Publishers' Circular Segmentation Policy

0.1   Nomenclature:

Term	Definition
Title	The title of the whole run of the periodical
Volume	A set of numbers, usually grouped annually and numbered sequentially.
Number	An individual issue.
Department	A division within a number such as correspondence or news.
Item	An article or component within a department.
Wrapper	An advertising wrapper that surrounds a number. It is usually paginated separately to distinguish it from the paginated letterpress.
Front Matter	It was common practice in the C19th to issue material to be bound in with the numbers at the end of the volume. These might include a frontispiece and a preface.
Index	It was also common practice to issue an index to the volume, to be bound in at the back.
Masthead	The name of the title, usually printed in a fancy typeface.
Dateline	The strap line that contains information such as volume, number, and date.

These are arranged in a hierarchical structure:

Edition
-        Title
         -        Volume #1
                  -        Front Matter
                  -        Number #1
                           -        wrapper [will be indicated through metadata if present]
                           -        number contents
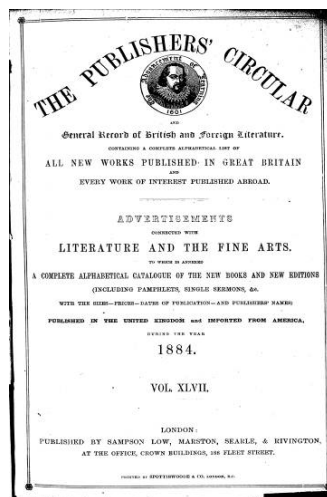   -   department #1
                                             -        item #1
                                             -        item #2
                                             -        etc…
                                    -        department #2
                                    -        department #3
                                    -        etc…
                  -        Number #2
                           -        wrapper [will be indicated through metadata if present]
                           -        number contents
                                    -        department #1
                                             -        item #1
                                             -        item #2
                                             -        etc…

0.0 Introductory note:

The *Publisher's Circular* is quite a complex title that consists of a number of different parts. Along with this document we have provided 3 marked-up PDFs that show how the rules should be applied across this content. Please note that in these PDFs items which contain department headers or mark the beginning of departments have been have been marked up with black boxes. Department headers have been indicated assuming that the results of the pending snippets demo will be positive and that we will accept the snippets plan for TOC entry. Please also note that in the PDFs it has sometimes been necessary to exclude dividing lines between items from indicated segments in order that they are not obscured by mark-up line. However, for the actual segmentation these should be included in segments at the end of the preceding item's segment.

0.1 Front Matter

This is usually material that is bound at the front of a volume and consists of volume frontispieces giving the title of the volume and the date. For example:

THE PUBLISHERS' CIRCULAR

General Record of British and Foreign Literature.

CONTAINING A COMPLETE ALPHABETICAL LIST OF

ALL NEW WORKS PUBLISHED· IN GREAT BRITAIN

AND

EVERY WORK OF INTEREST PUBLISHED ABROAD.

ADVERTISEMENTS

CONNECTED WITH

LITERATURE AND THE FINE ARTS.

TO WHICH IS ADDED

A COMPLETE ALPHABETICAL CATALOGUE OF THE NEW BOOKS AND NEW EDITIONS

(INCLUDING PAMPHLETS, SINGLE SERMONS, &c.

WITH THE SIZES—PRICES—DATES OF PUBLICATION—AND PUBLISHERS' NAMES)

PUBLISHED IN THE UNITED KINGDOM and IMPORTED FROM AMERICA,

DURING THE YEAR

1884.

VOL. XLVII.

LONDON:

PUBLISHED BY SAMPSON LOW, MARSTON, SEARLE, & RIVINGTON,

AT THE OFFICE, CROWN BUILDINGS, 188 FLEET STREET.

PRINTED BY SPOTTISWOODE & CO. LONDON, N.C.

