CONTRAST-in

Elsevier Documentation for Supplier Dataset Orders and Deliveries

Describing PTSIII Order DTD 1.51, and CONTRAST Journal EW Input Version 2019.1, Book EW Input Version 2019.1, Satellite EW Input Version 2019.1, Print Version 1.4, Ready Messages Version 3.0

Content and Data Architecture, Elsevier B.V.

Version 1.32.0

May 28, 2019

Correspondence to:
Jos Migchielsen
Content and Data Architecture, Operations
DTD Development and Maintenance
Elsevier
Radarweg 29
1043 NX Amsterdam
Netherlands
Email: j.migchielsen@elsevier.com

This document was created by Elsevier's DTD Development & Maintenance Team, the team responsible for development, maintenance and support of the Elsevier SGML and XML DTDs and XML content transport schemas. Comments about the schemas and their documentation, as well as change requests, can be sent to the team. Change requests will be considered for implementation in a future version.

The Elsevier CONTRAST standard schemas and a fully clickable PDF file of this documentation are available via http://www.elsevier.com/locate/xml.

This is version 1.32.0 of the documentation of the CONTRAST content transport standard. The document is still being expanded: mistakes are being corrected and extra clarification is given when needed. **It should be noted that the standard itself, including the collection of W3C schemas, is complete and final. This document already authoratively describes the CONTRAST standard.** The authors welcome comments and suggestions for improvement.

This document was typeset using pdfTEX and the MiKTEX 2.9 distribution.

Contents

Chapter 1
Introduction

This is the documentation of Elsevier's CONTRAST standard (*con*tent *tra*nsport *st*andard).

A *dataset* is the name for any delivery of Elsevier content. Traditionally, datasets have been delivered to and from the Electronic Warehouse in a format called EFFECT. This format defined a dataset directory structure, and an accompanying file describing the dataset called the `dataset.toc`.

EFFECT (Exchange Format For Electronic Components And Texts) is a standard to enable large-scale deliveries of electronic files. This standard was initially developed by Elsevier to support comprehensive electronic journal/article distribution from production systems at the publisher to distribution servers either at local libraries or at a remote host organization. The EFFECT standard describes how large amounts of electronic files can be structured and encased in datasets, and how the "packing list" (the file dataset.toc), which comes with the dataset, is structured. The standard was developed in the course of the TULIP project, a five-year research project (1991–1995) on digital libraries by (then) Elsevier Science and nine major universities in the USA. The material provided by Elsevier was used to create local current awareness and article delivery database systems.

The EFFECT standard has served us well while the CAP and PreCAP workflows evolved into a smooth process for journals. It was even (mis)used for certain varieties of books, that were forced into the journal model. EFFECT's limits were reached.

The introduction of new intrastructure, based on XML, was the incentive to develop a new XML-based content transport standard for journal and book content. CONTRAST (*con*tent *tra*nsport *st*andard) is the name of that standard. It consists of agreements about how datasets are organized and of W3C schemas that define a dataset description format.

Dataset creation is largely done by Elsevier's external suppliers. The suppliers validate the dataset with Elsevier's own validation tools before they deliver content to the Electronic Warehouse. For journals, the workflow is very automated and driven by XML orders from Elsevier's workflow system PTS.

This document describes " CONTRAST-in", i.e. the Elsevier–supplier interaction; it specifies the dataset requirements and documents the orders.

The Electronic Warehouse receives deliveries from the supplier and sends them on to online content repositories and other users of Elsevier's electronic content. For this, the EW can use EFFECT as well as CONTRAST. Outbound CONTRAST is equal to inbound CONTRAST in almost all respects. It is described in a separate document, "CONTRAST-out".

Chapter 2

Technical aspects

An important part of the CONTRAST transport standard are the W3C XML schemas that define the transport format.

This chapter contains some technical details about the CONTRAST schemas and the XML files that are structured according to these schemas.

2.1. CONTRAST versions

There are several versions of CONTRAST described in this document. The three major versions are CONTRAST for serial publications, CONTRAST for books and CONTRAST for ready messages. Each has a version number, and the version number is captured in the XML schema.

The correct way to determine the version number is by using the namespace of the `dataset` top element. For instance,

`http://www.elsevier.com/xml/schema/transport/journal-2016.6/s100`

is the namespace of version 2016.6 of the S100 journal schema.

As is usual, this is not an actual file. The location of the XML schema file itself is to be based on the schema location.

```
<dataset
  xmlns="http://www.elsevier.com/xml/schema/transport/journal-2016.6/s100"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation=
   "http://www.elsevier.com/xml/schema/transport/journal-2016.6/s100
    http://www.elsevier.com/xml/schema/transport/journal-2016.6/s100.xsd"
  schema-version="2016.6">
```

is the typical opening of a dataset.xml file that conforms to the above-mentioned schema. The convention is used that the schema location is almost identical to the namespace. The schema location is not an existing file on the Elsevier website either. It is expected that applications deploy XML catalogs in order to retrieve an instance of the schema.

2.2. Version history

2.2.1. Journal 1.0

The Journal 1.0 schemas were released on 31 October 2003.

2.2.2. Journal 1.1

The Journal 1.1 schemas were released on 8 March 2004.

- The S5 schema was added to support deliveries for that stage.
- The schemas are now also used for production of book series. This was reflected in the wording of various annotations and in the fact that optional element `isbn` was added to element `journal-issue-properties`.
- The `ml` version `JI 5.0.0` was replaced by `SI 5.1.0`.
- The PDF property `WRAPPED` was replaced by `WRAPPED OPTIMIZED`.
- Optional element `pdf-pages-web` was added to element `production-details`.
- The list of values for supplier production types was removed.

The schemas were later patched. The first time (17 March 2004) to improve the pattern used in the `suppl` element and the second time (4 June 2004) to make the element `page-fraction-trail` optional and contain only positive values.

2.2.3. Journal 1.2

The Journal 1.2 schemas were released on 22 October 2004.

- Several schemas were adapted to support the delivery of subitems. The S5, S100 and S200 schemas were changed to contain up to 512 journal items and the P100 schema was changed to contain up to 512 web PDFs. Furthermore the element `production-details` was made optional in the S100 and S200 schemas.
- The `jid` element was changed to allow JIDs starting with "`bs:`".
- The `suppl` element was changed to allow supplements numbered 10 or higher.

The schemas were later patched (17 December 2004) to allow for a new DOI prefix.

2.2.4. Journal 1.3

The Journal 1.3 schemas were released on 25 January 2005.

- Element `parent-item` was replaced by `batch` and `journal-item`'s attribute `parent` by `type` with values in the new list journal-item-types-list.
- An S300-H300-Consult schema was added to support deliveries of specific third-party journals for Consult. It is equal to the S300-H300 schema except that element `web-pdf` is optional.
- Value `MAIN-ABRIDGED` was added to the list of PDF purposes. The number of web-PDFs was changed to 5 in the S100 schema and to 4 in the S200 schema. In the P100 schema the number of web-PDFs was changed to 4 and possible PDF purposes `GRAPHICAL-ABSTRACT`, `STEREOCHEMISTRY-ABSTRACT` and `MAIN-ABRIDGED` were added.
- Elements `journal-issue-properties`, `volume-issue-number` and `suppl` were locally defined in the P100 schema, and `suppl` was changed in the P100 schema to make delivery of spin-off's possible.

The schemas were later patched (1 March 2005). The local definitions as described above were removed and instead a simple type "suppl-type" was added to the mother schema. The type was redefined in the other schemas except for the S5, S100 and S200 ones.

A second patch was released on 27 June 2005. In the mother schema the simple types "pii-pattern" and "doi-pattern" were added and used. A new schema was added for S350-H350-Retro deliveries. It is equal to the S350-H350 schema except that the two new types were redefined to allow for S350-H350 deliveries that contain SSDIs and SSDI-based DOIs.

These changes had not the desired result and were corrected with a third patch on 18 August 2005. In these schemas the SSDI patterns were added to "pii-pattern" and "doi-pattern" in the mother schema and these latter two simple types were restricted in the other schemas except in the retro schema.

2.2.5. Journal 1.4

The Journal 1.4 schemas were released on 31 August 2005.

- A new schema, based on the P100 and S300-H300 schemas, was added for Q300 deliveries.
- The new stage Q300 was added to the stages list.
- The pattern for version numbers was changed to include versions up to $6.n$.
- Support for ISBN-13 was implemented.
- A new PDF version was added, `1.4`.
- `vol-first` and `iss-first` were changed to accept non-negative integers.
- `pathname` was changed to exclude the use of "`..`" and accept only pathnames that end in an "8+3" filename.
- In the F300 schema the schema identifier for the print schema was changed to `print 1.0`.
- In the P100 schema, element `journal-issue/files-info/ml` was made optional (for spinoff issues only).

The schemas were later patched (18 October 2005) to improve the PII and DOI patterns used in the Retro schema.

2.2.6. Journal 1.5

The Journal 1.5 schemas were released on 21 December 2005.

- Support for embargoes was implemented.
- Various lists of patterns and lists of values were introduced. Some elements are no longer declared in the mother schema.
- To support deliveries of material for special projects three so-called project schemas were introduced. They replace the Consult and the Retro schemas. Much of the schema validation is removed from these project schemas and is performed elsewhere. Additionally element `project-id` was introduced and production process `CONVERSION` was changed to `PROJECT`.

The schemas were later patched to enable S350 `FULL-TEXT` deliveries and S300 and S350 project deliveries of DTD 4.5 material (23 January 2006) and to enable `HEAD-ONLY` project deliveries (27 January 2006).

2.2.7. Journal 1.6

The Journal 1.6 schemas were released on 12 June 2006.

- Web-PDF version 1.3 was removed.
- In F300 schemas spin-off issues were made possible.

2.2.8. Journal 1.7

The Journal 1.7 schemas were released on 29 September 2006.

- In S300-H300(-project) schemas delivery of main-abridged PDF files is now possible.
- In Q300(-project) schemas delivery without issue PDF files is now possible.
- Element `embargo` was changed slightly and element `embargo-until-date` and simple type `embargo-stages-list` were added. Element `embargo-until-date` was added to the S5, S100 and S200 schemas. Both elements are now subelements of `journal-issue-properties` or `journal-item-properties`.

2.2.9. Journal 1.8

The Journal 1.8 schemas were released on 25 January 2007.

- Support for JA DTD 5.0.2 was added.
- Three new (tombstone) PITs were added: `RET`, `REM` and `DUP`.
- Element `project-id` was added to the S5, S100 and S200 schemas.

2.2.10. Journal 1.9

The Journal 1.9 schemas were released on 2 August 2007.

- Support for delivery of RadCon auxiliary files with the three project schemas was added: value `AUXILIARY` was added to `ml-purposes-list`, value `INFOPATH` was added to `ml-versions-list` and simple type `ml-purposes-list-journal-item` was added.
- Element `embargo` was removed from the P100 and Q300 schemas.

2.2.11. Journal 1.10

The Journal 1.10 schemas were released on 7 December 2007.

- The Book PII pattern was added to `pii-patterns-current`, `pii-patterns-current-hub` and `pii-patterns-general`.
- The new Collection PII pattern was added to `pii-patterns-current-hub`.
- The Book DOI pattern was added to `doi-patterns-current` and `doi-patterns-current-hub`.
- The new Collection DOI pattern was added to `doi-patterns-current-hub`.
- The `version-number` was changed to allow stage-based version numbers.
- The `jid` element was changed to allow JIDs starting with "`BS:`".
- The weight `HEAD-ONLY` was made possible again by replacing ml-weights-list by ml-weights-list-general.

2.2.12. Journal 1.11

The Journal 1.10 schemas were released on 3 April 2008.

- Web PDF files conformant to the new web PDF specifications 6.0 will have a `pdf-version` of `1.4 6.0`.
- To prepare for thumbnails for MMCs, three new asset types were created: `IMAGE-MMC`, `IMAGE-MMC-DOWNSAMPLED` and `IMAGE-MMC-THUMBNAIL`.
- The JA DTD 5.0.1 was decommissioned, it can now only be used in project schemas.
- The value `PROJECT` for `production-process` is now allowed in the S5, S100 and S200 schemas.
- New project schemas S100-project and S200-project were added.

- The S300-H100 schema was removed.
- The new PDF purpose `EDITED-PROOF` was added in the S100 and S100-project schemas.
- Element `proof-uri` was added to element `production-details`.

2.2.13. Journal 1.12

The Journal 1.12 schemas were released on 3 September 2008.

- Support for new JA DTD 5.1.0 was added.
- ISBN-10s are now only allowed in projects.
- Longer pathnames are now possible.

2.2.14. Journal 1.13

The Journal 1.13 schemas were released on 15 December 2008.

- To support partial deliveries the value `PARTIAL-RELOAD` for element `dataset-action` was added.
- Further support for partial deliveries was added in the form of the new attribute `omitted` of element `journal-item`.
- The occurrence indicator of element `journal-item` was changed to "unbounded" in the P100, Q300(-project), S5, S100(-project) and S200(-project) schemas.
- The occurrence indicator of element `web-pdf` on issue level was changed to "unbounded" in the P100 and Q300(-project) schemas.
- Value `VIDEO-FLASH` was added as possible value for element `asset/type`.

2.2.15. Journal 1.14

The Journal 1.14 schemas (v1.14) were released on 7 May 2009.

- A new project schema s5-project was added.
- The pathname patterns were changed to improve the support of partial deliveries.

2.2.16. Journal 1.15

The Journal 1.15 schemas (v1.15) were released on 25 August 2009.

- Stage values S250 and H200 were added to the mother schema. The patterns for version numbers were adapted as well.
- The S250 schema was added to support deliveries for that stage.
- The pattern for `suppl` was changed to allow for double-digit part numbers.
- Web PDF files conformant to the new web PDF specifications 6.1 will have a `pdf-version` of `1.6 6.1` or `1.7 6.1`.
- Element `project-id` and `production-process` value `PROJECT` were added to the F300 and P100 schemas.

2.2.17. Journal 2010.2

The Journal 2010.2 schemas were released on 28 June 2010. With this release the naming convention is changed to match the EWII release names.

- Changed the pattern for element `jid` to remove "`bs:`" as an allowed prefix.
- Removed `IMAGE-NONCAP` from the list of possible value for element `asset/type`.
- Removed `1.6 6.1` from the list of possible value for element `pdf-version`.
- Element `web-pdf` with children `pathname` and `pdf-version` were added to support the delivery of journal-issue PDFs (TOCs) for S300 and S350 schemas.

2.2.18. Journal 2010.3

The Journal 2010.3 schemas were released on 26 November 2010.

- Value `1.7 6.2` was added to the list of possible values for element `pdf-version`.
- Values `1.4` and `1.4 6.0` were removed from the list of possible values for element `pdf-version`. The values can still be used for delivery of legacy material.
- The value `print 1.1` was added as a possible value for element `schema-version` in the F300 schema.

2.2.19. Journal 2011.2

The Journal 2011.2 schemas were released on 18 May 2011.

- Removed `1.7 6.1` from the list of possible value for element `pdf-version`.

The schemas were later patched (21 December 2011) to enable delivery of assets with a four-character extension.

2.2.20. Journal 2012.1

The Journal 2012.1 schemas were released on 10 May 2012.

- Added `print 1.2` to the list of allowed values for element `journal-issue/files-info/ml/schema-version`.

2.2.21. Journal 2012.2

The Journal 2012.2 schemas were released on 11 May 2012.

- Value `CRP` was added to the list of possible values for `pit`.
- Support for JA DTD 5.2.0 was added.
- Support for SI DTD 5.2.0 was added.
- Value `1.7 6.3` was added to the list of possible values for elements `pdf-version`.
- A new attribute named `cross-mark` was added to element `journal-item`.

2.2.22. Journal 2013.1

The Journal 2013.1 schemas were released on 18 October 2012.

- Value `1.7 6.4` was added to the list of possible values for elements `pdf-version`.

2.2.23. Journal 2013.3

The Journal 2013.3 schemas were released on 17 October 2013.

- Value `PAGEBREAK 5.0.0` was added to the list of possible values for elements `dtd-version`, `ml-versions-list` and `ml-versions-list-journal-item`.
- Value `PAGEBREAK` was added to the list of possible values for elements `ml-purposes-list` and `ml-purposes-list-journal-item`.
- Values `1.7 6.2` and `1.7 6.3` were removed from the list of possible values of element `web-pdf-versions-list`.

2.2.24. Journal 2014.5

The Journal 2014.5 schemas were released on 4 July 2014.

- Introduced a new pattern to support BS namespace for the element `doi-patterns-current`.

2.2.25. Journal 2014.6

The Journal 2014.6 schemas were released on 7 July 2014.

- Introduced a new pattern to support K-PII identifiers for the element `pii-patterns-current`, `pii-patterns-general` and `doi-patterns-current`.

The schemas were later patched (2 October 2014) to enable delivery of items with a DOI with a longer prefix as usual.

2.2.26. Journal 2015.1

The Journal 2015.1 schemas were released on 19 January 2015.

- Updated elements `ml-versions-list`, `ml-versions-list-journal-item` to support new DTD 5.4.0.
- Updated `journal-item-pits-list` and `dtd-version` to support new PITs.
- A new schema named S280 is created to support Virtual Special Issue (VSI) for journal content. It is available as regular and a project schema.
- Update elements `stages-list` and `version-number-patterns` to support S280
- Removed older version number pattern from `version-number-patterns`.

2.2.27. Journal 2015.3

The Journal 2015.3 schemas were released on 9 March 2015.

- Updated element `doi-patterns-current` to support Heliyon's new DOI pattern

2.2.28. Journal 2015.4

The Journal 2015.4 schemas were released on 1 July 2015.

- Updated elements `web-pdf-legacy-versions-list` and `web-pdf-versions-list` with a new value `1.7 6.5`.
- Updated element `schema-version` with a new value `print 1.3`.

2.2.29. Journal 2016.2

The Journal 2016.2 schemas were released on 11 March 2016.

- Updated pattern of elements `suppl-pattern` and `suppl-pattern-with-spinoff` to handle lengthy digits.

2.2.30. Journal 2016.6

The Journal 2016.6 schemas were released on 13 January 2017.

- Updated elements `ml-versions-list`, `ml-versions-list-journal-item` and element `dtd-version` to support new DTD 5.5.0.
- Updated `journal-item-pits-list` to support new PITs.

2.2.31. Journal 2018.1

The Journal 2018.1 schemas were released on 23 January 2018.

- Updated element `schema-version` in the F300 schema with a new value `print 1.4`.
- In the P100 schema element `journal-issue/files-info/web-pdf` was made optional.

2.2.32. Journal 2018.4

The Journal 2018.4 schemas were released on 9 May 2018.

- All DOI patterns were replaced by the general DOI pattern.

The schemas were later patched (22 August 2018) to include a journal item's `article-number`.

2.2.33. Journal 2018.6

The Journal 2018.6 schemas were released on 26 September 2018.

- Updated elements `ml-versions-list`, `ml-versions-list-journal-item` and element `dtd-version` to support new DTD 5.6.0.
- Updated `journal-item-pits-list` with new PITs introduced in DTD 5.6.0.

The schemas were later patched (14 January 2019) to support `article-numbers` in project datasets.

2.2.34. Journal 2019.1

The Journal 2019.1 schemas were released on 10 May 2019. With this release the version names no longer match the EWII release names.

- Value `1.7 7.0` was added to the list of possible values for element `web-pdf/pdf-version`.
- A list of possible values for element `print-pdf/pdf-version` was added.
- Element `print-pdf/pdf-version` was added in the F300 schema.

2.2.35. Book 1.0

The Book 1.0 schemas were released on 28 October 2003.

2.2.36. Book 1.1

The Book 1.1 schemas were released on 12 November 2003.

- The list of PITs was replaced by the list from the EHS Books DTD.
- The allowed version of the EHS Books DTD was changed to 5.1.0.

2.2.37. Book 1.2

The Book 1.2 schemas were released on 10 February 2004.

- New element `collection-item-id` was added to `book-item-unique-ids`. The new element contains two other new elements: `item-id` and `collection-id`.
- The list of values for `ml` versions was split into two lists, one for book projects and one for book items. The `FLA 4.3.1`, `MRW 1.2.0` and the `BS 5.0.0 *` values were removed while eight `EHS-BOOKS 5.1.1 *` values were added.

2.2.38. Book 1.3

The Book 1.3 schemas were released on 23 June 2005.

- Stage values `Q300`, `S350` and `H350` were added.
- Elements `isbn`, `pii` and `doi` were changed to allow for ISBN-13s.
- Element `version-number` was changed to allow for version numbers starting with 0.
- Values `JI 5.0.0` and `MRW 5.0.0 MRW` were removed as possible book project DTD versions while value `BOOK 5.2.0 BOOK` was added.

- Values `JA 5.0.0 *` and `MRW 5.0.0 *` were removed as possible book item DTD versions while eight `BOOK 5.2.0 *` values were added.
- PIT values `BK` and `SCP` were added.
- Asset type `IMAGE-COVER` was added.
- PDF purpose `COMPLETE` was added.
- The PDF property `WRAPPED` was replaced by `WRAPPED OPTIMIZED`.
- An S350-H350 schema was added to support PreCAP book deliveries. Apart from the obvious changes it is equal to the S300-H300 schema.
- In both the S300-H300 and S350-H350 schemas the element `dataset-properties` was locally defined.
- In the S300-H300 schema an optional element `web-pdf` was added to `book-project`'s subelement `files-info`.
- A Q300 schema was added for proof deliveries that precede S300-H300. It accepts assorted PDF files similar to journal P100 and its items are optional.

The schemas were later patched (12 July 2005). In the father schema the pattern for version numbers was changed to include versions up to $6.n$ and a new PDF version was added, `1.4`.

A second patch was released on 18 October 2005. Element `pathname` was changed to exclude the use of "`..`" and accept only pathnames that end in an "8+3" filename. The S350-H350 schema was changed to enable delivery of a web PDF for a book project and a mandatory raw-text file for each book item.

2.2.39. Book 1.4

The Book 1.4 schemas were released on 21 December 2005.

- Various lists of patterns and lists of values were introduced. Some elements are no longer declared in the father schema.
- To support deliveries of material for special projects three so-called project schemas were introduced. Much of the schema validation is removed from these project schemas and is performed elsewhere. Additionally element `project-id` was introduced and production process `CONVERSION` was changed to `PROJECT`.

The schemas were later patched (31 January 2006). Element `project-id` and production process `PROJECT` were added to the Q300 schema.

2.2.40. Book 1.5

The Book 1.5 schemas were released on 12 June 2006.

- Web-PDF version 1.3 was removed.

2.2.41. Book 1.6

The Book 1.6 schemas were released on 25 January 2007.

- A new schema, O300, was added for deliveries to online book stores.
- Stage O300 and production process `OLBS` were added.
- Support for new Book-metadata DTD was added.
- Support for MRW DTD 5 was added.
- Four new PITs were added: `DED`, `MRW`, `EDB` and `MIS`. The latter three were added for MRWs.

2.2.42. Book 1.6.1

The Book 1.6.1 schemas were released on 2 February 2007.

- Support for Book DTD 5.2.1 was added.

2.2.43. Book 1.7

The Book 1.7 schemas were released on 2 August 2007.

- Support for deliveries with pagebreak files (on issue and item level) was added: value `PAGEBREAK` was added to `ml-purposes-list`, `PAGEBREAK 5.0.0` was added to `ml-versions-list-item`.
- The occurrence of `ml` in the O300 schema was changed to 1..2.

2.2.44. Book 1.8

The Book 1.8 schemas were released on 7 December 2007.

- The new Collection PII pattern was added to `pii-patterns-current-hub`.
- The new Collection DOI pattern was added to `doi-patterns-current-hub`.
- The `version-number` was changed to allow stage-based version numbers.
- Element `item-id` was made optional.

2.2.45. Book 1.9

The Book 1.9 schemas were released on 3 April 2008.

- Web PDF files conformant to the new web PDF specifications 6.0 will have a `pdf-version` of `1.4 6.0`.
- To prepare for thumbnails for MMCs, three new asset types were created: `IMAGE-MMC`, `IMAGE-MMC-DOWNSAMPLED` and `IMAGE-MMC-THUMBNAIL`.

2.2.46. Book 1.10

The Book 1.10 schemas were released on 3 September 2008.

- Support for new Book-Metadata DTD 5.0.1 was added
- ISBN-10s are now only allowed in projects
- Longer pathnames are now possible
- Support for changes file was removed from the O300 schema

2.2.47. Book 1.11

The Book 1.11 schemas were released on 11 December 2008.

- To support partial deliveries the value `PARTIAL-RELOAD` for element `dataset-action` was added.
- Further support for partial deliveries was added in the form of the new attribute `omitted` of element `book-item`.
- The `MRW 5.0.0 *` values were removed as possible DTD versions.
- Ten `BOOK 5.3.0 *` values were added as possible DTD versions.
- The new PIT `COP` (copyright) was added.
- Value `VIDEO-FLASH` was added as possible value for element `asset/type`.

2.2.48. Book 1.12

The Book 1.12 schemas (v1.12) were released on 7 May 2009.

- The PIT `DCT` (dictionary) was added.
- The pathname patterns were changed to improve the support of partial deliveries.

2.2.49. Book 1.13

The Book 1.13 schemas (v1.13) were released on 25 August 2009.

- The `EHS-BOOKS 5.1.1 *` values were removed as possible DTD versions.
- Web PDF files conformant to the new web PDF specifications 6.1 will have a `pdf-version` of `1.6 6.1` or `1.7 6.1`.
- Element `project-id` and `production-process` value `PROJECT` were added to the O300 schema.

2.2.50. Book 2010.2

The Book 2010.2 schemas were released on 28 June 2010. With this release the naming convention is changed to match the EWII release names.

- Added new type `edition-pattern` for use as a restriction in element `book-project-properties/edition` for O300, Q300, S300, and S350 schemas.
- Removed `IMAGE-NONCAP` from the list of possible value for element `asset/type`.
- Removed `1.6 6.1` from the list of possible value for element `pdf-version`.

2.2.51. Book 2010.3

The Book 2010.3 schemas were released on 26 November 2010.

- Value `1.7 6.2` was added to the list of possible values for element `pdf-version`.
- Values `1.4` and `1.4 6.0` were removed from the list of possible values for element `pdf-version`. The values can still be used for delivery of legacy material.
- An F300 schema was added for deliveries of datasets with material for printing.
- A U300 and a U300-project schema were added for deliveries of (online) book updates.
- Element `item-id` was restricted to 8 digits.

The schemas were later patched (6 January 2011) to allow `item-id` to contain eight letters and digits. A second patch (2 February 2011) re-introduced the value `1.4` for element `pdf-version` in O300 datasets.

2.2.52. Book 2011.2

The Book 2011.2 schemas were released on 18 May 2011.

- Removed `1.7 6.1` from the list of possible value for element `pdf-version`.
- An E300 schema was added for deliveries of ePub files.
- The U300 and U300-project schema were removed.

The schemas were later patched (12 July 2011) to add element `book-project-unique-ids/isbn` to the F300 schema. A second patch (21 December 2011) enabled delivery of assets with a four-character extension.

2.2.53. Book 2012.1

The Book 2012.1 schemas were released on 10 May 2012.

- Added `COMPLETE-PF` and `COMPLETE-CE` to the lists `web-pdf-purposes-list` (for element `web-pdf`).
- Added `print 1.2` to the list of allowed values for element `book-project/files-info/ml/schema-version`.

2.2.54. Book 2012.2

The Book 2012.2 schemas were released on 11 May 2012.

- Value `1.7 6.3` was added to the list of possible values for elements `pdf-version`.

2.2.55. Book 2013.1

The Book 2013.1 schemas were released on 18 October 2012.

- Values `RET` and `OVW` were added to the list of possible values for `pit`.
- Added support for Book DTD 5.3.1.
- Added a new stage `S280` to the list of allowed values of element `stages-list`, and a new schema named `s280.xsd` added to support the delivery of book items without hub.
- Value `1.7 6.4` was added to the list of possible values for elements `pdf-version`.

2.2.56. Book 2013.3

The Book 2013.3 schemas were released on 17 October 2013.

- Values `1.4`, `1.7 6.2` and `1.7 6.3` were removed from the list of possible values for elements `web-pdf-versions-list` and `pdf-version`.

2.2.57. Book 2014.5

The Book 2014.5 schemas were released on 4 July 2014.

- Introduced a new pattern to support BS namespace for the element `doi-patterns-current`.

2.2.58. Book 2014.6

The Book 2014.6 schemas were released on 7 July 2014.

- Introduced a new pattern to support K-PII identifiers for the element `pii-patterns-current` and `doi-patterns-current`.

The schemas were later patched (2 October 2014) to enable delivery of items with a DOI with a longer prefix as usual.

2.2.59. Book 2015.1

The Book 2015.1 schemas were released on 19 January 2015.

- Updated elements `ml-versions-list`, `ml-versions-list-item`, `ml-versions-list-item-current`, `ml-versions-list-project`, `ml-versions-list-project-current` to support new DTD 5.4.0.
- Updated `pit-list` to support new PITs.
- Removed older version number pattern from `version-number-patterns`.
- A new schema named S200 is created to support chapter based publishing for book content. It is available as regular and a project schema.

2.2.60. Book 2015.3

The Book 2015.3 schemas were released on 9 March 2015. No changes made to this version.

2.2.61. Book 2015.4

The Book 2015.4 schemas were released on 1 July 2015.

- Updated elements `web-pdf-versions-list` and `pdf-version` with a new value `1.7 6.5`.
- Updated element `schema-version` with a new value `print 1.3`.

2.2.62. Book 2016.2

The Book 2016.2 schemas were released on 11 March 2016. No changes made to this version.

2.2.63. Book 2016.6

The Book 2016.6 schemas were released on 13 January 2017.

- Updated elements `ml-versions-list`, `ml-versions-list-item`, `ml-versions-list-item-current`, `ml-versions-list-project`, `ml-versions-list-project-current` to support new DTD 5.5.0.
- Updated `pit-list` to support new PITs.

2.2.64. Book 2018.1

The Book 2018.1 schemas were released on 23 January 2018.

- Updated element `schema-version` in the F300 schema with a new value `print 1.4`.

2.2.65. Book 2018.4

The Book 2018.4 schemas were released on 9 May 2018. No changes made to this version.

The schemas were later patched (22 August 2018). No changes made to this version.

2.2.66. Book 2018.6

The Book 2018.6 schemas were released on 26 September 2018.

- Updated elements `ml-versions-list`, `ml-versions-list-item`, `ml-versions-list-item-current`, `ml-versions-list-project`, `ml-versions-list-project-current` to support new DTD 5.6.0.
- Updated `pit-list` with new PITs introduced in DTD 5.6.0.

The schemas were later patched (14 January 2019). No changes made to the Book schemas.

2.2.67. Book 2019.1

The Book 2019.1 schemas were released on 10 May 2019. With this release the version names no longer match the EWII release names.

- Value `1.7 7.0` was added to the list of possible values for element `web-pdf/pdf-version`.
- A list of possible values for element `print-pdf/pdf-version` was added.
- Element `print-pdf/pdf-version` was added in the F300 schema.

2.2.68. Satellite 2010.3

The Satellite 2010.3 schema was released on 17 December 2010.

2.2.69. Satellite 2011.2

The Satellite 2011.2 schema was released on 18 May 2011.

- Element `dtd-version` was added.

The schema was later patched (21 December 2011) to enable delivery of assets with a four-character extension.

2.2.70. Satellite 2012.1

The Satellite 2012.1 schema was released on 10 May 2012.

- Added `project` to the list of allowed values for element `production-process`.
- Added an optional element `project-id`.
- New a300-project schema created.

2.2.71. Satellite 2012.2

The Satellite 2012.2 schema was released on 11 May 2012. There were no changes with respect to the previous version.

2.2.72. Satellite 2013.1

The Satellite 2013.1 schema was released on 18 October 2012. No changes made to this version.

2.2.73. Satellite 2013.3

The Satellite 2013.3 schema was released on 17 October 2013. No changes made to this version.

2.2.74. Satellite 2014.5

The Satellite 2014.5 schema was released on 4 July 2014. No changes made to this version.

2.2.75. Satellite 2014.6

The Satellite 2014.6 schemas were released on 07 July 2014.

- Introduced a new pattern to support K-PII identifiers for the element `pii-patterns-general`.

The Satellite 2014.6p1 schemas were released on 2 October 2014. There were no changes made to this version.

2.2.76. Satellite 2015.1

The Satellite 2015.1 schemas were released on 19 January 2015.

- Modified collection PII pattern of `pii-patterns-general`
- Removed PII pattern declaration `pii-patterns-general`

2.2.77. Satellite 2015.3

The Satellite 2015.3 schemas were released on 9 March 2015. No changes were made to this version.

2.2.78. Satellite 2015.4

The Satellite 2015.4 schemas were released on 1 July 2015. No changes were made to this version.

2.2.79. Satellite 2016.2

The Satellite 2016.2 schemas were released on 11 March 2016. No changes were made to this version.

2.2.80. Satellite 2016.6

The Satellite 2016.6 schemas were released on 13 January 2017. No changes were made to this version.

2.2.81. Satellite 2018.3

The Satellite 2018.3 schemas were released on 23 January 2018. No changes were made to this version.

2.2.82. Satellite 2018.4

The Satellite 2018.4 schemas were released on 9 May 2018. No changes made to this version.

The schemas were later patched (22 August 2018). No changes were made to this version.

2.2.83. Satellite 2018.6

The Satellite 2018.6 schemas were released on 26 September 2018. No changes were made to this version.

The schemas were later patched (14 January 2019). No changes made to the Satellite schemas.

2.2.84. Satellite 2019.1

The Satellite 2019.1 schemas were released on 10 May 2019. No changes were made to this version. With this release the version names no longer match the EWII release names.

2.2.85. Print 1.0

The Print 1.0 schema was released in August 2005.

2.2.86. Print 1.0.1

The Print 1.0.1 schema was released in July 2009.

- The print PDF type value `fpo` (“for position only”) was added.
- Element `spine-width` was corrected.
- Element `jid` was changed to accept JIDs starting with “BS”.

2.2.87. Print 1.1

The Print 1.1 schema was released in August 2010.

- The namespace was changed to contain the new version number as was the fixed value of top element `print`’s attribute `schema-version`.
- Element `jid` was changed and no longer accepts “`bs:`” as an allowed prefix.
- Optional element `fascicle/fascicle-isbn` was added.

2.2.88. Print 1.2

The Print 1.2 schema was released on 6 October 2014.

- Changed namespace and internal set-up.
- Added optional `serial-issue-properties/isbn`.
- Added optional elements `fascicle/exterior-pdf` and `fascicle/binding`.
- Added `trim-size` as mandatory subelement of `serial-issue-properties` and `book-project-properties`.
- Added mandatory subelements `print-pdf/purpose` and `print-pdf/pdf-property`.
- Changed value `cm` of `spine-width`'s attribute `unit` to `mm`.
- Added pattern to `trim-size`.

The schema was later patched (v1.2p1, 31 March 2015) to correct the trim-size pattern.

2.2.89. Print 1.3

The Print 1.3 schema was released on 18 June 2015.

- Added mandatory subelements `book-project-properties/volume-set` and `book-project-properties/pin-code`.
- Added five optional subelements to element `fascicle`: `text-paper-type`, `printing-colours`, `printing-quality`, `cover-stock` and `cover-lamination`.

The schema was later patched (v1.3p1, 21 April 2016) to extend the trim-size pattern.

2.2.90. Print 1.3.1

The Print 1.3.1 schema was released on 14 March 2017.

- Added optional subelements `print-pdf/pii` and `print-pdf/batch` (containing `batch-member/pii` elements).

2.2.91. Print 1.4

The Print 1.4.0 schema was released on 30 November 2017.

- Added mandatory subelement `serial-issue-properties/pii`.
- Added optional subelement `book-project-properties/pii`.
- Adapted patterns used for element `pii`.
- Added three optional subelements to element `serial-issue-properties`: `volume-set`, `pin-code` and `title`.
- Added optional subelements `page-totals/no-pages-colour` and `page-totals/no-pages-mono`.

Chapter 3

Deliverables of CAP and PreCAP

This chapter deals with the *deliverables* of the CAP and PreCAP workflows.

The purpose of the CAP (Computer-Aided Production) and PreCAP workflows is to produce a number of products at various stages of the lifetime of these products. Together, these products are used to build a publication, be it in print, online, or in any other media. These products are called the deliverables of the CAP and PreCAP workflows.

The CAP workflow for journals was implemented from 1996 onwards. When it began, the only deliverable was "S300plus", only for a limited number of full-length-article-like publication item types. S300plus meant that the PDF file of these issues, as well as the full-text SGML, was available electronically — very advanced at the time. This was subsequently scaled up to many more article types, delivered at various stages of the lifetime of the articles. It has by now evolved in a smooth operation, embedded in Elsevier's Global Production.

The PreCAP workflow was set up as a "quick win" beside CAP, in order to get content online fast. Printed journal issues were scanned and by means of OCR technology, SGML files of heads and, later, tails were created, until the journal was ready to move over to CAP. PreCAP still survives today, for journals that are produced outside regular Production workflows, e.g. camera-ready journals, and when back volumes need to be brought on line.

A "CAP for books" did not start until 2002, and is well under way to follow the success of CAP for journals.

This chapter defines which stages are recognised in the workflow and which version numbering applies to it. It also explains which items are delivered in electronic form.

3.1. Stages

3.1.1. Item stages

Items are the core content of journals and books. Journal articles, book chapters, editorials, indexes, glossaries, advertisements — all these are examples of items. Items are the smallest units that are, or can be, tracked in the workflow.

S5, S100, S200, S250, S280, P100, Q300 and S300 are the existing CAP deliverables for items, and S350 the PreCAP deliverable.

- S5 – The author's input material, accepted by the editorial board.
- S100 – The uncorrected proof.
- S200 – The final, corrected article.
- S250 – The final, corrected article, with final publication details (of an issue in progress).
- S280 – The final, corrected article, to support virtual collection (VSI).
- P100 – Proofs of items that are not suitable for online publication before S300, such as indexes and editorial boards.
- Q300 – Proof of the article with final publication details.
- S300 – The article with final publication details.
- S350 – A scanned copy of the printed article.

The letter "S" stands for "stage", "Q" for quality checks, "P" for proof. In order to decide what is an item, the rules of *contents entries* applies, described in Section 3.6.

Each deliverable can be delivered more than once with different version numbers. The precise composition of each deliverable is described in Chapters 6 and 7.

3.1.2. Issue stages

It is common to see a journal issue or a book series volume as a number of items packaged together. For CAP and PreCAP deliverables, however, we use this term also in a somewhat more abstract sense. An issue is the *information* needed to make up the issue. This consists of the issue's properties on the one hand, such as cover date and possible title and editors' names, and the hierarchy of the items that appear in the issue on the other hand.

Each issue that is published is a deliverable of either the CAP or the PreCAP workflow. We distinguish the deliverables H300, H350 and F300, and the deliverables P100 and Q300.

- H300 – The complete issue for electronic publication.
- H350 – The complete issue for electronic publication, derived (scanned) after the fact from the printed issue.
- F300 – The issue cover-to-cover for print publication.
- P100 – Proofs of issue items that are not contents entries (see Section 3.6), as well as items that are not desirable as S100 or S200, such as indexes and editorial boards.
- Q300 – Proof of the complete issue for electronic publication.

The letter "H" refers to "hub", as the main component of the deliverable is the issue hub, which connects the issue with its items in the proper hierarchy and contains the issue data. The letter "F" is associated with "fat" PDF files, i.e., high-quality PDF files suitable for print publication.

Implementation Note. 1. At present, S300 and S350 items will always be delivered as part of a complete issue delivery. Therefore, S300-H300 and S350-H350 deliveries have been defined that contain all the issue hub and all the items. In the future, we may see S300 items delivered on their own, e.g. if corrections need to be made to one item alone.

Remark. 1. In old terminology, S300 was used to denote the delivery of a complete issue. Formally, this is the deliverable S300-H300: the final and complete issue hub together with all its items.

3.1.3. Book project

The terminology in the books world is more ambiguous than in the journals area. There are volumes, parts, sections, and these may have multiple meanings. When these books (parts, volumes) are scheduled for production, it is decided how they will be produced. A major reference work consisting of three physical volumes can be published at the moment when one volume is ready, but also as one whole (which is currently the way). A *book project* is the term used for such a deliverable; it can, therefore, comprise one or more physical books, or it can be a continuation of an earlier book project, etc.

The CAP deliverables for book projects defined today are Q300, H300, H350 and O300.

- Q300 – The book project's main "hub" file with optional PDF proofs of pages of the printed book, all for proofing purposes in Production.
- H300 – The complete book project for electronic publication.
- H350 – The complete book project for electronic publication, derived (scanned) after the fact from the printed book(s).
- O300 – The complete book project for online book sellers.

3.1.4. Satellites

Satellites contain information about articles, issues or book chapters, or indeed about images, videos, etc. There are various types of satellites (e.g. annotations, enhancement fragments) which are all delivered in an A300 dataset.

- A300 – A collection of satellites for electronic publication.

Implementation Note. 1. An A300 dataset can contain up to 500 satellites. However, in the case of enhancement fragments only one satellite is allowed for now.

3.2. Versions

In the regular workflow, each deliverable (e.g., S100, S200, S300) will be delivered only once, but the CONTRAST standard allows for the possibility of redeliveries.

Each item, therefore, not only possesses a stage, but also a version number. The order to the supplier includes the version number that must be used.

The version number consists of two components. The first is related to the stage of the deliverable, the second is the sequence number of the delivery within that stage. However, no meaning should be derived from the version numbers — the only thing that matters is that they are unique.

As of April 2008 the version numbers will be based directly on the stage of the deliverable, e.g. S100.1, H300.3, O300.2, etc. In the past the version number was of the form $m.n$ where m was related to the stage. This system was abolished. (See [14] for more information.)

Implementation Note. 1. Within every stage every delivery except the first one can be accompanied by a changes-with-respect-to XML file. That is, when the second number of the version number is 2 or higher. This changes-with-respect-to file, describing the changes that took place, will be implemented at a later date.

