# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audience</td>
<td>v</td>
</tr>
<tr>
<td>Documentation Accessibility</td>
<td>v</td>
</tr>
<tr>
<td>Related Documents</td>
<td>vi</td>
</tr>
<tr>
<td>Conventions</td>
<td>vi</td>
</tr>
</tbody>
</table>

## 1 Introduction

1.1 Understanding Site Studio Publisher ........................................ 1-1
1.2 Installing Site Studio Publisher ........................................... 1-2
1.3 Starting Site Studio Publisher ............................................... 1-2
1.4 Guidelines for Publishing ..................................................... 1-2
1.5 Guidelines for Customizing Site Studio Pages ............................ 1-3
1.6 System Requirements .................................................................. 1-4

## 2 Publisher Tasks

2.1 About Publisher Tasks .............................................................. 2-1
2.2 Creating and Editing Tasks ...................................................... 2-1
2.3 Specifying Task Parameters .................................................... 2-1
2.3.1 Settings ............................................................................. 2-2
2.3.2 Dates .................................................................................. 2-5
2.3.3 Days .................................................................................. 2-6
2.3.4 Run Times ......................................................................... 2-6
2.3.5 Options ............................................................................. 2-6
2.3.6 Include List ....................................................................... 2-7
2.3.7 Exclude List ...................................................................... 2-7
2.3.8 Filter Sets ......................................................................... 2-7
2.3.9 Triggers ............................................................................ 2-8
2.4 Configuring Publisher .............................................................. 2-10

## A User Interface

A.1 Site Studio Publisher Tasks ...................................................... A-1
A.2 Add New Task/Edit Task .......................................................... A-2
A.2.1 Publisher Settings ............................................................... A-2
A.2.2 Dates ................................................................................ A-6
A.2.3 Days ............................................................................... A-6
A.2.4 Run Times ......................................................................... A-7
A.2.5 Options .......................................................................................................................... A-7
A.2.6 Include List .................................................................................................................. A-8
A.2.7 Exclude List ................................................................................................................. A-8
A.2.8 FilterSets ..................................................................................................................... A-9
A.2.9 Triggers ....................................................................................................................... A-10
A.3 Task Info ....................................................................................................................... A-12
A.4 Logs ................................................................................................................................ A-13
A.5 View Log ........................................................................................................................ A-13

Index
Preface

The User’s Guide for Site Studio Publisher contains information to assist administrators responsible for the publishing of Web sites managed by Site Studio.

Audience

This document is intended for those people identified in the organization who are responsible for publishing an organization’s Web site managed by Site Studio.

Documentation Accessibility

Our goal is to make Oracle products, services, and supporting documentation accessible to all users, including users that are disabled. To that end, our documentation includes features that make information available to users of assistive technology. This documentation is available in HTML format, and contains markup to facilitate access by the disabled community. Accessibility standards will continue to evolve over time, and Oracle is actively engaged with other market-leading technology vendors to address technical obstacles so that our documentation can be accessible to all of our customers. For more information, visit the Oracle Accessibility Program Web site at http://www.oracle.com/accessibility/.

Accessibility of Code Examples in Documentation

Screen readers may not always correctly read the code examples in this document. The conventions for writing code require that closing braces should appear on an otherwise empty line; however, some screen readers may not always read a line of text that consists solely of a bracket or brace.

Accessibility of Links to External Web Sites in Documentation

This documentation may contain links to Web sites of other companies or organizations that Oracle does not own or control. Oracle neither evaluates nor makes any representations regarding the accessibility of these Web sites.

Access to Oracle Support

Oracle customers have access to electronic support through My Oracle Support. For information, visit http://www.oracle.com/support/contact.html or visit http://www.oracle.com/accessibility/support.html if you are hearing impaired.
Related Documents

For more information, see the following documents in the Oracle Site Studio documentation set:

- User’s Guide for Site Studio Contributor
- User’s Guide for Site Studio Designer
- Administrator and Manager’s Guide for Site Studio
- Technical Reference Guide for Site Studio

Conventions

The following text conventions are used in this document:

<table>
<thead>
<tr>
<th>Convention</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>boldface</strong></td>
<td>Boldface type indicates graphical user interface elements associated with an action, or terms defined in text or the glossary.</td>
</tr>
<tr>
<td><em>italic</em></td>
<td>Italic type indicates book titles, emphasis, or placeholder variables for which you supply particular values.</td>
</tr>
<tr>
<td><strong>monospace</strong></td>
<td>Monospace type indicates commands within a paragraph, URLs, code in examples, text that appears on the screen, or text that you enter.</td>
</tr>
</tbody>
</table>
This section covers the following topics:
- "Understanding Site Studio Publisher” on page 1-1
- "Installing Site Studio Publisher” on page 1-2
- "Starting Site Studio Publisher” on page 1-2
- "Guidelines for Publishing” on page 1-2
- "Guidelines for Customizing Site Studio Pages” on page 1-3
- "System Requirements” on page 1-4

1.1 Understanding Site Studio Publisher

This guide describes how to use Site Studio Publisher to publish a Site Studio Web site. With Site Studio, you store the Web sites you build in Oracle Content Server. At some point, you may find that you would like to create a static copy of a Web site and transfer it from an Oracle Content Server environment to a pure web server environment that is not running an Oracle Content Server instance. This process is referred to as publishing.

Site Studio Publisher creates a static snapshot of a dynamic site by traversing all the links in a Web site (visiting all of the linked pages) and downloading a copy of each page and all of the resources (images, flash movies, and so on) on each page. Your entire Web site, including the content of queries, layout pages, fragments, contributor data files, and native documents, are then copied and published to the new server.

The most efficient use of a dynamic site would have immediate updates and automatically update only the changed areas of the Web site. Since Site Studio Web sites are dynamic and created using a large number of variables, it is very difficult to track and publish the affected areas of the Web site in a reliable and consistent manner. Site Studio Publisher can selectively process, or not process, pages on a Web site based on different control methods. When each asset does not need to be compared each time, then it is much easier to keep the static site up to date, and the site owner can easily control the process.

Site Studio Publisher operates from an initial list of available links, and then compares them to a previous version of links used by Site Studio Publisher. The use of controls and filters can control the list, and also limit Site Studio Publisher to using only certain parts of the list. The first time Site Studio Publisher examines a Web site, all links are examined and added to the list.

When Site Studio Publisher examines a link and determines that it has changed, the content is analyzed and the new set of links for that page is determined. If the
examination is skipped, either because no changes were found or because the controls and filters require that a specific link is skipped, then the previous list of links is used for comparison.

### 1.2 Installing Site Studio Publisher

Site Studio Publisher installation is simple. On Oracle Content Server 11g, the Site Studio Publisher component is pre-installed, and is enabled through the Content Server UI.

