

[\[ back to Help Center \]](#)

## Using Data Extract

---

- [What Data Extract is](#)
- [What to use Data Extract for](#)
- [Combined patent office and DWPI results](#)
- [Overview of the Data Extract control panel](#)
- [How to create a Data Extract](#)
- [Data Extract formats available \(with sample files\)](#)
- [Field Codes/Tags used in selected Data Extracts formats](#)
- [DTD \(Document Type Definition\) for XML format](#)
- [Pricing for Data Extracts containing only patent office fields](#)
- [Pricing for Data Extracts containing any DWPI fields](#)

### What Data Extract is:

Data Extract is a tool that extracts over 50 bibliographic fields — as well as the full text — of patent office and Derwent World Patents Index (DWPI) data and delivers it all to your desktop in file formats suitable for use in many popular applications like Derwent Analytics<sup>SM</sup>, Microsoft's Excel®, ISI ReasearchSoft's EndNote®, Reference Manager®, ProCite® and most standard text editors.

[\[back to top\]](#)

---

### What to use Data Extract for:

Use Data Extract to export patent records from your Work Files or Result Sets for use in other applications.

[\[back to top\]](#)

---

### Combined patent office and DWPI result:

When working with a patent office Result Set, you can select DWPI fields to be extracted. In this situation, the DWPI fields in your Data Extract will come from the DWPI record corresponding to the patent record you are extracting.

Conversely, when working with a DWPI Result Set, you can select patent office fields to be extracted. When you do this, the patent office fields in your Data Extract come from the patent office record corresponding to the basic patent in the patent family that each DWPI record describes.

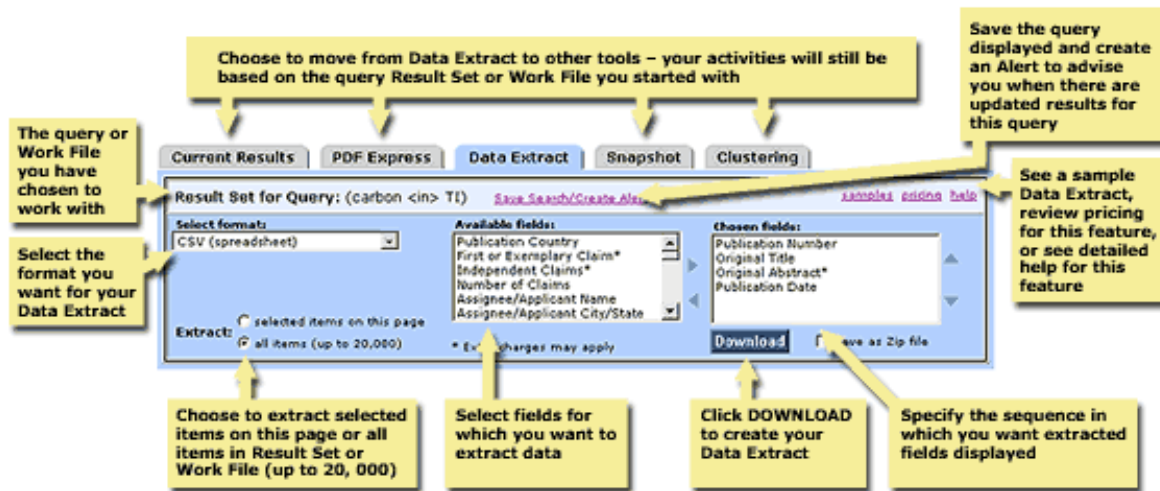
**Note:** If you are creating an extract for use in ISI ReasearchSoft's EndNote®, Reference Manager®, or ProCite®, you can create your extract from your DWPI search results or Work File, but you cannot extract DWPI fields. The fields you extract will be from the base patents corresponding to the selected DWPI records.

[\[back to top\]](#)

---

### Overview of the Data Extract control panel:

The following is a high-level overview of the Data Extract control panel:



[\[back to top\]](#)

### How to create a Data Extract:

Initiate the creation of a Data Extract from any query Result Set or any Work File:

1. In **Select format** box, indicate the format you want for your extract. Formats are described following.
2. Choose to **Extract** either **Selected items on this page** or **All items (up to 20,000)**.
3. If you are choosing to extract Selected Items from this Page, select those items from your Result Set.
4. If you have chosen one of the CSV or tagged file formats, the **Chosen fields** box shows a default set of fields based on the fields you have chosen to display on your Result Set — you can remove these fields if you choose.

