

# **A Methodology to Collect International Book Statistics**

**Framework – Indicators – Methodology & Strategies – Groundwork for a Test Run**

© UNESCO 2009

Document UNESCO CLT/CEI/CID/2009/PI/108

The ideas and opinions expressed in this work are those of the authors  
and do not necessarily reflect the views of UNESCO.

-----

**Text by Rüdiger Wischenbart and Holger Ehling**

ruediger@wischenbart.com

Information contact:

UNESCO

Culture Sector

Division of Cultural Expressions and Creative Industries

Section of Creative Industries for Development

1, rue Miollis

75732 Paris Cedex 15

France

Tel + 33 1 45 68 46 33

Fax + 33 1 45 68 55 95

[g.alonso@unesco.org](mailto:g.alonso@unesco.org)

[m.rosi@unesco.org](mailto:m.rosi@unesco.org)

## Table of Content

What for?

Why book statistics are a crucial key to sustain the quality and diversity of books.

|  |    |
|--|----|
| 4  |    |
| 1. Framework   | 5  |
| No hope for standard definitions in the future       | 5  |
| Solution strategies and practical approaches         | 6  |
| 2. Indicators  | 8  |
| Analysis and discussion of indicators                | 8  |
| Titles (number of)                                   | 8  |
| Categories (of books, by subject)                    | 11 |
| Publishers (number of)                               | 13 |
| Turnover (of publishers / publishing markets)        | 15 |
| Distribution (number of book stores)                 | 16 |
| Book sales (sales of combined retail channels)       | 18 |
| Exports (by publishers) and imports (by retail)      | 18 |
| Libraries (number of)                                | 20 |
| Library books (number of)                            | 20 |
| Library users (number of)                            | 21 |
| 3. Case studies                                      | 21 |
| Australia  | 21 |
| Africa   | 22 |
| Latin America  | 22 |
| 4. Conclusions                                       | 23 |
| A realistic, robust and cost effective strategy      | 23 |
| A sound foundation for a global model of the book    | 25 |
| 5. Next steps:                                       |    |
| Preparatory study, part 03 and part 04:              | 27 |
| A fast test run (Part 03)                            | 27 |
| List of suggested sample countries                   | 27 |
| MyBook Index or A global model of the book (Part 04) | 28 |

## **What for?**

### **Why book statistics are a crucial key to sustain the quality and diversity of books.**

Two sets of drivers of change work as tidal forces and exercise tremendous pressure on the book (and hence both on the publishing industry and book culture) as the defining format of knowledge in the global knowledge society:

The digital communication and network technologies, and  
The financial crisis of the global economy.

For all stakeholders and actors, navigating through the turbulences ahead, and also making the best of the opportunities deriving from imminent change will require tremendous skills, the courage and the luck of good leadership.

However, navigating probably rough waters will be even more daunting as for the book and for publishing, no reliable maps are available this far.

The lack of reliable – or even broadly realistic – data and analysis on the world's book culture and publishing markets has been deplored time and again. Yet, the gap has not been bridged.

With UNESCO inviting the International Publishers Association, the International Booksellers Federation and the International Federation of Library Associations to a joint effort, the chances are better than ever before to at last seriously chart the waters of the book world.

This paper argues for a realistic, robust and cost efficient way to

Collect and integrate data, as they are available, even inconsistent and incomplete;  
Organise the process in an incremental way which allows to starting quickly;  
Produce relevant preliminary findings fast;  
Provide sound foundations for a global model of the book.

A concept derived from this structured approach will also allow presenting a proposal to the main stakeholders from both governments, involved industries and concerned other stakeholders in order to raise the necessary funds to get the required work done in due time and quality.

## 1. Framework

The desire of major stake holders, including aside from UNESCO, the International Publishers Association (IPA), the International Federation of Booksellers (IBF), and the International Federation of Library Associations (IFLA), is confronted with not only a general lack of readily available data on books, publishing, book distribution and reading, but also by the fundamental inconsistencies of those data that are currently collected by various bodies in various scopes and with manifold goals, as we have shown in our previous study, "A Study on International Book Statistics".

But not only data and systematic data collection is missing. There are also no standard definitions even for the most basic items such as the book, the publisher or the bookseller available.

Even if, in the case of the book, UNESCO has issued a resolution which includes a definition of the book<sup>1</sup>, it is hardly used in trade statistics. A recent survey on the European book market coordinated by the Association of Italian Publishers (AIE) and the European Federation of Booksellers (EBF), among others, clearly showed that even within Europe, there is no agreement whatsoever on what is defined as a book. Even relevant legislation concerning the book, like VAT, does not always define the book that it is supposed to regulate.

### No hope for standard definitions in the future

Despite the general consensus that standard definitions would be desirable, we see no realistic way to create and implement them, for at least two reasons:

Current book statistics are deeply enrooted in various national cultural traditions or pragmatic trade routines or both, and we do not see any possibility to make the existing agents change their ways, as this would require that some countries or regions give up their approach only to replace them by adopting another country's or region's approach;

Digital change is about to grow its impact on the book and on publishing and book selling with great force, and one of the side effects of this change is the dilution of most current practices and entities: Almost any content can, with little effort, be turned into an e-Book, and anyone can act as a publisher, and in book retail, new books, used books, books produced on demand and, again, the emergence of e-Books heavily expand the scope of book distribution

At the same time, books as containers for content and knowledge will, for the foreseeable future, keep their privileged role at the core of the knowledge societies and for human development, regardless of how their physical appearance, production and distribution modes may change.