2. In the first implementation of CONTRAST the items' and issue's version within an S300-H300 dataset will be identical.

3.3. File and asset types

CAP and PreCAP deliverables consist of a variety of files. Which files belong to which deliverable depends on the stage. In the dataset.xml, these files are listed under the appropriate subelement of `files-info`.

3.3.1. SGML/XML components

The main content of items and issue and book hubs is contained in the XML (or SGML) file. However, not all items are fully captured in XML: for some less important items only the head and tail or even only the title are captured. This is called the *weight* of the XML files, see Section 3.5.

`ml` is used for XML files (or older SGML files). The specifications for XML files are divided over various documents. The *Tag by Tag* documentation [4, 5, 6] is the starting point.

XML files are not complete without the external files that they reference, such as strip-ins, images and electronic components. External files declared in the entity declaration of the XML files are called *assets*. See below for the definition of asset types.

3.3.2. Web PDF files

All items possess a "web" PDF file that is published online. By definition, the PDF file of an item contains all the pages that contain a portion of the item. If pages contain portions of other material as well, then these are not suppressed: they are also visible. Note that the page ranges may well be non-contiguous if the item is spread over the (printed) issue — in online products, they appear only once at their first occurrence. A special case is when colour images are collected in a colour plate section. The colour plate section itself is rarely a contents entry; the pages with plates belonging to the item are included in the item's PDF file per the definition above.

Electronic publishing made it possible to publish items with more online content than print content. These items are called "e-extra" and have two PDF files. The main PDF file contains all the material to be published online and the so-called main abridged PDF file contains all the material to be published in print (as per the specifications above). The latter PDF file is only delivered in stages S100, S200 and P100.

`web-pdf` is used for PDF files that satisfy the CAP criteria for "web PDF" files. This includes the PDF files containing the author query form, the graphical abstract and the stereo-chemistry abstract. For the web PDF specifications, see [2].

3.3.3. Print PDF files

`print-pdf` is used for PDF files that satisfy the CAP criteria for print publication (these files are also known as "fat PDF" files).

3.3.4. Raw text files

`raw-text` is used for ASCII files with the text of the document. This is a component of a PreCAP delivery.

3.3.5. Asset types

Assets are the external files associated with an XML file declared in the entity declaration, e.g. image files or electronic components. Exactly which kind of files may be associated with an item or issue/book hub and which criteria they need to satisfy is described in [7, 8, 9].

The following types of assets are defined.

- `APPLICATION` is used for files belonging to computer applications. It includes Microsoft Word, Microsoft Excel, Adobe PDF files, zip files. Note that a dataset (see Chapter 4) may contain at most 20 zip files.
- `AUDIO` is used for audio files, such as MP3 or WAV files.
- `IMAGE-CAP` is used for files that satisfy the CAP specifications for artwork, with the exception of cover images, [1].
- `IMAGE-DOWNSAMPLED` is used for images (of type `IMAGE-CAP`) that were downsampled at the Electronic Warehouse. This type is only to be used in projects.
- `IMAGE-MMC` is used for the original artwork (e.g. a movie still) that is converted to a thumbnail for an MMC, i.e., an asset which has a type not starting with IMAGE-.
- `IMAGE-MMC-DOWNSAMPLED` is similar to `IMAGE-DOWNSAMPLED` but it is created from an `IMAGE-MMC`. This type is only to be used in projects.
- `IMAGE-MMC-THUMBNAIL` is used for the thumbnail images created from images of type `IMAGE-MMC`. This type is only to be used in projects.
- `IMAGE-NONCAP` is used for artwork files that do not satisfy the specifications for artwork.
- `IMAGE-COVER` is used for images of issue or book covers according to the specifications for covers.
- `VIDEO` is used for movie files, such as MPEG files.
- `VIDEO-FLASH` is used for Flash movie files. These files are the result of a conversion of a movie file that must also be present in the dataset.
- `XML` is used for XML assets, such as SVG files.

3.4. Batches

A batch is a set of items. One item acts as the representative of the batch. This item can be a real item or a placeholder. The batch placeholder is only used in Production and will not be delivered to customers.

The two main purposes of batches are to support items with add-on items and to support sections of abstracts, news items, etc. In the first case the main item is the representative of the batch and in the second case a special item is created to act as the batch representative. For more information see [13].

There are two reasons for using batches:

- Simplified production: batch items need not be tracked individually but only as a batch. They do not need separate login or separate compilation into an issue.
- Better online contents lists, where batch item can be shown collapsed (only the batch representative is shown) or expanded.

Implementation Note. At present, it is required that a batch is delivered complete, i.e. the batch representative together with all the batch items. By convention, the batch item's directories are always subdirectories of the batch representative's directory (see Section 4.1).

Since batch items are in fact items in their own right, they possess an XML file and a web PDF file just like any item, as described in Section 3.3. Hence, the PDF file consists of all item pages that possess a portion of the batch item (portions of the batch representative and/or other batch items may be visible as well).

An item is a batch item if and only if the dataset description file, dataset.xml, contains a `batch` element for the batch item pointing to the representative item. Hence, `batch` must be present for all batch items.

When an item is a batch representative, its "type" (this attribute of `journal-item` is explained in Chapter 6) must be different from the default value `stand-alone`. In the case of add-on items the value must be `with-add-ons` and in the case of e.g. abstract sections the value must be `batch-placeholder`.

The batch representative item possesses a web PDF file that includes the item *and* all batch items.

The batch representative item and batch items must have the same publication item type (PIT).

Note about future expansion. The specification supports batch items whose batch representative is also a batch item for future expansion. This is to support advanced hierarchies of items.

3.5. Weights of CAP deliverables

CAP item deliverables exist in a number of weights: `FULL-TEXT`, `HEAD-AND-TAIL` (known as CAPLitePlus), and `CONTENTS-ENTRY-ONLY` (known as Ultralight).

The weight indicates to what extent the text of the item is captured in XML. The majority of Elsevier's electronic products is captured in full-text XML, a small number as CAPLitePlus.

It is not true that an XML file with only a head and a tail is necessarily of weight `HEAD-AND-TAIL`. This can only be verified by comparing the XML file with the PDF file of the item.

The *Tag by Tag* [4] describes which elements in the XML file belong to which weight.

3.5.1. Default serial item weight assignment

The default assignment for weights for items of serial publications (journals and book series) depends on the publication item type (PIT) and the production type and is as given in Table 1. PITs are described in [3].

Table 1: PIT + production type gives default weight

PIT	Production type	Default weight
ADD, BRV, CNF, COR, DIS, EDI, ERR, EXM, FLA, PRP, PRV, REQ, REV, SCO, SSU	`NON-CRC`	`FULL-TEXT` or `HEAD-AND-TAIL`[a]
ADD, BRV, CNF, COR, DIS, EDI, ERR, EXM, FLA, PRP, PRV, REQ, REV, SCO, SSU	`CRC`	`HEAD-AND-TAIL`
ABS, ADV, ANN, CAL, CON, EDB, IND, LIT, MIS, NWS, OCN, PNT, PUB	`NON-CRC`	`CONTENTS-ENTRY-ONLY`
ABS, ADV, ANN, CAL, CON, EDB, IND, LIT, MIS, NWS, OCN, PNT, PUB	`CRC`	`CONTENTS-ENTRY-ONLY`

[a] Depending on the product specification.

If Elsevier's default requirements are to deliver a certain weight, then it is never an error to deliver a heavier weight and always an error to deliver a lower weight. Delivery of heavier weight than the default will occur by separate agreements between Elsevier and the supplier.

In the next section, the concept of contents entries is described. Parts of the issue that are not contents entries, called ancillary material, are not part of any delivery. Weights do not apply to ancillary material, and therefore the rule that heavier-than-default is allowed does not apply either.

3.6. Contents entries

The printed journal and the electronic journal differ in that the printed journal contains cover pages, preliminary pages, etc., that do not belong to the online version. For instance, a web version of the journal may have a homepage for the journal with the Aims and Scope and the Instruction for Authors — in a way, the homepage and the other pages around the content are the online alternatives of the cover and preliminary pages.

To decide which items require electronic delivery, a simple rule of thumb applies: If the item appears in the (paper) table of contents, then it belongs to the electronic delivery. If it does not, it is not delivered. The editorial board, when it appears in the issue, is an exception to this rule: it is always delivered with publication item type `EDB` whether or not it is listed in the table of contents.

Items that appear in the electronic delivery are called *contents entries*. The remainder of the journal issue consists of *ancillary material*. That material, when compiled in the PTSIII system, receives the pseudo publication item type `ZZZ`. In particular, the table of contents itself, advertisements, and most front- and backmatter items do not appear online since they do not occur in the table of contents.

The precise rule for the supplier is to deliver an item to the Electronic Warehouse if and only if the PTS order states that its PIT is not `ZZZ`.

3.7. Split items

Items can be split. That is, an item's pages can be non-contiguous in the issue or book project. For instance, an item can appear on pp. 31–40 and p. 68. The latter page can contain for instance a colour plate belonging to the item, or just the last part of the item.

In the case of a journal item the issue order will contain the item's information in two "row"s. The two rows only differ in the page information. (See p. 107 for more information.) The issue order will contain the following:

```
<row>...
  <page-from>31</page-from><page-to>40</page-to>
  <pdf-pages>11</pdf-pages>...
</row>
...
<row>...
  <page-from>68</page-from><page-to>68</page-to>
  <pdf-pages>11</pdf-pages>...
</row>
```

In the dataset such an item appears as one item. The PDF file contains all the pages belonging to that item. The dataset.xml file will contain the following:

```
...
<production-details>
  ...
  <pdf-pages>11</pdf-pages>
  ...
</production-details>
...
```

In the issue hub it also appears as one item. However, there it appears with multiple page ranges:

```
...
<ce:include-item>
  ...
  <ce:pages><ce:first-page>31</ce:first-page>
            <ce:last-page>40</ce:last-page></ce:pages>
  <ce:pages><ce:first-page>68</ce:first-page></ce:pages>
</ce:include-item>
...
```

The place it appears is the place of the first occurrence of a part of the item.

Note that the page-ranges in the issue hub must be contracted where possible. Take for example an article that is printed on pp. 91–124 and where p. 123 contains a half-page advert. In PTS there are three rows for this article and this advert: one for the item, pp. 91–123, one for the advert, p. 123, and a second one for the item, p. 124. The PDF file for this item has 34 pages (one of which also contains the advert). In the issue hub there is *one* page-range: 91–124.

Chapter 4
CONTRAST datasets

CONTRAST (*con*tent *tra*nsport *st*andard) is Elsevier's standard for dataset deliveries from the supplier to the Electronic Warehouse.[1] A dataset is the name given to a collection of electronic content, transported from one place to the other, in particular from supplier to the Electronic Warehouse.

CONTRAST has the following three components:

- The dataset directory structure. The structure and naming conventions are different compared to EFFECT datasets.
- The dataset delivery protocol.
- The accompanying file describing the dataset, called dataset.xml.

The CONTRAST dataset.xml file is an XML file, that validates against a W3C schema. For each deliverable, there is a separate schema. We have used schemas instead of DTDs so that we could make use of their datatyping functionality.

1. CONTRAST is also used for dataset deliveries from EW to online repositories. That is described in other documentation.

4.1. CONTRAST directory structure

Content in a CONTRAST dataset is organized in a directory structure. In this section conventions for the directory structure are described.

4.1.1. Rules versus conventions

It is of utmost importance to bear in mind that these conventions are only there to help human inspection of a dataset. Systems must not draw any conclusion whatsoever from the directory names and the directory structure. The *only* reliable source of information is the dataset.xml file that is part of every dataset.

For instance, the convention may describe that items are stored within a directory based on the PII. The *rule*, however, is to look up the PII in the dataset.xml, and then read the `pathname` of the item, and use the file found there. The system should work equally well if the item resides in a directory `foo`.

Another example: the convention may be that the main XML file is called `main.xml`. Systems must, however, never go to the directory of the item and search for a file with that name. Instead, they must inspect the dataset.xml file and look for the appropriate manifestation. The `pathname` will then lead to the main XML file.

This small programming overhead makes content transport extensible and leaves room for future expansion or revision.

Note: The convention that the main XML file is called `main.xml` has become a rule.

4.1.2. Which files are listed in the dataset.xml file?

The following files are never listed in the dataset.xml file:

- the dataset.xml file itself, that must be present within the top-level directory of each dataset;
- the fingerprint files coming from the Elsevier validation tools;
- the strip-in files called in by XML files.

All other files are mentioned in the dataset.xml.

4.1.3. File and directory names

With *file name* we mean the file name inclusive extension, excluding any directory names. With *path name* we mean the full path name of the file. In case the file is present in the dataset it is the path name in the dataset (that is, relative to the top-level directory of the dataset). In case the file is not present in the dataset it is the path name in the system it resides in (that is, an absolute path name, starting with a server name).

File and directory names follow standard Unix rules. File and directory names are case sensitive. If it is specified below that a directory is called `fp` then it consists entirely of lowercase characters. If it is specified that the directory name is a PII, then the first letter and the check digit(s) `X` must be uppercase.

Note: To support "partial deliveries" most of the restrictions on element `pathname` were removed. However, names of files and directories should not contain the backslash character.

While the files must be named with the right case, it should be noted that datasets need to be valid on any computer platform. For this reason it is not allowed to have files or directories within a directory whose names differ only in the case of one of more letters.

4.1.4. Dataset package file

A dataset consists of a directory tree with files on different directory levels. The whole dataset, when delivered to and from the Electronic Warehouse is packed into a ZIP file or a (possibly gzipped) tar file, or, for very large datasets, it can be split over more than one of such files (see Chapter 10 for more information). The package file follows an 8+3 filename convention, because it may have to be transported via restrictive channels, e.g. on an ISO 9660 CDROM. All the files mentioned in the dataset.xml file need to follow that convention as well, except for issue PDFs in P100 and Q300 deliveries.

Convention. The dataset is named with the last eight characters of the content of `supplier-dataset-id`.

No directory within the dataset may be empty. Note that a dataset (see Section 3.3.5) may contain at most 20 zip files.

4.1.5. Dataset base directory

Each dataset must be contained within a single top-level directory, called the *base directory*. (All pathnames are relative to this directory.) The dataset.xml file is present within that directory and is called `dataset.xml`. Hence, when the ZIP or tar files are all unpacked, the result is a single directory (with files and subdirectories).

Convention. The top-level directory of the dataset has the same name as the content of `supplier-dataset-id`.

4.1.6. Fingerprints

The Elsevier validation tools check various files in the dataset, such as XML files, artwork files and PDF files. Their findings are captured within *fingerprint* files. These files are used by the recipient to check whether the files were all right, or whether they need to be looked at when a check was skipped.

Rule. Each directory that contains files that are subject to validation, contains a subdirectory `fp` containing the fingerprints belonging to those files.

As a consequence, `fp` subdirectories will appear within many directories.

4.1.7. XML files and their assets and strip-ins

XML files structured with any one of the DTD 5 family of DTDs come with zero or more strip-ins and zero or more *assets*. Assets are graphic files or electronic components called in by the XML file and declared as external entities in the DOCTYPE declaration of the XML file.

Rule. The strip-ins belonging to a file called *file*`.xml` are located in a subdirectory called *file*`.stripin`.

Rule. The assets belonging to a file called *file*`.xml` are located in a subdirectory called *file*`.assets`.

If there are no assets or strip-ins, then these directories are not present.

4.1.8. S5, S100, S200, S250 and S280 serial items

Convention. If items in serial publications are part of an S5, S100, S200, S250 or S280 delivery, then they reside in a directory, called the *item directory*, within the base directory. The name of that directory is the PII of the item without dashes and parentheses.

Therefore, a typical item dataset looks like this:

```
jmi00434/
    dataset.xml
    fp/dataset_xml_fp.xml
    S0022404903002780/main.xml
                      main.pdf
    S0022404903002780/fp/main_xml_fp.xml
                         main_pdf_fp.xml
    S0022404903002780/main.stripin/si1.gif
                                   ...
                                   si137.gif
    S0022404903002780/main.assets/gr1.tif
                                  ...
                                  gr17.jpg
    S0022404903002780/main.assets/fp/gr1_tif_fp.xml
                                     ...
                                     gr17_jpg_fp.xml
```

Convention. If the item is a *batch item* (see Section 3.4) then one of two possibilities may exist. The first possibility is that the batch item travels independently. Then the (batch) item directory resides within the base. The second possibility is that the batch item travels with the batch representative item and fellow batch items. In that case the batch item's directory resides within the batch representative item directory. In all these cases, the conventions for naming by PII and structure still apply.

Implementation Note. The option to let batch items travel alone is reserved for future expansion. Current implementations require the batch representative item and its batch items to be despatched together and hence only the second possibility mentioned above is allowed.

4.1.9. P100, Q300, S300 and S350 items, and serial issues

Convention. In case of an issue in a serial publication, i.e., a journal issue or a volume of a book series, the base directory contains a subdirectory with the ISSN of the serial without punctuation. Within that directory there is a directory, called the *issue directory* with the volume/issue number in VIS format. In this format, a directory name is created from the volume/issue number by prepending the name with a `v`, replacing the slash by `i`, and introducing an `s` before the `suppl`.

```
"v" vol-first [ "-" vol-last ]
  [ "i" iss-first [ "-" iss-last ] ]
  [ "s" suppl ]
```

Examples:

Volume/issue	Directory name
37C	v37sC
37/2	v37i2
37I2	v37sI2
37/1–3	v37i1-3
37S2	v37sS2
37S131	v37sS131
37–39C	v37-39sC
37/1S	v37i1sS
37PA	v37sPA

P100, Q300, S300 and S350 items reside within the issue directory. They are named with the item's PII number without punctuation and follow the conventions for item directories.

The issue hub file also resides in the issue directory. The issue hub file may or may not contain (inline) graphics or strip-in images, so there may be asset and strip-in directories belonging to the hub file (called `issue.assets` and `issue.stripin` if the issue hub file is called issue.xml).

Hence, a typical S300-H300 dataset looks like this (we have left out the `fp` and other subdirectories for brevity):

```
jmi00435/
    dataset.xml
    00224049/v188sC/issue.xml
    00224049/v188sC/issue.assets/cover.tif
    00224049/v188sC/S0022404903002159/...
    00224049/v188sC/S0022404903002068/...
    00224049/v188sC/S0022404903002172/...
    00224049/v188sC/S0022404903002160/...
    00224049/v188sC/S0022404903002081/...
    ...
```

Figure 1 shows the directory structure in an example S300-H300 dataset.

4.1.10. Book projects

Convention. The base directory contains a directory for the book project, named using the unformatted ISBN. Within that directory, the following can be found.

- `main.xml`, the hub file for the book project;
- `changes.xml`, an optional file describing changes with respect to an earlier version after a correction;
- `main.stripin`, the strip-in directory belonging to the hub file;
- `main.assets`, the assets directory belonging to the hub file;
- `changes.stripin`, the strip-in directory belonging to the changes XML file;
- `changes.assets`, the assets directory belonging to the changes XML file;
- `front`, a subdirectory containing all the book item directories within the frontmatter of the book project, named and structured as described above;
- `body`, a subdirectory containing all the book item directories in the body of the book project, named and structured as described above;
- `rear`, a subdirectory containing all the book item directories in the rear of the book project, such as any back-of-the-book index, named and structured as described above;

- `repository`, a subdirectory, with possible subdirectories, containing unlisted additional components, such as PDF files, typeset files and Word files, that may be requested by the Book Production departments.

Hence, a typical Books dataset looks like this (we have left out the `fp` and other subdirectories for brevity):

```
jmi00436/
    dataset.xml
    072163950X/main.xml
    072163950X/front/B072163950X100278/...
    072163950X/body/B072163950X100011/...
    072163950X/body/B072163950X100023/...
    072163950X/body/B072163950X100035/...
    ...
    072163950X/body/B072163950X100229/...
    072163950X/rear/B072163950X100230/...
    072163950X/rear/B072163950X100242/...
    072163950X/rear/B072163950X100254/...
    072163950X/rear/B072163950X100266/...
    072163950X/repository/...
```

Figure 1 shows the directory structure in an example Books dataset.

Figure 1: Left – An example of an S300-H300 serial issue dataset directory structure. Right – An example of a Books dataset directory structure.

4.2. The dataset.xml file

Each CONTRAST dataset has a dataset.xml file in its base directory. It validates against a schema. For each deliverable a different schema exists, e.g. the S100 journal item schema.

The dataset.xml file has `dataset` as its top element, with three children: `dataset-unique-ids`, `dataset-properties` and `dataset-content`. The latter element is described in the subsequent chapters, it is different for each deliverable.

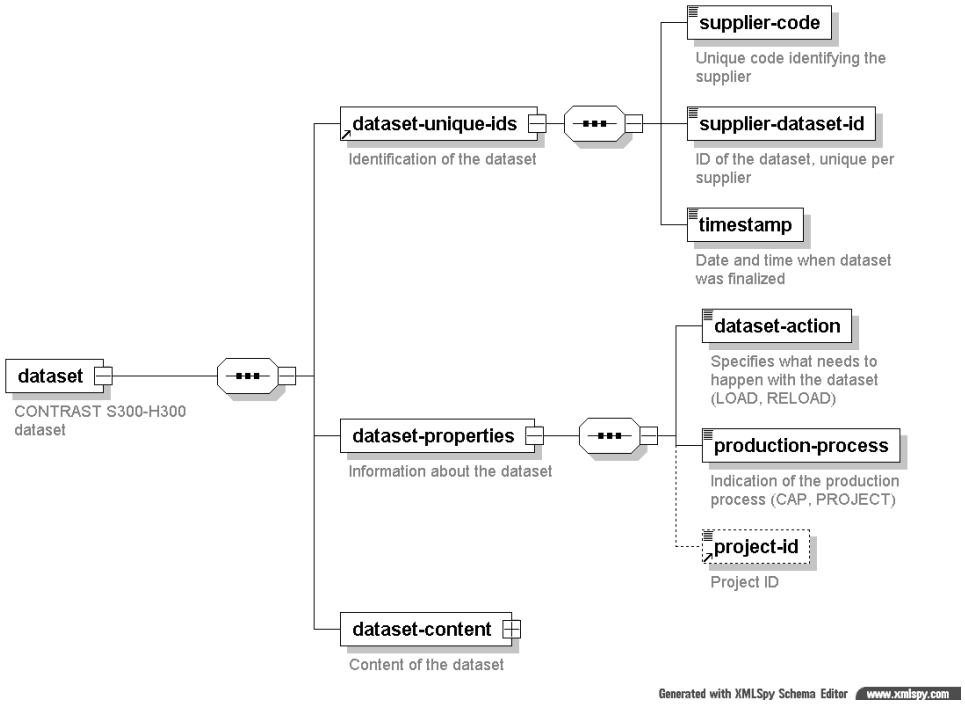


Figure 2: Top structure of the dataset.xml file.

Figure 2 shows the top structure of the dataset.xml.

4.2.1. Elements related to dataset identification

dataset-unique-ids

All datasets can be uniquely identified by a set of identifiers, which are contained within the element `dataset-unique-ids`. Alone, the three mandatory subelements `supplier-code`, `timestamp`, `supplier-dataset-id`, do not uniquely identify the dataset, but together they do.

dataset-unique-ids/supplier-code

The element `supplier-code` contains the unique code of the supplier of the dataset. This is by definition the sender of the dataset. This can be a typesetter, a data conversion house, but also the Electronic Warehouse. The Electronic Warehouse maintains the list of allowed supplier codes.

`supplier-code` is an unrestricted W3C schema token, `xs:token`.

dataset-unique-ids/supplier-dataset-id

Suppliers of datasets are responsible for assigning an identifier to each dataset they despatch. This identifier is contained in `supplier-dataset-id`. It is an unrestricted W3C schema token, `xs:token`. The identifier must be unique per `supplier-code`. It has been agreed that the dataset identifier contains at most 12 characters. Suppliers are free to devise any identification scheme they desire for their datasets.

Two different suppliers may well despatch a dataset with the same `supplier-dataset-id`, but the same supplier may never despatch two different datasets with the same id. Even if a dataset arrives corrupted, and a resubmission is requested, the resubmission must carry a new `supplier-dataset-id`.

dataset-unique-ids/timestamp

The timestamp when the creation of the dataset was completed is captured with `time-stamp`. Its content is in W3C schema DateTime format, `xs:dateTime` and should be accurate to the second. If the time zone is absent, UTC is assumed. Local time with offset from UTC is also allowed.

It is important that the timestamp is correctly captured. Say, for instance, in India a dataset is completed at 15:50 local time. It is then transported to the EW and imported, say ten minutes later. The EW import is located in Oxford, UK, which is $4\frac{1}{2}$ hours behind in the Summer. If no time zone and no UTC offset is used, the dataset will be imported (at 10:30 UTC) before it was finished (at 15:50 UTC). Therefore, it is imperative that the time is captured as

```
<timestamp>2005-05-13T15:50:00+05:30</timestamp>
```

4.2.2. Dataset properties

When a dataset arrives, a certain action must be performed by the recipient. Note that the action relates to the *whole* dataset. It is not possible to have different actions performed on different parts of the dataset.

Deliveries go by *stage*, e.g. S200 or S300. The recipient may decide what to do with existing content of an earlier stage. (The word “stage” is used not just to indicate a point in time, it has become the term for a certain CAP delivery, see Section 3.1.) Within each stage new versions may be despatched. Such a new version is called a *correction*. The correction retains the same stage and same PII, but has, of course, an increased version number.

An *update* takes place when, e.g., a chapter of a book is rewritten. Such an update requires a *correction* of the hub file and a new delivery of a chapter.

dataset-properties/dataset-action

The element `dataset-action` indicates the action to be performed. The value `LOAD` is used under ordinary circumstances. If the dataset is the result of a redelivery request, then the value `RELOAD` is to be used.

In case of a partial delivery of an issue or book, the value `PARTIAL-RELOAD` is to be used. Partial deliveries are not possible for S5, S100(-project), S200(-project), S250, P100 and O300 deliveries. Items that are not delivered should still be present in the dataset.xml file as

usual, except that element `journal-item` or `book-item` should have an attribute `omitted` with value "`true`". Files for those items should be omitted from the dataset. The hub file should always be delivered.

dataset-properties/production-process

The element `production-process` indicates the workflow used for the document. It can take the values `CAP`, `PRECAP`, `PROJECT` and `SCP`.

`PRECAP` is used only if the delivery is S350 and H350.

`PROJECT` can be used in S5, S100, S200, S250, S280, Q300, P100, S300-H300, S350-H350, F300, O300 and A300 deliveries, in conjunction with element `dataset-properties/project-id`.

`PROJECT` is used in project schemas where it is the only allowed value.

`SCP` (satellite CAP production) is used in the satellite schema.

dataset-properties/project-id

In case the dataset is delivered as part of a project this element must be used and contain the Project ID.

4.2.3. Example

XML

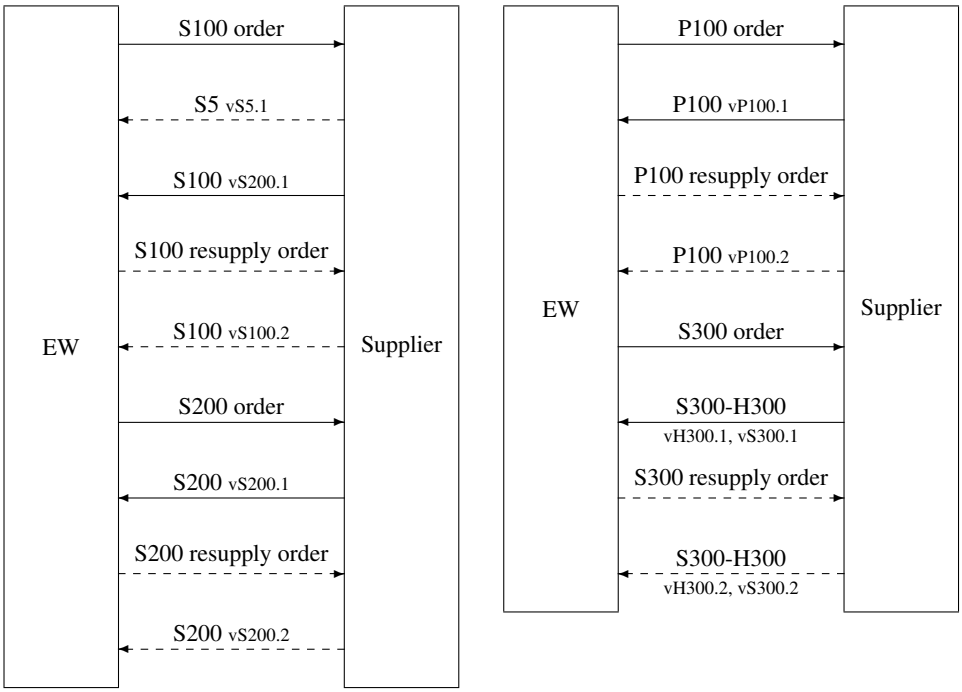
```
<dataset
  xmlns="http://www.elsevier.com/xml/schema/transport/journal-2015.3/s200"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation=
    "http://www.elsevier.com/xml/schema/transport/journal-2015.3/s200
     http://www.elsevier.com/xml/schema/transport/journal-2015.3/s200.xsd"
  schema-version="2015.3">

  <dataset-unique-ids>
    <supplier-code>Typesetter Code</supplier-code>
    <supplier-dataset-id>A20030617001</supplier-dataset-id>
    <timestamp>2014-12-07T12:33:00+02:00</timestamp>
  </dataset-unique-ids>
  <dataset-properties>
    <dataset-action>LOAD</dataset-action>
    <production-process>CAP</production-process>
  </dataset-properties>
  <dataset-content>
    <journal-item>
    .
    .
    .
    </journal-item>
  </dataset-content>
</dataset>
```

Chapter 5

CAP journal dataset orders

Each dataset *delivery* is preceded by a dataset *order*. Such an order can take different forms, from the ad-hoc agreements for conversion projects to the system-generated XML orders used in the CAP journal workflow that are described in this chapter.



The left-hand diagram above shows a typical item workflow, and the right-hand diagram a typical issue workflow. The dashed arrows indicate optional (and repeatable) steps. An S5 delivery is ordered along with the S100; the S100 order will indicate that S5 is also required. In the S300-H300 deliveries the items have a version number of the type S300.n while the hub files have a version number of the type H300.n.

5.1. Order components

An XML order file combined with source data needed to produce the ordered deliverable (when available) form the components that belong to an order. The EW places both data and order in the appropriate directory for the supplier (see [10] for a detailed description).

An XML order is an XML file stuctured according to the PTSIII Order DTD. It is used for distributing item metadata to the typesetters, the printers or the binders. When an order is received, the "executor" (the supplier) can continue with the following step in the workflow of the item or issue.

An item order is called *_PII_stage*.xml. For example:

_S0375960104001379_S100.xml
_S0375960104001379_S100RESUPPLY.xml
_S0375960104001379_S200.xml
_S0375960104001379_S200RESUPPLY.xml

Subitems do not have their own orders. If an item is a parent item, its order contains the information of all its subitems.

An issue order is called *issue_id_stage*.xml, where *issue_id* is the PTS physical issue ID. For example:

5608_P100.xml
5608_P100RESUPPLY.xml
5608_S300.xml
5608_S300RESUPPLY.xml
5608_F300.xml
5608_F300RESUPPLY.xml

Source data is the collection of files that the author submitted with text, artwork and other electronic material. The files are collected in a ZIP file for use by the supplier. The filename is that of the accompanying order. For example:

_S0375960104001379_S100.zip
_S0375960104001379_S100.xml

The source data is placed in the appropriate directory before the order.

- The S100 order is sent to the typesetter when the workflow step "Structure & Enrich" is started. The result of this step is a delivery of an S100 dataset with a structured (SGML or XML) file and accompanying PDF file.
- The S200 order is sent when the step "Perform Final Enrichment" is started. The result of this step is a delivery of an S200 dataset with a corrected SGML or XML file and the accompanying PDF file.
- The P100 order is sent to the typesetter when the step "Physical Issue Compilation" is started. The result of this step is a P100 dataset, i.e. a dataset containing proof material for an issue (cover, index, etc.). Sometimes this contains the full issue.
- The S300 order is sent when the step "Finalize Issue Electronic" is started. This results in an S300 dataset with unchanged SGML or XML files and paginated accompanying PDF files.
- The last order sent to the typesetter is the F300 order when the step "Finalize Issue Fat PDF" is started. The result is an F300 dataset containg the PDF files necessary for printing an issue.

- A print order is sent to the printer when the step “Print Issue” is started.
- A bind order is sent to the binder when the step “Bind Issue” is started.
- The offprint order is sent to the (offprint-)printer when the step “Print & Finish” is started.
- Finally the “send” orders are sent to the warehouse when the step “Send” is started.

Note that in case of some steps a problem can be raised and solved, and the step is then started again. In cases where the printer and the binder are the same, there is one step in the workflow but two orders are sent.

In case of a workflow reset a new order is sent with a higher version number. This is not a resupply.

The EW can also request a new supply, for instance when a dataset is corrupted during file transfer. This is not a resupply. The same dataset should be delivered again with the same version numbers and action.

The general part of the orders is described below; the S100, S200 and F300 orders are described in Section 6.1 (p. 50); the P100 and S300 orders are described in Section 6.12 (p. 99); the print/bind orders are described in Section 6.20 (p. 155).

5.2. General supplier order

The top element of this DTD is element `orders` which contains one or more `orders`. In practice, it will always contain exactly one order. Below, we traverse the DTD starting from the element `order`.

order

The element `order` contains subelements `time`, `po-number`, `due-date` (optional), `prod-site`, `opco`, `stage`, `bam-stage` (optional), `executor` (possibly more than one), followed by one of `item-info`, `issue-info`, `print-bind-info` or `issue-labels-info`.

order/time

This empty element contains the (local) time of order creation in six attributes.

```
<time day="22" month="04" yr="2004" hr="14" min="42" sec="52"/>
```

order/po-number

The element `po-number` contains the purchase order number. This is a unique number for each typeset order (prefix "T"), print/bind order (prefix "P") and offprint order (prefix "R"). This number has to be displayed on all relevant invoices.

order/due-date

The optional element `due-date` contains in subelement `time` the date and time the order is due. This date is in UTC (i.e. in GMT).

```
<due-date>
  <time day="07" month="04" yr="2004" hr="12" min="00" sec="00"/>
</due-date>
```

order/prod-site

The element `prod-site` contains the Elsevier Production location where the journal is produced. The following codes are used:

Code	Name
CELL	Elsevier Cell Press New York
ESEO	Elsevier UK Primary Production
ESI	Elsevier New York
ESIL	Elsevier Shannon
ESJE	Elsevier Jena, Germany
ESME	Elsevier SAS
ESNL	Elsevier B.V.
ESPH	Elsevier Philadelphia
ESSD	Elsevier San Diego
ESST	Elsevier St. Louis

order/opco

This element contains the “Operating Company”, i.e. the Elsevier company that receives the invoice.

order/stage

The element `stage` defines a deliverable that is being ordered. It is empty but has a mandatory attribute `step` which can have one of the following 34 values:

- `RSVP`: No delivery is needed, only a signal to close this step.
- `CU`: This is the order to check the completeness and usability of S0 files. In this case, the element `item-info` is present in element `order`.
- `PROOF`: This is an item order and takes the place of S5/S100 in the first phase of 24h publication.
- `S5`, `S100`, `S200`, `S250`: In these cases, the element `item-info` is present in element `order`. For historical reasons the order is called the “e-coversheet”.
- `S300`, `P100`, `Q300`: In these cases, the element `issue-info` is present in element `order`. The order is called the “issue pagination sheet”.
- `S350`: The order is for this value as yet undefined.
- `F300`: This is the order for “Fat PDF” files. In this case, the element `item-info` is present in element `order`.
- `S5RESUPPLY`, `S100RESUPPLY`, `S200RESUPPLY`, `S250RESUPPLY`: Apart from the value of `step` this order is equal to the `S5`/`S100`/`S200` order. Note that in these cases element `dataset-action` in the dataset.xml file should have value `RELOAD`.
- `S300RESUPPLY`, `P100RESUPPLY`, `Q300RESUPPLY`, `F300RESUPPLY`: Apart from the value of `step` this order is equal to the `S300`/`P100`/`Q300`/`F300` order. Note that in these cases element `dataset-action` in the dataset.xml file should have value `RELOAD`.
- `FINALXML`, `FINALXMLRESUPPLY`: This is the order for creating the XML file only. The PDF file is created by a dedicated service, based on the XML file.
- `PRINT`, `BIND`: In these cases, the element `print-bind-info` is present in element `order`. The order is called the “print/bind order”.
- `OFFPRINTS`: In this case, the element `issue-info` is present in element `order`. The order is called the “offprint sheet”. Note that, even though the content is the same as step S300, the PTSIII stylesheet will show only information relevant to offprints.
- `E-OFFPRINT`: In this case, the element `item-info` is present in element `order`. The order is sent to the EW who then will deliver an S300 PDF file to the Author Gateway who in turn will send it to the author.
- `ISSUE-LABELS`: In this case, the element `issue-info` is present in element `order`. The order is called the “issue processing form”. Again, the PTSIII stylesheet will show only information relevant to issue labels. This order is not sent to suppliers.
- `SEND-ISSUE`, `SEND-OFF-SYSTEM-ISSUE`, `SEND-OFFPRINTS`: In these cases, the element `issue-info` is present in element `order`. In the second case the issue labels arrive from an outside source, e.g. a society. These orders are sent to the warehouse. These values are included for future expansion and will not be used until further notice.
- `PUBACC`, `PUBACCRESUPPLY`: In these cases, the element `item-info` is present in element `order`. This order is sent via the EW to a dedicated supplier who will upload PDF files to a special site. No delivery of material to the EW is necessary.
- `ITEM-OFFPRINTS`, `SEND-ITEM-OFFPRINTS`: In these cases, the element `item-info` is present in element `order`. The first order is for printing item-based offprints, on demand. The second order is for despatch of these printed-on-demand offprints.

order/bam-stage

This element contains the VTW stage of the item (in the VTW captured in the property `bam:stage`). The values are `final` and `corrected_proof`.

order/executor

The element `executor` contains the code, the name and affiliation of the executor in subelements `exec-code`, `exec-name` and `aff`, respectively. It has a mandatory attribute `type` with possible values `ES` (Elsevier), `PSP` (Production Supervisor Printer or Local Supplier Manager), `AUTHOR-GATEWAY`, `TYPESETTER`, `COMPOSITOR`, `PRINTER`, `BINDER`, `WAREHOUSE`, `OFFPRINTS-SENDER`, `PUBACC-SUPPLIER`, `CU-SUPPLIER` and `FINISHER`. It has a second mandatory attribute `addressee` with possible values `yes` and `no` (the default). An order generally contains several executors. A value of "`yes`" indicates that the order is meant for that executor.

order/executor/aff

The affiliation is captured in (optional) subelements `organization`, `institute`, `inst-contd`, `address`, `address-contd`, `zipcode`, `cty`, `cny`, `tel`, `fax` and `ead` (of which there may be more than one). These correspond to fields in the PTS database.

Element `zipcode` has an attribute `zipcode-pos` indicating where the zipcode should be represented with respect to the elements `cty` and `cny`. Its values are `NONE`, `BEFORECTY`, `AFTERCTY`, `BEFORECNY` and `AFTERCNY`.

Chapter 6
Serial issues and serial items

A serial publication is a journal or a book series. It is divided into “issues”: journal issues or book series volumes.

This chapter describes the interaction between Elsevier and the supplier for the production of serial publications. This includes descriptions of the dataset requests (orders) for CAP journals and the dataset deliveries for each stage for all CAP and PreCAP journals and book series.

6.1. Item orders

This section describes the S100, S200 and S250 orders and the corresponding resupply orders.

order/item-info

Element `item-info` contains all the information on an item. It contains ID information in subelements `version-no` (optional), `jid`, `issn`, `aid`, `pii` and `doi` (optional).

order/item-info/version-no

The element `version-no` contains the version number of the item, as described in Section 3.2. The content of this element must be used in the dataset.xml file in CONTRAST deliveries as the content of the element `dataset/dataset-content/journal-item/version/version-number`.

order/item-info/language

The element `language` contains the language of the item, written out in full.

`<language>English</language>`

order/item-info/crossmark

The element `crossmark` indicates if the item is "CrossMarked" (value `yes`) or not (value `no`). The content of this element must be used in the dataset.xml file in CONTRAST deliveries for the content of the attribute `dataset/dataset-content/journal-item/@cross-mark`. In case of `yes` the attribute must contain `true`, in case of `no` it must contain `false`.

order/item-info/jid

The content of this element must be used in DTD 5.x files as the content of the element `article/item-info/jid` and in the dataset.xml file in CONTRAST deliveries as the content of the element `dataset/dataset-content/journal-item/journal-item-unique-ids/jid-aid/jid`.

order/item-info/issn

The mandatory element `issn` contains the ISSN of the journal.

order/item-info/aid

The content of this element must be used in DTD 5.x files as the content of the element `article/item-info/aid` and in the dataset.xml file in CONTRAST deliveries as the content of the element `dataset/dataset-content/journal-item/journal-item-unique-ids/jid-aid/aid`.

order/item-info/pii

The content of this element must be used in DTD 5.x files as the content of the element `article/item-info/ce:pii`, and in the dataset.xml file in CONTRAST deliveries

as the content of the element dataset/dataset-content/journal-item/journal-item-unique-ids/pii.

order/item-info/doi

The content of this element must be used in DTD 5.x files as the content of the element article/item-info/ce:doi, and in the dataset.xml file in CONTRAST deliveries as the content of the element dataset/dataset-content/journal-item/journal-item-unique-ids/doi.

order/item-info/assigned-to-issue

An item in stage S250 is assigned to a specific issue. The issue PII, the issue cover date and the volume-issue information is present in subelements pii, effect-cover-date, vol-from, vol-to, iss-from, iss-to and supp. Elements pii, issn, effect-cover-date and vol-from are mandatory, the other ones are optional.

