

# EPrints Preservation – Preservation Action

## Table of Contents

---

Table of Contents .....	1
1 Introduction .....	1
2 Tutorial - Aim.....	1
3 The Formats/Risks Screen.....	1
3.1 Exercise 1 – Finding your GIF Collection EPrint.....	2
3.2 Exercise 2 – Uploading your Preservation Plan .....	2
3.3 Exercise 3 – Viewing the Result (via abstract page) .....	3
3.4 Exercise 4 – Viewing the Result (via Format/Risks) .....	3
4 Re-Enacting a Plan on New Files (Exercise 5).....	4
5 Provenance .....	5
5.1 Exercise 6 - Viewing Document Relations.....	5

## 1 Introduction

---

This tutorial is designed to follow the File Format Identification and Risk Analysis Exercise. It is also recommended that you follow the Planets example for Plato in migrating the set of GIF files that we looked at in the last section. Having followed these 2 exercises you should ideally be coming to this workshop with a preservation plan (XML) given to you by Plato.

## 2 Tutorial - Aim

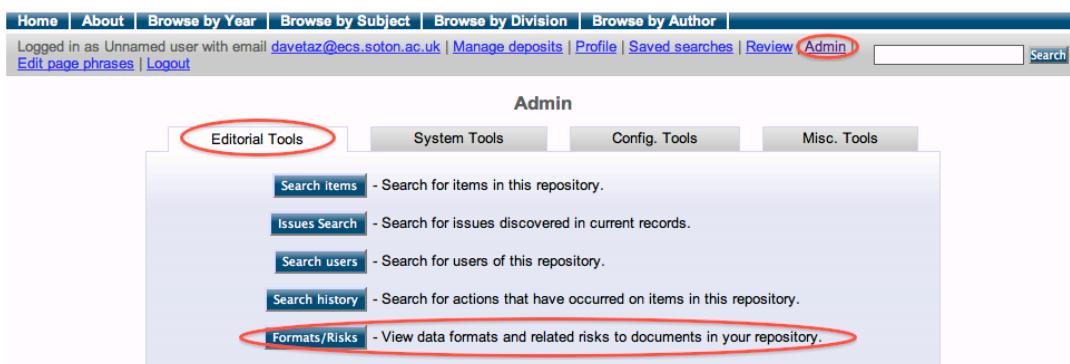
---

The aim of this tutorial is to demonstrate the methods available in EPrints for applying a preservation plan to a set of objects. With the need to re-view this preservation plan in the future, we also demonstrate the tools available for viewing the plan, reversing any actions which resulted from this plan and removing it. We also look at the provenance data stored by EPrints which relates to your preservation choices.

## 3 The Formats/Risks Screen

---

This screen will be our main reference point throughout the exercise. Available via the **Admin** interface (shown below) we can view the file types in our repository and any related risk scores.



The screenshot shows the EPrints Admin interface. At the top, there is a navigation bar with links: Home, About, Browse by Year, Browse by Subject, Browse by Division, Browse by Author, and a login message: 'Logged in as Unnamed user with email davetaz@ecs.soton.ac.uk | Manage deposits | Profile | Saved searches | Review Admin | Edit page phrases | Logout'. Below the navigation bar is a search bar with a 'Search' button. The main area is titled 'Admin' and contains a tab bar with 'Editorial Tools' (which is highlighted with a red circle), 'System Tools', 'Config. Tools', and 'Misc. Tools'. Under the 'Editorial Tools' tab, there are several buttons with descriptions: 'Search items' (Search for items in this repository), 'Issues Search' (Search for issues discovered in current records), 'Search users' (Search for users of this repository), 'Search history' (Search for actions that have occurred on items in this repository), and 'Formats/Risks' (View data formats and related risks to documents in your repository). The 'Formats/Risks' button is also highlighted with a red circle.