---

<sup>1</sup> "A book is a non-periodical printed publication of at least 49 pages, exclusive of the cover pages, published in the country and made available to the public"; "A title is a term used to designate a printed publication which forms a separate whole, whether issued in one or several volumes."  
UNESCO Resolution of 1964, p 145

Therefore, it is probably of greater relevance than ever to try to measure the field, and to analyse it, based on empirical information as good as it can be produced.

### **Solution strategies and practical approaches**

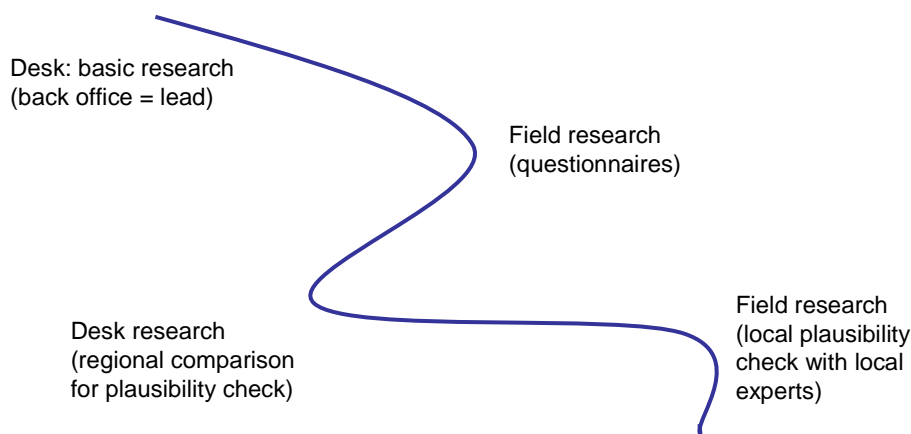
In our initial report “A Study on International Book Statistics”, we suggested already an approach that works on two levels and thereby can allow, under specific conditions,

The collection of a variety of different (inconsistent) data by a combination of desk research and field research,

The integration of data from relevant existing book statistic projects, and

A data integration and quality control process in a project centre (or back office) process of data integration.

### **Desk > field research A continuous process**



Contact: [ruediger@wischenbart.com](mailto:ruediger@wischenbart.com)

17

But even this incremental approach, while allowing the compilation and integration of a broad variety of information and information sources, does not fully acknowledge the fact that some countries will provide much more detailed and reliable data, while others will remain largely opaque.

In order to make the best of this new complexity, and to avoid to compromise the more differentiated findings for some countries with the lack of information on others, we want to outline out two complementary sets of approaches, for

Data rich countries

Data poor countries

In practical terms, we want to suggest to define, for each indicator several layers of information depth, so that one can, according to availability, get in every single country and for each indicator, those data that are available, yet in a structured way

which will allow a process of data integration in the project's back office and reporting.

The approach is modelled after a layered cake:



Taking the example of the number of titles in a given country, one can define e.g. the following layers:

Printed Book titles: new titles (data as collected e.g. by national library, book trade association, statistical institutions, etc.)

Printed Book titles: available titles in the ISBN catalogue (data as collected e.g. by national library, book trade association, statistical institutions, ISBN agency, etc.)

e-Book titles with ISBN

e-Book titles without ISBN

Various bodies of "grey literature"

Various bodies of public domain literature

Results from (a) and (b) will be available for a much larger number of countries than the rest, yet with clearly defined separate layers that allow comparisons for each layer (or depth of information).

## 2. Indicators

Based on input particularly from IPA, we consider the following indicators for the book statistics project:

Titles (number of)  
Categories (of books, by subject)  
Publishers (number of)  
Turnover (of publishers / publishing markets)  
Exports (by publishers)  
Distribution (number of book stores)  
Book sales (sales of combined retail channels)  
Imports (by retail and publishers)  
Libraries (number of)  
Library books (number of)  
Library users (number of)

We consider these indicators as meaningful and adequate, yet are fully aware that each of these indicators carries at least several issues and challenges.

We understand that these industry specific indicators can and should be completed by more general information about countries, markets, and cultural practices, e.g. demographic information, basic cultural, media and information indicators, and ideally even about reading habits. But those data will be retrieved from other, existing sources, while we focus on a core set of factors as closely related to books, the book industry and book culture as possible.

For an assessment of the indicators, and for the development of a data collecting and data integration methodology, we designed a set of Excel spread sheet templates to frame main issues and challenges for each indicator, which we shall discuss below, one indicator at a time.

### **Analysis and discussion of indicators**

#### *Titles (number of)*

"Titles" is generally used in book statistics as synonymous to "individual titles of books": A (new) title of author A is released as a book (a hard cover or paperback or an e-Book) and printed at a print run of x copies. A "new" title is the first edition of a specific title, with further editions to follow eventually.

UNESCO issued a standard definition of the book and the title in its resolution of 1964:

"A book is a non-periodical printed publication of at least 49 pages, exclusive of the cover pages, published in the country and made available to the public"; "A title is a term used to designate a printed publication which forms a separate whole, whether issued in one or several volumes."

UNESCO Resolution of 1964, p 145

However, usage of this definition is limited and can either vary in some countries, or according to different purposes (e.g. for statistical uses, or for tax collection if VAT is put on the retail price of a book). In many usages, no specific definition of a book is given at all.

A recent book market study by several European associations, published in 2007, compiled a broad number of book definitions, yet in the statistical survey which is the main purpose of the study, those definitions and the differences between them seem not to have played any significant role.<sup>2</sup>

Discussion and challenges:

The UNESCO definition has the advantage of being clear and formally simple. It is largely coherent with a common sense understanding of what is a book and a title.

