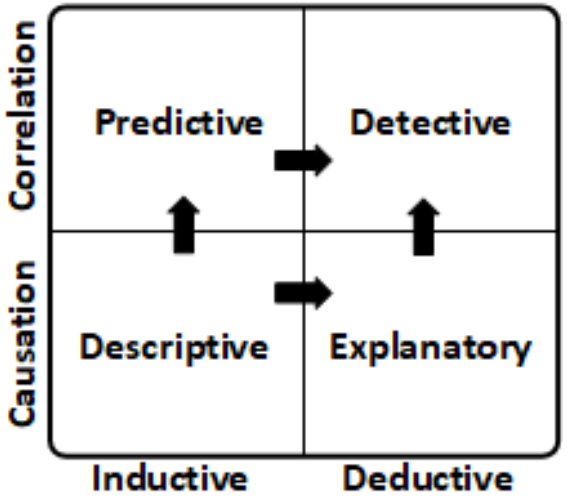
4. Operational Security Science

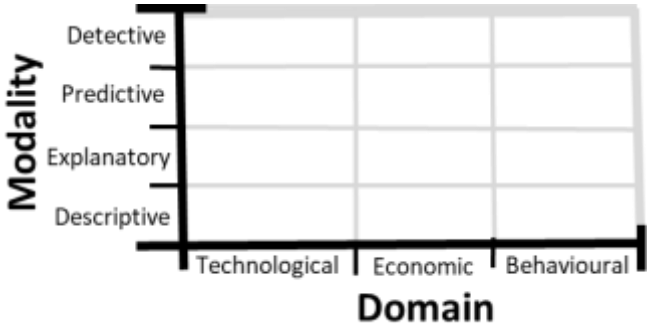


CSDS as a Process: Discovery and Detection

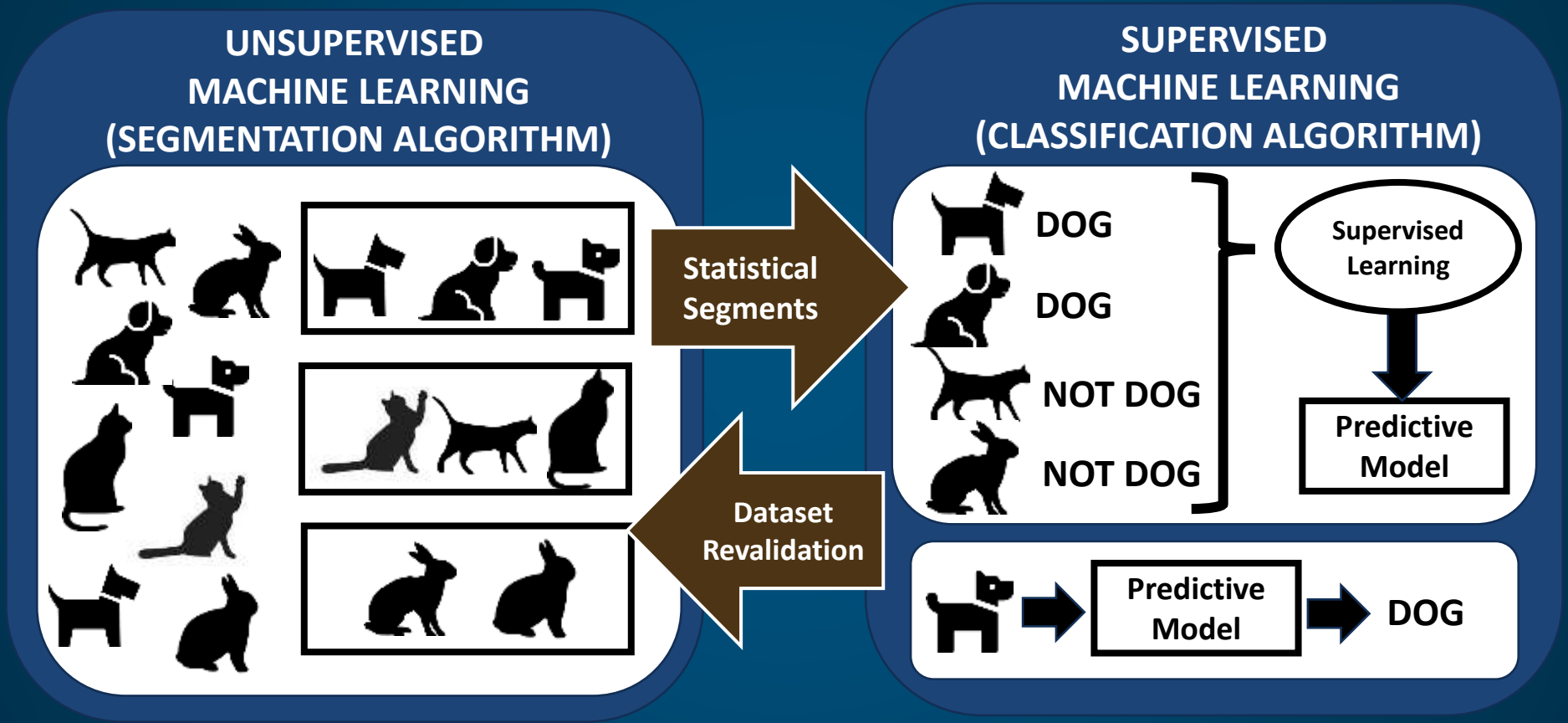
Cyber Analytics: Staged Discovery Process







