

Finding the Right Data in an Ocean of Bits

Presented by Glyn Bowden, HPE

Speaker Bio

Glyn is CTO for the AI & Data Practice for HPE Pointnext & GreenLake Advisory & Professional Services group. His technical experience has spanned many industries from global finance, national security and high technology. With a background in high performance compute & storage, cloud native computing and emerging technologies such as blockchain and machine learning, Glyn's goal is to make high technology solutions accessible to all.



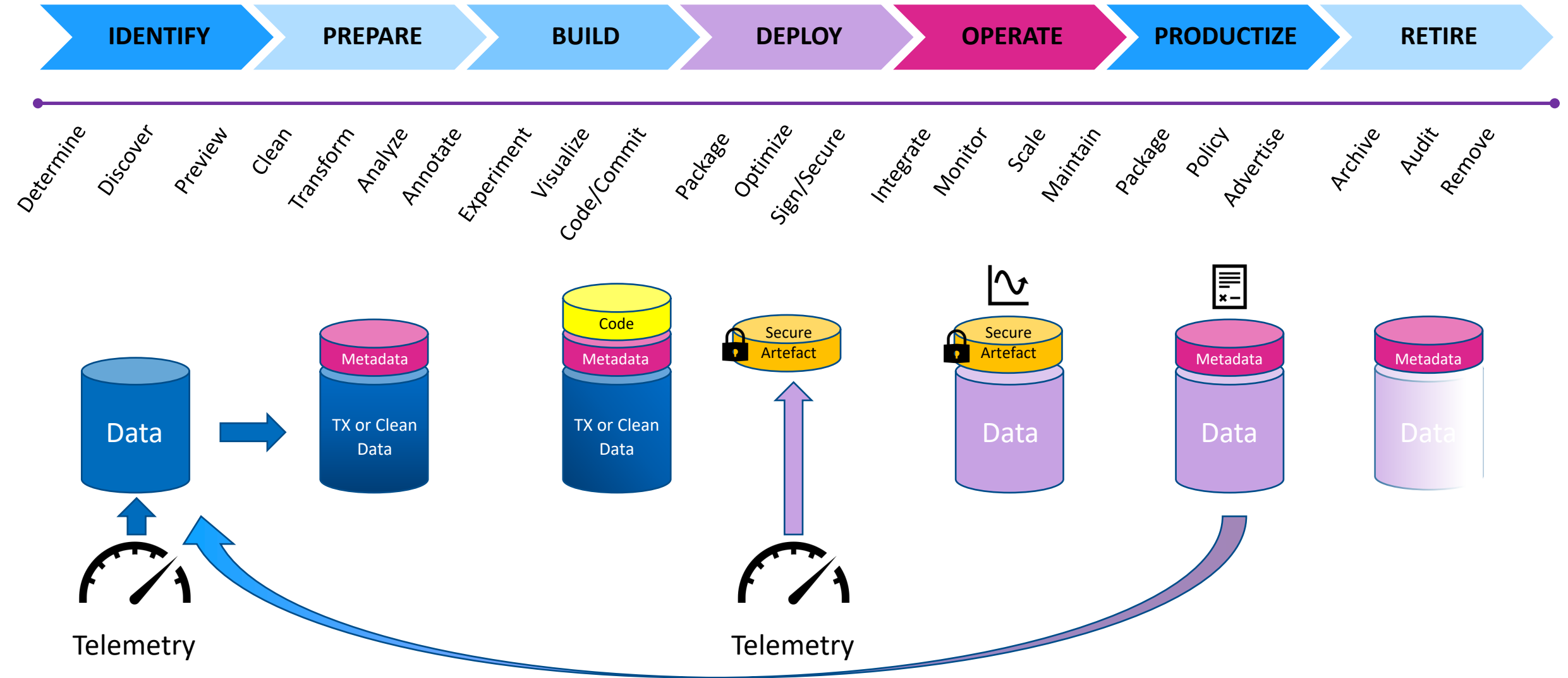
Questions about today's topic? Reach me in chat / Q&A Live (yes I'm right here!) or drop me a note using chat or profile within the conference, or my email address will be at the end.

(spoiler, it's gjb@hpe.com)

The Data Workflow

Where it comes from, and where it goes

The Data (Scientist) Workflow



Data begat Data

- Data production is often cyclical
- By analyzing data, you produce new data (insights)
 - Value add of data
 - Insights can be shared without sharing the data
- Very quickly can become overwhelming!
 - what created what?
 - who can see what?
 - why did this particular inference get made?

