Elsevier EMBASE XML Data Files

Delivery process and file description

Version 3

November 2025

Table of Contents

Overview

Elsevier's EMBASE is a comprehensive biomedical research database with extensive and deep indexing of full-text content according to the authoritative life science thesaurus EMTREE.

This user guide provides information on the process of the EMBASE full corpus in XML format (flat file) and a description of the structure and elements in the downloaded data files.

1. Flat file delivery process

EMBASE flat file delivery grants users access to the most up-to-date and complete version of the EMBASE full corpus as XML files. The zipped files are downloaded via sFTP from an AWS S3 bucket. The available corpus includes the EMBASE, EMBASE Classic and EMBASE French Collections and is updated after each Emtree release. The full corpus and data updates are available in two forms based on user subscription. In one form, records are enriched with EMTREE parent terms and in the other they are not.

For details on flat file delivery, please see the sFTP User Guide.

1.1. Frequency of updates

Daily updates made to EMBASE are additionally available as a delta set on the same AWS S3 bucket enabling consumption of changes on a particular day. The daily update folder separates "updates" and "deletions" and remains in the bucket until the next two full corpus become available.

In this way, users will always have an up-to-date data set by first retrieving the most recent full corpus (provided after each Emtree release) and then supplementing this with daily updates.

The user defines the frequency of the updates. If updates are not done daily, however, the downloaded data set may no longer be in line with the most current data set on embase.com. For example, if the daily update is done weekly, then the full set in hands of the user will be one week behind the data set available on embase.com.

A downloaded data set can then be completed by downloading and integrating all relevant daily updates – if still available. If the files are no longer available, the full corpus and any subsequent updates must be downloaded.

1.2. Updates to EMTREE

With every new release of EMTREE, the EMBASE flat files are reprocessed to reflect the updated enrichment (a process called backposting). New releases of EMTREE occur 3 times a year.

2. Structure of the XML data files

The XML data files that can be downloaded from EMBASE consist of a list of literature records wrapped in the element bibdataset. Each record of the list is contained in an element bibrecord and includes bibliographic, non-bibliographic, abstract and indexing information organized into 4 broad elements:

Schema location	Schema information can be found at the start of XML.
item-info	Contains the **non-bibliographic information about a record**, such as date of creation, identifiers, and copyright information.
head	Encompasses the **abstract and indexing information** of the record, including authors and affiliations, type of document, source, and indexing terms.
tail	Includes a single child element that lists the number of bibliographic records listed in the document of the record. The full bibliography (references cited) of the document in a record is not provided.
explosions	XML files enriched with EMTREE terms include this element, which provides all parent elements of each subject index term based on EMTREE hierarchies.

Note that at the start of XML the following schema information can be found:

xmlns:ait=http://www.elsevier.com/xml/ani/ait

xmlns:ce=http://www.elsevier.com/xml/ani/common

xmlns:xsi=http://www.w3.org/2001/XMLSchema-instance

xmlns=http://www.elsevier.com/xml/ani/ani

xmlns:ani=http://www.elsevier.com/xml/ani/ani

xsi:schemaLocation= http://www.elsevier.com/xml/ani/ani http://www.elsevier.com/xml/ani/embase_daas.xsd

2.1. Record information in item-info

The child elements of item-info provide information about the copyright, history and, in the case of a non-Elsevier origin (e.g., MEDLINE), the external source (external-source) of a record. Also included are a range of record and document identifiers, like the digital object identifier or DOI (itemidlist).

Example of information included in item-info

```
<item-info>
  <copyright type="Elsevier">Copyright 2022 Elsevier B.V., All rights reserved.</copyright>
  <itemidlist>
    <ce:doi>10.3748/wjg.v28.i1.140</ce:doi>
    <itemid idtype="PUI">2016661994</itemid>
    <itemid idtype="CAR-ID">938170524</itemid>
    <itemid idtype="MEDL">35125824</itemid>
    <itemid idtype="COLLECTIONS">/opt/embase/workarea/embase_fabrication/fabrication/current/daily/20220201194528_00000/embasecom_new_20220131_20014_421_2.zip</itemid>
    <itemid idtype="COLLECTIONS">/opt/embase/workarea/embase_fabrication/fabrication/current/daily/20220207194527_00000/embasecom_out_full_20220204_16934_651_1.zip</itemid>
    <itemid idtype="COLLECTIONS">/opt/embase/workarea/embase_fabrication/fabrication/current/daily/20220211194527_00000/embasecom_out_full_update_20220210_16935_660_1.zip</itemid>
  </itemidlist>
  <history>
    <date-created year="2022" month="02" day="04"/>
    <date-aip-loaded year="2022" month="02" day="01"/>
    <date-loaded year="2022" month="02" day="01"/>
    <date-loaded year="2022" month="02" day="07"/>
    <record-update-date="2022" month="02" day="07"/>
    <record-update-date="2022" month="05" day="09"/>
  </history>
  <dbcollection>EMBASE</dbcollection>
  <dbcollection>MEDLINE</dbcollection>
  <external-source>In-Process</external-source>
</item-info>
```

2.2. Abstract and indexing information in head

The element head encompasses the bulk of information contained in an EMBASE record, including the citation information and the deep and extensive indexing of a document. This information is organized into the child elements described below.

citation-info	Encompasses descriptive information about the document in a record, such as language, document type, keywords from the authors.
citation-title	Lists the original title of the document in a record and any translations.
author-group	Provides information about the authors of the document in a record, organized by author affiliation (i.e., one or more authors that share a common affiliation). Authors publishing under one name (e.g., consortia, collaborations) are also included in this element, as the name of the collaboration plus the names of individual members.
correspondence	Gives details of the corresponding author on the document in a record.
abstracts	Provides the original and any other abstracts or translations of the abstract generated for the document in a record.
source	Includes all details about the source of the document in a record, be it a book series, a conference, a journal or other. This element includes ISBN, volume, issue, and pages for reference, as well as publication date, publisher name and location. Also listed are location and dates of conferences.
enhancement	Contains a wealth of additional indexing or classification added to a record, such as controlled subject index terms organized by term type, tradenames included in the document and manufacturers of devices and drugs mentioned in the document. Also includes EMTREE terms describing the context in which index terms describe the content of the document. The attribute *status* provides a 3-digit code that indicates the indexing status of a record as either intermediate (001) or final (002). The attribute *type* indicates the type of indexing performed; 17 for automatic indexing and 8 for manual indexing. In EMBASE, after automatic indexing generates a record of intermediate status, it is taken to final status by manual indexing.