Unsupervised Category Extraction => Targets

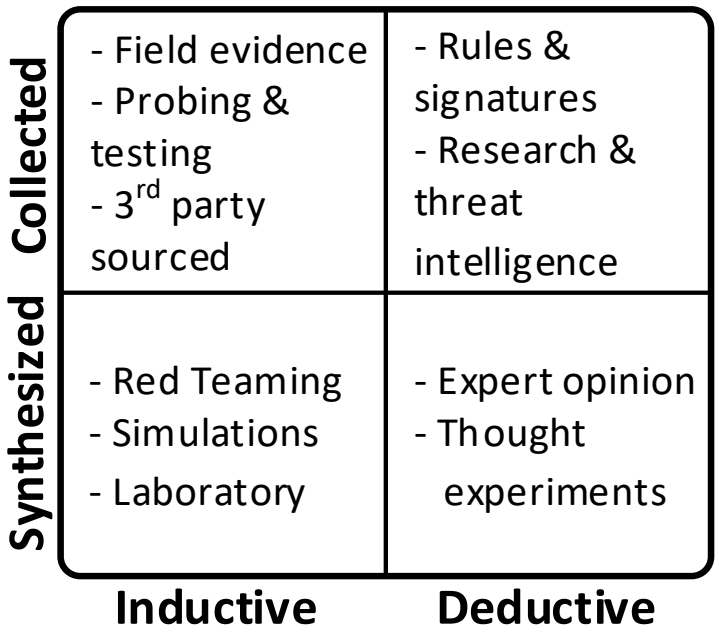


Research Methods for Cybersecurity

- *Experimental*
  - ➢ i.e. hypothetical-deductive and quasi-experimental
- *Applied*
  - ➢ i.e. applied experiments and observational studies
- *Mathematical*
  - ➢ i.e. theoretical and simulation-based
- *Observational*
  - ➢ i.e. exploratory, descriptive, machine learning-based

Manz, D. and Edgar, T. (2017)
Research Methods for Cyber Security

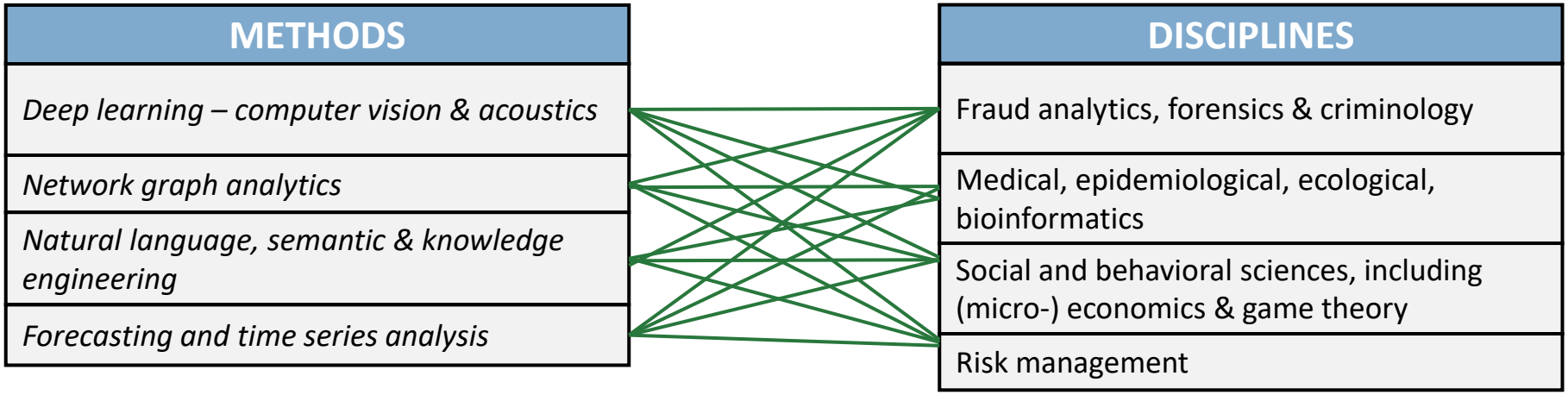
Labels: What constitutes 'evidence'?



