Hoppla 2.0

User Guide Version 0.5

Contents

1 Introduction

We are proud to announce that the first public version of the Hoppla - Home Office Painless Persistent Archiving tool is now available. This document will give you an overview how to use Hoppla and its main features. Your feedback will be very much appreciated! Please note that the software is a research prototype and the use of Hoppla is undertaken at your own risk!

More information about Hoppla can be found at:

```
http://www.ifs.tuwien.ac.at/dp/hoppla
```

2 Requirements

This section gives an overview of the minimal requirements of Hoppla and the external software needed for optimal use and experience. If you are new to Hoppla, please take a brief look into it.

2.1 Minimum Requirements

Hoppla is platform independent and can be run on Windows, Linux and MacOS platforms. The only strong requirement that your system has to meet is an installed JRE (Java Runtime Environment)

- JRE 1.6 - Java Runtime Environment `http://java.sun.com/javase/downloads/index.jsp`

- 1GB RAM

2.2 Recommended

We recommend to have at least 1.5 GB RAM if you plan to preserve larger collections and the following tools installed in order to enable the logical preservation features of Hoppla.

- JRE 1.6_u17 or above

- 1.5 GB RAM or above

- Open Office 3 or above `http://download.openoffice.org/`

- Imagemagick `http://www.imagemagick.org/script/index.php`

- SoX - Sound eXchange `http://sourceforge.net/projects/sox/`

- FLAC `http://flac.sourceforge.net/download.html`
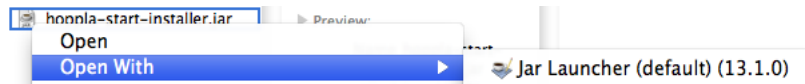
- MEncoder

- ps2pdf

Figure 1: start the installation via mouse click

3 Installation

In order to install Hoppla on your system you have to download the installation jar *hoppla-start-installer.jar*, which can be found at `http://www.ifs.tuwien.ac.at/dp/hoppla/release`. To start the installation process you have to run the jar either by clicking with the right mouse button as in figure 1 and select open with... Sun 6 JRE, or from the command line as in figure 2.

Under windows you have to open the command prompt and under Unix the terminal. Navigate to the folder in which you downloaded the installation jar and run *java -jar hoppla-start-installer.jar*

Figure 2: start the installation via the command line

After that you have to follow some easy steps in order to install the program. When you are finished you are set to run the program. To do so navigate to the installation folder, chosen in the previous step and run the application with one of the start scripts.

If you are under windows you have to double click on the hoppla.bat file and if you are on a unix platform you have to execute the hoppla.sh shell script as in figure 3.

Figure 3: start Hoppla via the terminal

4   Usage

4.1   Setting up a project

When you first start Hoppla you will see the main window (as shown in Figure 8). Clicking on *File > New Project* will start a new project dialog where you can set your first Hoppla project in a few easy steps.

First of all you can choose a name and optionally a description for your project, e.g. 'MyDocumentsProject' and 'This project contains all my documents' (as shown in Figure 4).
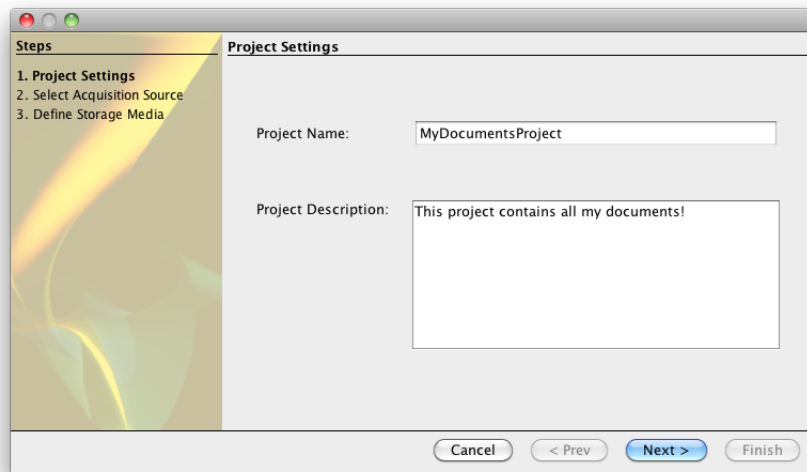

Figure 4: setup a project - set name and description

In the next step you can add one or more sources which Hoppla will manage. A source is a location on your local hard drive or on a remote server which contains important data. This could include e.g. your vacation photos, your work documents or any other digital file collection that is important to you. Of course you can have one source that includes all these types of files, however we recommend that you split your important files in different folders, i.e. categories, as Hoppla will allow you to manage them easily and it will ease future retrieval and access (but more on that later).

