

E-books, e-lending and more

- Setting the scene
- The e-book market
- Acquisition models
- Our acquisition policy for e-books
- What do we buy? What we subscribe to?
- How we expose them?
- Demand-driven acquisition
- Outlook
- Epilogue: text and data mining

Tullio Basaglia, Library Services

CERN Scientific Information Service

16/4/2015 - T. Basaglia - GS-SIS



Setting the scene: CERN

- 21 members countries
- More than 100 countries represented
- 2,400 staff + 10,000 particle physicists
- 98% of the physicists are not staff: nomadic community
- physicists, but also computer scientists, engineers, administrative staff
- “Early adopters”
- Used to non-intermediated access to information
- “Open Access culture”
- Prepublications are e-only since the mid-Nineties. Our subscriptions to journals are e-only since Jan. 2008

The e-book market

- Few actors, worldwide publishers: some are owned by investment groups, which are heavily business-oriented.
- Concentration due to mergers and acquisitions – monopolistic market.
- Competition between publishers and suppliers who are aggregating content from multiple sources.
- Strong pressure on libraries to buy packages, instead of doing a selection.

Acquisition models

- Perpetually purchased content is *usually* available in the form of a DRM-free PDF file (access control is done at the Institution level, IP-number based)
- ‘virtual loans’ consist in a non-permanent download of a DRM-protected PDF file, which also implies some limitations in copying, sharing, printing etc. Subscribed content is in general DRM-protected
- Acquisition model A:
 - Perpetual purchase
 - “Pick and choose”
 - Package: all or nothing
 - “e-lending”
 - Pay-as-you-go
 - Flat fee subscription
- Acquisition model B (the future?): Demand-Driven Acquisition

Library does the selection

Our acquisition policy for e-books

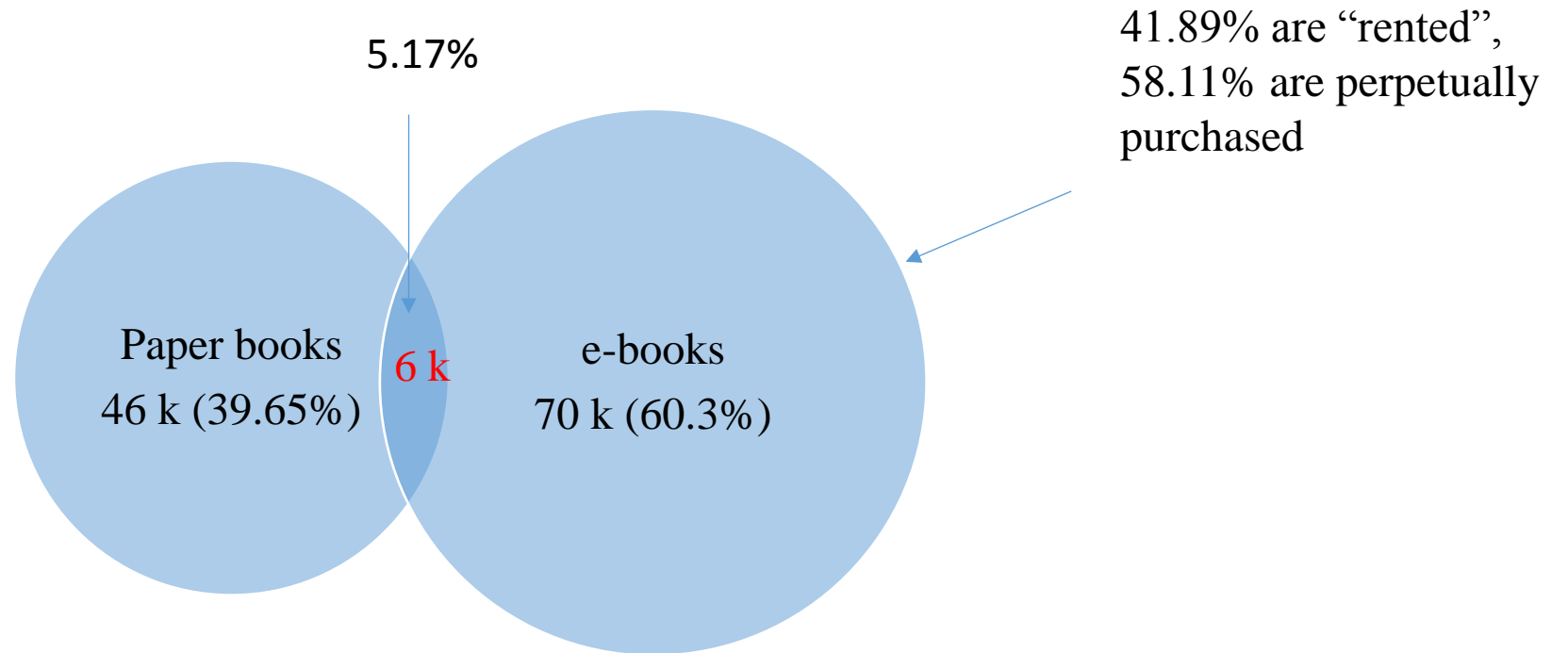
- “Content [=quality and subject coverage] is king” when selecting a supplier
- Formats available (PDF, Mobi, Kindle etc.) are a key aspect when selecting a supplier. Is a plugin needed to read them? Are they DRM-protected? Compatibility issues (content <-> support)!
- Perpetual purchase of e-books from the publishers in core subject areas, complemented by pay per usage access in other subject areas
- Avoid as much as possible ‘package purchases’ (Springer is an exception!), pick-and-choose preferred
- Purchase a title on paper **and** e-book if it is on a ‘core subject’ or by CERN author(s), otherwise we tend to rely on the e-book version
- E-book is the preferred format for IT titles. Paper bought if there is an explicit request
- ‘Books for consultation’ are purchased – if there is an explicit request – e-only

