

Cookbook for Open Access books

Sebastian Nordhoff

Language Science Press

Language Science Press Guidelines



Cookbook for Open Access books

Sebastian Nordhoff

Language Science Press



Sebastian Nordhoff. 2018. *Cookbook for Open Access books*. Berlin: Language Science Press.

This title can be downloaded at:

<http://langsci-press.org/catalog/book/opendata>

© 2018, Sebastian NordhoffLanguage Science Press

Published under the Creative Commons Attribution 4.0 Licence (CC BY 4.0):

<http://creativecommons.org/licenses/by/4.0/>

ISBN: 978-3-96110-089-7 (Digital)

978-1721065851 (Softcover)

DOI:10.5281/zenodo.1286925

Source code available from www.github.com/langsci/opendata

Collaborative reading:

<https://paperhive.org/documents/remote?type=langsci&id=cookbook>

Cover and concept of design: Ulrike Harbort

Fonts: Linux Libertine, Libertinus Math, Arimo, DejaVu Sans Mono

Typesetting software: Xe_{La}TeX

Language Science Press

Unter den Linden 6

10099 Berlin, Germany

langsci-press.org

Storage and cataloguing done by FU Berlin



Contents

1	Introduction	1
1.1	Scope	1
2	Prime the pump	3
3	Tasks	5
3.1	Creating prestige	6
3.1.1	Bold claims	6
3.1.2	Selectivity	6
3.1.3	Big names	7
3.1.4	Number of supporters	7
3.1.5	CI	7
3.1.6	Innovation	8
3.1.7	Content	8
3.2	Acquisition	9
3.3	Discoverability: Getting readers and getting read	10
3.3.1	Website	11
3.3.2	Search engines	11
3.3.3	Platforms	11
3.3.4	Blog and social media	12
3.3.5	Identifiers	13
3.3.6	Keywords	13
3.3.7	Paper copies	14
3.3.8	Book reviews	14
3.3.9	JSTOR and Project Muse	14
3.3.10	Readers become authors	14
3.4	Quality control	15
3.5	Requirements	15
3.5.1	Requirements for a web platform	15
3.5.2	Requirements for book production	16
3.5.3	Required elements	17
3.5.4	Requirements for distribution	19

Contents

3.5.5	Requirements for archiving	20
3.6	Automation	21
3.6.1	Ingestion	21
3.6.2	Generation	21
3.6.3	Outlets	23
3.6.4	Archiving	24
3.7	Solutions	24
3.7.1	In-house	24
3.7.2	Web services	26
3.7.3	Outsourced	27
3.8	Accountability	27
3.8.1	Numbers	27
3.8.2	Technological	28
3.8.3	Organisational	28
3.9	Money	29
3.9.1	Revenue	29
3.9.2	Expenses	30
3.9.3	Liquidity	30
3.9.4	Taxation	30
3.9.5	Bank accounts	31
3.10	Legal	32
3.10.1	Legal forms	32
3.10.2	Brands	32
3.10.3	Intellectual property rights (IPR)	33
3.10.4	Hosting	34
3.11	Governance	36
3.11.1	Internal	36
3.11.2	Relations to home institutions	37
3.11.3	Organisational matters	38
3.12	Community involvement	38
3.12.1	Overview	38
3.12.2	Recognition	39
3.12.3	Social media	40
3.12.4	Inquiries	40
3.12.5	Newsletters	40
3.12.6	Merchandise	41
3.12.7	Series editors meeting	41

3.13	Branding and promotion	41
3.13.1	Values and claims	41
3.13.2	Design	42
3.13.3	Social media	42
3.13.4	Conferences	42
3.14	Network	42
3.15	Competition	43
4	Readings	45
5	Timeline	47
6	Time sinks and lessons learnt	49
7	Conclusion	51

1 Introduction

This book details the experiences and lessons learnt when setting up Language Science Press, in the period from 2012–2018. It makes recommendations for how to set up an open access book publisher. These recommendations are drawn from personal experience and have a practitioner's view. As a correlate, theoretical considerations and empirical evidence are backgrounded. This is on purpose.

Language Science Press is based in Germany. The cookbook tries to be generic and give general principles, only using the particular German context for exemplification. When this is done, this is explicitly indicated, but it might be the case that there are some passages which we failed to identify as peculiar to Germany. If you are based elsewhere and a passage does not make sense to you, you might have found such a passage.

It is one of our goals that many XYZ Science Presses will spring up in the future, in other fields ranging from Archeology to Zoology. Obviously, every scientific field is different and has its own culture and its own communities of practice. It is hoped that this cookbook will help these initiatives in calibrating their setup based on our experiences.

Some of the recommendations given here might sound reasonably trivial, others might sound opinionated. But we have chosen to be explicit, at the risk of sounding banal in order to give as complete a picture about the things we have learned in the course of this project.

1.1 Scope

Language Science Press publishes open access books in linguistics. We do not have insights into closed access books or journal articles, although some of our findings can probably be transferred.

Linguistics as a discipline is typically considered part of the humanities. This being said, it is often dubbed the “humanities discipline closest to the natural sciences”. A number of practices in linguistics are closer to the natural sciences than to the humanities in general: linguists publish a lot of journal articles; peer review is generally required; linguistic articles can contain a lot of diagrams, charts, tables, and theorems; and the scientific community is organised on a world-wide

1 Introduction

scale, with English as the normal language of publication. It may or may not be the case that the leading role that linguistics has in open access among the humanities is somehow related to these practices. In general, authors in linguistics do not have to be convinced to publish open access; they demand it. This seems to be different in other fields of the humanities. This particular position of linguistics has to be taken into account when applying the recipes presented here.