As already mentioned, you can choose between source folders on your local hard drive, a mailbox and a remote SSH enabled machine. This is easily done by selecting the plugin type of your source. You have to specify a name of the source e.g 'MyDocuments' or 'e-mails work'. If you feel that some of the files in the source folder are not so important but you don't want to physically move them, you can exclude some sub folders, that Hoppla shall skip during the acquisition process and thus save some extra storage space. You can optionally

add customized filters to enhance the preservation process furtherly. After you are ready the setup window should look similar to the one in figure 5.
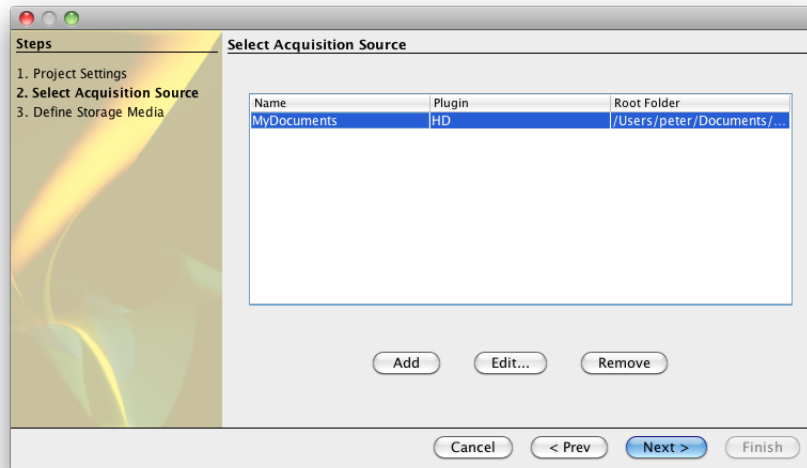

Figure 5: setup a project - choose a source

In the third step you have to specify the output media for Hoppla. These are the media where Hoppla will store all backups and migrations of your own collection. All storage media folders shall be initially empty. For now Hoppla supports (external) hard drives and SSH remote servers (*Please refer to 6* to see what other types will be supported in the future versions). In this project we will choose a local folder on the same system. Give a name of the medium and click finish. The window should look similar to figure 6. Of course it is recommended to choose more than one different storage media, e.g. a local hard drive and an external hard drive, as this will decrease the chance of data loss.

In the last step 'Preservation Level' you will see a window with 5 sliders with scales from 0 to 1 as in figure 7. One for each of the following 5 categories: *Images*, *Audio*, *Video*, *Text* and *Other*. With these sliders you can specify the importance of logical preservation for these media types. The rating ranges from zero (no preservation actions) to one (all available actions) E.g. if you have a project for your important work documents, you will obviously want to select *Text* with a high rating.

For a low rating (0,1 - 0,4) only essential migration for objects that are at imminent risk of becoming obsolete are performed. For a medium rating (0,4 - 0,7) migrations are done for formats that are no longer in wide-spread use, outmoded (e.g. old versions of an application) or the migration of proprietary format to open source formats. High ratings (0,7 - 1,0) initiate additional multiple migration pathways for object formats to support different future usage
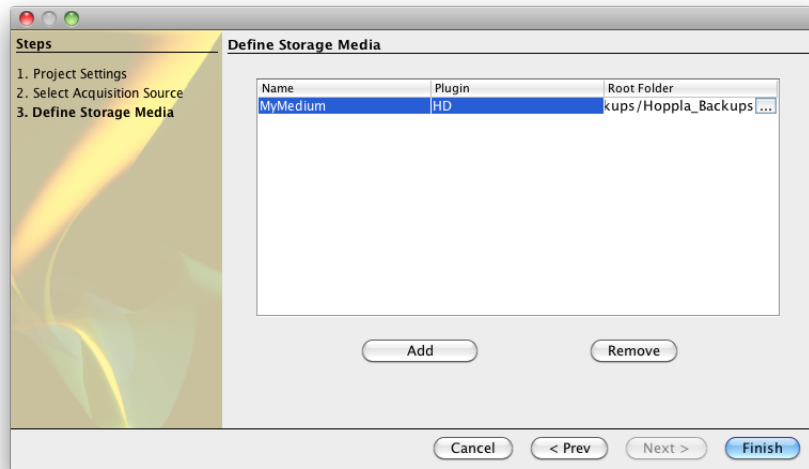
Figure 6: setup a project - choose a storage medium and complete the setup

scenarios e.g. migration of Microsoft Word objects to PDF/A, OpenOffice Format(ODT) and plain text format (preserving the layout, the editability and allowing computer-assisted processing).