EXAMPLES OF SECURITY EVIDENCE

1. Field evidence (e.g. observed incidents)
2. Sourcing own data from field testing (e.g. local experiments)
3. Honeypots
4. IDSs (Intrusion Detection Systems)
5. Simulation findings
6. Laboratory testing (e.g. malware in a staged environment)
7. Stepwise discovery (iterative interventions)
8. Pen testing (attempts to penetrate the network)
9. Red teaming (staged attacks to achieve particular goals)
10. Incidents (records associated with confirmed incidents)
11. Reinforcement learning (self-improving ML to achieve a goal)
12. Research examples (datasets recording attacks from research)
13. Expert review (opinion and guidance from experts)
14. Intelligence feed (indications from a 3rd party service)
15. Thought experiments (e.g. boundary conditions, counterfactuals)

CSDS Methods & Approaches



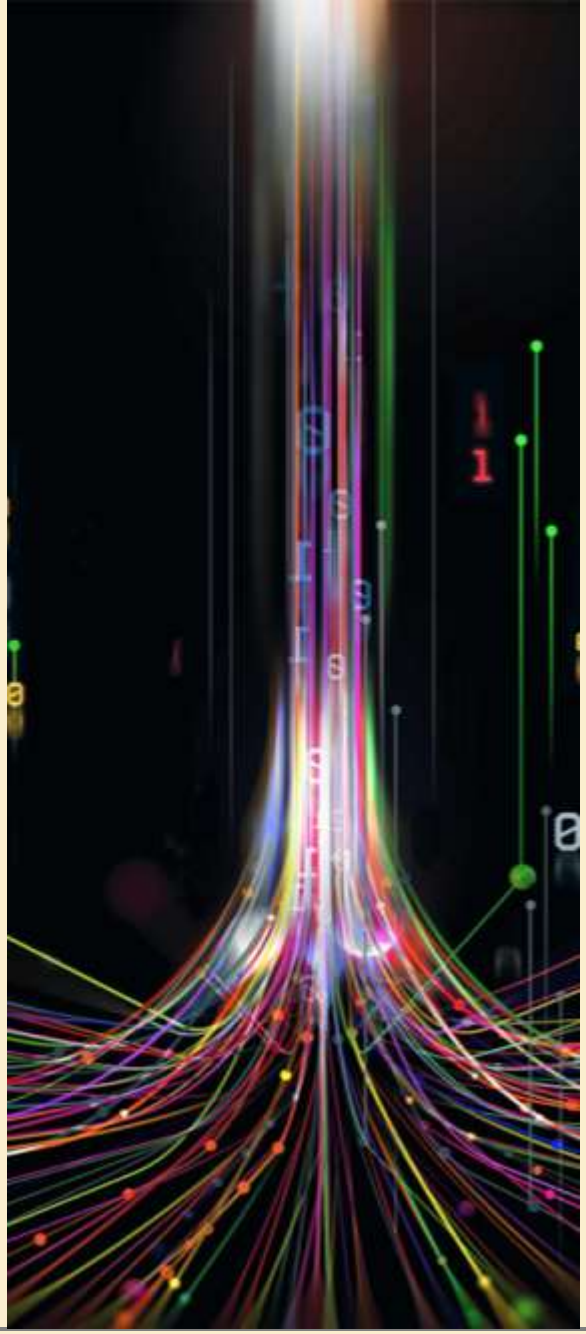
NOVEL METHODS

- Pattern recognition
- Classification
- Feature extrapolation
- Diagnostics
- Multivariate inferential statistics