What do we buy – what do we subscribe to?

- We **buy** perpetually **core subjects’ titles** from a few key publishers: Springer (book packages: P&A, Mathematics, Engineering), and from others on a ‘pick-and-choose’ base: CUP, OUP, T&F, Wiley, World Scientific...etc.
- For **non-core subjects’ titles**, we **subscribe** to two aggregators’ platforms: EBL (Ebook Library = Proquest), www.ebllib.com and Safari proquest.tech.safaribooksonline.de (Proquest!).
- Access: on EBL, the end-user can browse for 5’ or 10’, then s/he has 2 options: stop browsing or do a 2-weeks ‘virtual loan’ (=non-permanent download of a DRM-protected PDF). Copying and printing limitations.
- *Note:* on EBL, the reader could also trigger the acquisition of the ebook. We do **not** use this functionality. We prefer to **acquire** the ebook from the publisher.
- Access: on Safari, online reading **only** is possible. No copying, printing limitations.
- For EBL, we pay a low platform fee per year, and we pay per usage (=a fee is charged at every 2-weeks ‘virtual loan’, 10-15% of the list price). Monthly invoicing.
- For Safari, we pay a platform fee

CERN Library books collection

- On the whole, 116,000 book titles: 70,000 e-books and 46,000 on paper. There are 6,000 hybrid titles.



From acquisitions to collection development

- Dichotomy between the efforts towards a sound, balanced development and the pressure to buy ‘en masse’ coming from the market
- Advantage of the subscription-based platforms (EBL and Safari): we can expand and reduce our collection indefinitely, especially in ‘non-core’ subject areas (e.g. engineering, materials science, some sectors of physics), where many small communities have very differentiated needs.
- We are a research library. We (also) need to liaise with experts. Different paths exist: email, web forms, personal contacts... However, ‘discovery patterns’ need to be created (no printed publisher’s catalogue circulation!), that should be adapted to the way our users work
- To achieve this goal, we have to develop our own tools, because we cannot rely on services provided by traditional book vendors/agencies (we are not their clients)
- Constraints:
 - Some readers still prefer the paper version
 - DRM is not compatible with some operating systems, Adobe Digital Editions does not work under UNIX

How we expose them?

- We import records into our online catalogue. No manual cataloguing, no ‘list of online resources’. The record is on cds.cern.ch (=our online catalogue), and the links point to the publishers’ or aggregators’ servers.
- Ingestion of book records into our catalogue is not painless...
- We periodically **import** records in batch, from the publishers’ or aggregators’ web or ftp sites.
- We import **all** records we buy, we **selectively** import records from aggregators (subject profile defined at the aggregators’ end).
- We **convert** the records (generally available in XML) into MARC.
- We run a script to **match** the imported records against **already existing records** in CDS, because we want a **single record** for paper and e-book.
- We **delete** records from aggregators, when we detect that there is no usage, except for core subjects’ titles.

Can we do better to expose them? How?

