

# Liquid Book: reuse and sharing of multifacet content for evolving books

Fabio Casati, Azzurra Ragone

November 23, 2009

## Abstract

Liquid Books (LB) are a new way of thinking about books in the Internet era. LB are collaborative books, composed by multifaceted content, which stay up to date with the current state of art as they continuously evolve over time. Each author can produce different editions of the LB, tailored w.r.t. the needs of prospective readers. LB can be seen as a *big pot* where all the authors put the content they want to share and then they create their personal view on the LB, selecting just some of the content from the cauldron. In this paper we address the challenges related to such a new model of creating and publish knowledge, like the extent of modification and reuse allowed, credit attribution policies and the IT infrastructure required to enable the sharing, reuse and rendering of the LB content.

## 1 Introduction to Liquid Books

Liquid books (LB) are collaborative, evolutionary, possibly open-source and multifacet versions of the traditional books. The main goal behind the LB idea is to enable authors to easily share and reuse their content, giving them at the same time the guarantee that their content will be used appropriately and so producing versions of the book tailored for the needs of the prospective readers (students, consultants, etc.) LB can be open source, partially open, up to the case of traditional closed book. In our model we aim at covering all the possibilities, leaving to authors the final choice of how distribute their content.

A primary - but not the only - application scenario we focus on is that of textbooks, but all the concepts described for this scenario can be also applied to different ones, like books for professionals with more advanced topics w.r.t. the ones intended for students.

When teaching, professors and students need a book which is tailored for the class, which is up to date, which includes (or points to web-based) exercises, which has a companion web site with additional readings, etc. And, at the basis, a book that is correct and clear.

Today, this rarely happen: books are written by a small set of authors, require a lot of coordination actions among them and because updating books often requires a big

effort, by both authors and publishers, then usually new versions contain just minimal changes or more chapters at the end.

From the perspective of professors, it is very hard to find a book tailored for their class, meaning that they have to resort to a collection of books, with the related problems (cost, inconsistent description, difficulty in exposing a coherent set of topics in a coherent way, difficulties in linking topics and classes and exercises, etc..).

The main idea behind LB is let professors compose and create a customized (and as liquid, evolving) textbook for their class by putting together modular (but "refinable") content from other colleagues.

A LB is an evolving and multi-author collection of materials on a given topic that unlike traditional books/e-books, is composed of different kinds of artifacts (documents, slides, exercises, code, experiments etc.), is collaboratively written by several, possibly (but not necessarily) cooperating authors, and has various author/publisher-defined copyright models (see Fig. 1).

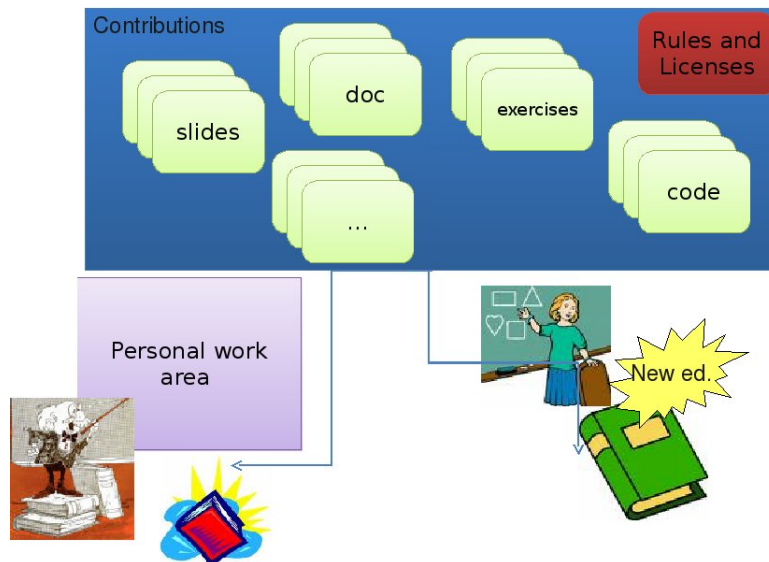


Figure 1: The Liquid book scenario

The idea is to have a few professors that share their teaching materials (slides, sections on given topics, exercises), as they want to produce a book for their class, and then each one of them or some of them produce a different version of the book, as they may have a different vision on how to organize the content.

