BUILDING AN AGILE KNOWLEDGE MANAGEMENT DATA CENTER

Eric Stith

812 TSS/ENTI

DISTRIBUTION STATEMENT A - Approved for public release; distribution is unlimited. 412TW-PA-21211

CROSS PLATFORM DATA CENTER

- Background
- Goals
- How can Agile processes and tools help the Cross-Platform Data Center's Knowledge Management

CROSS PLATFORM DATA CENTER

"Test & Evaluation data is currently isolated & compartmentalized with little discovery or reuse outside of each respective acquisition program, minimizing its long-term applicability and effectiveness." TRMC

Cross Platform Data Center is an Air Force Investment Program that has been funded to help address the issue stated above by implementing some of the tools developed by Test Resource Management Center.

WHY CROSS PLATFORM DATA CENTER

- **Most investment has been on *test* needs, but this is an *evaluation* need.**
  - Need an evaluation strategy to not only analyze today's questions but future questions, even after fielding.
- **T&E quality is inadequate for our needs**
  - More data is being collected than can be properly analyzed
  - Only a tiny fraction of data is looked at
  - Only simplistic analysis is being done on a small fraction of data
  - Global view of the collected data is rarely, if ever done
  - No systematic anomaly detection, trend analysis, regression analysis, causality analysis, pattern recognition, simulation/test comparisons, perceived truth/ground truth comparisons are being done.

WHY CROSS PLATFORM DATA CENTER

- **T&E timeliness is inadequate for our needs**
  - Analyst retrieval of test data in some cases takes more than a week
  - Sometimes it's easier (though not cheaper) to just re-run a test rather than find old data that may answer your question
  - Long data ingest times prevent:
    - Proper and timely debriefing of test participants after a test is over
    - Correlation of data in real time
- **How can T&E be accomplished with improved effectiveness?**
  - Can fewer tests be run with an efficient process to analyze data?
  - Cross-program lessons learned are not being made, except anecdotally, can this be changed?

CROSS PLATFORM DATA CENTER

ANYTHING ALARMING ?

Do you have a central location for all types of data related to a single test?

CROSS PLATFORM DATA CENTER

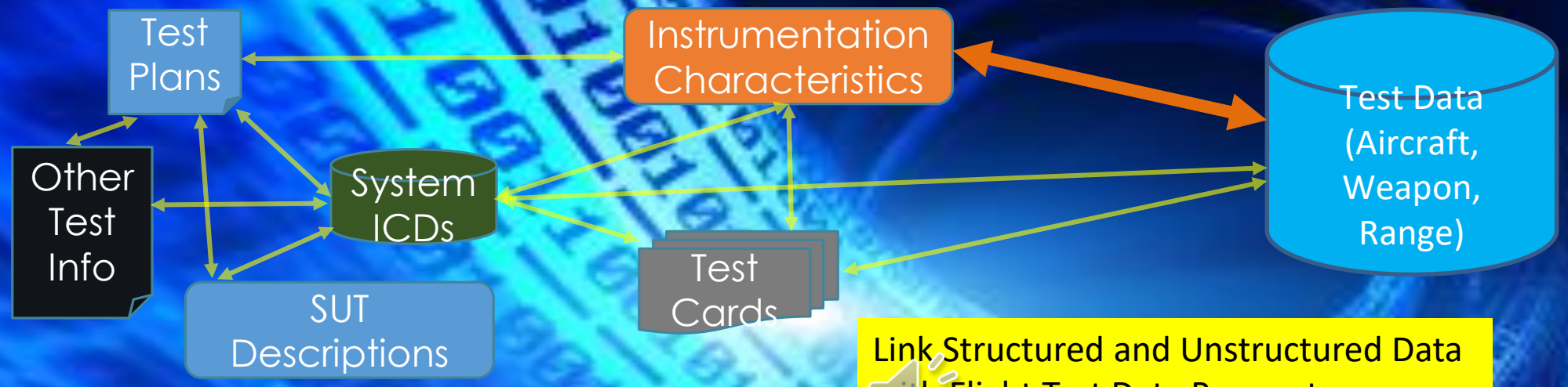
What is Big Data?