However it comes with a number of challenges:

Challenges (1):

The UNESCO resolution includes with regard to "book statistics" also the following:  
Government Publications,  
School textbooks,  
University Theses,  
Offprints (that is e.g. "reprints of a part of a book");

Discussion:

For most countries, there is no realistic way to systematically assess the number of such items as government publications (and if one would include public domain information, it would be only reasonable to also include relevant reports issued by private companies and non government organisations).

However there are loose tentative efforts in some countries, to define and survey "grey literature" which is a term used for all of these book related formats of information.<sup>3</sup>

Challenges (2):

The UNESCO definition of 1964 does not refer at all to digital publications (namely database based publications, online publications, eBooks) nor does it refer to the growth of published titles due to print on demand services (POD) which together are about to very strongly expand the number of titles published in one country.

Currently, the distinction between books as items clearly limited by two covers, and comparatively 'open' formats of online publications is growingly fluid, and with POD, the numbers of titles published is growing significantly in selected markets, e.g. the USA, but also in Asia.<sup>4</sup>

---

<sup>2</sup> Cooperation Agreement: Study on the Book Market. Final Version. Coordinated by AIE and EBF, 20 June 2007, <http://www.booksineurope.org>

<sup>3</sup> For details, see [www.greynet.org](http://www.greynet.org)

<sup>4</sup> "While traditional book publishing was basically flat last year, there was a staggering rise in the reported number of "On Demand" and short-run books to 134,773, pushing the grand total for projected 2007 U.S. book output to 411,422 books. To maintain the continuity of statistics, Bowker is

E-Books open a huge field of book type publishing formats, some carrying ISBN, some not, some being sold through ordinary online book stores, but many more being distributed, with or without a retail price, from a wide range of web sites.

Many National Libraries already collect electronic and/or specifically online publications which will gain in relevance over the foreseeable future.

No viable definition of an e-Book exists at this point. The UNESCO standard definition as quoted in a UNESCO questionnaire for libraries is not applicable as it would include any larger HTML page online:

“Electronic book (eBook):

Digital document, licensed or not, where searchable text is prevalent and which can be seen in analogy to a print book (monograph).

Measurement: Count the number of purchased or licensed eBooks (titles) in the collection at the end of the year. The number of titles can be higher than the number of subscriptions, as there may be several titles comprised in one eBook.”<sup>5</sup>

Discussion:

In the near future, it will be important for any statistical survey to be able to include and represent various types of electronic publications, but at this point, change is still wildly unpredictable. Nevertheless, the indicator “titles” or “books” needs to be ready to include e-Books.

Challenges (3):

While ISBN is a good way of hinting that any given item carrying it is to be considered a book, but not all books – under any definition – carry an ISBN. In many countries, ISBNs are also used for non-book items to facilitate their distribution through the book retailing channels. It has also to be noted that ISBNs are often subscribed to in bulk by publishers, so that the relation between ISBNs issued and books published can be skewed.

Thus, there is no way to work backwards from the number of ISBN issued in a country in a given period of time, to the number of books actually published within that realm.

Discussion:

ISBN is useful, notably with e-Books, to partly define layers, but not good enough to count the number of published titles, or books.

Available sources and data (‘lowest hanging fruit’)

Despite all of the above mentioned challenges, the following data seem to exist for the largest number of countries:

---

excluding this output from its traditional reporting and has begun tracking the On Demand industry segment separately.”, Bowker, Press release, May 28, 2008, <http://www.bowker.com/index.php/press-releases/526-bowker-reports-us-book-production-flat-in-2007>

<sup>5</sup> For details and context, see [http://www.uis.unesco.org/template/pdf/cscl/Qre2007\\_Libraries\\_EN.pdf](http://www.uis.unesco.org/template/pdf/cscl/Qre2007_Libraries_EN.pdf)

New titles (that is titles published, without re-editions, in a given year in one country; this refers usually to printed books carrying an ISBN)

Books in print, or Available titles (titles published and distributed, mostly printed books, normally based on a “Books in print” database, hosted e.g. by a National library)

Copies printed (from books included in the above – even if it may be unclear whether POD copies will be accounted for as well)

Discussion:

The numbers described above are probably the most reliable and solid core data available for the largest number of countries.

However, their representative value will predictably shrink for major – mainly OECD – countries when e-Books and other electronic formats will expand just as non-traditional channels of distribution (e.g. mobile platforms) will become relevant.

Definitions of layers (draft)

The printed book on paper, according to the UNESCO definition, yet excluding government and private publications, theses, etc., and including school textbooks, POD and other paper bound publications only as far as they carry an ISBN (a certain issue will remain with books issued directly by governments that may not carry an ISBN)

Digital publications that carry the main elements of a printed book (as defined in a) and an ISBN, yet are not paper bound

Digital publications that carry the main elements of a printed book (as defined in a), yet carry no ISBN and are not paper bound and are identified and collected e.g. by National Libraries

Government or NGO or other publications that comply with the UNESCO standard definition, yet carry no ISBN and are not available through book retail channels (which will have an overlap with (c) )

*Categories (of books, by subject)*

Counting books immediately brings up questions about defining categories and sub-categories in order to understand the various sub-sets of the book markets as well as book culture.

Challenge 01:

There is no standardised system of categorisation and, what makes the issue even more complex, there are at least two lines of approaches which cater to largely different goals and expectations:

Classification systems aiming at libraries, and  
Classification systems aiming at the book trade.

Classification systems for libraries tend to have a much finer granularity, as they need to allow direct access by each title or volume e.g. in a repository; and they are based on what subjects are on offer.