Then we envision that each professor at the start of the class will "create" a version of the book that is tailored for his/her class by reusing the shared material, either composing it when possible, or taking it and modifying it when needed — and when allowed by the contractual framework they agreed on before to start the LB. We expect modifications to existing content both because when we compose it with other material we need to make sure that the text flows properly and that terms are used consistently,

but possibly also to add/edit content.

There may exist several versions of a LB, per year, per professor and per class. Versions are different from editions: **version** are different view on the LB content, i.e. the content can be assembled in different way depending on the need of readers: book for a class of master students or phd students, book to improve my personal knowledge of a topic, not for teaching purpose, etc. Instead, **editions** are "*solidification*" of the LB, whose content remain stable and fix during the years: editions are LB which is possible to refer to via an ISBN number or a DOI.

We also expect companion web site to be possibly created as a result of the process.

Note that while editions are static collections generated from specific versions of the LB (they do not change anymore), the LB itself remains in its evolving state. As a result, authors can progressively improve and maintain the book, generating new editions as they reach the delta they consider relevant. This model is not that far from open source software (and these may indeed be considered open source books) with the difference that we have potentially many personalized releases, that it may not be fully open but restricted to a community of collaborating professors, and that the business models may be different, with the editor (who could be one of the author) possibly entitled to take content and sell the book – details are left to be specified in suitable licensing models or agreements between authors. LB are like a shared framework: the owners decide who can do what, and if the power to transfer these rights is granted. What we envision for now are LB with 3-20 authors, not like millions, though that is a possibility.

## 2 Liquid Books for Teaching

While LB are intended for all kinds of content, one of the specific areas we focus on is textbooks, both because they seem to be a class of books that can particularly benefits from liquidity and openness, and also because it is one of the experiments we plan to carry on. In this section we will refer to a running example, analyzing different scenarios, in order to highlight the challenges related to each one of them.

### 2.1 Creating a LB: not so simple as they thought

Imagine that Fabio, Stefan and Paco decide to create a new LB on "Smart Web 2.0 applications", each one of them has a different area of expertise and they will put together their competences and materials, that already have or that are writing, to "compose" the LB. The three authors have many overlapping competences, although each of them has a field in which he is stronger. They all teach classes related to this topic, with different foci (focuses) and levels, and they want to share knowledge and teaching material and to prepare new material for their classes, tailored for their students and the goals of their class. Teaching material they have includes slides, lecture notes, exercises, reference literature, etc.. We point out that in this first scenario we do not care about: (i) how content is organized and how is retrieved (at a conceptual level); (ii) the IT platform where the content is actually hosted (svn, launchpad, LP platform); (iii) the multi-faced form of content.

Indeed, in this first scenario we have a small set of authors, not so much content (authors just started the LB) and we suppose they use whatever system they prefer to share content and collaborate. Moreover, we consider, for simplicity, that the final output is just a pdf with a companion web site.

Each professor wants to obtain **tailored** and **packaged** teaching material for the class. The idea sounds great, the three researchers trust each other, they feel that by sharing their books/lecture notes can only get better. As soon as they think about how to do this, problems start to arise. In particular, it is unrealistic to assume that an author can just take somebody else's content and put it into their book without any change. Some (perhaps minimal) change is always needed maybe as simple as removing forward/backward **references**, adjusting **terminology** with the rest of the edition where this part will go into, etc. Once an author begins to use and change a contribution by another, many issues need to be discussed, as the followings:

**Extent of modification and reuse allowed:** as an author, what do I allow the other authors to change? Only small backward/forward references to make it consistent or minor notation, figures, or also terminology? Or also to remove content or add other content? Or add comments and other personal opinions/perspectives to the text I wrote? Do I allow you to publish a summary version of my chapter, or the same concepts written differently? Here a need arises to specify **what** each author can change, more than **how** much.

Indeed, before sharing content, authors would like to have some guarantee that "their" content will be used appropriately. However this is a nice idea in principle, but difficult to put into practice, as no one want "to loose control" on their content, and would like to have "the final word" before the content will be used and published. Therefore, another question arises: does everybody need to agree on all editions that anybody publishes? If Stefan changes Paco content (given that Stefan is allowed to do that) to adapt that to his class, what if Paco does not like that content and does not approve that change?