Big Data is whatever amount of data breaks your current system

Questions the CPDC seeks to address:

- Anomaly Detection – Did something go wrong?
- Causality Detection – What contributed to it?
- Trend Analysis – What's happening over time?
- Predicting Equipment Function and Failure – When will something go wrong?
- Regression Analysis – How is today's data different than the past?
- Data Set Comparison – Are these two large data sets equivalent?
- Pattern Recognition – Are there any recognizable patterns in the data?

WHAT IS CROSS PLATFORM DATA CENTER?



Link Structured and Unstructured Data with Flight Test Data Parameters

- **Main Objectives**
  - **Create a government-owned and -operated data center with capabilities for the CTFs that:**
    - **Uses a sustainable Agile development cycle that encourages collaboration and adoption of common tools**
    - **Enables a Knowledge Management cycle for T&E data**

WHAT IS AGILE? IS IT THE SAME AS CI/CD OR DEVOPS?

- They aren't the same thing, but they're related and are often used together.
- Agile is focused on the development process, and it is therefore the most broadly applicable.
- DevOps/DevSecOps/Dev*Ops are focused on the culture of how specialties interact with each other.
- CI/CD is a specific practice related to software release cycles.
- Because Agile is a process, it is the broadest term, but a good Agile process will often require the use of DevOps and CI/CD. We will discuss how culture and practice can contribute to a healthy knowledge management cycle.

AGILE VERSUS WATERFALL

- Waterfall is the process that envisions development as a set of stages when one flows into the next until completion, rather than a cycle of development.
  - Long release cycles
  - Large changes
  - Long, difficult deployments
  - All requirements must be known at the outset
  - Difficult to reverse back to a previous stage
- Agile prioritizes a cyclic development process that progresses in small increments toward the desired solution.
  - Short release cycles (sprints)
  - Fast delivery of new features
  - Easier deployments and migrations
  - Encourages teamwork
  - Testing occurs throughout development, which helps to find problems early
  - Sustainability

CPDC AGILE DEVELOPMENT PROCESS

- Requirements Gathering
- Design
- Develop
- Test
- Deploy
- Review
- **ITERATE!** – Go back to the beginning

CPDC AGILE DEVELOPMENT PROCESS

- Requirements Gathering
  - Gather known customer requirements
  - Define system specifications to meet given performance needs
  - Leverage past experience to anticipate unspecified needs from past needs

CPDC AGILE DEVELOPMENT PROCESS

- Design
  - Select tools that are standardized and scalable
  - Encourage the use of newer tools and technologies, especially open source
  - Pick common tools to automate and standardize deployments

CPDC AGILE DEVELOPMENT PROCESS

- Develop
  - Prototype a software solution on commodity hardware
  - Enable the use of existing tools and a variety of languages (C/C++, C#, R, Python, etc.)
  - Standard DevOps Automation tools

CPDC AGILE DEVELOPMENT PROCESS

- Test
  - Unlike Waterfall, testing isn't a separate phase, but rather a part of the development process, and the two occur concurrently.
  - Always test with carefully cultivated test data that tests all code branches
  - Use code coverage analysis tools
  - Continuous Integration (CI)

CPDC AGILE DEVELOPMENT PROCESS

- Deploy
  - Deploy to a production system (single classification/caveat)
  - Incorporate real data
  - Do this frequently in order to find problems and complete the cycle
  - Continuous Delivery or Continuous Deployment (CD)

CPDC AGILE DEVELOPMENT PROCESS

- Review
  - Figure out what worked well and what didn't
  - Lessons learned
  - Interact with the customers