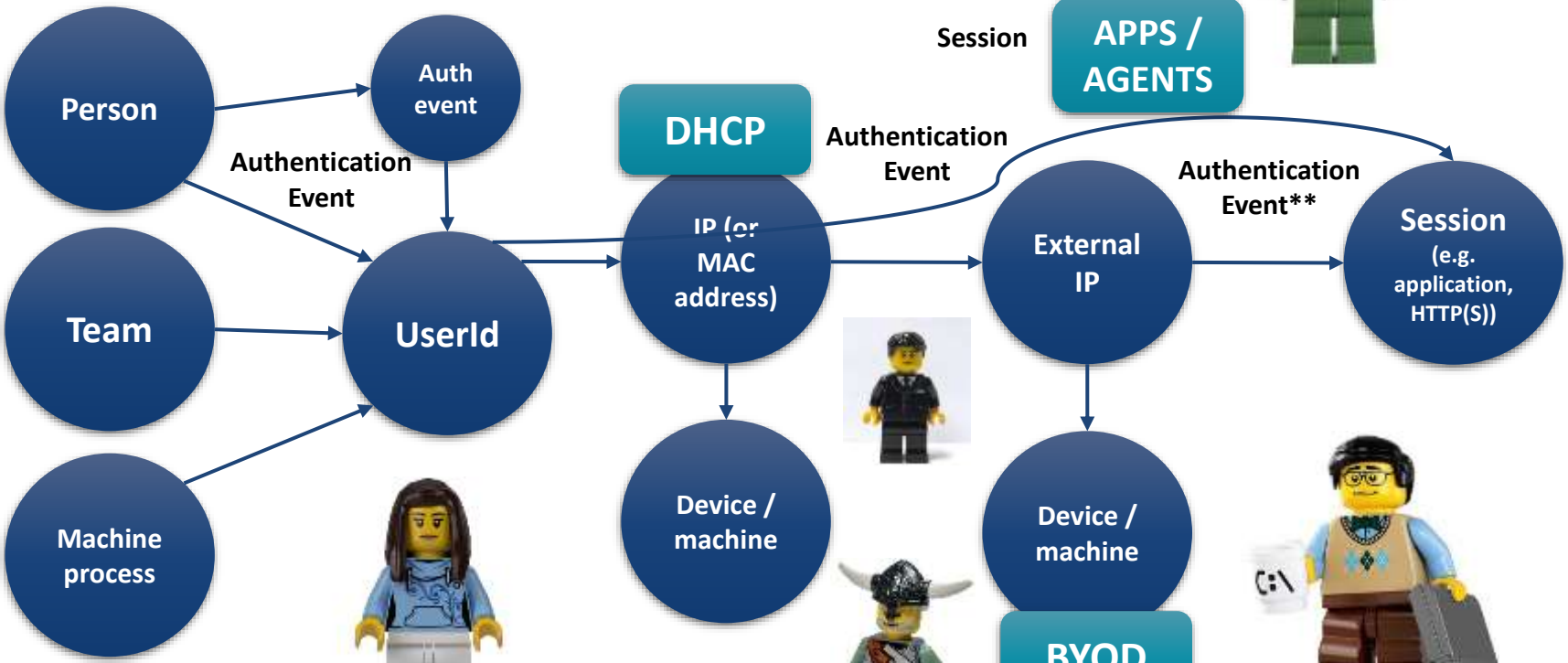
NOVEL CONTEXTS

- Complex multi-domain models
- Technological
- Economic
- Behavioral
- Techno-economic-behavioral context
- Deterministic + latent variables