Classification systems for the book trade instead need to distinguish much more broadly and in pragmatic terms what is available, in order to e.g. manage an inventory, and to better understand the interaction between publishers and booksellers on the one side, and the customers' desires on the other.

Discussion:

UNESCO builds its classification (in the resolution of 1964) on Paul Otlet's Universal Decimal Classification (UDC) of 1904/1907 (based on the US Dewey classification and the Leibnitz decimal categorization system).

UDC is "faceted classification system"<sup>6</sup>, made of 10 categories (numbered 0-9) to classify any kind of facts or items. UDC is also used by libraries as a classification system. Its top categories (0-9) are:

- 0 generalities
- 1 philosophy, psychology
- 2 religion, theology
- 3 social sciences
- 4 vacant (*philology*)
- 5 natural sciences
- 6 technology
- 7 the arts
- 8 language, linguistics, literature
- 9 geography, biography, history

The beauty of UDC lies in its clarity and simplicity, with only 10 top categories (which, for instance, are enhanced with subcategories to 23 in the UNESCO classification system).

UDC has severe shortcomings however as it reflects not only a late 19<sup>th</sup> century understanding of the world and its order of knowledge – which has clearly changed by today.

The derived UNESCO classification of 1964 does not make for these short comings in several respects as it:

Does not reflect any of the implications from today's changing user as well as product patterns (notably the consequences of globalisation, trans-cultural interactions, and of today's information and communication technologies (ICT));

Does not allow to properly classify at least two segments that are very prominent in any trade oriented classification system: educational (text) books and children's books;

Is too complex with the 23 categories it foresees.

---

<sup>6</sup> [http://en.wikipedia.org/wiki/Faceted\\_classification](http://en.wikipedia.org/wiki/Faceted_classification)

On the other hand, any selection of classification systems as used by the book industry are less coherent, or more complex (e.g. with often even more categories than UNESCO's), or both.

For the sake of simplicity, Annex 01 shows only a comparison of only three, the English language "BISAC Subject headings", the French "Secteurs editoriaux", and the German "Warengruppen", yet comparable systems of classification are used in all major markets with more or less the same basic structure

Solution strategy:

A comparative analysis of the trade classification systems, and of available trade statistics overview points nevertheless to pragmatic approaches or possibilities for a work around that may be good enough at least initially – hoping that in the long term, conversion keys at least for a few main categories can be developed.

In detail: The colouring in Annex 01 highlights related categories, which can be discussed as follows, (yet this absolutely requires verification with actual market data, which has not been part of our remit):

Green: Generalities (UDC 0): Coincides perhaps with general trade non-fiction,;  
Yellow: Pale yellow covers largely "education" and children (and UNESCO suggests to introduce accordingly sub-categories (a) and (b) for school textbooks, and children's books<sup>7</sup>; however, in UDC, this coincides with the much broader "Social Sciences" (dark yellow)

Mauve: Arts, including new forms of illustrated books, e.g. graphic novels;

Pink: Literature / fiction: Includes "linguistics and philology" in UDC, however, the number of titles on linguistics and philology will hardly produce a serious distortion of the overall picture.

Definitions of layers (draft)

The approach as lined out above should be good enough for at least a general break down to assess estimated volumes for the following categories:

General trade fiction

General trade non-fiction

Children's books (yet with inaccuracies with regard to the inclusion / exclusion of juvenile / young adult reading

Education

Science

*Publishers (number of)*

For definitions of what is a publisher, much applies of what had been said about the definition of a title and a book.

---

<sup>7</sup> See UNESCO Resolution of 1964, page 145: "School textbooks and children's books already identified in the above-mentioned subject groups should also be counted separately in the two following additional groups: (a) school textbooks and (b) children's books."

Everyone who publishes a book can (or even has to be) considered a publisher. With new production and distribution technologies, the number of actors will increase dramatically, as will the number of titles produced, and the number of actors who publish only coincidentally.

There is no way of drawing a solid line in order to separate a core group of (e.g. professional) publishers from all the others.

#### Challenge 01

Every person is potentially a publisher. In the USA, most of the European Union and in a growing number of countries, this is already a reality.

#### Discussion

There are few conceivable ways to limit this number.

Counting the number of recipients of sets of ISBN. However this is a hugely inflated number, as many individuals or institutions acquire one or sets of ISBN, yet do not publish on a continuous basis.

Counting the number of members of a country's publishers association or related bodies. This is in many cases most likely the most realistic approach. But even so, many occasional publishers may be included, while all who, for whatever reason, publish outside of the ISBN, or who for whatever reason chose not to be a member of the national association will be excluded, and many countries actually have several associations, namely in countries with several e.g. languages and culturally distinct regions.

#### Challenge 02:

A very limited number of publishers, because a publisher needs a state license to publish.

#### Discussion

In a number of countries (e.g. PR China), publishers need a license granted by the state (in PR China, only 578 companies currently hold such a license). At the same time, thousands of organisational entities of various kinds – sub-organisations, or even independent privately held companies – pursue activities that include almost everything of a publisher's activities, with the exception of having its own allocated ISBN, which instead is provided by one of the state controlled companies.

So there may well be an exact number of 'publishers' in such a country, this provides no way to even estimate the number of actors (or agencies).

#### Solution strategy:

Define sub-sets of actors which are each clearly defined so that they are comparable between countries and allow consolidation.

## Definitions of layers (draft)

Members of representative publishers' associations in a given country (where 'representative' may be understood by e.g. being members of regional federations, or even by representing members with a combined minimum market share of e.g. 10 percent of a publishing market.

Recipients of at least 10 ISBN for the current or the previous year.

Recognised actors in electronic book type publishing (it is probably too early to be defined more in detail as of now, but is a place holder for an additional category/ layer in the near future.