To enable Site Studio Publisher, perform these tasks:

1. Log in to the Oracle Content Admin Server page with administrator credentials.
2. Go to the Component Manager.
3. Click the checkbox next to Site Studio Publisher under the Web Content Management section.
4. Click Update at the bottom of the page.
   A dialog will appear confirming your selection.
5. The server must be restarted to complete enabling Site Studio Publisher.

To install and enable the Site Studio Publisher component on a 10gR3 Oracle Content Server instance, see the Content Server 10gR3 documentation.

### 1.3 Starting Site Studio Publisher

Site Studio Publisher is run from the Site Studio Administration section of the Oracle Content Server user interface.

To get to the Site Studio Publisher interface, perform these tasks:

1. Log in to the Oracle Content Server with administrator credentials.
2. Open the Administration tray or page, depending on your layout.
3. Select Site Studio Administration.
4. Select Site Studio Publisher.

### 1.4 Guidelines for Publishing

To publish a Web site from an Oracle Content Server environment to a web server, all the content of your site is copied by Site Studio Publisher. However, the copy of the site must be both complete and self-contained. That is, the copy must contain all of the resources of the original site, and pages in the copy must refer only to resources within the copy itself. To create a self-contained copy, Site Studio Publisher rewrites any absolute (full) URLs within the original site to relative URLs to allow the static copy to be hosted on a different hosting instance.

Site Studio Publisher uses the following procedures to identify and evaluate links and resources contained in a particular page:

- When handling HTML, Site Studio Publisher looks for any attributes in tags that may contain links (for example, the 'HREF' attribute in an 'A' tag or the 'SRC' attribute in an 'IMG' tag), and downloads these images. If the link is absolute, the link in the copied site is updated to refer to the downloaded image (through a relative URL) instead of the original image.
When handling JavaScript, Site Studio Publisher looks for specific JavaScript commands that are commonly used to load a new image or change what is being displayed on the page. When those specific commands are encountered, it examines the JavaScript and interprets each command. However, an unrecognized JavaScript command may cause Site Studio Publisher to not include a referenced page or fail to download an image.

Be careful when appending URL parameters to a page URL. By default, in Site Studio Publisher, extra page parameters that result in a variant of the dynamic page are not used when naming the static page. Site Studio pages should not be customized by appending URL parameters to a node reference, as any appended parameters are ignored (that is, any node parameters not defined using Site Studio are not used in naming the pages).

### 1.5 Guidelines for Customizing Site Studio Pages

When customizing Site Studio-generated JavaScript or writing your own components, be sure to use relative URLs for all images, movies, and so on, that are referenced in the script; avoid concatenating strings to build up paths to resources. This makes it easier for Site Studio Publisher to properly identify and evaluate the external references.

Pay special attention to those items included in a dynamic list or using JavaScript patterns, and also to naming conventions for published pages.

#### Using JavaScript in Links for Published Pages

You may want to specify a JavaScript pattern in a link to be used in your published Web pages. Certain patterns must be used for Site Studio Publisher to identify and evaluate JavaScript references and links in order for these references to be converted to relative references in the published version.

Site Studio Publisher looks for specific patterns in the embedded JavaScript of Web pages. When these patterns are encountered, Site Studio Publisher examines the JavaScript and interprets the command.

- Site Studio Publisher evaluates general JavaScript patterns that are commonly used to load a new image or change what is being displayed on the page.
- Site Studio Publisher also evaluates certain Site Studio--specific JavaScript patterns.

**Note:** An unrecognized JavaScript command may cause Site Studio Publisher to not include a referenced page or fail to download an image.

In all patterns listed below, double quotation marks ("') and single quotation marks (') are treated as equivalent. The crawler attempts to match any valid URL within the quotes.

#### General JavaScript Patterns

<table>
<thead>
<tr>
<th>JavaScript fragment</th>
<th>Link to be traversed</th>
<th>Replacement</th>
</tr>
</thead>
</table>
1.6 System Requirements

The Site Studio Publisher application requires Java version 5 (JRE 1.5) or higher.
Publisher Tasks

Site Studio Publisher allows you to define multiple tasks used in crawling and publishing the sites you maintain. These tasks are used to define how much of the site is crawled and how often the results are published.

Additionally, some default value of Site Studio Publisher, such as queue size, can be modified in the configuration file.

This section covers the following topics:

- "About Publisher Tasks" on page 2-1
- "Creating and Editing Tasks" on page 2-1
- "Configuring Publisher" on page 2-10

2.1 About Publisher Tasks

Tasks in Publisher are individual commands used to crawl a certain URL and then publish the gathered content. Each task can be made to crawl across an entire Web site, or across just a few pages.

Each task crawls the Web site (or portion of the Web site) at the time marked in the task. Tasks can also be created to run only on demand.

2.2 Creating and Editing Tasks

Creating a task requires little more than a URL of the site to crawl. Tasks are easy to create and edit because of the simple interface.

2.3 Specifying Task Parameters

The different parameters for a task are relatively straightforward. You need to have the URL for the Web site, and username and password, and you must set the parameters for what days the task will crawl the site. Some of the parameters, such as triggers, are more complex.

The parameters are grouped into separate sections on the page.

This section deals with the following topics:

- "Settings" on page 2-2
- "Dates" on page 2-5
- "Days" on page 2-6
- "Run Times" on page 2-6
2.3.1 Settings

The Settings of a task include some of the most basic information of a Publisher task. This includes the location of the site to crawl, any necessary login information and what errors are critical.

The Settings section has advanced settings, which can easily be seen when you click the Show Advanced Settings link. The advanced settings expand and display in the same section.

Description
The name of the task is used to differentiate it from other tasks. It can also contain some information about how often the task runs, or what site it crawls, and so forth.

Manifest Url
This is the URL for Site Studio Publisher to crawl and then publish. The server, instance, and site must be filled in by the user.

Output Path
The local directory where the content is downloaded. Important state information is stored here as well. This is relative to the SSPHome location.

Username
The user ID, for password-protected sites.

Password
The password, for password-protected sites. The password is encrypted in the .hda file.

Authentication
The type of authentication used on the site, for password-protected sites.

- LoginForm is used for Oracle Content Server 11g servers.
- BasicAuth is used for Oracle Content Server 10g servers.
- ExtranetLook is used for Oracle Content Server 10g servers with the ExtranetLook component.
- NTLM is used for Oracle Content Server 10g servers configured to use NTLM.
- CustomForm is used to customize a different form-based login.