Following on from the previous exercise we should have a series of GIF files listed in the high risk category:

**High Risk Objects**

Graphics Interchange Format (Version 1987a)   6

Title	User
<a href="#">Triumph_Spitfire_1500.gif (54Kb)</a>	Unnamed user with email <a href="mailto:davetaz@ecs.soton.ac.uk">davetaz@ecs.soton.ac.uk</a>
<b>Title:</b> TTest	
<b>EPrint ID:</b> 104	<b>User:</b> Unnamed user with email <a href="mailto:davetaz@ecs.soton.ac.uk">davetaz@ecs.soton.ac.uk</a>
<a href="#">Volvo_264_GL.gif (61Kb)</a>	
<b>Title:</b> TTest	
<b>EPrint ID:</b> 104	<b>User:</b> Unnamed user with email <a href="mailto:davetaz@ecs.soton.ac.uk">davetaz@ecs.soton.ac.uk</a>
<a href="#">VW_Golf.gif (56Kb)</a>	
<b>Title:</b> TTest	
<b>EPrint ID:</b> 104	<b>User:</b> Unnamed user with email <a href="mailto:davetaz@ecs.soton.ac.uk">davetaz@ecs.soton.ac.uk</a>
<a href="#">Zastava_750.gif (49Kb)</a>	
<b>Title:</b> TTest	
<b>EPrint ID:</b> 104	<b>User:</b> Unnamed user with email <a href="mailto:davetaz@ecs.soton.ac.uk">davetaz@ecs.soton.ac.uk</a>

**Preservation Actions**

---

**Download File Selection**

No. of Files:

---

**Upload Preservation Plan**

no file selected

### 3.1 Exercise 1 – Finding your GIF Collection EPrint

For the purposes of this exercise we need to be able to view our EPrint containing the high risk GIF files. This can be done easily via the **Format Risks** interface by finding the **EPrint ID** listed in the left hand column of the format details. From the screenshot above, our EPrintID is 104, but it is most likely to be different in your repository.

Once you have found your EPrint ID, open the EPrint (by clicking the ID) in a new window or tab for later use. At the moment it should look something like the following:

**Article #21**

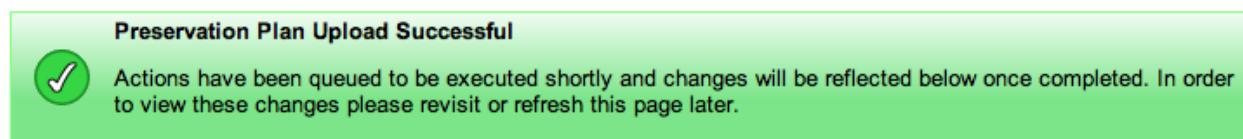
**UNSPECIFIED UNSPECIFIED**

	Image (GIF) <a href="#">Download (76Kb)</a>   <a href="#">Preview</a>
	Image (GIF) <a href="#">Download (79Kb)</a>   <a href="#">Preview</a>
	Image (GIF) <a href="#">Download (79Kb)</a>   <a href="#">Preview</a>
	Image (GIF) <a href="#">Download (77Kb)</a>   <a href="#">Preview</a>
	Image (GIF) <a href="#">Download (51Kb)</a>   <a href="#">Preview</a>

### 3.2 Exercise 2 – Uploading your Preservation Plan

In this exercise we are going to upload our preservation plan. This can be done by simply browsing for the plan and uploading it via the **Preservation Actions** panel.

Upon successful upload you should see the following message at the top of your screen:



EPrints will now be parsing and executing any actions defined by the plan in the background. This is necessary as you may be performing an operation on potential 100s or even 1000s of files.

It is important to realize that EPrints will attempt to execute the actions defined in the plan on all the files listed in the file format group through which you uploaded the plan. In this case this should be our collection of GIF files.

### 3.3 Exercise 3 – Viewing the Result (via abstract page)

Viewing the result can be done in two ways, via the abstract page and via the Format/Risks screen. In this section we are going to simply refresh the abstract page of the EPrint which you opened in **Exercise 1**. You may need to refresh this a few times and wait a while for the actions to happen.