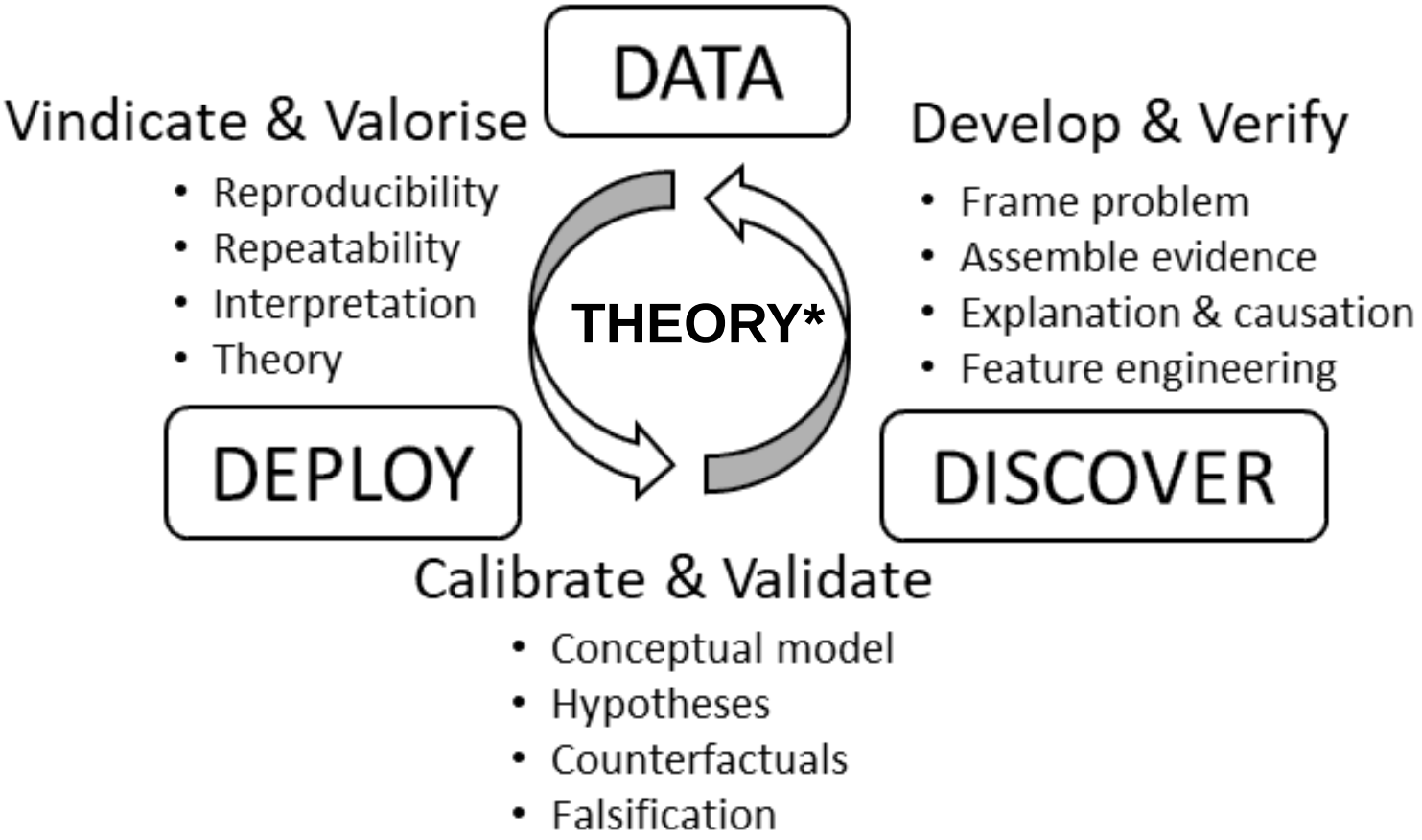
5. Where do we go from here?



What is a ‘User’?

