**NIST Big Data Working Group (NBD-WD)**

**NBD-WD-2013/M0240**

**Source: <NBD-WG | Security and Privacy>**

**Status: <Proposal>**

**Title: [New Section] Roles and Methods for BG Security Assurance**

**Author: Mark Underwood, Krypton Brothers LLC**

***Abstract***

I suggested after applying the template to our use cases that perhaps we had not provided adequately for specific types of metadata that are particular to bigdata: (1) temporal (point in time); (2) transformational ("fusion" in DoD parlance). Anything suggested that would add to the existing diagrams (except the first one) would add complexity, and there's plenty of that in merging the ideas across the subgroups. Perhaps we could add a separate slide and label it "S&P Assurance Best Practices", and include a list of recommended practices that apply to big data -- rather than trying to work it into the existing diagram. E.g., maybe a checklist like this:

* Provide system auditors with forward and backward provenance traceability from Data Provider to Data Consumer
* For each Transformation Provider, identify lost provenance or degraded privacy. Further, build verifiability for provenance and privacy transformations with elective audit
* Data Providers should provide configuration metadata & temporal markings so that long timeframe analytics across devices and platforms can be supported. Metadata is also needed for Transformation Providers, since they must also be identified by configuration at the time of transformation
* Transformation Provider metadata should include point in time dimension constraints, such as sampling rate, aggregation algorithm, anonymization, and throttling (or, e.g., compression) required to accommodate velocity, volume or other dimension-specific measurements
* Provider configuration metadata identifies constraints associated with velocity, volume and other instrumentation-specific dimensions. E.g., sampling frequency for audio, image resolution, camera type, FFT, etc. Out-of-bounds values and other error mitigation algorithms could be provided as metadata rather than remediation.
* A Big Data system designed to include a temporal dimension along with configuration metadata will support more uses over a longer period of time.
* Transformation Provider metadata discloses its pre- and post-effect on privacy and provenance.
* Transformation Providers that perform fusion across multiple Data Providers, such as for enhanced situation awareness or complex event processing should provide adequate metadata for Data Consumers to query reliability and verifiability, as well as to reproduce TP behavior in standalone test environments.

Re the provenance issue, we are approach-neutral of course, but I found this discussion of [Spade](http://spade.csl.sri.com/SPADE/Papers/Papers.html) at SRI useful. I attached one of their notional diagrams below.

Ashish Gehani and Dawood Tariq, SPADE: Support for Provenance Auditing in Distributed Environments, 13th

ACM/IFIP/USENIX International Conference on Middleware, 2012

