

# Design of the indexer

- [Overview](#)
- [Flow diagram](#)

## Overview

A command-line application that can be gracefully interrupted and restarted at any time, with little or no loss of progress.

A full rewrite from the Beta release

- Moved from the Scala Actors framework to a Hadoop-compatible threaded model.

Highly modular, and configurable.

- Open to contributions from the community.
- Configuration file that determines at startup which modules are used.
- Rule-based configuration using the Digester component from Apache Commons

## Flow diagram

- Configuration
- Evaluation, Assessment
  - Scheduling
    - Determine what discovery is to be done
  - Discovery
    - Visit the client sites to build lists of URIs of Individuals for the index
    - Will the client site give us the last modified date for the individual?
  - Synchronization, update
    - Record the results of discovery in the search index
    - Remove any URIs which are no longer viable
    - Add any new URIs.
    - Record that discovery was done.
- Population, Retrieval, scan, enactment, evaluation, fulfillment,
  - Ranking, Prioritization
    - Inspect the index to see which records should be updated.
    - Build a to-do list.
  - Assembly - for each URI in the do-do list:
    - Modeling
      - Use LOD requests to build the model for the URI
    - Indexing
      - Build an index record from the model, and write to the search index.