Publish Now
Publish Now is a feature that marks pages with an implied high priority. As a contributor edits a Web page, they may mark a page by selecting "Publish Now" from Contributor. Publisher obtains a list of the marked pages from the server when it requests the list at the beginning of its operation.
When a crawl is executed, if the Publish Now control is set to true, Publisher only compares and publishes the marked, changed pages. If the altered content includes new links, Publisher also publishes those links. This crawl is suitable to run frequently.

**Force Download**
When Publisher typically examines links, the current server state is compared with the record of what Publisher previously found. What Publisher does not check is if the asset previously found is still in place on the local file system. The Force Download control is used to override in case the output file system has missing or corrupt assets.

**Force Analyze**
The Force Analyze control is useful when you have implemented filters in your task specification. This control is only invoked during analysis, so use of this control forces the analysis pass on those files marked for examination on this crawl. If no controls are used to limit the examination operation, all files are re-analyzed with this setting.

This control is intended for use during development. The best use of the control is when filterset and filter controls within the task specification are used. If a filter is changed, then this usually affects some link rewriting behaviors. Normally, if the dynamic page has not changed, the filters are not re-applied. The result is that when a dynamic page has not changed, the links are not re-written using the changed filters.

**Treat Home Page Errors As Critical**
This control allows you to stop the task if errors are found while retrieving the home page. If the box is checked, then any error on the home page will abort the task.

**Treat Manifest Errors as Critical**
This control allows you to stop the task if there are errors in the URLs listed in the site manifest. If the box is checked, then any errors found retrieving the URLs in the site manifest will abort the task.

**Use Cache Control**
Cache Control is a control that allows the server to compare the Max Age (caching) section property controlling the CacheControl:max-age http header. Using this control, Publisher only selects and crawls those pages where the max-age value has not expired.

This allows various parts of the site to be labeled as "rarely changed" and still create, control, and execute positive checks at controlled intervals. This enables you to run weekly, monthly, quarterly, or even annual publishing controls to enable timely execution of the publication of the Web site.

**Use Last Modified**
The Last Modified control is used to compare the http header Last-Modified returned from the web server. The web server normally returns this header for all resources accessed with a weblayout URL. (This response header is not provided for a dynamic Site Studio page.)

If useLastModified is checked, the Last-Modified value is re-submitted in an If-Modified-Since request header the next time that resource is retrieved. That then
allows the web server to return a 304 - Not Modified response if the resource is unchanged. If this is not checked, the resource is downloaded and compared with the previously retrieved content.

You would only choose to uncheck this control if you found that your web server was returning unreliable results for Last-Modified.

**Use SSPETag**
The SSPETag control is used to identify changes to files retrieved by the GET_FILE service. Use of this mechanism allows the content server to return a 304 - Not Modified response when applicable and so avoid unnecessary downloads.

Without this mechanism, file content is retrieved and compared with the previous version.

**Default Filename**
The default filename is the filename used for URLs that do not have an explicitly-specified filename.

**Page Extension**
The extension that is added to page urls that do not otherwise specify an extension. For example, a typical reference to a document in the dynamic site might look like:

http://myServer/mySite/Section1/DocumentX

Where DocumentX is the dDocName of the target item in Content Server. For use on a static site, you must add an extension.

**User Agent**
This control enables you to specify a value for the User-Agent http request header used by Site Studio Publisher when crawling the site.

**Friendly Url Parameters**
This control enables you to select a comma-separated list of additional parameter names to honor. Friendly URLs already honor parameters ending in NextRow or dcPageNum, and use the parameter name and value to construct the filename for the crawled page.

If a URL Parameter affects the appearance of the page, then you need to capture a different copy of the page for each combination of parameter values.

**Dynamic Url Parameters**
This control enables you to select a comma-separated list of additional parameter names to honor. Dynamic URLs using the GET_PAGE service already honor dID, dDocName, RevisionSelectionMethod, and Rendition.

If a URL Parameter affects the appearance of the page, then you need to capture a different copy of the page for each combination of parameter values.

**Additional Services**
This control enables you to select additional services to be crawled. There is built-in support for those services that are expected to generate meaningful static content: SS_GET_PAGE, GET_FILE, and GET_DYNAMIC_CONVERSION. Any services entered will be added to those three as the crawled services. There is not a method to prevent the crawl from using one of the three default services.
For more information on the Site Studio Services, see the *Technical Reference Guide for Site Studio*.

**Soft Error Threshold**
This control determines the number of soft errors allowed. If this number is exceeded, publishing fails.

**Hard Error Threshold**
This control determines the number of hard errors allowed. If this number is exceeded, publishing fails.

**Delete Threshold**
This control determines the number of missing or deleted objects found while crawling. If this number is exceeded, publishing fails. Entering \(-1\) in this field allows any number of missing objects.

**Soft Error Codes**
This control enables you to specify which server error codes are treated as soft errors. No wild cards are allowed; each code must be entered explicitly in a comma-separated list.

**Hard Error Codes**
This control enables you to specify which server error codes are treated as hard errors. No wild cards are allowed; each code must be entered explicitly in a comma-separated list.

**Ignore Error Codes**
This control enables you to specify individual error codes that are ignored. These codes will not affect the Site Studio Publisher crawl. No wild cards are allowed; each code must be entered explicitly in a comma-separated list.

### 2.3.2 Dates

The scheduling of a task starts with the Dates section. Here, a task is set as either a task that runs according to a schedule, or is run only on demand.

If you select that a task is run on demand - that is, by selecting **Manual** - then the Days (see "Days" on page 2-6) and Run Times (see "Run Times" on page 2-6) sections will be unavailable.

**Manual / Range**
This control enables you to set if the task will be run only when selected, or if it will run at regular, scheduled intervals within the selected date range.

If you select Manual, then the **Start Date** and **End Date** fields, and the Days (see "Days" on page 2-6) and Run Times (see "Run Times" on page 2-6) sections will be made unavailable.

**Start Date / End Date**
This control enables you to set the start date and end date of the scheduling period for this task. There is a calendar control next to each date field.
2.3.3 Days

The Days section controls the scheduling of the task by day. Here you can set if the task is run on certain days of the week (such as every Monday and Thursday) or on certain days of the month (such as the 10th and 20th of the month).

The options in this section will be available only if **Range** was selected in the Dates section (see "Dates" on page 2-5).

**Days of Week**
This control enables you to specify which days of the week the task will run, regardless of what day of the month that day is.

**Days of Month**
This control enables you to specify which days of the week the task will run. This list can be a range, or a comma-separated list of specific dates, or a combination. For instance, 1, 15 and 10–20 and 1, 7–10, 15–19, 23 are all valid entries.

You can also use the word **last** in this field to denote the last day of the month. It can be used in a list, such as 15, last to run on the 15th of the month as well as the last day of the month.