Volume frontispiece

0.2 *English Catalogue of Books*

This is an annual supplement to the *Publisher's Circular* and is usually bound at the beginning of a volume, but might also appear at the back. It takes the form of an index, rather like a book index, and is in two parts, for example:
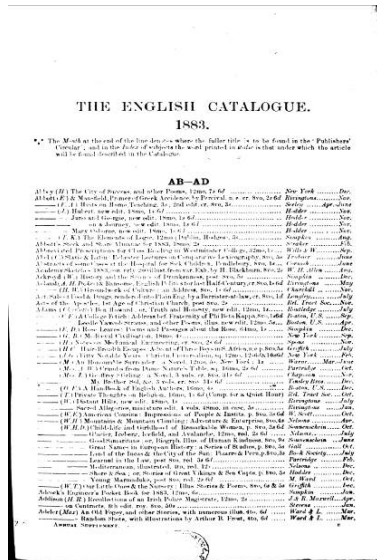
English Catalogue, **Catalogue**
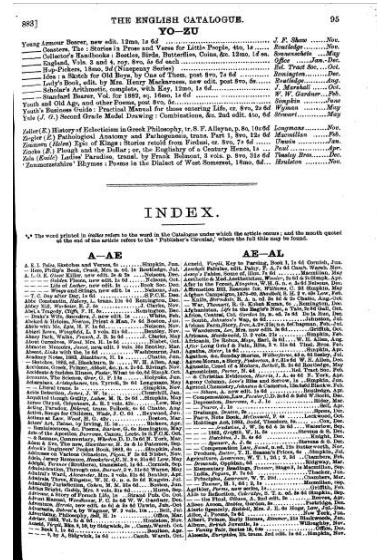
English Catalogue, **Index**

0.3 Wrappers

These often appear around a number on separately paginated sheets (pp. i-iv) and mainly carry advertisements. Below are the four pages of wrapper from 1 February 1888. The actual number would come between the front wrapper at p. ii and the back wrapper at p. iii:

Wrapper, p. i

Wrapper, p. ii

Wrapper, p. iii

Wrapper, p. iv

0.4   Numbers

Each issue of the *Publishers Circular* contains a wide variety of formats of material, including normal letterpress, small text-based advertisements with simple typography and pictures, large fancy full page adverts with complex typographical settings and pictures, and separate sequences of illustrations in the special extended numbers issues at Christmas time. For the purposes of specifying this policy we have made a distinction between **Type 1** content and **Type 2** content. Type 1 content includes all the content described above apart from the large full page advertisements. Full examples of these are given under the appropriate sections below.

1. 0   Item Segmentation Policy

1.1   Type 1 Content

Most items in numbers of the *Publisher's Circular* are separated by horizontal lines like this:

Horizontal line dividing items 1886

This means these can be used to mark the end of one item and the beginning of the one that follows it. However, there are a number of different types of horizontal line in the run of the *Publisher's Circular*. For instance:

Horizontal line dividing items in 1886

Horizontal line dividing items in 1880

Horizontal line dividing items in 1890

Horizontal line dividing items in 1890

However, the rule remains the same: divide into items where horizontal lines occur. This rule should be applied across all types of content, except full page advertisements and the insides of wrappers as delimited below.

As in the *Leader*, there is the problem that items that end at the end of a column or begin at the top of a column will lack any dividing lines. The same procedure should be followed as per the *Leader* so that these are marked as separate items.

1.2   Type 2 Content

Relying on horizontal lines to segment items is problematic in the *Publishers' Circular* advertising pages as many advertisements use horizontal lines to underline text or divide up content.   For instance:

Two adjacent pages from 1885, showing use of lines in advertisements

Each of the above pages is actually a single advertisement, and so we don't really want them divided any further. Therefore we are proposing to treat all full-page advertisements as a different form of content, **Type 2** content. Rather than segment these pages into items, we think it would be easier to treat each advert as a single item. This depends on Olive being able to identify advertisements, perhaps by the variety of text sizes that are frequently employed. We could also supply the page numbers upon which this segmentation policy should be applied. Below are some examples of Type 2 content:

Full-page advertisement with text

Full-page advertisement with fancy type

Full page advertisement with image

### 1.2.1 Problems with Type 2 Content

Occasionally there is more than one advertisement on a page. This regularly happens in the wrappers (see above under **0.3**), and can also occur in the advertisement sections themselves, for instance:
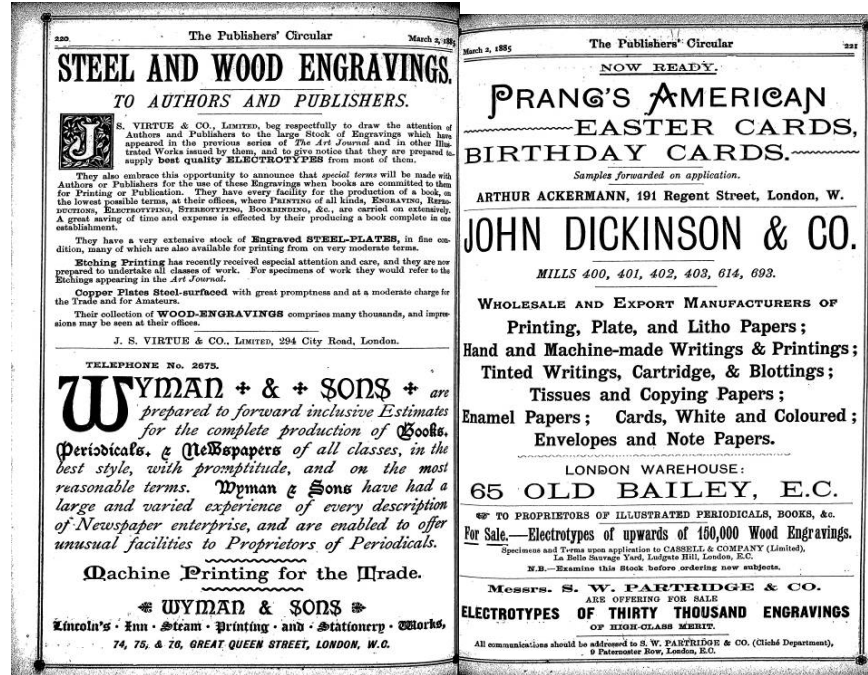
These adjacent pages from 1885 show 6 different advertisements divided by lines

Ideally, the lines that distinguish one advertisement from another could be distinguished from those that are within the advertisements. However, we know this is difficult so, if this Is not possible, we suggest treating these pages as Type 2 content, and not dividing them up.

There are also potential problems with page orientation and the title's use of double page spreads:
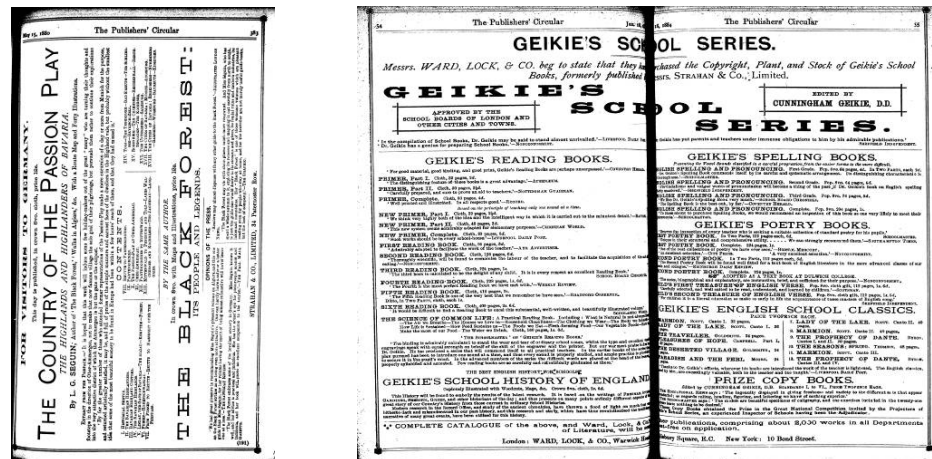
Page oriented 90° in original Double page spread in 1884

The page printed at 90° should still be segmented as a single item. The same applies to the double page spread: the item should continue from one page to the other.

2.0 Department Header Recognition

As in the *Leader*, we will try and use our segmentation policy in order to generate a ToC that only contains departments. Below we have stipulated rules that will allow us to identify which items should appear in the ToC. All other items should be called "Untitled Article" and will not appear in the ToC of the final application.

In Type 1 content department headers can be recognised by their appearance after a double horizontal rule. As we are segmenting Type 2 content as single pages, there will be no divisions and so it will form one department. Double lines look like this:

A double line marks the division between the item above and the department "Index to the Books published between February 16 and 28"

A double line marks the division between the contents and the first department, which is untitled on the page but called "Literary Intelligence" in the contents.

2.1 Exceptions

Although this is quite a regular feature throughout the run, the title does use double lines occasionally where there is no department. There are three regular instances where this occurs:

2.1.1 Dateline

The dateline on the front page is always contained within a double line. This should be treated as an item, not a department.

Front page for 6 December 1886

2.1.2 Borders and Running Headers

Every page of the *Publishers' Circular* and the wrapper of the *English Catalogue* is contained in a border, which unfortunately is made from a double line. There is also a double line separating the running header from the page. Just as in the *Leader*, the running header should be treated as an item, not a department.
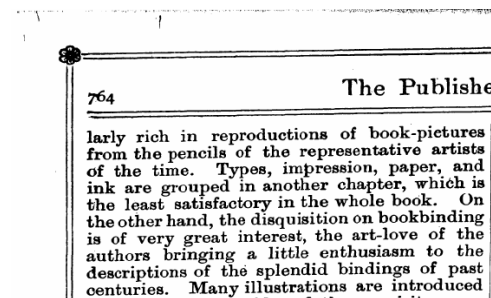
Detail from a page in 1887 that shows the double line in the border and   beneath the running header

2.1.3 Publishers Imprint

There is also a double line used to separate the publisher's imprint from the number on the last page of each issue. Like the running header, this should be treated as an item rather than a department.