Example of data included in citation-info and citation-title

```
<head>
  <citation-info>
    <citation-type code="ar"/>
    <citation-language xml:lang="en"/>
    <abstract-language xml:lang="en"/>
    <author-keywords>
      <author-keyword>Cytokine</author-keyword>
      <author-keyword>Hepatitis C virus</author-keyword>
      <author-keyword>Hepatocellular carcinoma</author-keyword>
      <author-keyword>Sustained virologic response</author-keyword>
      <author-keyword>Tumor necrosis factor-like weak inducer of apoptosis</author-keyword>
      <author-keyword>Tumor necrosis factor-α</author-keyword>
    </author-keywords>
  </citation-info>
  <citation-title>
    <titletext xml:lang="en" original="y">Dynamics of cytokines predicts risk of hepatocellular carcinoma among chronic hepatitis C patients after viral eradication</titletext>
  </citation-title>
```

Example of data included in author-group and correspondence

```
<author-group>
  <author seq="22">
    <ce:initials>M.-L.</ce:initials>
    <ce:indexed-name>Yu M.-L.</ce:indexed-name>
    <ce:surname>Yu</ce:surname>
    <ce:given-name>Ming-Lung</ce:given-name>
    <ce:e-address>fish6069@gmail.com</ce:e-address>
  </author>
  <affiliation country="twn">
    <organization>Institute of Biomedical Sciences, National Sun Yat-Sen University</organization>
    <city>Kaohsiung</city>
  </affiliation>
</author-group>
<correspondence>
  <person>
    <ce:initials>M.-L.</ce:initials>
    <ce:indexed-name>Yu M.-L.</ce:indexed-name>
    <ce:surname>Yu</ce:surname>
    <ce:given-name>Ming-Lung</ce:given-name>
  </person>
  <affiliation country="twn">
    <organization>Hepatitis Center and Hepatobiliary Division, Department of Internal Medicine, Kaohsiung Med
ical University Hospital</organization>
    <address-part>No. 100 Shih-Chuan 1st Road, Sanmin District</address-part>
    <city>Kaohsiung</city>
  </affiliation>
  <ce:e-address>fish6069@gmail.com</ce:e-address>
</correspondence>
```

Example of data included in abstracts

<abstracts>
<abstract xml:lang="en" original="y">
<publishercopyright>© 2022 Baishideng Publishing Group Co., Limited. All rights reserved.</publishercopyright>
<ce:para>BACKGROUND Chronic hepatitis C virus (HCV) infection induces profound alterations in the cytokine and chemokine signatures in peripheral blood. Clearance of HCV by antivirals results in host immune modification, which may interfere with immune-mediated cancer surveillance. Identifying HCV patients who remain at risk of hepatocellular carcinoma (HCC) following HCV eradication remains an unmet need. We hypothesized that antiviral therapy-induced immune reconstruction may be relevant to HCC development. AIM To investigate the impact of differential dynamics of cytokine expression on the development of HCC following successful antiviral therapy. METHODS One hundred treatment-naïve HCV patients with advanced fibrosis (F3/4) treated with direct-acting antivirals (DAAs) or peginterferon/ribavirin who achieved sustained virologic response [SVR, defined as undetectable HCV RNA throughout 12 wk (SVR12) for the DAA group or 24 wk (SVR24) for the interferon group after completion of antiviral therapy] were enrolled since 2003. The primary endpoint was the development of new-onset HCC. Standard HCC surveillance (abdominal ultrasound and α-fetoprotein) was performed every six months during the followup. Overall, 64 serum cytokines were detected by the multiplex immunoassay at baseline and 24 wk after end-of-treatment. RESULTS HCC developed in 12 of the 97 patients over 459 person-years after HCV eradication. In univariate analysis, the Fibrosis-4 index (FIB-4), hemoglobin A1c (HbA1c), the dynamics of tumor necrosis factor-α (TNF-α), and TNF-like weak inducer of apoptosis (TWEAK) after antiviral therapy were significant HCC predictors. The multivariate Cox regression model showed that ΔTNF-α (≤ -5.7 pg/mL) was the most important risk factor for HCC (HR = 11.54, 95%CI: 2.27- 58.72, P = 0.003 in overall cases; HR = 9.98, 95%CI: 1.88-52.87, P = 0.007 in the interferon group). An HCC predictive model comprising FIB-4, HbA1c, ΔTNF-α, and ΔTWEAK had excellent performance, with 3-, 5-, 10-, and 13-year areas under the curve of 0.882, 0.864, 0.903, and 1.000, respectively. The 5-year accumulative risks of HCC were 0%, 16.9%, and 40.0% in the low-, intermediate-, and high-risk groups, respectively. CONCLUSION Downregulation of serum TNF-α significantly increases the risk of HCC after HCV eradication. A predictive model consisting of cytokine kinetics could ameliorate personalized HCC surveillance strategies for post-SVR HCV patients.</ce:para>
</abstract>
</abstracts>

