

Archimedes Bundle Structure

Mark J. Schiefsky, Malcolm D. Hyman

October 9, 2003

1 Purpose of the Archimedes Bundle Conventions

An Archimedes bundle is a collection of one or more source texts and related objects from the Archimedes Project Repository, arranged according to the conventions specified in this document. Grouping repository objects into bundles provides a convenient, meaningful association of objects for upload to and download from the repository. In addition, Arboreal implements certain default behaviors that simplify working with primary texts in bundles:

1. The bundle structure enables the Archimedes DTD and essential related files to be stored separately from the primary text files and to be referenced automatically by Arboreal and any XML-compliant software.
2. When a primary text in an Archimedes bundle is loaded, Arboreal automatically loads the main catalog file in the bundle.
3. Morphology files in an Archimedes bundle are automatically loaded as needed.
4. The bundle structure allows Arboreal to find large page images and thumbnail figures automatically.

2 Components of an Archimedes Bundle

Each Archimedes bundle must contain the following components:

- **The file `catalog.xml`**

The file `catalog.xml` is the main catalog file for the entire bundle. The structure of a catalog file is described in 'The Arboreal Catalog System'.

- **The folder `dtd`**

The folder `dtd` contains the Archimedes DTD and related files. The following files are currently required in the `dtd` folder:

```
archimedes.dtd
archimedes.pen
ISOLat1.pen
ISOLat2.pen
```

- **The folder `texts`**

The folder `texts` contains primary XML source texts. These texts **must** be placed in the top level of this folder.

- **The folder `images`**

The folder `images` may contain any number of subfolders, each of which contains a set of page images or figures. Subfolders are named according to the following conventions:

1. The name of a folder containing a series of page images is formed by concatenating the ID portion of the locator of those images with the string `pageimg`. The separator character is a hyphen. For example, the folder containing page image series 01 for text 007.xml will have the name `007-01-pageimg`.
2. The name of a folder containing a series of figures is formed by concatenating the ID portion of the locator of those images with the string `pageimg`. The separator character is a hyphen. For example, the folder containing the figures from page image series 01 for text 007.xml will have the name `007-01-figures`.

In addition, a page image folder may contain a subfolder `large`, which contains large page images. A figure folder may contain a subfolder `thumbs`, which contains thumbnail-sized figures.

- **The folder `morphology`**

The folder `morphology` contains morphology files generated by the Donatus system. These files have the canonical names assigned by Donatus. For example, the morphology files for the text `monte_mecha_01_1a_1577.xml` will be:

```
- monte_mecha_01_1a_1577.morph.xml
```

- monte_mecha_01_la_1577.unparsed.xml

- **The folder matching**

The folder matching contains matching files that relate the primary texts in the bundle. Matching files are named according to the conventions specified in the document 'Archimedes Project Repository'. For example, the file matching texts 049.xml, 056.xml, and 063.xml is named 049-056-063.xml.

3 Creation of Arboreal Bundles

A skeleton bundle may be downloaded from the Archimedes Project Server at Harvard (<http://archimedes.fas.harvard.edu/bundle.tar.gz>). The skeleton bundle contains the dtd folder and other (empty) folders required in an Archimedes bundle.

Morphology files can be generated using Donatus (<http://archimedes.fas.harvard.edu/cgi-bin/donatus>). Starting with version 4.0, Arboreal offers a built-in interface for Donatus.

Arboreal currently offers limited support for the generation of catalog files. A simple program called `catalog.command` is also available for OS X; it may be used to generate catalog files for all texts and image files in a directory. (The program is available on the server **Projekt I/1 Archimedes**.) For further information see the related document, 'The Arboreal Catalog System'.

4 Acknowledgments

Thanks to Dan Lassiter, our man in Berlin, for pointing out snafus in an earlier version of this document.