2.3.4 Run Times

The Run Times section is used to specify how often the task runs each day.

The options in this section will be available only if **Range** was selected in the Dates section (see "Dates" on page 2-5).

**Multiple Times per Day / Once per Day**
This control enables you to select how often the task runs each day, and during what times of the day.

If you select Multiple Times per Day, you should then enter how often the task will run using the **Hours** and **Minutes** fields. Additionally, you must enter the time of day that the task will run during.

If you select Once per Day, the task will run only at one time of day, which is entered in the box.

2.3.5 Options

The Options section is used to set the priority and log levels of the task. You can also set an email address to receive notification after the task runs.

**Priority**
This control enables you to select if the task will be run at a priority level. If unchecked, the task will run at the normal level.

Priority tasks are run in a different queue than normal tasks. Since most tasks are expected to be set to normal priority, the expectation is that few, if any, tasks will be run with a high priority, enabling those that are set at high priority to run sooner and end sooner. Additionally, priority tasks would check only part of a site, by using Publish Now or Include List.

**Log Level**
This control enables you to set how much information is entered in the log.
Each item in the drop-down list includes the logging levels above it. For example, selecting **INFO** includes logs of not only **INFO** items but also **WARN** and **ERROR**.

**Email Notification**
This control enables you to enter an email address that Publisher will send an email message to once the task has completed (or failed due to errors).

### 2.3.6 Include List
This control enables you to select individual site URLs which will be specifically included in the task. If a URL entered here is also listed in an exclude filter in the Filter set (see "Filter Sets" on page 2-7), the filter set takes precedence, and the URL will **not** be included in publication.

Each time you click **Add New Item**, an additional field for entering a URL will display.

### 2.3.7 Exclude List
This control enables you to select individual site URLs which will be specifically excluded from the task.

Each time you click **Add New Item**, an additional field for entering a URL will display.

### 2.3.8 Filter Sets
A filter element describes a transformation that is applied to a downloaded content file. A filterset contains filter elements, and each filter is applied in the order the filter elements appear in the filterset. Several types of filter can be specified. Specification of a filter element depends on the kind of filter used, as specified by the type attribute.

Filtersets are used in different contexts: when applying filters to a downloaded file, for instance, or when deciding whether a given URL should be downloaded.

Each attribute of a filterset is compared to the relevant characteristics of the source content. Each and every attribute must match for the filterset itself to be considered a match. The filtersets are evaluated in the order in which they appear in the job description file, and the first filterset match is applied.

**Type**
This filterset attribute specifies the content that the filterset will invoke. The types available are:

- **transform-content**: for URLs that match this filterset, the child filter elements are applied during download, transforming the content in the manner specified by these filters.
- **transform-link**: for URLs that match this filterset, the child filter elements are applied during download. In this case, the filter elements are only applied to links found in the current downloaded file, not to the entire content.
- **exclude**: for URLs that would normally be included, match this filterset, and the URL matches, content is not downloaded.

**Path**
This filterset attribute specifies a wildcard pattern to match the file path of the URL (the part following the URL’s host name) currently downloaded. This attribute uses Java regular expression syntax for pattern matching.
Use the following reference for Java syntax:
http://java.sun.com/j2se/1.4.2/docs/api/java/util/regex/Pattern.html

**Host Name**
This filterset attribute is a value used to match the URL’s host name.

**Port**
This filterset attribute is a value used to match the URL’s port number.

**Mime Type**
This filterset attribute is a value used to match the URL’s MIME-type. The URL values that starts with the entered filter value are considered a match; text matches both text/html and text/xml.

Each filterset is a collection of at least one filter. These filters act in the order they are listed within the filterset, just as the filtersets act in the order they are listed.

The filters have the following attributes:

**Regex**
This control is the regular expression used to identify text to replace in the context of the file or link.

**Replacement**
This is the replacement value for the identified regex string.

**Path Regex**
This control is the regular expression used to identify a path and file to replace in the context of the current file. The Path Regex and Path Replacement are the best method of finding and renaming output files.

Using path regex is only meaningful within an enclosing transform-content filterset. Additionally, it does not use the Global and Ignore Case checkboxes. It is always case-sensitive, and replaces only the first match.

**Path Replacement**
This is the replacement value for the identified path regex string.

**Global**
This attribute is used to control the behavior of the substitution. If Global is checked, every match for the regex expression is replaced. If unchecked, only the first occurrence is replaced.

**Ignore Case**
This attribute is used to control the behavior of the regex expression. If Ignore Case is selected, then the case of the expression is ignored when comparing the regex value.

### 2.3.9 Triggers

Triggers enable the replication engine to run commands after the download is complete. A trigger is defined by a trigger element.
Username/Password
The username and password used for authentication. You cannot specify different authentications for each trigger. All triggers in a task must have the same credentials and authentication method.

Authentication
Select to authenticate the entered username and password either with basic authentication or with a custom form.

Type
The type control determines the type of command functionality.
A cmd trigger allows a shell script or an operating system command to be run as a trigger. The command is passed to the operating system.
The http-post trigger sends a command to the specified URL using HTTP_POST functionality.
The http-get trigger sends a command to the specified URL using HTTP_GET functionality.
The http-soap trigger sends a command, encoded as a SOAP package, to the specified URL using HTTP_POST functionality.

Command
This trigger control holds a command-line if the type is set to cmd. For the other http trigger types, the control holds a URL.
This is the command-line argument to pass through the trigger.
Examples of a command can be:
  ping MyServer
  cmd /C ren %%d\*.xml *.html

The first example (in Windows) would simply ping the MyServer server. The second is an example of a Windows command-line trigger renames any file with the filename extension of .xml to use the filename extension .html.
To run this trigger in the directory where content files have been downloaded, you must use %%d, followed by \ in a Windows environment or / in a UNIX environment, to represent the name of the download directory.

Command Data
This control is used to specify the path, relative to the task’s output path, of the file to be uploaded.
This is valid only when the trigger type is http-post or http-soap.
For the http-post, is the application/x-www-form-urlencoded data to be submitted with the POST. Essentially, this is a list of name-value pairs separated by the & character. For example:
  objectType=package&actionType=update&packageName=NAME

For http-soap, it is the location of a file containing the SOAP commands to be executed.
2.4 Configuring Publisher

There are some changes you can make to the default values of Publisher. The config.cfg file can be edited to allow different default actions and paths to be used.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Purpose</th>
<th>Defaults</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSPHome=D:/SSPHome</td>
<td>This option sets the &quot;home&quot; directory for Publisher's relative paths for output and auxiliary files.</td>
<td>data/SiteStudioPublisher</td>
</tr>
<tr>
<td>SSPLockFileTimeout=30</td>
<td>This option controls the time that the Publisher output directory is locked. This value defines how old a file must be before it will be considered &quot;stale&quot; and be overridden.</td>
<td>30 (seconds)</td>
</tr>
<tr>
<td>SSPOverrideStaleLockFile</td>
<td>This option enables or disables the SSPLockFileTimeout mechanism.</td>
<td>true (enabled)</td>
</tr>
<tr>
<td>SSPDeleteLockFile</td>
<td>This option determines if Publisher will delete the &quot;lock&quot; file at the end of a run.</td>
<td>true</td>
</tr>
</tbody>
</table>
The `config.cfg` file is accessed in the **General Configuration** section of the Admin Server, on the **Additional Configuration Variables** screen.