Example of data included in source

```
<source type="j" country="chn">
  <sourcetitle>World Journal of Gastroenterology</sourcetitle>
  <sourcetitle-abbrev>World J. Gastroenterol.</sourcetitle-abbrev>
  <issn type="electronic">22192840</issn>
  <issn type="print">10079327</issn>
  <codencode>WJGAF</codencode>
  <volisspag>
    <voliss volume="28" issue="1"/>
    <pagerange first="140" last="153"/>
  </volisspag>
  <publicationyear first="2022"/>
  <publicationdate>
    <year>2022</year>
    <month>01</month>
    <day>07</day>
  </publicationdate>
  <website>
    <ce:e-address>https://www.wjgnet.com/1007-9327/full/v28/i1/140.htm</ce:e-address>
  </website>
  <publisher>
    <publishername>Baishideng Publishing Group Inc</publishername>
  </publisher>
</source>
```

Example of data included in enhancement

```
<enhancement status="002" type="8">
  <descriptorgroup>
    <descriptors type="MDV" controlled="y">
      <descriptor>
        <devterm>ELISA kit</devterm>
        <mainterm>ELISA kit</mainterm>
        <weight>b</weight>
      </descriptor>
    <descriptors type="DRG" controlled="y">
      <descriptor>
        <mainterm>peginterferon alpha2a plus ribavirin</mainterm>
        <drgterm>peginterferon alpha2a plus ribavirin</drgterm>
        <link>
          <linkterm>drug therapy</linkterm>
          <sublink>
              <sublinkterm>chronic hepatitis C</sublinkterm>
          </sublink>
        </link>
        <weight>b</weight>
      </descriptor>
    <descriptors ="MED" controlled="y">
      <descriptor>
        <mainterm>adult</mainterm>
        <weight>b</weight>
      </descriptor>
      <descriptor>
        <mainterm>alpha fetoprotein blood level</mainterm>
        <weight>b</weight>
      </descriptor>
      <descriptor>
        <mainterm>antiviral therapy</mainterm>
        <weight>b</weight>
      </descriptor>
      <descriptor>
        <mainterm>article</mainterm>
        <weight>b</weight>
      </descriptor>
      <descriptor>
        <mainterm>cancer risk</mainterm>
        <weight>a</weight>
      </descriptor>
      <descriptor>
        <mainterm>chronic hepatitis C</mainterm>
        <disterm>chronic hepatitis C</disterm>
        <link>
          <linkterm>diagnosis</linkterm>
        </link>
        <link>
          <linkterm>drug therapy</linkterm>
          <sublink>
            <sublinkterm>peginterferon alpha2a plus ribavirin</sublinkterm>
```

```
          </sublink>
        </link>
        <weight>a</weight>
      </descriptor>
      <descriptor>
        <mainterm>confidence interval</mainterm>
        <weight>b</weight>
      </descriptor>
    </descriptors>
  </descriptorgroup>
  <classificationgroup>
    <classifications type="EMCLASS">
      <classification>
        <classification-code>16</classification-code>
      </classification>
      <classification>
        <classification-code>37</classification-code>
      </classification>
      <classification>
        <classification-code>4</classification-code>
      </classification>
      <classification>
        <classification-code>48</classification-code>
      </classification>
      <classification>
        <classification-code>5</classification-code>
      </classification>
      <classification>
        <classification-code>6</classification-code>
      </classification>
    </classifications>
  </classificationgroup>
  <manufacturergroup>
    <manufacturers type="MNV">
      <manufacturer country="usa">Abbott</manufacturer>
      <manufacturer country="usa">Abbott Molecular</manufacturer>
    </manufacturers>
  </manufacturergroup>
  <tradenamegroup>
    <tradenames type="TNV">
      <trademanuitem>
        <tradename>RealTime HCV</tradename>
        <manufacturer country="usa">Abbott Molecular</manufacturer>
      </trademanuitem>
      <trademanuitem>
        <tradename>RealTime HCV Genotype II</tradename>
        <manufacturer country="usa">Abbott Molecular</manufacturer>
      </trademanuitem>
    </tradenames>
  </tradenamegroup>
  <chemicalgroup>
      <chemicals source="esbd">
      <chemical>
```

```
        <chemical-name>hemoglobin A1c</chemical-name>
        <cas-registry-number>62572-11-6</cas-registry-number>
      </chemical>
    </chemicals>
  </chemicalgroup>
</enhancement>
```

2.3. EMTREE-enriched records

EMTREE-enriched records include exploded index terms with a list of all controlled parent subject terms for each subject index term assigned to a record. The full list is wrapped within the element explosions. The subject index term is included in the child element mainterm, and grouped with its parent terms, each contained in a child element ancestor.

