

System Design

This is a document describing the design of the NetarchiveSuite software

This document only describes the underlying design of the NetarchiveSuite software, i.e. it does not describe how to install, run, or use NetarchiveSuite, for that see the [Installation Manual](#) and the [User Manual](#).

The first section gives an overview, and the remainder of the document gives more details about the design.

The code is available through the [downloaded site](#) or from our [Subversion repository](#) .

Contents

- [Overall Systems Design](#)
- [Settings](#)
- [Localization](#)
- [JMS Channels](#)
- [JSP](#)
- [Pluggable parts](#)
- [XML handling by Deploy](#)
- [Archive Design](#) — The Archive Design Description contains the description of overview of how the archive works, describes Indexing and caching and describes CrawlLogIndexCache.
- [Harvester design](#)
- [ViewerProxy Design](#) — This section describes the viewerproxy control resolver, the special viewerproxy access via urls and the observer resolver.

Audience

The reader is expected to be familiar with Java programming and have an understanding of the core issues involved in large-scale web harvesting. Having used Heritrix before is a definite plus, and an elementary understanding of SQL databases is required for some parts.

Search manual

[Download as pdf](#)

