



Working with Disparate Data Resources using Carta

Branden Kappes

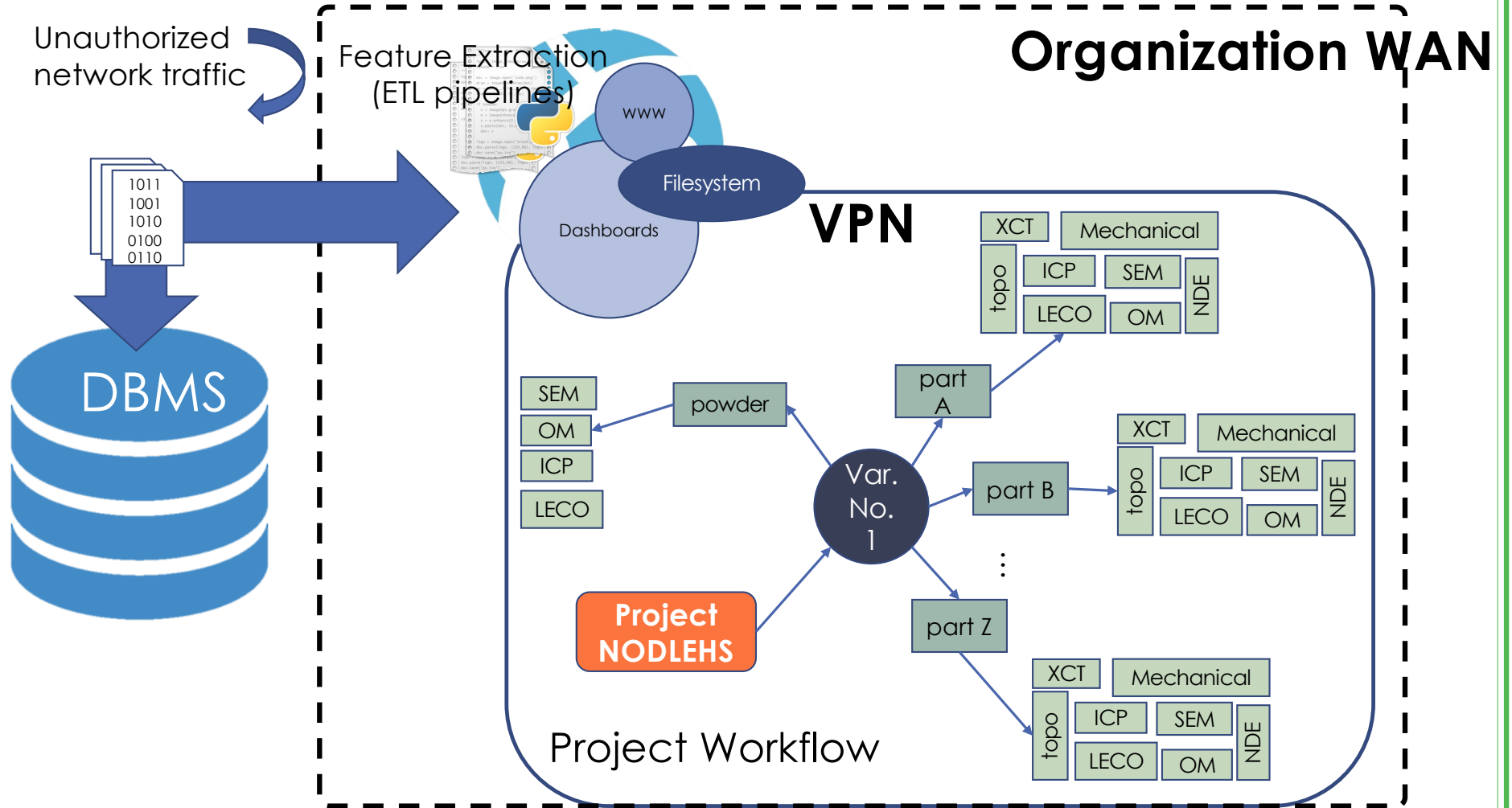
Contextualize, LLC

Castle Rock, CO 80108



Outline

- Generalizing a Materials Research Data Infrastructure
- Improve data consistency through a flexible but consistent data collection strategy.
- Mitigate data demand through registration.
- Work with individual groups to establish organizational POC; establish connections to existing data resources; and optimize data collection strategies.