## *Turnover (of publishers / publishing markets)*

For many countries, industry statistics are broken down in a way, that allow to assess turnovers for the publishing industry. However, the challenges mentioned above from the lack of a clear definition of what is a publisher, result in a broad spectre of inconsistencies when it comes to measure the industry turnover.

## Challenge 01

Generally it will be necessary to distinguish between turnover data that

Include or exclude VAT (mostly, industry statistics are excl. VAT, but this needs to be checked);

Represent publishers sales (as opposed to sales through retail channels or consumer spending, as this comes with a difference of up to 50 % from the retail price of books due to inclusion or exclusion of retail discounts);

Include or exclude direct distribution of books or other published materials through other channels, e.g. educational material produced and disseminated directly through government agencies;

Include the distribution of a significant number of imported titles through the book retailing channels (Latin America, Commonwealth, Francophonie, former Soviet Union states)

The importance of those clarifications mentioned above will even grow, as direct sales by publishers, but also new, non-traditional distribution channels increase dynamically, both for printed publications as well as for electronic publications.

## Challenge 02

National as well as international industry statistics differ considerably in what types of publishing they include or exclude in their break downs, notably

Book publishing (through publishers, printed on paper or electronically, yet always with an ISBN attached);

Learned journals and professional periodicals;

Various kinds of digitally born and distributed information (professional and scientific information, wire services, and other database based information online and, to a lesser degree, offline);

Magazines;

Directories;

Public domain publications and various grey literature.

In addition, it will not always be clear to what extent publishers' turnover data include or exclude exported publications, which can make sensible differences in the case of countries with a high amount of cultural exports – e.g. the UK, France, Germany, USA.

## Discussion

Despite of these challenges, the approach chosen for this methodology, to work with layers – or subsets for each indicator – will be helpful once again to obtain comparable and realistic data for the turnover of publishers – and in similar ways for book sales (see below).

### Solution strategy:

Define sub-sets of segments in the publishing industry which are each clearly defined so that they are comparable between countries and allow consolidation.

### Definitions of layers (draft)

It will be required to build a matrix to represent both challenges as described above:

#### Types of data:

Incl. / excl. VAT

Publishers' sales

Other sales

Types of publications

Books by publishers with ISBN

Learned journals and professional periodicals;

Digitally born and distributed information that is comparable to books (see definition of title)

Public domain publications and various grey literature.

### *Distribution (number of book stores)*

What has been discussed in terms of complexities for publishers' and their turnover applies even more strongly for retail.

### Challenge 01

A recent European study coordinated by the Italian Publishers Association AIE and the European Booksellers Federation in 2007 clearly stated

“It is virtually impossible to give the number of publishing houses or of bookshops throughout

Europe. In the case of Germany for example, all the publishers are members of the Borseverein des Deutschen Buchhandels (2000) [which actually is not the case, rw/hge] In France though, the Syndicat National de l'Édition represents approximately

400 member companies who combine the most important part of the business and turnover of the French publishing industry.”<sup>8</sup>

The picture is even more challenging than this statement emphasizes as books are not only distributed through traditional retail channels which may be represented by national booksellers’ associations, but, to a growing extent through

Department stores

Non-book retail channels

Direct sales of publishers

Online book retail channels which often are not represented in available book sales data, and often are organized on an international scale and, in many cases are difficult to track.

In most EU member countries as well as in the US, book sales through traditional retail channels have dipped below or are hovering around the 50 per cent barrier. In addition to these more or less traditional and in their majority professionally organized retail channels, in different countries widely diverse ways have been developed in order to distribute books, and many of these escape any accounting, notably

Dispersed and fragmented (often ‘private’) online sales (e.g. eBay);

Local and national book fairs as centres of distribution

Street vendors

Used book channels (online and offline and combinations of those) which represent a growing share of the market.

## Discussion

It is clear that in most cases, it will, if at all, only be possible to collect or retrieve data for the first list of predominantly professionally organised retail channels.

For the many countries where private and/or non-traditional ways of distribution prevail, it will be advisable and possible though to have text descriptions about how books get to the readers, in order to have a fair understanding of cultural patterns and market structures.

This will be even more important for an overall understanding of book distribution as electronic distribution of books and book type information will spread in the years to come. However, such direct dissemination of content online will, in many cases, escape the traditional definition of a “book store” or outlet.

For data rich countries, instead, a much finer grid of information and structures will be available, with informative computations such as numbers of various types of outlets per population.

## Definitions of layers (draft)

---

<sup>8</sup> Cooperation Agreement. Study on the Book Market. Final Version. Coordinated by AIE and EBD, <http://www.european-writers-congress.org/upload/mcastudy.pdf>

It will be required to focus on layered definitions which represent largely the traditional retail channels for printed books, and eventually, if necessary, add additional layers as data and insight builds up:

Book retailers who are members in a national or regional association, broken further down – at least for data rich countries – into

Chain book stores

Independent book stores

Book stores in other retail structures, including department stores

Online retailers

Other, non-traditional retail channels.

### *Book sales (sales of combined retail channels)*

Much of what has been discussed with regard to the number of book outlets explains why an overall number for book sales will always come with a significant degree of inconsistency.

However, as the statistics project will evolve, and data are available and analysed for more and more countries, it will be reasonably conceivable to develop computations to produce solid estimates even for data poor countries, notably by computing and consolidating title productions, publishers' turnover and imports.

For estimates in OECD countries, in addition of the data sources discussed here – notably public industry statistics – it is desirable as well as realistic to receive, for research purposes, also data from commercial market survey provided by companies like Nielsen or GfK, and they will allow to significantly calibrate the estimates.

