**Volume 1 version 2 Topics needing enhancement – Ann and Nancy**

12/15/15

1. Revisit all **Data Science** Definitions (2.2)

Many edits were made at the last minute, and the definitions don’t flow. Probably needs a major re-write.

1. Revisit “**data lifecycle**” which should probably be “analytics lifecycle” (3.4 and references throughout)
2. Be more specific on **data-information-knowledge** definitions and usage (3.4 and references throughout)
3. **Governance** (3.8)
4. **Provenance**
5. **Management** (not there yet)
6. **Security and Privacy** (3.7) *– Mark*
7. **Value** (probably 2.2)

Note: (Societal) Impacts would be in a separate whitepaper

1. **Patterns** (4)

Bob has done some good work on IoT, we should look to pull into pattern descriptions

1. **Concurrency** (3.3.1 but perhaps should be in 4)
2. **Emergence** (new topic, probably 3.3.3)

There is emergent behavior in Big Data, the most obvious example is w.r.t. PII.

**Volume** – given enough data, I can begin to learn a lot about you (where you work and live given your cell phone position data, or Netflix KDD prize or recommender systems)

**Variety** – combining datasets that don’t have explicit PII, you can determine individuals (internet “PI” sites can give background information on individuals…where lived, who family is)

**Velocity** – real-time information

1. Any other topics we’ve missed