Setting Up a Data Collection Template

Simplifying consistent data collection



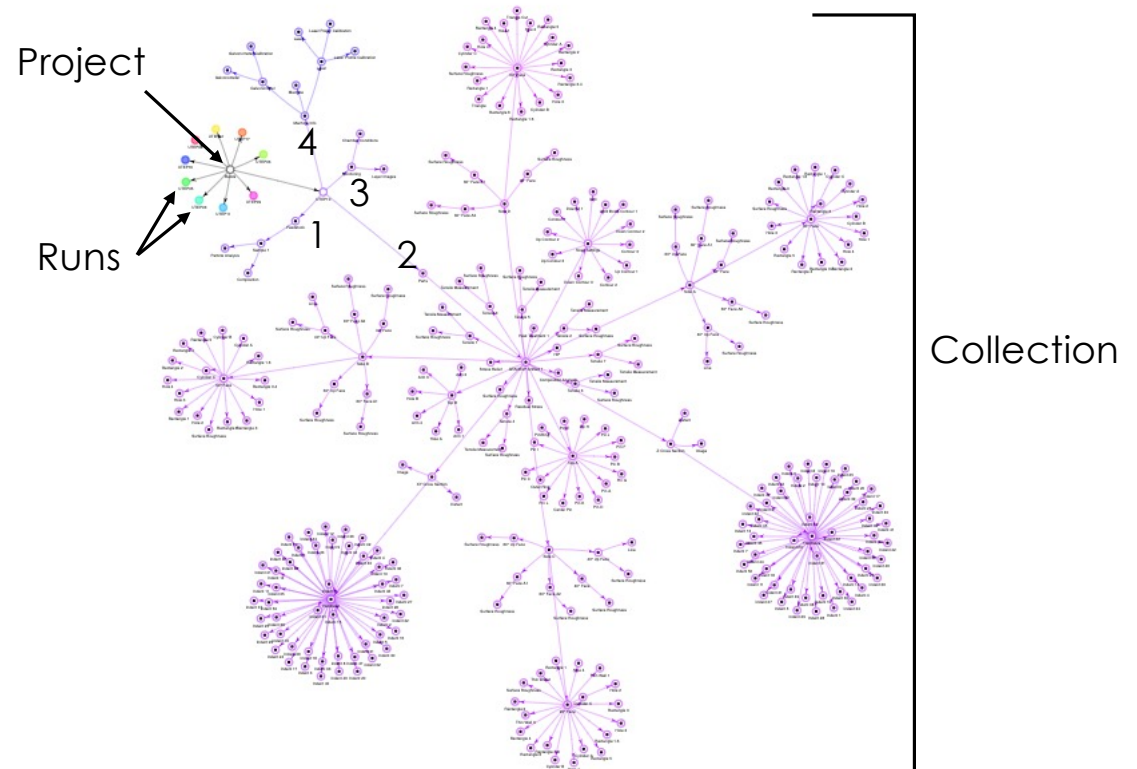
Outline and Problem Statement

- Each group has an established, data-intensive process.
 - Collection SOP is well established, at least conceptually if not practically.
- Data are collected incrementally by multiple specialists.
 - Maintaining data consistency is critical to streamline downstream efforts.
- Data processing and analysis are asynchronous, fluid, and multistep.
 - Unlike data collection, which is more-or-less rote, data featurization (processing/ETL) and analysis is exploratory.
 - Exploration and analysis *will not be performed by data scientists*.
 - Coordination between experts ranges from cumbersome to impossible.
- Modeling, while an eventual goal, is premature; predicated on development of successful ETL and analysis pipelines.



Ambitious Data Problems Demand Ambitious Data Collection

- The multi-year, investigatory project will collect a projected, but flexible number of runs.
- Each run has a N types of observations
- Each observation type requires one-or-more samples to carry out a characterization.
- Each sample requires one-or-more (typically, more) measurement modalities for acceptable characterization.