Example of data included in explosions

```
<explosions>
  <explosion>
    <ancestors>
      <ancestor>chronic hepatitis</ancestor>
      <ancestor>chronic hepatitis c</ancestor>
      <ancestor>chronic viral hepatitis</ancestor>
      <ancestor>digestive system disease</ancestor>
      <ancestor>digestive system infection</ancestor>
      <ancestor>digestive system inflammation</ancestor>
      <ancestor>diseases</ancestor>
      <ancestor>flaviviridae infection</ancestor>
      <ancestor>hepatitis</ancestor>
      <ancestor>hepatitis c</ancestor>
      <ancestor>hepatobiliary disease</ancestor>
      <ancestor>hepatobiliary system infection</ancestor>
      <ancestor>hepatobiliary system inflammation</ancestor>
      <ancestor>infection</ancestor>
      <ancestor>inflammation</ancestor>
      <ancestor>liver disease</ancestor>
      <ancestor>physical disease</ancestor>
      <ancestor>physical disease by anatomical structure</ancestor>
      <ancestor>physical disease by etiology and pathogenesis</ancestor>
      <ancestor>rna virus infection</ancestor>
      <ancestor>virus hepatitis</ancestor>
      <ancestor>virus infection</ancestor>
    </ancestors>
    <disterm>chronic hepatitis C</disterm>
    <link>
      <linkterm>diagnosis</linkterm>
    </link>
    <link>
      <linkterm>drug therapy</linkterm>
```

```
      <sublink>
        <sublinkterm>peginterferon alpha2a plus ribavirin</sublinkterm>
      </sublink>
    </link>
    <mainterm>chronic hepatitis C</mainterm>
    <weight>a</weight>
  </explosion>
  <explosion>
    <ancestors>
      <ancestor>chemical, physical and mathematical phenomena</ancestor>
      <ancestor>confidence interval</ancestor>
      <ancestor>mathematical parameters</ancestor>
      <ancestor>mathematical phenomena</ancestor>
      <ancestor>parameters</ancestor>
      <ancestor>procedures, parameters and devices</ancestor>
      <ancestor>statistical concepts</ancestor>
      <ancestor>statistical parameters</ancestor>
    </ancestors>
    <mainterm>confidence interval</mainterm>
    <weight>b</weight>
  </explosion>
  <explosion>
    <ancestors>
      <ancestor>controlled study</ancestor>
      <ancestor>types of article or study</ancestor>
      <ancestor>types of study</ancestor>
    </ancestors>
    <mainterm>controlled study</mainterm>
    <weight>b</weight>
  </explosion>
  <explosion>
    <ancestors>
      <ancestor>assessment of humans</ancestor>
      <ancestor>diagnostic procedure</ancestor>
      <ancestor>disease assessment</ancestor>
      <ancestor>disease risk assessment</ancestor>
      <ancestor>medical procedures</ancestor>
      <ancestor>procedures</ancestor>
      <ancestor>procedures, parameters and devices</ancestor>
    </ancestors>
    <mainterm>disease risk assessment</mainterm>
    <weight>a</weight>
  </explosion>
  <explosion>
    <ancestors>
      <ancestor>female</ancestor>
      <ancestor>groups by age and sex</ancestor>
      <ancestor>groups by sex</ancestor>
    </ancestors>
    <mainterm>female</mainterm>
    <weight>b</weight>
  </explosion>
  <explosion>
```

```
    <ancestors>
      <ancestor>assessment of humans</ancestor>
      <ancestor>diagnostic procedure</ancestor>
      <ancestor>digestive system disease assessment</ancestor>
      <ancestor>disease assessment</ancestor>
      <ancestor>fibrosis-4 index</ancestor>
      <ancestor>medical procedures</ancestor>
      <ancestor>procedures</ancestor>
      <ancestor>procedures, parameters and devices</ancestor>
    </ancestors>
    <mainterm>Fibrosis-4 Index</mainterm>
    <weight>b</weight>
  </explosion>
</explosions>
```

3. Description of elements in an EMBASE XML data file

		Parent	Contains code/tag	Contains data	Brief description
1	**bibdataset**				Top-level wrapper element of Elsevier Abstract and Indexing DTD.
2	**item**	1			Top-level wrapper element.
3	**ait:process-info**	2			Contains information relevant to the processing of the record, organized into child elements ait:date-delivered, ait:date-sort, and ait:status.
	ait:date-delivered	3		•	Date given in the attributes *year*, *month*, and *day*, on which a record was processed for delivery. For internal purposes only.
	ait:date-sort	3		•	The date (given in the attributes year, month, and *day*) of a record used for sorting. Data in publicationdate are standardized in format by day, month and year using predefined input rules so that documents with unstructured dates of publication can also be sorted.
	ait:status	3	•		A multi-attribute element that indicates the status of a record as an item in EMBASE. The attribute *type* is set to "core" to indicate that the item is a full bibliographic record. The attribute *state* may be "new" to indicate that the item has been delivered to EMBASE for the first time, "update" to indicate that it replaces a previously delivered item, or "delete" to indicate that a previously delivered item should be deleted. The attribute *stage* indicates an uncorrected document in the record ("S100"), a corrected document ("S200") or a published document ("S300"). The default value is "S300".
4	**bibrecord**	2			Top-level element containing all information of a record, organized into child elements head, item-info, and tail.
5	**item-info**	4			Includes the non-bibliographic information of a record, organized into child elements itemidlist for identifiers, history for record history, copyright for copyright information, and dbcollection for codes of database collections.
	copyright	5		•	Contains the copyright notice from Elsevier B.V. (with year of most recent update to include the record) and/or a third party (e.g., Medline). Origin of the copyright notices is given in the attribute *type*.
6	**itemidlist**	5			Includes known identifiers for the document of a record, such as Publisher Item Identifier (ce:pii) and Digital Object Identifier (ce:doi). Other identifiers are contained in child element itemid. Each identifier is contained it a dedicated element.
	ce:pii	6		•	Contains the publisher item identifier (PII) of the document in a record.
	ce:doi	6		•	Contains the digital object identifier (DOI) of the document in the record. Empty for documents of a print-only journal.
	itemid	6	•	•	Contains other identifiers. The nature of the identifier is given in the attribute idtype, and can be: idtype=PUI: a unique EMBASE identification number assigned to a record idtype=ABSN: the abstract number if the document in a record is a conference abstract idtype=CAR-ID: an internal Elsevier CAR ID number that uniquely identifies each record created in Elsevier systems idtype=EMBASE: a record accession number that is specific to EMBASE. This identifier is no longer in use

