# Harvester Manual

## Index Data

## Overview

This document describes the Harvester using the following notation:

- Resources to harvest and on what schedule
- Harvester Job, the actual process of harvesting a resource at a given time, a term used in the document
- Storages (or Local Unified Indexes)
- Transformation Steps, which transform data from one format to another or do transformations on the same format
- Transformation Pipelines, which consist of multiple Transformation Steps

A Harvester Job fetches data via http from a remote Resource, does some initial extraction/conversion of records depending on the protocol and configuration, splits the data into minor chunks (necessary on larger datasets due to memory constraints), sends this data through a Transformation Pipeline, and then sends the data to a Storage (Local Unified Index) for indexing.

The indexed resources are exposed in a MasterKey 2 Toroid to be used as targets for a MasterKey 2 meta search solution.