CPDC AGILE DEVELOPMENT PROCESS

- **ITERATE!** – Go back to the beginning

AGILE PRACTICES AND TOOLS

A *very* incomplete list…

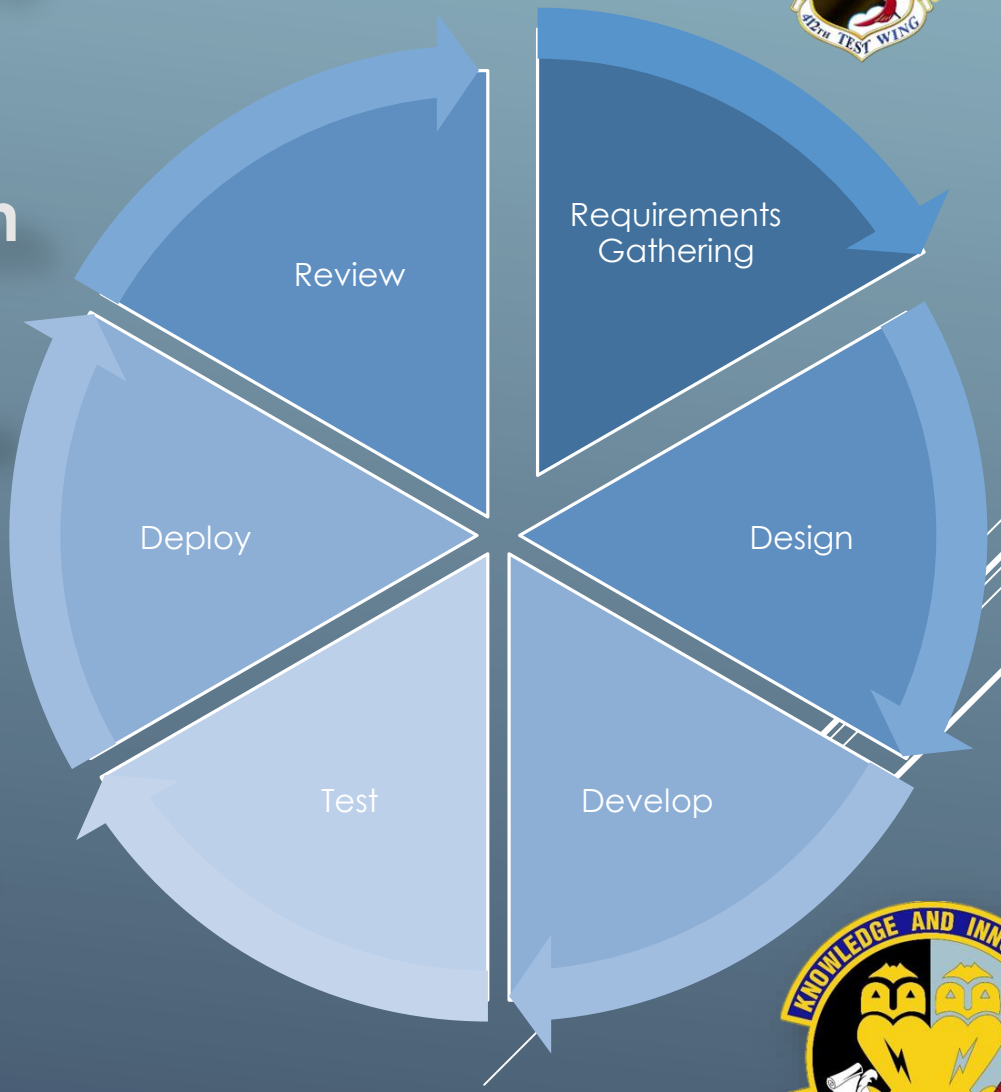
- Sprints – phases of product delivery, based on defined timeframes.
- GitOps – Using Git as the source of truth for deployments
- Clustering software like Kubernetes (k8s)
- Immutable Systems and Infrastructure as Code
- Containerization – Making software that runs in regularized environments.

AGILE PRACTICES AND TOOLS

- Sprints – phases of product delivery, based on defined timeframes.
  - Adapt more quickly to a changing software landscape
  - Integrate new requirements more quickly
  - Reduces complex, difficult deployments and migrations

AGILE PRACTICES AND TOOLS

- GitOps – Using Git as the source of truth for deployments
  - Systems are described in desired state
  - Use of tools like Flux to automate deployment through pull requests
  - Operation on standalone networks requires a git repository on the network, and perhaps Gitlab or other development platform



AGILE PRACTICES AND TOOLS

- Clustering software like Kubernetes (k8s)
  - Building tools for k8s helps to standardize them
  - Software can move between private clusters, public clouds, hybrid clouds more easily.
  - Choice of monitoring tools and other cluster tools

AGILE PRACTICES AND TOOLS

- Immutable Systems and Infrastructure as Code
  - Servers are not modified after they are deployed.
  - Servers are described in code, and are fully part of the release cycle
  - All modifications are made in code, and rolled out by deploying
  - Containerization – Making software that runs in regularized environments.

