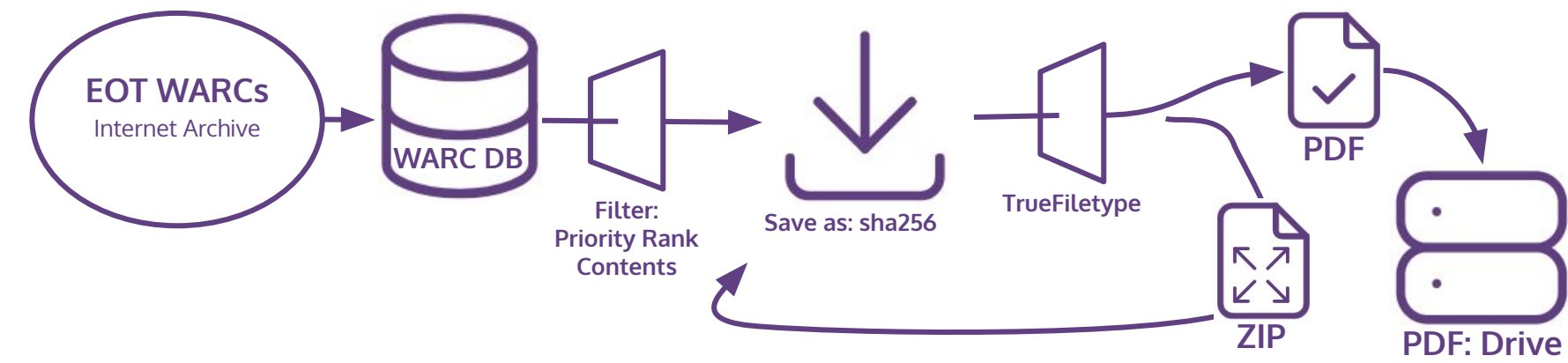
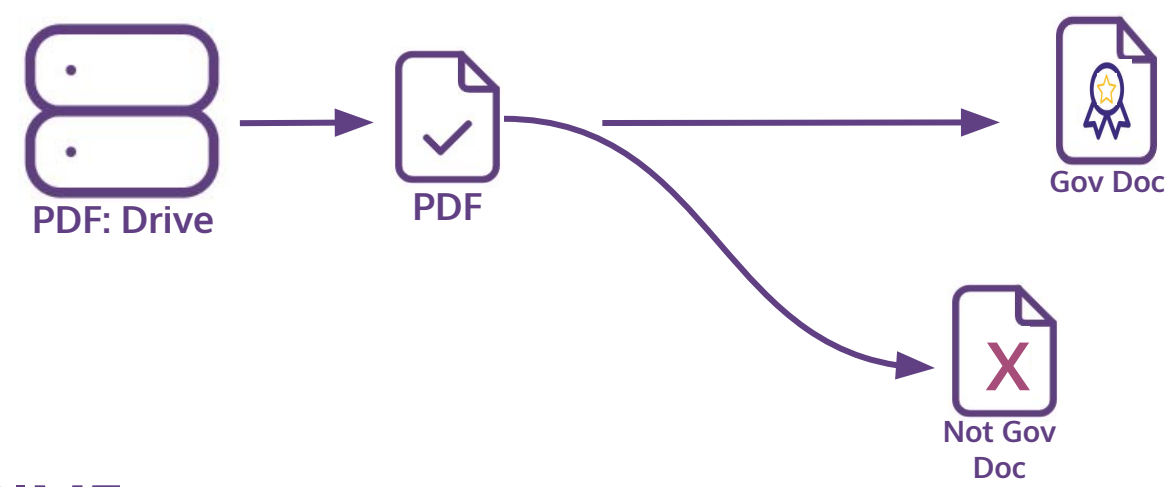