This information is to be copied over to the S250 signal and S250 dataset.

order/item-info/s250-sequence-number

Once the number of PDF pages are known for an S250 item it receives an issue sequence number which is captured with s250-sequence-number. This information is to be copied over to the S250 dataset.

order/item-info/prefix

This optional element contains the prefix for the page numbers in page-range.

order/item-info/suffix

This optional element contains the suffix for the page numbers in page-range.

order/item-info/page-range

Once the number of PDF pages are known for an S250 item it receives a pagination. page-range captures the pagination in first-page and last-page (optional). This information is to be copied over to the S250 dataset.

```
<assigned-to-issue>
  <pii>S0020-0190(10)X0013-8</pii>
  <effect-cover-date>20100930</effect-cover-date>
  <vol-from>110</vol-from>
  <iss-from>20</iss-from>
  <s250-sequence-number>10</s250-sequence-number>
  <page-range>
    <first-page>893</first-page>
    <last-page>897</last-page>
  <page-range>
</assigned-to-issue>
```

order/item-info/additional-issue-info

This optional element contains additonal info on the issue in mandatory subelements `paper-type-interior`, `paper-type-cover`, `trimmed-size`, `head-margin`, `back-margin` and `typeset-model`.

```
<additional-issue-info>
  <paper-type-interior>TEP50</paper-type-interior>
  <paper-type-cover>HEG250</paper-type-cover>
  <trimmed-size>210x280mm</trimmed-size>
  <head-margin>13</head-margin>
  <back-margin>18</back-margin>
  <typeset-model>EU7</typeset-model>
</additional-issue-info>
```

order/item-info/pdf-pages

This optional element contains the number of pages in the PDF file of the item.

order/item-info/no-offprints-paid

This optional element contains the number of paid offprints.

order/item-info/covers

This optional element indicates whether the item offprints are printed with or without covers.

order/item-info/offprint-type

This element contains the type of offprint, e.g. `Normal`.

order/item-info/embargo

The optional element `embargo` contains a date/time before which the item is not allowed to be published. The content of this element must be used in the dataset.xml file in CONTRAST deliveries as the content of the element `dataset/dataset-content/journal-item/journal-item-properties/embargo`.

order/item-info/embargo-until-stage

The optional element `embargo-until-stage` contains a stage before which the item is not allowed to be published. The possible values are `S100`, `S200` and `S300`. The content of this element must be used in the dataset.xml file in CONTRAST deliveries as the content of the element `dataset/dataset-content/journal-item/journal-item-properties/embargo-until-stage`.

order/item-info/batch

This optional element will be used when a number of items are treated as one. For instance, an item with add-on items (e.g. commentaries) or a number of abstracts. In the latter case the batch item only has a role in Production, it is a placeholder.

Element `batch` consists of one or more elements `batch-member` each of which consists of elements `aid`, `pit`, `pii` and `doi` (optional). The content of these latter four elements must be used similar to the way described elsewhere in this section. In case of a batch

placeholder the PIT will be "BPH". For practical purposes all batch-member items will receive the same PIT.

See also the section on subitems for more information (Section 3.4, p. 29).

order/item-info/funding

This optional element contains information on the funding of the research reported on in the article. It is contained in the elements `funded-by`, `grant-number`, `nihms-id` and `pubmedcentral-id`. Only element `funded-by` is mandatory, while `grant-number` is repeatable.

`nihms-id` contains the National Institute of Health ID for the item, while `pubmedcentral-id` contains the PubMed Central ID. These IDs are sent to the supplier and in case of a redelivery can help identify the item at e.g. PubMed Central.

order/item-info/refers-to-document

This element contains the PII (subelement `pii`), the DOI (subelement `doi`), the journal ID (subelement `jid`), the article ID (subelement `aid`), and the PIT (subelement `pit`) of the document which is referred to from the item. Except for `pii` the subelements are optional. Element `refers-to-document` is optional and can appear any number of times.

order/item-info/refers-to-document/pii

The content of this element must be used in DTD 5.x files as the content of the element `ce:document-thread/ce:refers-to-document/ce:pii`.

order/item-info/refers-to-document/doi

The content of this element must be used in DTD 5.x files as the content of the element `ce:document-thread/ce:refers-to-document/ce:doi`.

order/item-info/item-group

Items may belong to an "item group", e.g. a special issue. This optional element contains the item group code.

```
<item-group>IG002099</item-group>
<item-group-description>HPCE '04 S.I.</item-group-description>
```

order/item-info/item-group-description

This optional element contains a description of the item-group.

order/item-info/item-title

This element contains the item's title. Note that this is the PTS title which may differ from the real title (accented characters for instance are not used in orders). The title from the item's source files should be leading.

order/item-info/section

Some journals divide their content into sections. The journal *Nuclear Physics A* for instance has a section "Hadronic Physics". This optional element contains the section the item belongs to. Sections are described by codes, which are defined per journal.

```
<section>HP</section>
```

order/item-info/dochead

This optional element contains the document heading or article type of the article.

```
<dochead>Short Communication</dochead>
```

order/item-info/eo-item-nr

This optional element contains the editorial office item identification number.

```
<eo-item-nr>TCS_RSolis-Oba_1938</eo-item-nr>
<e-submission-item-nr>TCS_solis.AT.csd.uwo.
  ca_20020716/1</e-submission-item-nr>
<editor>Dr. G. Ausiello</editor>
```

order/item-info/e-submission-item-nr

This optional element contains the electronic submission item identification number.

order/item-info/editor

This optional element contains the name of the editor who submitted the item.

order/item-info/received-date

This optional element contains the date of receipt of the item in subelement `date`. That subelement must be used in DTD 5.x files to populate the element `ce:date-received`.

order/item-info/revised-date

This optional element contains the date of revision of the item in subelement `date`. That subelement must be used in DTD 5.x files to populate the element `ce:date-revised`.

order/item-info/accept-date

This optional element contains the date of acceptance of the item in subelement `date`. That subelement must be used in DTD 5.x files to populate the element `ce:date-accepted`.

order/item-info/cpc

This optional element indicates whether a journal is rolled out to the Centralized Page Composition (CPC) workflow. It can have three values, `Supplier` (CPC tactical), `Y` (CPC strategic) and `N` (not rolled out to CPC).

order/item-info/in-scope-cpc

This optional element is an empty element indicating for items in a journal which is rolled out to CPC if it is in scope for CPC or not. Mandatory attribute `type` has values `Y` and `N`.

order/item-info/expiry-date

This element contains in subelement `date` the date of expiry or withdrawal of an item.

order/item-info/prd-type-as-sent

This element contains the production type of the item as sent by Elsevier to the supplier. It can contain one of the following production type codes:

Code	Description
CRC	Camera ready copy
FLC	LaTeX e-submission. Mns sent via FTP/EWII. Copy edit
FLF	LaTeX e-submission. Mns sent via FTP/EWII. Fast track
FLP	LaTeX e-submission. Mns sent via FTP/EWII.
FLS	LaTeX e-submission. Mns sent via FTP/EWII. Selective S200
FLX	LaTeX e-submission. Mns sent via FTP/EWII. Special handling
FS5	E-submission. Mns sent via FTP/EWII. S5 delivery req'd
FTC	E-submission. Mns sent via FTP/EWII. Copy edit
FTF	E-submission. Mns sent via FTP/EWII. Fast track
FTP	E-submission. Mns sent via FTP/EWII.
FTQ	E-submission. Mns sent via FTP/EWII. Quick S100
FTS	E-submission. Mns sent via FTP/EWII. Selective S200
FTX	E-submission. Mns sent via FTP/EWII. Special handling
GEN	Automatically generated
NNN	Default Value for No Corrections
PRS	Press-set
PXF	PXE4 Fast
TYC	Scanned or hard copy. Mns sent by post. Copy edit
TYF	Scanned or hard copy. Mns sent by post. Fast track
TYP	Scanned or hard copy. Mns sent by post.
TYS	Scanned or hard copy. Mns sent by post. Selective S200
TYX	Scanned or hard copy. Mns sent by post. Spec'l handling
ZZZ	Issue-item - Default value
COM	Source files available. Mns sent by post.
TEX	LaTeX files available. Mns sent by post.
FCP	[CPC only] Super-fast-track

Note: Any datasets delivered in response to a PTS order should use the "supplier production type" FTP. This is captured in element `supp-prod-type` in a Contrast-in dataset, see the S100 example on p. 75.

order/item-info/online-publ-date

This optional element contains the online publication date of the item in subelement `date`.

order/item-info/online-version

This empty element describes the online version of an item. It has an attribute `type` with four possible values: `e-only`, `e-extra`, `e-appended` and `print`. The latter is the default.

order/item-info/pit

This element contains the production item type, the PIT, of the item. Currently there are 31 PITs defined (see [3]). Apart from these PITs there is the PIT ZZZ which should not result in a delivery.

The content of this element must be used in DTD 5.x files as the value of the top element's attribute `docsubtype` and in the dataset.xml file in CONTRAST deliveries as the content of the element `dataset/dataset-content/journal-item/journal-item-properties/pit`. Note that PITs are always written uppercase in dataset.xml files and lowercase in DTD 5.x files.

order/item-info/copy-edit-content

This optional element is an empty element. It has a mandatory attribute `required` with values `yes` and `no`, signifying if the item needs copy-editing or not, respectively.

order/item-info/no-mns-pages

This element contains the number of manuscript pages.

order/item-info/no-phys-figs

This element contains the actual number of figures. A figure with two subfigures on two pages counts as one physical figure.

```
<no-phys-figs>5</no-phys-figs>
<no-bw-figs>1</no-bw-figs>
<no-web-colour-figs>0</no-web-colour-figs>
<no-colour-figs>2</no-colour-figs>
<colour-fig-nr-print>2</colour-fig-nr-print>
<colour-fig-nr-print>5</colour-fig-nr-print>
<no-e-components>0</no-e-components>
```

order/item-info/no-bw-figs

This optional element contains the number of figures that should appear in black-and-white both in print and on the web.

order/item-info/no-web-colour-figs

This optional element contains the number of figures that should appear in colour on the web but in black-and-white in print.

order/item-info/no-colour-figs

This element contains the number of colour figures.

order/item-info/colour-fig-nr-print

This optional element is present for every colour figure and contains the number of that figure.

order/item-info/no-e-components

This element contains the number of electronic components, e-components.

order/item-info/physical-figures

This optional element contains information on the physical figures. It consists of one or more elements `figure`. Element `figure` consists of subelements `figure-nr`, `figure-type`, `figure-production-type` (optional), `graphical-abstract` (optional), `file-name` (optional) and `figure-remarks` (optional). Element `file-name` may appear any number of times.

```
<physical-figures>
  <figure>
    <figure-nr>2</figure-nr>
    <figure-type>BW</figure-type>
    <file-name>fig2a.jpg</file-name>
    <file-name>fig2b.jpg</file-name>
    <figure-remarks>Fig. 2 on 2 pp.</figure-remarks>
  </figure>
</physical-figures>
```

order/item-info/physical-figures/figure/figure-type

The element `figure-type` can contain one of the following four codes:

`BW`	Black and White
`COLOUR`	Image will appear in colour online and B&W in print
`COLOUR-IN-PRINT`	Image will be colour online and colour in print
`E-ONLY-COLOUR`	Image will appear online only (will not get printed)

order/item-info/physical-figures/figure/figure-production-type

Figure production types are not yet implemented and will not appear in the orders.

order/item-info/physical-figures/figure/graphical-abstract

The element `graphical-abstract` is an empty element. It has a mandatory attribute `grabs` with possible values `yes` and `no`, indicating if the figure is a graphical abstract or not.

order/item-info/physical-figures/figure/file-name

In case a figure is spread over more than one file, more than `file-name` can appear, as is shown in the above example. There can also be more than one `file-name` if different versions of the figure exist:

```
<file-name>fig1bw.jpg</file-name>
<file-name>fig1c.jpg</file-name>
```

order/item-info/e-components

This element contains information on the electronic components. It consists of one or more elements `e-component`. Each `e-component` consists of `e-component-nr`, `file-name` (optional, it can appear any number of times), `e-component-format` and `e-component-remarks` (optional).

```
<e-components>
  <e-component>
    <e-component-nr>1</e-component-nr>
    <e-component-format>APPLICATION</e-component-format>
    <e-component-remarks>Table 2</e-component-remarks>
  </e-component>
  <e-component>
    <e-component-nr>2</e-component-nr>
    <e-component-format>APPLICATION</e-component-format>
    <e-component-remarks>Table 3</e-component-remarks>
  </e-component>
</e-components>
```

order/item-info/e-components/e-component/e-component-format

The element `e-component-format` can contain one of the following four codes: `EAPP` (Application), `EAUD` (Audio), `EIMG` (Image), `EVID` (Video).

order/item-info/righthand-start

This element contains "yes" if articles start on righthand pages and "no" if articles can start on lefthand pages.

order/item-info/copyright-status

This element contains the item's copyright status. It can contain one of the following statuses: 001, . . . , 009. The status together with the article's PIT and the journal base data determines the basic copyright line (see [11]). In case the status is 009, the copyright line is extended based on the employer and the license information.

order/item-info/employer

In case the author's Employing Institution retains the copyright this element contains the name of the institution.

order/item-info/copyright-recd-date

This optional element contains the copyright transfer form receipt date in subelement `date`.

order/item-info/license

In case the item is published under a Creative Commons license this element contains the license code (e.g. "CC BY-NC-ND").

order/item-info/first-author

This optional element contains information on the first author in subelements `degree` (optional), `fnm` (optional), `orcid` (optional) and `snm`. Note that the information from the item's source files should be leading.

order/item-info/corr-author

This optional element contains information on the corresponding author in subelements `degree` (optional), `fnm` (optional), `snm`, `orcid` (optional) and `aff`.

order/item-info/item-remarks

This optional element contains remarks on the item in one or more subelements `item-remark`.

order/item-info/item-remarks/item-remark

This element contains a remark on the item. The type of the remark is in subelement `remark-type`. It can contain one of the following texts:

EW Resupply	
Late corrections	
Revised ms/artwork	
Typesetter/Elsevier error	
Rerun request	
Agreed workflow/exception	
problem	S300
offprints	financial realization
general	advert details
figures	extra copies
setter	reprinted issue
not complete	printer
copyright	electronic annex
author	binder
S100	late corrections
S200	grace copies
compile issue	
covers and prelims	

The remark itself is in subelement `remark` and a response to the remark is in (optional) subelement `response`.

```
<item-remark>
  <remark-type>problem</remark-type>
  <remark>3-JUN-2004 (F37) Please note that we have received 16
    figures(Fig.S1-S16) and one table(table S1) as supporting
    information but the order file mentions 0 e-component/s.
    Please check and update the order file.
  </remark>
  <response>FDALHUIJ 07-JUN-2004 08:54
    coversheet has been updated
  </response>
</item-remark>
```

order/item-info/corrections

This optional empty element contains information on corrections. It has an attribute `type` which can have the following values:

Value	Description
FTP	sent to the supplier by FTP
POST	sent to the supplier by post
FTPANDPOST	sent to the supplier by FTP and post
PROOFSYS	
FTPANDPROOFSYS	corrections uploaded to proofing system, additional corrections sent to supplier by FTP
MASTERFTP	master copy sent to the supplier by FTP
MASTERPOST	master copy sent to the supplier by post
MASTERFTPANDPOST	master copy sent to the supplier by FTP and post
MASTERREMARKS	corrections sent via Remarks field in the order
MASTERNONE	master copy sent to supplier, no corrections received
MASTERPROOFSYS	corrections uploaded to proofing system
REMARKS	
NONE	

order/item-info/corrections-uri

This optional element contains a link, a URI, to the corrections.

order/item-info/revised-proof

This optional element indicates whether the supplier needs to send a revised proof for approval prior to delivering the final version of the item.

order/item-info/offprint-payment

This optional element indicates if payment for offprints is received. It has an attribute `payment` with values `yes` and `no`.

6.2. S5 deliveries

An S5 serial item is an author's original document packed as a CAP deliverable.

The S5 dataset directory structure is defined in Section 4.1.8. In this section we describe what is expected in the dataset and in the dataset.xml.

Below, we traverse the S5 schema for serial publications starting from the element `dataset-content`. The elements `dataset-unique-ids` and `dataset-properties` are explained in Section 4.2. The element `dataset-content` contains a list of `journal-item` subelements. Items in the dataset may belong to different journals or book series, but they must be restricted to one content type.

journal-item

The element `journal-item` contains all the information pertaining to an item in a serial publication. It can have an attribute named `type`.

The `type` attribute takes the values `stand-alone` (default), `with-add-ons` and `batch-placeholder`. In Section 3.4 (p. 29) an explanation is given when `type` is used.

journal-item/version

journal-item/version/version-number

The element `version-number` contains the version number of the item, as described in Section 3.2. The version number is assigned by Elsevier. For S5, it will not always be included in the order.

journal-item/version/stage

The value of `stage` is `S5`.

journal-item/journal-item-unique-ids

journal-item/journal-item-unique-ids/pii

The element `pii` contains the (formatted) PII of the item. In the case of book series this is a book PII.

journal-item/journal-item-unique-ids/doi

The element `doi` contains the DOI of the item, if any. In the case of book series the DOI may be based on a book PII. Only items that will appear online may have a DOI.

journal-item/journal-item-unique-ids/jid-aid

The element `jid-aid` contains up to four subelements: `jid` is the Elsevier system code and `issn` the (formatted) ISSN of the serial publication to which the item belongs; `aid` is the Elsevier system article ID and `article-number` is the item's article number.

The first three elements are mandatory while the last one is optional. The `jid`, `aid` and `article-number` must be equal to the same elements in the item's XML file(s). Book series JIDs should begin with "BS:".

journal-item/journal-item-properties

journal-item/journal-item-properties/pit

The element `pit` contains the publication item type (PIT) of the item. The value of `pit` must be identical to the value of the top-level attribute `docsubtype` in the XML file, except for the case: PITs are always written uppercase in dataset.xml files and lowercase in DTD 5.x files.

journal-item/journal-item-properties/production-type

The element `production-type` can take the values `NON-CRC` and `CRC`.

journal-item/journal-item-properties/embargo

The optional element `embargo` contains a date/time before which the item is not allowed to be published. Its content is in W3C schema DateTime format, `xs:dateTime`, and should be accurate to the minute. To avoid confusion the date/time should be in UTC, signified by a "Z". That is, the embargo date/time should be in the following format: "*yyyy*-*mm*-*dd*T*hh*:*mm*:00Z".

```
<embargo>2006-04-11T12:00:00Z</embargo>
```

journal-item/journal-item-properties/embargo-until-stage

The optional element `embargo-until-stage` contains a stage before which the item is not allowed to be published. That is, if the value is `S300` the item may not be published if it has stage S5, S100 or S200. The possible values are `S5`, `S100`, `S200`, `S300` and `S350`.

journal-item/journal-item-properties/batch

If a journal item is part of a batch, then element `batch` must be present. It contains `pii`, the PII of the journal item that represents the batch, and `doi`, its optional DOI. Note that the item that is referred to, itself an independent item, has the attribute `type` set to a non-default value (see Section 3.4, p. 29).

journal-item/files-info

The element `files-info` contains all the information needed to process the files belonging to the item.

journal-item/files-info/ml

One or two XML files can be associated with the item. These are listed under `ml`.

journal-item/files-info/ml/pathname

`pathname` is the pathname, relative to dataset.xml (that is, relative to the base directory), of the XML file.

journal-item/files-info/ml/purpose

purpose indicates what the XML file is for. There must always be one XML file with purpose MAIN. The optional second XML file is a changes-with-respect-to file with purpose CHANGES.

journal-item/files-info/ml/dtd-version

The version of the DTD and the top-level element (doctype) used to capture the item is contained in dtd-version. This must, of course, be identical to the declaration in the XML file.

journal-item/files-info/ml/weight

The weight of the XML file can be FULL-TEXT, HEAD-AND-TAIL, HEAD-ONLY or CONTENTS-ENTRY-ONLY. The weight of the item indicates which parts of the text are captured in XML. For S5 it is unlikely, but not impossible, that it will be more than CONTENTS-ENTRY-ONLY.

journal-item/files-info/ml/asset

All assets belonging to the item are listed under asset. This includes all external files declared in the XML file; it does not include any strip-ins or fingerprints.

journal-item/files-info/ml/asset/pathname

pathname is the pathname, relative to dataset.xml, of the asset file.

journal-item/files-info/ml/asset/type

The type of the asset can have the values APPLICATION, IMAGE-CAP, AUDIO, VIDEO, VIDEO-FLASH, XML. These types are defined in Section 3.3.

journal-item/files-info/web-pdf

An S5 item must have at exactly one web PDF file associated with it.

journal-item/files-info/web-pdf/pathname

pathname is the pathname, relative to dataset.xml, of the web PDF file.

journal-item/files-info/web-pdf/purpose

purpose has the value MAIN.

journal-item/files-info/web-pdf/pdf-version

pdf-version is the version of the web-PDF file, the allowed values are 1.7 6.4, 1.7 6.5, and 1.7 7.0.

journal-item/files-info/web-pdf/pdf-property

The element `pdf-property` describes the nature of the web PDF file and has one of the following values: `WRAPPED OPTIMIZED`, `DISTILLED OPTIMIZED`, `DISTILLED OPTIMIZED BOOKMARKED`.

The value `WRAPPED OPTIMIZED` is only allowed if the item's production type is `CRC`.

6.3. S5 project deliveries

The S5-Project schema supports deliveries of material for special projects. It is based on the regular S5 schema. Some pattern validation has been removed. This validation will be performed elsewhere and is not described below. In fact, below is assumed that this validation does take place. Additionally some lists of values have been expanded.

An S5 serial item is an author's original document packed as a CAP deliverable. The S5 Project schema is mainly used internally to "tombstone" items that have not made it past the S5 stage.

The S5 dataset directory structure is defined in Section 4.1.8. In this section we describe what is expected in the dataset and in the dataset.xml.

Below, we traverse the S5 schema for serial publications starting from the element `dataset-content`. The elements `dataset-unique-ids` and `dataset-properties` are explained in Section 4.2. The element `dataset-content` contains a list of `journal-item` subelements. Items in the dataset may belong to different journals or book series, but they must be restricted to one content type.

journal-item

The element `journal-item` contains all the information pertaining to an item in a serial publication. It can have an attribute named `type`.

The `type` attribute takes the values `stand-alone` (default), `with-add-ons` and `batch-placeholder`. In Section 3.4 (p. 29) an explanation is given when `type` is used.

journal-item/version

journal-item/version/version-number

The element `version-number` contains the version number of the item, as described in Section 3.2. The version number is assigned by Elsevier. For S5, it will not always be included in the order.

journal-item/version/stage

The value of `stage` is `S5`.

journal-item/journal-item-unique-ids

journal-item/journal-item-unique-ids/pii

The element `pii` contains the (formatted) PII of the item. In the case of book series this is a book PII.

journal-item/journal-item-unique-ids/doi

The element `doi` contains the DOI of the item, if any. In the case of book series the DOI may be based on a book PII. Only items that will appear online may have a DOI.

journal-item/journal-item-unique-ids/jid-aid

The element `jid-aid` contains up to four subelements: `jid` is the Elsevier system code and `issn` the (formatted) ISSN of the serial publication to which the item belongs; `aid` is the Elsevier system article ID and `article-number` is the item's article number.

The first three elements are mandatory while the last one is optional. The `jid`, `aid` and `article-number` must be equal to the same elements in the item's XML file(s). Book series JIDs should begin with "BS:".

journal-item/journal-item-properties

journal-item/journal-item-properties/pit

The element `pit` contains the publication item type (PIT) of the item. The value of `pit` must be identical to the value of the top-level attribute `docsubtype` in the XML file, except for the case: PITs are always written uppercase in dataset.xml files and lowercase in DTD 5.x files.

journal-item/journal-item-properties/production-type

The element `production-type` can take the values `NON-CRC` and `CRC`.

journal-item/journal-item-properties/embargo

The optional element `embargo` contains a date/time before which the item is not allowed to be published. Its content is in W3C schema DateTime format, `xs:dateTime`, and should be accurate to the minute. To avoid confusion the date/time should be in UTC, signified by a "Z". That is, the embargo date/time should be in the following format: "*yyyy-mm-dd*T*hh*:*mm*:00Z".

```
<embargo>2006-04-11T12:00:00Z</embargo>
```

journal-item/journal-item-properties/embargo-until-stage

The optional element `embargo-until-stage` contains a stage before which the item is not allowed to be published. That is, if the value is `S300` the item may not be published if it has stage S5, S100 or S200. The possible values are `S5`, `S100`, `S200`, `S300` and `S350`.

journal-item/journal-item-properties/batch

If a journal item is part of a batch, then element `batch` must be present. It contains `pii`, the PII of the journal item that represents the batch, and `doi`, its optional DOI. Note that the item that is referred to, itself an independent item, has the attribute `type` set to a non-default value (see Section 3.4, p. 29).

journal-item/files-info

The element `files-info` contains all the information needed to process the files belonging to the item.

journal-item/files-info/ml

One or two XML files can be associated with the item. These are listed under `ml`.

journal-item/files-info/ml/pathname

`pathname` is the pathname, relative to dataset.xml (that is, relative to the base directory), of the XML file.

journal-item/files-info/ml/purpose

`purpose` indicates what the XML file is for. There must always be one XML file with purpose `MAIN`. The optional second XML file is a changes-with-respect-to file with purpose `CHANGES`.

journal-item/files-info/ml/dtd-version

The version of the DTD and the top-level element (doctype) used to capture the item is contained in `dtd-version`. This must, of course, be identical to the declaration in the XML file.

journal-item/files-info/ml/weight

The weight of the XML file can be `FULL-TEXT`, `HEAD-AND-TAIL`, `HEAD-ONLY` or `CONTENTS-ENTRY-ONLY`. The weight of the item indicates which parts of the text are captured in XML. For S5 it is unlikely, but not impossible, that it will be more than `CONTENTS-ENTRY-ONLY`.

journal-item/files-info/ml/asset

All assets belonging to the item are listed under `asset`. This includes all external files declared in the XML file; it does not include any strip-ins or fingerprints.

journal-item/files-info/ml/asset/pathname

`pathname` is the pathname, relative to dataset.xml, of the asset file.

journal-item/files-info/ml/asset/type

The `type` of the asset can have the values `APPLICATION`, `IMAGE-CAP`, `IMAGE-NONCAP`, `AUDIO`, `VIDEO`, `VIDEO-FLASH`, `XML`. These types are defined in Section 3.3.

journal-item/files-info/web-pdf

An S5 item must have at exactly one web PDF file associated with it.

journal-item/files-info/web-pdf/pathname

`pathname` is the pathname, relative to dataset.xml, of the web PDF file.

journal-item/files-info/web-pdf/purpose

`purpose` has the value `MAIN`.

journal-item/files-info/web-pdf/pdf-version

`pdf-version` is the version of the web-PDF file, the allowed values are `1.7 6.4`, `1.7 6.5`, and `1.7 7.0`.

journal-item/files-info/web-pdf/pdf-property

The element `pdf-property` describes the nature of the web PDF file and has one of the following values: `WRAPPED OPTIMIZED`, `DISTILLED OPTIMIZED`, `DISTILLED OPTIMIZED BOOKMARKED`.

The value `WRAPPED OPTIMIZED` is only allowed if the item's production type is `CRC`.

6.4. S100 deliveries

An S100 serial item is an uncorrected proof. The S100 deliverable is created from the author's input material, but it is not yet corrected and approved by the author and/or the editors.

The S100 dataset directory structure is defined in Section 4.1.8. In this section we describe what is expected in the dataset and in the dataset.xml.

Below, we traverse the S100 schema for serial publications starting from the element `dataset-content`. The elements `dataset-unique-ids` and `dataset-properties` are explained in Section 4.2. The element `dataset-content` contains a list of `journal-item` subelements. Items in the dataset may belong to different journals or book series, but they must be restricted to one content type.

journal-item

The element `journal-item` contains all the information pertaining to an item in a serial publication. It can have two attributes, `type` and `cross-mark`.

The first one, `type`, takes the values `stand-alone` (default), `with-add-ons` and `batch-placeholder`. In Section 3.4 (p. 29) an explanation is given when `type` is used.

The second one, `cross-mark`, takes the values `false` and `true`.

journal-item/version

journal-item/version/version-number

The element `version-number` contains the version number of the item, as described in Section 3.2. The version number is assigned by Elsevier and is included in the order.

journal-item/version/stage

The value of `stage` is `S100`.

journal-item/journal-item-unique-ids

journal-item/journal-item-unique-ids/pii

The element `pii` contains the (formatted) PII of the item. In the case of book series this is a book PII. It must be equal to the PII in the XML file and the PII in the document properties of the main PDF file (if appropriate).

journal-item/journal-item-unique-ids/doi

The element `doi` contains the DOI of the item, if any. In the case of book series the DOI may be based on a book PII. Only items that will appear online may have a DOI. It must be equal to the DOI in the XML file and the DOI in the document properties of the main PDF file (if appropriate).

journal-item/journal-item-unique-ids/jid-aid

```
<?xml version="1.0" encoding="UTF-8" ?>
<dataset
  xmlns="http://www.elsevier.com/xml/schema/transport/journal-2016.2/s100"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.elsevier.com/xml/schema/transport/journal-2016.2/s100
    http://www.elsevier.com/xml/schema/transport/journal-2016.2/s100/s100.xsd"
  schema-version="2016.2">
  <dataset-unique-ids>
    <supplier-code>ALD</supplier-code>
    <supplier-dataset-id>ALD68477</supplier-dataset-id>
    <timestamp>2016-04-07T09:30:47+02:00</timestamp>
  </dataset-unique-ids>
  <dataset-properties>
    <dataset-action>LOAD</dataset-action>
    <production-process>CAP</production-process>
  </dataset-properties>
  <dataset-content>
    <journal-item>
      <version>
        <version-number>S100.1</version-number>
        <stage>S100</stage>
      </version>
      <journal-item-unique-ids>
        <pii>S0040-4020(03)01057-3</pii>
        <doi>10.1016/S0040-4020(03)01057-3</doi>
        <jid-aid>
          <jid>TET</jid><issn>0040-4020</issn><aid>11699</aid>
        </jid-aid>
      </journal-item-unique-ids>
      <journal-item-properties>
        <pit>FLA</pit>
        <production-type>NON-CRC</production-type>
        <production-details>
          <pdf-pages>11</pdf-pages>
          <page-fraction-body>11</page-fraction-body>
          <page-fraction-trail>1</page-fraction-trail>
          <supp-prod-type>FTP</supp-prod-type>
          <proof-uri>/supplier/S0040-4020(03)01057-3/eproof.pdf</proof-uri>
        </production-details>
      </journal-item-properties>
      <files-info>
        <ml>
          <pathname>S0040402003010573/main.xml</pathname>
          <purpose>MAIN</purpose>
          <dtd-version>JA 5.0.2 ARTICLE</dtd-version>
          <weight>FULL-TEXT</weight>
          <asset>
            <pathname>S0040402003010573/main.assets/fx1.tif</pathname><type>IMAGE-CAP</type>
          </asset>
          <asset>
            <pathname>S0040402003010573/main.assets/fx2.tif</pathname><type>IMAGE-CAP</type>
          </asset>
          <asset>
            <pathname>S0040402003010573/main.assets/mmc1.doc</pathname><type>APPLICATION</type>
          </asset>
        </ml>
        <web-pdf>
          <pathname>S0040402003010573/tx1.pdf</pathname>
          <purpose>MAIN</purpose>
          <pdf-version>1.4 6.0</pdf-version>
          <pdf-property>DISTILLED OPTIMIZED BOOKMARKED</pdf-property>
         </web-pdf>
        <web-pdf>
          <pathname>S0040402003010573/query.pdf</pathname>
          <purpose>AUTHOR-QUERY</purpose>
          <pdf-version>1.4</pdf-version>
          <pdf-property>DISTILLED OPTIMIZED</pdf-property>
        </web-pdf>
      </files-info>
    </journal-item>
  </dataset-content>
</dataset>
```

Figure 3: Sample S100 dataset.xml.

The element `jid-aid` contains up to four subelements: `jid` is the Elsevier system code and `issn` the (formatted) ISSN of the serial publication to which the item belongs; `aid` is the Elsevier system article ID and `article-number` is the item's article number.

The first three elements are mandatory while the last one is optional. The `jid`, `aid` and `article-number` must be equal to the same elements in the item's XML file(s). Book series JIDs should begin with "BS:".

journal-item/journal-item-properties

journal-item/journal-item-properties/pit

The element `pit` contains the publication item type (PIT) of the item. The value of `pit` must be identical to the value of the top-level attribute `docsubtype` in the XML file, except for the case: PITs are always written uppercase in dataset.xml files and lowercase in DTD 5.x files.

journal-item/journal-item-properties/production-type

The element `production-type` can take the values `NON-CRC` and `CRC`.

journal-item/journal-item-properties/embargo

The optional element `embargo` contains a date/time before which the item is not allowed to be published. Its content is in W3C schema DateTime format, `xs:dateTime`, and should be accurate to the minute. To avoid confusion the date/time should be in UTC, signified by a "Z". That is, the embargo date/time should be in the following format: "*yyyy-mm-dd*T*hh*:*mm*:00Z".

```
<embargo>2006-04-11T12:00:00Z</embargo>
```

journal-item/journal-item-properties/embargo-until-stage

The optional element `embargo-until-stage` contains a stage before which the item is not allowed to be published. That is, if the value is `S300` the item may not be published if it has stage S5, S100 or S200. The possible values are `S5`, `S100`, `S200`, `S300` and `S350`.

journal-item/journal-item-properties/batch

If a journal item is part of a batch, then element `batch` must be present. It contains `pii`, the PII of the journal item that represents the batch, and `doi`, its optional DOI. Note that the item that is referred to, itself an independent item, has the attribute `type` set to a non-default value (see Section 3.4, p. 29).

journal-item/journal-item-properties/production-details

The optional element `production-details` is used to pass back information to Elsevier systems; it is not to be used for subitems. It contains the following system-related fields: `pdf-pages` (the number of pages in the PDF file), `pdf-pages-web` (the optional number of pages in the PDF file for the online product, if different), `page-fraction-body` and `page-fraction-trail` (the so-called stock pages of the item), and `supp-prod-type` (`FTP`, see the description of the Order element `prd-type-as-sent` on p. 55). It also contains the URI of the item's proof in element `proof-uri`.

journal-item/files-info

The element `files-info` contains all the information needed to process the files belonging to the item.

journal-item/files-info/ml

One or two XML files can be associated with the item. These are listed under `ml`.

journal-item/files-info/ml/pathname

`pathname` is the pathname, relative to dataset.xml (that is, relative to the base directory), of the XML file.

journal-item/files-info/ml/purpose

`purpose` indicates what the XML file is for. There must always be one XML file with purpose `MAIN`. The optional second XML file is a changes-with-respect-to file with purpose `CHANGES`. A third optional XML file is a file containing pagebreak information, the purpose is `PAGEBREAK`.

journal-item/files-info/ml/dtd-version

The version of the DTD and the top-level element (doctype) used to capture the item is contained in `dtd-version`. This must, of course, be identical to the declaration in the XML file.

journal-item/files-info/ml/weight

The weight of the XML file can be `FULL-TEXT`, `HEAD-AND-TAIL`, `HEAD-ONLY` or `CONTENTS-ENTRY-ONLY`. The weight of the item indicates which parts of the text are captured in XML.

journal-item/files-info/ml/asset

All assets belonging to the item are listed under `asset`. This includes all external files declared in the XML file; it does not include any strip-ins or fingerprints.

journal-item/files-info/ml/asset/pathname

`pathname` is the pathname, relative to dataset.xml, of the asset file.

journal-item/files-info/ml/asset/type

The `type` of the asset can have the values `APPLICATION`, `IMAGE-CAP`, `AUDIO`, `VIDEO`, `VIDEO-FLASH`, `XML`. These types are defined in Section 3.3.

journal-item/files-info/web-pdf

An S100 item must have at least one, and can have up to five, web PDF files associated with it.

journal-item/files-info/web-pdf/pathname

pathname is the pathname, relative to dataset.xml, of the web PDF file.

journal-item/files-info/web-pdf/purpose

purpose indicates what the web PDF file is for. There may not be any PDF files with the same purpose.

There must always be one PDF file with purpose MAIN. This contains the PDF file of the item.

The other optional web PDF files have values EDITED-PROOF (contains an editable version of the proof), MAIN-ABRIDGED (contains the abridged print version), GRAPHICAL-ABSTRACT (contains a separate graphical abstract), STEREO-CHEMISTRY-ABSTRACT (contains a separate set of stereochemistry abstracts), AUTHOR-QUERY (contains author queries).

journal-item/files-info/web-pdf/pdf-version

pdf-version is the version of the web-PDF file, the allowed values are 1.7 6.4, 1.7 6.5, and 1.7 7.0.

journal-item/files-info/web-pdf/pdf-property

The element pdf-property describes the nature of the web PDF file and has one of the following values: WRAPPED OPTIMIZED, DISTILLED OPTIMIZED, DISTILLED OPTIMIZED BOOKMARKED.

The value WRAPPED OPTIMIZED is only allowed if the item's production type is CRC.

If the purpose of the PDF file is not MAIN, then the value DISTILLED OPTIMIZED must be used.

6.5. S100 Project deliveries

The S100-Project schema supports deliveries of material for special projects. It is based on the regular S100 schema. Some pattern validation has been removed. This validation will be performed elsewhere and is not described below. In fact, below is assumed that this validation does take place. Additionally some lists of values have been expanded.

An S100 serial item is an uncorrected proof. The S100 deliverable is created from the author's input material, but it is not yet corrected and approved by the author and/or the editors.

The S100 dataset directory structure is defined in Section 4.1.8. In this section we describe what is expected in the dataset and in the dataset.xml.

Below, we traverse the S100 schema for serial publications starting from the element `dataset-content`. The elements `dataset-unique-ids` and `dataset-properties` are explained in Section 4.2. The element `dataset-content` contains a list of `journal-item` subelements. Items in the dataset may belong to different journals or book series, but they must be restricted to one content type.

journal-item

The element `journal-item` contains all the information pertaining to an item in a serial publication. It can have two attributes, `type` and `cross-mark`.

The first one, `type`, takes the values `stand-alone` (default), `with-add-ons` and `batch-placeholder`. In Section 3.4 (p. 29) an explanation is given when `type` is used.

The second one, `cross-mark`, takes the values `false` and `true`.

journal-item/version

journal-item/version/version-number

The element `version-number` contains the version number of the item, as described in Section 3.2. The version number is assigned by Elsevier and is included in the order.

journal-item/version/stage

The value of `stage` is `S100`.

journal-item/journal-item-unique-ids

journal-item/journal-item-unique-ids/pii

The element `pii` contains the (formatted) PII of the item. In the case of book series this is a book PII. It must be equal to the PII in the XML file and the PII in the document properties of the main PDF file (if appropriate).

journal-item/journal-item-unique-ids/doi

The element `doi` contains the DOI of the item, if any. In the case of book series the DOI may be based on a book PII. Only items that will appear online may have a DOI. It must

```
<?xml version="1.0" encoding="UTF-8" ?>
<dataset
  xmlns="http://www.elsevier.com/xml/schema/transport/journal-2016.2/s100"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.elsevier.com/xml/schema/transport/journal-2016.2/s100
    http://www.elsevier.com/xml/schema/transport/journal-2016.2/s100/s100.xsd"
  schema-version="2016.2">
  <dataset-unique-ids>
    <supplier-code>ALD</supplier-code>
    <supplier-dataset-id>ALD68477</supplier-dataset-id>
    <timestamp>2016-04-07T09:30:47+02:00</timestamp>
  </dataset-unique-ids>
  <dataset-properties>
    <dataset-action>LOAD</dataset-action>
    <production-process>PROJECT</production-process>
    <project-id>ABC</project-id>
  </dataset-properties>
  <dataset-content>
    <journal-item>
      <version>
        <version-number>1.1</version-number>
        <stage>S100</stage>
      </version>
      <journal-item-unique-ids>
        <pii>S0040-4020(03)01057-3</pii>
        <doi>10.1016/S0040-4020(03)01057-3</doi>
        <jid-aid>
          <jid>TET</jid><issn>0040-4020</issn><aid>11699</aid>
        </jid-aid>
      </journal-item-unique-ids>
      <journal-item-properties>
        <pit>FLA</pit>
        <production-type>NON-CRC</production-type>
        <production-details>
          <pdf-pages>11</pdf-pages>
          <page-fraction-body>11</page-fraction-body>
          <page-fraction-trail>1</page-fraction-trail>
          <supp-prod-type>FTP</supp-prod-type>
          <proof-uri>/supplier/S0040-4020(03)01057-3/eproof.pdf</proof-uri>
        </production-details>
      </journal-item-properties>
      <files-info>
        <ml>
          <pathname>S0040402003010573/main.xml</pathname>
          <purpose>MAIN</purpose>
          <dtd-version>JA 5.0.2 ARTICLE</dtd-version>
          <weight>FULL-TEXT</weight>
          <asset>
            <pathname>S0040402003010573/main.assets/fx1.tif</pathname><type>IMAGE-CAP</type>
          </asset>
          <asset>
            <pathname>S0040402003010573/main.assets/fx2.tif</pathname><type>IMAGE-CAP</type>
          </asset>
          <asset>
            <pathname>S0040402003010573/main.assets/mmc1.doc</pathname><type>APPLICATION</type>
          </asset>
        </ml>
        <web-pdf>
          <pathname>S0040402003010573/tx1.pdf</pathname>
          <purpose>MAIN</purpose>
          <pdf-version>1.4 6.0</pdf-version>
          <pdf-property>DISTILLED OPTIMIZED BOOKMARKED</pdf-property>
        </web-pdf>
        <web-pdf>
          <pathname>S0040402003010573/query.pdf</pathname>
          <purpose>AUTHOR-QUERY</purpose>
          <pdf-version>1.4</pdf-version>
          <pdf-property>DISTILLED OPTIMIZED</pdf-property>
        </web-pdf>
      </files-info>
    </journal-item>
  </dataset-content>
</dataset>
```

Figure 4: Sample S100 dataset.xml.

be equal to the DOI in the XML file and the DOI in the document properties of the main PDF file (if appropriate).

journal-item/journal-item-unique-ids/jid-aid

The element `jid-aid` contains up to four subelements: `jid` is the Elsevier system code and `issn` the (formatted) ISSN of the serial publication to which the item belongs; `aid` is the Elsevier system article ID and `article-number` is the item's article number.