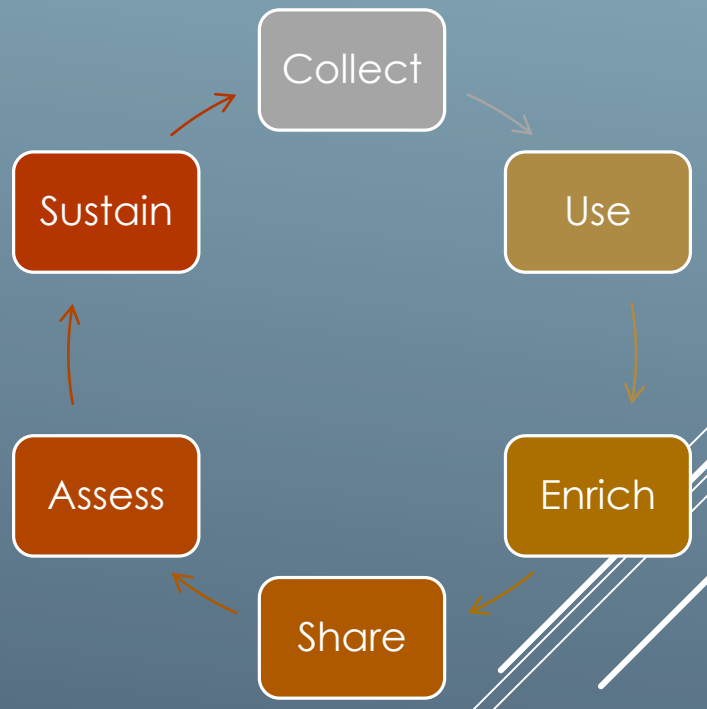
AGILE PRACTICES AND TOOLS

- Containerization – Making software that runs in regularized environments.
  - Uses tools such as Docker to run near-standalone stripped down OS images without full virtualization.
  - Allows services to run inside clearly controlled and specified environments.
  - Runs on top of most operating systems as well as clusters.

WHAT IS KNOWLEDGE MANAGEMENT?

Knowledge Management is a cycle. It's a continuous process that has to be sustained. There's no single definition of that the cycle that everyone uses, but generally it follows a few steps.

- Collect – gather data that is of interest.
  - Aircraft
  - Range Sensors
  - Test Project Docs
- Use – Assess the knowledge and adapt to be used
  - Engineering unit conversion
  - Other derived data
- Enrich – Replace or extend existing data
  - Augment open air range test data with data from other sources
  - Longitudinal Data Analyses
- Share – Make the knowledge collected available to other interested parties
  - Define Standard Data Access methods where possible
    - File Formats
    - Data Schemas
    - Common Access API
    - File systems and object stores (data lakes)
  - Select data to push to systems where it's most useful
- Assess – determine future needs and how to improve the resources available
  - Able to compare data from current and previous tests to support regression testing and analyses
  - Able to identify from the historical data all times when an event occurred
    - e.g., Determine all events where F-22 exceeding 3.5 g's
- Sustain – Keep the information available and determine how to maintain its relevance

