



netarchive.dk

# Plug-ins

---

Lars Clausen  
Netarkivet

# Rationale

---

- Expandability, of course
- Modularity
- Flexibility
- Adaptability

# ArcRepositoryClient plug-in

---

- ❑ Front-end for entire archive (2 impls)
- ❑ `get(arcFile, index)`
- ❑ `getFile(fileName, location, toFile)`
- ❑ `store(file)`
- ❑ `batch(fileBatchJob, location)`
- ❑ `removeAndGetFile(file, location, checksum, credentials)`
- ❑ `updateAdminData(file, location, state)`
- ❑ `updateAdminChecksum(file, location, checksum)`

# JobIndexCache plug-in

---

- ❑ Serves indexes (two implementations)
- ❑ File `getIndex(Set<Long> jobIDs)`
- ❑ `<init>(RequestType)`
- ❑ Skips unknown jobs
- ❑ Generates sorted indexes
- ❑ IndexClientFactory makes three kinds:
  - CDX, DedupCrawlLog, FullCrawlLog

# DBSpecifics plug-in

---

- Handles different databases (3 impls)
- Detailed in database section

# RemoteFile plug-in

---

- ❑ Handles moving large files (2.5 impls)
- ❑ `<init>(file, useChecksums, fileDeletable, multipleDownloads);`
- ❑ `InputStream getInputStream();`
- ❑ `String getChecksum();`
- ❑ `void cleanup();`
- ❑ HTTP version is nicest, but get-only
- ❑ Handling local moves, 0-byte files

# Notifications plug-in

---

- How to alert for serious errors (1 impl)
- Decouples logging and alerts
- `errorEvent(String message, Throwable e);`

# SiteSection plug-in

---

- Has multiple implementations
- Constructor calls superclass with:
  - Section title (i18n)
  - Section name
  - # pages visible in menu
  - Array of page suffix + title (i18N)
  - Directory name
  - Translation bundle



# Plug-in architecture

---

- ❑ Plug-in initialized via SettingsFactory
- ❑ Either getInstance or constructor called
- ❑ Arguments possible
- ❑ A settings group per plug-in
- ❑ Specifies the class and other arguments
- ❑ All settings strings hardcoded in Settings:(

# Likely future plug-ins

---

- Harvester – encapsulate Heritrix better
- Job splitter – different job splittings
- Metadata wrapper – what to put where
- Domains – encapsulate domain concept

# Improvements in plug-in architecture

---

- Better settings approach
  - Class as an attribute
  - Arguments decoupled from Settings.java
  - Validity checking for parts of settings.xml
- More than one class per plug-in
  - Batch job reduce methods?
  - Method of selecting?