Here different issues need to be addressed, related to the track or enable of:

- *micro* changes: references, notations, terminology, etc.
- *macro* changes: add/remove new content, rewrite content in different ways
- changes concerning the *meaning* of what is written
- personal changes concerning *personal opinions* or *perspectives*

**Notification of changes:** If reuse is allowed and also changes of different nature, then there is the need to track these changes and to notify authors when they happen. Let us imagine that Fabio has written a document A1 (see Figure 2), then Stefan takes this document and modifies it (doc A2), then Paco takes this content A2 and modify it again (doc A3). The authors should be notified of these changes. Then, if Fabio takes his content A1 and modify it, creating content AB, Paco and Stefan should be notified that the seed document has been modified.

**Policies:** The above mentioned issues lead directly to the problem of define different policies, as the following:

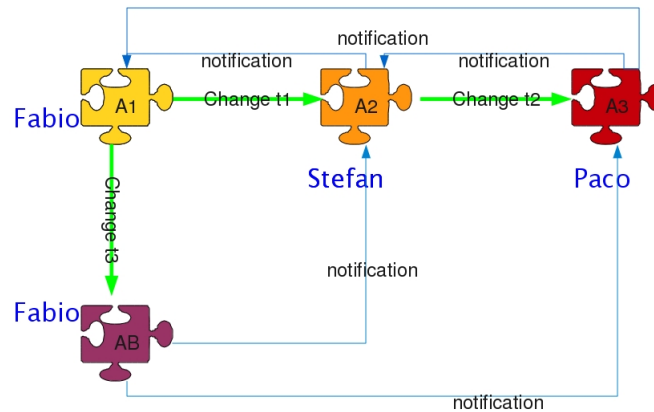


Figure 2: Example of reuse of a contribution

- **Approval policies:** once somebody reuses my content, do I require my approval on the final version to see that my content has been used appropriately and interpreted correctly?
- **Policies to opt out:** if I do not want to be part of this team anymore, and I want to opt out, what happens to my content? I probably cannot just delete my content, also because other authors have already taken that and modified it and included it in their books. So some rules have to be defined in order to handle these cases.

**Authorship:** the reuse of content leads to the problem of who decides authorship. Once Stefan has taken some content from Paco's and Fabio's one, he will have edited that, changed/added something and then arranged the new content in a new book tailored for his class of "Web 2.0 applications". Now the problem is: who is the author of the book? And, how to decide this? It is clear that Stefan did the main work, but he took content that Fabio and Paco have written before.

Possible solutions to disentangle the authorship problem could be:

- Stefan is listed as main author (a sort of editor), and Fabio and Paco are listed as contributors (but on the main cover);
- The three authors are listed simply "as authors", without any distinction;
- Stefan is listed as the only author, but inside the book, for each section there is the indication of who wrote/inspired/initiated the section itself.

- There is authorship with the description of what each contributed, what has been reused, extent of changes, etc.

**Royalties:** A big problem in this scenario is how to decide royalties (yeah, it's about money). The problem of authorship is not only a problem of "intellectual property rights" or credit attribution, as to establish the authors of a book, or how much each author contributed to that, it is essential in order to establish royalties: who gets the money. This cannot be solved in a "qualitative" manner as in the authorship, as here we need to get numbers/percentages.

**IT platform:** it is essential to establish, given such a scenario, which are the features the adopted IT platform should have in order to support the authors in the creation process of the LB. Given the above issues the platform should be able to track changes and to keep track of different branches that spring out from the LB. The scenario is not really different from the Open Source Software (OSS) one. There are several platforms available, which support the collaborative writing of documents, from gdocs <sup>1</sup>, wiki <sup>2</sup>, which are very simple and intuitive, to more complex one, like subversion systems (svn), launchpad <sup>3</sup>, which, conversely, provide more features, but often require a learning effort that discourage the user to deal with that.

Therefore, Fabio, Paco and Stefan, discouraged by the lack of contractual frameworks, that can regulate their joint effort, can protect their intellectual property and give them the right credit for what they contribute to, and not having a platform that can support their work, decide to give up, and to go via traditional channels, they will contact a publisher, who will handle all the legal issues and the dissemination of the book.