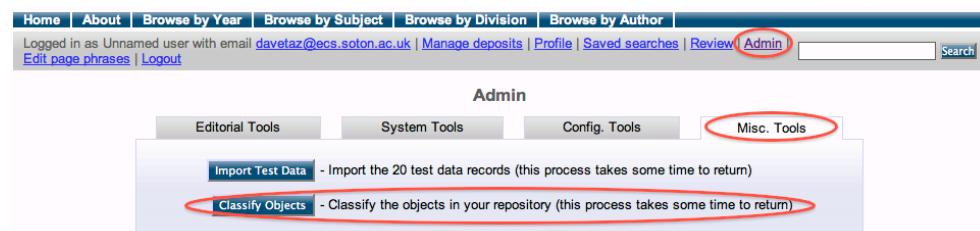
If the process ran successfully, your abstract page will have a number of items which look similar to that shown below:



Here we can see that the migrated version is now the primary document. The original GIF is still available but has lost its prominent position on the page.

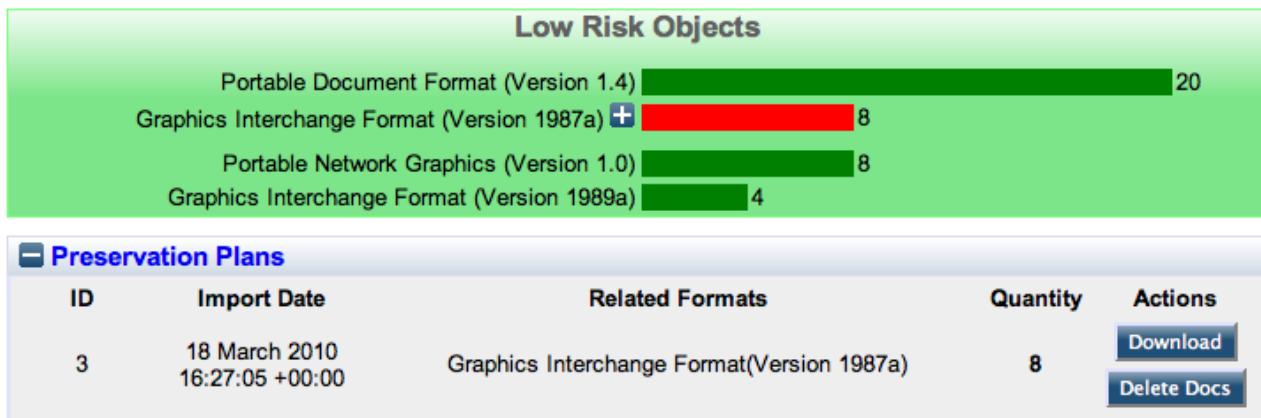
### 3.4 Exercise 4 – Viewing the Result (via Format/Risks)

The above process has added documents to our publication records, so we now need to re-classify them. This operation be performed using the **Classify Objects** button available via the **Misc Tools** tab in the **Admin** interface.



In practice, classifying the objects would be run on the server as a background task, but we have provided a button for this tutorial. In a fully populated repository, this task could take considerable time.

Once complete your **Format/Risks** screen will look slightly different from before:



Now we can see our migrated GIF files listed in the Low Risks category (still with a red bar), but we can also see our 8 new PNG files.

Below the file format information we now have a **Preservation Plans** box. Expanding this shows information relating to the plan, such as the time it was imported, the format it relates to and the number of documents which were created as result of this plan.