		Parent	Contains code/tag	Contains data	Brief description
					idtype=EMBACK: a record accession number that is specific to EMBASE Classic. This identifier is no longer in use idtype=MEDL: the Medline PubMed Unique Identifier (PMID) of the document in a record idtype=COLLECTIONS: a code related to the Elsevier collections infrastructure used for internal processing purposes idtype=ASSOCIATEDPUI: If a MEDLINE record is accessioned before an EMBASE record is created, the MEDLINE record is assigned an associated EMBASE identification number contained in this element. Once the EMBASE record is created, it is assigned a new EMBASE identification number (PUI) that is found in the element ce:pui but the associated PUI is retained.
7	**history**	5			Provides the date on which an automatically and fully indexed record were delivered (date-created); (date-aip-loaded; date-loaded); (record-updated-date); into EMBASE. Each child element structures the date into attributes *year*, *month*, and *day*.
	date-aip-loaded	7		•	Provides the date on which an automatically indexed record (e.g., article in press) is loaded into EMBASE. Besides conference abstracts and preprints, all automatically indexed records are subsequently indexed manually and loaded again into EMBASE.
	date-created	7		•	Provides the date on which a fully indexed (automatically and manually) record was delivered to EMBASE.
	date-loaded	7		•	Provides the date on which a fully indexed (automatically and manually) record is loaded into EMBASE.
	record-updated-date	7		•	A multi-value date field that captures the history of the record, listing each date when an update is made to the record.
	dbcollection	5		•	Contains the Elsevier database collection(s) to which a record belongs. such as EMBASE, EMBASE Classic, EMBFRA, MEDLINE, NURSING, PREPRINT and CLINICAL TRIAL. A record can belong to 1 or more collections.
	external-source	5		•	Includes additional information about the external (non-Elsevier) origin of a record.
8	**head**	4			Contains the abstract and indexing information of a record, organized into the child elements abstracts, author-group, citation-info, citation-title, correspondence, enhancement, and source.
9	**citation-info**	8			Provides the descriptive information of the document in a record, organized into the child elements abstract-language, author-keywords, citation-language, citation-type.
	citation-type	9		•	Identifies the type of document in a record: abstract report (ab),article (ar), book (bk), book review (br), business article (bz), conference abstract (cb), chapter (ch), conference paper (cp), conference review (cr), clinical trial (ct), dissertation (di), data paper (dp), editorial (ed), erratum (er), article in press (ip), letter (le), multimedia (mm), note (no), patent (pa), preprint (pp), press release (pr), review (re), chapter with references only (rf), report (rp), short survey (sh), standard (st), tombstone (tb), working paper (wp).
	citation-language	9	•		Provides the language of the original document in a record. In cases where a document is published in parallel in other languages, up to three languages may be provided.
	abstract-language	9, 22	•		Provides the language of the abstracts or summaries published in the original document. Up to three languages are captured.

		Parent	Contains code/tag	Contains data	Brief description
10	**author-keywords**	9			Contains the uncontrolled keywords assigned by the author(s) to the document in a record. These may be in the original language or translated.
	author-keyword	10		•	Contains a single uncontrolled keyword assigned by the author(s) to the document in a record.
11	**citation-title**	8			Contains the original and every translation of the title of a document in a record, organized into child element titletext.
	titletext	11	•	•	Contains the title of the document in a record, original or translated. The attribute *lang* provides a 2-letter or 3-letter code for the language of the recorded document title. The attribute *original* is set to "y" if that language is the original language used in the document of a record.
12	**author-group**	8			Contains name and affiliation of the author(s) of the document in a record, organized by affiliation and into the child elements author, collaboration, affiliation.
13	**author**	12			Contains personal information about an author of the document in a record, organized into the child elements ce:initials, ce:indexed-name, ce:degrees, ce:surname, ce:given-name, ce:suffix, ce:e-address. The attribute *seq* contains a sequential numbering indicating the order of authors in the document of a record.
	ce:initials	13, 17, 25		•	Provides the initials of an author or contributor, as part of their name.
	ce:indexed-name	13, 14, 17		•	Contains a sortable variant of an author's name or a collaboration name. In the case of an author, the entry consists of the concatenated values of the elements surname and initials, with all special characters removed.
	ce:degrees	13, 17, 25		•	An optional element that provides any degrees listed for an author or contributor.
	ce:surname	13, 17, 25		•	Provides the family name of an author or contributor. Names that cannot be easily parsed into given name, initials and surname (e.g., Chinese names) are contained completely in this element.
	ce:given-name	13, 17, 25		•	Provides the forename(s) or first name(s) of an author or contributor.
	ce:suffix	13,		•	An optional element containing any suffix to the name of an author, such as junior, senior, III, etc.

		Parent	Contains code/tag	Contains data	Brief description
		17			
	ce:e-address	13, 16, 26		•	Provides one or more electronic mail addresses of an author, a corresponding person, or a publisher.
14	**collaboration**	12			Contains within the child element ce:indexed-name the designation of a group of authors that present themselves under a common name.
15	**affilitation**	12, 16, 24, 26			Contains details about the affiliation of an author group, corresponding person, contributor, or publisher, like name and address. Includes the child elements ce:text, organization, address-part, city-group, city, state. A 3-letter code in the attribute *country* indicates the country of the author's affiliation.
	ce:text	15		•	An optional container element for text used for an *unstructured* address of an author group's affiliation. Empty if the affiliation has a structured address.
	organization	15		•	Provides the name of the organization for an author group's affiliation, which can include institution, department, or other.
	address-part	15		•	For the structured address of an author group's affiliation, contains a street or P.O. box.
	city-group	15, 30		•	An optional container element for city information of an author group's affiliation or of the venue for a conference.
	city	15, 30		•	Contains a city name for the structured address of an author group's affiliation or for the venue of a conference.
	state	15		•	Provides the name of the state in the structured address of an author group's affiliation.
16	**correspondence**	8			Contains details about the corresponding author of the document in a record, organized into child elements person, affiliation, ce:e-address.
17	**person**	16			Provides the name and personal details of the corresponding author, organized into child elements initials, indexed-name, degrees, surname, given-name, and suffix.
18	**abstracts**	8			Contains one or more abstracts of the document, the original author summary and any possible translations or other alternatives.
19	**abstract**	18			Each abstract of the document in a record is listed in abstract, with information in child elements publishercopyright and ce:para. The attribute *lang* provides a 2-letter or 3-letter code for the language of the abstract and *original* is set to "y" if that language is the original used in the document. The attribute *perspective* indicates if the abstract was auto-translated.
	publishercopyright	19		•	Provides the copyright statement of the publisher for the document in a record.