## 2.2 Browsing a LB: search and retrieval of content in the big pot

Differently from the previous scenario, we can imagine to have more than three authors, maybe six and with more content shared among that. Then the need arises to organize this content, such that, then, it could be easily retrieved in order to be included in the personal view of one or more authors. We envision a tag-based system gmail-like, with labels. Basically, authors can tag content, so that then it will be easily browsed and retrieved. Authors look for content through tags, and the retrieval system is substantially a keyword-based one. Tags can appear on a tag cloud and that could be the starting point for searching for content. The system will take track of *who tags what*, so that an author can choose to visualize on the tag cloud just his tags, or Paco's tags, or tag of everyone. Then the system will allow the use of *distinguished tags*, that is tags who have a shared meaning (shared semantics) among authors, that are used, for instance, to tag *outdated* content, *advanced* topics, *preliminary* content (*i.e.* content that cannot be published as it is), etc. The tag-based system we envision is a sort of baseline, chosen to keep things simple, at least in this first stage, to propose to authors a system that look familiar to the ones they use everyday and that do not require to them

---

<sup>1</sup>docs.google.com

<sup>2</sup><http://www.wiki.com/>

<sup>3</sup><https://launchpad.net>

to learn new technologies. Obviously, the browsing and retrieval system could then be improved using different approaches not only based on tags.

### 2.3 Multi-facet content: LB as kaleidoscope

LB for their nature are composed by multi-facet content and each author can choose what kind of (in which form) content to present to readers. Following our textbooks example ( but the issue remains valid for each type of LB) , teaching material can include slides, lecture notes, videos (*e.g.* videos of recorded lectures, videos of interesting talks), exercises, code libraries, web pages, reference literature, etc.). So authors should select which one of the multiple forms of content to select and present to the reader, indeed there are different representation of the same content and they can decide to include for a given content *e.g.* just slides, or chapter with code libraries and videos. Therefore, one issue that should be addressed is how to link content, how to render that and in which order. Consequently, the authors need, in order to handle the multi-facet content, an IT platform that can support the management of different representation of content and then the rendering of that in the desired way.

### 2.4 Evolving content: LB and the power of the community

Here, following the story of our three authors, Fabio, Stefan and Paco, we can imagine that, after some years, they come again together, when LB is finally an almost established paradigm and so they can choose among different contractual frameworks the one more suitable for their needs. For the sake of simplicity, here we list some very simple rules they decide to agree on <sup>4</sup>:

- each one of the authors can transfer the right to modify the content to other trusted (potential) contributors
- each modification of the content should be approved by the author/s, who contribute to that content
- each one of the authors can "open" (part of) the content to comments of people they authorize (from small network of people to all the users of the LP platform)
- the authors together decide the preferred way for distributing the content (completely free access, fee payment, etc.)
- each one of the authors can give the "*status of authors*" to contributors who contribute substantially to the LB (see Claudio's example below)

Suppose Fabio wants to create a LB for his course on "Web 2.0 applications", he takes some content from the "big pot", he chooses for each topic the content he wants to present (text, slides, code, etc.) and, then, he gives an order to the contents, creating an index (table of content) for his book, etc.

---

<sup>4</sup>Obviously, in a real case the contractual framework will be more detailed and more complex than this one.

Now Fabio, before distributing his LB to his students ask for feedback and comments to his colleagues, not only his co-authors Stefan and Paco, but also some other professors he trust. Therefore, as allowed by the initial agreement, he gives access to his version (rendering) of the book to Claudio, Maurizio and David and also gives them the permission to leave comments and to Claudio the possibility to modify the content itself and add new content. After this first "review round" the (liquid) content is changed both because Claudio added new content/updated existing one and because Fabio himself added new content/updated existing one. Now a new version is ready. As Claudio contributes substantially to the new version, adding new content on sponsored-search auctions and optimization algorithms, Fabio proposes to him to become an author of the LB, under the condition to accept the agreement they (him, Paco and Stefan) signed at the beginning. Claudio accepts that and Fabio grants him the "status of authors" of the LB. As Maurizio and David provided some useful insights and comments, Fabio decides to cite them as a "contributors" of that particular version of the LB, to give credit to them for their work. Now, that there is a more mature version, Fabio decides to make this new version visible to his students, granting them the right to write comments on the content. During the semester, following hints coming from students (but not only that), Fabio modifies the content, i.e. updating the slides he is using during his lessons, and adding new contents too, like code coming from the projects he is doing with the students, and experiments he is doing by himself. At the end of the semester, based on comments he got from students, Fabio decides to add new content as well as to revise existing one, he also decides to include in the new version new material which came up from the course (experiments, code libraries, etc.). Now that he has a most stable version, Fabio contacts a publisher (Springer) to publish a physical "Edition" of the book. Springer will take care of the "dissemination" of the book (libraries, online selling, etc.). The book could be distributed through different channels and in different forms: from the traditional way (hard-copy, pdf, etc) to new ways, like give different types of access (base/premium access) to different part of content (only to text and slides, or to code and experiments, etc.)

