Goal

Establish the policies that will govern your data collection policy. This should include (at minimum):

* What are the criteria that make a dataset, or collection, relevant to your protocol? What, if anything, might exclude a relevant dataset?
  + Scope: United States Civil Lawsuits involving government entities
    - Must US
    - Must be civil claim
    - Must involve government entity as a defendant
  + Court Documents within a Case Filing
    - Initial claim including docket sheet
    - Any additional claims
    - Resolution to case
      * Partial or full, including settlement amount if disclosed
* What will you require to be included for a minimum viable package of data? (e.g. the data must have a clear license, a format that is (or can be converted to) an open standard, and enough contextual documentation that structured metadata can be created).
  + Documents will be converted into PDF format, Machine Readable, Text encoded.
  + Documents must have been requested and received from HRDC from a public records department
* What formats will you accept, and what formats will not you accept?
  + Any format that is convertible into PDF
  + ALL documents MUST be PDF before being tagged and include.
* How big can a collection be (in size)?
  + No larger than 100MB per document
* What kind of preservation guarantees will you make for each format?
  + Fixity