		Parent	Contains code/tag	Contains data	Brief description
	ce:para	19		•	The abstract of the document in a record may consist of one or more paragraphs. Each is contained in this child element ce:para.
20	**source**	8			Contains information about the source of the document in a record, organized into child elements sourcetitle, sourcetitle-abbrev, issuetitle, issn, isbn, codencode, volisspag, article-number, publicationyear, publicationdate, website, contributor-group, publisher, additional-srcinfo, and bib-text. The element has two attributes: *type* assigns a code to indicate the type of source as journal (j), book series (k) or trade journal (d). *country* provides the 3-letter code of the country where the source is published.
	sourcetitle	20		•	Contains the full title of the source (e.g., journal, book, conference proceeding, report) for the document in a record.
	sourcetitle-abbrev	20		•	Contains the abbreviated title of the source for the document in a record.
	issuetitle	20		•	Provides the name of the issue in which the document of a record appears. Issues of journals, books, conference proceedings or reports occasionally have their own title.
	issn	20			Provides one or more international standard serial numbers (ISSN) of the source for the document in a record. The attribute *type* indicates the type of ISSN (e.g., electronic, print).
	isbn	20		•	Provides one or more 10-digit or 13-digit International Standard Book Number (ISBN) assigned to the source of the document in a record. Note: 10-digit ISBN were discontinued at the end of 2006. The attribute *length* indicates if the ISBN is 10 or 13 digits. The attribute *level* indicates the level of the ISBN assigned to the source of the document in a record, such as a book set or an individual book volume. Unique 13-digit ISBN are assigned to each format of a source. Thus, 13-digit ISBN include the attribute *type* to indicate format, such as hardback or paperback.
	codencode	20		•	Provides the CODEN code that Chemical Abstracts Service (CAS) uses to uniquely identify the source of the document in a record.
21	**volisspag**	20			Contains details about the volume, issue and pages of a source where the document of a record appears. Includes child elements voliss, supplement, pagerange, pages, pagecount.
	voliss	21		•	Provides the volume and issue of the source in which the document of a record was published in the respective attributes *volume* and *issue*.
	supplement	21		•	Indicates that the document of a record is located in a supplement issue. May include supplement number.
	pagerange	21		•	If pages of a source are numeric, provides the page on which the document of a record begins in attribute *first* and where it ends in attribute *last*.
	pages	21		•	An optional element used to provide page information when the first and last pages are not completely numeric.
	pagecount	21		•	Provides the number of pages in the document of a record, when relevant for the document type, such as a book.
	article-number	20		•	Provides the item number assigned to the document by the publisher.
	publicationyear	20		•	Provides the year of publication as reported in the document of a record. The attribute *first* provides the year of first appearance (e.g., e-publication). The attribute *last* is used when, for example, final distribution occurs in another year.

		Parent	Contains code/tag	Contains data	Brief description
22	**publicationdate**	20			Specifies a more complete date of publication as reported in the document of a record, either in child elements year, month, and day, or in child element date-text.
	year	22		•	Indicates the year of publication.
	month	22		•	Indicates the month of publication.
	day	22		•	Indicates the day of publication.
	pubdatetext	22		•	Provides the unstructured publication date (e.g., Spring 2002) of the document in a record.
23	**website**	20			Contains information about the website where the document in a record is published. May contain a URL or an electronic mail address
	ce:e-address	23		•	Provides the URL of the website or a general email for the source website where the document in a record is published. If a URL, the attribute *type* is set to "url".
24	**contributor-group**	20			Contains information on one or more contributors to the source of a document in a record, organized into child elements contributor and affiliation. Contributors are mainly editors.
25	**contributor**	24			Contains details about a contributor to the source of a document in a record, organized into child elements ce:initials, ce:degrees, ce:surname, and ce:given-name. The attribute *role* indicates the function of a contributor (mainly editors for book series) and the attribute *seq* contains a sequential numbering indicating the order of contributors in the source.
26	**publisher**	20			Contains information about the publisher of the document in a record; at least the publishername and optionally the address either in the child element publisheraddress or affiliation. If available, an email address is provided in the child element ce:e-address.
	publishername	26		•	Lists the name of the publisher of the document in a record.
	publisheraddress	26		•	Optionally provides the unstructured address of the publisher of the document in a record. A structured address is provided in the element affiliation.
27	**additional-srcinfo**	20			Provides details on any additional source information, such as the name and date of a conference that generated the document in a record.
28	**conferenceinfo**	27			If the source of a document in a record is associated with a conference, this element provides details in child element confevent.
29	**confevent**	28			Contains information about a conference event, organized into child elements confname, conflocation, confdate, and confcode,
	confname	29		•	Provides the name of a conference associated with the source of the document in a record.
30	**conflocation**	29			Provides the venue and address of a conference associated with the source of the document in a record, structured into child elements venue, city-group, city. The attribute *country* provides the 3-letter code for the country of a conference event.
	venue	30		•	Lists the name of the hotel, center, or other place where a conference event is held.
31	**confdate**	29			Provides the begin and end dates of a conference associated with the source of the document in a record.

		Parent	Contains code/tag	Contains data	Brief description
	startdate	31		•	Lists the year, month, and day of the start of a conference event as corresponding attributes.
	enddate	31		•	Lists the year, month, and day of the end of a conference event as corresponding attributes.
	confcode	29		•	Contains a unique code assigned by Elsevier to a conference associated with the source of the document in a record.
	bib-text	20		•	Optional element used when bibliographic information about a source is unstructured.
32	**enhancement**	8			Contains additional indexing or classification of a record. Includes child elements descriptorgroup, classificationgroup, manufacturergroup, tradenamegroup, sequencebanks, chemicalgroup. The attribute *status* provides a 3-digit code that indicates the indexing status of a record as either intermediate (001) or final (002). The attribute *type* indicates the type of indexing performed; 17 for automatic indexing and 8 for manual indexing. In EMBASE, after automatic indexing generates a record of intermediate status, it is taken to final status by manual indexing.
33	**descriptorgroup**	32			Contains a list of subject index terms describing the contents of the document in the record. The subject index terms may be controlled by a thesaurus or uncontrolled and are grouped into the child element descriptors by type (e.g., controlled drug terms, controlled medical terms).
34	**descriptors**	33			Contains a set of subject index terms of a specific type, each in the child element descriptor, that describe the contents of the document in a record. The attribute *controlled* is set to "y" if the subject index terms included in a descriptors element are controlled. The attribute *type* contains the type of subject index terms and can be either MDV (controlled medical device terms), DRG (controlled drug terms) or MED (controlled medical terms).
35	**descriptor**	34			Contains one subject index term of a specific type describing the contents of the document in a record. Expanded information about the subject index term is provided in child elements devterm, disterm, drgterm, mainterm, weight, link.
	devterm	35, 56		•	Provides a controlled medical device term as defined by EMTREE.
	disterm	35, 56		•	Provides a controlled disease term as defined by EMTREE.
	drgterm	35, 56		•	Provides a controlled drug term as defined by EMTREE.
	mainterm	35, 56		•	Contains a principle subject index term assigned to the document in a record. Context about how the "main term" is used in the document is provided in child element link. The attribute *candidate* is set to "true" when the subject index term in mainterm is controlled but still a candidate term.
	weight	35, 56	•		Designates the subject index term in mainterm as a major term (weight = a) or a minor term (weight = b).