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Purpose</th>
<th>Defaults</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSPTimerTickInterval=5000</td>
<td>This option determines, in milliseconds, the interval at which Publisher will examine the task schedule and queues to run eligible tasks and update status.</td>
<td>5000 (milliseconds)</td>
</tr>
<tr>
<td>SSPTimerTickInterval=5000</td>
<td><strong>SSPViewLogPageLength=5000</strong> This option lets you control the number of lines displayed per page when viewing a log file. By default, Publisher breaks a log file into 5000 line chunks for viewing in the &quot;View Log&quot; page (see &quot;View Log&quot; on page A-13).</td>
<td>5000 (lines)</td>
</tr>
<tr>
<td>SSPLogFileAgeLimit=7</td>
<td>This option determines how long log files are stored. Publisher will keep log files spanning up to 7 days for each task. Old logs for a specific task are discarded whenever that task is run.</td>
<td>7 (days)</td>
</tr>
<tr>
<td>SSPPriorityTaskThreads=1</td>
<td>This option determines the number of tasks marked &quot;Priority&quot; that run at the same time. The sum of the <em>Priority</em> and <em>Normal</em> tasks should not exceed 4.</td>
<td>1</td>
</tr>
<tr>
<td>SSPNormalTaskThreads=1</td>
<td>This option determines the number of tasks not marked &quot;Priority&quot; that run at the same time. The sum of the <em>Priority</em> and <em>Normal</em> tasks should not exceed 4.</td>
<td>1</td>
</tr>
<tr>
<td>SSPProxyHost</td>
<td>This option lists the address of an http proxy server, if used. If set, this will override any JVM values used.</td>
<td>No default value. (Example: <code>proxy.example.com</code>)</td>
</tr>
<tr>
<td>SSPProxyPort</td>
<td>This option lists the port used for the http proxy server, if used. If set, this will override any JVM values used.</td>
<td>80</td>
</tr>
<tr>
<td>SSPNonProxyHosts</td>
<td>This option lists any servers that should be connected to directly and not through the proxy listed in <code>SSPProxyHost</code>.</td>
<td>No default value. (Example: `MyServer</td>
</tr>
</tbody>
</table>
The Oracle Content Server interface includes various pages that you can use to manage the publishing process of your Site Studio Web sites:

- "Site Studio Publisher Tasks" on page A-1
- "Task Info" on page A-12
- "Logs" on page A-13
- "View Log" on page A-13

A.1 Site Studio Publisher Tasks

This screen lists the publishing tasks that have been defined, and details about their operation.

*Figure A–1 Main Site Studio Publisher Screen*

<table>
<thead>
<tr>
<th>Element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pause Publishing</td>
<td>Click to stop the scheduling process from starting new tasks. This will pause scheduling new tasks, but will not pause tasks that are currently running. This button only appears when tasks are not paused.</td>
</tr>
<tr>
<td>Resume Publishing</td>
<td>Click to allow the scheduling process to start new tasks. This button only appears when tasks are paused.</td>
</tr>
<tr>
<td>Add New Task</td>
<td>Opens the Add New Task screen (see &quot;Add New Task/Edit Task&quot; on page A-2) to enter a task for publishing.</td>
</tr>
<tr>
<td>Refresh</td>
<td>Refreshes the list of tasks.</td>
</tr>
<tr>
<td>Description</td>
<td>The name of the task.</td>
</tr>
</tbody>
</table>
Add New Task/Edit Task

The page used for adding and editing tasks to run with publisher has many parts to it. It is here that you can specify how often you publish a site, what site to publish, and even which sections of the site to include or exclude as well as other filters. Triggers can be set as well.

This section covers the following topics:

- "Publisher Settings" on page A-2
- "Dates" on page A-6
- "Days" on page A-6
- "Run Times" on page A-7
- "Options" on page A-7
- "Include List" on page A-8
- "Exclude List" on page A-8
- "FilterSets" on page A-9
- "Triggers" on page A-10

**A.2 Add New Task/Edit Task**

The page used for adding and editing tasks to run with publisher has many parts to it. It is here that you can specify how often you publish a site, what site to publish, and even which sections of the site to include or exclude as well as other filters. Triggers can be set as well.

### A.2.1 Publisher Settings

The Publisher Settings section is used to set the basic information of the task. The location of the site to publish, the location of the output, the username and password, and so forth.
The Settings section has advanced settings, which can easily be seen when you click the **Show Advanced Settings** link. The advanced settings expand and display in the same section.

**Figure A–2 Publisher Settings Section (Basic)**

- **Description**: Enter a name of the task.
- **Manifest Url**: http://SERVER/INSTANCE/idcplg?idcService=SS_GET_SITE
- **Output Path**: output/SITE
- **Username**:
- **Password**: ************
- **Authentication**: BasicAuth
- **Show Advanced Settings**: Publish Now

**Figure A–3 Publisher Settings Section (Expanded)**

- **Description**: WeeklyTuesdayAutomated Publish
- **Manifest Url**: http://myvmware/stellent/idcplg?idcService=SS_GET_SITE_f
- **Output Path**: output/SITE
- **Username**: sysadmin
- **Password**: ************
- **Authentication**: BasicAuth
- **Hide Details**: Publish Now, Force Download, Force Analyze, Treat home page errors as critical, Treat manifest errors as critical, Use Cache Control, Use LastModified, Use SSLTag
- **Default Filename**: index.html
- **Page Extension**: html
- **User Agent**: Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.0)
- **Friendly Url Parameters**:
- **Dynamic Url Parameters**:
- **Additional Services**:
- **Soft Error Threshold**: 99
- **Hard Error Threshold**: 99
- **Delete Threshold**: 1
- **Soft Error Codes**: 403
- **Hard Error Codes**: 404,501,502
- **Ignore Error Codes**:

<table>
<thead>
<tr>
<th>Element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>Enter a name of the task.</td>
</tr>
</tbody>
</table>
### Element Description