Be careful when you select the preservation levels as they will have an effect on the migrations Hoppla will recommend for your collection.

Now you are all set up to preserve your data. Just click finish and you are ready to go.

4.2 Workflows

Hoppla provides you three different workflows for your digital collection. These are **Backup**, **Backup and Migrate**, and **Migration on Archive**. All three are handled within a few button clicks and are discussed in the following subsections. In order to start one of these workflows click the corresponding button in the main window as in Figure 8. Before going through these individually let's take a look into a brief explanation of some keywords.

- Version - A version is an exact copy of an element (a file) at a given time. It's basically like a snapshot of the file. An element can have many versions spread across different storage media. There may be even more than one copy of a single version across different storage media.

- Migration - A migration is a special form of a version. It represents a copy of a file but in another format (representation). For example if you have a MS-Word file and migrate it into a PDF/A file. (Please note, that
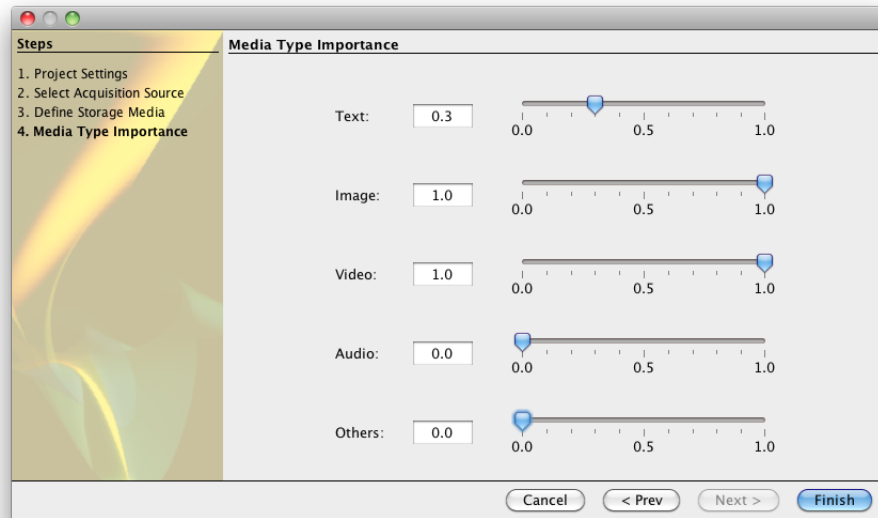
6

Figure 7: setup a project - select the media type importance

the newly created migrations are a modification of the original data and always include the risk of loosing functionalities or content of the original data.)

- Source - A source represents your original collection on a host system which Hoppla will preserve.

- Archive - When we use archive here, we mean your digital collection managed within Hoppla.

4.2.1 Backup

The first workflow or preservation strategy is the *Backup* workflow. If you choose the backup option, Hoppla will make a copy of all files from the source and store them on the selected storage media. For hard drive source media the File/Folder structure will be preserved exactly as on the source media. Hoppla also creates a new copy (a new version) of files that already exist in the archive and have changed since the last backup. Using this simple replication method will decrease the risk of the physical loss of your data (including the versions of the objects).
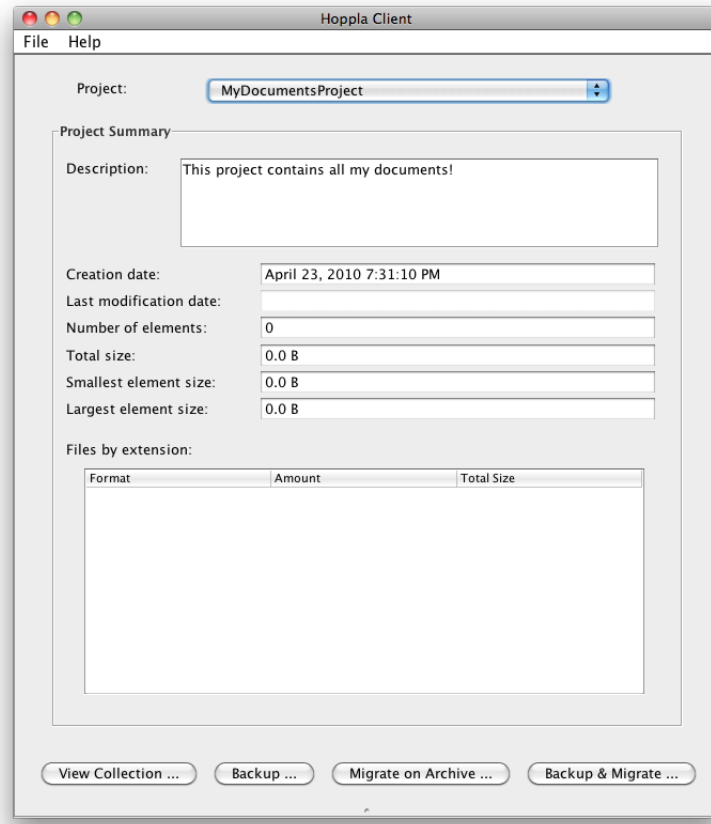
Figure 8: Hoppla's Main Window