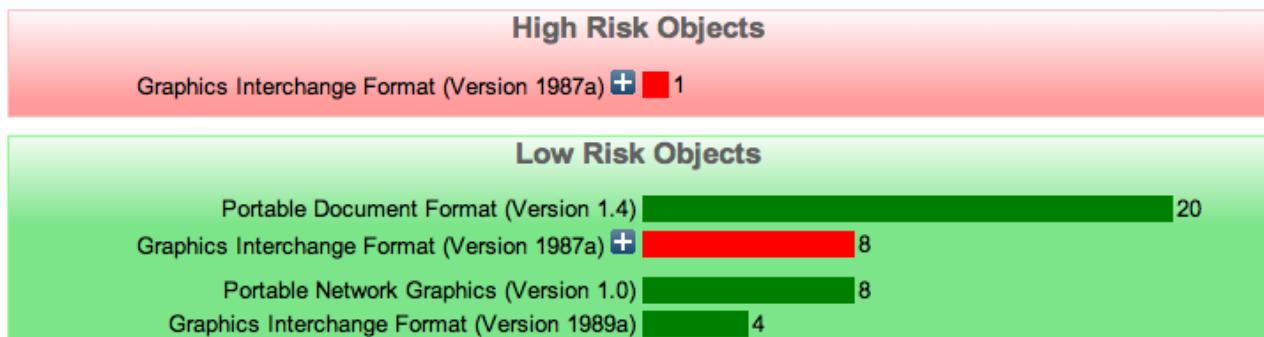
#### 4 Re-Enacting a Plan on New Files (Exercise 5)

Although this should be an automated process, for the purposes of the tutorial we will be manually controlling what would happen if a new at risk GIF was added to the repository.

In order to do this we need to create a new EPrint containing at least one of the at risk GIF files. As before the set of files are available at [http://www.eprints.org/software/training/3.2/march\\_2010/](http://www.eprints.org/software/training/3.2/march_2010/). We recommend that you download the zip file containing these onto your computer so you can view and upload them into a new EPrint. The file "VW\_Golf.gif" is a perfect example of a GIF which will be at risk.

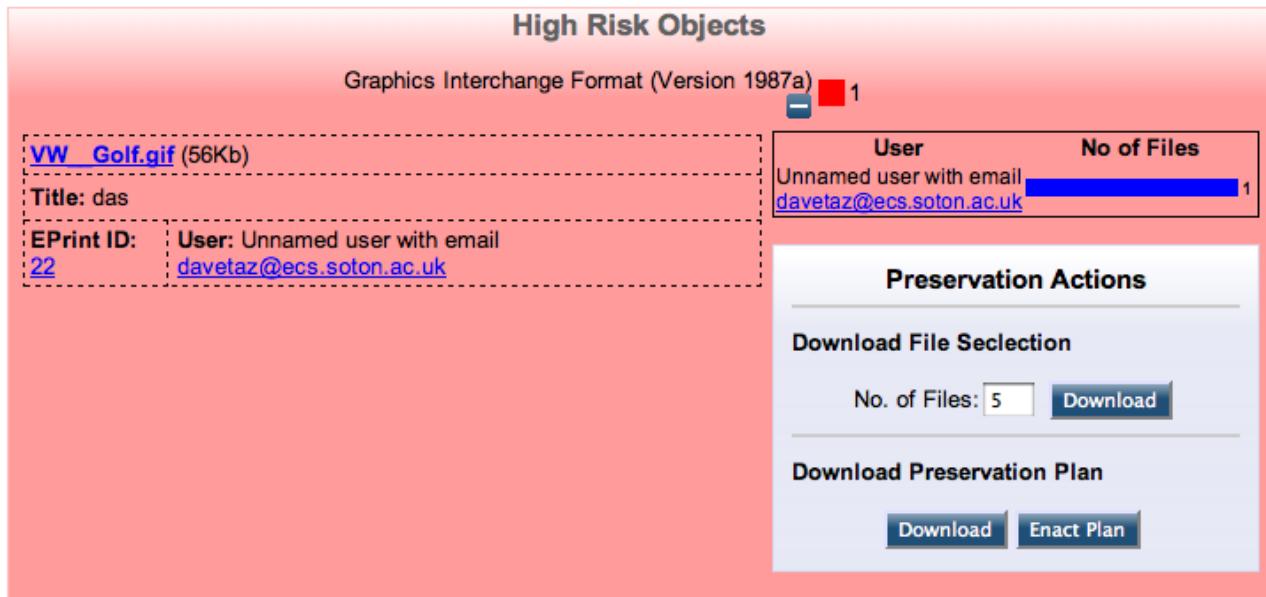
Once you have created a new EPrint, you will need to re-classify the objects in the repository (see **Exercise 3**).