In the **Available fields** box, highlight any fields you want added to your extract and click the blue arrow that points to the right to add them to the **Chosen fields** box.

To remove a field from the extract parameters, highlight it in the **Chosen fields** box and click the blue arrow that points left.

5. If you have chosen RIS format for EndNote, RefMan or ProCite, the required fields are shown in the **Chosen fields** box, you cannot remove any of the required fields and there are no additional fields that can be selected.
6. If you have chosen Derwent Analytics, the **Chosen fields** box already displays the required fields and you cannot remove any of them.

In the **Available fields** box, highlight the fields you want added to your extract and click the blue arrow that points to the right to add them to the **Chosen fields** box.

To remove an optional field from the extract parameters, highlight it in the **Chosen fields** box and click the blue arrow that points left.

7. If you have chosen one of the XML options: in the **Available fields** box, choose either **Bibliographic** or **Full text with bibliographic**.
8. For CSV, tagged and Derwent Analytics downloads, optionally choose to **zip** your files for a faster download. Both XML options automatically download zipped.

For RIS downloads, you can **zip** your files or choose to directly export into EndNote, Reference Manager or ProCite. To directly export, when you are prompted to either **Open** or **Save** your file, choose **Open**. You will then be asked which program you want to open the data with. Select the program (Endnote, Reference Manager or ProCite) then the integrated Direct Export function completes the process.

9. Click **DOWNLOAD** to create Data Extract.

10. Wizard screens will prompt you to open your Data Extract file directly or save it to a disk.

[\[back to top\]](#)

#### Data Extract formats available with file samples:

Format	Description	File Type
CSV	<p>Comma separated values, with multiple occurrences of a field shown all in one field.</p> <p>In most PC operating systems, .csv files are defaulted to open in Microsoft Excel or another spreadsheet program. If your spreadsheet program does not open automatically, open the spreadsheet program first and then open the .csv file. These files can also be opened in a regular text editor.</p> <p>The fields you requested for your extract will be shown in the order you requested with the data items separated by commas. The field names will be shown once at the top of each column.</p> <p>A list of all field codes and their tags is shown following.</p>	<p>.csv</p> <p><a href="#">Sample file</a></p> <p><a href="#">Graphical example</a></p>
CSV Multi-Fields/Codes	<p>Comma separated value with field codes, with multiple occurrences of a field shown as separate fields.</p> <p>In most PC operating systems, .csv files are defaulted to open in Microsoft Excel or another spreadsheet program. If your spreadsheet program does not open automatically, open the spreadsheet program first and then open the .csv file. These files can also be opened in a regular text editor.</p> <p>The fields you requested for your extract will be shown in the order you requested with the data items separated by commas. Each field will begin with a field code that represents the patent data shown in that field. Multiple occurrences of a field will be shown as separate fields and the field code will be suffixed with a sequence number (i.e., PA1 will be the first Patent Assignee listed, PA2 will be the second Patent Assignee listed, etc.).</p> <p>If there are not multiple occurrences in a field that could potentially have multiple occurrences, then the sequence number will still be used but there will only be one occurrence of it (i.e., there will be a PA1 without a PA2).</p> <p>This format is not recommended for use with the following fields:</p> <ul style="list-style-type: none"> <li>● Abstract</li> <li>● First or Exemplary Claim</li> <li>● Independent Claims</li> <li>● Other References</li> <li>● Assignee/Inventor</li> <li>● Inventor City/State</li> <li>● Attorney Name</li> <li>● Field of Search</li> <li>● Examiner: Primary</li> <li>● Examiner: Assistant</li> </ul> <p>A list of all field codes and their tags is shown following.</p>	<p>.csv</p> <p><a href="#">Sample file</a></p> <p><a href="#">Graphical example</a></p>
Tagged	<p>Tagged entries are suitable for import into a tagged file application, with multiple occurrences of a field shown all in one field.</p> <p>The fields you requested for your extract will be shown in the order you requested. Each field will begin with a tag that identifies the patent data shown in that field.</p> <p>A list of all field codes and their tags is shown following.</p>	<p>.tag</p> <p><a href="#">Sample file</a></p> <p><a href="#">Graphical example</a></p>