### Definitions of layers (draft)

It will be required, just as for the number of retail outlets, to focus on layer definitions which represent largely the traditional retail channels for printed books, and eventually, if necessary, add additional layers as data and insight builds up:

Book retailers who are members in a national or regional association, broken further down – at least for data rich countries – into

Chain book stores

Independent book stores

Book stores in other retail structures, including department stores

Online retailers

Other, non-traditional retail channels.

### *Exports (by publishers) and imports (by retail)*

Exports of publishers have the pragmatic advantage that they usually appear in detailed government statistics and, even more so, that the main exporting countries

tend to be few in number and rather data rich and interested in their own interest to produce these figures.

So it is realistic to build a fairly representative basic grid of data from export statistics of the various global and regional main export countries and work from there.

#### Challenge 01

Government statistics have not necessarily the same break downs – detailing e.g. printed books – that are required for the book statistic project. Hence the same inconsistencies apply as have been discussed for the turnover of publishers will apply.

#### Challenge 02

Theoretically, one may expect to be able to match export figures with respective imports between given countries. Even between data rich and strong exporting countries such as the USA and UK, this is not the case, as a limited test sample has clearly shown. This is probably due to two sets of differences

Differences in definitions and break downs as mentioned in Challenge 01, which would result in inconsistencies in the numbers between exported and imported items; Differences between exporting and importing agents, as discussed above with regard to challenges for publishers' turnover and retail's sales figures, which, when combined, would necessarily result in inconsistencies in the numbers between exported and imported values.

Nevertheless, even if absolute figures would differ, the underlying pattern should match in many cases.

#### Discussion

Even as a simple matching procedure – which would also include the ample UNESCO data on the global exchange of cultural goods – seems not realistic, it seems feasible to start building a global model of the flows of books.

This model of the flows of books, based on export and import data, is of particular relevance as it is probably the strongest toolset for measuring and analysing the global exchange of books, and the basis for any global map and analysis of the book world, hence key for the desired overall assessments and policy statements in a future “Global Book Report” which is the strategic goal of this book statistics project.

#### Solution strategy

By the fact that exports and imports – just as translations – follow the strictly hierarchical pattern of a cascade, with few countries as the main sources for exports, and many countries in a overwhelmingly receiving role,

As a consequence, a top down approach should be chosen to build a global dynamic model and, as a result, a map of book flows:

Identify at first the top 10 exporting countries and their export statistics;  
Compare to what extent figures, or at least underlying patterns, between exports and imports in those top 10 countries match;  
Compare and calibrate these statistics with UNESCO's own data on the trade of cultural goods;  
Do the same with the 10 sample countries picked for the test run of the book statistic project;  
Work further down, region by region, for the rest of the world.

### *Libraries (number of)*

Prima facie, the question of what is a library seems a simple one to answer. Compared to the publishing and retailing sectors, the number of agents in any given country is limited and, because of the strong external funding requirements, should be traceable by looking thoroughly at the accounts and budgets of national, regional and local funding institutions. IFLA and its national members also tend to be on the safe side when it comes to distinguishing the major types of libraries that are relevant for our survey, namely:

public libraries – open to any citizen

school libraries – exclusive to students and teachers in primary and secondary education

academic/research libraries – exclusive to students and teachers in tertiary education and further research.

Most national library institutions have for some time now included electronic publications in their counting systems and have devised methods to distinguish and number them in order to make them available to library users.

Classifications of books in the individual library holdings have for some time been based on universal systems, such as UDC. This will make the analysis and comparison of data much easier.

However, as pointed out before, the classifications used by libraries to record their holdings tend to differ significantly from the classifications used by publishers and retailers. To consolidate these methodologies will constitute a significant part of the work that has to be done to reach adequate findings within the scope of the survey.

While we are confident that with the cooperation of IFLA and the national library associations, basic figures for the number of libraries in the individual countries can be established quickly, we need to issue a caveat. During our preliminary "test drive" we found that data can be generated for countries in Africa and Latin America much easier than for Europe – a finding that owes to the fact that in Europe, libraries are more often than not run by local or regional entities, with figures not aggregated on a national basis, while in Latin America and Africa, central organisations seem to be the norm, which makes access to data much easier.

### *Library books (number of)*

Much of what has been said for the number of libraries also goes for the number of titles held in them. While some national statistics for library holdings are available, in many countries these numbers will have to be generated through field and desk research by aggregating numbers from local and regional institutions. Once again, the basic separation of electronic and paper-based holdings will have to be taken into consideration, and what has been said about the problems of defining “book” is at the heart of the challenge we are faced with.

### *Library users (number of)*

The number of library users can only be gleaned from the numbers of registered users in the various types of libraries. Especially for school and academic/research libraries, the numbers of registered users and the data for actual library usage will be very much at odds.

A solution may be to take a look at the number of titles lent from the libraries and relating them to the number of registered users. Thus a fair assessment of library usage may be gleaned.

## **3. Case studies**

### **Australia**

In Australia, book statistics for the publishing and retailing sectors are put together by the Australian Bureau of Statistics, normally on a bi-annual basis.

While this approach lacks closeness to the industry players, it certainly provides for a largely unbiased way of going about the matter, and data are gleaned from all sorts of sources, ranging from tax receipts to customs documentation.