<table>
<thead>
<tr>
<th>Element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manifest Url</td>
<td>Enter the URL for Site Studio Publisher to crawl. This is a specially formatted URL that gets information about the Web site to initiate the crawl.</td>
</tr>
<tr>
<td>Output Path</td>
<td>Enter the path to the local directory where content is downloaded. This is relative to the SSPHome location.</td>
</tr>
<tr>
<td>Username</td>
<td>Enter the user ID for password-protected sites.</td>
</tr>
<tr>
<td>Password</td>
<td>Enter the encoded password for password-protected sites.</td>
</tr>
<tr>
<td>Authentication</td>
<td>Select the type of authentication for the site.</td>
</tr>
<tr>
<td>Loginform</td>
<td>is used for Oracle Content Server 11g servers.</td>
</tr>
<tr>
<td>BasicAuth</td>
<td>is used for Oracle Content Server 10g servers.</td>
</tr>
<tr>
<td>ExtranetLook</td>
<td>is used for Oracle Content Server 10g servers with the ExtranetLook component.</td>
</tr>
<tr>
<td>NTLM</td>
<td>is used for Oracle Content Server 10g servers configured to use NTLM.</td>
</tr>
<tr>
<td>CustomForm</td>
<td>is used to customize a different form-based login.</td>
</tr>
<tr>
<td>Publish Now</td>
<td>Select to compare and publish only the marked, changed pages. If the altered content includes new links, Site Studio Publisher also publishes those links.</td>
</tr>
<tr>
<td>Show Advanced Settings</td>
<td>Click to display the Advanced Settings.</td>
</tr>
<tr>
<td>Hide Details</td>
<td>Click to collapse the Advanced Settings.</td>
</tr>
<tr>
<td>Force Download</td>
<td>Select to force a download of the Web page whether there are found changes or not.</td>
</tr>
</tbody>
</table>
| Force Analyze   | Select to force the analysis pass of the files listed in the filter parameters. If no files are listed in the filter, then all pages are analyzed.  
|                 | This is used only with filters.                                                                                                                 |
| Treat home page errors as critical | Select to specify if any error retrieving the home page aborts the job.                                                                           |
| Treat manifest errors as critical | Select to specify if any error retrieving any of the page URLs listed in the site manifest aborts the job.                                           |
| Use Cache Control | Select to compare the Max Age section property. Site Studio Publisher only selects and crawls those pages where the max-age value has not expired. |
| Use Last Modified | Select to use the value of the http header Last-Modified returned from the web server.                                                         
<p>|                 | The web server normally returns this header for all resources accessed with a weblayout URL. (This response header is not provided for a dynamic Site Studio page.) |
|                 | If useLastModified is checked, the Last-Modified value is re-submitted in an If-Modified-Since request header the next time that resource is retrieved. That then allows the web server to return a 304 - Not Modified response if the resource is unchanged. |
|                 | If this is not checked, the resource is downloaded and compared with the previously retrieved content.                                           |
|                 | You would only choose to do this if you found that your web server was returning unreliable results for Last-Modified.                       |</p>
<table>
<thead>
<tr>
<th>Element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use SSPETag</td>
<td>This option controls a proprietary mechanism used to identify changes to files retrieved by the GET_FILE service. Without this mechanism, file content is retrieved and compared with the previous version. Use of this mechanism allows the content server to return a 304 - Not Modified response and so avoid unnecessary downloads.</td>
</tr>
<tr>
<td>Default Filename</td>
<td>Specify the filename to be used for URLs where there is no filename explicitly specified.</td>
</tr>
<tr>
<td>Page Extension</td>
<td>The extension that is added to page urls that do not otherwise specify an extension. For example, a typical reference to a document in the dynamic site might look like: <a href="http://myServer/mySite/Section1/DocumentX">http://myServer/mySite/Section1/DocumentX</a> Where DocumentX is the dDocName of the target item in Content Server. For use on a static site, you must add an extension.</td>
</tr>
<tr>
<td>User Agent</td>
<td>Enables you to specify a value for the User-Agent http request header used by Site Studio Publisher when crawling the site.</td>
</tr>
<tr>
<td>Friendly Url Parameters</td>
<td>Specify a comma separated list of additional parameter names to honor. Site Studio Publisher supports a few built-in values. Friendly Urls already honor parameters ending in _NextRow or _dcPageNum and use the parameter name and value to construct the filename for the crawled page. If a URL Parameter affects the appearance of the page, then you need to capture a different copy of the page for each combination of parameter values.</td>
</tr>
<tr>
<td>Dynamic Url Parameters</td>
<td>Specify a comma separated list of additional parameter names to honor. Dynamic Urls using the GET_PAGE service already honor _dID, _dDocName, _RevisionSelectionMethod, and _Rendition. If a URL Parameter affects the appearance of the page, then you need to capture a different copy of the page for each combination of parameter values.</td>
</tr>
<tr>
<td>Additional Services</td>
<td>Select to allow additional services to be crawled. This element enables you to control the service calls that Site Studio Publisher attempts to crawl. There is built-in support for those services that are expected to generate meaningful static content: SS_GET_PAGE, GET_FILE, and GET_DYNAMIC_CONVERSION.</td>
</tr>
<tr>
<td>Soft Error Threshold</td>
<td>The number of soft errors allowed. If the defined number is exceeded, publishing fails.</td>
</tr>
<tr>
<td>Hard Error Threshold</td>
<td>The number of hard errors allowed. If the defined number is exceeded, publishing fails.</td>
</tr>
<tr>
<td>Delete Threshold</td>
<td>The number of objects that can be missing before the crawl is failed. If this number is exceeded, publishing fails.</td>
</tr>
<tr>
<td>Soft ErrorCodes</td>
<td>Enter a list of codes to specify individual error codes that are treated as soft errors. There are no wildcards allowed in this list, you must list each error code explicitly (separated by a comma).</td>
</tr>
</tbody>
</table>
A.2.2 Dates

The Dates section is used to specify a range of dates that the task will run.

If you select Manual, then the Days section options (see "Days" on page A-6) and the Run Time section options (see "Run Times" on page A-7) are greyed out. This is because selecting Manual will mean that the task is a task with will run only on-demand.

**Figure A–4 Dates Section**

<table>
<thead>
<tr>
<th>Element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hard Error Codes</td>
<td>Enter a list of codes to specify individual error codes that are treated as hard errors. There are no wildcards allowed in this list, you must list each error code explicitly (separated by a comma).</td>
</tr>
<tr>
<td>Ignore Error Codes</td>
<td>Enter a list of codes to specify individual error codes that are ignored. These codes will not affect the Site Studio Publisher crawl. There are no wildcards allowed in this list, you must list each error code explicitly (separated by a comma).</td>
</tr>
</tbody>
</table>

---

A.2.3 Days

The Days section is used to select if the task will run on certain days of the week, or on certain days of the month.

