

# Publish to Solr Application Bundle

- [Configuration](#)
- [Example Configuration](#)
- [Edit Xsl](#)
  - [Add metadata field](#)
  - [Change the document ID](#)
  - [Advanced Edit](#)

The Publish to Solr application sends document feeds, to the Solr index update servlet, of metadata and content of files extracted by Aspire connectors. The feed to Solr can be customized by editing the XSL transformation file provided by the user.

Publish to Solr	
Factory Name	com.searchtechnologies.aspire:app-publish-to-solr
subType	default
Inputs	<a href="#">AspireObject</a> from a connector's subjob with metadata and content extracted from a specific file/folder.
Outputs	An XML transformation of the <a href="#">AspireObject</a> sent to the Solr's xmlfeed URL.
Versions	3.0, 3.1
Type Flags	job-input

## Configuration

This section lists all configuration parameters available to configure the Publish to Solr component.

Property	Type	Default	Description
SolrNoUrl	boolean	true	Indicates if the publisher must use a Url or build one from the host and port entered.
SolrPort	integer	8983	Solr port where to send the feeds
solrHost	string	none	Solr hostname or IP adress. e.g. solr.domain.com
SolrUrl	string	none	Complete Url where the feeds are going to be send. e.g. http://localhost:8983/solr/core/update
aspireToSolrXsl	string	\${appbundle.home}/config/xsl/aspireToSolr.xsl	Location of the XSL to transform the job data to a Solr feed. See <a href="#">Edit Xsl</a> .
maxResults ( <a href="#">2.1 Release</a> )	int	1000000	( <a href="#">Index dump</a> ) How many documents can be fetched by the search engine for the same query
pageSize ( <a href="#">2.1 Release</a> )	int	10000	( <a href="#">Index dump</a> ) How many documents to fetch per page
urlField ( <a href="#">2.1 Release</a> )	String	displayUrl	( <a href="#">Index dump</a> ) Field used to store the url in the search engine
idField ( <a href="#">2.1 Release</a> )	String	id	( <a href="#">Index dump</a> ) Field used to store the id in the search engine.
timestampField ( <a href="#">2.1 Release</a> )	String	submitTS	( <a href="#">Index dump</a> ) The name of the timestamp field holding the index timestamp of every document.

## Example Configuration

```
<application config="com.searchtechnologies.aspire:app-publish-to-gsa">
  <properties>
    <SolrNoUrl>true</SolrNoUrl>
    <SolrHost>localhost</SolrHost>
    <SolrPort>8983</SolrPort>
    <aspireToSolrXsl>${appbundle.home}/config/xsl/aspireToSolr.xsl</aspireToSolrXsl>
    <debug>false</debug>
  </properties>
</application>
```



Any optional properties can be removed from the configuration to use the default value described on the table above.

## Edit Xsl

The default XSL transformation file can be found in [AspireToSolr.xsl](#).

The default transformation XSL file provided by the publisher expects metadata as described in [Connector AspireObject Metadata](#).

## Add metadata field

To add a new metadata field extracted by an Aspire Connector add an XSL element under the **<doc>** tag.

```
<field name="metafieldNameInSolr_t">
  <xsl:value-of select="metafieldNameFromAspireObject" />
</field>
```

Notice that the dynamic field `_t` is being used by default. If you have a Solr schema that supports your field, then just enter the field name as defined in the schema.

## Change the document ID

The **id** of a Solr document is used to uniquely identify a file in the index. By default, Publish To Solr will use the following fields from the Aspire document in order of precedence (if one is missing, then the next will be used):

- `fetchUrl`
- `url`
- `displayUrl`
- `id`

If you want to change this behavior, edit or create a new XSL file which has the following element:

```
<field name="id">
  <xsl:value-of select="idFieldNameFromAspireObject" />
</field>
```

## Advanced Edit

More advanced changes can be accomplished reading the [Solr Update XML Messages](#) wiki.