The first three elements are mandatory while the last one is optional. The `jid`, `aid` and `article-number` must be equal to the same elements in the item's XML file(s). Book series JIDs should begin with "BS:".

journal-item/journal-item-properties

journal-item/journal-item-properties/pit

The element `pit` contains the publication item type (PIT) of the item. The value of `pit` must be identical to the value of the top-level attribute `docsubtype` in the XML file, except for the case: PITs are always written uppercase in dataset.xml files and lowercase in DTD 5.x files.

journal-item/journal-item-properties/production-type

The element `production-type` can take the values `NON-CRC` and `CRC`.

journal-item/journal-item-properties/embargo

The optional element `embargo` contains a date/time before which the item is not allowed to be published. Its content is in W3C schema DateTime format, `xs:dateTime`, and should be accurate to the minute. To avoid confusion the date/time should be in UTC, signified by a "Z". That is, the embargo date/time should be in the following format: "*yyyy-mm-dd*T*hh*:*mm*:00Z".

```
<embargo>2006-04-11T12:00:00Z</embargo>
```

journal-item/journal-item-properties/embargo-until-stage

The optional element `embargo-until-stage` contains a stage before which the item is not allowed to be published. That is, if the value is `S300` the item may not be published if it has stage S5, S100 or S200. The possible values are `S5`, `S100`, `S200`, `S300` and `S350`.

journal-item/journal-item-properties/batch

If a journal item is part of a batch, then element `batch` must be present. It contains `pii`, the PII of the journal item that represents the batch, and `doi`, its optional DOI. Note that the item that is referred to, itself an independent item, has the attribute `type` set to a non-default value (see Section 3.4, p. 29).

journal-item/journal-item-properties/production-details

The optional element `production-details` is used to pass back information to Elsevier systems; it is not to be used for subitems. It contains the following system-related fields: `pdf-pages` (the number of pages in the PDF file), `pdf-pages-web` (the optional number

of pages in the PDF file for the online product, if different), page-fraction-body and page-fraction-trail (the so-called stock pages of the item), and supp-prod-type (FTP, see the description of the Order element prd-type-as-sent on p. 55). It also contains the URI of the item's proof in element proof-uri.

journal-item/files-info

The element files-info contains all the information needed to process the files belonging to the item.

journal-item/files-info/ml

One or two XML files can be associated with the item. These are listed under ml.

journal-item/files-info/ml/pathname

pathname is the pathname, relative to dataset.xml (that is, relative to the base directory), of the XML file.

journal-item/files-info/ml/purpose

purpose indicates what the XML file is for. There must always be one XML file with purpose MAIN. The optional second XML file is a changes-with-respect-to file with purpose CHANGES. A third optional XML file is a file containing pagebreak information, the purpose is PAGEBREAK.

journal-item/files-info/ml/dtd-version

The version of the DTD and the top-level element (doctype) used to capture the item is contained in dtd-version. This must, of course, be identical to the declaration in the XML file.

journal-item/files-info/ml/weight

The weight of the XML file can be FULL-TEXT, HEAD-AND-TAIL, HEAD-ONLY or CONTENTS-ENTRY-ONLY. The weight of the item indicates which parts of the text are captured in XML.

journal-item/files-info/ml/asset

All assets belonging to the item are listed under asset. This includes all external files declared in the XML file; it does not include any strip-ins or fingerprints.

journal-item/files-info/ml/asset/pathname

pathname is the pathname, relative to dataset.xml, of the asset file.

journal-item/files-info/ml/asset/type

The type of the asset can have the values APPLICATION, IMAGE-CAP, IMAGE-NONCAP, IMAGE-DOWNSAMPLED, IMAGE-THUMBNAIL, IMAGE-MMC, IMAGE-MMC-DOWNSAMPLED, IMAGE-MMC-THUMBNAIL, AUDIO, VIDEO, VIDEO-FLASH, XML. These types are defined in Section 3.3.

journal-item/files-info/web-pdf

An S100 item must have at least one, and can have up to five, web PDF files associated with it.

journal-item/files-info/web-pdf/pathname

`pathname` is the pathname, relative to dataset.xml, of the web PDF file.

journal-item/files-info/web-pdf/purpose

`purpose` indicates what the web PDF file is for. There may not be any PDF files with the same purpose.

There must always be one PDF file with purpose `MAIN`. This contains the PDF file of the item.

The other optional web PDF files have values `EDITED-PROOF` (contains an editable version of the proof), `MAIN-ABRIDGED` (contains the abridged print version), `GRAPHICAL-ABSTRACT` (contains a separate graphical abstract), `STEREO-CHEMISTRY-ABSTRACT` (contains a separate set of stereochemistry abstracts), `AUTHOR-QUERY` (contains author queries).

journal-item/files-info/web-pdf/pdf-version

`pdf-version` is the version of the web-PDF file, the allowed values are `1.7 6.4`, `1.7 6.5`, and `1.7 7.0`.

journal-item/files-info/web-pdf/pdf-property

The element `pdf-property` describes the nature of the web PDF file and has one of the following values: `WRAPPED OPTIMIZED`, `DISTILLED OPTIMIZED`, `DISTILLED OPTIMIZED BOOKMARKED`.

The value `WRAPPED OPTIMIZED` is only allowed if the item's production type is `CRC`.

If the purpose of the PDF file is not `MAIN`, then the value `DISTILLED OPTIMIZED` must be used.

6.6. S100-Proof deliveries

The S100-Proof schema is used to transport S100 datasets from suppliers to ProofCentral. The schema is equal to the S100 schema except that some elements were made optional as ProofCentral doesn't use that information. The most important change is that the PDF files are no longer mandatory.

The following elements were made optional:

- `files-info/web-pdf`
- `journal-item-properties/production-details/pdf-pages`
- `journal-item-properties/production-details/page-fraction-body`
- `journal-item-properties/production-details/supp-prod-type`

The S100-Proof schema cannot be used to deliver material to the Electronic Warehouse.

6.7. S200 deliveries

An S200 item in a serial publication is a corrected proof, authorized by the authors of the article and/or the editors.

The S200 dataset directory structure is defined in Section 4.1.8.

Below, we traverse the S200 schema for serial publications starting from the element `dataset-content`. The elements `dataset-unique-ids` and `dataset-properties` are explained in Section 4.2. The element `dataset-content` contains a list of `journal-item` subelements. The items may belong to different journals or book series, but they must belong to the same content type.

journal-item

The element `journal-item` contains all the information pertaining to an item in a serial publication. It can have two attributes, `type` and `cross-mark`.

The first one, `type`, takes the values `stand-alone` (default), `with-add-ons` and `batch-placeholder`. In Section 3.4 (p. 29) an explanation is given when `type` is used.

The second one, `cross-mark`, takes the values `false` and `true`.

journal-item/version

journal-item/version/version-number

The element `version-number` contains the version number of the item, as described in Section 3.2. The version number is assigned by Elsevier and is included in the order.

journal-item/version/stage

The value of `stage` is `S200`.

journal-item/journal-item-unique-ids

journal-item/journal-item-unique-ids/pii

The element `pii` contains the (formatted) PII of the item. In the case of book series this is a book PII. It must be equal to the PII in the XML file and the PII in the document properties of the main PDF file (if appropriate).

journal-item/journal-item-unique-ids/doi

The element `doi` contains the DOI of the item, if any. In the case of book series the DOI may be based on a book PII. Only items that will appear online may have a DOI. It must be equal to the DOI in the XML file and the DOI in the document properties of the main PDF file (if appropriate).

journal-item/journal-item-unique-ids/jid-aid

The element `jid-aid` contains up to four subelements: `jid` is the Elsevier system code and `issn` the (formatted) ISSN of the serial publication to which the item belongs; `aid` is the Elsevier system article ID and `article-number` is the item's article number.

The first three elements are mandatory while the last one is optional. The `jid`, `aid` and `article-number` must be equal to the same elements in the item's XML file(s). Book series JIDs should begin with "BS:".

journal-item/journal-item-properties

journal-item/journal-item-properties/pit

The element `pit` contains the publication item type of the item. The value of `pit` must be identical to the value of the top-level attribute `docsubtype` in the XML file, except for the case: PITs are always written uppercase in dataset.xml files and lowercase in DTD 5.x files.

journal-item/journal-item-properties/production-type

The element `production-type` can take the values `NON-CRC` and `CRC`.

journal-item/journal-item-properties/embargo

The optional element `embargo` contains a date/time before which the item is not allowed to be published. Its content is in W3C schema DateTime format, `xs:dateTime`, and should be accurate to the minute. To avoid confusion the date/time should be in UTC, signified by a "Z". That is, the embargo date/time should be in the following format: "*yyyy-mm-dd*T*hh*:*mm*:00Z".

```
<embargo>2006-04-11T12:00:00Z</embargo>
```

journal-item/journal-item-properties/embargo-until-stage

The optional element `embargo-until-stage` contains a stage before which the item is not allowed to be published. That is, if the value is `S300` the item may not be published if it has stage S5, S100 or S200. The possible values are `S5`, `S100`, `S200`, `S300` and `S350`.

journal-item/journal-item-properties/batch

If a journal item is part of a batch, then element `batch` must be present. It contains `pii`, the PII of the journal item that represents the batch, and `doi`, its optional DOI. Note that the item that is referred to, itself an independent item, has the attribute `type` set to a non-default value (see Section 3.4, p. 29).

journal-item/journal-item-properties/production-details

The optional element `production-details` is used to pass back information to Elsevier systems; it is not to be used for subitems. It contains the following PTSIII-related fields: `pdf-pages` (the number of pages in the PDF file), `pdf-pages-web` (the number of pages in the PDF file for the online version, if different), `page-fraction-body` and `page-fraction-trail` (the so-called stock pages of the item), and `supp-prod-type` (the production type that the supplier used to create the item: `CRC`, `COM`, `COX`, `FTP`, `FTX`, `GEN`, `PRS`, `TEX`, `TYP`, `TYX`). It also contains the URI of the item's proof in element `proof-uri`.

journal-item/files-info

The element `files-info` contains all the information needed to process the files belonging to the item.

journal-item/files-info/ml

One or two XML files can be associated with the item. These are listed under `ml`.

journal-item/files-info/ml/pathname

`pathname` is the pathname, relative to dataset.xml (that is, relative to the base directory), of the XML file.

journal-item/files-info/ml/purpose

`purpose` indicates what the XML file is for. There must always be one XML file with purpose `MAIN`. The optional second XML file is a changes-with-respect-to file with purpose `CHANGES`. A third optional XML file is a file containing pagebreak information, the purpose is `PAGEBREAK`.

journal-item/files-info/ml/dtd-version

The version of the DTD and the top-level element (doctype) used to capture the item is contained in `dtd-version`. This must, of course, be identical to the declaration in the XML file.

journal-item/files-info/ml/weight

The weight of the XML file can be `FULL-TEXT`, `HEAD-AND-TAIL`, `HEAD-ONLY` or `CONTENTS-ENTRY-ONLY`. The weight of the item indicates which parts of the text are captured in XML.

journal-item/files-info/ml/asset

All assets belonging to the item are listed under `asset`. This includes all external files declared in the XML file; it does not include any strip-ins or fingerprints.

journal-item/files-info/ml/asset/pathname

`pathname` is the pathname, relative to dataset.xml, of the asset file.

journal-item/files-info/ml/asset/type

The `type` of the asset can have the values `APPLICATION`, `IMAGE-CAP`, `AUDIO`, `VIDEO`, `VIDEO-FLASH`, `XML`. These types are defined in Section 3.3.

journal-item/files-info/web-pdf

An S200 item must have at least one, and can have up to four, web PDF files associated with it.

journal-item/files-info/web-pdf/pathname

pathname is the pathname, relative to dataset.xml, of the web PDF file.

journal-item/files-info/web-pdf/purpose

purpose indicates what the web PDF file is for. There may not be any PDF files with the same purpose.

There must always be one PDF file with purpose MAIN. This contains the PDF file of the item.

The other optional web PDF files have values MAIN-ABRIDGED (contains the abridged print version), GRAPHICAL-ABSTRACT (contains a separate graphical abstract), STEREO-CHEMISTRY-ABSTRACT (contains a separate set of stereochemistry abstracts). Author queries cannot exist at this stage.

journal-item/files-info/web-pdf/pdf-version

pdf-version is the version of the web-PDF file, the allowed values are 1.7 6.4, 1.7 6.5, and 1.7 7.0.

journal-item/files-info/web-pdf/pdf-property

The element pdf-property describes the nature of the web PDF file and has one of the following values: WRAPPED OPTIMIZED, DISTILLED OPTIMIZED, DISTILLED OPTIMIZED BOOKMARKED.

The value WRAPPED OPTIMIZED is only allowed if the item's production type is CRC.

If the purpose of the PDF file is not MAIN, then the value DISTILLED OPTIMIZED must be used.

6.8. S200 Project deliveries

The S200-Project schema supports deliveries of material for special projects. It is based on the regular S200 schema. Some pattern validation has been removed. This validation will be performed elsewhere and is not described below. In fact, below is assumed that this validation does take place. Additionally some lists of values have been expanded.

An S200 item in a serial publication is a corrected proof, authorized by the authors of the article and/or the editors.

The S200 dataset directory structure is defined in Section 4.1.8.

Below, we traverse the S200 schema for serial publications starting from the element `dataset-content`. The elements `dataset-unique-ids` and `dataset-properties` are explained in Section 4.2. The element `dataset-content` contains a list of `journal-item` subelements. The items may belong to different journals or book series, but they must belong to the same content type.

journal-item

The element `journal-item` contains all the information pertaining to an item in a serial publication. It can have two attributes, `type` and `cross-mark`.

The first one, `type`, takes the values `stand-alone` (default), `with-add-ons` and `batch-placeholder`. In Section 3.4 (p. 29) an explanation is given when `type` is used.

The second one, `cross-mark`, takes the values `false` and `true`.

journal-item/version

journal-item/version/version-number

The element `version-number` contains the version number of the item, as described in Section 3.2. The version number is assigned by Elsevier and is included in the order.

journal-item/version/stage

The value of `stage` is `S200`.

journal-item/journal-item-unique-ids

journal-item/journal-item-unique-ids/pii

The element `pii` contains the (formatted) PII of the item. In the case of book series this is a book PII. It must be equal to the PII in the XML file and the PII in the document properties of the main PDF file (if appropriate).

journal-item/journal-item-unique-ids/doi

The element `doi` contains the DOI of the item, if any. In the case of book series the DOI may be based on a book PII. Only items that will appear online may have a DOI. It must be equal to the DOI in the XML file and the DOI in the document properties of the main PDF file (if appropriate).

journal-item/journal-item-unique-ids/jid-aid

The element `jid-aid` contains up to four subelements: `jid` is the Elsevier system code and `issn` the (formatted) ISSN of the serial publication to which the item belongs; `aid` is the Elsevier system article ID and `article-number` is the item's article number.

The first three elements are mandatory while the last one is optional. The `jid`, `aid` and `article-number` must be equal to the same elements in the item's XML file(s). Book series JIDs should begin with "BS:".

journal-item/journal-item-properties

journal-item/journal-item-properties/pit

The element `pit` contains the publication item type of the item. The value of `pit` must be identical to the value of the top-level attribute `docsubtype` in the XML file, except for the case: PITs are always written uppercase in dataset.xml files and lowercase in DTD 5.*x* files.

journal-item/journal-item-properties/production-type

The element `production-type` can take the values `NON-CRC` and `CRC`.

journal-item/journal-item-properties/embargo

The optional element `embargo` contains a date/time before which the item is not allowed to be published. Its content is in W3C schema DateTime format, `xs:dateTime`, and should be accurate to the minute. To avoid confusion the date/time should be in UTC, signified by a "Z". That is, the embargo date/time should be in the following format: "*yyyy-mm-dd*T*hh*:*mm*:00Z".

```
<embargo>2006-04-11T12:00:00Z</embargo>
```

journal-item/journal-item-properties/embargo-until-stage

The optional element `embargo-until-stage` contains a stage before which the item is not allowed to be published. That is, if the value is `S300` the item may not be published if it has stage S5, S100 or S200. The possible values are `S5`, `S100`, `S200`, `S300` and `S350`.

journal-item/journal-item-properties/batch

If a journal item is part of a batch, then element `batch` must be present. It contains `pii`, the PII of the journal item that represents the batch, and `doi`, its optional DOI. Note that the item that is referred to, itself an independent item, has the attribute `type` set to a non-default value (see Section 3.4, p. 29).

journal-item/journal-item-properties/production-details

The optional element `production-details` is used to pass back information to Elsevier systems; it is not to be used for subitems. It contains the following PTSIII-related fields: `pdf-pages` (the number of pages in the PDF file), `pdf-pages-web` (the number of pages

in the PDF file for the online version, if different), page-fraction-body and page-fraction-trail (the so-called stock pages of the item), and supp-prod-type (the production type that the supplier used to create the item: CRC, COM, COX, FTP, FTX, GEN, PRS, TEX, TYP, TYX). It also contains the URI of the item's proof in element proof-uri.

journal-item/files-info

The element files-info contains all the information needed to process the files belonging to the item.

journal-item/files-info/ml

One or two XML files can be associated with the item. These are listed under ml.

journal-item/files-info/ml/pathname

pathname is the pathname, relative to dataset.xml (that is, relative to the base directory), of the XML file.

journal-item/files-info/ml/purpose

purpose indicates what the XML file is for. There must always be one XML file with purpose MAIN. The optional second XML file is a changes-with-respect-to file with purpose CHANGES. A third optional XML file is a file containing pagebreak information, the purpose is PAGEBREAK.

journal-item/files-info/ml/dtd-version

The version of the DTD and the top-level element (doctype) used to capture the item is contained in dtd-version. This must, of course, be identical to the declaration in the XML file.

journal-item/files-info/ml/weight

The weight of the XML file can be FULL-TEXT, HEAD-AND-TAIL, HEAD-ONLY or CONTENTS-ENTRY-ONLY. The weight of the item indicates which parts of the text are captured in XML.

journal-item/files-info/ml/asset

All assets belonging to the item are listed under asset. This includes all external files declared in the XML file; it does not include any strip-ins or fingerprints.

journal-item/files-info/ml/asset/pathname

pathname is the pathname, relative to dataset.xml, of the asset file.

journal-item/files-info/ml/asset/type

The type of the asset can have the values APPLICATION, IMAGE-CAP, IMAGE-NONCAP, IMAGE-DOWNSAMPLED, IMAGE-THUMBNAIL, IMAGE-MMC, IMAGE-MMC-DOWNSAMPLED, IMAGE-MMC-THUMBNAIL, AUDIO, VIDEO, VIDEO-FLASH, XML. These types are defined in Section 3.3.

journal-item/files-info/web-pdf

An S200 item must have at least one, and can have up to four, web PDF files associated with it.

journal-item/files-info/web-pdf/pathname

`pathname` is the pathname, relative to dataset.xml, of the web PDF file.

journal-item/files-info/web-pdf/purpose

`purpose` indicates what the web PDF file is for. There may not be any PDF files with the same purpose.

There must always be one PDF file with purpose `MAIN`. This contains the PDF file of the item.

The other optional web PDF files have values `MAIN-ABRIDGED` (contains the abridged print version), `GRAPHICAL-ABSTRACT` (contains a separate graphical abstract), `STEREO-CHEMISTRY-ABSTRACT` (contains a separate set of stereochemistry abstracts). Author queries cannot exist at this stage.

journal-item/files-info/web-pdf/pdf-version

`pdf-version` is the version of the web-PDF file, the allowed values are `1.7 6.4`, `1.7 6.5`, and `1.7 7.0`.

journal-item/files-info/web-pdf/pdf-property

The element `pdf-property` describes the nature of the web PDF file and has one of the following values: `WRAPPED OPTIMIZED`, `DISTILLED OPTIMIZED`, `DISTILLED OPTIMIZED BOOKMARKED`.

The value `WRAPPED OPTIMIZED` is only allowed if the item's production type is `CRC`.

If the purpose of the PDF file is not `MAIN`, then the value `DISTILLED OPTIMIZED` must be used.

6.9. S250 deliveries

An S250 item in a serial publication is a corrected proof, which has already been placed in an issue. Hence the volume-issue information, the page numbers and the sequence number of the item in the issue are known. This information is to be copied from the orders and placed in the appropriate elements in the schema described below.

The S250 dataset directory structure is defined in Section 4.1.8.

Below, we traverse the S250 schema for serial publications starting from the element `dataset-content`. The elements `dataset-unique-ids` and `dataset-properties` are explained in Section 4.2. The element `dataset-content` contains a list of `journal-item` subelements. The items may belong to different journals or book series, but they must belong to the same content type.

journal-issue

The element `journal-issue` contains information pertaining to the issue the item(s) belong to. There are no files associated with the issue.

journal-issue/journal-issue-unique-ids

journal-issue/journal-issue-unique-ids/pii

The element `pii` contains the PII of the issue. In the case of book series this is a book hub PII or a collection PII.

journal-issue/journal-issue-unique-ids/doi

The element `doi` contains the DOI of the issue, if any. In the case of book series the DOI may be based on a book hub PII or a collection PII. Only issues that will appear online may have a DOI.

journal-issue/journal-issue-properties

journal-issue/journal-issue-properties/jid

The element `jid` contains the Elsevier system code of the serial publication to which the issue belongs. Book series JIDs should begin with “BS:”.

journal-issue/journal-issue-properties/issn

The element `issn` contains the ISSN of the serial publication to which the issue belongs.

journal-issue/journal-issue-properties/volume-issue-number

The volume/issue number of the issue is captured in `volume-issue-number`. It consists of `vol-first`, `vol-last`, `iss-first`, `iss-last` and `suppl`.

journal-item

The element `journal-item` contains all the information pertaining to an item in a serial publication. It can have two attributes, `type` and `cross-mark`.

The first one, `type`, takes the values `stand-alone` (default), `with-add-ons` and `batch-placeholder`. In Section 3.4 (p. 29) an explanation is given when `type` is used.

The second one, `cross-mark`, takes the values `false` and `true`.

journal-item/version

journal-item/version/version-number

The element `version-number` contains the version number of the item, as described in Section 3.2. The version number is assigned by Elsevier and is included in the order.

journal-item/version/stage

The value of `stage` is `S250`.

journal-item/journal-item-unique-ids

journal-item/journal-item-unique-ids/pii

The element `pii` contains the (formatted) PII of the item. In the case of book series this is a book PII. It must be equal to the PII in the XML file and the PII in the document properties of the main PDF file (if appropriate).

journal-item/journal-item-unique-ids/doi

The element `doi` contains the DOI of the item, if any. In the case of book series the DOI may be based on a book PII. Only items that will appear online may have a DOI. It must be equal to the DOI in the XML file and the DOI in the document properties of the main PDF file (if appropriate).

journal-item/journal-item-unique-ids/jid-aid

The element `jid-aid` contains up to four subelements: `jid` is the Elsevier system code and `issn` the (formatted) ISSN of the serial publication to which the item belongs; `aid` is the Elsevier system article ID and `article-number` is the item's article number.

The first three elements are mandatory while the last one is optional. The `jid`, `aid` and `article-number` must be equal to the same elements in the item's XML file(s). Book series JIDs should begin with "BS:".

journal-item/journal-item-properties

journal-item/journal-item-properties/pit

The element `pit` contains the publication item type of the item. The value of `pit` must be identical to the value of the top-level attribute `docsubtype` in the XML file, except for the case: PITs are always written uppercase in dataset.xml files and lowercase in DTD 5.x files.

journal-item/journal-item-properties/production-type

The element `production-type` can take the values `NON-CRC` and `CRC`.

journal-item/journal-item-properties/embargo

The optional element `embargo` contains a date/time before which the item is not allowed to be published. It will not be used for stage S250.

journal-item/journal-item-properties/embargo-until-stage

The optional element `embargo-until-stage` contains a stage before which the item is not allowed to be published. It will not be used for stage S250.

journal-item/journal-item-properties/s250-sequence-number

This element contains the sequence number of the item in the issue.

journal-item/journal-item-properties/page-range

The element `page-range` contains the page range of the item in `first-page` and `last-page`.

journal-item/journal-item-properties/batch

If a journal item is part of a batch, then element `batch` must be present. It contains `pii`, the PII of the journal item that represents the batch, and `doi`, its optional DOI. Note that the item that is referred to, itself an independent item, has the attribute `type` set to a non-default value (see Section 3.4, p. 29).

journal-item/files-info

The element `files-info` contains all the information needed to process the files belonging to the item.

journal-item/files-info/ml

One or two XML files can be associated with the item. These are listed under `ml`.

journal-item/files-info/ml/pathname

`pathname` is the pathname, relative to dataset.xml (that is, relative to the base directory), of the XML file.

journal-item/files-info/ml/purpose

`purpose` indicates what the XML file is for. There must always be one XML file with purpose `MAIN`. The optional second XML file is a changes-with-respect-to file with purpose `CHANGES`. A third optional XML file is a file containing pagebreak information, the purpose is `PAGEBREAK`.

journal-item/files-info/ml/dtd-version

The version of the DTD and the top-level element (doctype) used to capture the item is contained in `dtd-version`. This must, of course, be identical to the declaration in the XML file.

journal-item/files-info/ml/weight

The weight of the XML file can be `FULL-TEXT`, `HEAD-AND-TAIL`, `HEAD-ONLY` or `CONTENTS-ENTRY-ONLY`. The weight of the item indicates which parts of the text are captured in XML.

journal-item/files-info/ml/asset

All assets belonging to the item are listed under `asset`. This includes all external files declared in the XML file; it does not include any strip-ins or fingerprints.

journal-item/files-info/ml/asset/pathname

`pathname` is the pathname, relative to dataset.xml, of the asset file.

journal-item/files-info/ml/asset/type

The `type` of the asset can have the values `APPLICATION`, `IMAGE-CAP`, `AUDIO`, `VIDEO`, `VIDEO-FLASH`, `XML`. These types are defined in Section 3.3.

journal-item/files-info/web-pdf

An S250 item has one or two web PDF files associated with it.

journal-item/files-info/web-pdf/pathname

`pathname` is the pathname, relative to dataset.xml, of the web PDF file.

journal-item/files-info/web-pdf/purpose

`purpose` indicates what the web PDF file is for. There may not be any PDF files with the same purpose.

There must always be one PDF file with purpose `MAIN`. This contains the PDF file of the item.

The other optional web PDF files have values `MAIN-ABRIDGED` (contains the abridged print version), `GRAPHICAL-ABSTRACT` (contains a separate graphical abstract), `STEREO-CHEMISTRY-ABSTRACT` (contains a separate set of stereochemistry abstracts). Author queries cannot exist at this stage.

journal-item/files-info/web-pdf/pdf-version

`pdf-version` is the version of the web-PDF file, the allowed values are `1.7 6.4`, `1.7 6.5`, and `1.7 7.0`.

journal-item/files-info/web-pdf/pdf-property

The element `pdf-property` describes the nature of the web PDF file and has one of the following values: `WRAPPED OPTIMIZED`, `DISTILLED OPTIMIZED`, `DISTILLED OPTIMIZED BOOKMARKED`.

The value `WRAPPED OPTIMIZED` is only allowed if the item's production type is `CRC`.

If the purpose of the PDF file is not `MAIN`, then the value `DISTILLED OPTIMIZED` must be used.

6.10. S280 deliveries

An S280 deliverable is introduced to support the delivery of virtual item collections to serve any defined classification.

The S280 dataset directory structure is defined in Section 4.1.8.

Below, we traverse the S280 schema for serial publications starting from the element `dataset-content`. The elements `dataset-unique-ids` and `dataset-properties` are explained in Section 4.2. The element `dataset-content` contains a list of `journal-item` subelements. The items may belong to different journals or book series, but they must belong to the same content type.

journal-item

The element `journal-item` contains all the information pertaining to an item in a serial publication. It can have two attributes, `type` and `cross-mark`.

The first one, `type`, takes the values `stand-alone` (default), `with-add-ons` and `batch-placeholder`. In Section 3.4 (p. 29) an explanation is given when `type` is used.

The second one, `cross-mark`, takes the values `false` and `true`.

journal-item/version

journal-item/version/version-number

The element `version-number` contains the version number of the item, as described in Section 3.2. The version number is assigned by Elsevier and is included in the order.

journal-item/version/stage

The value of `stage` is `S280`.

journal-item/journal-item-unique-ids

journal-item/journal-item-unique-ids/pii

The element `pii` contains the (formatted) PII of the item. In the case of book series this is a book PII. It must be equal to the PII in the XML file and the PII in the document properties of the main PDF file (if appropriate).

journal-item/journal-item-unique-ids/doi

The element `doi` contains the DOI of the item, if any. In the case of book series the DOI may be based on a book PII. Only items that will appear online may have a DOI. It must be equal to the DOI in the XML file and the DOI in the document properties of the main PDF file (if appropriate).

journal-item/journal-item-unique-ids/jid-aid

The element `jid-aid` contains up to four subelements: `jid` is the Elsevier system code and `issn` the (formatted) ISSN of the serial publication to which the item belongs; `aid` is the Elsevier system article ID and `article-number` is the item's article number.

The first three elements are mandatory while the last one is optional. The `jid`, `aid` and `article-number` must be equal to the same elements in the item's XML file(s). Book series JIDs should begin with "BS:".

journal-item/journal-item-properties

journal-item/journal-item-properties/pit

The element `pit` contains the publication item type of the item. The value of `pit` must be identical to the value of the top-level attribute `docsubtype` in the XML file, except for the case: PITs are always written uppercase in dataset.xml files and lowercase in DTD 5.x files.

journal-item/journal-item-properties/production-type

The element `production-type` can take the values `NON-CRC` and `CRC`.

journal-item/files-info

The element `files-info` contains all the information needed to process the files belonging to the item.

journal-item/files-info/ml

One or two XML files can be associated with the item. These are listed under `ml`.

journal-item/files-info/ml/pathname

`pathname` is the pathname, relative to dataset.xml (that is, relative to the base directory), of the XML file.

journal-item/files-info/ml/purpose

`purpose` indicates what the XML file is for. There must always be one XML file with purpose `MAIN`. The optional second XML file is a changes-with-respect-to file with purpose `CHANGES`. A third optional XML file is a file containing pagebreak information, the purpose is `PAGEBREAK`.

journal-item/files-info/ml/dtd-version

The version of the DTD and the top-level element (doctype) used to capture the item is contained in `dtd-version`. This must, of course, be identical to the declaration in the XML file.

journal-item/files-info/ml/weight

The weight of the XML file can be `FULL-TEXT`, `HEAD-AND-TAIL`, `HEAD-ONLY` or `CONTENTS-ENTRY-ONLY`. The weight of the item indicates which parts of the text are captured in XML.

journal-item/files-info/ml/asset

All assets belonging to the item are listed under `asset`. This includes all external files declared in the XML file; it does not include any strip-ins or fingerprints.

journal-item/files-info/ml/asset/pathname

`pathname` is the pathname, relative to dataset.xml, of the asset file.

journal-item/files-info/ml/asset/type

The `type` of the asset can have the values `APPLICATION`, `IMAGE-CAP`, `AUDIO`, `VIDEO`, `VIDEO-FLASH`, `XML`. These types are defined in Section 3.3.

journal-item/files-info/web-pdf

An S280 item has one or two web PDF files associated with it.

journal-item/files-info/web-pdf/pathname

`pathname` is the pathname, relative to dataset.xml, of the web PDF file.

journal-item/files-info/web-pdf/purpose

`purpose` indicates what the web PDF file is for. There may not be any PDF files with the same purpose.

There must always be one PDF file with purpose `MAIN`. This contains the PDF file of the item.

The other optional web PDF files have values `MAIN-ABRIDGED` (contains the abridged print version), `GRAPHICAL-ABSTRACT` (contains a separate graphical abstract), `STEREO-CHEMISTRY-ABSTRACT` (contains a separate set of stereochemistry abstracts). Author queries cannot exist at this stage.

journal-item/files-info/web-pdf/pdf-version

`pdf-version` is the version of the web-PDF file, the allowed values are `1.7 6.4`, `1.7 6.5`, and `1.7 7.0`.

journal-item/files-info/web-pdf/pdf-property

The element `pdf-property` describes the nature of the web PDF file and has one of the following values: `WRAPPED OPTIMIZED`, `DISTILLED OPTIMIZED`, `DISTILLED OPTIMIZED BOOKMARKED`.

The value `WRAPPED OPTIMIZED` is only allowed if the item's production type is `CRC`.

If the purpose of the PDF file is not `MAIN`, then the value `DISTILLED OPTIMIZED` must be used.

6.11. S280 Project deliveries

The S280-Project schema supports deliveries of material for special projects. It is based on the regular S280 schema. Some pattern validation has been removed. This validation will be performed elsewhere and is not described below. In fact, below is assumed that this validation does take place. Additionally some lists of values have been expanded.

An S280 deliverable is introduced to support the delivery of virtual item collections to serve any defined classification.

The S280 dataset directory structure is defined in Section 4.1.8.

Below, we traverse the S280 schema for serial publications starting from the element `dataset-content`. The elements `dataset-unique-ids` and `dataset-properties` are explained in Section 4.2. The element `dataset-content` contains a list of `journal-item` subelements. The items may belong to different journals or book series, but they must belong to the same content type.

journal-item

The element `journal-item` contains all the information pertaining to an item in a serial publication. It can have two attributes, `type` and `cross-mark`.

The first one, `type`, takes the values `stand-alone` (default), `with-add-ons` and `batch-placeholder`. In Section 3.4 (p. 29) an explanation is given when `type` is used.

The second one, `cross-mark`, takes the values `false` and `true`.

journal-item/version

journal-item/version/version-number

The element `version-number` contains the version number of the item, as described in Section 3.2. The version number is assigned by Elsevier and is included in the order.

journal-item/version/stage

The value of `stage` is `S280`.

journal-item/journal-item-unique-ids

journal-item/journal-item-unique-ids/pii

The element `pii` contains the (formatted) PII of the item. In the case of book series this is a book PII. It must be equal to the PII in the XML file and the PII in the document properties of the main PDF file (if appropriate).

journal-item/journal-item-unique-ids/doi

The element `doi` contains the DOI of the item, if any. In the case of book series the DOI may be based on a book PII. Only items that will appear online may have a DOI. It must be equal to the DOI in the XML file and the DOI in the document properties of the main PDF file (if appropriate).

journal-item/journal-item-unique-ids/jid-aid

The element `jid-aid` contains up to four subelements: `jid` is the Elsevier system code and `issn` the (formatted) ISSN of the serial publication to which the item belongs; `aid` is the Elsevier system article ID and `article-number` is the item's article number.

The first three elements are mandatory while the last one is optional. The `jid`, `aid` and `article-number` must be equal to the same elements in the item's XML file(s). Book series JIDs should begin with "BS:".

journal-item/journal-item-properties

journal-item/journal-item-properties/pit

The element `pit` contains the publication item type of the item. The value of `pit` must be identical to the value of the top-level attribute `docsubtype` in the XML file, except for the case: PITs are always written uppercase in dataset.xml files and lowercase in DTD 5.x files.

journal-item/journal-item-properties/production-type

The element `production-type` can take the values `NON-CRC` and `CRC`.

journal-item/files-info

The element `files-info` contains all the information needed to process the files belonging to the item.

journal-item/files-info/ml

One or two XML files can be associated with the item. These are listed under `ml`.

journal-item/files-info/ml/pathname

`pathname` is the pathname, relative to dataset.xml (that is, relative to the base directory), of the XML file.

journal-item/files-info/ml/purpose

`purpose` indicates what the XML file is for. There must always be one XML file with purpose `MAIN`. The optional second XML file is a changes-with-respect-to file with purpose `CHANGES`. A third optional XML file is a file containing pagebreak information, the purpose is `PAGEBREAK`.

journal-item/files-info/ml/dtd-version

The version of the DTD and the top-level element (doctype) used to capture the item is contained in `dtd-version`. This must, of course, be identical to the declaration in the XML file.

journal-item/files-info/ml/weight

The weight of the XML file can be FULL-TEXT, HEAD-AND-TAIL, HEAD-ONLY or CONTENTS-ENTRY-ONLY. The weight of the item indicates which parts of the text are captured in XML.

journal-item/files-info/ml/asset

All assets belonging to the item are listed under asset. This includes all external files declared in the XML file; it does not include any strip-ins or fingerprints.

journal-item/files-info/ml/asset/pathname

pathname is the pathname, relative to dataset.xml, of the asset file.

journal-item/files-info/ml/asset/type

The type of the asset can have the values APPLICATION, IMAGE-CAP, AUDIO, VIDEO, VIDEO-FLASH, XML. These types are defined in Section 3.3.

journal-item/files-info/web-pdf

An S280 item has one or two web PDF files associated with it.

journal-item/files-info/web-pdf/pathname

pathname is the pathname, relative to dataset.xml, of the web PDF file.

journal-item/files-info/web-pdf/purpose

purpose indicates what the web PDF file is for. There may not be any PDF files with the same purpose.

There must always be one PDF file with purpose MAIN. This contains the PDF file of the item.

The other optional web PDF files have values MAIN-ABRIDGED (contains the abridged print version), GRAPHICAL-ABSTRACT (contains a separate graphical abstract), STEREO-CHEMISTRY-ABSTRACT (contains a separate set of stereochemistry abstracts). Author queries cannot exist at this stage.

journal-item/files-info/web-pdf/pdf-version

pdf-version is the version of the web-PDF file, the allowed values are 1.7 6.4, 1.7 6.5, and 1.7 7.0.

journal-item/files-info/web-pdf/pdf-property

The element pdf-property describes the nature of the web PDF file and has one of the following values: WRAPPED OPTIMIZED, DISTILLED OPTIMIZED, DISTILLED OPTIMIZED BOOKMARKED.

The value WRAPPED OPTIMIZED is only allowed if the item's production type is CRC.

If the purpose of the PDF file is not MAIN, then the value DISTILLED OPTIMIZED must be used.

6.12. Issue orders

This section describes the S300 order.

order/issue-info

Element `issue-info` contains all the critical metadata of an issue. It consists of subelements `general-info`, `issue-content` and `issue-remarks` (optional).

order/issue-info/general-info

This element contains general issue information. It consists of elements `version-no` (optional), `pii` (optional) and `doi` (optional), followed by elements `article-based-publishing` (optional), `embargo` (optional), `issue-production-type`, `buffer-status`, `hold-until-date` (optional), `zero-warehousing` (optional), `jid`, `journal-no`, `issn`, `isbn` (optional), `journal-title` and `pmg` (optional) containing information about the journal the issue belongs to.

These are followed by elements `vol-from`, `vol-to`, `iss-from`, `iss-to` and `supp` containg volume-issue information (except `vol-from` these elements are optional).

These are in turn followed by elements containing information necessary for printing: `paper-type-interior`, `paper-type-cover`, `cover-finishing`, `cover-print-type`, `print-type`, `binding-type`, `offprint-type`, `trimmed-size`, `head-margin`, `back-margin`, `typeset-model`, `righthand-start`, `issue-weight`, `spine-width` (the first three of these are mandatory, the other ones are optional).

These are followed by optional elements `effect-cover-date`, `cover-date`, `cover-date-printed`, `cover-label`, `cover-copyright` and `cover-caption` containing cover-date information and optional element `special-issue` containing any special issue data. See below for more information on all these elements.

Element `general-info` ends with elements containing information on the different kinds and amounts of pages: `no-pages-prelims` (optional), `no-pages-interior`, `no-pages-extra` (optional), `no-pages-insert` (optional), `no-pages-bm` (optional), `no-pages-print`, `no-pages-web` (optional), `no-pages-total`, `no-pages-blank`, `no-pages-adverts` and `page-ranges`.

`general-info` concludes with mandatory element `paid-ads` and optional elements `corrections` and `corrections-uri`.

order/issue-info/general-info/version-no

The element `version-no` contains the version number of the issue, as described in Section 3.2. The content of this element can be used in the dataset.xml file in CONTRAST deliveries as the content of the element `dataset/dataset-content/journal-issue/version/version-number`.

order/issue-info/general-info/pii

This element contains an issue PII. The content must be used as the content of element `issue-info/ce:pii` in the issue hub file.

order/issue-info/general-info/doi

Issue DOIs are not yet implemented and will not appear in the orders.

order/issue-info/general-info/article-based-publishing

This optional empty element has an attribute `abp` with values `yes` and `no`, signifying if the article is ABP or not, respectively.

order/issue-info/general-info/embargo

The optional element `embargo` contains a date/time before which the issue is not allowed to be published. The content of this element must be used in the dataset.xml file in CONTRAST deliveries as the content of the element `dataset/dataset-content/journal-issue/embargo`.

order/issue-info/general-info/issue-production-type

This element contains the production type of the issue as sent by Elsevier to the supplier, e.g. `CRC` or `ELE`.

order/issue-info/general-info/buffer-status

This empty element has a mandatory attribute `status` with values `yes` and `no`. An issue with `status yes` should not be despatched to the warehouse until an instruction from the relevant production site is received, or the `status` is changed to `no`. Note that this element is not rendered by the PTS stylesheet. See also the description of this element in the print/bind order (p. 155).

order/issue-info/general-info/hold-until-date

An issue may only be sent to the warehouse after this date (in subelement `date`). Note that this element is not rendered by the PTS stylesheet. See also the description of this element in the print/bind order (p. 155).

order/issue-info/general-info/zero-warehousing

This optional empty element has an attribute `status` with values `yes` and `no`, signifying if the journal is participating in the Zero-Warehousing Project or not, respectively.

order/issue-info/general-info/jid

The element `jid` contains the JID of the journal. The content of this element can be used in the dataset.xml file in CONTRAST deliveries as the content of the element `dataset/dataset-content/journal-issue/journal-issue-properties/jid`.

order/issue-info/general-info/journal-no

The element `journal-no` contains the PTS ID of this journal.

order/issue-info/general-info/issn

The element issn contains the ISSN of the journal. The content of this element can be used in the dataset.xml file in CONTRAST deliveries as the content of the element dataset/dataset-content/journal-issue/journal-issue-properties/issn.

order/issue-info/general-info/isbn

The element isbn contains the ISBN of the journal volume the article is part of. The content of this element can be used in the dataset.xml file in CONTRAST deliveries as the content of the element dataset/dataset-content/journal-issue/journal-issue-properties/isbn.

order/issue-info/general-info/journal-title

The element journal-title contains the title of this journal.

order/issue-info/general-info/pmg

The element pmg contains the Product Market Group, an Elsevier journal classification code, the journal belongs to. The PMG code is a three-digit number.

order/issue-info/general-info/vol-from

This element together with the optional elements vol-to, iss-from, iss-to and supp uniquely define an issue of a particular journal.