The options in this section will be available only if Range was selected in the Dates section (see "Dates" on page A-6).

**Figure A–5 Days Section**

<table>
<thead>
<tr>
<th>Element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manual / Range</td>
<td>Select to either have a task that runs only when commanded (manual) or to run at regular intervals at least once within a certain range of dates.</td>
</tr>
<tr>
<td>Start Date</td>
<td>The first date of a range that the task will run.</td>
</tr>
<tr>
<td>End Date</td>
<td>The final date of a range that the task will be run.</td>
</tr>
</tbody>
</table>
A.2.4 Run Times

The Run Times section is used to specify how often the task runs each day.

The options in this section will be available only if **Range** was selected in the Dates section (see "Dates" on page A-6).

**Figure A–6 Run Times Section**

<table>
<thead>
<tr>
<th>Element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Days of Week</td>
<td>Select to run the task on the selected days of the week. The task will run on the days selected that are within the range selected in the Dates section (see &quot;Dates&quot; on page A-6).</td>
</tr>
<tr>
<td>Days of Month</td>
<td>Select to run the task on particular days of the month. Days can be listed individually, by a range, or in a combination of both, for example: 1, 2, 7-13, 25. The word <strong>last</strong> is also a valid entry, to denote the last day of a month. It can be entered by itself or in combination as above.</td>
</tr>
</tbody>
</table>

A.2.5 Options

The Options section is where you set the priority and log levels of the task. You can also set an email address to receive notification after the task runs.

**Figure A–7 Options Section**

<table>
<thead>
<tr>
<th>Element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple Times Per Day</td>
<td>Select to have a task that runs either at regular intervals each day.</td>
</tr>
<tr>
<td>Once per Day</td>
<td>Select to have a task that runs only once per day. Enter a time in the box.</td>
</tr>
<tr>
<td>Hours Minutes</td>
<td>Select how often the task should run between the Start Time and the End Time.</td>
</tr>
<tr>
<td>Start Time</td>
<td>The time of day the task will start running.</td>
</tr>
<tr>
<td>End Time</td>
<td>The time of day the task will stop running.</td>
</tr>
</tbody>
</table>
A.2.6 Include List

The Include List section is used to specify which parts of the site (by URL) should be included in publishing.

If you list a URL for inclusion in this section, and it appears in a listed FilterSet type 'exclude', then it will not be included in the publishing.

**Figure A–8 Include List Section**

<table>
<thead>
<tr>
<th>Element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Priority</td>
<td>Select to mark this task as a priority task. This is used to ensure that the more vital tasks are run at a priority over the other tasks.</td>
</tr>
<tr>
<td>Log Level</td>
<td>Select the level of log information to write to the logs. Each item in the drop-down list includes the logging levels above it. For example, selecting INFO includes logs of not only INFO items but also WARN and ERROR.</td>
</tr>
<tr>
<td>Email Notification</td>
<td>Enter an email address to send notification of the task completion (or error) to.</td>
</tr>
</tbody>
</table>

A.2.7 Exclude List

The Exclude List is used to specify which parts of the site (by URL) should be excluded from publishing.

**Figure A–9 Exclude List Section**

<table>
<thead>
<tr>
<th>Element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hide Exclude List</td>
<td>Click to minimize the Exclude List section.</td>
</tr>
<tr>
<td>URL</td>
<td>Enter a URL to place on the Exclude List for Site Studio Publisher to avoid. This can be a regular expression.</td>
</tr>
<tr>
<td>Remove</td>
<td>Click to remove the URL from the list.</td>
</tr>
<tr>
<td>Add New Item</td>
<td>Click to add another field to enter an additional URL for the Include List.</td>
</tr>
</tbody>
</table>
A.2.8 FilterSets

Filtersets are controls used to transform content once it is downloaded from crawling, before it is published. Filters are applied in the order that they appear in the Filterset.

**Figure A–10 FilterSets Section**

<table>
<thead>
<tr>
<th>Element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Remove</td>
<td>Click to remove the URL from the list.</td>
</tr>
<tr>
<td>Add New Item</td>
<td>Click to add another field to enter an additional URL for the Exclude List.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Add New Filterset</td>
<td>Expands to display an additional set of fields to enter information for another filterset.</td>
</tr>
<tr>
<td>Move Up</td>
<td>Move Down</td>
</tr>
</tbody>
</table>
| Type    | The `type` attribute specifies the context in which the filterset should be invoked. The types available are:  

  - **transform-content**: for URLs that match this filterset, the child filter elements are applied during download, transforming the content in the manner specified by these filters.  
  - **transform-link**: for URLs that match this filterset, the child filter elements are applied during download. In this case, the filter elements are only applied to links found in the current downloaded file, not to the entire content.  
  - **exclude**: for URLs that would normally be included, match this filterset, and the URL matches, content is not downloaded. |
A.2.9 Triggers

Triggers enable the replication engine to run commands either before a package is downloaded or after the download is complete. The types of triggers available are cmd, http-post, http-get, and http-soap.
### Figure A–11  Triggers Section

<table>
<thead>
<tr>
<th>Element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Show Triggers</td>
<td>Displays the fields required to enter the most basic information for a trigger.</td>
</tr>
<tr>
<td>Add New Trigger</td>
<td>Displays additional information to add multiple types of triggers.</td>
</tr>
<tr>
<td>Username</td>
<td>The username used for authentication. You cannot specify different authentications for each trigger. All triggers in a task must have the same credentials and authentication method.</td>
</tr>
<tr>
<td>Password</td>
<td>The password used for authentication. You cannot specify different authentications for each trigger. All triggers in a task must have the same credentials and authentication method.</td>
</tr>
<tr>
<td>Authentication</td>
<td>Select to authenticate the user either with basic authentication or with a custom form.</td>
</tr>
<tr>
<td>Move Up</td>
<td>Move Down</td>
</tr>
<tr>
<td>Type</td>
<td>Specifies the type of trigger.</td>
</tr>
<tr>
<td>Command</td>
<td>Command line arguments to pass through the trigger. Only used if the type is set to <code>cmd</code>.</td>
</tr>
<tr>
<td>Command Data</td>
<td>The path, relative to the task’s output path, of the file to be uploaded. Only used if the type is set to <code>http-post</code> or <code>http-soap</code>.</td>
</tr>
<tr>
<td>Response File</td>
<td>Specifies the full path to a file to capture the response.</td>
</tr>
<tr>
<td>SOAPAction</td>
<td>Sets the value for the SOAPAction HTTP request header field. Only used if the type is set to <code>http-soap</code>.</td>
</tr>
<tr>
<td>Run the trigger even if the crawl failed</td>
<td>Specifies whether the command should be run, even if the job encountered errors.</td>
</tr>
<tr>
<td>Run the trigger if there were no changes</td>
<td>If checked, the trigger will run even if there has been no change in content.</td>
</tr>
<tr>
<td>Ignore trigger failure</td>
<td>Specifies whether further trigger execution will happen. If checked, further triggers will run, even if the current trigger fails.</td>
</tr>
<tr>
<td>Log the response</td>
<td>Specifies whether to write the response to the log.</td>
</tr>
</tbody>
</table>
A.3 Task Info

The Task Info screen is used to view information about the most recent task initiating the specified Site Studio Publisher crawl.