		Parent	Contains code/tag	Contains data	Brief description
36	**link**	35, 56			Contains the child elements linkterm and sublink with EMTREE terms used to describe the context in which the subject index term in mainterm is used. They are concept modifiers.
	linkterm	36		•	Provides an EMTREE term used to describe the context in which the subject index term in mainterm is used when assigned to the document in a record.
37	**sublink**	36			Contains the child element sublinkterm with an EMTREE term that further qualifies the context in which the subject index term in mainterm is used. The term is subordinate to the EMTREE term in linkterm.
	sublinkterm	37		•	Provides an EMTREE term subordinate to the term in linkterm.
38	**classificationgroup**	32			Provides information on content categories to which a document in a record is assigned. These are assigned according to the EMBASE section headings (see Appendix).
39	**classifications**	38			Contains one or more child elements classification with content categories of a specific type to which a document in a record is assigned. The *type* is indicated in the corresponding attribute.
40	**classification**	39			Contains the child element classification-code which provides the numerical code of a content category assigned to the document in a record based on the EMBASE section headings.
	classification-code	40	•		Provides the numerical code of a category in the EMBASE section headings (see Appendix) to which a document in a record is assigned.
41	**manufacturergroup**	32			Contains information about one or more manufacturers mentioned in the document of a record, listed in child element manufacturers.
42	**manufacturers**	41			Lists details about one or more manufacturers of a specific type mentioned in the document of a record, each in child element manufacturer. The type is provided in the attribute *type* and can be either MNF for drug manufacturers or MNV for device manufacturers.
	manufacturer	42, 45		•	Provides the name of a manufacturer mentioned in the document of a record according to a controlled list. The country of the manufacturer is indicated in the attribute *country*.
43	**tradenamegroup**	32			Contains information on tradenames mentioned in the document of a record, organized in the child element tradenames.
44	**tradenames**	43			Lists one or more tradenames of a specific type mentioned in the document of a record, each in child element trademanuitem. The attribute *type* provides a 3-letter code that indicates the type of tradenames listed. These can be TNM for a drug tradename or TNV for a device tradename.
45	**trademanuitem**	44			Provides details about a tradename, and optionally its manufacturer, organized into child elements tradename, and manufacturer.
	tradename	45		•	Provides the tradename for a drug, device that is mentioned in the document of a record. The trademark symbol is not used.
46	**sequencebanks**	32			Contains information about one or more nucleotide or amino acid sequences defined or mentioned in the document of a record, organized into child element sequencebank.

		Parent	Contains code/tag	Contains data	Brief description
47	**sequencebank**	46			Contains the accession numbers of one or more nucleotide or amino acid sequences defined or mentioned in the document of a record. All accession numbers listed in one sequencebank element come from one sequence bank. The attribute *name* provides the name of the sequence bank.
	sequence-number	47		•	Provides the accession number of individual nucleotide or amino acid sequences from a specific sequence bank that are mentioned in the document of a record.
48	**chemicalgroup**	32			Contains information on chemicals mentioned in the document of a record. Sets of chemicals are organized into child element chemicals according to the source of the information.
49	**chemicals**	48			Contains a list of child element chemical, each containing the association of a chemical name with one or more CAS registry numbers. The source of that information is common to the list and provided in the attribute *source* as either "nlm" for the National Library of Medicine or "esbd" for the Elsevier Bibliographic Databases Division.
50	**chemical**	49			Contains the association of a chemical name with one or more CAS Registry Numbers and/or Enzyme Commission Number, organized into child elements chemical-name, cas-registry-number, and enzyme-commission-number.
	chemical-name	50		•	Provides the name of a chemical substance that is mentioned in the document of a record.
	cas-registry-number	50		•	Provides one CAS Registry Number for the chemical substance named in element chemical. A chemical substance may have more than one CAS Registry Number.
	enzyme-commission-number	50		•	Provides one Enzyme Commission Number (EC number) for the chemical substance named in element chemical, provided it is an enzyme. A chemical substance may have more than one Enzyme Commission Number.
51	**tail**	4			Contains in the child element bibliography the number of references in the bibliography of the document in a record.
	bibliography	51		•	Contains in the attribute *refcount* the number of references in the bibliography of the document in a record.
52	**explosions**	4			Contains information on all EMTREE parent terms for each of the index terms (mainterm, devterm, drgterm, disterm) assigned to the document of a record, grouped by index term in the child element explosion.
53	**explosion**	52			Contains information on all EMTREE parent terms for one index term assigned to the document of a record, organized into child elements ancestors, devterm, disterm, drgterm, link, mainterm, and weight.
54	**ancestors**	53			Lists each EMTREE parent term of one index term in a child element ancestor.
	ancestor	54		•	Provides one EMTREE parent term of an index term assigned to the document of a record.