The content of these elements must be used in the dataset.xml file in CONTRAST deliveries as the content of elements vol-first, vol-last, iss-first, iss-last and suppl, respectively. (Note the differences in element names.)

They are omitted when they are superfluous. Some examples:

- Vol. 10, No. 2 (single volume, single issue):
  `<vol-from>10</vol-from><iss-from>2</iss-from>`
- Vol. 67, Nos 3–7 (single volume, multiple issues):
  `<vol-from>67</vol-from><iss-from>3</iss-from><iss-to>7</iss-to>`
- Vol.67C (single and complete volume):
  `<vol-from>67</vol-from><supp>C</supp>`
- Vols. 34–36C (multiple complete volumes):
  `<vol-from>34</vol-from><vol-to>36</vol-to><supp>C</supp>`

To obtain the “supplement” to one of the above possibilities, the element supp is added with appropriate content. In case of a supplement to a complete volume or to multiple volumes, the completeness is implied. The element supp can have the following content (here n is one of A,...,Z,1,2,3,..., i.e. a capital letter or an integer):

- C: indicating completeness of the volume or multiple volumes
- I: indicating a master index published as a supplement
- S: indicating a supplement
- In,Sn: indicating the nth index or supplement
- Pn: indicating the nth part of a volume or multiple volume (used when a volume or multiple volume is published physically in two or more bound parts)

See also Section 4.1.9 (p. 36).

order/issue-info/general-info/paper-type-interior

This element contains the paper type of the interior pages.

```
<general-info>
  ...
  <paper-type-interior>AG40</paper-type-interior>
  <paper-type-cover>CEG275</paper-type-cover>
  <cover-finishing>varnished</cover-finishing>
  <cover-print-type>Litho</cover-print-type>
  <print-type>Litho</print-type>
  <binding-type>perfect</binding-type>
  <trimmed-size>8_1/16x10_7/8in</trimmed-size>
  <head-margin>2 cm</head-margin>
  <back-margin>3 cm</back-margin>
  <typeset-model>Non-standard</typeset-model>
  <righthand-start>yes</righthand-start>
  <issue-weight>45.37</issue-weight>
  <spine-width>5</spine-width>
  ...
</general-info>
```

order/issue-info/general-info/paper-type-cover

This element contains the paper type of the cover pages.

order/issue-info/general-info/cover-finishing

This element contains the type of cover finishing.

order/issue-info/general-info/cover-print-type

This element contains the type of cover printing. It can have the following two values: `Digital` and `Litho`.

order/issue-info/general-info/print-type

This element contains the type of printing. It can have the following six values: `E-Only`, `HP`, `Litho`, `Litho Sheet`, `Litho Web` and `Oce`.

order/issue-info/general-info/binding-type

This element contains the type of binding.

order/issue-info/general-info/offprint-type

This element contains the type of offprint. It can have the following two values: `Normal`, `High Quality`.

order/issue-info/general-info/trimmed-size

This element contains the trimmed size of the issue.

order/issue-info/general-info/head-margin

This element contains the head margin of the issue pages.

order/issue-info/general-info/back-margin

This element contains the back margin of the issue pages.

order/issue-info/general-info/typeset-model

This element indicates the typeset model.

order/issue-info/general-info/righthand-start

This element contains "yes" if articles start on righthand pages and "no" if articles can start on lefthand pages.

order/issue-info/general-info/issue-weight

This element contains the weight of the issue (in grams).

order/issue-info/general-info/spine-width

This element contains the width of the spin (in mm).

order/issue-info/general-info/effect-cover-date

This optional element contains the cover-date in "Effect" format. That is, in one of the following six formats: $yyyy$, $yyyymm$, $yyyymm/mm$, $yyyymmdd$, $yyyymmdd/dd$ or $yyyymmdd/mmdd$, where $yyyy$ is a year, mm is a month and dd is a day.

mm can have one of the following values: $01, \ldots, 12, 21, 22, 23, 24, 31, 32, 33, 34$. The values $01, \ldots, 12$ are used for the months, the values $21, 22, 23, 24$ are used for the seasons (21: Spring, 22: Summer, 23: Autumn, 24: Winter), and the values $31, 32, 33, 34$ are used for the quarters (31: 1st Quarter, 32: 2nd Quarter, 33: 3rd Quarter, 34: 4th Quarter).

order/issue-info/general-info/cover-date-printed

This optional element contains the cover-date as it is printed on the issue cover.

order/issue-info/general-info/cover-label

This optional element contains the label of the cover image. The content must be used to populate element `issue-data/cover-image/ce:figure/ce:label` in the issue hub file.

order/issue-info/general-info/cover-copyright

This optional element contains the copyright of the cover image. The content must be used to populate element `issue-data/cover-image/ce:figure/ce:copyright` in the issue hub file.

order/issue-info/general-info/cover-caption

This optional element contains the caption of the cover image. The content must be used to populate element `issue-data/cover-image/ce:figure/ce:caption` in the issue hub file.

order/issue-info/general-info/cover-date

Element `cover-date` contains the coverdate (also contained in `effect-cover-date`) in subelement `date-range`. This subelement contains the coverdate in subelements `start-date` and `end-date` (optional). These latter two elements contain dates in one of the following three "Effect" formats: $yyyy$, $yyyymm$, $yyyymmdd$.

The content must be used to populate element `issue-data/cover-date` in the issue hub file.

order/issue-info/general-info/special-issue

This optional element contains special-issue information in subelements `special-issue-id` (containing the internal working title), `full-name`, `subtitle`, `alt-title`, `alt-subtitle`, `conference` and `editors` or `si-editors`. Only the first is mandatory.

The content of elements `full-name`, `subtitle`, `alt-title` and `alt-subtitle` must be used to populate elements `title-editors-group/ce:title`, `title-editors-group/ce:subtitle`, `title-editors-group/ce:alt-title` and `title-editors-group/ce:alt-subtitle`, respectively, in the issue hub file.

```
<special-issue>
  <special-issue-id>MC</special-issue-id>
  <full-name>Special Issue on the Conference on the
    Numerical Solution of Markov Chains 2003</full-name>
  <conference>
    <conf-name>International Conference on the
      Numerical Solution of Markov Chains</conf-name>
    <abbr-name>MC 2003</abbr-name>
    <venue>Urbana-Champaign, IL, USA</venue>
    <conference-date>
      <date-range>
        <start-date>20030903</start-date>
        <end-date>20030905</end-date>
      </date-range>
    </conference-date>
  </conference>
  <si-editors>
    <si-editor>
      <given-name>W.</given-name><surname>Grassmann</surname>
    </si-editor>
    <si-editor>
      <given-name>C.</given-name><surname>Meyer</surname>
    </si-editor>
    <si-editor>
      <given-name>B.</given-name><surname>Stewart</surname>
    </si-editor>
    <si-editor>
      <given-name>D.</given-name><surname>Szyld</surname>
    </si-editor>
  </si-editors>
</special-issue>
```

order/issue-info/general-info/special-issue/conference

In case of a conference, this optional element contains extra information in optional subelements conf-name, abbr-name, venue, conference-date and sponsor.

The content of these four elements must be used to populate elements conference-info/full-name, conference-info/abbr-name, conference-info/venue, conference-info/date-range and title-editors-group/sponsors, respectively. Note that element sponsor may contain more than one sponsor whereas in the issue hub file sponsors are captured in separate elements sponsors/sponsor.

order/issue-info/general-info/special-issue/editors

This optional element contains the editors of the special issue in an unstructured form. The content must be used to populate element issue-data/title-editors-group/editors in the issue hub file.

order/issue-info/general-info/special-issue/si-editors

This optional element contains the editors of the special issue in a structured form. It contains one or more subelements si-editor. The latter element has subelements degrees, given-name, surname, suffix, degrees, and affiliation, of which only surname is mandatory. The content must be used to populate element issue-data/title-editors-group/ce:editors and its subelements in the issue hub file.

order/issue-info/general-info/no-pages-extra

Element no-pages-extra contains the number of unpaginated pages in the issue. These pages, e.g. advertisements, have prefix "EX" to the page numbers. When an article ends on e.g. p. 20 and is followed by two advertisement pages (pages EX1 and EX2), the subsequent article will start on p. 21.

order/issue-info/general-info/no-pages-print

Element no-pages-print contains the total number of printed pages, i.e. without the e-only and e-extra pages.

order/issue-info/general-info/no-pages-web

Element no-pages-web contains the total number of pages that will appear on the web, i.e. including the e-only and e-extra pages.

order/issue-info/general-info/no-pages-total

Element no-pages-total contains the total number of pages in an issue, excluding the covers. It is the sum of no-pages-prelims, no-pages-interior, no-pages-extra, no-pages-insert and no-pages-bm.

order/issue-info/general-info/no-pages-adverts

Element no-pages-adverts contains the number of pages with paid advertisements.

order/issue-info/general-info/page-ranges

This optional element contains the page-ranges the issue consists of, contained in one or more elements page-range. Each element page-range consists of two elements, first-page and last-page. It has a mandatory attribute type with values PRELIM, INTERIOR, EXTRA, INSERT and BACKMATTER. The element will not be present in case all articles have an article-number.

```
<page-ranges>
  <page-range type="PRELIM">
    <first-page>i</first-page>
    <last-page>x</last-page>
  </page-range>
  <page-range type="INTERIOR">
    <first-page>1</first-page>
    <last-page>150</last-page>
  </page-range>
  <page-range type="BACKMATTER">
    <first-page>I</first-page>
    <last-page>IV</last-page>
  </page-range>
</page-ranges>
```

order/issue-info/general-info/paid-ads

Element paid-ads has an attribute paid-ads with values yes and no indicating if the issue has paid advertisements or not.

order/issue-info/general-info/corrections

This optional empty element contains information on corrections. It has an attribute type which can have the following values:

Value	Description
FTP	sent to the supplier by FTP
POST	sent to the supplier by post
REMARKSSUPELSERROR	corrections sent via Remarks field, corrections due to supplier and EW errors
REMARKSSUPERROR	corrections sent via Remarks field, corrections due to supplier errors
FTPSUPELSERROR	corrections sent via EW, corrections due to supplier and EW errors
FTPSUPERROR	corrections sent via EW, corrections due to supplier errors
REMARKS	
NONE	

order/issue-info/issue-content

This element contains the content of an issue in one or more row elements. Every row corresponds to a row in the PTSIII issue line-up.

order/issue-info/issue-content/row

Element `row` contains the information of the item (or part of the item) that makes up the row. The type of row is indicated by the mandatory attribute `type` which has the following possible values:

- `ce`: a contents entry, i.e. an item that appears in the contents and hence in the dataset
- `non-ce`: a non-contents entry, i.e. an item that does not appear in the contents
- `blank`: a blank page
- `advert`: an advert
- `remark`
- `h1`: a level-1 heading
- `h2`: a level-2 heading
- `h3`: a level-3 heading
- `h4`: a level-4 heading
- `he`: "heading end" to indicate an h1, h2, h3 or h4 section is ended

A prime example where an `he` row must be used is when an issue's content ends with a section followed by an index. The row then indicates that the section is closed and that the index is not part of it. Such an `he` row need not be used if the section is followed by another section (i.e. an h1, h2, h3 or h4 row).

The element `row` consists of elements `version-no`, `crossmark`, `aid`, `pii`, `article-number`, `doi`, `s250-sequence-number`, `cpc`, `embargo`, `embargo-until-stage`, `batch` and `item-title`, followed by `copyright-status`, `employer`, `copyright-recd-date`, `license`, `corr-author`, `eo-item-nr`, `pit` and `prd-type`.

These are followed by elements containing information on (e-)pages, offprints, etc.: `prefix`, `suffix`, `first-e-page`, `last-e-page`, `page-from`, `page-to`, `pdf-pages`, `online-publ-date`, `online-version`, `no-issues-free`, `no-issues-paid`, `no-offprints-tot`, `no-offprints-paid`, `no-offprints-free`, `page-charge`, `covers` and `e-suite`.

The element ends with elements `no-colour-figs`, `colour-fig-nr-print`, `refers-to-document`, `remark` and `offprint-payment`.

All the above-mentioned elements, except `item-title`, `online-version` and `e-suite`, are optional. Elements `colour-fig-nr-print` and `refers-to-document` can occur any number of times.

Split items: Items can be split. That is, an item's pages can be non-contiguous in the issue. For instance, an item can appear on pp. 31–40 and p. 68. In this case the item information appears in two rows. The content of these rows are identical except for the elements `page-from` and `page-to`.

In the dataset such an item appears as one item. The PDF file contains all the pages belonging to that item. In the issue hub it also appears as one item, and with multiple page ranges. The place it appears is the place of the first occurrence of a part of the item.

Items can of course be split in more than two parts. (See Section 3.7, p. 32, for more information.)

order/issue-info/issue-content/row/version-no

The element `version-no` contains the version number of the item, as described in Section 3.2. The content of this element can be used in the dataset.xml file in CONTRAST

deliveries as the content of the element `dataset/dataset-content/journal-item/version/version-number`.

order/issue-info/issue-content/row/crossmark

The element `crossmark` indicates if the item is "CrossMarked" (value `yes`) or not (value `no`). The content of this element must be used in the dataset.xml file in CONTRAST deliveries for the content of the attribute `dataset/dataset-content/journal-item/@cross-mark`. In case of `yes` the attribute must contain `true`, in case of `no` it must contain `false`.

order/issue-info/issue-content/row/pii

The content of this element must be used to populate element `ce:include-item/ce:pii` in the issue hub file.

order/issue-info/issue-content/row/article-number

The content of this element must be used to populate element `ce:include-item/ce:article-number` in the issue hub file. Note that this will only be possible as of JA DTD 5.6.0.

order/issue-info/issue-content/row/doi

The content of this element must be used to populate element `ce:include-item/ce:doi` in the issue hub file.

order/issue-info/issue-content/row/s250-sequence-number

This optional element contains the item's sequence number in the H200 issue. It must be used to create the order of the items in the H200 issue hub file.

order/issue-info/issue-content/row/cpc

This optional element indicates whether a journal is rolled out to the Centralized Page Composition (CPC) workflow. It can have three values, `Supplier` (CPC tactical), `Y` (CPC strategic) and `N` (not rolled out to CPC). It is used in the CPC tactical workflow to support mixed issues.

order/issue-info/issue-content/row/embargo

The optional element `embargo` contains a date/time before which the issue is not allowed to be published. The content of this element must be used in the dataset.xml file in CONTRAST deliveries as the content of the element `dataset/dataset-content/journal-item/journal-item-properties/embargo`.

order/issue-info/issue-content/row/embargo-until-stage

The optional element `embargo-until-stage` contains a stage before which the item is not allowed to be published. The possible values are `S100`, `S200` and `S300`.

order/issue-info/issue-content/row/batch

This optional element will be used when a number of items are treated as one. For instance, an item with add-on items (e.g. commentaries) or a number of abstracts. In the latter case the batch item only has a role in Production, it is a placeholder.

Element `batch` consists of one or more elements `batch-member` each of which consists of elements `aid`, `pit`, `pii` and `doi` (optional). The content of these latter four elements must be used similar to the way described elsewhere in this section. In case of a batch placeholder the PIT will be "BPH". For practical purposes all batch-member items will receive the same PIT.

See also the section on subitems for more information (Section 3.4, p. 29).

order/issue-info/issue-content/row/item-title

This element contains the item's title. Note that this is the PTS title which may differ from the real title (accented characters for instance are not possible). The title from the item's source files should be leading.

order/issue-info/issue-content/row/page-from

The content of `page-from` and `page-to` must be used to populate the elements `ce:include-item/ce:pages/ce:first-page` and `ce:include-item/ce:pages/ce:last-page` in the issue hub file. Note that the first page and the last page in the order may be the same; this is not allowed in the issue hub file.

```
<prefix>CO</prefix>
<page-from>CO1</page-from>
<page-to>CO1</page-to>
...
<prefix>FM</prefix>
<page-from>ii</page-from>
<page-to>ii</page-to>
...
<page-from>1</page-from>
<page-to>6</page-to>
```

order/issue-info/issue-content/row/online-publ-date

This element contain the online publication date in subelement `date`. The content is to be used for the PDF online publication date (see [12]).

order/issue-info/issue-content/row/online-version

This empty element describes the online version of an item. It has an attribute `type` with three possible values: `e-only`, `e-extra`, `e-appended` and `print`. The latter is the default.

order/issue-info/issue-content/row/covers

This optional element indicates whether the item offprints are printed with or without covers.

order/issue-info/issue-content/row/e-suite

This mandatory element indicates if the item follows the e-suite workflow. It has values `yes` and `no`.

order/issue-info/issue-content/row/refers-to-document

This element contains the PII (subelement `pii`), the DOI (subelement `doi`), the journal ID (subelement `jid`), the article ID (subelement `aid`), and the PIT (subelement `pit`) of the document which is referred to from the item. Except for `pii` the subelements are optional. Element `refers-to-document` is optional and can appear any number of times.

order/issue-info/issue-remarks

This optional element contains remarks on the issue in one or more subelements `issue-remark`. For more information see the description of element `order/item-info/item-remarks` (p. 59).

6.13. P100 deliveries

Material that does not follow the complete item and issue workflow goes through the stage P100 before it is complete. This can include non-contents-entries, such as covers and prelim pages that need to be checked prior to (print) publication, and items such as editorial boards and author and subject indexes that are generated after issue compilation. The P100 deliverable is not meant for online publication. It is used in supplier–Elsevier communication.

The P100 dataset directory structure is defined in Section 4.1.9.

Below, we traverse the P100 schema for journals starting from the element `dataset-content`. The elements `dataset-unique-ids` and `dataset-properties` are explained in Section 4.2.

A P100 delivery contains proof material for precisely one issue. It contains the hub file and a number of issue-level PDF files, and possibly some items. The items must belong to the issue. It can also contain spin-offs so that these can be proofed.

journal-issue

The element `journal-issue` contains all the information pertaining to the journal issue in the dataset. The files associated with the issue are the issue hub with possible assets.

journal-issue/version

journal-issue/version/version-number

The element `version-number` contains the version number of the issue, as described in Section 3.2. The version number is assigned by Elsevier and is included in the order.

journal-issue/version/stage

The value of `stage` is `P100`.

journal-issue/journal-issue-unique-ids

journal-issue/journal-issue-unique-ids/pii

The element `pii` contains the PII of the issue. In the case of book series this is a book hub PII or a collection PII. It must be equal to the PII in the XML hub file.

journal-issue/journal-issue-unique-ids/doi

The element `doi` contains the DOI of the issue, if any. In the case of book series the DOI may be based on a book hub PII or a collection PII. Only issues that will appear online may have a DOI. It must be equal to the DOI in the XML hub file.

journal-issue/journal-issue-properties

journal-issue/journal-issue-properties/jid

The element `jid` contains the Elsevier system code of the serial publication to which the issue belongs. Book series JIDs should begin with "BS:".

journal-issue/journal-issue-properties/issn

The element `issn` contains the ISSN of the serial publication to which the issue belongs.

journal-issue/journal-issue-properties/volume-issue-number

The volume/issue number of the issue is captured in `volume-issue-number`. It consists of `vol-first`, `vol-last`, `iss-first`, `iss-last` and `suppl`.

journal-issue/journal-issue-properties/isbn

The issue's optional ISBN, an ISBN-13, is captured with `isbn`.

journal-issue/files-info

The element `files-info` contains all the information needed to process the files belonging to the issue.

journal-issue/files-info/ml

The issue hub is an XML file associated with the issue. Its details are listed under `ml`. Although this element is optional it is only allowed to be absent in case the P100 delivery contains a spin-off issue.

journal-issue/files-info/ml/pathname

`pathname` is the pathname, relative to dataset.xml (that is, relative to the base directory), of the hub XML file.

journal-issue/files-info/ml/purpose

The element `purpose` always contains the value `MAIN`.

journal-issue/files-info/ml/dtd-version

The version of the DTD of the hub file is contained in `dtd-version`. The version must, of course, be identical to the declaration in the XML file.

journal-issue/files-info/ml/asset

All assets belonging to the item are listed under `asset`. This includes all external files declared in the XML file; it does not include any strip-ins or fingerprints.

journal-issue/files-info/ml/asset/pathname

`pathname` is the pathname, relative to dataset.xml, of the asset file.

journal-issue/files-info/ml/asset/type

The `type` of the asset can have the values `IMAGE-CAP`, `IMAGE-COVER`. These types are defined in Section 3.3.

journal-issue/files-info/web-pdf

Beside the hub XML file, there can be any number of issue-level PDF files. These may contain covers, prelim pages and other material that needs to be checked but will not be part of an online delivery. These PDF files do not possess a “purpose” field.

The element is optional to support P100 delivery of e-only issues which do not have any web PDF files.

journal-issue/files-info/web-pdf/pathname

`pathname` is the pathname, relative to dataset.xml, of the web PDF file.

journal-issue/files-info/web-pdf/pdf-version

`pdf-version` is the version of the web-PDF file, the allowed values are `1.7 6.4`, `1.7 6.5`, and `1.7 7.0`.

journal-item

The element `journal-item` contains all the information pertaining to an item in a serial publication. It can have two attributes, `type` and `cross-mark`.

The first one, `type`, takes the values `stand-alone` (default), `with-add-ons` and `batch-placeholder`. In Section 3.4 (p. 29) an explanation is given when `type` is used.

The second one, `cross-mark`, takes the values `false` and `true`.

journal-item/version

journal-item/version/version-number

The element `version-number` contains the version number of the item, as described in Section 3.2. The version number is assigned by Elsevier and is included in the order.

journal-item/version/stage

The value of `stage` is `P100`.

journal-item/journal-item-unique-ids

journal-item/journal-item-unique-ids/pii

The element `pii` contains the PII of the item. In the case of book series this is a book PII. It must be equal to the PII in the XML file and the PII in the document properties of the main PDF file (if appropriate).

journal-item/journal-item-unique-ids/doi

The element `doi` contains the DOI of the item, if any. In the case of book series the DOI may be based on a book PII. Only items that will appear online may have a DOI. It must be equal to the DOI in the XML file and the DOI in the document properties of the main PDF file (if appropriate).

journal-item/journal-item-unique-ids/jid-aid

The element `jid-aid` contains up to four subelements: `jid` is the Elsevier system code and `issn` the (formatted) ISSN of the serial publication to which the item belongs; `aid` is the Elsevier system article ID and `article-number` is the item's article number.

The first three elements are mandatory while the last one is optional. The `jid`, `aid` and `article-number` must be equal to the same elements in the item's XML file(s). Book series JIDs should begin with "BS:".

journal-item/journal-item-properties

journal-item/journal-item-properties/pit

The element `pit` contains the publication item type of the item. The value of `pit` must be identical to the value of the top-level attribute `docsubtype` in the XML file, except for the case: PITs are always written uppercase in dataset.xml files and lowercase in DTD 5.x files.

journal-item/journal-item-properties/production-type

The element `production-type` can take the values `NON-CRC` and `CRC`.

journal-item/journal-item-properties/batch

If a journal item is part of a batch, then element `batch` must be present. It contains `pii`, the PII of the journal item that represents the batch, and `doi`, its optional DOI. Note that the item that is referred to, itself an independent item, has the attribute `type` set to a non-default value (see Section 3.4, p. 29).

journal-item/files-info

The element `files-info` contains all the information needed to process the files belonging to the item.

journal-item/files-info/ml

One or two XML files can be associated with the item. These are listed under `ml`.

journal-item/files-info/ml/pathname

`pathname` is the pathname, relative to dataset.xml, of the XML file.

journal-item/files-info/ml/purpose

`purpose` indicates what the XML file is for. There must always be one XML file with purpose `MAIN`. The optional second XML file is a changes-with-respect-to file with purpose `CHANGES`. A third optional XML file is a file containing pagebreak information, the purpose is `PAGEBREAK`.

journal-item/files-info/ml/dtd-version

The version of the DTD and the top-level element (doctype) used to capture the item is contained in `dtd-version`. This must, of course, be identical to the declaration in the XML file.

journal-item/files-info/ml/weight

The weight of the XML file can be `FULL-TEXT`, `HEAD-AND-TAIL`, `HEAD-ONLY` or `CONTENTS-ENTRY-ONLY`. The weight of the item indicates which parts of the text are captured in XML.

journal-item/files-info/ml/asset

All assets belonging to the item are listed under `asset`. This includes all external files declared in the XML file; it does not include any strip-ins or fingerprints.

journal-item/files-info/ml/asset/pathname

`pathname` is the pathname, relative to dataset.xml, of the asset file.

journal-item/files-info/ml/asset/type

The `type` of the asset can have the values `APPLICATION`, `IMAGE-CAP`, `AUDIO`, `VIDEO`, `VIDEO-FLASH`, `XML`. These types are defined in Section 3.3.

journal-item/files-info/web-pdf

A P100 item must have at least one, and can have up to four, web PDF files associated with it.

journal-item/files-info/web-pdf/pathname

`pathname` is the pathname, relative to dataset.xml, of the web PDF file.

journal-item/files-info/web-pdf/purpose

`purpose` indicates what the web PDF file is for. There may not be any PDF files with the same purpose.

There must always be one PDF file with purpose `MAIN`. This contains the PDF file of the item.

The other optional web PDF files have values `MAIN-ABRIDGED` (contains the abridged print version), `GRAPHICAL-ABSTRACT` (contains a separate graphical abstract), `STEREO-CHEMISTRY-ABSTRACT` (contains a separate set of stereochemistry abstracts).

journal-item/files-info/web-pdf/pdf-version

`pdf-version` is the version of the web-PDF file, the allowed values are `1.7 6.4`, `1.7 6.5`, and `1.7 7.0`.

journal-item/files-info/web-pdf/pdf-property

The element `pdf-property` describes the nature of the web PDF file and has one of the following values: `WRAPPED OPTIMIZED`, `DISTILLED OPTIMIZED`, `DISTILLED OPTIMIZED BOOKMARKED`.

6.14. Q300 deliveries

A Q300 journal issue is a deliverable of the CAP process that precedes the S300-H300 delivery. It is used for proofing the dataset prior to publication. In particular the issue's hub file can be proofed this way.

The Q300 dataset directory structure is defined in Section 4.1.9.

Below, we traverse the Q300 schema for serials starting from the element `dataset-content`. The elements `dataset-unique-ids` and `dataset-properties` are explained in Section 4.2.

A Q300 schema contains one journal issue. It can always be accompanied with all or some of its items, and it has the option to add web PDF files of issue pages for the purpose of proofing.

journal-issue

The element `journal-issue` contains all the information pertaining to the issue in the dataset. The files associated with the issue are the issue hub with possible assets.

journal-issue/version

journal-issue/version/version-number

The element `version-number` contains the version number of the issue, as described in Section 3.2. The version number is assigned by Elsevier and is included in the order.

journal-issue/version/stage

The value of `stage` is `Q300`.

journal-issue/journal-issue-unique-ids

journal-issue/journal-issue-unique-ids/pii

The element `pii` contains the PII of the issue. In the case of book series this is a book hub PII or a collection PII. It must be equal to the PII in the XML hub file.

journal-issue/journal-issue-unique-ids/doi

The element `doi` contains the DOI of the issue, if any. In the case of book series the DOI may be based on a book hub PII or a collection PII. Only issues that will appear online may have a DOI. It must be equal to the DOI in the XML hub file.

journal-issue/journal-issue-properties

journal-issue/journal-issue-properties/jid

The element `jid` contains the Elsevier system code of the serial publication to which the issue belongs. Book series JIDs should begin with "BS:".

journal-issue/journal-issue-properties/issn

The element `issn` contains the ISSN of the serial publication to which the issue belongs.

journal-issue/journal-issue-properties/volume-issue-number

The volume/issue number of the issue is captured in `volume-issue-number`. It consists of `vol-first`, `vol-last`, `iss-first`, `iss-last` and `suppl`.

journal-issue/journal-issue-properties/isbn

The issue's optional ISBN, an ISBN-13, is captured with `isbn`.

journal-issue/files-info

The element `files-info` contains all the information needed to process the files belonging to the issue.

journal-issue/files-info/ml

The issue hub is an XML file associated with the issue. Its details are listed under `ml`.

journal-issue/files-info/ml/pathname

`pathname` is the pathname, relative to dataset.xml (that is, relative to the base directory), of the hub XML file.

journal-issue/files-info/ml/purpose

The element `purpose` always contains the value `MAIN`.

journal-issue/files-info/ml/dtd-version

The version of the DTD of the hub file is contained in `dtd-version`. The version must, of course, be identical to the declaration in the XML file.

journal-issue/files-info/ml/asset

All assets belonging to the item are listed under `asset`. This includes all external files declared in the XML file; it does not include any strip-ins or fingerprints.

journal-issue/files-info/ml/asset/pathname

`pathname` is the pathname, relative to dataset.xml, of the asset file.

journal-issue/files-info/ml/asset/type

The `type` of the asset can have the values `IMAGE-CAP`, `IMAGE-COVER`. These types are defined in Section 3.3.

journal-issue/files-info/web-pdf

Beside the hub XML file, there can be any number of issue-level PDF files. These PDF files are untyped, i.e. there is no subelement `purpose`. Element `web-pdf` is optional.

journal-issue/files-info/web-pdf/pathname

`pathname` is the pathname, relative to dataset.xml, of the web PDF file.

journal-issue/files-info/web-pdf/pdf-version

`pdf-version` is the version of the web-PDF file, the allowed values are `1.7 6.4`, `1.7 6.5`, and `1.7 7.0`.

journal-item

The element `journal-item` contains all the information pertaining to an item in a serial publication. It can have three attributes, `type`, `omitted` and `cross-mark`.

The first one, `type`, takes the values `stand-alone` (default), `with-add-ons` and `batch-placeholder`. In Section 3.4 (p. 29) an explanation is given when `type` is used.

The second one, `omitted`, takes the values `false` (default) and `true`. The latter value is to be used if an item is omitted from the dataset. It should only be used if the `dataset-action` has value `PARTIAL-RELOAD`. (See also p. 41 for information on partial deliveries.)

The third one, `cross-mark`, takes the values `false` and `true`.

journal-item/version

journal-item/version/version-number

The element `version-number` contains the version number of the item, as described in Section 3.2. The version number is assigned by Elsevier and is included in the order.

journal-item/version/stage

The value of `stage` is `Q300`.

journal-item/journal-item-unique-ids

journal-item/journal-item-unique-ids/pii

The element `pii` contains the PII of the item. In the case of book series this is a book PII. It must be equal to the PII in the XML file and the PII in the document properties of the main PDF file (if appropriate).

journal-item/journal-item-unique-ids/doi

The element `doi` contains the DOI of the item, if any. In the case of book series the DOI may be based on a book PII. Only items that will appear online may have a DOI. It must be equal to the DOI in the XML file and the DOI in the document properties of the main PDF file (if appropriate).

journal-item/journal-item-unique-ids/jid-aid

The element `jid-aid` contains up to four subelements: `jid` is the Elsevier system code and `issn` the (formatted) ISSN of the serial publication to which the item belongs; `aid` is the Elsevier system article ID and `article-number` is the item's article number.

The first three elements are mandatory while the last one is optional. The `jid`, `aid` and `article-number` must be equal to the same elements in the item's XML file(s). Book series JIDs should begin with "BS:".

journal-item/journal-item-properties

journal-item/journal-item-properties/pit

The element `pit` contains the publication item type of the item. The value of `pit` must be identical to the value of the top-level attribute `docsubtype` in the XML file, except for the case: PITs are always written uppercase in dataset.xml files and lowercase in DTD 5.x files.

journal-item/journal-item-properties/production-type

The element `production-type` can take the values `NON-CRC` and `CRC`.

journal-item/journal-item-properties/batch

If a journal item is part of a batch, then element `batch` must be present. It contains `pii`, the PII of the journal item that represents the batch, and `doi`, its optional DOI. Note that the item that is referred to, itself an independent item, has the attribute `type` set to a non-default value (see Section 3.4, p. 29).

journal-item/files-info

The element `files-info` contains all the information needed to process the files belonging to the item.

journal-item/files-info/ml

One or two XML files can be associated with the item. These are listed under `ml`.

journal-item/files-info/ml/pathname

In case `journal-item`'s attribute `omitted` is not present or has the value `false`, `pathname` is the pathname, relative to dataset.xml, of the XML file.

In case `journal-item`'s attribute `omitted` has the value `true`, `pathname` is the pathname of the XML file in the system it resides in (an absolute pathname starting with a server name).

journal-item/files-info/ml/purpose

`purpose` indicates what the XML file is for. There must always be one XML file with purpose `MAIN`. The optional second XML file is a changes-with-respect-to file with purpose `CHANGES`. A third optional XML file is a file containing pagebreak information, the purpose is `PAGEBREAK`.

journal-item/files-info/ml/dtd-version

The version of the DTD and the top-level element (doctype) used to capture the item is contained in `dtd-version`. This must, of course, be identical to the declaration in the XML file.

journal-item/files-info/ml/weight

The weight of the XML file can be `FULL-TEXT`, `HEAD-AND-TAIL`, `HEAD-ONLY` or `CONTENTS-ENTRY-ONLY`. The weight of the item indicates which parts of the text are captured in XML.

journal-item/files-info/ml/asset

All assets belonging to the item are listed under `asset`. This includes all external files declared in the XML file; it does not include any strip-ins or fingerprints.

journal-item/files-info/ml/asset/pathname

In case `journal-item`'s attribute `omitted` is not present or has the value `false`, `pathname` is the pathname, relative to dataset.xml, of the XML file.

In case `journal-item`'s attribute `omitted` has the value `true`, `pathname` is the pathname of the XML file in the system it resides in (an absolute pathname starting with a server name).

journal-item/files-info/ml/asset/type

The `type` of the asset can have the values `APPLICATION`, `IMAGE-CAP`, `AUDIO`, `VIDEO`, `VIDEO-FLASH`, `XML`. These types are defined in Section 3.3.

journal-item/files-info/web-pdf

A Q300 item must have at least one, and can have up to four, web PDF files associated with it.

journal-item/files-info/web-pdf/pathname

In case `journal-item`'s attribute `omitted` is not present or has the value `false`, `pathname` is the pathname, relative to dataset.xml, of the XML file.

In case `journal-item`'s attribute `omitted` has the value `true`, `pathname` is the pathname of the XML file in the system it resides in (an absolute pathname starting with a server name).

journal-item/files-info/web-pdf/purpose

`purpose` indicates what the web PDF file is for. There may not be any PDF files with the same purpose.

There must always be one PDF file with purpose `MAIN`. This contains the PDF file of the item.

The other optional web PDF files have values `MAIN-ABRIDGED` (contains the abridged print version), `GRAPHICAL-ABSTRACT` (contains a separate graphical abstract), `STEREO-CHEMISTRY-ABSTRACT` (contains a separate set of stereochemistry abstracts).

journal-item/files-info/web-pdf/pdf-version

`pdf-version` is the version of the web-PDF file, the allowed values are `1.7 6.4`, `1.7 6.5`, and `1.7 7.0`.

journal-item/files-info/web-pdf/pdf-property

The element `pdf-property` describes the nature of the web PDF file and has one of the following values: `WRAPPED OPTIMIZED`, `DISTILLED OPTIMIZED`, `DISTILLED OPTIMIZED BOOKMARKED`.

6.15. Q300 Project deliveries

A Q300 journal issue is a deliverable of the CAP process that precedes the S300-H300 delivery. It is used for proofing the dataset prior to publication. In particular the issue's hub file can be proofed this way.

The Q300-Project schema supports deliveries of material for special projects. It is based on the regular Q300 schema. Some pattern validation has been removed. This validation will be performed elsewhere and is not described below. In fact, below is assumed that this validation does take place. Additionally some lists of values have been expanded.

The Q300 dataset directory structure is defined in Section 4.1.9.

Below, we traverse the Q300-Project schema for serials starting from the element `dataset-content`. The elements `dataset-unique-ids` and `dataset-properties` are explained in Section 4.2.

A Q300-Project schema contains one journal issue. It can always be accompanied with all or some of its items, and it has the option to add web PDF files of issue pages for the purpose of proofing.

journal-issue

The element `journal-issue` contains all the information pertaining to the issue in the dataset. The files associated with the issue are the issue hub with possible assets.

journal-issue/version

journal-issue/version/version-number

The element `version-number` contains the version number of the issue, as described in Section 3.2. The version number is assigned by Elsevier and is included in the order.

journal-issue/version/stage

The value of `stage` is `Q300`.

journal-issue/journal-issue-unique-ids

journal-issue/journal-issue-unique-ids/pii

The element `pii` contains the PII of the issue. In the case of book series this is a book hub PII or a collection PII. It must be equal to the PII in the XML hub file.

journal-issue/journal-issue-unique-ids/doi

The element `doi` contains the DOI of the issue, if any. In the case of book series the DOI may be based on a book hub PII or a collection PII. Only issues that will appear online may have a DOI. It must be equal to the DOI in the XML hub file.

journal-issue/journal-issue-properties

journal-issue/journal-issue-properties/jid

The element `jid` contains the Elsevier system code of the serial publication to which the issue belongs. Book series JIDs should begin with "BS:".

journal-issue/journal-issue-properties/issn

The element `issn` contains the ISSN of the serial publication to which the issue belongs.

journal-issue/journal-issue-properties/volume-issue-number

The volume/issue number of the issue is captured in `volume-issue-number`. It consists of `vol-first`, `vol-last`, `iss-first`, `iss-last` and `suppl`.

journal-issue/journal-issue-properties/embargo

The optional element `embargo` contains a date/time before which the issue is not allowed to be published. Its content is in W3C schema DateTime format, `xs:dateTime`, and should be accurate to the minute. To avoid confusion the date/time should be in UTC, signified by a "Z". That is, the embargo date/time should be in the following format: "*yyyy-mm-dd*T*hh*:*mm*:00Z".

```
<embargo>2006-04-11T12:00:00Z</embargo>
```

journal-issue/journal-issue-properties/isbn

The issue's optional ISBN, an ISBN-10 or an ISBN-13, is captured with `isbn`.

journal-issue/files-info

The element `files-info` contains all the information needed to process the files belonging to the issue.

journal-issue/files-info/ml

The issue hub is an XML file associated with the issue. Its details are listed under `ml`.

journal-issue/files-info/ml/pathname

`pathname` is the pathname, relative to dataset.xml (that is, relative to the base directory), of the hub XML file.

journal-issue/files-info/ml/purpose

The element `purpose` always contains the value `MAIN`.

journal-issue/files-info/ml/dtd-version

The version of the DTD of the hub file is contained in `dtd-version`. The version must, of course, be identical to the declaration in the XML file.

journal-issue/files-info/ml/asset

All assets belonging to the item are listed under `asset`. This includes all external files declared in the XML file; it does not include any strip-ins or fingerprints.

journal-issue/files-info/ml/asset/pathname

`pathname` is the pathname, relative to dataset.xml, of the asset file.

journal-issue/files-info/ml/asset/type

The `type` of the asset can have the values `IMAGE-CAP`, `IMAGE-COVER`. These types are defined in Section 3.3.

journal-issue/files-info/web-pdf

Beside the hub XML file, there can be any number of issue-level PDF files. These PDF files are untyped, i.e. there is no subelement `purpose`. Element `web-pdf` is optional.

journal-issue/files-info/web-pdf/pathname

`pathname` is the pathname, relative to dataset.xml, of the web PDF file.

journal-issue/files-info/web-pdf/pdf-version

`pdf-version` is the version of the web-PDF file, the allowed values are `1.7 6.4`, `1.7 6.5`, and `1.7 7.0`.

journal-item

The element `journal-item` contains all the information pertaining to an item in a serial publication. It can have three attributes, `type`, `omitted` and `cross-mark`.

The first one, `type`, takes the values `stand-alone` (default), `with-add-ons` and `batch-placeholder`. In Section 3.4 (p. 29) an explanation is given when `type` is used.

The second one, `omitted`, takes the values `false` (default) and `true`. The latter value is to be used if an item is omitted from the dataset. It should only be used if the `dataset-action` has value `PARTIAL-RELOAD`. (See also p. 41 for information on partial deliveries.)

The third one, `cross-mark`, takes the values `false` and `true`.

journal-item/version

journal-item/version/version-number

The element `version-number` contains the version number of the item, as described in Section 3.2. The version number is assigned by Elsevier and is included in the order.

journal-item/version/stage

The value of `stage` is `Q300`.

journal-item/embargo

The optional element `embargo` contains a date/time before which the item is not allowed to be published. Its content is in W3C schema DateTime format, `xs:dateTime`, and should be accurate to the minute. To avoid confusion the date/time should be in UTC, signified by a "Z". That is, the embargo date/time should be in the following format: "*yyyy-mm-dd*T*hh*:*mm*:00Z".

```
<embargo>2006-04-11T12:00:00Z</embargo>
```

journal-item/journal-item-unique-ids

journal-item/journal-item-unique-ids/pii

The element `pii` contains the PII of the item. In the case of book series this is a book PII. It must be equal to the PII in the XML file and the PII in the document properties of the main PDF file (if appropriate).

journal-item/journal-item-unique-ids/doi

The element `doi` contains the DOI of the item, if any. In the case of book series the DOI may be based on a book PII. Only items that will appear online may have a DOI. It must be equal to the DOI in the XML file and the DOI in the document properties of the main PDF file (if appropriate).

journal-item/journal-item-unique-ids/jid-aid

The element `jid-aid` contains up to four subelements: `jid` is the Elsevier system code and `issn` the (formatted) ISSN of the serial publication to which the item belongs; `aid` is the Elsevier system article ID and `article-number` is the item's article number.