Automation Improves Consistency

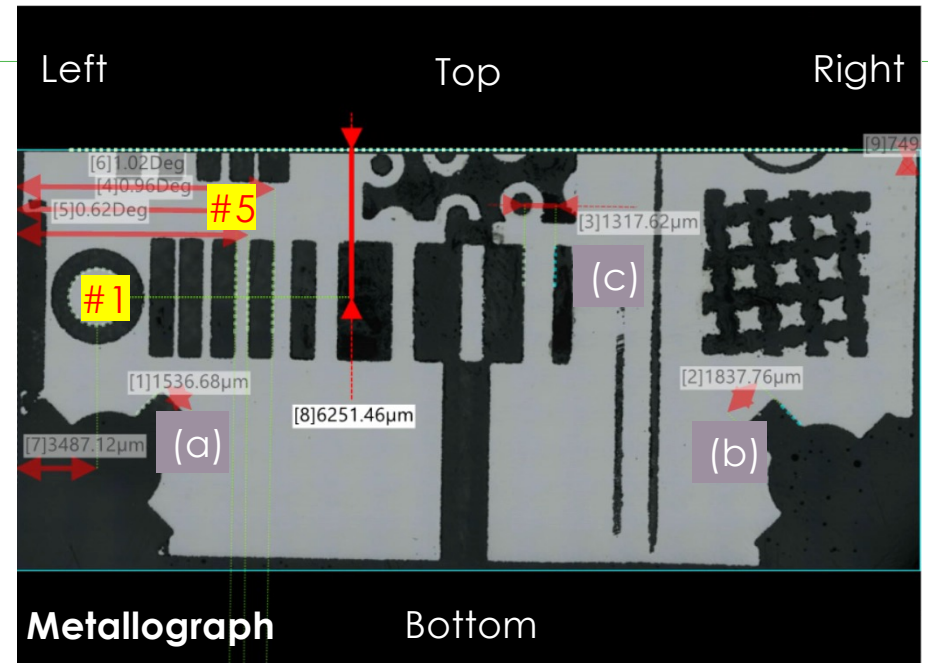
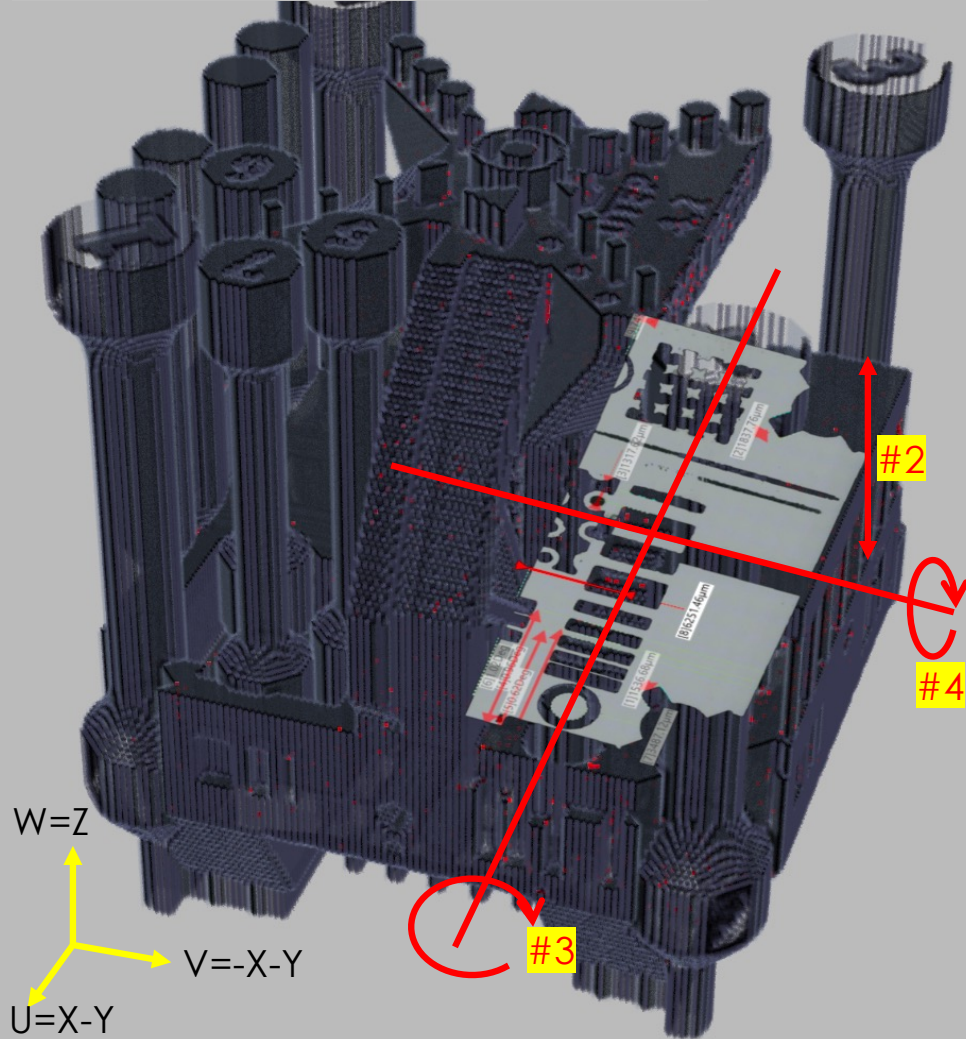
- With an error rate of about 1% for data entry, manual construction would introduce approximately 30 errors and omissions per construction and another 30 during data collection.
- Construction errors would likely compound if previous constructions are used to template manual construction of additional runs.
 - Example: The first construction introduces 15 mislabels and 15 omissions (1%). The second construction fixes 10 of the initial mislabels but introduces 15 more, leaving 20 mislabels; 15 additional omissions, now 30 total.