### 3 LB, non only textbooks

Also Paco wants to create his personal "view" of the LB, so he decide to create a personal view of the LB (*i.e.* a new version). Differently from Fabio's book, his target are not students, he wants to create a LB for people who develop web 2.0 applications for commercial use. Paco selects some content from the "big pot" and renders that putting more "advanced topics" in his LB, and more experiments with real data and application. Then, he makes the version visible on the LP platform <sup>5</sup>, allowing comments from readers who are registered on the platform (here, we can imagine that he gives access only to friends, or friends of friends, in LP network). Through the access right management system provided by the platform, he can discriminate among people and give to someone only the right to write comments, to other the right to modify content and to other the right to update new content and (maybe) the right to tag content. Paco

---

<sup>5</sup>Here, we envision that LB are hosted by the LiquidPub platform (<http://project.liquidpub.org/>).

wants to leverage the power of the LP community, Paco give access to people he trusts, people in his network or, alternatively, he can just decide to invite people, maybe people who run the most famous LJ in the field, or people that have an high reputation in the LP platform.

Basically, here the content is created more in a collaborative way, with a lot of contributors and Paco who has mainly the role of **editor**, that is to collect contributions, addressing comments, organize contents and rendering them in a proper way. At the end, after several review phases, when the book will be mature enough (wrt to Paco judgment), Paco can decide to finalize the LB and make it visible to all through the LP platform, that is to publish that, do an edition that can be referred to using an ISBN or DOI. The ones who have left useful comments and insights will get credit for their contributions, as they will be cited in the final edition (but also in the liquid version who remains in his evolving state) as contributors. Here Paco does not involve the publisher in the process, using the LP platform to disseminate the book (and maybe other channels like amazon), and using his reputation as an expert in the field (maybe gained from the first story) and the reputation of the contributors to get credibility. For instance, among the contributors there are people considered as experts and who work for leading companies in the field. Also in this case he can decide with the other authors (Fabio and Stefan) if the book will be completely open or if some part will be open to all and if some fee will be requested to access to other part of the book. If the LB is on the LP platform, then we can compute some metrics based on the usage of the LB, for instance how many users have downloaded contents from the platform. On the other hand, we can give the possibility to users to rate the content and write reviews (as in amazon), and then we can compute other metrics based on this.

## **4 Liquid Book: what is new, what is different, what are the benefits**

### **4.1 Benefits and Differences wrt Traditional Books**

The main difference of LB w.r.t. traditional books is the fact that LB have the possibility to continuously evolve and to instantly adapt to different needs and wishes. Differently from traditional books, that are constrained by editions, and so hardly updated, LB can be updated (possibly) every single days without such a big effort. Keeping the example of textbooks, they can easily adapts to different requirements of different professors, who can compose their courses using only some of the material proposed in the LB, or using view/perspective of other authors. Moreover, LB are multi-facet, not only composed by traditional chapters and exercises, but also by slides, code libraries, experiments, etc.

We also propose a new way of collaboration among authors, the content, once shared, can be edited/changed by other authors, if the original contributors allows so. We can envision a small communities of authors (e.g. four or five) who create a "seminal" LB, whose content is then (partially) taken by other authors who can rearrange it, adding new contributions, correct some flaws, until a new delta is reached to have a new version. We can imagine new version of the LB which branch out from a "root"

one, and that can include different authors for each version. Such a scenario requires the definition of different copyright models, or, generally speaking, contractual frameworks that can rule the interaction among authors in the LB framework.

Moreover, the fact that contributions are annotated allows one to easily browse the content of LB, and also to "create view" that are compliant w.r.t. the authors' needs (e.g. teaching purpose or learning). Different from traditional books, as LB are *modular* books, the reuse and sharing of content is made easier, as well as the *personalization* of that. Furthermore the modular and multi-facet nature of that allows the LB authors to choose from different way to distribute content, from totally open to partially open to completely closed book.