Viewing the **Formats/Risks** screen should then display the following:



Here we can see that we have GIFs files of the same version listed in both high and low risk categories. This is because the one which is high risk has not yet been migrated.

To manually enact our preservation action on this object, we can open up the high risk GIF to view it's details and click the **enact plan** button.



The object of this exercise is to show that EPrints can recognize the fact that it already has a plan that you previously related to this file format.

As before, reload the item's abstract page, or re-classify your objects in order to confirm that the process ran successfully.

## 5 Provenance

---

EPrints 3.2 now includes an arbitrary relational model allowing you to add relations between objects and other objects in a Linked Data fashion.

As EPrints 3.2 now treats thumbnails as documents, which are “volatile versions” of the main document they relate to. This is one example of relational metadata that is stored.

In terms of the preservation data, when a file is migrated, the original document is given a new relation which links the two. The new migrated file is related back to the original, and also related to the plan which was responsible for causing its creation.

### 5.1 Exercise 6 - Viewing Document Relations

One way to view document relations is to customize your EPrint submission workflow. This can be done via the **View Configuration** tool in the **Config. Tools** tab of the **Admin** screen.

From the list of configuration files we need to find **workflows / eprint / default.xml**. Click on the file to view it and then edit it.

In this file, find the following line ion the **Documents** section:

```
<!-- <field ref="relation" />-->
```

and change it to:

```
<field ref="relation" />
```

After saving this file, from the **Config Tools** tab of the **Admin** interface click **Reload Configuration**.

We now need to find the abstract page of our EPrint from Exercise 1, scroll to the bottom and click **View Item**. If we then follow the **Edit** link to get to the **Upload** stage in the workflow, we can now click the **show options** link on a document to view the relations which that document holds.

By showing options on a previously at risk GIF we should see the following:

Relationships:			
7.	<a href="http://eprints.org/relation/hasVolatileVersion">http://eprints.org/relation/hasVolatileVersion</a>	/id/document/68	<input checked="" type="checkbox"/> <input type="checkbox"/>
8.	<a href="http://eprints.org/relation/hasVersion">http://eprints.org/relation/hasVersion</a>	/id/document/68	<input checked="" type="checkbox"/> <input type="checkbox"/>
9.	<a href="http://eprints.org/relation/hasPreviewThumbnailVersion">http://eprints.org/relation/hasPreviewThumbnailVersion</a>	/id/document/68	<input checked="" type="checkbox"/> <input type="checkbox"/>
10.	<a href="http://eprints.org/relation/hasMigratedVersion">http://eprints.org/relation/hasMigratedVersion</a>	/id/document/125	<input checked="" type="checkbox"/> <input type="checkbox"/>
11.	<a href="http://eprints.org/relation/hasVersion">http://eprints.org/relation/hasVersion</a>	/id/document/125	<input checked="" type="checkbox"/> <input type="checkbox"/>

Here we can see that this file has a migrated document of ID 125.

By finding that document (**which cannot be done via the ID**, you will need to find the document which has the same filename and is a PNG) we can see both the inverse relation and the plan which the migration was a result of.

1.	<a href="http://eprints.org/relation/isVersionOf">http://eprints.org/relation/isVersionOf</a>	/id/document/64	<input checked="" type="checkbox"/>
2.	<a href="http://eprints.org/relation/isMigratedVersionOf">http://eprints.org/relation/isMigratedVersionOf</a>	/id/document/64	<input checked="" type="checkbox"/> <input type="checkbox"/>
3.	<a href="http://eprints.org/relation/isMigrationResultOf">http://eprints.org/relation/isMigrationResultOf</a>	/id/preservation_plan/3	<input checked="" type="checkbox"/> <input type="checkbox"/>
4.	<a href="http://eprints.org/relation/hasVolatileVersion">http://eprints.org/relation/hasVolatileVersion</a>	/id/document/148	<input checked="" type="checkbox"/> <input type="checkbox"/>

Note that it is not currently recommended to expose the relations contained in your repository in this way. Reversing the earlier change to the workflow and re-loading the configuration will hide them again.