

# Snapshot Harvests

- Creating/editing a snapshot harvest



## Menu

Definitions

- Selective Harvests
- Snapshot Harvests
- Schedules
- Find Domain(s)
- Create Domain
- Domain Statistics
- Alias Summary
- Edit Harvest Templates
- Global Crawler Traps
- Extended Fields

Harvest status

- Harvest Channels
- Bitpreservation
- Quality Assurance
- Systemstate

## Snapshot Harvests

Harvest definition	Max objects	Max bytes	Status	Commands
snapshot_6oZ	-1	1,000,000	Active	Edit History

[Create new snapshot harvest definition](#)

On the **Snapshot Harvests** page new snapshot harvests are started, harvesting all domains known to the system in their default configurations. An overview of all snapshot harvests is also provided.

The link **Create new harvestdefinition** opens the template below.

## Creating/editing a snapshot harvest



## Menu

Definitions

- Selective Harvests
- Snapshot Harvests
- Schedules
- Find Domain(s)
- Create Domain
- Domain Statistics
- Alias Summary
- Edit Harvest Templates
- Global Crawler Traps
- Extended Fields

Harvest status

- Harvest Channels
- Bitpreservation
- Quality Assurance
- Systemstate

## Snapshot Harvest

Harvest name:

Max number of objects per domain:

Max number of bytes per domain:

Max number of seconds for each job:

Comments:

**Harvest only domains that were not completely harvest in a previous harvest:**

None



This page is used to define name and size (max. bytes per domain and/or max objects per domain ) of the harvest. The default object limit for harvests if using object limits rather than bytelimits is -1 (-1 here meaning unlimited).

It is recommended to systematize the naming for clarity, e.g. 2007-1, 2007-2 etc.

The size of the harvest can be defined in two ways: at the harvest definition **Snapshot Harvests** or at the configuration of the single domain. It will always be the lower size limit stopping the harvesting of a domain.

Comments can be added freely.

Snapshot harvests can be based on previous snapshots in the sense that it can be limited to only harvest domains that hit the max number of bytes limit in a previous harvest.

The domains completely finished (not hitting the max number of bytes limit – either on the configuration level or on the snapshot harvest level) in the first harvest will not be included in the second. Domains included in harvests which were aborted through the Heritrix GUI or otherwise stopped unclearly (for example by a crash of a harvester machine) will also not be included.

All other domains will be harvested from the beginning in the second harvest.

**Save** saves the harvest definition and returns to **Snapshot harvests**.

After defining a snapshot harvest the harvest is activated with the **Activate** button on the snapshot frontpage. Harvest will not start until you press **Activate**. Status then changes to 'Active'.

**Deactivate** is not relevant in Snapshot Harvests because they only run once. By **Edit** the Snapshot Definition can be changed but only before activation. Parameters changeable are size, commentary and if previous harvest start point should be used.

**History** provides an overview of the specific harvest: see [Harvest Status](#)