**Figure A–12 Task Info screen**

<table>
<thead>
<tr>
<th>Element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task ID</td>
<td>The ID number of the task.</td>
</tr>
<tr>
<td>Description</td>
<td>The description of the task.</td>
</tr>
<tr>
<td>Source Url</td>
<td>The source URL of the site being published.</td>
</tr>
<tr>
<td>Start Time</td>
<td>The time the task started.</td>
</tr>
<tr>
<td>End Time</td>
<td>The time the task completed.</td>
</tr>
<tr>
<td>Status</td>
<td>The completion status of the task.</td>
</tr>
<tr>
<td>Total</td>
<td>Total number of URLs (including files such as CSS, JavaScript, and so forth) that make up the crawled site.</td>
</tr>
<tr>
<td>Modified</td>
<td>Number of URLs that were modified since the most recent crawl of the site.</td>
</tr>
<tr>
<td>Added</td>
<td>Number of URLs added since the most recent crawl of the site.</td>
</tr>
<tr>
<td>Deleted</td>
<td>Number of URLs deleted since the most recent crawl of the site.</td>
</tr>
<tr>
<td>Examined</td>
<td>The number of URLs followed while crawling the site.</td>
</tr>
<tr>
<td>Redirected</td>
<td>The number of URLs redirected while crawling the site.</td>
</tr>
<tr>
<td>Downloaded</td>
<td>The number of examined links that were actually downloaded.</td>
</tr>
</tbody>
</table>
A.4 Logs

Lists the job summary and different log files generated for the specific task.

**Figure A–13  Log Files**

<table>
<thead>
<tr>
<th>Element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not found</td>
<td>The number of links followed that received an error.</td>
</tr>
<tr>
<td>Hard Error Count</td>
<td>Total number of &quot;hard&quot; errors found. The number in parenthesis is the maximum allowed.</td>
</tr>
<tr>
<td>Soft Error Count</td>
<td>Total number of &quot;soft&quot; errors found. The number in parenthesis is the maximum allowed.</td>
</tr>
<tr>
<td>View Log</td>
<td>Opens the log file for the task in the View Log screen (see &quot;View Log&quot; on page A-13).</td>
</tr>
</tbody>
</table>

### A.5 View Log

The View Log screen is where the log file can be viewed. When the log file is long, the Previous and Next buttons are available to page through the file.

<table>
<thead>
<tr>
<th>Element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task ID</td>
<td>The ID number of the task.</td>
</tr>
<tr>
<td>Description</td>
<td>The description of the task, as entered in the Publisher Settings section on the Add/Edit Task page (see &quot;Publisher Settings&quot; on page A-2).</td>
</tr>
<tr>
<td>Log Files</td>
<td>A link for each date and time that the task was run. Each link opens the log file for that specific run on the View Log page (see &quot;View Log&quot; on page A-13). If the log file is very long, it will be split across multiple files, and each file will have a link with the same timestamp.</td>
</tr>
</tbody>
</table>
**Figure A–14 View Log Screen**

<table>
<thead>
<tr>
<th>Element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Previous</td>
<td>Takes you to the previous page in the log.</td>
</tr>
<tr>
<td>Next</td>
<td>Takes you to the next page in the log.</td>
</tr>
</tbody>
</table>
Index

A
absolute URL, 1-2
Add New Task screen, A-2
   Dates section, 2-5, A-6
   Days section, 2-6, A-6
   Exclude list section, 2-7, A-8
   FilterSets section, 2-7, A-9
   Include List section, 2-7, A-8
   Options section, 2-6, A-7
   Publisher Settings section, 2-2, A-2
   Run Times section, 2-6, A-7
   Triggers section, 2-8, A-10
appended parameters, 1-3

C
cmd trigger, 2-9
Content Server environment, 1-2
Contributor
   Publish Now feature, 2-2

D
Dates section, 2-5, A-6
Days section, 2-6, A-6

E
Edit Task screen, A-2
   Dates section, 2-5, A-6
   Days section, 2-6, A-6
   Exclude list section, 2-7, A-8
   FilterSets section, 2-7, A-9
   Include List section, 2-7, A-8
   Options section, 2-6, A-7
   Publisher Settings section, 2-2, A-2
   Run Times section, 2-6, A-7
   Triggers section, 2-8, A-10
editing a task
   scheduling, 2-5, 2-6
Exclude List section, 2-7, A-8

F
filter
   re-applying, 2-3
filter application, 2-7
filter element, 2-7
filters, 2-7, A-9
Filters section, 2-7, A-9
filterset
   evaluation, 2-7
FilterSets section, 2-7, A-9

H
HTML
   handling in Publisher, 1-2
http-get, 2-9
http-post, 2-9
http-soap, 2-9

I
Include List section, 2-7, A-8
installation, 1-2

J
JavaScript
   general patterns, 1-3
   generated by Site Studio, 1-3
   handling in Publisher, 1-3
   parameters, 1-4
   patterns used by Publisher, 1-3
   Site Studio-specific patterns, 1-3, 1-4

L
Logs screen, A-13

M
minimum requirements, 1-4

O
Options section, 2-6, A-7

P
properties
   URL Directory Name, 1-4
URL Page Name, 1-4
Publisher control
  Publish Now, 2-2
  Publish This Page, 2-2
Publisher Settings section, 2-2, A-2
Publisher Tasks screen, A-1

R
relative URL, 1-2, 1-3
Run Times section, 2-6, A-7

S
screen
  Add New Task, A-2
  Edit Task, A-2
  Publisher Tasks, A-1
Site Studio Contributor
  Publish Now feature, 2-2
SOAP, 2-9
system requirements, 1-4

T
Task Info screen, A-12
task scheduling, 2-5, 2-6
trigger, 2-8
cmd, 2-9
  http-get, 2-9
  http-post, 2-9
http-soap, 2-9
Triggers section, 2-8, A-10

U
URL
  absolute, 1-2
  relative, 1-2, 1-3
URL Directory Name property, 1-4
URL Page Name property, 1-4

V
View Log screen, A-13