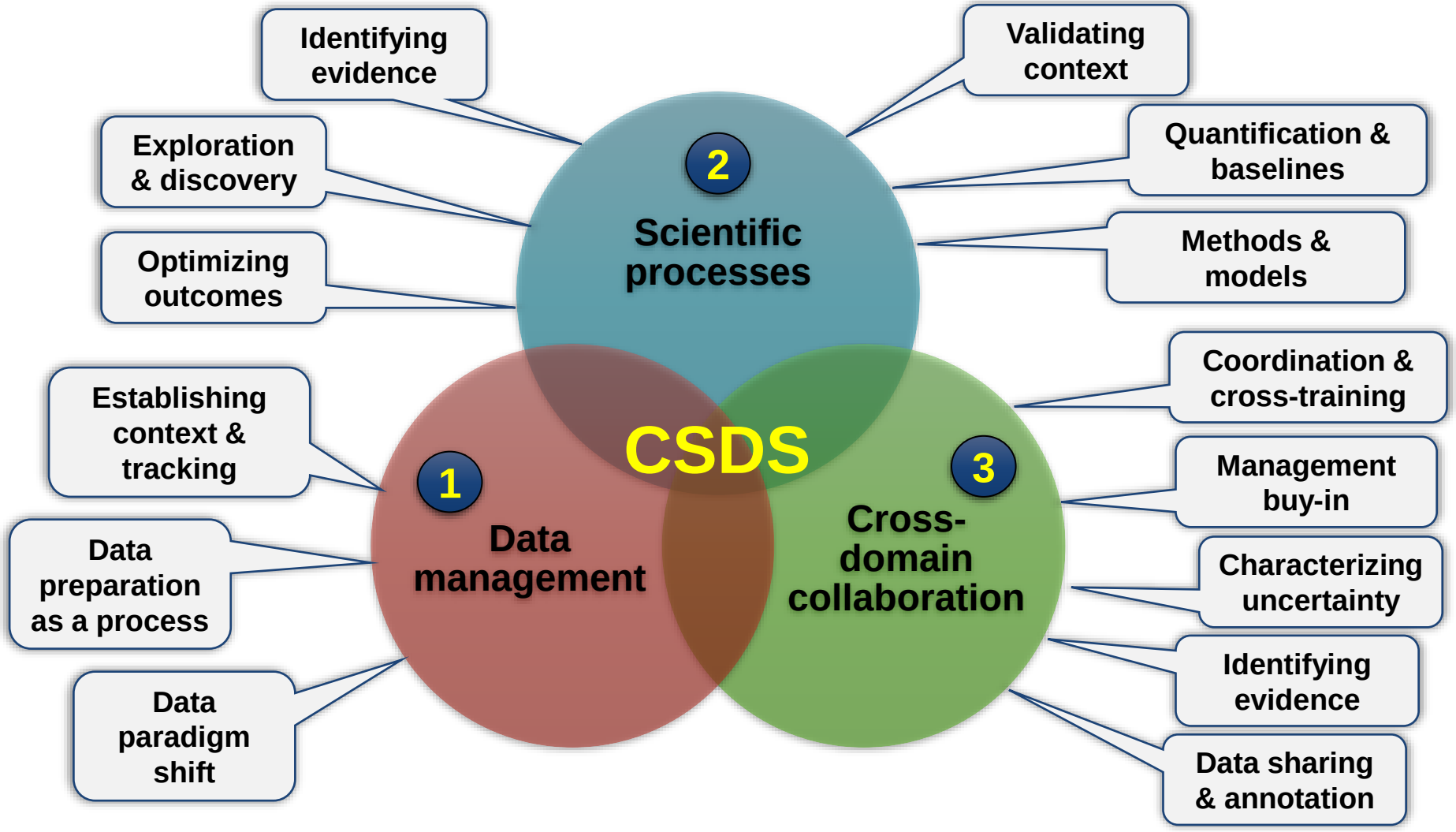


CSDS as a Unified Process



** As cumulative, validated, socialized knowledge*

CSDS: A Roadmap for Rigor



FloCon 2022

18th Annual Open Forum for Large-Scale Data Analytics

Using Data to Defend

What Do We Mean by a Science of Security?