The ABS method does not use a detailed classification system such as the UNESCO resolution or UDC, but rather simplifies the matter by using the following analysis tree:

Layer One  
Electronic  
Paper

Layer Two (Books only)  
General (hardback, paperback, mass market)  
Textbooks (primary, secondary, tertiary)  
Children’s Books

This simplified system may not *prima facie* lead to the depth of insight into the structures of the Australian market, reader’s behaviour etc. that a more detailed approach would offer, but it comes with a compelling argument: it is easy to implement and non-controversial

## **Africa**

Africa and the 46 sub-Saharan countries can be deemed as the prototype for a data-poor region.

With the exception of South Africa, national statistics for the book industries and its sectors are either not compiled or are seemingly without relevant relation to the economic truth. From our knowledge and recent discussions with industry participants on the development of the publishing industries in these countries we conclude that the industries as such are growing, especially in the educational sector, but there are no tangible data to back up this assessment.

The most recent serious approach to collecting book industry data on a continental basis was undertaken by APNET and ADEA at the turn of the century. For this, the national publishers' institutions were issued with a detailed questionnaire, which asked for data on a range of 29 topics, from demographic to taxes. Just over 20 countries responded to this questionnaire, and a qualified follow-up concerning data research or training of respondents was not undertaken.

This has led to the ADEA/APNET report being seriously flawed, and an objective assessment of the continent's publishing industry is still a desideratum.

The lack of reliable data has been a negative factor constraining both the development of the individual publishing industries as well as the private and public (donor/NGO) investment in the sector.

## **Latin America**

Latin America is the subject of research for the most ambitious regional UNESCO driven book industry statistics project.

Through the work of CERLALC, based in Columbia, a number of surveys and diagnostics of the state of the book industries in Latin America have been undertaken, and CERLALC has also issued a thorough paper on its methodology.

While the work of CERLALC cannot be praised highly enough, it is obvious from the analysis of the most recent diagnostic paper (for 2006) and related national reports that the results tend to show large inconsistencies: be it the number of publishers in countries such as Mexico and Columbia, or the number of titles, it seems that the CERLALC method has yet to be adopted in most countries.

One major reason for the lack of backing for the CERLALC method in Latin American countries seems to be the basic questionnaire which is being sent out to the national partners, which asks for data on more than thirty points, and which seems to strain the information gathering resources on the ground beyond the point where useful data can be compiled.

## 4. Conclusions

At a first glance, this paper, just as its precursor in June 2008, may appear as a desperate list of difficulties rather than a treatise of methods and solutions to collect and compile international book statistics.

Nothing could be more wrong.

### **A realistic, robust and cost effective strategy**

What we argue for is in fact a solution strategy that is simply realistic in the acknowledgement of the complexities in the goal of collecting and integrating book and publishing statistics on a global scale, and the proposition of a methodology that, as realistically, encourages a start with findings that are at arms' length, but from there engage in a learning system that quickly will allow to draw a map of the global publishing industry and of book culture as data are coming in and are sorted out.

We strongly emphasize that the following basic indicators are a valid starting point and good to organize a vast amount of findings and insights on the global publishing industry and the culture of the book.

These indicators are

- Titles (number of)
- Categories (of books, by subject)
- Publishers (number of)
- Turnover (of publishers / publishing markets)
- Exports (by publishers)
- Distribution (number of book stores)
- Book sales (sales of combined retail channels)
- Imports (by retail and publishers)
- Libraries (number of)
- Library books (number of)
- Library users (number of)

We also stress that even incomplete data on these indicators per country or other entity will allow meaningful analysis, and in the process of aggregation of findings between countries and regions across the globe will quickly evolve into a tightly woven grid that allows a mapping of the world's book and publishing landscapes.

For each indicator, a number of sub-definitions, or 'layers', will be defined (and re-fined as the statistics project further evolves), allowing to identify and "tag" (or reference) each set of data as it comes in.

|              | Layer 01 | Layer 02 | Layer 03 | Layer 04 | Layer 05 |
|--------------|----------|----------|----------|----------|----------|
| Indicator 01 | x        |          |          |          |          |
| Indicator 02 | x        |          | z        |          |          |

|              |   |   |   |   |  |
|--------------|---|---|---|---|--|
| Indicator 03 | x | y | z |   |  |
| Indicator 04 | x | y | z | o |  |
| Indicator 05 | x |   | z |   |  |
| Indicator 06 | x | y | z | o |  |
| Indicator 07 |   | y | z | o |  |
| Indicator 08 | x |   |   | o |  |
| Indicator 09 |   | y |   |   |  |
| Indicator 10 | x | y |   |   |  |