Improving Data Utilization through Data Registration

Enabled through structured data collection

Generated from LPBF laser vectors



#1. The center of this cylinder gives us the (X, Y) position relative to the GTA origin.

#2. What provides the elevation (Z coordinate). A. Measurements (a), (b), and (c) are 3 mm triangular fiducials

#3. What provides rotation about U? A. Measurements (a), (b), and (c) are 3 mm triangular fiducials

#4. What provides rotation about V? A. Measurements (a), (b), and (c) are 3 mm triangular fiducials

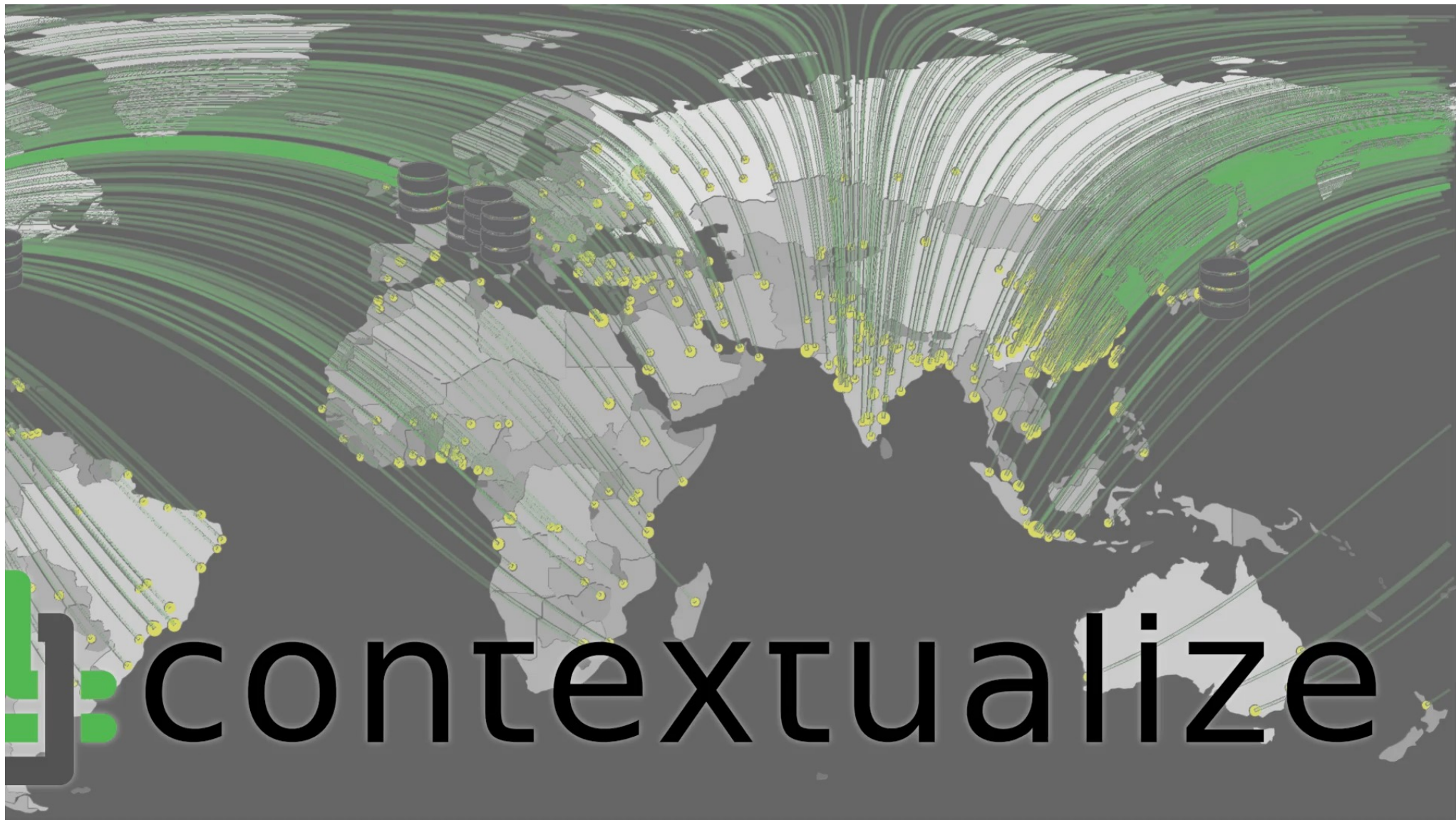
#5. These angles provide rotation about W.



Establishing a Cooperative Arrangement for Data Collection and Dissemination

Branden Kappes

branden.kappes@contextualize.us.com



 contextualize