Tagged Multi-Fields	<p>Tagged entries suitable for import into a tagged file application, with multiple occurrences of a field shown as separate fields.</p> <p>The fields you requested for your extract will be shown in the order you requested. Each field will begin with a tag that identifies the patent data shown in that field. Multiple occurrences of a field will be shown as separate fields and the tag will be suffixed with a sequence number (i.e., PA1 will be the first Patent Assignee listed, PA2 will be the second Patent Assignee listed, etc.).</p> <p>If there are not multiple occurrences in a field that could potentially have multiple occurrences, then the sequence number will still be used but there will only be one occurrence of it (i.e., there will be a PA1 without a PA2).</p> <p>A list of all field codes and their tags is shown following.</p>	<p>.tag</p> <p><a href="#">Sample file</a></p> <p><a href="#">Graphical example</a></p>
EndNote, RefMan, ProCite (RIS)	<p>This file was designed primarily for use in ISI ReasearchSoft's EndNote®, Reference Manager® and ProCite®.</p> <p>The fields for this extract include a default set of 10 (shown following). No additional/optional fields can be chosen.</p> <ul style="list-style-type: none"> <li>● Publication Number</li> <li>● Title</li> <li>● Abstract</li> <li>● Inventor</li> <li>● Assignee</li> <li>● Application Date</li> <li>● IPC Codes</li> <li>● National Class</li> <li>● Domestic Citations</li> <li>● Image Information</li> </ul> <p>A list of all field codes and their tags is shown following.</p>	<p>.ris</p> <p><a href="#">Sample file</a></p> <p><a href="#">Graphical example</a></p>
Patent office data for Derwent Analytics	<p>This file was designed specifically for exporting patent office data for use in Derwent Analytics. Derwent Analytics is a data mining and visualization tool from Thomson that is designed for the desktop. <a href="#">More about Derwent Analytics.</a></p> <p>The fields for this extract include a default set of 20 (shown following) required for the Derwent Analytics application. When you are creating your Data Extract, there are more than 30 additional/optional fields you can choose to add to your extract.</p> <ul style="list-style-type: none"> <li>● Abstract</li> <li>● Application Country</li> <li>● Application Date</li> <li>● Application Number</li> <li>● Assignee/Applicant Country</li> <li>● Assignee/Applicant Name</li> <li>● Claim - First or Exemplary</li> <li>● Field of Search</li> <li>● Inventor Country</li> <li>● Inventor Name</li> <li>● Main IPC (1st 4 Digits)</li> <li>● Main IPC</li> <li>● Number of Claims</li> <li>● Priority Country</li> <li>● Priority Date</li> <li>● Priority Number</li> <li>● Publication Country</li> <li>● Publication Date</li> <li>● Publication Number</li> <li>● Title</li> </ul> <p>A list of all field codes and their tags is shown following.</p>	<p>.tag</p> <p><a href="#">Sample file</a></p> <p><a href="#">Graphical example</a></p>
DWPI data for Derwent Analytics	<p>This file was designed specifically for exporting DWPI data for use in Derwent Analytics. Derwent Analytics is a data mining and visualization tool designed for the desktop. <a href="#">More about Derwent Analytics.</a></p> <p>The fields for this extract include a default set of 16 (shown following) required for the Derwent Analytics application. When you are creating your Data Extract, there are more than 40 additional/optional fields you can choose to add to your extract.</p>	<p>.tag</p> <p><a href="#">Sample file</a></p> <p><a href="#">Graphical example</a></p>

	<ul style="list-style-type: none"> <li>● Derwent Abstract</li> <li>● Derwent Accession Number</li> <li>● Derwent Assignee</li> <li>● Derwent Classes</li> <li>● Derwent Family</li> <li>● Derwent Inventor</li> <li>● Derwent IPC Codes</li> <li>● Derwent Main Class</li> <li>● Derwent Main IPC</li> </ul>	<ul style="list-style-type: none"> <li>● Derwent Manual Codes</li> <li>● Derwent Title</li> <li>● Derwent Update</li> <li>● Priority Country</li> <li>● Priority Date</li> <li>● Priority Number</li> <li>● Publication Date</li> <li>● Publication Number</li> </ul>	
	A list of all field codes and their tags is shown following.		
XML: All patents in one file	XML is a flexible, commonly used markup language. XML files downloaded from Data Extract can be used in a variety of applications and programs.		.xml inside a .zip <a href="#">Sample file</a>
OR	Field codes and tags are not used for XML extracts. The XML file is constructed according to the DTD (Document Type Definition). Use this link to download the <a href="#">DTD</a> .		<a href="#">Graphical example</a>
XML: One file per patent	When choosing XML, you have the option of extracting full-text fields (which includes bibliographic fields) or just bibliographic fields. If you choose bibliographic fields only, the following fields will <i>not</i> be included in your extract:	<ul style="list-style-type: none"> <li>● federalResearchStatement</li> <li>● backgroundOfInvention</li> <li>● briefDescriptionOfDrawings</li> <li>● generalDescription</li> <li>● embodiments</li> <li>● claims</li> </ul>	OR .xml per patent inside a .zip <a href="#">Sample file</a> <a href="#">Graphical example</a>