The first three elements are mandatory while the last one is optional. The `jid`, `aid` and `article-number` must be equal to the same elements in the item's XML file(s). Book series JIDs should begin with "BS:".

journal-item/journal-item-properties

journal-item/journal-item-properties/pit

The element `pit` contains the publication item type of the item. The value of `pit` must be identical to the value of the top-level attribute `docsubtype` in the XML file, except for the case: PITs are always written uppercase in dataset.xml files and lowercase in DTD 5.x files.

journal-item/journal-item-properties/production-type

The element `production-type` can take the values `NON-CRC` and `CRC`.

journal-item/journal-item-properties/batch

If a journal item is part of a batch, then element `batch` must be present. It contains `pii`, the PII of the journal item that represents the batch, and `doi`, its optional DOI. Note that the item that is referred to, itself an independent item, has the attribute `type` set to a non-default value (see Section 3.4, p. 29).

journal-item/files-info

The element `files-info` contains all the information needed to process the files belonging to the item.

journal-item/files-info/ml

One or two XML files can be associated with the item. These are listed under `ml`.

journal-item/files-info/ml/pathname

In case `journal-item`'s attribute `omitted` is not present or has the value `false`, `pathname` is the pathname, relative to dataset.xml, of the XML file.

In case `journal-item`'s attribute `omitted` has the value `true`, `pathname` is the pathname of the XML file in the system it resides in (an absolute pathname starting with a server name).

journal-item/files-info/ml/purpose

`purpose` indicates what the XML file is for. There must always be one XML file with purpose `MAIN`. An optional second XML file is a changes-with-respect-to file with purpose `CHANGES`. A third optional XML file is a file containing pagebreak information, the purpose is `PAGEBREAK`. A fourth optional XML file is a file containing metadata for the RadCon project, the purpose is `AUXILIARY`. See the example in Fig. 6 (p. 140).

journal-item/files-info/ml/dtd-version

The version of the DTD and the top-level element (doctype) used to capture the item is contained in `dtd-version`. This must, of course, be identical to the declaration in the XML file.

This is not the case for RadCon metadata files. These files are structured according to a MicroSoft Infopath schema and the dtd-version is `INFOPATH`. See the example in Fig. 6 (p. 140).

journal-item/files-info/ml/weight

The weight of the XML file can be `FULL-TEXT`, `HEAD-AND-TAIL`, `HEAD-ONLY` or `CONTENTS-ENTRY-ONLY`. The weight of the item indicates which parts of the text are captured in XML.

journal-item/files-info/ml/asset

All assets belonging to the item are listed under `asset`. This includes all external files declared in the XML file; it does not include any strip-ins or fingerprints.

journal-item/files-info/ml/asset/pathname

In case `journal-item`'s attribute `omitted` is not present or has the value `false`, `pathname` is the pathname, relative to dataset.xml, of the XML file.

In case `journal-item`'s attribute `omitted` has the value `true`, `pathname` is the pathname of the XML file in the system it resides in (an absolute pathname starting with a server name).

journal-item/files-info/ml/asset/type

The `type` of the asset can have the values `APPLICATION`, `IMAGE-CAP`, `IMAGE-NONCAP`, `IMAGE-DOWNSAMPLED`, `IMAGE-THUMBNAIL`, `IMAGE-MMC`, `IMAGE-MMC-DOWNSAMPLED`, `IMAGE-MMC-THUMBNAIL`, `AUDIO`, `VIDEO`, `VIDEO-FLASH`, `XML`. These types are defined in Section 3.3.

journal-item/files-info/web-pdf

A Q300 item must have at least one, and can have up to four, web PDF files associated with it.

journal-item/files-info/web-pdf/pathname

In case `journal-item`'s attribute `omitted` is not present or has the value `false`, `pathname` is the pathname, relative to dataset.xml, of the XML file.

In case `journal-item`'s attribute `omitted` has the value `true`, `pathname` is the pathname of the XML file in the system it resides in (an absolute pathname starting with a server name).

journal-item/files-info/web-pdf/purpose

`purpose` indicates what the web PDF file is for. There may not be any PDF files with the same purpose.

There must always be one PDF file with purpose `MAIN`. This contains the PDF file of the item.

The other optional web PDF files have values `MAIN-ABRIDGED` (contains the abridged print version), `GRAPHICAL-ABSTRACT` (contains a separate graphical abstract), `STEREO-CHEMISTRY-ABSTRACT` (contains a separate set of stereochemistry abstracts).

journal-item/files-info/web-pdf/pdf-version

`pdf-version` is the version of the web-PDF file, the allowed values are `1.7 6.4`, `1.7 6.5`, and `1.7 7.0`.

journal-item/files-info/web-pdf/pdf-property

The element `pdf-property` describes the nature of the web PDF file and has one of the following values: `WRAPPED OPTIMIZED`, `DISTILLED OPTIMIZED`, `DISTILLED OPTIMIZED BOOKMARKED`.

6.16. S300-H300 deliveries

An S300 item in a serial publication is a corrected proof, authorized by the authors of the article and/or the editors of the publication, which has received its confirmed publication details (volume and issue number and page range).

An H300 issue is an issue whose table of contents is complete. Each of its items has reached the S300 stage.

The S300-H300 dataset directory structure for serials is defined in Section 4.1.9.

Below, we traverse the S300-H300 schema for serials starting from the element `dataset-content`. The elements `dataset-unique-ids` and `dataset-properties` are explained in Section 4.2.

An S300-H300 schema contains precisely one issue, accompanied by one of more items. The items must belong to the issue. If the dataset action is `LOAD`, then the issue must be accompanied by all its items.

journal-issue

The element `journal-issue` contains all the information pertaining to the issue in the dataset. The files associated with the issue are the issue hub with possible assets.

journal-issue/version

journal-issue/version/version-number

The element `version-number` contains the version number of the issue, as described in Section 3.2. The version number is assigned by Elsevier and is included in the order.

journal-issue/version/stage

The value of `stage` is `H300`.

journal-issue/journal-issue-unique-ids

journal-issue/journal-issue-unique-ids/pii

The element `pii` contains the PII of the issue. In the case of book series this is a book hub PII or a collection PII. It must be equal to the PII in the XML hub file.

journal-issue/journal-issue-unique-ids/doi

The element `doi` contains the DOI of the issue, if any. In the case of book series the DOI may be based on a book hub PII or a collection PII. Only issues that will appear online may have a DOI. It must be equal to the DOI in the XML hub file.

journal-issue/journal-issue-properties

journal-issue/journal-issue-properties/jid

The element `jid` contains the Elsevier system code of the serial publication to which the issue belongs. Book series JIDs should begin with "BS:".

journal-issue/journal-issue-properties/issn

The element `issn` contains the ISSN of the serial publication to which the issue belongs.

journal-issue/journal-issue-properties/volume-issue-number

The volume/issue number of the issue is captured in `volume-issue-number`. It consists of `vol-first`, `vol-last`, `iss-first`, `iss-last` and `suppl`.

journal-issue/journal-issue-properties/embargo

The optional element `embargo` contains a date/time before which the issue is not allowed to be published. Its content is in W3C schema DateTime format, `xs:dateTime`, and should be accurate to the minute. To avoid confusion the date/time should be in UTC, signified by a "Z". That is, the embargo date/time should be in the following format: "*yyyy-mm-dd*T*hh*:*mm*:00Z".

```
<embargo>2006-04-11T12:00:00Z</embargo>
```

journal-issue/journal-issue-properties/isbn

The issue's optional ISBN, an ISBN-13, is captured with `isbn`.

journal-issue/files-info

The element `files-info` contains all the information needed to process the files belonging to the issue.

journal-issue/files-info/ml

The issue hub is an XML file associated with the issue. Its details are listed under `ml`.

journal-issue/files-info/ml/pathname

`pathname` is the pathname, relative to dataset.xml (that is, relative to the base directory), of the hub XML file.

journal-issue/files-info/ml/purpose

The element `purpose` always contains the value `MAIN`.

journal-issue/files-info/ml/dtd-version

The version of the DTD of the hub file is contained in `dtd-version`. The version must, of course, be identical to the declaration in the XML file.

journal-issue/files-info/ml/asset

All assets belonging to the item are listed under `asset`. This includes all external files declared in the XML file; it does not include any strip-ins or fingerprints.

journal-issue/files-info/ml/asset/pathname

`pathname` is the pathname, relative to dataset.xml, of the asset file.

journal-issue/files-info/ml/asset/type

The `type` of the asset can have the values `IMAGE-CAP`, `IMAGE-COVER`. These types are defined in Section 3.3.

journal-issue/files-info/web-pdf

Beside the hub XML file, there can be an issue-level PDF file which contains Tables of Contents. The PDF files are untyped, i.e. there is no subelement `purpose`. Element `web-pdf` is optional.

journal-issue/files-info/web-pdf/pathname

`pathname` is the pathname, relative to dataset.xml, of the web PDF file.

journal-issue/files-info/web-pdf/pdf-version

`pdf-version` is the version of the web-PDF file, the allowed values are `1.7 6.4`, `1.7 6.5`, and `1.7 7.0`.

journal-item

The element `journal-item` contains all the information pertaining to an item in a serial publication. It can have three attributes, `type`, `omitted` and `cross-mark`.

The first one, `type`, takes the values `stand-alone` (default), `with-add-ons` and `batch-placeholder`. In Section 3.4 (p. 29) an explanation is given when `type` is used.

The second one, `omitted`, takes the values `false` (default) and `true`. The latter value is to be used if an item is omitted from the dataset. It should only be used if the `dataset-action` has value `PARTIAL-RELOAD`. (See also p. 41 for information on partial deliveries.)

The third one, `cross-mark`, takes the values `false` and `true`.

journal-item/version

journal-item/version/version-number

The element `version-number` contains the version number of the item, as described in Section 3.2. The version number is assigned by Elsevier and is included in the order.

journal-item/version/stage

The value of `stage` is `S300`.

journal-item/journal-item-unique-ids

journal-item/journal-item-unique-ids/pii

The element pii contains the PII of the item. In the case of book series this is a book PII. It must be equal to the PII in the XML file and the PII in the document properties of the main PDF file (if appropriate).

journal-item/journal-item-unique-ids/doi

The element doi contains the DOI of the item, if any. In the case of book series the DOI may be based on a book PII. Only items that will appear online may have a DOI. It must be equal to the DOI in the XML file and the DOI in the document properties of the main PDF file (if appropriate).

journal-item/journal-item-unique-ids/jid-aid

The element jid-aid contains up to four subelements: jid is the Elsevier system code and issn the (formatted) ISSN of the serial publication to which the item belongs; aid is the Elsevier system article ID and article-number is the item's article number.

The first three elements are mandatory while the last one is optional. The jid, aid and article-number must be equal to the same elements in the item's XML file(s). Book series JIDs should begin with "BS:".

journal-item/journal-item-properties

journal-item/journal-item-properties/pit

The element pit contains the publication item type of the item. The value of pit must be identical to the value of the top-level attribute docsubtype in the XML file, except for the case: PITs are always written uppercase in dataset.xml files and lowercase in DTD 5.x files.

journal-item/journal-item-properties/production-type

The element production-type can take the values NON-CRC and CRC.

journal-item/journal-item-properties/batch

If a journal item is part of a batch, then element batch must be present. It contains pii, the PII of the journal item that represents the batch, and doi, its optional DOI. Note that the item that is referred to, itself an independent item, has the attribute type set to a non-default value (see Section 3.4, p. 29).

journal-item/files-info

The element files-info contains all the information needed to process the files belonging to the item.

journal-item/files-info/ml

One or two XML files can be associated with the item. These are listed under ml.

journal-item/files-info/ml/pathname

In case journal-item's attribute omitted is not present or has the value false, pathname is the pathname, relative to dataset.xml, of the XML file.

In case journal-item's attribute omitted has the value true, pathname is the pathname of the XML file in the system it resides in (an absolute pathname starting with a server name).

journal-item/files-info/ml/purpose

purpose indicates what the XML file is for. There must always be one XML file with purpose MAIN. The optional second XML file is a changes-with-respect-to file with purpose CHANGES. A third optional XML file is a file containing pagebreak information, the purpose is PAGEBREAK.

journal-item/files-info/ml/dtd-version

The version of the DTD and the top-level element (doctype) used to capture the item is contained in dtd-version. This must, of course, be identical to the declaration in the XML file.

journal-item/files-info/ml/weight

The weight of the XML file can be FULL-TEXT, HEAD-AND-TAIL, HEAD-ONLY or CONTENTS-ENTRY-ONLY. The weight of the item indicates which parts of the text are captured in XML.

journal-item/files-info/ml/asset

All assets belonging to the item are listed under asset. This includes all external files declared in the XML file; it does not include any strip-ins or fingerprints.

journal-item/files-info/ml/asset/pathname

In case journal-item's attribute omitted is not present or has the value false, pathname is the pathname, relative to dataset.xml, of the XML file.

In case journal-item's attribute omitted has the value true, pathname is the pathname of the XML file in the system it resides in (an absolute pathname starting with a server name).

journal-item/files-info/ml/asset/type

The type of the asset can have the values APPLICATION, IMAGE-CAP, AUDIO, VIDEO, VIDEO-FLASH, XML. These types are defined in Section 3.3.

journal-item/files-info/web-pdf

An S300 item has one or two web PDF files associated with it.

journal-item/files-info/web-pdf/pathname

In case journal-item's attribute omitted is not present or has the value false, pathname is the pathname, relative to dataset.xml, of the XML file.

In case journal-item's attribute omitted has the value true, pathname is the pathname of the XML file in the system it resides in (an absolute pathname starting with a server name).

journal-item/files-info/web-pdf/purpose

purpose indicates what the web PDF file is for. There may not be any PDF files with the same purpose.

There must always be one PDF file with purpose MAIN. This contains the PDF file of the item.

The other optional web PDF file has value MAIN-ABRIDGED (contains the abridged print version).

journal-item/files-info/web-pdf/pdf-version

pdf-version is the version of the web-PDF file, the allowed values are 1.7 6.4, 1.7 6.5, and 1.7 7.0.

journal-item/files-info/web-pdf/pdf-property

The element pdf-property describes the nature of the web PDF file and has one of the following values: WRAPPED OPTIMIZED, DISTILLED OPTIMIZED, DISTILLED OPTIMIZED BOOKMARKED.

```
<?xml version="1.0" encoding="UTF-8" ?>
<dataset
  xmlns="http://www.elsevier.com/xml/schema/transport/journal-1.12/s300-h300"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.elsevier.com/xml/schema/transport/journal-1.12/s300-h300
    http://www.elsevier.com/xml/schema/transport/journal-1.12/s300-h300.xsd"
  schema-version="1.12">
  <dataset-unique-ids>
    <supplier-code>aldmm</supplier-code>
    <supplier-dataset-id>ALDMM0004391</supplier-dataset-id>
    <timestamp>2004-06-10T16:04:26</timestamp>
  </dataset-unique-ids>
  <dataset-properties>
    <dataset-action>LOAD</dataset-action>
    <production-process>CAP</production-process>
  </dataset-properties>
  <dataset-content>
    <journal-issue>
      <version>
        <version-number>H300.1</version-number>
        <stage>H300</stage>
      </version>
      <journal-issue-unique-ids>
        <pii>S9999-9994(03)X7607-2</pii>
      </journal-issue-unique-ids>
      <journal-issue-properties>
        <jid>MANSC</jid>
        <issn>0276-8976</issn>
        <volume-issue-number>
          <vol-first>11</vol-first>
          <iss-first>1</iss-first>
        </volume-issue-number>
        <isbn>978-0-7623-1095-1</isbn>
      </journal-issue-properties>
      <files-info>
        <ml>
          <pathname>02768976/v11i1/issue.xml</pathname>
          <purpose>MAIN</purpose>
          <dtd-version>SI 5.1.0</dtd-version>
          <asset>
            <pathname>02768976/v11i1/issue.assets/cover.tif</pathname>
            <type>IMAGE-COVER</type>
          </asset>
        </ml>
      </files-info>
    </journal-issue>
    <journal-item>
      <version>
        <version-number>S300.1</version-number>
        <stage>S300</stage>
      </version>
      <journal-item-unique-ids>
        <pii>S0276-8976(04)11001-8</pii>
        <doi>10.1016/S0276-8976(04)11001-8</doi>
        <jid-aid>
          <jid>MANSC</jid>
          <issn>0276-8976</issn>
          <aid>11001</aid>
        </jid-aid>
      </journal-item-unique-ids>
      <journal-item-properties>
        <pit>REV</pit>
        <production-type>NON-CRC</production-type>
      </journal-item-properties>
```

Figure 5: Sample S300-H300 dataset.xml.

```
      <files-info>
        <ml>
          <pathname>02768976/v11i1/S0276897604110018/main.xml</pathname>
          <purpose>MAIN</purpose>
          <dtd-version>JA 5.0.2 ARTICLE</dtd-version>
          <weight>FULL-TEXT</weight>
          <asset>
            <pathname>02768976/v11i1/S0276897604110018/main.assets/gr1.tif</pathname>
            <type>IMAGE-CAP</type>
          </asset>
        </ml>
        <web-pdf>
          <pathname>02768976/v11i1/S0276897604110018/main.pdf</pathname>
          <purpose>MAIN</purpose>
          <pdf-version>1.4 6.0</pdf-version>
          <pdf-property>DISTILLED OPTIMIZED BOOKMARKED</pdf-property>
        </web-pdf>
      </files-info>
    </journal-item>
    <journal-item>
        <version>
            <version-number>S300.5</version-number>
            <stage>S300</stage>
        </version>
        <journal-item-unique-ids>
            <pii>S0276-8976(04)11002-X</pii>
            <doi>10.1016/S0276-8976(04)11002-X</doi>
            <jid-aid>
                <jid>MANSC</jid>
                <issn>0276-8976</issn>
                <aid>11002</aid>
            </jid-aid>
        </journal-item-unique-ids>
        <journal-item-properties>
            <pit>REV</pit>
            <production-type>NON-CRC</production-type>
        </journal-item-properties>
        <files-info>
            <ml>
                <pathname>02768976/v11i1/S027689760411002X/main.xml</pathname>
                <purpose>MAIN</purpose>
                <dtd-version>JA 5.0.2 ARTICLE</dtd-version>
                <weight>FULL-TEXT</weight>
                <asset>
                    <pathname>02768976/v11i1/S027689760411002X/main.assets/gr1.tif</pathname>
                    <type>IMAGE-CAP</type>
                </asset>
                <asset>
                    <pathname>02768976/v11i1/S027689760411002X/main.assets/gr2.jc3</pathname>
                    <type>IMAGE-CAP</type>
                </asset>
                <asset>
                    <pathname>02768976/v11i1/S027689760411002X/main.assets/gr3.jc3</pathname>
                    <type>IMAGE-CAP</type>
                </asset>
            </ml>
            <web-pdf>
                <pathname>02768976/v11i1/S027689760411002X/main.pdf</pathname>
                <purpose>MAIN</purpose>
                <pdf-version>1.4 6.0</pdf-version>
                <pdf-property>DISTILLED OPTIMIZED BOOKMARKED</pdf-property>
            </web-pdf>
        </files-info>
    </journal-item>
    .
    .
    .
  </dataset-content>
</dataset>
```

Figure 5: Sample S300-H300 dataset.xml (continued).

6.17. S300-H300 Project deliveries

The S300-H300-Project schema supports deliveries of material for special projects. It is based on the regular S300-H300 schema. Some pattern validation has been removed. This validation will be performed elsewhere and is not described below. In fact, below is assumed that this validation does take place. Additionally some lists of values have been expanded.

An S300-H300-Project schema contains precisely one issue, accompanied by one of more items. The items must belong to the issue. If the dataset action is `LOAD`, then the issue must be accompanied by all its items.

journal-issue

The element `journal-issue` contains all the information pertaining to the issue in the dataset. The files associated with the issue are the issue hub with possible assets.

journal-issue/version

journal-issue/version/version-number

The element `version-number` contains the version number of the issue, as described in Section 3.2. The version number is assigned by Elsevier and is included in the order.

journal-issue/version/stage

The value of `stage` is `H300`.

journal-issue/journal-issue-unique-ids

journal-issue/journal-issue-unique-ids/pii

The element `pii` contains the PII of the issue. In the case of book series this is a book hub PII or a collection PII. It must be equal to the PII in the XML hub file.

journal-issue/journal-issue-unique-ids/doi

The element `doi` contains the DOI of the issue, if any. In the case of book series the DOI may be based on a book hub PII or a collection PII. Only issues that will appear online may have a DOI. It must be equal to the DOI in the XML hub file.

journal-issue/journal-issue-properties

journal-issue/journal-issue-properties/jid

The element `jid` contains the Elsevier system code of the serial publication to which the issue belongs. Book series JIDs should begin with “BS:”.

journal-issue/journal-issue-properties/issn

The element `issn` contains the ISSN of the serial publication to which the issue belongs.

journal-issue/journal-issue-properties/volume-issue-number

The volume/issue number of the issue is captured in `volume-issue-number`. It consists of `vol-first`, `vol-last`, `iss-first`, `iss-last` and `suppl`.

journal-issue/journal-issue-properties/embargo

The optional element `embargo` contains a date/time before which the issue is not allowed to be published. Its content is in W3C schema DateTime format, `xs:dateTime`, and should be accurate to the minute. To avoid confusion the date/time should be in UTC, signified by a "Z". That is, the embargo date/time should be in the following format: "*yyyy-mm-dd*T*hh*:*mm*:00Z".

```
<embargo>2006-04-11T12:00:00Z</embargo>
```

journal-issue/journal-issue-properties/isbn

The issue's optional ISBN, an ISBN-10 or an ISBN-13, is captured with `isbn`.

journal-issue/files-info

The element `files-info` contains all the information needed to process the files belonging to the issue.

journal-issue/files-info/ml

The issue hub is an XML file associated with the issue. Its details are listed under `ml`.

journal-issue/files-info/ml/pathname

`pathname` is the pathname, relative to dataset.xml (that is, relative to the base directory), of the hub XML file.

journal-issue/files-info/ml/purpose

The element `purpose` always contains the value `MAIN`.

journal-issue/files-info/ml/dtd-version

The version of the DTD of the hub file is contained in `dtd-version`. The version must, of course, be identical to the declaration in the XML file.

journal-issue/files-info/ml/asset

All assets belonging to the item are listed under `asset`. This includes all external files declared in the XML file; it does not include any strip-ins or fingerprints.

journal-issue/files-info/ml/asset/pathname

`pathname` is the pathname, relative to dataset.xml, of the asset file.

journal-issue/files-info/ml/asset/type

The `type` of the asset can have the values `IMAGE-CAP`, `IMAGE-COVER`. These types are defined in Section 3.3.

journal-issue/files-info/web-pdf

Beside the hub XML file, there can be an issue-level PDF file which contains Tables of Contents. The PDF files are untyped, i.e. there is no subelement `purpose`. Element `web-pdf` is optional.

journal-issue/files-info/web-pdf/pathname

`pathname` is the pathname, relative to dataset.xml, of the web PDF file.

journal-issue/files-info/web-pdf/pdf-version

`pdf-version` is the version of the web-PDF file, the allowed values are `1.7 6.4`, `1.7 6.5`, and `1.7 7.0`.

journal-item

The element `journal-item` contains all the information pertaining to an item in a serial publication. It can have three attributes, `type`, `omitted` and `cross-mark`.

The first one, `type`, takes the values `stand-alone` (default), `with-add-ons` and `batch-placeholder`. In Section 3.4 (p. 29) an explanation is given when `type` is used.

The second one, `omitted`, takes the values `false` (default) and `true`. The latter value is to be used if an item is omitted from the dataset. It should only be used if the `dataset-action` has value `PARTIAL-RELOAD`. (See also p. 41 for information on partial deliveries.)

The third one, `cross-mark`, takes the values `false` and `true`.

journal-item/version

journal-item/version/version-number

The element `version-number` contains the version number of the item, as described in Section 3.2. The version number is assigned by Elsevier and is included in the order.

journal-item/version/stage

The value of `stage` is `S300`.

journal-item/journal-item-unique-ids

journal-item/journal-item-unique-ids/pii

The element `pii` contains the PII of the item. In the case of book series this is a book PII. It must be equal to the PII in the XML file and the PII in the document properties of the main PDF file (if appropriate).

journal-item/journal-item-unique-ids/doi

The element `doi` contains the DOI of the item, if any. In the case of book series the DOI may be based on a book PII. Only items that will appear online may have a DOI. It must be equal to the DOI in the XML file and the DOI in the document properties of the main PDF file (if appropriate).

journal-item/journal-item-unique-ids/jid-aid

The element `jid-aid` contains up to four subelements: `jid` is the Elsevier system code and `issn` the (formatted) ISSN of the serial publication to which the item belongs; `aid` is the Elsevier system article ID and `article-number` is the item's article number.

The first three elements are mandatory while the last one is optional. The `jid`, `aid` and `article-number` must be equal to the same elements in the item's XML file(s). Book series JIDs should begin with "BS:".

journal-item/journal-item-properties

journal-item/journal-item-properties/pit

The element `pit` contains the publication item type of the item. The value of `pit` must be identical to the value of the top-level attribute `docsubtype` in the XML file, except for the case: PITs are always written uppercase in dataset.xml files and lowercase in DTD 5.x files.

journal-item/journal-item-properties/production-type

The element `production-type` can take the values `NON-CRC` and `CRC`.

journal-item/journal-item-properties/batch

If a journal item is part of a batch, then element `batch` must be present. It contains `pii`, the PII of the journal item that represents the batch, and `doi`, its optional DOI. Note that the item that is referred to, itself an independent item, has the attribute `type` set to a non-default value (see Section 3.4, p. 29).

journal-item/files-info

The element `files-info` contains all the information needed to process the files belonging to the item.

journal-item/files-info/ml

One or two XML files can be associated with the item. These are listed under `ml`.

journal-item/files-info/ml/pathname

In case `journal-item`'s attribute `omitted` is not present or has the value `false`, `pathname` is the pathname, relative to dataset.xml, of the XML file.

In case `journal-item`'s attribute `omitted` has the value `true`, `pathname` is the pathname of the XML file in the system it resides in (an absolute pathname starting with a server name).

journal-item/files-info/ml/purpose

`purpose` indicates what the XML file is for. There must always be one XML file with purpose `MAIN`. An optional second XML file is a changes-with-respect-to file with purpose `CHANGES`. A third optional XML file is a file containing pagebreak information, the purpose

```
<journal-item>
  <version>
    <version-number>3.1</version-number>
    <stage>S300</stage>
  </version>
  <journal-item-unique-ids>
    <pii>S1933-0332(06)70948-X</pii>
    <doi>10.1016/S1933-0332(06)70948-X</doi>
    <jid-aid>
      <jid>BS:RADCON</jid>
      <issn>1933-0332</issn>
      <aid>70948</aid>
    </jid-aid>
  </journal-item-unique-ids>
  <journal-item-properties>
    <pit>MIS</pit>
    <production-type>NON-CRC</production-type>
  </journal-item-properties>
  <files-info>
    <ml>
      <pathname>19330332/v1i25/S193303320670948X/main.xml</pathname>
      <purpose>MAIN</purpose>
      <dtd-version>JA 5.0.1 SIMPLE-ARTICLE</dtd-version>
      <weight>FULL-TEXT</weight>
      <asset>
        <pathname>19330332/v1i25/S193303320670948X/main.assets/gr1.jc3</pathname>
        <type>IMAGE-CAP</type>
      </asset>
      <asset>
        <pathname>19330332/v1i25/S193303320670948X/main.assets/gr2.jpg</pathname>
        <type>IMAGE-CAP</type>
      </asset>
      <asset>
        <pathname>19330332/v1i25/S193303320670948X/main.assets/gr3.jc3</pathname>
        <type>IMAGE-CAP</type>
      </asset>
    </ml>
    <ml>
      <pathname>19330332/v1i25/S193303320670948X/metadata.xml</pathname>
      <purpose>AUXILIARY</purpose>
      <dtd-version>INFOPATH</dtd-version>
      <weight>FULL-TEXT</weight>
    </ml>
  </files-info>
</journal-item>
```

Figure 6: Example of (part of) a S300-H300 Project dataset.xml for a RadCon (a.k.a. ImageConsult) delivery.

is `PAGEBREAK`. A fourth optional XML file is a file containing metadata for the RadCon project, the purpose is `AUXILIARY`. See the example in Fig. 6 (p. 140).

journal-item/files-info/ml/dtd-version

The version of the DTD and the top-level element (doctype) used to capture the item is contained in `dtd-version`. This must, of course, be identical to the declaration in the XML file.

This is not the case for RadCon metadata files. These files are structured according to a MicroSoft Infopath schema and the dtd-version is `INFOPATH`. See the example in Fig. 6 (p. 140).

journal-item/files-info/ml/weight

The weight of the XML file can be `FULL-TEXT`, `HEAD-AND-TAIL`, `HEAD-ONLY` or `CONTENTS-`

ENTRY-ONLY. The weight of the item indicates which parts of the text are captured in XML.

journal-item/files-info/ml/asset

All assets belonging to the item are listed under asset. This includes all external files declared in the XML file; it does not include any strip-ins or fingerprints.

journal-item/files-info/ml/asset/pathname

In case journal-item's attribute omitted is not present or has the value false, pathname is the pathname, relative to dataset.xml, of the XML file.

In case journal-item's attribute omitted has the value true, pathname is the pathname of the XML file in the system it resides in (an absolute pathname starting with a server name).

journal-item/files-info/ml/asset/type

The type of the asset can have the values APPLICATION, IMAGE-CAP, IMAGE-NONCAP, IMAGE-DOWNSAMPLED, IMAGE-THUMBNAIL, IMAGE-MMC, IMAGE-MMC-DOWNSAMPLED, IMAGE-MMC-THUMBNAIL, AUDIO, VIDEO, VIDEO-FLASH, XML. These types are defined in Section 3.3.

journal-item/files-info/web-pdf

An S300 item has one or two web PDF files associated with it. Although the element is optional it may only be absent if the project allows for it.

journal-item/files-info/web-pdf/pathname

In case journal-item's attribute omitted is not present or has the value false, pathname is the pathname, relative to dataset.xml, of the XML file.

In case journal-item's attribute omitted has the value true, pathname is the pathname of the XML file in the system it resides in (an absolute pathname starting with a server name).

journal-item/files-info/web-pdf/purpose

purpose indicates what the web PDF file is for. There may not be any PDF files with the same purpose.

There must always be one PDF file with purpose MAIN. This contains the PDF file of the item.

The other optional web PDF file has value MAIN-ABRIDGED (contains the abridged print version).

journal-item/files-info/web-pdf/pdf-version

pdf-version is the version of the web-PDF file, the allowed values are 1.7 6.4, 1.7 6.5, and 1.7 7.0.

journal-item/files-info/web-pdf/pdf-property

The element `pdf-property` describes the nature of the web PDF file and has one of the following values: `WRAPPED OPTIMIZED`, `DISTILLED OPTIMIZED`, `DISTILLED OPTIMIZED BOOKMARKED`.

6.18. S350-H350 deliveries

An H350 issue is a deliverable of the PreCAP process. In the PreCAP process, printed issues are scanned and by means of OCR technology, datasets are created. S350 items are items in H350 issues.

The S350-H350 dataset directory structure is defined in Section 4.1.9.

Below, we traverse the S350-H350 schema for serial publications starting from the element `dataset-content`. The elements `dataset-unique-ids` and `dataset-properties` are explained in Section 4.2.

An S350-H350 schema contains one or more issues. Each issue must always be accompanied with all its items. The items, whose details are contained in `journal-item`, belong to the `journal-issue` that immediately precedes them.

journal-issue

The element `journal-issue` contains all the information pertaining to the issue in the dataset. The files associated with the issue are the issue hub with possible assets.

journal-issue/version

journal-issue/version/version-number

The element `version-number` contains the version number of the issue, as described in Section 3.2.

journal-issue/version/stage

The value of `stage` is `H350`.

journal-issue/journal-issue-unique-ids

journal-issue/journal-issue-unique-ids/pii

The element `pii` contains the PII of the issue. In the case of book series this is a book hub PII or a collection PII. It must be equal to the PII in the XML hub file.

journal-issue/journal-issue-unique-ids/doi

The element `doi` contains the DOI of the item, if any. In the case of book series the DOI may be based on a book hub PII or a collection PII. Only items that will appear online may have a DOI. It must be equal to the DOI in the XML hub file.

journal-issue/journal-issue-properties

journal-issue/journal-issue-properties/jid

The element `jid` contains the Elsevier system code of the serial publication to which the issue belongs. Book series JIDs should begin with “BS:”.

journal-issue/journal-issue-properties/issn

The element `issn` contains the ISSN of the serial publication to which the issue belongs.

journal-issue/journal-issue-properties/volume-issue-number

The volume/issue number of the issue is captured in `volume-issue-number`. It consists of `vol-first`, `vol-last`, `iss-first`, `iss-last` and `suppl`.

journal-issue/journal-issue-properties/embargo

The optional element `embargo` contains a date/time before which the issue is not allowed to be published. Its content is in W3C schema DateTime format, `xs:dateTime`, and should be accurate to the minute. To avoid confusion the date/time should be in UTC, signified by a "Z". That is, the embargo date/time should be in the following format: "*yyyy-mm-dd*T*hh*:*mm*:00Z".

```
<embargo>2006-04-11T12:00:00Z</embargo>
```

journal-issue/journal-issue-properties/isbn

The issue's optional ISBN, an ISBN-13, is captured with `isbn`.

journal-issue/files-info

The element `files-info` contains all the information needed to process the files belonging to the issue.

journal-issue/files-info/ml

The issue hub is an XML file associated with the issue. Its details are listed under `ml`.

journal-issue/files-info/ml/pathname

`pathname` is the pathname, relative to dataset.xml (that is, relative to the base directory), of the hub XML file.

journal-issue/files-info/ml/purpose

The element `purpose` always contains the value `MAIN`.

journal-issue/files-info/ml/dtd-version

The version of the DTD of the hub file is contained in `dtd-version`. The version must, of course, be identical to the declaration in the XML file.

journal-issue/files-info/ml/asset

All assets belonging to the item are listed under `asset`. This includes all external files declared in the XML file; it does not include any strip-ins or fingerprints.

journal-issue/files-info/ml/asset/pathname

`pathname` is the pathname, relative to dataset.xml, of the asset file.

journal-issue/files-info/ml/asset/type

The `type` of the asset can have the values `IMAGE-CAP`, `IMAGE-COVER`. These types are defined in Section 3.3.

journal-issue/files-info/web-pdf

Beside the hub XML file, there can be an issue-level PDF file which contains Tables of Contents. The PDF files are untyped, i.e. there is no subelement `purpose`. Element `web-pdf` is optional.

journal-issue/files-info/web-pdf/pathname

`pathname` is the pathname, relative to dataset.xml, of the web PDF file.

journal-issue/files-info/web-pdf/pdf-version

`pdf-version` is the version of the web-PDF file, the allowed values are `1.7 6.4`, `1.7 6.5`, and `1.7 7.0`.

journal-item

The element `journal-item` contains all the information pertaining to an item in a serial publication. It can have three attributes, `type`, `omitted` and `cross-mark`.

The first one, `type`, takes the values `stand-alone` (default), `with-add-ons` and `batch-placeholder`. In Section 3.4 (p. 29) an explanation is given when `type` is used.

The second one, `omitted`, takes the values `false` (default) and `true`. The latter value is to be used if an item is omitted from the dataset. It should only be used if the `dataset-action` has value `PARTIAL-RELOAD`. (See also p. 41 for information on partial deliveries.)

The third one, `cross-mark`, takes the values `false` and `true`.

journal-item/version/version-number

The element `version-number` contains the version number of the item, as described in Section 3.2. The version number is assigned by Elsevier and is included in the order.

journal-item/version/stage

The value of `stage` is `S350`.

journal-item/journal-item-unique-ids

journal-item/journal-item-unique-ids/pii

The element `pii` contains the PII of the item. In the case of book series this is a book PII. It must be equal to the PII in the XML file and the PII in the document properties of the PDF file (if appropriate).

journal-item/journal-item-unique-ids/doi

The element `doi` contains the DOI of the item, if any. In the case of book series the DOI may be based on a book PII. Only items that will appear online may have a DOI. It must

be equal to the DOI in the XML file and the DOI in the document properties of the PDF file (if appropriate).

journal-item/journal-item-unique-ids/jid-aid

The element `jid-aid` contains up to four subelements: `jid` is the Elsevier system code and `issn` the (formatted) ISSN of the serial publication to which the item belongs; `aid` is the Elsevier system article ID and `article-number` is the item's article number.

The first three elements are mandatory while the last one is optional. The `jid`, `aid` and `article-number` must be equal to the same elements in the item's XML file(s). Book series JIDs should begin with "BS:".

journal-item/journal-item-properties

journal-item/journal-item-properties/pit

The element `pit` contains the publication item type of the item. The value of `pit` must be identical to the value of the top-level attribute `docsubtype` in the XML file, except for the case: PITs are always written uppercase in dataset.xml files and lowercase in DTD 5.x files.)

journal-item/journal-item-properties/production-type

The value of `production-type` is `CRC`.

journal-item/journal-item-properties/batch

If a journal item is part of a batch, then element `batch` must be present. It contains `pii`, the PII of the journal item that represents the batch, and `doi`, its optional DOI. Note that the item that is referred to, itself an independent item, has the attribute `type` set to a non-default value (see Section 3.4, p. 29).

journal-item/files-info

The element `files-info` contains all the information needed to process the files belonging to the item.

journal-item/files-info/ml

One or two XML files can be associated with the item. These are listed under `ml`.

journal-item/files-info/ml/pathname

In case `journal-item`'s attribute `omitted` is not present or has the value `false`, `pathname` is the pathname, relative to dataset.xml, of the XML file.

In case `journal-item`'s attribute `omitted` has the value `true`, `pathname` is the pathname of the XML file in the system it resides in (an absolute pathname starting with a server name).

journal-item/files-info/ml/purpose

purpose indicates what the XML file is for. There must always be one XML file with purpose MAIN. The optional second XML file is a changes-with-respect-to file with purpose CHANGES. A third optional XML file is a file containing pagebreak information, the purpose is PAGEBREAK.

journal-item/files-info/ml/dtd-version

The version of the DTD and the top-level element (doctype) used to capture the item is contained in dtd-version. This must, of course, be identical to the declaration in the XML file.

journal-item/files-info/ml/weight

The weight of the XML file can be FULL-TEXT, HEAD-AND-TAIL, HEAD-ONLY or CONTENTS-ENTRY-ONLY. The weight of the item indicates which portions of the text are captured in XML.

journal-item/files-info/ml/asset

All assets belonging to the item are listed under asset. This includes all external files declared in the XML file; it does not include any strip-ins or fingerprints.

journal-item/files-info/ml/asset/pathname

In case journal-item's attribute omitted is not present or has the value false, pathname is the pathname, relative to dataset.xml, of the XML file.

In case journal-item's attribute omitted has the value true, pathname is the pathname of the XML file in the system it resides in (an absolute pathname starting with a server name).

journal-item/files-info/ml/asset/type

The type of the asset can have the values APPLICATION, IMAGE-CAP, AUDIO, VIDEO, VIDEO-FLASH, XML. These types are defined in Section 3.3.

journal-item/files-info/web-pdf

An S350 item has exactly one web PDF file associated with it, the main PDF file which contains the item.

journal-item/files-info/web-pdf/pathname

In case journal-item's attribute omitted is not present or has the value false, pathname is the pathname, relative to dataset.xml, of the XML file.

In case journal-item's attribute omitted has the value true, pathname is the pathname of the XML file in the system it resides in (an absolute pathname starting with a server name).

journal-item/files-info/web-pdf/purpose

purpose indicates what the web PDF file is for, it must be MAIN.

journal-item/files-info/web-pdf/pdf-version

`pdf-version` is the version of the web-PDF file, the only values currently allowed are `1.2`, `1.3`, `1.4`, `1.4 6.0`, `1.7 6.2`, `1.7 6.3`, `1.7 6.4`, `1.7 6.5`, and `1.7 7.0`.

journal-item/files-info/web-pdf/pdf-property

The element `pdf-property` describes the nature of the web PDF file and can only have value `WRAPPED OPTIMIZED`.

journal-item/files-info/raw-text

An S350 item also possesses a raw text manifestation, which is obtained by optical character recognition from the scanned PDF file.

journal-item/files-info/raw-text/pathname

In case `journal-item`'s attribute `omitted` is not present or has the value `false`, `pathname` is the pathname, relative to dataset.xml, of the XML file.

In case `journal-item`'s attribute `omitted` has the value `true`, `pathname` is the pathname of the XML file in the system it resides in (an absolute pathname starting with a server name).

6.19. S350-H350 Project deliveries

The S350-H350-Project schema supports deliveries of material for special projects. It is based on the regular S350-H350 schema. Some pattern validation has been removed. This validation will be performed elsewhere and is not described below. In fact, below is assumed that this validation does take place. Additionally some lists of values have been expanded.

An H350 issue is a deliverable of the PreCAP process. In the PreCAP process, printed issues are scanned and by means of OCR technology, datasets are created. S350 items are items in H350 issues.

The S350-H350 dataset directory structure is defined in Section 4.1.9.

Below, we traverse the S350-H350-Project schema for serial publications starting from the element `dataset-content`. The elements `dataset-unique-ids` and `dataset-properties` are explained in Section 4.2.

