

smiths detection

Data Desensitization

Matthew Merzbacher
5 November 2019



So [redacted], [DeID] cares?

- “Data isn’t everything... it’s the only thing” – Red Sanders-Merzbacher
 - World is increasingly data-driven
 - Algorithm development relies on “real” threat scans
 - Scans of threats are sensitive, proprietary, expensive, and contain unneeded information
- [How] Can data be modified to allow broader (faster, greater) access?
 - **Step 0: Identify problem to be solved**
 - **Step 1: Determine what needs to be shared at a gross level**
 - **Step 2: Desensitize remainder (expert or automatic), by changing**
 - Scanner
 - Threats (in data space or in threat space), while retaining key characteristics
 - Problem
 - Remove what cannot be shared
 - **Step 3: Create data control mechanisms**
 - Pave the road to sharing, including faster permissions / vetting
- Learn from other fields, such as Medical De-Identification of PHI