[\[back to top\]](#)

#### Field code tags used in selected Data Extracts formats:

Field Code/Tag	Field Name
AC	Application Country
AD	Application Date
AN	Application Number
PAx	Assignee
ASx	Assignee City/State
CAx	Assignee Country
AX	Assistant Examiner
AG	Attorney
CUx	US Citations
OCx	Other Citations
CC	Company Code
ABDW	Derwent Abstract
ACCDW	Derwent Accession Number
ASDW	Derwent Assignee
CLDW	Derwent Classes
FDW	Derwent Family
INDW	Derwent Inventor
ICDW	Derwent IPC Codes
CMDW	Derwent Main Class
ICMDW	Derwent Main IPC
MCDW	Derwent Manual Codes
TIDW	Derwent Title
TTDW	Derwent Title Terms
UPDW	Derwent Update
ICx	IPC Codes

NCx	UPC Codes
ECx	ECLA Codes
DN	Designated States National
DR	Designated States Regional
FIx	Family Information
FS	Field of Search
FCx	Forward References
FRx	Foreign References
INx	Inventor
ISx	Inventor City/State
CIx	Inventor Country
OUx	Original UPC
PD	Patent Date
PC	Patent Country
PN	Patent Number
PX	Primary Examiner
CP	Priority Country
DP	Priority Date
PR	Priority Number
TI	Title

**NOTE:** "x" is a sequence number added when you select to download multiple occurrences of the field as separate fields. If you selected the option that downloads multiple occurrences in a field and there are not multiple occurrences in that field for that patent, then the sequence number will still be used but there will only be one occurrence of it (i.e., there will be a PA1 without a PA2). If you select to download the data as one field, the sequence number is not included in the field code.

[\[back to top\]](#)

#### **DTD (Document Type Definition) for XML format:**

Use this link to download the DTD that defines the format used for files extract in XML.

[DTD](#)

[\[back to top\]](#)

#### **Pricing for Data Extracts containing only patent office fields:**

These prices apply to Data Extracts that contain only patent office fields — even if you are working with the results of a DWPI search.

For Delphion Unlimited subscribers, Data Extract is free for extracts that do not exceed 500 records. For extracts exceeding 500 records, the following charges apply:

- \$10 per additional set of 500 non-XML records that includes abstract, claims or both in the extract (NOTE: abstract and claims are automatically included when extracting for Derwent Analytics; abstract is automatically included in extracts for ISI ReasearchSoft's EndNote®, Reference Manager®, and ProCite®).
- \$15 per additional set of 500 records if XML is the download format (either bibliographic fields or full text).
- \$5 per additional set of 500 records for all other fields in standard (non-XML) formats.

For Delphion Premier subscribers, each set of 500 records is charged as follows:

- \$10 per set of 500 non-XML records that includes abstract, claims or both in the extract

(NOTE: abstract and claims are automatically included when extracting for Derwent Analytics; abstract is automatically included in extracts for ISI ReasearchSoft's EndNote®, Reference Manager®, and ProCite®).

- \$15 per set of 500 records if XML is the download format (either bibliographic fields or full text).
- \$5 per set of 500 records for all other fields in standard (non-XML) formats.

#### Pricing for Data Extracts containing DWPI fields:

These prices apply when your Data Extract contains one or more DWPI fields — even if you are working with the results of a patent office search.

Charges are based on the number of records downloaded in the extract — a set of 500 records is called a *block*.

Charges are calculated by first determining the number of complete 500-record blocks. Then, if the remainder is more than 200 records, you will be charged for one additional 500-record block. If the remainder is less than 200 records, then this is the number of individual records you will be charged for in addition to the number of blocks.

The actual per-record and per-block charges differ depending on whether you are using DWPI on a pay-per-use basis or you are a DWPI subscriber through your group/contract account.

	Transactional, per record (see note following)	500 record block
Delphion Unlimited or Premier user, using pay-per-use DWPI	\$ 3	\$ 600
DWPI on Delphion subscriber, through a group/contract account	\$ 1	\$ 200

**Note:** Several patent office records could correspond to the same DWPI record. If you are working with the results of a patent office search and this occurs, you will only be charged for unique DWPI records.

[\[back to top\]](#)