As a result, each subset of data, from

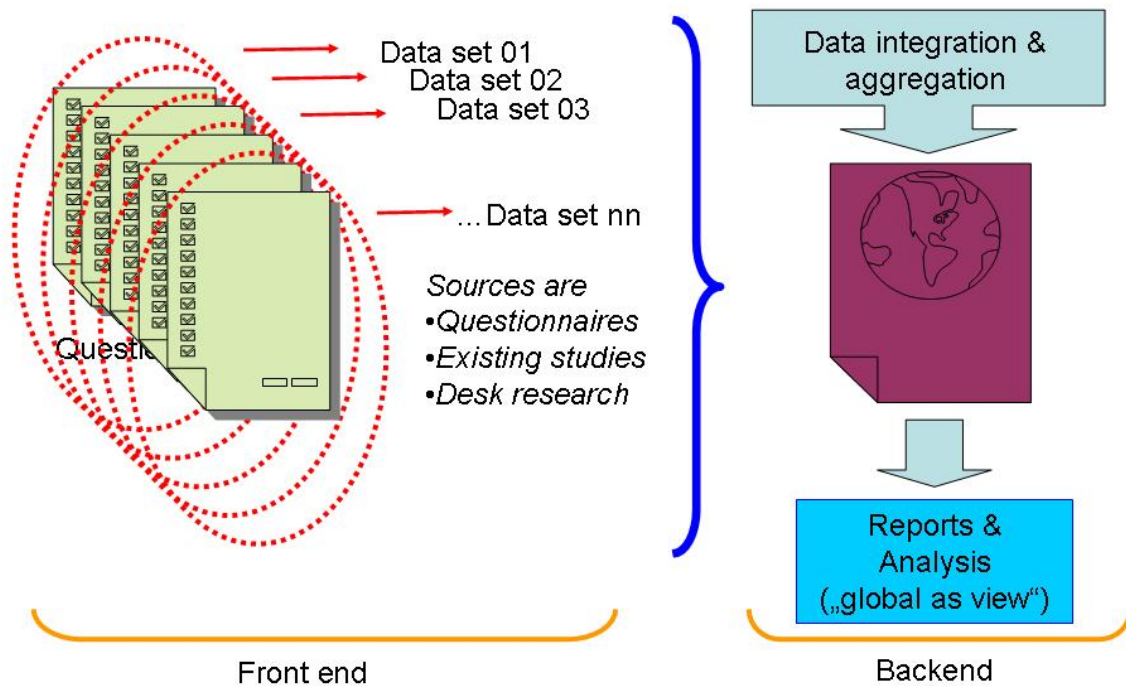
Existing statistics projects;

Desk research (or desk driven benchmarking); or

Field research (notably by specifically sent out questionnaires to national or regional actors)

can be qualified (or tagged) and therefore integrated into a systematic and incremental process.

## Data integration and data merging



Contact: [ruediger@wischenbart.com](mailto:ruediger@wischenbart.com)

27

In this perspective, we need to stress that

Much more data, complete and incomplete, more or less representative, are available out there than one would suspect;

Those data are certainly inconsistent and incomplete, but a structured and calibrated approach as outlined above allows to qualify, tag, and in the end integrate these data in an elaborate and systematic process (which makes our approach robust as it allows to working with what one can get);

It is an incremental process, so even first rough or incomplete sets of data return already telling results – and are, on top, valuable as benchmarks for subsequent project phases;

It is an integrative approach, as it strongly relies on taking in (and calibrating) results from other, already existing statistical projects;

Field research and questionnaires are only one track for the collection of data, which can be completed, and fact checked, for quality assurance, with various other kinds of data and structured information;

Cost efficiency and immediacy of first returns are a given, as opposed to any approach that depends exclusively on its own and exclusive research.

### A sound foundation for a global model of the book

The ultimate goal of an international book statistics project is certainly to develop a global model of the book (of the publishing industry and of book culture).

These statistics and subsequent analysis will form a sound foundation for such an undertaking.

Even if it is beyond the scope of this paper which is focused exclusively on the methodology for collecting and integrating data, we can hint at the fact that the resulting data and insights are crafted in ways to allow to

Describe (or map) book production and distribution as well as book culture within geographical entities (countries, regions, world), and to  
Relay (or model) the interaction between domestic book industries and cultures with global trends and patterns.

Details of such a modelling concept will be dealt with in a separate paper.

## **5. Next steps: Preparatory study, part 03 and part 04:**

After the initial study on an international book statistics project to UNESCO, IPA, IBF and IFLA in Paris in June 2008, this paper singles out indicators and proposes a methodology for collecting and integrating complex and largely inconsistent sets of relevant data.

In a next step (part 03), we propose to start a test run with actual data for ca. 10 countries with the ambition to aggregate and compute as many data as one can get.

### **A fast test run (Part 03)**

This fast test run should

Be largely based on desk research, completed with simple queries to relevant actors in these countries to input data as they readily are available;

Start a structured discussion on data integration approaches with independent data experts, and identify ca. 3 comparable case studies for the integration of highly complex and inconsistent data from other fields as “best practices”;

Work out a detailed working and budget plan for a complete run of the book statistics project over a project duration of ca. 3 years;

Work out a finance plan, notably by identifying potential sponsors (and stakeholders), and detail both their commitments and their benefits.

At the Paris meeting in June 2008, we suggested already a list of countries for such a test run.

The goal was to identify countries as “low hanging fruit”, where we already had a realistic assessment of the data situation. In this perspective, we deliberately accepted a significant number of OECD countries as compared to expectedly ‘data poor’ countries, as we clearly aim at results that both need to be representative on a global scale, and yet produce as data rich results as conceivable in order to check and refine our methodology.

#### *List of suggested sample countries*

Australia  
PR China  
Mexico  
Columbia  
France  
Germany  
Czech Republic  
Sweden  
UK  
Nigeria  
South Africa

Of course, this country list is subject to further debate and calibration, but one should bear in mind that the key target of the fast test run is to produce relevant data and fast results.

### **MyBook Index or A global model of the book (Part 04)**

All these statistics and insights will only matter if they allow to developing a “global model of the book”, or, some sort of global “**MyBook Index**” (as mentioned already in the Paris presentation in June 2008) which allows to having significant comparisons between national, regional or even global book industries and book cultures.

Again, this perspective confronts not only statistical barriers, but much more complex conceptual oddities.

How can one reasonably compare countries of not only extremely diverse wealth, but also of various cultural or political systems, with perhaps largely diverging cultural, political or economic strategies?

If the ambition is to assess more than the blunt availability of containers (or books) in a given territory, but to also allow measuring cultural knowledge interdependence through books, one will need to correlate the respective domestic structure around the book with the global book market and its overarching dynamics.

The methodology elaborated in this paper and the resulting data will allow the creation of such a global model of the book, or a “MyBook Index”.

Vienna and Frankfurt, in October 2008