END

Dr Scott Mongeau

scott@sark7.com

Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA 15213

APPENDIX

Scientific methods in cybersecurity

Cybersecurity Data Science (CSDS) Corpus

April 12, 2020 Advocacy, Best practices, Management, Methods, Research, Theory

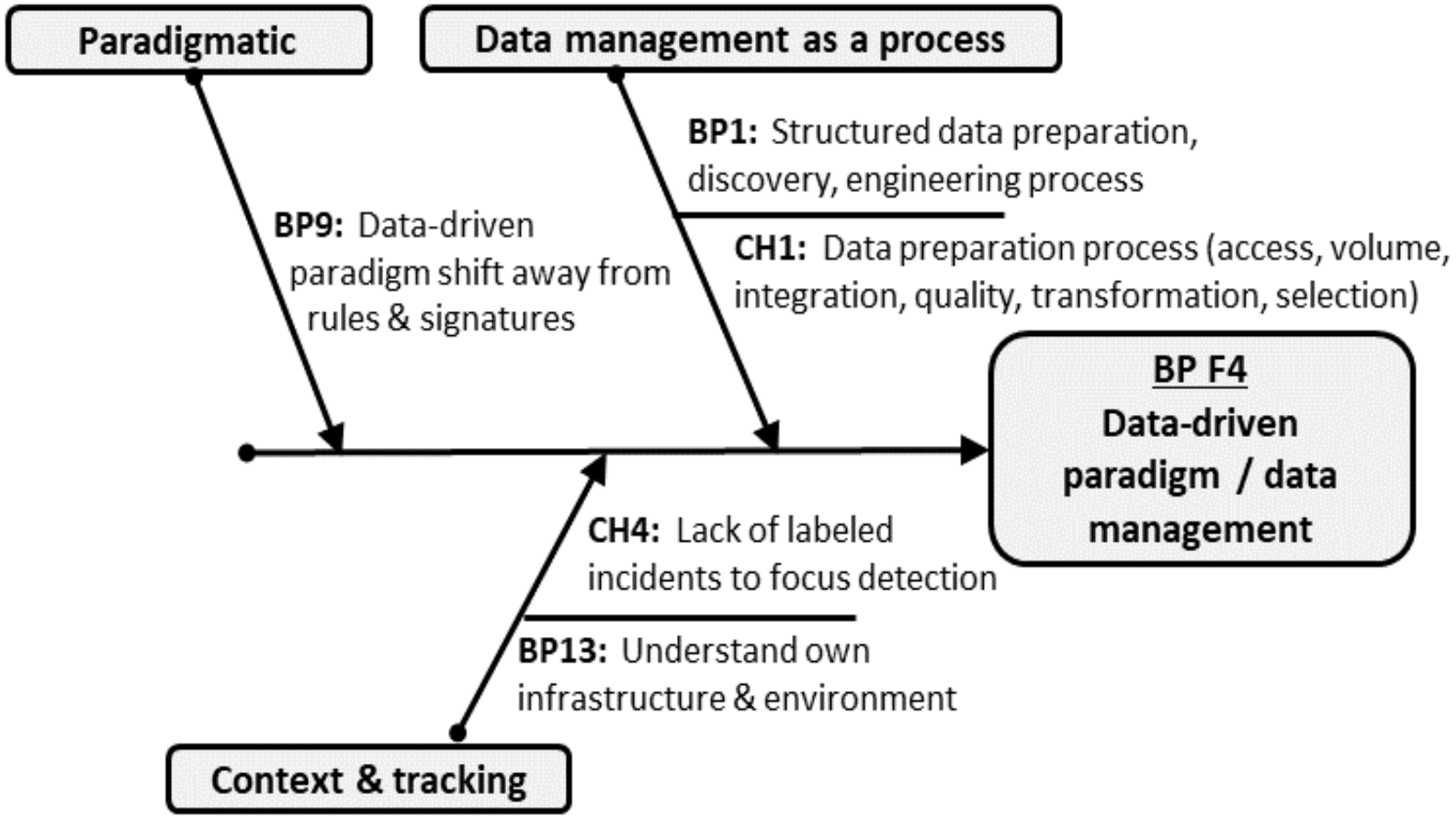
OVERVIEW

For those interested in the rapidly emerging field of cybersecurity data science (CSDS), below is a corpus of 33 book-length works. The list covers publications going back to 2001, although two-thirds of the works (22 out of 33) were published in the last five years (2016 to 2020).

CSDS Corpus

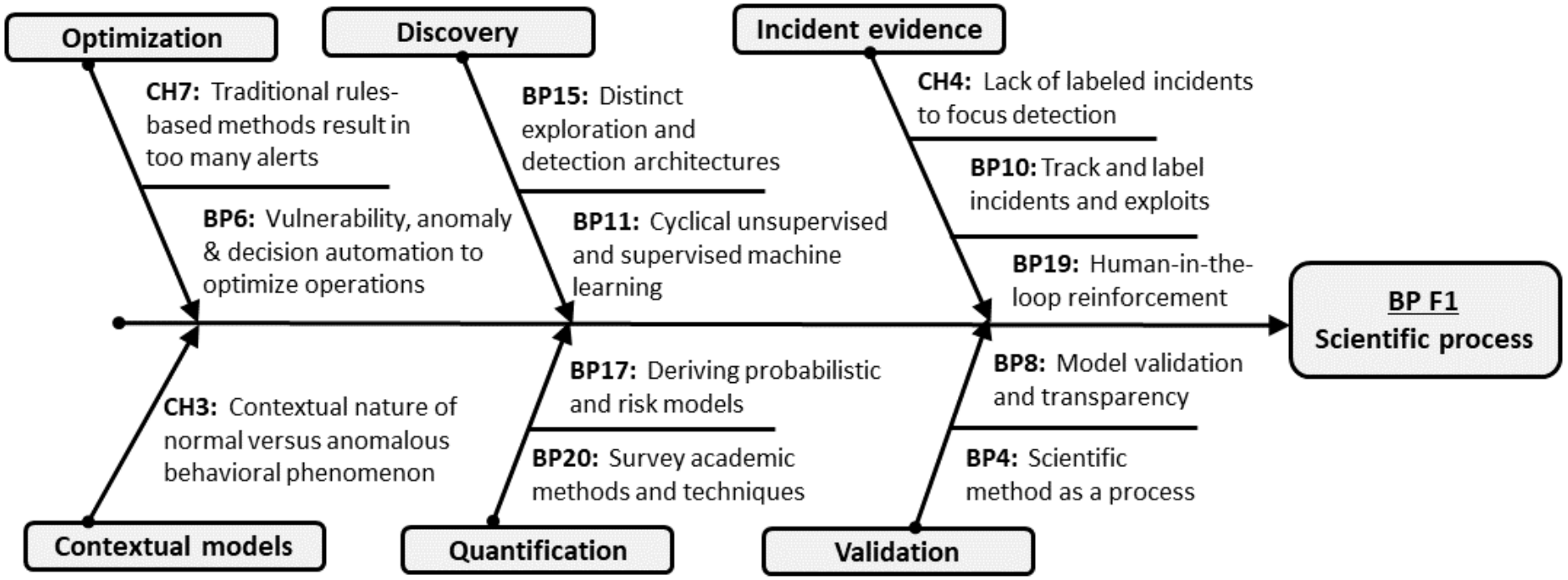
BOOK	AUTHORS
Data-Driven Security	Jacobs & Rudis, 2014
Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques	Baesens et al., 2015
Essential Cybersecurity Science	Dykstra, 2016
Dynamic Networks and Cyber-Security	Adams & Heard (Eds.), 2016
How to Measure Anything in Cybersecurity Risk	Hubbard & Seiersen, 2016
Research Methods for Cyber Security	Edgar & Manz, 2017
Big Data Analytics in Cybersecurity	Savas & Deng (Eds.), 2017
Cybersecurity Analytics	Verma & Marchette, 2020