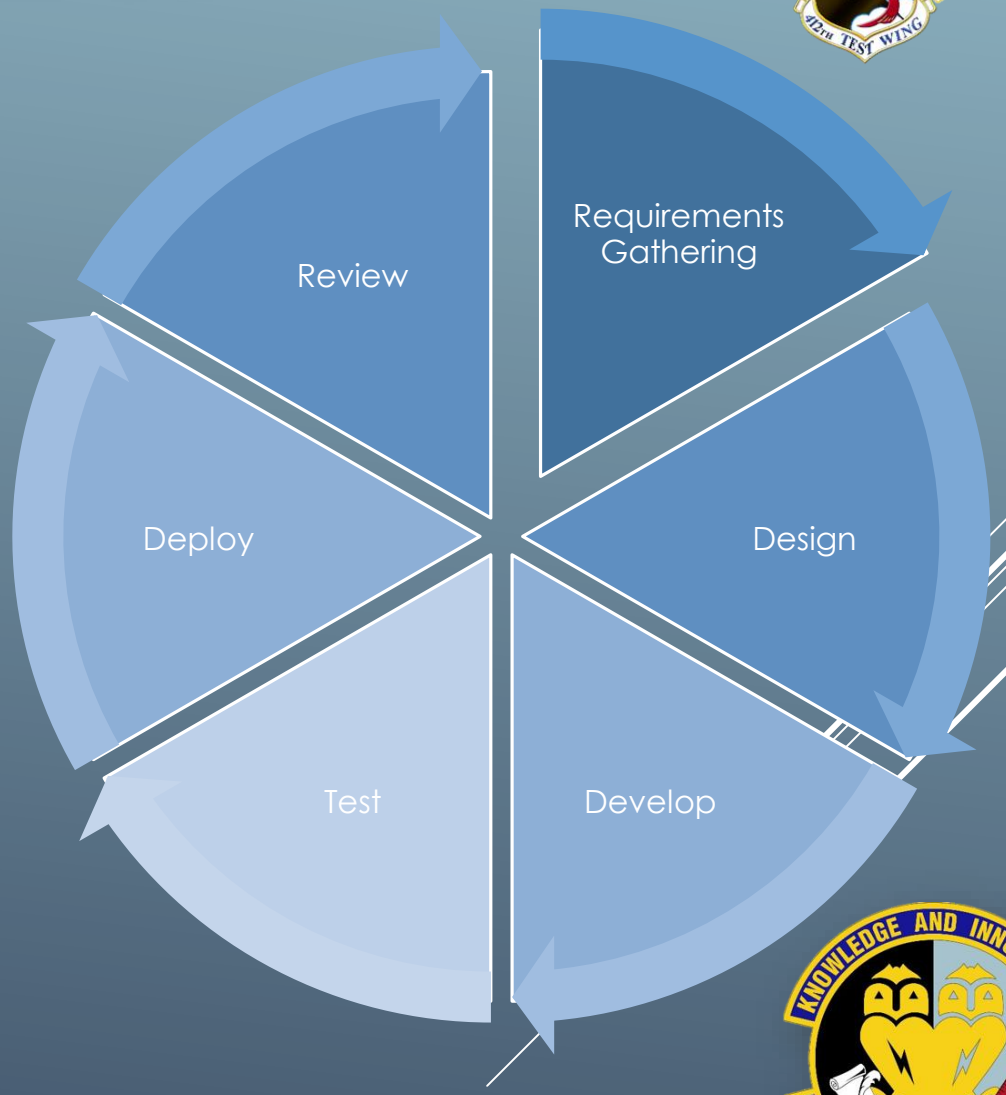


WHY USE AGILE FOR KNOWLEDGE MANAGEMENT?

- Many customers will struggle to define all **requirements** at the outset.
  - **Agile** allows rapid iteration and refinement
  - Progress does not stall just because there are unknown requirements

WHY USE AGILE FOR KNOWLEDGE MANAGEMENT?

- Smaller, incremental changes are easier to **review**, **approve**, and **deploy**



WHY USE AGILE FOR KNOWLEDGE MANAGEMENT?

- Knowledge Management changes quickly
  - KM is a large and fast-changing discipline, and different people will prefer different tools
  - What's popular today will be deprecated tomorrow
  - If you can't adapt quickly, you will always be using outdated tools and preventing cutting edge work
  - Software needs will never fully converge
    - Hadoop is largely being replaced by Apache Spark
    - some users will prefer Dask over Spark
    - all sorts of different databases with different use cases exist

CROSS PLATFORM DATA CENTER

Questions?

eric.stith.ctr@us.af.mil