## 4.2 Differences wrt Wikibooks

Here, we list some differences between LB and Wikibooks (WB)

- WB are open-content and freely distributed, LB are not necessarily open-content, if authors do not allow so. They can decide to make freely available just some part of the book, or nothing at all. WB are "public" everybody can see everything, while LB are not necessarily visible to everyone, readers of the book do not have necessary access to the "sources" of the LB. Moreover, in LB still there is (possibly) the notion of publisher, editions with ISBN, DOI etc.

- In WB anyone can contribute adding new content, editing or deleting existing one. Indeed WB adopt a GFDL license <sup>6</sup>, which is a Creative Commons license which ensures that content remains copyrighted to the authors, while at the same time the copyleft licensing allows to freely distribute and reproduce such a content. Whereas in LB authors decide *who* can modify *what* and *if* their content can be edited or if just comments are allowed. Moreover the framework will give the possibility to choose the licensing schema more suitable for each case, but authors will not be forced to choose the Creative Commons license, neither any other license, leaving to them the choice about how to distribute their content and how to deal with intellectual property rights. In WB everybody can participate, so a WB can have a thousand of authors, in LB we do not exclude this possibility, but this is just a possible way of collaborating. Others could be: only few authors start a LB and then give the permission to only "trusted authors" to modify the content.

- The main difference is in the fact that WB are *static* collection of content in the sense that it is not possible to have different views on the WB, while LB are modular multi-facet content, that can be personalized depending on the different needs of authors.

At the end, LB and WB are different as their ultimate aim is different. WB are intended to make easy the spread of content in an open way without posing any restriction to possible contributors. LB are more focused on *reuse, sharing and evolution* of content among authors and to propose to prospective readers multi-facet content tailored for their needs. Therefore, in WB the focus is exclusively on the creation process, in LB not only on that, but also on organizing the content (view) and rendering it in a suitable way for different readers.

---

<sup>6</sup><http://it.wikipedia.org/wiki/GFDL>

### 4.2.1 Benefits

**Benefits for readers:** book is up to date and is tailored to their needs (especially for a class this is important). There is additional material clearly linked, companion web site, exercises if textbooks, etc. Also readers could buy only contributions they are really interested on, instead of the entire LB edition. This is tied with the possibilities to have different business models.

**Benefits for authors:** reduced effort to create and constantly keeping update books, easy to increment and to add content. Moreover, the reuse make possible not to start from zero every time I want to put new content in my LB. Having a legal framework makes easy to collaborate and reuse material.

**Benefits for publishers:** more editions, and so potentially more money. Having more editions can be a way to provide additional value with respect to printing services by supporting a legal and technical framework for LB and for additional material.

## 4.3 Difference between Liquid Book and Liquid Journal

In this section we highlight some of the differences between Liquid Book and Liquid Journal (LJ), the other use case we have in the LiquidPub project. LB are mainly centered on authors, rather than on readers as LJ, indeed in the LB case we do not focus on readers, who rate, comment, bookmark content, but more on "users" with rights and duties, that can have different types of access to LB content, for instance: view, use (e.g. for teaching), share/redistribute, edit/change.

In LJ people collect *existing* content, without modify it, by explicit queries, by following recommendations, or by adding new content manually. The purpose of LJ is to help people to *find* content for the sake of *reading*. Then, the side effect is that they help rate and filter content. While in LB is not only a matter of *collecting* existing content and (possibly) bring out "interesting" one, but the focus is on the creation process: *creating* new content in a new collaborative way. Actually, in LB we deal with a "continuous creation process", as they evolve over time, so the focus is on both the creation and the update process.

LB are a way for people to collaborate in creating content, content that needs to be organized, indeed in LB there are specific notions of lessons, exercises, companion web site, etc.

Moreover, an issue we do not have in LJ, but that is central in LB is the notion of copyright/licensing, permissions that need to be defined, together with contractual frameworks among authors, and between authors and publishers.

LJ can be created by publisher, as well as "common people", so basically by everyone, and there is no need for licenses or permissions as they rely on licenses/permissions that already exist (permissions related to e.g. papers they decide to include in the LJ). While in LB there is the notion of author, contributor, editor, publisher that strictly agree to some contractual framework in order to manage their collaboration and to share, distribute content. Therefore, we aim in the LB use case to define several types of contractual frameworks that will be necessarily different from the existing ones, as will reflect the above mentioned differences between LB and traditional books.