CSDS Best Practice I: Data Management



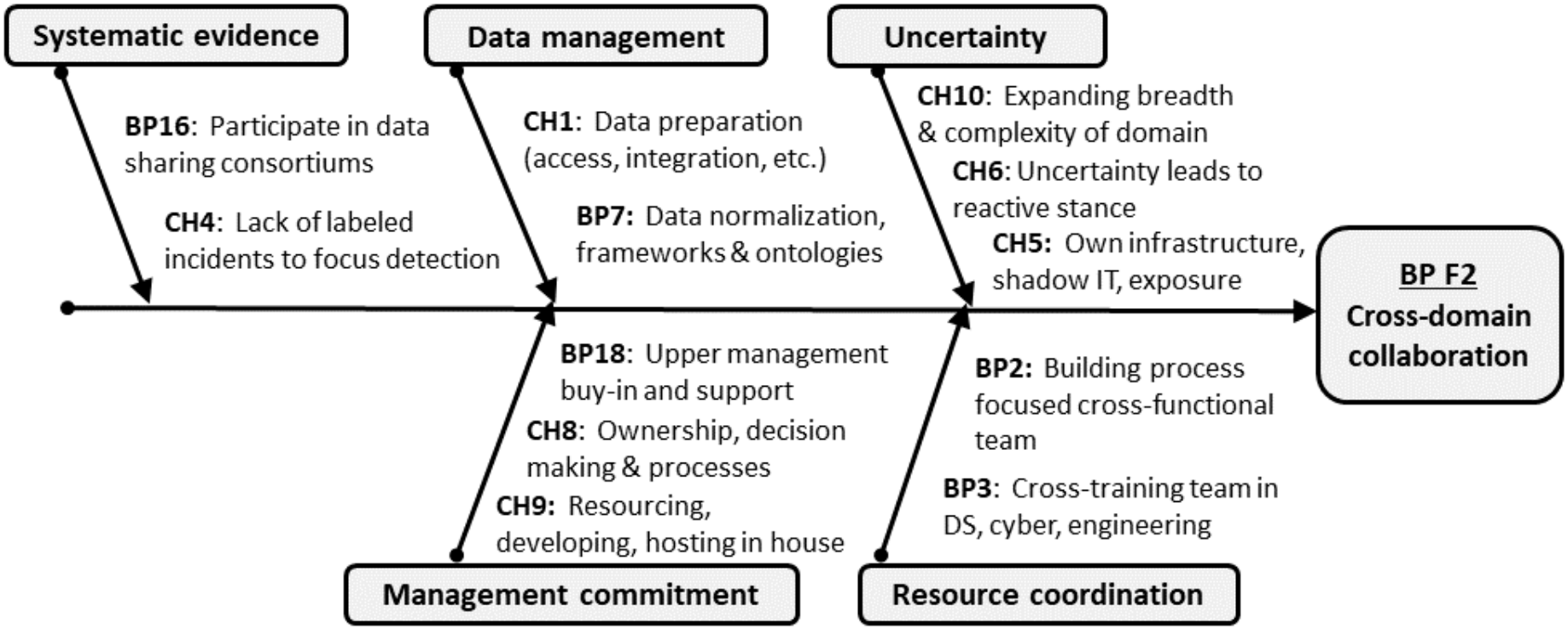
Mongeau 2021 '[Cybersecurity Data Science]' Springer

CSDS Best Practice II: Scientific Process



Mongeau 2021 '[Cybersecurity Data Science]' Springer

CSDS Best Practice III: Cross-Domain Collaboration



Mongeau 2021 '[Cybersecurity Data Science]' Springer

This slide has intentionally been left blank.