An S350-H350-Project schema contains one or more issues. Each issue must always be accompanied with all its items. The items, whose details are contained in `journal-item`, belong to the `journal-issue` that immediately precedes them.

journal-issue

The element `journal-issue` contains all the information pertaining to the issue in the dataset. The files associated with the issue are the issue hub with possible assets.

journal-issue/version

journal-issue/version/version-number

The element `version-number` contains the version number of the issue, as described in Section 3.2.

journal-issue/version/stage

The value of `stage` is `H350`.

journal-issue/journal-issue-unique-ids

journal-issue/journal-issue-unique-ids/pii

The element `pii` contains the PII of the issue. In the case of book series this is a book hub PII or a collection PII. It must be equal to the PII in the XML hub file.

journal-issue/journal-issue-unique-ids/doi

The element `doi` contains the DOI of the item, if any. In the case of book series the DOI may be based on a book hub PII or a collection PII. Only items that will appear online may have a DOI. It must be equal to the DOI in the XML hub file.

journal-issue/journal-issue-properties

journal-issue/journal-issue-properties/jid

The element `jid` contains the Elsevier system code of the serial publication to which the issue belongs. Book series JIDs should begin with "BS:".

journal-issue/journal-issue-properties/issn

The element `issn` contains the ISSN of the serial publication to which the issue belongs.

journal-issue/journal-issue-properties/volume-issue-number

The volume/issue number of the issue is captured in `volume-issue-number`. It consists of `vol-first`, `vol-last`, `iss-first`, `iss-last` and `suppl`.

journal-issue/journal-issue-properties/embargo

The optional element `embargo` contains a date/time before which the issue is not allowed to be published. Its content is in W3C schema DateTime format, `xs:dateTime`, and should be accurate to the minute. To avoid confusion the date/time should be in UTC, signified by a "Z". That is, the embargo date/time should be in the following format: "*yyyy-mm-dd*T*hh*:*mm*:00Z".

```
<embargo>2006-04-11T12:00:00Z</embargo>
```

journal-issue/journal-issue-properties/isbn

The issue's optional ISBN, an ISBN-10 or an ISBN-13, is captured with `isbn`.

journal-issue/files-info

The element `files-info` contains all the information needed to process the files belonging to the issue.

journal-issue/files-info/ml

The issue hub is an XML file associated with the issue. Its details are listed under `ml`.

journal-issue/files-info/ml/pathname

`pathname` is the pathname, relative to dataset.xml (that is, relative to the base directory), of the hub XML file.

journal-issue/files-info/ml/purpose

The element `purpose` always contains the value `MAIN`.

journal-issue/files-info/ml/dtd-version

The version of the DTD of the hub file is contained in `dtd-version`. The version must, of course, be identical to the declaration in the XML file.

journal-issue/files-info/ml/asset

All assets belonging to the item are listed under `asset`. This includes all external files declared in the XML file; it does not include any strip-ins or fingerprints.

journal-issue/files-info/ml/asset/pathname

`pathname` is the pathname, relative to dataset.xml, of the asset file.

journal-issue/files-info/ml/asset/type

The `type` of the asset can have the values `IMAGE-CAP`, `IMAGE-COVER`. These types are defined in Section 3.3.

journal-issue/files-info/web-pdf

Beside the hub XML file, there can be an issue-level PDF file which contains Tables of Contents. The PDF files are untyped, i.e. there is no subelement `purpose`. Element `web-pdf` is optional.

journal-issue/files-info/web-pdf/pathname

`pathname` is the pathname, relative to dataset.xml, of the web PDF file.

journal-issue/files-info/web-pdf/pdf-version

`pdf-version` is the version of the web-PDF file, the allowed values are `1.7 6.4`, `1.7 6.5`, and `1.7 7.0`.

journal-item

The element `journal-item` contains all the information pertaining to an item in a serial publication. It can have three attributes, `type`, `omitted` and `cross-mark`.

The first one, `type`, takes the values `stand-alone` (default), `with-add-ons` and `batch-placeholder`. In Section 3.4 (p. 29) an explanation is given when `type` is used.

The second one, `omitted`, takes the values `false` (default) and `true`. The latter value is to be used if an item is omitted from the dataset. It should only be used if the `dataset-action` has value `PARTIAL-RELOAD`. (See also p. 41 for information on partial deliveries.)

The third one, `cross-mark`, takes the values `false` and `true`.

journal-item/version

journal-item/version/version-number

The element `version-number` contains the version number of the item, as described in Section 3.2. The version number is assigned by Elsevier and is included in the order.

journal-item/version/stage

The value of `stage` is `S350`.

journal-item/journal-item-unique-ids

journal-item/journal-item-unique-ids/pii

The element pii contains the PII of the item. In the case of book series this is a book PII. It must be equal to the PII in the XML file and the PII in the document properties of the PDF file (if appropriate).

journal-item/journal-item-unique-ids/doi

The element doi contains the DOI of the item, if any. In the case of book series the DOI may be based on a book PII. Only items that will appear online may have a DOI. It must be equal to the DOI in the XML file and the DOI in the document properties of the PDF file (if appropriate).

journal-item/journal-item-unique-ids/jid-aid

The element jid-aid contains up to four subelements: jid is the Elsevier system code and issn the (formatted) ISSN of the serial publication to which the item belongs; aid is the Elsevier system article ID and article-number is the item's article number.

The first three elements are mandatory while the last one is optional. The jid, aid and article-number must be equal to the same elements in the item's XML file(s). Book series JIDs should begin with "BS:".

journal-item/journal-item-properties

journal-item/journal-item-properties/pit

The element pit contains the publication item type of the item. The value of pit must be identical to the value of the top-level attribute docsubtype in the XML file, except for the case: PITs are always written uppercase in dataset.xml files and lowercase in DTD 5.x files.)

journal-item/journal-item-properties/production-type

The value of production-type is CRC.

journal-item/journal-item-properties/batch

If a journal item is part of a batch, then element batch must be present. It contains pii, the PII of the journal item that represents the batch, and doi, its optional DOI. Note that the item that is referred to, itself an independent item, has the attribute type set to a non-default value (see Section 3.4, p. 29).

journal-item/files-info

The element files-info contains all the information needed to process the files belonging to the item.

journal-item/files-info/ml

One or two XML files can be associated with the item. These are listed under ml.

journal-item/files-info/ml/pathname

In case journal-item's attribute omitted is not present or has the value false, pathname is the pathname, relative to dataset.xml, of the XML file.

In case journal-item's attribute omitted has the value true, pathname is the pathname of the XML file in the system it resides in (an absolute pathname starting with a server name).

journal-item/files-info/ml/purpose

purpose indicates what the XML file is for. There must always be one XML file with purpose MAIN. An optional second XML file is a changes-with-respect-to file with purpose CHANGES. A third optional XML file is a file containing pagebreak information, the purpose is PAGEBREAK. A fourth optional XML file is a file containing metadata for the RadCon project, the purpose is AUXILIARY. See the example in Fig. 6 (p. 140).

journal-item/files-info/ml/dtd-version

The version of the DTD and the top-level element (doctype) used to capture the item is contained in dtd-version. This must, of course, be identical to the declaration in the XML file.

This is not the case for RadCon metadata files. These files are structured according to a MicroSoft Infopath schema and the dtd-version is INFOPATH. See the example in Fig. 6 (p. 140).

journal-item/files-info/ml/weight

The weight of the XML file can be FULL-TEXT, HEAD-AND-TAIL, HEAD-ONLY or CONTENTS-ENTRY-ONLY. The weight of the item indicates which portions of the text are captured in XML.

journal-item/files-info/ml/asset

All assets belonging to the item are listed under asset. This includes all external files declared in the XML file; it does not include any strip-ins or fingerprints.

journal-item/files-info/ml/asset/pathname

In case journal-item's attribute omitted is not present or has the value false, pathname is the pathname, relative to dataset.xml, of the XML file.

In case journal-item's attribute omitted has the value true, pathname is the pathname of the XML file in the system it resides in (an absolute pathname starting with a server name).

journal-item/files-info/ml/asset/type

The type of the asset can have the values APPLICATION, IMAGE-CAP, IMAGE-NONCAP, IMAGE-DOWNSAMPLED, IMAGE-THUMBNAIL, IMAGE-MMC, IMAGE-MMC-DOWNSAMPLED, IMAGE-MMC-THUMBNAIL, AUDIO, VIDEO, VIDEO-FLASH, XML.

journal-item/files-info/web-pdf

An S350 item has exactly one web PDF file associated with it, the main PDF file which contains the item. Although the element is optional it may only be absent if the project allows for it.

journal-item/files-info/web-pdf/pathname

In case `journal-item`'s attribute `omitted` is not present or has the value `false`, `pathname` is the pathname, relative to dataset.xml, of the XML file.

In case `journal-item`'s attribute `omitted` has the value `true`, `pathname` is the pathname of the XML file in the system it resides in (an absolute pathname starting with a server name).

journal-item/files-info/web-pdf/purpose

`purpose` indicates what the web PDF file is for, it must be `MAIN`.

journal-item/files-info/web-pdf/pdf-version

`pdf-version` is the version of the web-PDF file, the allowed values are `1.7 6.4`, `1.7 6.5`, and `1.7 7.0`.

journal-item/files-info/web-pdf/pdf-property

The element `pdf-property` describes the nature of the web PDF file and can only have value `WRAPPED OPTIMIZED`.

journal-item/files-info/raw-text

An S350 item also possesses a raw text manifestation, which is obtained by optical character recognition from the scanned PDF file.

journal-item/files-info/raw-text/pathname

In case `journal-item`'s attribute `omitted` is not present or has the value `false`, `pathname` is the pathname, relative to dataset.xml, of the XML file.

In case `journal-item`'s attribute `omitted` has the value `true`, `pathname` is the pathname of the XML file in the system it resides in (an absolute pathname starting with a server name).

6.20. Print/bind orders

This section describes the print/bind order and the issue processing form.

6.20.1. Print/bind order

order/print-bind-info

This element contains information for the printer and binder in subelements `general-info`, `print-details`, `issue-content` and `issue-remarks` (optional).

order/print-bind-info/general-info

See the description of `order/issue-info/general-info` (p. 99).

order/print-bind-info/general-info/buffer-status

This empty element has a mandatory attribute `status` with values `yes` and `no`. An issue with `status yes` should not be despatched to the warehouse until an instruction from the relevant production site is received, or the `status` is changed to `no`. The PTSIII stylesheet adds this instruction to the rendering of the order.

order/print-bind-info/general-info/hold-until-date

An issue may only be sent to the warehouse after this date (in subelement `date`). The PTSIII stylesheet adds this instruction to the rendering of the order.

order/print-bind-info/print-details

This element contains issue information necessary for the printer and/or binder in the following subelements: `no-copies`, `no-author-copies`, `no-mark-prom-copies`, `no-voucher-copies`, `no-grace-copies`, `no-extra-copies`, `total-print-run`, `no-pages-cover`, `no-pages-coated`. `split-print-run` (optional) and `hold-despatch-date` (optional).

```
<print-details>
  <no-copies>430</no-copies>
  <no-author-copies>0</no-author-copies>
  <no-mark-prom-copies>200</no-mark-prom-copies>
  <no-voucher-copies>0</no-voucher-copies>
  <no-grace-copies>0</no-grace-copies>
  <no-extra-copies>250</no-extra-copies>
  <total-print-run>880</total-print-run>
  <no-pages-cover>4</no-pages-cover>
  <no-pages-coated>0</no-pages-coated>
</print-details>
```

order/print-bind-info/print-details/no-copies

This element contains the number of copies to be printed. This is the standard print run.

order/print-bind-info/print-details/no-author-copies

This element contains the number of extra “author” copies to be printed.

order/print-bind-info/print-details/no-mark-prom-copies

This element contains the number of extra marketing promotion copies to be printed.

order/print-bind-info/print-details/no-voucher-copies

This element contains the number of extra voucher copies to be printed.

order/print-bind-info/print-details/no-grace-copies

This element contains the number of extra “grace” copies to be printed.

order/print-bind-info/print-details/no-extra-copies

This element contains the number of other extra copies to be printed.

order/print-bind-info/print-details/total-print-run

This element contains the total number of copies to be printed. It is the sum of the six previous numbers.

order/print-bind-info/print-details/no-pages-cover

This element contains the number of cover pages.

order/print-bind-info/print-details/no-pages-coated

This element contains the number of pages coated different than the regular pages (e.g. containing colour figures).

order/print-bind-info/print-details/split-print-run

The element `split-print-run` is an empty element. It has a mandatory attribute `split` with possible values `yes` and `no`, indicating if the print run is split or not.

order/print-bind-info/print-details/hold-despatch-date

This element is not used; instead the element `hold-until-date` (p. 155) is used.

order/print-bind-info/issue-content

See the description of `order/issue-info/issue-content` (p. 106).

order/print-bind-info/issue-remarks

This optional element contains remarks on the issue in one or more subelements `issue-remark`. For more information see the description of element `order/item-info/item-remarks` (p. 59).

6.20.2. Issue labels order

order/issue-labels-info

This element is needed to trigger the production of issue despatch labels. It contains information in subelements `general-info` (see p. 99), `print-details` (see p. 155), `issue-label-details` and `issue-remarks` (optional, see p. 59).

order/issue-labels-info/issue-labels-details

This element contains in subelement `weight` the weight of the issue (in grams).

```
<issue-labels-details>
  <issue-weight>220</issue-weight>
</issue-labels-details>
```

6.21. F300 deliveries

An F300 dataset is a dataset containing "fat PDF" files, suitable for printing an entire issue.

journal-issue

The element `journal-issue` contains all the information pertaining to the issue in the dataset, as well as the locations within the dataset of the fat PDF files.

journal-issue/version

journal-issue/version/version-number

The element `version-number` contains the version number of the issue, as described in Section 3.2. The version number is assigned by Elsevier and is included in the order.

journal-issue/version/stage

The value of `stage` is `F300`.

journal-issue/journal-issue-unique-ids

journal-issue/journal-issue-unique-ids/pii

The element `pii` contains the PII of the issue. In the case of book series this is a book hub PII or a collection PII.

journal-issue/journal-issue-unique-ids/doi

The element `doi` contains the DOI of the issue, if any. In the case of book series the DOI may be based on a book hub PII or a collection PII. Only issues that will appear online may have a DOI.

journal-issue/journal-issue-properties

journal-issue/journal-issue-properties/jid

The element `jid` contains the Elsevier system code of the serial publication to which the issue belongs. Book series JIDs should begin with "BS:".

journal-issue/journal-issue-properties/issn

The element `issn` contains the ISSN of the serial publication to which the issue belongs.

journal-issue/journal-issue-properties/volume-issue-number

The volume/issue number of the issue is captured in `volume-issue-number`. It consists of `vol-first`, `vol-last`, `iss-first`, `iss-last` and `suppl`. A spin-off issue is possible in an F300 dataset.

journal-issue/journal-issue-properties/embargo

The optional element `embargo` contains a date/time before which the issue is not allowed to be published. Its content is in W3C schema DateTime format, `xs:dateTime`, and should be accurate to the minute. To avoid confusion the date/time should be in UTC, signified by a "Z". That is, the embargo date/time should be in the following format: "*yyyy-mm-dd*T*hh*:*mm*:00Z".

```
<embargo>2006-04-11T12:00:00Z</embargo>
```

journal-issue/journal-issue-properties/isbn

The issue's optional ISBN, an ISBN-13, is captured with `isbn`.

journal-issue/files-info

The element `files-info` contains all the information needed to process the files belonging to the issue.

journal-issue/files-info/ml

The XML file in the F300 dataset provides a full description to the printer about how to print the issue. It explains all the fat (print) PDF files in the dataset.

journal-issue/files-info/ml/pathname

`pathname` is the pathname, relative to dataset.xml (that is, relative to the base directory), of the print description XML file.

journal-issue/files-info/ml/schema-version

The version of the W3C schema of the print description file is contained in `schema-version`. The allowed values are `print 1.0`, `print 1.1`, `print 1.2`, `print 1.3` and `print 1.4`. The version must, of course, be identical to the declaration in the XML file.

journal-issue/files-info/print-pdf

Print PDF files, also known as "fat PDF" files, are used by the printer to print the issue.

journal-issue/files-info/print-pdf/pathname

`pathname` is the pathname, relative to dataset.xml, of the print PDF file.

journal-issue/files-info/print-pdf/purpose

`purpose` indicates the suitability of the PDF file for printing and/or digital use. It has three possible values: `SUITABILITY-OFFSET-AND-DIGITAL` if the PDF file is suitable for printing and for digital use, `SUITABILITY-DIGITAL` if the PDF file is suitable for digital use, while `SUITABILITY-NONE` is used when no statement on suitability can be made.

journal-issue/files-info/print-pdf/pdf-version

`pdf-version` is the version of the print-PDF file, the allowed values are `1.3 1.0`, and `1.6 2.0`.

journal-issue/files-info/print-pdf/pdf-property

`pdf-property` indicates how the PDF file is created. It has two possible values: `SCANNED` if the PDF file is the result of scanning, and `NOT SCANNED` if the PDF file is created in another way.

Chapter 7

Book projects and book items

This chapter describes all the elements of the transport schema for book projects and book items.

7.1. S200 deliveries

An S200 item in a chapter publication is a corrected proof, authorized by the authors of the book.

The S200 dataset directory structure is defined in Section 4.1.8.

Below, we traverse the S200 schema for chapter publications starting from the element `dataset-content`. The elements `dataset-unique-ids` and `dataset-properties` are explained in Section 4.2. The element `dataset-content` contains a list of `book-item` subelements.

book-item

The element book-item contains all of the metadata elements needed to uniquely identify the book item as well as the location of the book item's asset files in the dataset. It contains the following elements: `version`, `book-item-unique-ids`, `book-item-properties`, and `files-info`.

It can have one attribute, `omitted`, which takes values `false` (default) and `true`. The latter value is to be used if an item is omitted from the dataset. It should only be used if the `dataset-action` has value `PARTIAL-RELOAD`. (See also p. 41 for information on partial deliveries.)

Sample tagging of `book-item` can be found in Fig. 11 (p. 180).

book-item/version

book-item/version/version-number

The element `version-number` contains the version number of the item, as described in Section 3.2. The version number is assigned by Elsevier.

book-item/version/stage

The value of `stage` is `S200`.

book-item/book-item-unique-ids

The element `book-item-unique-ids` contains a required `pii` element, with an optional `doi` element. Each should be properly formatted with dashes, slashes and parentheses and should be identical to the same identifiers in the XML and PDF files. Additionally it contains required element `collection-item-id`.

book-item/book-item-unique-ids/collection-item-id

The element `collection-item-id` contains element `isbn` or element `issn` optionally followed by `isbn` containing the ISSN and ISBN of the book project or book series. It has an optional element `collection-id` containing the ID of the collection the item belongs to and an optional `item-id` element containing the ID of the item. The latter two elements are the counterparts of elements `jid` and `aid` in the journal schemas. The element `collection-id` is only to be used for book series.

book-item/book-item-unique-ids/collection-item-id/item-id

The element `item-id` contains the ID of the item. It may not be more than 8 characters long.

book-item/book-item-properties

The element `book-item-properties` contains three mandatory elements: `branch`, `pit`, and `production-type`.

book-item/book-item-properties/branch

The `branch` element is used to identify in which portion of the book hierarchy the item occurs. There are three possible values: `FRONT`, `BODY` or `REAR`.

book-item/book-item-properties/pit

The `pit` element is a list of different types of items which can appear in books, similar to journal Publishing Item Types. The list of possible values is contained in the "pit-list" in the schema `book-item-project.xsd`. The value of `pit` must be identical to the value of the top-level attribute `docsubtype` in the XML file. (Except for the case, that is. PITs are always written uppercase in dataset.xml files and lowercase in DTD 5.x files.)

book-item/book-item-properties/production-type

The `production-type` element contains one of two possible values: `CRC` or `NON-CRC`.

The value `WRAPPED OPTIMIZED` is only allowed if the book item's production type is `CRC`.

If the purpose of the book item is not `MAIN`, then the value `DISTILLED OPTIMIZED` must be used.

book-item/files-info

The element `files-info` contains all the information needed to process the files belonging to the item.

book-item/files-info/ml

One or two XML files can be associated with the item. These are listed under `ml`.

book-item/files-info/ml/pathname

In case `book-item`'s attribute `omitted` is not present or has the value `false`, `pathname` is the pathname, relative to dataset.xml, of the XML file.

In case `book-item`'s attribute `omitted` has the value `true`, `pathname` is the pathname of the XML file in the system it resides in (an absolute pathname starting with a server name).

book-item/files-info/ml/purpose

The `purpose` element contains the purpose of the item. There must be always one XML file with purpose `MAIN`. An optional second XML file is a changes-with-respect-to file with

purpose CHANGES. A third optional XML file is a file containing pagebreak information, the purpose is PAGEBREAK.

book-item/files-info/ml/dtd-version

The version of the DTD and the top-level element (doctype) used to capture the item is contained in dtd-version. This must, of course, be identical to the declaration in the XML file.

book-item/files-info/ml/weight

The weight of the XML file can be FULL-TEXT, HEAD-AND-TAIL or CONTENTS-ENTRY-ONLY. The weight of the item indicates which parts of the text are captured in XML.

book-item/files-info/ml/asset

All assets belonging to the item are listed under asset. This includes all external files declared in the XML file, it does not include any strip-ins or fingerprints.

book-item/files-info/ml/asset/pathname

In case book-item's attribute omitted is not present or has the value false, pathname is the pathname, relative to dataset.xml, of the XML file.

In case book-item's attribute omitted has the value true, pathname is the pathname of the XML file in the system it resides in (an absolute pathname starting with a server name).

book-item/files-info/ml/asset/type

The type of the asset can have the values APPLICATION, IMAGE-CAP, IMAGE-MMC, AUDIO, VIDEO, VIDEO-FLASH, XML. These types are defined in Section 3.3.

book-item/files-info/web-pdf

An S300 item has exactly one web PDF file associated with it, the main PDF file which contains the item.

book-item/files-info/web-pdf/pathname

In case book-item's attribute omitted is not present or has the value false, pathname is the pathname, relative to dataset.xml, of the XML file.

In case book-item's attribute omitted has the value true, pathname is the pathname of the XML file in the system it resides in (an absolute pathname starting with a server name).

book-item/files-info/web-pdf/purpose

purpose indicates what the web PDF file is for, it must be MAIN.

book-item/files-info/web-pdf/pdf-version

```
<book-item>
  <version>
    <version-number>S200.1</version-number>
    <stage>S200</stage>
  </version>
  <book-item-unique-ids>
    <pii>B0-323-01195-0/00002-X</pii>
  </book-item-unique-ids>
  <book-item-properties>
    <branch>BODY</branch>
    <pit>CHP</pit>
    <production-type>NON-CRC</production-type>
  </book-item-properties>
  <files-info>
    <ml>
      <pathname>0323011950/body/B032301195000002X/main.xml</pathname>
      <purpose>MAIN</purpose>
      <dtd-version>EHS-BOOKS 5.1.1 CHAPTER</dtd-version>
      <weight>FULL-TEXT</weight>
      <asset>
        <pathname>0323011950/body/B032301195000002X/
          main.assets/gr1.jpg</pathname>
        <type>IMAGE-CAP</type>
      <asset>
    </ml>
    <web-pdf>
      <pathname>0323011950/body/B032301195000002X/main.pdf</pathname>
      <purpose>MAIN</purpose>
      <pdf-version>1.4 6.0</pdf-version>
      <pdf-property>DISTILLED OPTIMIZED BOOKMARKED</pdf-property>
    </web-pdf>
  </files-info>
</book-item>
```

Figure 7: Example tagging of a book-item.

`pdf-version` is the version of the web-PDF file, the allowed values are `1.7 6.4`, `1.7 6.5`, and `1.7 7.0`.

book-item/files-info/web-pdf/pdf-property

The element `pdf-property` describes the nature of the web PDF file and has one of the following values: `WRAPPED OPTIMIZED`, `DISTILLED OPTIMIZED`, `DISTILLED OPTIMIZED BOOKMARKED`.

7.2. S280 deliveries

An S280 book item is introduced to support the delivery of book/module items without a hub.

Below, we traverse the S280 schema for books starting from the element `dataset-content`. The elements `dataset-unique-ids` and `dataset-properties` are explained in Section 4.2.

The element `dataset-content` is the container element for all of the metadata covering all of the items in the book dataset delivery. It consists of one or more book-item elements belonging to each book being delivered.

book-item

The element book-item contains all of the metadata elements needed to uniquely identify the book item as well as the location of the book item's asset files in the dataset. It contains the following elements: `version`, `book-item-unique-ids`, `book-item-properties`, and `files-info`.

It can have one attribute, `omitted`, which takes values `false` (default) and `true`. The latter value is to be used if an item is omitted from the dataset. It should only be used if the `dataset-action` has value `PARTIAL-RELOAD`. (See also p. 41 for information on partial deliveries.)

book-item/version

book-item/version/version-number

The element `version-number` contains the version number of the item, as described in Section 3.2. The version number is assigned by Elsevier.

book-item/version/stage

The value of `stage` is `S280`.

book-item/book-item-unique-ids

The element `book-item-unique-ids` contains a required `pii` element, with an optional `doi` element. Each should be properly formatted with dashes, slashes and parentheses and should be identical to the same identifiers in the XML and PDF files. Additionally it contains required element `collection-item-id`.

book-item/book-item-unique-ids/collection-item-id

The element `collection-item-id` contains element `pii` to hold the collection PII, followed by element `isbn` or element `issn` optionally followed by `isbn` containing the ISSN and ISBN of the book project or book series. It has an optional element `collection-id` containing the ID of the collection the item belongs to and an optional `item-id` element containing the ID of the item. The latter two elements are the counterparts of elements `jid` and `aid` in the journal schemas.

book-item/book-item-unique-ids/collection-item-id/item-id

The element `item-id` contains the ID of the item. It may not be more than 8 characters long.

book-item/book-item-properties

The element `book-item-properties` contains two mandatory elements: `pit`, and `production-type`.

book-item/book-item-properties/pit

The `pit` element is a list of different types of items which can appear in books, similar to journal Publishing Item Types. The list of possible values is contained in the "pit-list" in the schema `book-item-project.xsd`. The value of `pit` must be identical to the value of the top-level attribute `docsubtype` in the XML file. (Except for the case, that is. PITs are always written uppercase in dataset.xml files and lowercase in DTD 5.x files.)

book-item/book-item-properties/production-type

The `production-type` element contains one of two possible values: `CRC` or `NON-CRC`.

The value `WRAPPED OPTIMIZED` is only allowed if the book item's production type is `CRC`.

If the purpose of the book item is not `MAIN`, then the value `DISTILLED OPTIMIZED` must be used.

book-item/files-info

The element `files-info` contains all the information needed to process the files belonging to the item.

book-item/files-info/ml

One or two XML files can be associated with the item. These are listed under `ml`.

book-item/files-info/ml/pathname

In case `book-item`'s attribute `omitted` is not present or has the value `false`, `pathname` is the pathname, relative to dataset.xml, of the XML file.

In case `book-item`'s attribute `omitted` has the value `true`, `pathname` is the pathname of the XML file in the system it resides in (an absolute pathname starting with a server name).

book-item/files-info/ml/purpose

The `purpose` element contains the purpose of the item. There must be always one XML file with purpose `MAIN`. An optional second XML file is a changes-with-respect-to file with purpose `CHANGES`. A third optional XML file is a file containing pagebreak information, the purpose is `PAGEBREAK`.

book-item/files-info/ml/dtd-version

The version of the DTD and the top-level element (doctype) used to capture the item is contained in `dtd-version`. This must, of course, be identical to the declaration in the XML file.

book-item/files-info/ml/weight

The weight of the XML file can be `FULL-TEXT`, `HEAD-AND-TAIL` or `CONTENTS-ENTRY-ONLY`. The weight of the item indicates which parts of the text are captured in XML.

book-item/files-info/ml/asset

All assets belonging to the item are listed under `asset`. This includes all external files declared in the XML file, it does not include any strip-ins or fingerprints.

book-item/files-info/ml/asset/pathname

In case `book-item`'s attribute `omitted` is not present or has the value `false`, `pathname` is the pathname, relative to dataset.xml, of the XML file.

In case `book-item`'s attribute `omitted` has the value `true`, `pathname` is the pathname of the XML file in the system it resides in (an absolute pathname starting with a server name).

book-item/files-info/ml/asset/type

The `type` of the asset can have the values `APPLICATION`, `IMAGE-CAP`, `IMAGE-MMC`, `AUDIO`, `VIDEO`, `VIDEO-FLASH`, `XML`. These types are defined in Section 3.3.

book-item/files-info/web-pdf

An S280 item has exactly one web PDF file associated with it, the main PDF file which contains the item.

book-item/files-info/web-pdf/pathname

In case `book-item`'s attribute `omitted` is not present or has the value `false`, `pathname` is the pathname, relative to dataset.xml, of the XML file.

In case `book-item`'s attribute `omitted` has the value `true`, `pathname` is the pathname of the XML file in the system it resides in (an absolute pathname starting with a server name).

book-item/files-info/web-pdf/purpose

`purpose` indicates what the web PDF file is for, it must be `MAIN`.

book-item/files-info/web-pdf/pdf-version

`pdf-version` is the version of the web-PDF file, the allowed values are `1.7 6.4`, `1.7 6.5`, and `1.7 7.0`.

book-item/files-info/web-pdf/pdf-property

The element `pdf-property` describes the nature of the web PDF file and has one of the following values: `WRAPPED OPTIMIZED`, `DISTILLED OPTIMIZED`, `DISTILLED OPTIMIZED BOOKMARKED`.

7.3. S280-Proof deliveries

The S280-Proof schema is used to transport S280 datasets from suppliers to ProofCentral. The schema is equal to the S280 schema except that the element `web-pdf` is made optional as ProofCentral doesn't use the Web PDF files.

The S280-Proof schema cannot be used to deliver material to the Electronic Warehouse.

7.4. Q300 deliveries

A Q300 book project is a deliverable of the CAP process that precedes the S300-H300 delivery. It is used for proofing the dataset prior to publication. In particular the book's hub file can be proofed this way.

The Q300 dataset directory structure is defined in Section 4.1.10.

A Q300 schema contains one book project. It can always be accompanied with all or some of its items, and it has the option to add web PDF files of book pages for the purpose of proofing.

The schema is equal to the S300-H300 schema with the following exceptions:

- The value of `book-project/version/stage` is `Q300`.
- Element `book-project/files-info/web-pdf` is different. It allows a collection of untyped web PDF files (i.e. there is no subelement `purpose`) to be included in the delivery.
- Element `book-item` is optional.

7.5. S300-H300 deliveries

An S300 book item is a complete book chapter or "non-chapter" (e.g., a preface, glossary, index or other item in the front- or backmatter) that has been authorized by the authors and/or the editors.

An H300 book project is a complete book project, with the hub file for the book project.

The S300-H300 dataset directory structure for books is defined in Section 4.1.10. The schema for the dataset.xml of S300 book item and H300 book project deliveries is illustrated in Figs. 8 and 9.

Below, we traverse the S300-H300 schema for books starting from the element `dataset-content`. The elements `dataset-unique-ids` and `dataset-properties` are explained in Section 4.2.

The element `dataset-content` dataset-content is the container element for all of the metadata covering all of the items in the book dataset delivery. It consists of precisely one book-project element one or more book-item elements belonging to each book being delivered.

book-project

The element `book-project` contains all of the metadata elements needed to uniquely identify the book. It also contains the location of the book's hub file in the dataset. It contains the following elements: `version`, `book-project-unique-ids`, `book-project-properties`, and `files-info`.

Sample tagging of a book project is found in Fig. 10 (p. 176).

book-project/version

book-project/version/version-number

At the moment only S300-H300 is envisioned for book deliveries, the content of the version-number element will be S300.x (S300.1 for new submissions, S300.2 for 1st resupply, S300.3 for 2nd resupply, etc.).

book-project/version/stage

The content of the stage element will always be H300 for book projects.

book-project/book-project-unique-ids

The element `book-project-unique-ids` contains required `isbn` and `pii` elements, with an optional `doi` element. The PII may be a collection PII and the DOI may be based on one. All should be properly formatted with dashes, slashes and parentheses. Note that the value of the `isbn` element must match exactly the value of the ISBN already present in the EWII basedata for the book. If there is a mismatch, an error will be generated by EWII and the dataset for that book will be unable to be imported.

book-project/book-project-properties

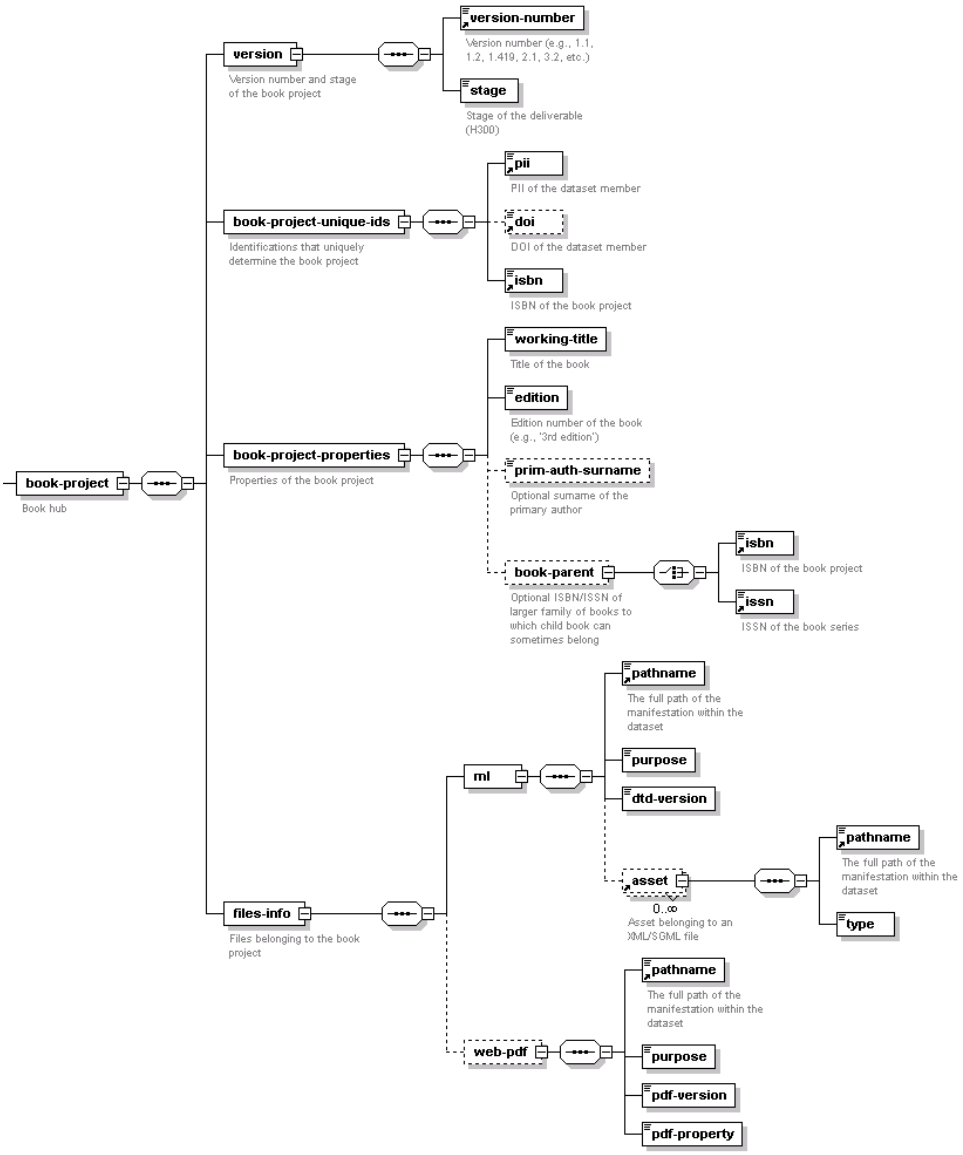


Figure 8: S300-H300 book schema, part 1.

The element `book-project-properties` contains mandatory `working-title` and `edition` elements, with optional `prim-auth-surname` and `book-parent` elements.

book-project/book-project-properties/working-title

The `working-title` element contains the title of the book (present to assist human readability and processing of the dataset.xml file). This version of the title is not to be confused with the official title for the book, which is in the hub file and may already be present in EWII as basedata.

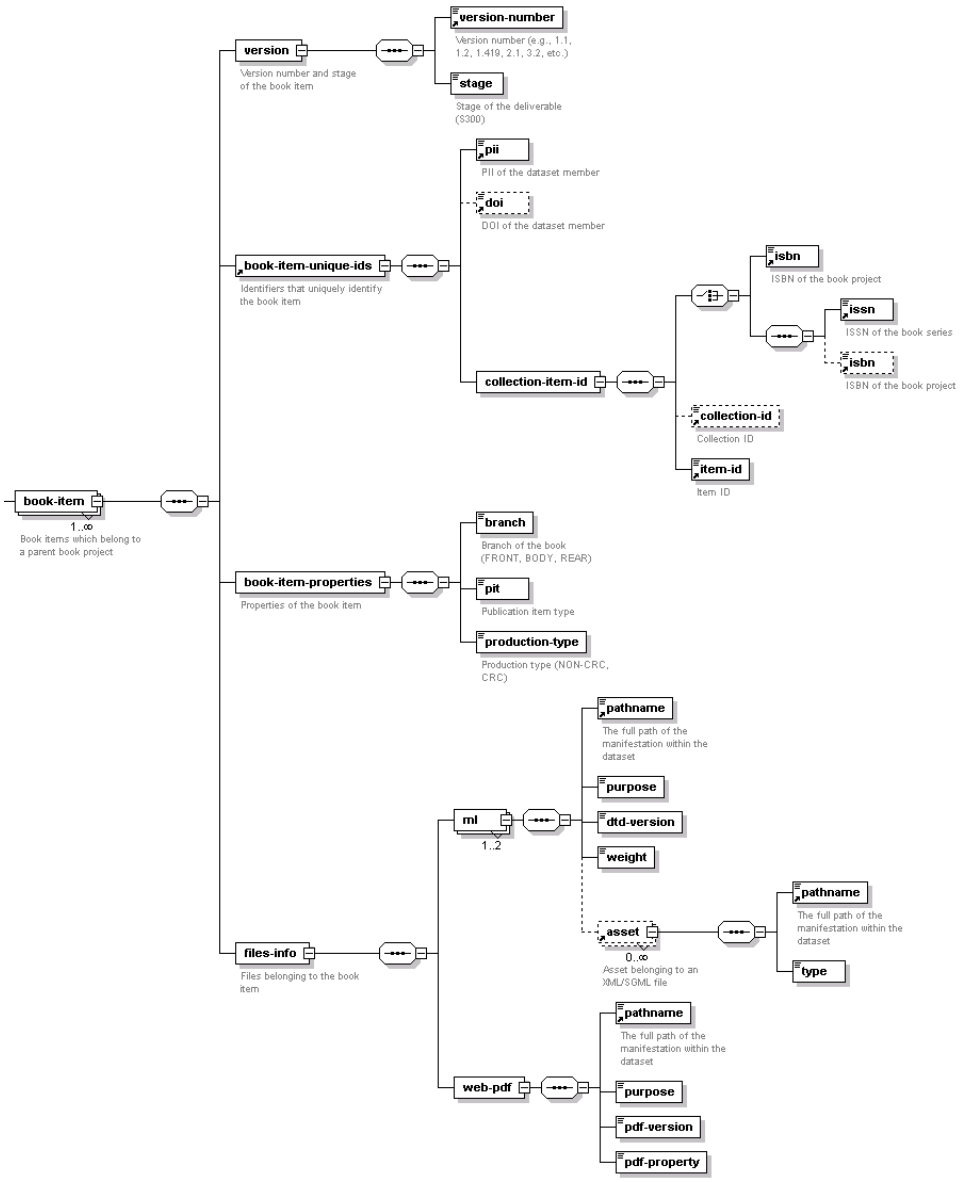


Figure 9: S300-H300 book schema, part 2.

book-project/book-project-properties/edition

The `edition` element contains a numerical value and not text version of the edition number for the book. Textual representations of the edition will live only in the hub file.

book-project/book-project-properties/prim-auth-surname

The `prim-auth-surname` element contains the surname of the book's primary author or editor. Within editorial and book production, books are often referred to in this manner (e.g., the "Watson" book).

book-project/book-project-properties/book-parent

The `book-parent` element (with child elements either `isbn` or `issn`) is only used when a book project belongs to a larger collection of books. If the book project being submitted is a book serial, the ISSN of the series must appears in the `issn` element. If the book project being submitted is a CONTRAST ultra-light delivery of an ancillary product (e.g., an instructor's manual) issued as a companion product to a textbook, then the ISBN of the textbook should appear in the `isbn` element.

book-project/files-info

The element `files-info` contains one `ml` subelement and an optional `web-pdf` subelement.

book-project/files-info/ml

The element `ml` contains mandatory `pathname`, `purpose`, and `dtd-version` elements, with optional/repeatable `asset` elements. This element describes the location and purpose of all the hub XML file and its assets.

book-project/files-info/ml/pathname

`pathname` is the pathname, relative to dataset.xml (that is, relative to the base directory), of the book project hub file.

book-project/files-info/ml/purpose

The `purpose` element contains the purpose of the item. There must be always one XML file with purpose `MAIN`. An optional second XML file is a changes-with-respect-to file with purpose `CHANGES`. A third optional XML file is a file containing pagebreak information, the purpose is `PAGEBREAK`.

book-project/files-info/ml/dtd-version

The `dtd-version` element contains a value from the "ml-versions-list" in the S300-H300 schema. The version must be identical to the version as declared in the XML file.

book-project/files-info/ml/asset

All assets belonging to the item are listed under `asset`. This includes all external files declared in the XML file, it does not include any strip-ins or fingerprints.

book-project/files-info/ml/asset/pathname

`pathname` is the pathname, relative to dataset.xml, of the asset file.

book-project/files-info/ml/asset/type

The `type` of the asset can have the values `IMAGE-CAP`, `IMAGE-COVER`. These types are defined in Section 3.3.

book-project/files-info/web-pdf

```
<book-project>
  <version>
    <version-number>H300.1</version-number>
    <stage>H300</stage>
  </version>
  <book-project-unique-ids>
    <pii>B0-323-01195-0/00000-X</pii>
    <isbn>0-323-01195-0</isbn>
  </book-project-unique-ids>
  <book-project-properties>
    <working-title>Mosby's Clinical Nursing</working-title>
    <edition>5</edition>
    <prim-auth-surname>Thompson</prim-auth-surname>
  </book-project-properties>
  <files-info>
    <ml>
      <pathname>0323011950/main.xml</pathname>
      <purpose>MAIN</purpose>
      <dtd-version>EHS-BOOKS 5.1.1 EHS-BOOK</dtd-version>
    </ml>
  </files-info>
</book-project>
```

Figure 10: Example tagging of a book project.