Data Provenance

- Like fine art, data has a provenance
 - A creator / artist (might be a sensor or camera for example)
 - A creation process / path. Source of canvas & paint, what process was used
 - The final artefact, Sign your work!
- Also, like fine art, data can inspire new data
 - The artefact can be used to create new data
 - The inspiration is important, must be recorded

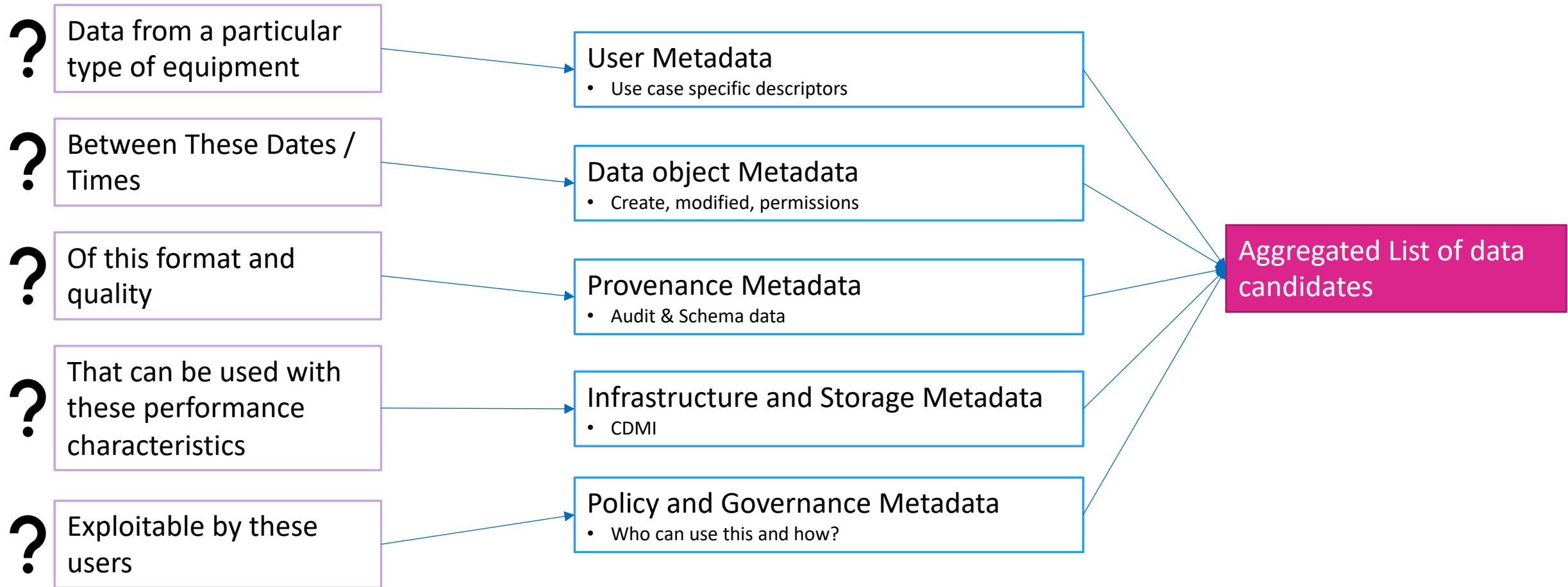
How to track provenance?

- Metadata! Data about data
- Metadata is multi-level...
 - Infrastructure and Storage metadata (CDMI)
 - Data object metadata (create, modified, permissions)
 - User metadata (use case specific descriptors)
 - *NEW(ish)* Provenance Metadata (multi-object spanning audit data with schema and create process)
 - Policy & Governance Metadata (who can use the data or artefact and how)

Back to the question... Finding the Right Data

- We know how the ocean was created
- We've started labelling and tagging our “fish” with metadata
- Create filters and searches on that metadata to select the appropriate data
- Multi-layered Search!

Multi-layered search with metadata



Summary

- Processing data, creates more data
 - Processing data, creates more data
 - Processing data, creates more data
 - ...
- Metadata describes the data...
 - and the process...
 - and the policy...
 - and the...
- Finding the right data relies on multi-layered queries.
 - And to find it again, make sure to annotate the right metadata

Thank you!

Glyn Bowden, CTO AI & Data, HPE Pointnext & GreenLake

gjb@hpe.com

@bowdengl



Please take a moment to rate this session.

Your feedback is important to us.