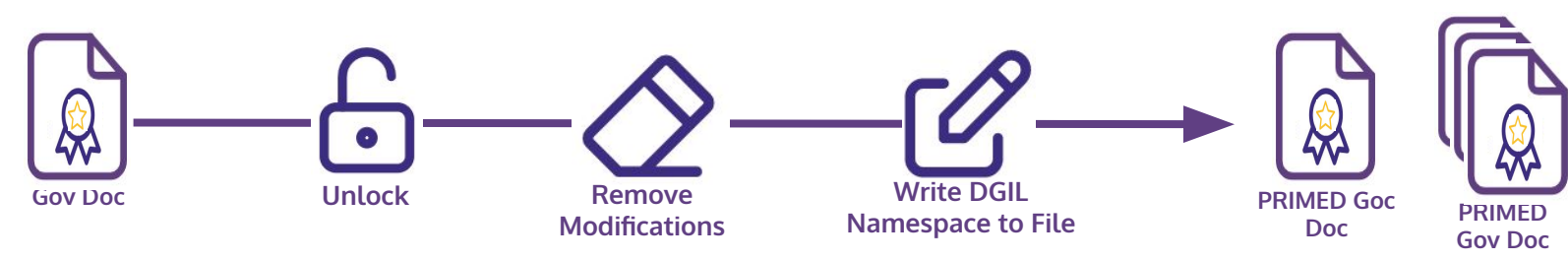
CRAWL



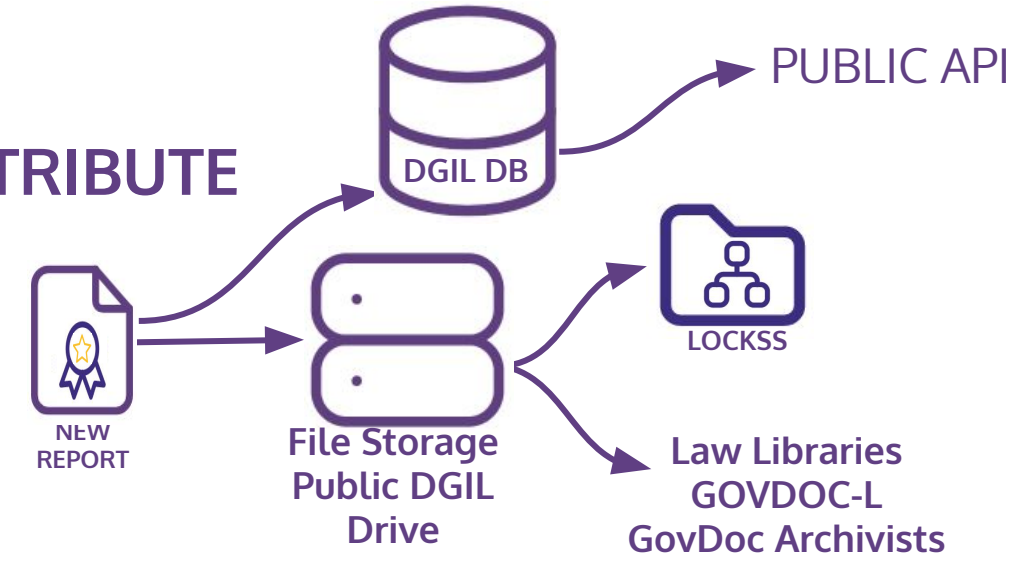
IDENTIFY



PRIME

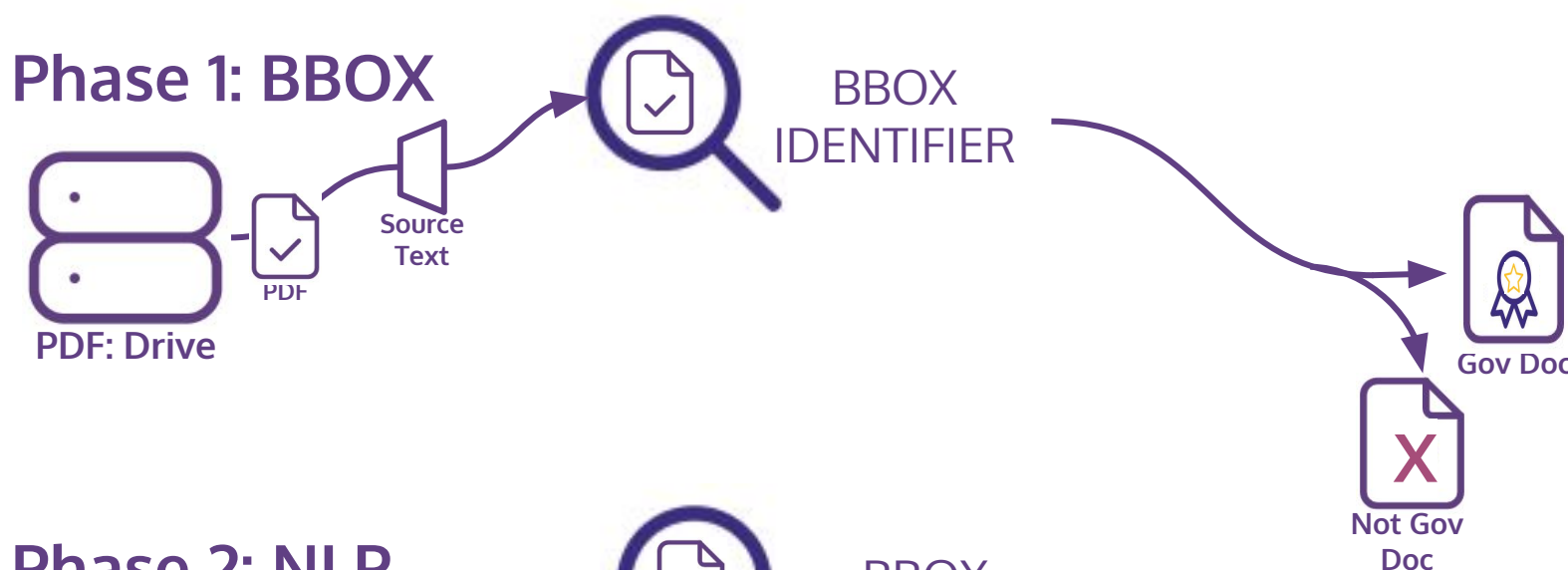


DISTRIBUTE

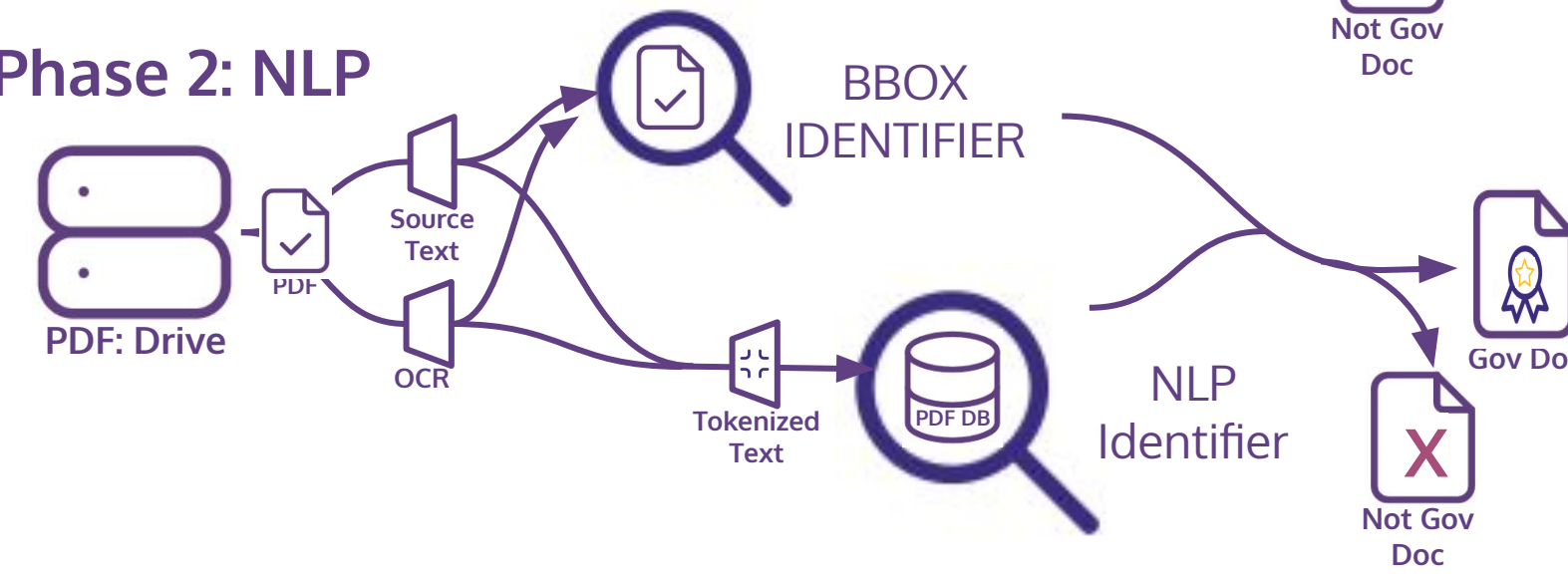


\*Notes 2021-09-11  
This DGIL map could be expanded to include multiple phase for each stage. In addition it needs to be expanded to include HTML identification. HTML will have greater emphasize on NLP, and statements of time of capture  
The definition of Gov Doc needs to be made. To start PEGI team has directed a focus on Environmental Gov Doc identification. Starting with top down US centric we have the Federal EPA, etc. then State agencies, the inclusion of government contracted documents NEPA and the CEQA likes should be identified with notes on authorship, likewise international equivalents should be identified.  
The negative "Not Gov Doc" results need to be reprocessed when new agency docs are added, likewise the image training classifier must evolve for greater inclusion.

IDENTIFY Phase 1: BBOX



IDENTIFY Phase 2: NLP



IDENTIFY Phase 3: Image Classifier