- **Enrichment of records** -> book TOCs and abstracts are imported from external sources: ebook suppliers or external library catalogues (27.500 titles)
- **Visibility:** digital signage and book reviews on the CERN Courier



- ‘Crowdsourcing of acquisition proposals’ (=Demand-Driven Acquisition) depends heavily on enhancements of the **layout of the records** (=e.g. virtual browsing of shelves), or of **search hits** (=facets, subject tree)
- The librarian has still a role to play!

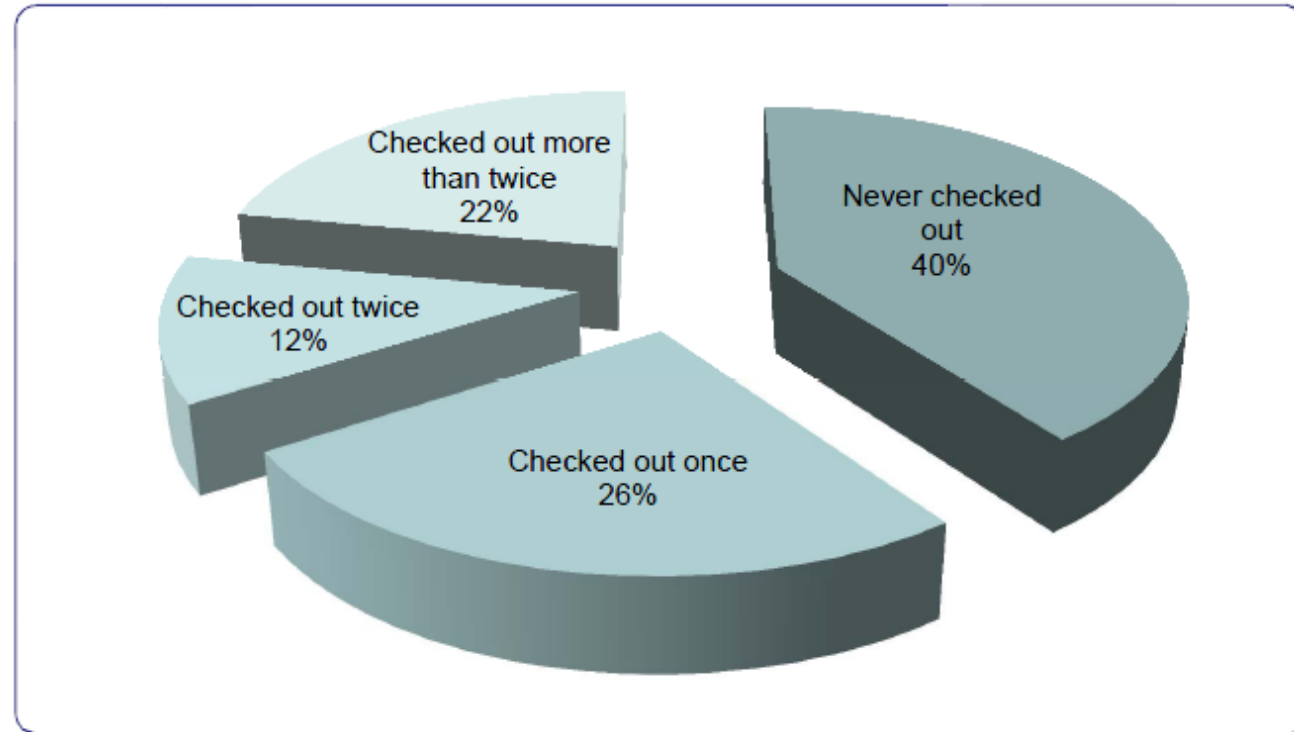
Demand-driven acquisition: why?

Publishers

Communication

Group is a sales and marketing consulting firm. They ran a survey among 74 US libraries in March 2010, asking which percentage of their books bought in 2005-09 was borrowed

Standard circulation data (4 years)



Data from D. Banush, Brown University

Four year circulation data: Approvals and Firm Orders 2005 -2009



Demand-Driven Acquisition: how does it work?

- Also called patron-driven acquisition
- Model usually linked to e-books acquisition, but it could also be applied to print books acquisition
- Aims at optimizing resources and shaping collections on the base of users' needs. Collections can be hugely expanded, because a very low flat fee 'per annum' allows the importation of an infinite amount of records
- Spend money only on books that will be used
- How does it work?