4. Appendix: Classifications (EMBASE section headings)

Assigned number	Classification (EMBASE section heading)
1	Anatomy, anthropology, embryology and histology
2	Physiology
3	Endocrinology
4	Microbiology: bacteriology, mycology, parasitology and virology (includes section 47 as of 1992)
5	General pathology and pathological anatomy
6	Internal medicine
7	Pediatrics and pediatric surgery
8	Neurology and neurosurgery
9	Surgery (includes section 34 as of 1992)
10	Obstetrics and gynecology
11	Otorhinolaryngology
12	Ophthalmology
13	Dermatology and venereology
14	Radiology
15	Chest diseases, thoracic surgery and tuberculosis
16	Cancer
17	Public health, social medicine and epidemiology
18	Cardiovascular diseases and cardiovascular surgery
19	Rehabilitation and physical medicine
20	Gerontology and geriatrics
21	Developmental biology and teratology
22	Human genetics
23	Nuclear medicine
24	Anesthesiology
25	Hematology
26	Immunology, serology and transplantation
27	Biophysics, bioengineering and medical instrumentation
28	Urology and nephrology

Assigned number	Classification (EMBASE section heading)
29	Clinical and experimental biochemistry
30	Clinical and experimental pharmacology
31	Arthritis and rheumatism
32	Psychiatry
33	Orthopedic surgery
34	Plastic surgery (1974–1991. Incorporated into section 9 in 1992)
35	Occupational health and industrial medicine
36	Health policy, economics and management
37	Drug literature
38	Adverse reaction titles
39	Pharmacy (introduced in 1997)
40	Drug dependence, alcohol abuse and alcoholism
46	Environmental health and pollution control
47	Virology (1974–1991. Incorporated into section 4 in 1992)
48	Gastroenterology
49	Forensic science abstracts
50	Epilepsy abstracts
51	Leprosy and other mycobacterial diseases (1979–1988. Incorporated into section 4 in 1989)
52	Toxicology (introduced in 1983)

5. Appendix: list of letter codes for language

Following two languages are kept as 3-letters:

Language codes that stay with 3 letters
"scr", "srp"

The ones that now have a 2-letter code are:

2-letter correspondance for previously 3-letter codes
AFRK("afr", "af", "Afrikaans") // Afrikaans
ALBA("alb", "sq", "Albanian") // Albanian
ARAB("ara", "ar", "Arabic") // Arabic
ARME("arm", "hy", "Armenian") // Armenian
AZRB("aze", "az", "Azerbaijani") // Azerbaijani
BASQ("baq", "eu", "Basque") // Basque
BYEL("bel", "be", "Belarusian") // Belarusian
BENG("ben", "bn", "Bengali") // Bengali
BOSN("bos", "bs", "Bosnian") // Bosnian
BULG("bul", "bg", "Bulgarian") // Bulgarian
BURM("bur", "my", "Burmese") // Burmese
CATA("cat", "ca", "Catalan") // Catalan
CHIN("chi", "zh", "Chinese") // Chinese
CROA("hrv", "hr", "Croatian") // Croatian
CZCH("cze", "cs", "Czech") // Czech
DANS("dan", "da", "Danish") // Danish
DUTC("dut", "nl", "Dutch") // Dutch
ENGL("eng", "en", "English") // English
ESPE("epo", "eo", "Esperanto") // Esperanto
ESTO("est", "et", "Estonian") // Estonian
FINN("fin", "fi", "Finnish") // Finnish
FREN("fre", "fr", "French") // French
GALL("glg", "gl", "Gallegan") // Gallegan
GEOR("geo", "ka", "Georgian") // Georgian
GERM("ger", "de", "German") // German
GREK("gre", "el", "Greek") // Greek
HEBR("heb", "he", "Hebrew") // Hebrew
INDI("hin", "hi", "Hindi") // Hindi
HUNG("hun", "hu", "Hungarian") // Hungarian
ICEL("ice", "is", "Icelandic") // Icelandic
BAHA("ind", "id", "Indonesian") // Indonesian
GAEL("gle", "ga", "Irish Gaelic") // Irish Gaelic
ITAL("ita", "it", "Italian") // Italian
JAPA("jpn", "ja", "Japanese") // Japanese
KORA("kor", "ko", "Korean") // Korean
LATIN("lat", "la", "Latin") // Latin
LATV("lav", "lv", "Latvian") // Latvian
LITH("lit", "lt", "Lithuanian") // Lithuanian
MACE("mac", "mk", "Macedonian") // Macedonian
MALA("may", "ms", "Malay") // Malay
MAOR("mao", "mi", "Maori") // Maori
MONG("mon", "mn", "Mongolian") // Mongolian
NORW("nor", "no", "Norwegian") // Norwegian
PERS("per", "fa", "Persian") // Persian
POLS("pol", "pl", "Polish") // Polish
POLY("map", "xx", "Polyglot") // Polyglot
PORT("por", "pt", "Portuguese") // Portuguese
PUSH("pus", "ps", "Pushto") // Pushto
RUMA("rum", "ro", "Romanian") // Romanian
RUSS("rus", "ru", "Russian") // Russian
SCOT("gla", "gd", "Scottish Gaelic") // Scottish Gaelic
SERB("scc", "sr", "Serbian") // Serbian
SINH("sin", "si", "Sinhalese") // Sinhalese
SLVK("slo", "sk", "Slovak") // Slovak
SLVN("slv", "sl", "Slovenian") // Slovenian
SPAN("spa", "es", "Spanish") // Spanish
SWED("swe", "sv", "Swedish") // Swedish
TAGA("tgl", "tl", "Tagalog") // Tagalog
THAI("tha", "th", "Thai") // Thai
TURK("tur", "tr", "Turkish") // Turkish
UKRA("ukr", "uk", "Ukrainian") // Ukrainian
URDU("urd", "ur", "Urdu") // Urdu
UZBE("uzb", "uz", "Uzbek") // Uzbek
VIET("vie", "vi", "Vietnamese") // Vietnamese