An H300 item can have one web PDF file associated with it, the PDF file which contains the complete book project.

book-project/files-info/web-pdf/pathname

`pathname` is the pathname, relative to dataset.xml, of the PDF file.

book-project/files-info/web-pdf/purpose

`purpose` indicates what the web PDF file is for, it must be `COMPLETE`.

book-project/files-info/web-pdf/pdf-version

`pdf-version` is the version of the web-PDF file, the allowed values are `1.7 6.4`, `1.7 6.5`, and `1.7 7.0`.

book-project/files-info/web-pdf/pdf-property

The element `pdf-property` describes the nature of the web PDF file and has one of the following values: `WRAPPED OPTIMIZED`, `DISTILLED OPTIMIZED`, `DISTILLED OPTIMIZED BOOKMARKED`.

book-item

The element book-item contains all of the metadata elements needed to uniquely identify the book item as well as the location of the book item's asset files in the dataset. It contains

the following elements: version, book-item-unique-ids, book-item-properties, and files-info.

It can have one attribute, omitted, which takes values false (default) and true. The latter value is to be used if an item is omitted from the dataset. It should only be used if the dataset-action has value PARTIAL-RELOAD. (See also p. 41 for information on partial deliveries.)

Sample tagging of book-item can be found in Fig. 11 (p. 180).

book-item/version

book-item/version/version-number

The element version-number contains the version number of the item, as described in Section 3.2. The version number is assigned by Elsevier.

book-item/version/stage

The value of stage is S300.

book-item/book-item-unique-ids

The element book-item-unique-ids contains a required pii element, with an optional doi element. Each should be properly formatted with dashes, slashes and parentheses and should be identical to the same identifiers in the XML and PDF files. Additionally it contains required element collection-item-id.

book-item/book-item-unique-ids/collection-item-id

The element collection-item-id contains element isbn or element issn optionally followed by isbn containing the ISSN and ISBN of the book project or book series. It has an optional element collection-id containing the ID of the collection the item belongs to and an optional item-id element containing the ID of the item. The latter two elements are the counterparts of elements jid and aid in the journal schemas. The element collection-id is only to be used for book series.

book-item/book-item-unique-ids/collection-item-id/item-id

The element item-id contains the ID of the item. It may not be more than 8 characters long.

book-item/book-item-properties

The element book-item-properties contains three mandatory elements: branch, pit, and production-type.

book-item/book-item-properties/branch

The branch element is used to identify in which portion of the book hierarchy the item occurs. There are three possible values: FRONT, BODY or REAR.

book-item/book-item-properties/pit

The pit element is a list of different types of items which can appear in books, similar to journal Publishing Item Types. The list of possible values is contained in the "pit-list" in the schema book-item-project.xsd. The value of pit must be identical to the value of the top-level attribute docsubtype in the XML file. (Except for the case, that is. PITs are always written uppercase in dataset.xml files and lowercase in DTD 5.x files.)

book-item/book-item-properties/production-type

The production-type element contains one of two possible values: CRC or NON-CRC.

The value WRAPPED OPTIMIZED is only allowed if the book item's production type is CRC.

If the purpose of the book item is not MAIN, then the value DISTILLED OPTIMIZED must be used.

book-item/files-info

The element files-info contains all the information needed to process the files belonging to the item.

book-item/files-info/ml

One or two XML files can be associated with the item. These are listed under ml.

book-item/files-info/ml/pathname

In case book-item's attribute omitted is not present or has the value false, pathname is the pathname, relative to dataset.xml, of the XML file.

In case book-item's attribute omitted has the value true, pathname is the pathname of the XML file in the system it resides in (an absolute pathname starting with a server name).

book-item/files-info/ml/purpose

The purpose element contains the purpose of the item. There must be always one XML file with purpose MAIN. An optional second XML file is a changes-with-respect-to file with purpose CHANGES. A third optional XML file is a file containing pagebreak information, the purpose is PAGEBREAK.

book-item/files-info/ml/dtd-version

The version of the DTD and the top-level element (doctype) used to capture the item is contained in dtd-version. This must, of course, be identical to the declaration in the XML file.

book-item/files-info/ml/weight

The weight of the XML file can be FULL-TEXT, HEAD-AND-TAIL or CONTENTS-ENTRY-ONLY. The weight of the item indicates which parts of the text are captured in XML.

book-item/files-info/ml/asset

All assets belonging to the item are listed under asset. This includes all external files declared in the XML file, it does not include any strip-ins or fingerprints.

book-item/files-info/ml/asset/pathname

In case book-item's attribute omitted is not present or has the value false, pathname is the pathname, relative to dataset.xml, of the XML file.

In case book-item's attribute omitted has the value true, pathname is the pathname of the XML file in the system it resides in (an absolute pathname starting with a server name).

book-item/files-info/ml/asset/type

The type of the asset can have the values APPLICATION, IMAGE-CAP, IMAGE-MMC, AUDIO, VIDEO, VIDEO-FLASH, XML. These types are defined in Section 3.3.

book-item/files-info/web-pdf

An S300 item has exactly one web PDF file associated with it, the main PDF file which contains the item.

book-item/files-info/web-pdf/pathname

In case book-item's attribute omitted is not present or has the value false, pathname is the pathname, relative to dataset.xml, of the XML file.

In case book-item's attribute omitted has the value true, pathname is the pathname of the XML file in the system it resides in (an absolute pathname starting with a server name).

book-item/files-info/web-pdf/purpose

purpose indicates what the web PDF file is for, it must be MAIN.

book-item/files-info/web-pdf/pdf-version

pdf-version is the version of the web-PDF file, the allowed values are 1.7 6.4, 1.7 6.5, and 1.7 7.0.

book-item/files-info/web-pdf/pdf-property

The element pdf-property describes the nature of the web PDF file and has one of the following values: WRAPPED OPTIMIZED, DISTILLED OPTIMIZED, DISTILLED OPTIMIZED BOOKMARKED.

```
<book-item>
  <version>
    <version-number>S300.1</version-number>
    <stage>S300</stage>
  </version>
  <book-item-unique-ids>
    <pii>B0-323-01195-0/00002-X</pii>
  </book-item-unique-ids>
  <book-item-properties>
    <branch>BODY</branch>
    <pit>CHP</pit>
    <production-type>NON-CRC</production-type>
  </book-item-properties>
  <files-info>
    <ml>
      <pathname>0323011950/body/B032301195000002X/main.xml</pathname>
      <purpose>MAIN</purpose>
      <dtd-version>EHS-BOOKS 5.1.1 CHAPTER</dtd-version>
      <weight>FULL-TEXT</weight>
      <asset>
        <pathname>0323011950/body/B032301195000002X/
          main.assets/gr1.jpg</pathname>
        <type>IMAGE-CAP</type>
      <asset>
    </ml>
    <web-pdf>
      <pathname>0323011950/body/B032301195000002X/main.pdf</pathname>
      <purpose>MAIN</purpose>
      <pdf-version>1.4 6.0</pdf-version>
      <pdf-property>DISTILLED OPTIMIZED BOOKMARKED</pdf-property>
    </web-pdf>
  </files-info>
</book-item>
```

Figure 11: Example tagging of a book-item.

7.6. S350-H350 deliveries

An H350 book project is a deliverable of the PreCAP process. In the PreCAP process, printed books are scanned and by means of OCR technology, datasets are created. S350 items are items in H350 book projects.

The S350-H350 dataset directory structure is defined in Section 4.1.10.

An S350-H350 schema contains one book project. It must always be accompanied with all its items. The schema is equal to the S300-H300 schema with the following obvious exceptions:

- The value of `dataset-properties/production-process` is `PRECAP`.
- The value of `book-project/version/stage` is `H350`.
- The value of `book-item/version/stage` is `S350`.

In addition there is the following change:

book-item/files-info/raw-text

An S350 item also possesses a raw text manifestation, which is obtained by optical character recognition from the scanned PDF file.

7.7. O300 deliveries

An O300 book project is a deliverable of the CAP process that is delivered after the S300-H300 delivery. It contains a complete book in PDF format and is meant for online book sellers. It is accompanied by a small XML file containing some metadata.

The O300 dataset directory structure is defined in Section 4.1.10.

Below, we traverse the O300 schema for books starting from the element `dataset-content`. The elements `dataset-unique-ids` and `dataset-properties` are explained in Section 4.2.

The element `dataset-content` is the container element for all of the metadata covering all of the items in the book dataset delivery. It consists of precisely one `book-project` element.

book-project

The element `book-project` contains all of the metadata elements needed to uniquely identify the book. It also contains the location of the book's "hub" file in the dataset. It contains the following elements: `version`, `book-project-unique-ids`, `book-project-properties`, and `files-info`.

Sample tagging of a book project is found in Fig. 12 (p. 185).

book-project/version

book-project/version/version-number

The element `version-number` contains the version number of the book, as described in Section 3.2. For O300 deliveries it will be `O300.`x.

book-project/version/stage

The element `stage` always contains the value `O300`.

book-project/book-project-unique-ids

The element `book-project-unique-ids` contains required `pii` and `isbn` elements. The PII may be a collection PII. All should be properly formatted with dashes and periods.

book-project/book-project-properties

The element `book-project-properties` contains mandatory `working-title` and `edition` elements, with optional `prim-auth-surname` and `book-parent` elements.

book-project/book-project-properties/working-title

The `working-title` element contains the title of the book (present to assist human readability and processing of the dataset.xml file). This version of the title is not to be confused with the official title for the book, which is in the book's metadata file.

book-project/book-project-properties/edition

The edition element contains a numerical value and not the text version of the edition number for the book. Textual representations of the edition will live only in the hub file.

book-project/book-project-properties/prim-auth-surname

The prim-auth-surname element contains the surname of the book's primary author or editor. Within editorial and book production, books are often referred to in this manner (e.g., the "Watson" book).

book-project/book-project-properties/book-parent

The book-parent element (with child elements either isbn or issn) is only used when a book project belongs to a larger collection of books. If the book project being submitted is a book serial, the ISSN of the series must appears in the issn element.

book-project/files-info

The element files-info contains all the information needed to process the files belonging to the item.

book-project/files-info/ml

Exactly one XML file is associated with the book project. It is listed under ml. The file contains the metadata for the book.

book-project/files-info/ml/pathname

pathname is the pathname, relative to dataset.xml, of the XML file.

book-project/files-info/ml/purpose

The element purpose always contains the value MAIN.

book-project/files-info/ml/dtd-version

The version of the DTD and the top-level element (doctype) used to capture the item is contained in dtd-version. It always contains the value BOOK-METADATA 5.0.0 BOOK-METADATA or BOOK-METADATA 5.0.1 BOOK-METADATA. This must, of course, be identical to the declaration in the XML file.

book-project/files-info/ml/asset

There is one asset belonging to the item: the cover-image. It is listed under asset.

book-project/files-info/ml/asset/pathname

pathname is the pathname, relative to dataset.xml, of the asset file.

book-project/files-info/ml/asset/type

The type of the asset must have the value IMAGE-COVER.

book-project/files-info/web-pdf

An O300 book project has exactly one web PDF file associated with it. The PDF file contains the complete book.

book-project/files-info/web-pdf/pathname

`pathname` is the pathname, relative to dataset.xml, of the web PDF file.

book-project/files-info/web-pdf/purpose

`purpose` indicates what the web PDF file is for. The allowed values are `COMPLETE`, `COMPLETE-PF` and `COMPLETE-CE`.

book-project/files-info/web-pdf/pdf-version

`pdf-version` is the version of the web-PDF file, the allowed values are `1.7 6.4`, `1.7 6.5`, and `1.7 7.0`.

```
<book-project>
  <version>
    <version-number>O300.2</version-number>
    <stage>O300</stage>
  </version>
  <book-project-unique-ids>
    <pii>B978-0-323-01195-2.X0001-4</pii>
    <isbn>978-0-323-01195-2</isbn>
  </book-project-unique-ids>
  <book-project-properties>
    <working-title>Mosby's Clinical Nursing</working-title>
    <edition>5</edition>
    <prim-auth-surname>Thompson</prim-auth-surname>
  </book-project-properties>
  <files-info>
    <ml>
      <pathname>9780323011952/book-metadata.xml</pathname>
      <purpose>MAIN</purpose>
      <dtd-version>BOOK-METADATA 5.0.1 BOOK-METADATA</dtd-version>
      <asset>
        <pathname>9780323011952/book-metadata.assets/
          e9780323011952_cover.jpg</pathname>
        <type>IMAGE-COVER</type>
      </asset>
    </ml>
    <web-pdf>
      <pathname>9780323011952/e9780323011952.pdf</pathname>
      <purpose>COMPLETE</purpose>
      <pdf-version>1.7 6.2</pdf-version>
    </web-pdf>
  </files-info>
</book-project>
```

Figure 12: Example tagging of an O300 book project (a second delivery).

7.8. F300 deliveries

An F300 dataset is a dataset containing "fat PDF" files, suitable for printing an entire book.

book-project

The element `book-project` contains all the information pertaining to the book in the dataset, as well as the locations within the dataset of the fat PDF files.

book-project/version

book-project/version/version-number

The element `version-number` contains the version number of the book, as described in Section 3.2. The version number is assigned by Elsevier and is included in the order.

book-project/version/stage

The value of `stage` is `F300`.

book-project/book-project-unique-ids

book-project/book-project-unique-ids/pii

The element `pii` contains the PII of the book.

book-project/book-project-unique-ids/doi

The element `doi` contains the DOI of the book, if any. Only books that will appear online may have a DOI.

book-project/book-project-unique-ids/isbn

The book project's ISBN, an ISBN-13, is captured with `isbn`.

book-project/book-project-properties

book-project/book-project-properties/working-title

The `working-title` element contains the title of the book (present to assist human readability and processing of the dataset.xml file). This version of the title is not to be confused with the official title for the book (which is present in the book's hub file).

book-project/book-project-properties/edition

The `edition` element contains the edition number of the book (as a numerical value).

book-project/book-project-properties/prim-auth-surname

The `prim-auth-surname` element contains the surname of the book's primary author or editor. Within editorial and book production, books are often referred to in this manner (e.g., the "Watson" book).

book-project/book-project-properties/book-parent

The `book-parent` element (with child element either `isbn` or `issn`) is only used when a book project belongs to a larger collection of books. If the book project being submitted is a book serial, the ISSN of the series must appears in the `issn` element.

book-project/book-project-properties/book-parent/isbn

book-project/book-project-properties/book-parent/issn

book-project/files-info

The element `files-info` contains all the information needed to process the files belonging to the issue.

book-project/files-info/ml

The XML file in the F300 dataset provides a full description to the printer about how to print the book. It explains all the fat (print) PDF files in the dataset.

book-project/files-info/ml/pathname

`pathname` is the pathname, relative to dataset.xml (that is, relative to the base directory), of the print description XML file.

book-project/files-info/ml/schema-version

The version of the W3C schema of the print description file is contained in `schema-version`. The allowed values are `print 1.0`, `print 1.1`, `print 1.2` and `print 1.3`. The version must, of course, be identical to the declaration in the XML file.

book-project/files-info/print-pdf

Print PDF files, also known as “fat PDF” files, are used by the printer to print the book.

book-project/files-info/print-pdf/pathname

`pathname` is the pathname, relative to dataset.xml, of the print PDF file.

book-project/files-info/print-pdf/purpose

`purpose` indicates the suitability of the PDF file for printing and/or digital use. It has three possible values: `SUITABILITY-OFFSET-AND-DIGITAL` if the PDF file is suitable for printing and for digital use, `SUITABILITY-DIGITAL` if the PDF file is suitable for digital use, while `SUITABILITY-NONE` is used when no statement on suitability can be made.

book-project/files-info/print-pdf/pdf-version

`pdf-version` is the version of the print-PDF file, the allowed values are `1.3 1.0`, and `1.6 2.0`.

book-project/files-info/print-pdf/pdf-property

`pdf-property` indicates how the PDF file is created. It has two possible values: `SCANNED` if the PDF file is the result of scanning, and `NOT SCANNED` if the PDF file is created in another way.

7.9. Book Project deliveries

Deliveries of material for special projects is to be done with the so-called project schemas. For book deliveries of type S200, S280, Q300, S300-H300 and S350-H350 there is a project variant. The project schemas are based on the regular schemas. Some pattern validation has been removed. This validation will be performed elsewhere and is not described here. In fact, one should assume that the schemas are the same.

Chapter 8
The Print schema

This chapter describes all the elements of the print schema for “fat” PDF deliveries to printers.

Journals

Datasets are supplied as a result of an order sent by PTS. When the step “Finalize Issue” is started, the supplier receives an S300 and an F300 order and creates S300 and F300 datasets. An F300 dataset contains an issue cover-to-cover for print publication and consists of so-called “fat” PDFs for printing.

S300 datasets are sent to EWII which closes the above-mentioned step. The F300 datasets are delivered to the EWII. In the future the EWII will then send the dataset to the printer accompanied by a print (F300) order. The dataset will also contain a file that enables the printer to print the issue, the print.xml. Currently the EW sends the dataset for journal Print on Demand printers only. The initial print-run journal printers receive the F300 dataset via a different route.

The print.xml file will contain information from the F300 order, the supplier’s metadata and the dataset.xml file. It will be stored in the EWII, thus enabling printing of the issue at any time in the future.

Books

Changes were made to the schema (versions 1.2 and 1.3) to make it usable for “Print on Demand” via the VTW and for delivery of F300 Book datasets to the EW and the printers.

In version 1.4 changes were made to be able to better deliver book series volumes. Also, journal issue and book PIIs were added for better identification, as well as two more page totals.

8.1. Print deliveries

This section contains a description of elements from the print.xsd schema, version 1.4. The print.xml file will be created by the typesetters.

In the case of journals the content of certain elements is based on the F300 order. Note that in some of those cases the element names are different:

F300 order	Print schema
`vol-from`	`vol-first`
`vol-to`	`vol-last`
`iss-from`	`iss-first`
`iss-to`	`iss-last`
`supp`	`suppl`

print

This is the top-level element. It contains all the information to enable the printer to print the serial issue, book series volume or book project. It contains the metadata in either `serial-issue-properties` or `book-project-properties` which is followed by `print-content`. It has an attribute `schema-version` with a fixed value `1.4`.

print/serial-issue-properties

This element contains metadata of the serial issue or book series volume to be printed.

print/serial-issue-properties/pii

The element `pii` contains the formatted PII for the journal issue hub PII or the book series volume PII.

print/serial-issue-properties/jid

The element `jid` contains the Elsevier system code of the serial publication to which the issue belongs. Book series JIDs should begin with "BS:" (note that book series are not in PTS). This is taken from the F300 order.

print/serial-issue-properties/issn

The element `issn` contains the ISSN of the serial publication to which the issue belongs. This is taken from the F300 order.

print/serial-issue-properties/isbn

The (optional) element `isbn` contains the ISBN of the journal issue or the book series volume. This is taken from the F300 order.

print/serial-issue-properties/title

This (optional) element contains the title of the journal issue or the book series volume.

print/serial-issue-properties/volume-issue-number

The volume/serial-issue number of the issue is captured in `volume-issue-number`. It consists of `vol-first`, `vol-last`, `iss-first`, `iss-last` and `suppl`. The values are taken from the F300 order (elements `vol-from`, `vol-to`, `iss-from`, `iss-to`, `supp`, respectively).

print/serial-issue-properties/cover-date-printed

Element `cover-date-printed` contains the cover-date of the journal issue as it is printed on the cover. This is taken from the F300 order.

print/serial-issue-properties/trim-size

This element contains the trim size. It has a mandatory attribute `unit` with possible values `mm` and `in`.

```
<trim-size unit="mm">210x280</trim-size>
<trim-size unit="in">8_1/2x11_1/4<trim-size>
```

print/serial-issue-properties/page-totals

This element has five mandatory and two optional subelements:

- `no-pages-prelims`, the total number of pages in the prelims of the issue
- `no-pages-interior`, the total number of pages in the interior of the issue (including blank pages)
- `no-pages-extra`, the total number of additional pages in the issue (e.g. advertisements)
- `no-pages-bm`, the total number of pages in the back matter of the issue
- `no-pages-total`, the sum of the above totals
- `no-pages-colour`, optional, the number of pages with colour
- `no-pages-mono`, optional, the number of pages without colour

All totals except `no-pages-total` can have the value zero. The values are taken from the F300 order.

print/serial-issue-properties/volume-set

This element indicates if a book series volume is part of a volume set. Its values are `true` and `false`.

print/serial-issue-properties/pin-code

This (optional) element indicates if there is an additional PIN code included in the book series volume. Its values are `true` and `false`.

print/book-project-properties

This element contains metadata of the book project to be printed.

print/book-project-properties/pii

The element `pii` contains the formatted PII for the book hub PII.

print/book-project-properties/isbn

This element contains the ISBN of the book project.

print/book-project-properties/title

This element contains the title of the book project.

print/book-project-properties/trim-size

This element contains the trim size. It has a mandatory attribute `unit` with possible values `mm` and `in`.

print/book-project-properties/page-totals

This element contains the various page totals. For more information see the above description.

print/book-project-properties/volume-set

This element indicates if the book is part of a volume set. Its values are `true` and `false`.

print/book-project-properties/pin-code

This element indicates if there is an additional PIN code included in the book. Its values are `true` and `false`.

print/print-content

This element contains the information for an issue or book project in subelement(s) `fascicle`. For each printed form of the issue or part of a book project there is one such element.

print/print-content/fascicle

A journal issue or book series volume can have several printed forms. This applies mostly to the cover. For instance you may have a special version for members of a society that has a different cover. There may be other cases where the same issue is produced in more than one form. The information for each printed form of an issue is captured in a separate `fascicle`.[2] This means that PDF information (`print-pdf`) is duplicated across more than one `fascicle`. In case of more than one printed form of the issue, subelement `description` must be used to describe the issue in short.

A book project may be printed in separate parts. The information for each part of the book project is captured in a separate `fascicle`.

Similar to journal issues book projects or parts thereof may be printed in different forms. For every printed form a separate `fascicle` will be present. These different forms often have their own ISBN. This can be captured in element `fascicle-isbn`.

For every PDF file in the delivery there is a subelement `print-pdf` or `exterior-pdf` containing information about that PDF file.

2. Fascicle: A small bundle, or one of the parts of a book published in separate sections.

print/print-content/fascicle/description

This (optional) element contains a short description of the issue or book project. In case of a serial issue it is only to be used when there are more than one printed forms for this issue.

```
<description>Society member's copy</description>
```

print/print-content/fascicle/fascicle-isbn

This (optional) element contains the ISBN of the book project contained in the fascicle. It is used for distinguishing the various fascicles.

print/print-content/fascicle/spine-width

This (optional) element contains the width of the spine. Mandatory attribute `unit` indicates which units are to be used: `mm` or `in`.

print/print-content/fascicle/binding

This (optional) element contains the binding information. It has a mandatory attribute `type` with possible values `hardback` and `paperback`.

print/print-content/fascicle/text-paper-type

This (optional) element contains the paper type of the text pages. Typical values are `90gsm silk`, `60gsm silk` and `115gsm gloss`.

print/print-content/fascicle/printing-colours

This (optional) element contains the number of basic colours of the text. Typical values are `4-colors`, `1-color with 4-color sections` and `1-color`.

print/print-content/fascicle/printing-quality

This (optional) element contains a quality indication of the printed fascicle. Its values are `normal` and `high`.

print/print-content/fascicle/cover-stock

This (optional) element contains values that describe the cover materials.

print/print-content/fascicle/cover-lamination

This (optional) element contains the cover lamination. Typical values are `gloss film` and `matte film`.

print/print-content/fascicle/print-pdf

This element contains information about a PDF file. It has a mandatory attribute `type` which can have the values `normal`, `cover`, `blank`, `advert`, `fold-out` and `fpo`. The value is taken from the F300 order (in case of journals) and the typesetter's metadata.

Value `fpo`, "for position only", is used for advertisements, or similar material, that are delivered directly to the printer. A properly named and sized (blank) PDF file should still

be present. See [16] for more information. This value is not used for “Print on Demand” for books.

print/print-content/fascicle/print-pdf/sequence-no

Element `sequence-no` is assigned by the typesetter. This number is used to order the PDF files. In case of journals it is based on the F300 order. The sequence numbering starts anew in every fasicle.

print/print-content/fascicle/print-pdf/pathname

This element contains the pathname of the PDF file, relative to the base directory of the dataset.

print/print-content/fascicle/print-pdf/purpose

This element indicates the suitability of the PDF file for printing and/or digital use. It has three possible values: `SUITABILITY-OFFSET-AND-DIGITAL` if the PDF file is suitable for printing and for digital use, `SUITABILITY-DIGITAL` if the PDF file is suitable for digital use, while `SUITABILITY-NONE` is used when no statement on suitability can be made.

print/print-content/fascicle/print-pdf/pdf-property

This element indicates how the PDF file is created. It has two possible values: `SCANNED` if the PDF file is the result of scanning, and `NOT SCANNED` if the PDF file is created in another way.

print/print-content/fascicle/print-pdf/nr-physical-pages

This element contains the number of physical pages in the PDF file. This is supplied by the typesetter.

print/print-content/fascicle/print-pdf/pagination

Element `pagination` contains the page numbers in subelements `first-page` and (optional) `last-page`. These numbers are supplied by the typesetter.

print/print-content/fascicle/print-pdf/colour-pages

A colour page is a page with more than one print colour. This includes spot-colour pages.

This element must be present if the PDF contains colour pages. Element `colour-pages` contains one or more `colour-page` elements, each containing a (printed) page number of a colour page. This information is taken from the typesetter’s metadata.

print/print-content/fascicle/print-pdf/pii

This optional element contains the formatted PII of the journal or book item the print PDF file belongs to. The value is taken from the F300 order (in case of journals) and the typesetter’s metadata.

print/print-content/fascicle/print-pdf/batch

In case the print PDF file belongs to a batch of items (for instance a number of abstracts, or an article with comments) this optional element is used. It contains the formatted PIIs of the batch-members, each in its own subelement `batch-member/pii`. The values are taken from the F300 order (in case of journals) and the typesetter's metadata.

print/print-content/fascicle/print-pdf/id

This (optional) element contains the ID of the item the PDF belongs to. In case of journals this is the AID which is taken from the F300 order. In case of books this can be the item ID. If there is no ID, the element must not be present.

print/print-content/fascicle/print-pdf/remarks

This (optional) element contains the item remarks. These are taken from the F300 order (in the case of journals), possibly added with remarks from the typesetter's metadata.

print/print-content/fascicle/exterior-pdf

This (optional) element contains information about a so-called exterior PDF file. These PDF files are not part of the printed journal issue or printed book. The element has an attribute `type` which can have the values `card`, `colour-plate-inserts`, `dust-jacket`, `end-sheet` and `transparency`.

print/print-content/fascicle/exterior-pdf/pathname

This element contains the pathname of the exterior PDF file, relative to the base directory of the dataset.

print/print-content/fascicle/exterior-pdf/id

This (optional) element contains the ID of the exterior PDF file. If there is no ID, the element must not be present.

print/print-content/fascicle/exterior-pdf/remarks

This (optional) element contains the item remarks. These are taken from the F300 order (in the case of journals), possibly added with remarks from the typesetter's metadata.

Chapter 9
The Satellite schema

This chapter describes all the elements of the satellite schema for deliveries of satellites to the EW.

9.1. Satellite deliveries

Satellites contain information about articles, issues or book chapters. They are attached to those articles, issues or chapters. A prime example is a satellite containing annotations. In the future satellites can also contain information about other publications. Satellites are delivered to the EW in an A300 dataset, "A" for attach or annotate.

Convention. Satellite files are stored in their own *satellite directory*, within the base directory. The preferred format for the names of the satellite directories is "satellite1", "satellite2", etc. The current maximum number of satellites in one dataset is 500.

Therefore, a typical satellite dataset looks like this:

```
jmi20445/
  dataset.xml
  fp/dataset_xml_fp.xml
  satellite1/lancetcoll-acme-S014067360566388X.rdf
            /fp/lancetcoll-acme-S014067360566388X_rdf_xsl2191_fp.xml
  satellite2/lancetcoll-acme-S0304416503X20954.rdf
            /fp/lancetcoll-acme-S0304416503X20954_rdf_xsl2191_fp.xml
  ...
```

Below, we traverse the A300 schema for for satellite deliveries starting from the element `dataset-content`. The elements `dataset-unique-ids` and `dataset-properties` are explained in Section 4.2.

dataset-content

Element `dataset-content` contains information on the satellites in the delivery, in subelements `satellite`. The schema allows for more than one satellite per dataset, although for the time being only one satellite per dataset is accepted.

dataset-content/satellite

Every `satellite` element contains information on a satellite in four subelements.

dataset-content/satellite/version

The version information is stored in two elements. Subelement `version-number` contains the version number of the satellite, as described in Section 3.2. It is of the form `A300.`n. The stage is stored in subelement `stage` and is always equal to `A300`.

dataset-content/satellite/satellite-unique-ids

This element contains the unique ID of the satellite. It is a URI and is stored in subelement `uri`.

```
<satellite-unique-ids>
    <uri>http://data.elsevier.com/annotation/Tag-Annot-1/⌒
        Lancet-JMi-1/DOI:10.1016/S0140-6736(10)61969-1</uri>
</satellite-unique-ids>
```

dataset-content/satellite/satellite-properties

This element contains various properties of the satellite.

dataset-content/satellite/satellite-properties/satellite-group

Satellites are divided into groups, for example `ANNOTATION` or `KML`. The group is stored in `satellite-group` and the value will be supplied by Elsevier.

dataset-content/satellite/satellite-properties/satellite-code

Satellites are further subdivided into types which are represented by a code. The code is stored in `satellite-code` and the value will be supplied by Elsevier.

dataset-content/satellite/satellite-properties/parent-collection

A satellite belongs to an article, an issue or a book chapter. The optional element `parent-collection` contains the PII, ISSN or ISBN of the collection (i.e. the journal or book) that article, issue or book chapter is part of.

In case the satellite belongs to an article or an issue, the collection is identified by a PII, an ISSN or both (subelements `pii` and `issn`). In case the satellite belongs to a book chapter, the collection is identified by an ISBN-13 (subelement `isbn`).

dataset-content/satellite/satellite-properties/linked-issues

This optional element contains the issue(s) the satellite is linked to. They are identified in one or more subelements `pii`.

dataset-content/satellite/satellite-properties/linked-items

This optional element contains the item(s) the satellite is linked to. They are identified in one or more subelements `pii`.

dataset-content/satellite/files-info

The element `files-info` contains all the information needed to process the files belonging to the item.

dataset-content/satellite/files-info/ml

One XML file is associated with the item. This is listed under `ml`. This element describes the location and purpose of all the hub XML file and its assets. The element `ml` contains mandatory elements `pathname`, `purpose`, and either `dtd-version` or `schema-version`, followed by optional and repeatable `asset` elements.

dataset-content/satellite/files-info/ml/pathname

`pathname` is the pathname, relative to dataset.xml, of the XML file.

dataset-content/satellite/files-info/ml/purpose

`purpose` indicates what the XML file is for. For satellite files the purpose is always `SATELLITE`.

dataset-content/satellite/files-info/ml/dtd-version

This element is used if the satellite files adhere to a DTD. Currently there is only one possible value, `EF 5.0.0 ENHANCEMENT-FRAGMENT`.

dataset-content/satellite/files-info/ml/schema-version

This element is used if the satellite files adhere to a W3C schema. It contains the namespace of the schema.

```
<schema-version>http://www.elsevier.com/xml/schema/
    rdf/Lancet-JMi-1 ljmi_v010.xsd</schema-version>
```

dataset-content/satellite/files-info/ml/asset

All assets belonging to the item are listed under `asset`. This includes all external files declared in the XML file; it does not include any strip-ins or fingerprints.

The pathname of the XML file, relative to dataset.xml, is stored in subelement `pathname`. The type of the asset is stored in subelement `type`. Possible values for the type will be supplied by Elsevier.

Note: Assets for satellite files are not implemented at first.

```
<?xml version="1.0" encoding="UTF-8"?>
<dataset xmlns="http://www.elsevier.com/xml/schema/transport/satellite-2010.3/a300"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://www.elsevier.com/xml/schema/transport/satellite-2010.3/a300
    http://www.elsevier.com/xml/schema/transport/satellite-2010.3/a300.xsd"
  schema-version="2010.3">

  <dataset-unique-ids>
    <supplier-code>SUP1</supplier-code>
    <supplier-dataset-id>JMI20444</supplier-dataset-id>
    <timestamp>2010-12-08T15:10:00</timestamp>
  </dataset-unique-ids>

  <dataset-properties>
    <dataset-action>LOAD</dataset-action>
    <production-process>SCP</production-process>
  </dataset-properties>

  <dataset-content>
    <satellite>
      <version>
        <version-number>A300.1</version-number>
        <stage>A300</stage>
      </version>
      <satellite-unique-ids>
        <uri>http://data.elsevier.com/annotation/Tag-Annot1/
          Lancet-JMi-1/DOI:10.1016/j.joca.2006.06.016</uri>
      </satellite-unique-ids>
      <satellite-properties>
        <satellite-group>ANNOTATION</satellite-group>
        <satellite-code>LJMI1</satellite-code>
        <linked-items>
          <pii>S1063-4584(06)00204-4</pii>
        </linked-items>
      </satellite-properties>
      <files-info>
        <ml>
          <pathname>satellite/JMi-SampleRDF-3.rdf</pathname>
          <purpose>SATELLITE</purpose>
          <schema-version>http://www.elsevier.com/xml/schema/
            rdf/Lancet-JMi-1 ljmi_v010.xsd</schema-version>
        </ml>
      </files-info>
    </satellite>
  </dataset-content>

</dataset>
```

Figure 13: Sample A300 dataset.xml.

9.2. Satellite Project deliveries

Deliveries of material for special projects is to be done with the so-called project schemas. For satellite deliveries of type A300 there is a project variant. The project schemas are based on the regular schemas. Few pattern validation has been removed. This validation will be performed elsewhere and is not described here. In fact, one should assume that the schemas are the same.

Chapter 10
Dataset delivery protocol

This chapter deals with the protocol for delivering CONTRAST datasets to the Electronic Warehouse.

10.1. Network delivery to the Electronic Warehouse

The method used for network delivery is FTP transfer according to the push model, in which suppliers send datasets via FTP to a dedicated area on the Electronic Warehouse drop zone, and subsequently — after the transfer of the dataset is fully complete — transfer a so-called *ready message* which is stored next to the dataset on the EW drop zone. Details on the ready messages, which contain some basic information about the dataset in an XML format, are given below.

Each journal-related dataset is the result of one or more PTSIII orders. The workflow system has an open task for each deliverable. Immediately after delivery of the ready message to the EW, the suppliers can consider the tasks related to the components within the dataset closed.

The complete dataset including its dataset.xml is packed into one file according to either the zip or the tar format (where the latter may be compressed using gzip), as described in Chapter 4.1. Its filename is equal to the last eight characters of the dataset identifier affixed with one of the extensions `.zip`, `.tar` or `.tgz`, for example `x0001829.zip`. It is not possible to use the same dataset identifier twice. It has been agreed that the dataset identifier contains at most 12 characters.

There is no limit to the size of a dataset. However, there is a limit to the size of the dataset package files. If the size of the delivery is very large, it is allowed to split the dataset over more than one ZIP, tar or gzipped tar file. These dataset package files must be valid ZIP, tar or gzipped tar files in their own right, i.e. methods such as ZIP spanning or physically cutting a file in pieces are not to be used.

The splitting of a large dataset over more files should of course be done after the Elsevier validation tool has been run. The size of these files should be between 512 and 1024 MB. Merging the decompressed and/or de-archived files results in the original dataset.

10.2. CD, DVD, tape

The same structure is found on other delivery media: a “ready” file and a ZIP, tar or gzipped tar file.

10.3. Ready messages

The ready messages are in an XML format which validates against a very simple W3C schema, depicted in Fig. 14.

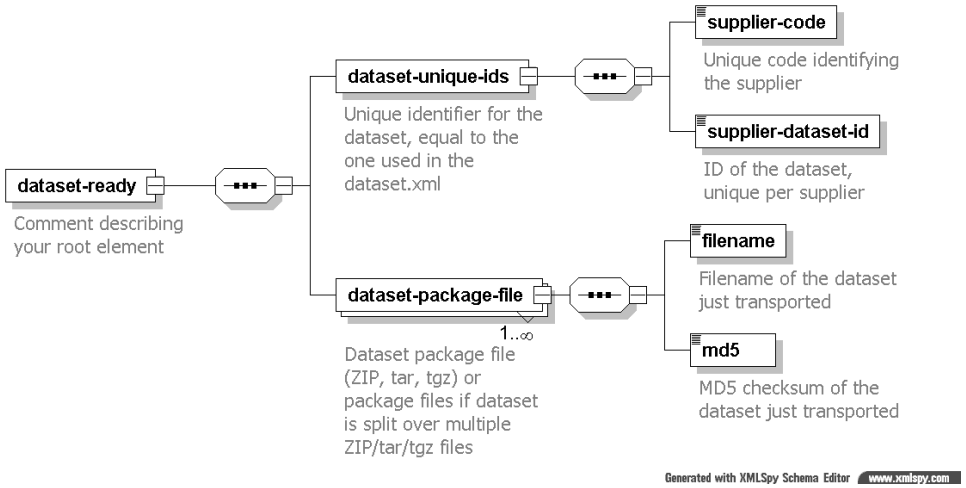


Figure 14: Ready XML schema.

- `dataset-ready`, the top-level element;
- `dataset-unique-ids`, element similar to the one in the transport schema, containing two subelements identifying the dataset as well as the file name of the dataset, i.e. it contains the same information as the dataset.xml with exception of the timestamp;
- `dataset-package-file`, with two subelements identifying the ZIP or tar file;
- `filename`, the filename of the package file;
- `md5`, the MD5 checksum of the dataset package file.

The content of `supplier-code` and `supplier-dataset-id` is case-sensitive, which means that the dataset ID as used in the ready message should match the ID used in the dataset.xml exactly, and the name of the package file put on the EW drop zone should match the file name given in the ready message exactly.

The ready message should contain nothing else apart from the XML information. In particular, it should contain no additional text such as an explanation of a resupply delivery. The file name of the ready message should be identical to the dataset identifier, followed by the extension "`.ready.xml`". The following is an example of a network delivery ready message:

XML

```
<?xml version="1.0" encoding="UTF-8"?>
<dataset-ready
  xmlns="http://www.elsevier.com/xml/schema/transport/ready-3.0/ready"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation=
    "http://www.elsevier.com/xml/schema/transport/ready-3.0/ready
     http://www.elsevier.com/xml/schema/transport/ready-3.0/ready.xsd"
  version="3.0">

  <dataset-unique-ids>
    <supplier-code>MACM</supplier-code>
    <supplier-dataset-id>x0001245437</supplier-dataset-id>
  </dataset-unique-ids>
```

```
    <dataset-package-file>
      <filename>01245437.tar</filename>
      <md5>58f6fe00b23d175815fb1a18105bbf9e</md5>
    </dataset-package-file>

  </dataset-ready>
```

The file name of this message would be `x0001245437.ready.xml`.

The ready message should be transferred to the EW drop zone after the transfer of the packed dataset file is completed. In the above example, two files would be present on the EW drop zone: `01245437.tar` and `x0001245437.ready.xml`.

If the dataset is delivered in more than one file, the first one that is mentioned in the ready message must contain the dataset.xml file. Note that the filenames of these files should still have eight characters. The following format is suggested (using the above example): `1245437a.tar`, `1245437b.tar`, etc.

The EW scans the EW drop zone for ready messages, reads them out, and then retrieves and processes the corresponding datasets.

10.4. Error handling

In case a dataset cannot be opened by the EW, then a message is sent by EW to the supplier with an indication of what went wrong, e.g., “file not found”, “dataset already registered”. The PTS task is then still open, and a renewed delivery attempt must be made with a new dataset ID, but the content retains its version number(s).

If an error is found after EW has opened the dataset, e.g., an error is found in a fingerprint, then a resupply order will be sent and a resupply task will be opened in PTS. As for all orders, a new dataset ID is used when the delivery is made and the version numbers given in the resupply order must be used.

10.5. Delivery from the Electronic Warehouse

As described in Chapter 5 a ZIP file with source material is placed on the drop zone for the supplier along with the XML order file. The ZIP file will be placed first, then the XML order, which therefore acts as a “ready” message.

Bibliography

[1] CAP specs for artwork, see Supplier Artwork Extranet.
[2] Specification for Elsevier "Web" PDF files (pdfreq40.pdf)
[3] Publication item types (rs98022.pdf).
[4] Tag by Tag Documentation of the Common Element Pool and the JA DTD (2004).
[5] Tag by Tag Documentation of the Serial Issue DTD (2004).
[6] Tag by Tag Documentation of the Health Sciences DTD (2004).
[7] The CAP Guide for MFC Activities (2003).
[8] MultiMedia Components (MMCs) – instructions for full-service suppliers (2003).
[9] Handling supplementary files in DTD 5.0 – instructions for full-service suppliers (2003).
[10] TBA
[11] Article copyright lines – by status and PIT (2003).
[12] Online publication dates in print (2004).
[13] Subitems Mk III – Ducks and ducklings (2005).
[14] Version number and PTS order DTD changes for EW / PTS 2008.1 (2008).
[15] Archival Requirements for Book Typesetters, v1.15, 2008.
[16] Strictly Foxtrot: F300 to EWII and printers — FAQ, v1.5 (2009).

Index