









Unstructured, redundant and noisy data

Cleaned and documented data

Collaboratively curated and structured data