4.2.2   Backup and Migrate

In addition to bit preservation (backup) this workflow provides logical preservation for endangered file formats for all new source files. If new source files are in formats that are at risk to become obsolete Hoppla suggests migration strategies for the objects. A dialog will open and present you the suggested migrations.

The result of this worflow is a bit preservation of all new and changed objects and also migrations of endangered objects. The migrations are also stored on the storage media. This is very important since it enables you to access your data even if a format becomes obsolete or the software to render the object is not any longer compatible with your system.

4.2.3 Migrate on Archive

The Migrate on Archive workflow performs migrations on objects that are already in your archive, for example if new migration strategies become available. The new migration will also be stored on storage media. Your should perform this workflow at regular intervals to keep your digital archive up to date.

4.3 Preservation Tools

In order to migrate files that are in obsolete or endangered formats Hoppla uses a set of installed tools on your host system.

The tool manager can be opened in *File > Manage Preservation Tools...*. In the window you can view which tools are supported. The list indicates tools which are already found by the discovery service integrated within Hoppla. If tools are installed on the system and not found by Hoppla you can specify the installation path of the tool in the list. After that, just trigger a new search by hitting the button below. This will activate the tool and it will be used by Hoppla from this moment on.

4.4 Collection Browser

Well, setting up projects and preserving data is good but eventually you will want to look at your archive and you sometimes will have to restore some of your original files if they have become lost, damaged or obsolete. There is where the collection browser comes in. It allows you to browse through your digital archive. The features of the browser include filtering based on different criteria, exploration of object metadata and history (version and migration) and different perspectives of the collection (list, hierarchic, tree, etc.). You can also access your archived objects form the collection browser (Hoppla needs to have access to the storage media for this feature so if the archive is stored on an external hard drive you'll have to plug it in). If a file or even a whole folder goes missing or damaged from your source you can select it in the collection browser and right-click it. Simply select 'data recovery' in the pop-up menu and Hoppla will ask you where you want to recover the files.

5 Privacy

Hoppla is implemented with regard to your privacy and non of its features threatens it or exposes your files to our web update server or any third parties. During the migration process Hoppla creates a collection profile containing the file formats of your source collection, a list of available migration tools and a user profile based on the project you manage. It contains no file names, encoded files, part of files, user names, passwords or any other data of private information. Based on the information sent the server is able to create a migration plan that is most suitable to your needs. This migration plan contains all information required to migrate the endangered formats.

6   Future Work

Hoppla 2.0 is the first official version of Hoppla. We plan to add many new features and repair bugs and flaws. If you found one please feel free the submit it at hoppla@ifs.tuwien.ac.at. The following list outlines some of the features that we plan to add in the next versions.

- new source types, e.g. social networks

- new storage medium types, e.g. CDs, DVDs

- special support for complex objects, e.g. html pages, zip files, ISOs, etc.

7   Feedback

Hoppla is still in its early stages and we will be greatful about any kind of feedback. If you have any questions, found a bug or just want to criticize please don't hesitate to drop us a line at hoppla@ifs.tuwien.ac.at.

8   Legal

Your use of Hoppla is undertaken at your risk. To the fullest extent permissible pursuant to applicable law, neither DME nor any other party involved in creating, producing, or delivering Hoppla is liable for any direct, incidental, consequential, indirect, or punitive damages arising out of your use of the Hoppla Software or the first-party and third-party services that are accessible through the Hoppla software. Without limiting the foregoing, everything within the Hoppla software is provided to you "as is," without warranty of any kind, either expressed or implied, including, but not limited to, the implied warranties or merchantability, fitness for a particular purpose, or noninfringement.

DME does not guarantee the accuracy or completeness of any information provided within the Hoppla software. DME and its partner are not responsible for the consequences resulting from the use of such information.