Detail showing the double line that separates the imprint from the content above it – in this case an advertisement.

2.2 Irregular exceptions

Unfortunately, there are some irregular exceptions to this rule, that might cause problem for automating the identification of departments in the ToC.

2.2.1 Departments beginning at the top of a page

Although double lines always appear to mark where a department starts, this does not happen if a department starts at the top of a page. The same problem occurs with the horizontal lines that divide items and – as we saw the same problem in the *Leader* – is quite common practice in nineteenth-century periodicals.
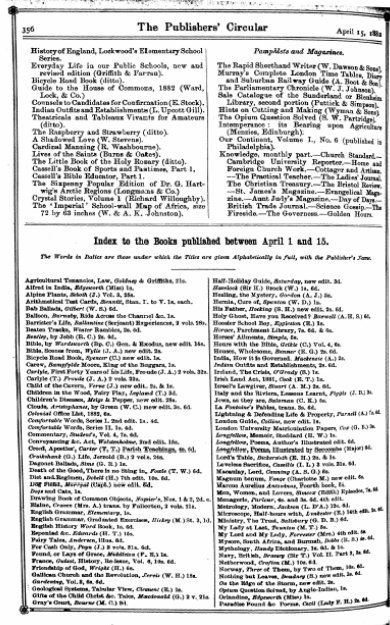
Page showing the department "Index to the Books published between April 1 and 15"
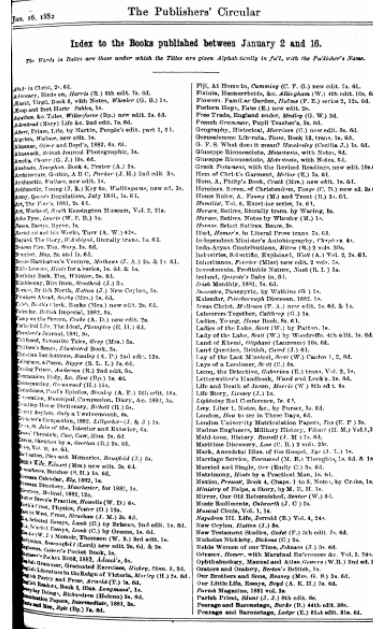
Page showing the department "Index to the Books published between January 2 and 16"

These examples from 1882 show the same department, "Index to the Books Published" in two different numbers. The one on the left has a double line separating "Index to the Books Published" from the content above; however, the example on the right has no double line as it begins at the top of the page.

2.2.2 Double lines used in Type 2 content

The double line is a typographical device that is sometimes used in advertisements. Any double lines in Type 2 content such as that below should be ignored.

Examples of double lines being used in Type 2 content from 1885 and 1888 respectively.

2.2.3 Double lines used to mark the end of a department

Double lines are used consistently in the *Publishers' Circular* to mark the beginnings and ends of departments. However, there are instances where a clear department ends (with a double line), some content is inserted as filler, then another double line marks the beginning of the next department. Ideally, this filler content should not be considered a department and so not appear in the ToC. However, we appreciate that it is difficult to register this automatically. In the example below "Books for Sale" and "Books Wanted to Purchase" are departments, and the 3 advertisements simply fill the space between them.

Double lines mark the end of "Books for Sale," but the advertisements do not constitute a department in their own right.

3.0   Pictures

The *Publishers' Circular* is the most illustrated of all our titles. Images are best handled as part of their respective items. We do not want pictures segmented separately, unless they are followed by a horizontal line like other items. By keeping pictures as part of their items, it means that the text that accompanies them provides something that can be searched upon, and that the segment in the viewer window won't have large white spaces where the picture / text was. Therefore pictures should not be treated in any special way, but segmented as part of items according to horizontal lines. As Type 2 content is segmented by page, pictures will simply be part of this larger unit. An example of a segmented page with a picture from Type 1 content is below. As you can see, the items are divided by horizontal lines, and the picture is part of the item:

Type 1 content showing segmentation

Type 2 content showing segmentation

Jim Mussell and Suzanne Paylor.