Book metadata imported and exposed → user discovers the e-book

S/he triggers the purchase by the library

Note: on EBL, the reader could also trigger the acquisition of the ebook. We do not use this functionality. We are experimenting DDA with paper books ...

Experimenting with DDA

- We periodically import records from several Amazons. The importation derives from a subject-based query
- The imported records go to a special ‘collection’: *book proposals*
- The end-user will be able to suggest the acquisition
- We will approve it, trigger the acquisition and transform the record into a ‘normal’ book record, or we will refuse it, proposing e.g. an ILL

Book Proposals

Search 1,301 records for:

Search

[Search Tips](#)
[Advanced Search](#)

[Add to Search](#) +

Latest additions:

2015-03-29

00:27



Building the H bomb : a personal history / Ford, Kenneth W.

World Scientific Publishing Co Pte Ltd, 2015. - 250 p.

[This book at Amazon](#)

[Purchase it for me!](#) - [Suggest for library](#)

[Detailed record](#) - [Similar records](#)

DDA: caveats

- As said, it requires effective strategies to promote discoverability, but also effective mechanisms of budget control!
- When selecting a platform, the flexibility of the administrative interface is crucial. On EBL e.g., we can:
 - set up a cost threshold for the virtual loans (=a loan should not cost more than x). The threshold can be 0 (=all loans require approval from the librarian) up to “no limit” (=no approval needed). This can be modified at any time
 - We can select the length of virtual loans
 - We can monitor at any given moment the usage of the collection

Open Access Ebooks

- SpringerOpen Books:
 - Books, conference proceedings (~50 titles at present)
 - OA fee are charged. Authors at member institutions get a 15% discount on the OA fee.
 - An e-book is produced and, if one wishes, a print on demand book can be purchased
 - Books are published under the Creative Commons Non-Commercial (CC-BY NC) license
 - They are on the Directory of Open Access Books (<http://www.doabooks.org>)
- De Gruyter (€ 10,000 OA fee + royalties on print sales), Taylor & Francis
- Proceedings of conferences – a crucial publication outlet in Part. Phys.:
 - EPJ Web of Conferences (EDP Science)
 - Journal of Physics: Conference Series (IOP)
 - Joint Accelerator Conference Website (JACoW)
 - Nuclear and Particle Physics Proceedings (Elsevier)
 - Physics Procedia (Elsevier)
 - Proceedings of Science (PoS) (SISSA) - GS-SIS



Outlook

- 24/24 7/7 365/365 access and remote authentication are necessary (=nomadic community) but not sufficient; increasing demand for barrier-free and easily ‘portable’ e-books
- Today: smart devices, dumb content
- Future: “layered” content (multimedia integration, data integration)?
- E-lending for research libraries through Amazon?
- The main challenge consists in leveraging the huge amount of content we are now adding to our collections. In short: **text-mining** of the well-structured e-books content

Text and Data Mining

- “Text and data mining (TDM) is the process of deriving information from machine-read material. It works by copying large quantities of material, extracting the data, and recombining it to identify patterns.”
- “LIBER (Ligue Européenne des Bibliothèques de Recherche) advocates for a more flexible copyright system that would allow TDM to be used to its full potential”:

<http://libereurope.eu/wp-content/uploads/2014/11/Liber-TDM-Factsheet-v2.pdf> (factsheet)

- “The Hague declaration on knowledge discovery in the digital age” was drafted at the beginning of 2015 by a panel of experts brought together by LIBER. It will be presented officially on the 6th of May in Brussels
- Key principle: copyright not intended to govern access to facts, ideas and data, nor should it

Library 2020

- “Ci sono segni di una serie di quiete e modeste rivoluzioni che avranno conseguenze di lungo termine. Le biblioteche pubbliche nel 2020 saranno messe nella condizione di non poter piu’ offrire accesso ai bestsellers; questo crea una grande opportunità, quella di portare altri contenuti all’attenzione del loro pubblico. Piccoli editori (scientifici ma anche commerciali di nicchia), insieme a scrittori indipendenti, negoziano condizioni mutualmente vantaggiose per pubblicare le loro opere attraverso le biblioteche pubbliche, *particolarmente* in formato elettronico.[...] Autori locali, musicisti locali, drammaturghi e registi sono alla ricerca di un pubblico, e le biblioteche pubbliche sono in grado di creare questo pubblico.”

Clifford Lynch, in “Library 2020: today’s leading visionaries describe tomorrow’s library”, ed by J. Janes, Scarecrow Press, 2013

Questions?