2 Prime the pump

The most important take-home message is: you have about 2-3 years after the launch to build up enough prestige to keep rolling. After that period, you either have a steady stream of submissions, or your project is all but dead.

Success begets success, and nothing attracts the crowd like the crowd. You want to be seen as the place to be, as the place where everyone flocks to. Every new publication should lead to a number of follow-up proposals from readers who have become interested in the press. The problem with books is that they take a year to write and a year to review and produce. This means that the follow-up books will come out about 2 years after the book which prompted them. In other words, the period from generation to generation is two years.

For Language Science Press, we can distinguish three “waves” of proposals:

1. Starting books (7). These are manuscripts invited by the press directors and are available for publication within the first year after launch.
2. Second wave (12). These are submissions from authors who have seen the announcements or publications of the first books. (Ideally within one year after launch.)
3. Third wave and beyond (50): Submissions from authors who have seen a second-wave or later book (Roughly 3 years after launch.)

Your operations can be seen as steady when you have reached the third generation of proposals, i.e. follow-up submissions on follow-up books on the starting books. At that point in time, the generations become less clearly distinguished, and quick books from the fourth generation might come out earlier than late-comers from the third generation. Your initial funding should allow you to get to the publication of a third generation book.

A number of conclusions follow from these premises:

1. You should have a book publication out ONE DAY AFTER your initial funding starts in order to attract second-generation proposals early.

2 *Prime the pump*

2. Your first publications should be of very high quality, and should be as speedy as possible without sacrificing quality. A delay in first-generation books will set you back for all future generations, in terms of time and in terms of quantity.
3. You must make sure that the starting books get maximal exposure in all relevant venues in order to produce follow-up submissions. For the first three generations, you want exponential growth. After that, linear growth will do.

Normally, you will not be able to adjust your staffing according to the number of books. This means that in the beginning, there will be too much staff, i.e. relatively fewer books per person to work on than later on. Use this leeway to create beautiful hand-crafted books and as pleasant as an experience for the authors as possible. Once the ball gets rolling and you attract more submissions, you will have to scale down that level of dedication, but at that point in time, you will hopefully be in a position to be selective with regard to the submissions you accept.

In the beginning, we had some edited volumes where chapters were written in various versions of Microsoft Word, with no particular template, and no consistent or coherent formatting. We adapted all of those to a uniform style and spent enormous amounts of time on this. As a result, we attracted more submissions, and we are now able to enforce our templates and guidelines, and can afford refusing submissions which do not comply. As the number of books in production goes up, the time we have available for each book goes down, but at the same time, the submissions are in much better shape and require less time overall as well.

3 Tasks

This chapter will detail a number of tasks which a fledgling project has to deal with. For each task, we will give some background and give some recommendations. The tasks to be discussed are:

1. Creating prestige (Section 3.1)
2. Acquisition of manuscripts (Section 3.2)
3. Discoverability (Section 3.3)
4. Book production
 - Quality control (Section 3.4)
 - Requirements (Section 3.5)
 - Automation (Section 3.6)
 - Solutions (Section 3.7)
5. Accountability (Section 3.8)
6. Financial matters (Section 3.9)
7. Legal matters (Section 3.10)
8. Governance structure (Section 3.11)
9. Community building (Section 3.12)
10. Branding (Section 3.13)
11. Building a network (Section 3.14)
12. Competition (Section 3.15)

3.1 Creating prestige

This task is the most important one. It is the single item which will determine success or failure of your enterprise. Prestige is a virtuous circle: with better prestige, you get better submissions, which lead to better prestige etc. The reverse is also true. You must invest as much as you can in starting as high up on the prestige ladder as you can possibly afford. Try to take over the publishing landscape working from the top of the quality pyramid to the bottom, rather than the other way round.

Obviously, in the beginning, you have zero books. This means that you have nothing to show, and you must signal prestige in another way. We can name six ways: bold claims, selectivity, big names, large crowd, CI, innovation, content.

3.1.1 Bold claims

Make your claim to prestige very clear, very loud, and very often. The important thing is that you are aware that this is a CLAIM, not a truth. In the case of Language Science Press, we asserted that we want to play in the Champions League of publishers and that we want to be better than de Gruyter (a leading publisher in linguistics). Have a couple of talking points where you want to be better in case someone asks you. For instance, faster turnaround time; better typography; better integration of multimedia. It is OK if people think the claim is preposterous, actually, this shows that your claim is just right.

3.1.2 Selectivity

Assert that you are selective and that not everybody will be able to publish with you. Highlight your rigorous quality assurance principles. Disclose your rejection rates. The point is not so much in rejecting bad manuscripts, but in deterring authors from submitting low quality manuscripts in the first place. The clearer you make that low quality manuscripts will be rejected, the less you will get of them. And low quality manuscripts cause more work than better ones, for less reward (or none at all if rejected). At all costs you must avoid as being seen a publisher of last resort. The first books will set the standards for the future submissions, so better be strict. When you get a good submission, do everything to publish it quickly to attract follow-ups, but do not try to generate follow-ups on bad submissions: they will be bad as well. Language Science Press has explicitly stated that it will not publish theses or festschrifts, as both of these are seen as low prestige in linguistics. We do publish books based on revised theses, though,

but it was important to make the distinction with a run-of-the-mill university press for thesis printouts.

3.1.3 **Big names**

Obviously, you have to back up your claim. Remember, you still have zero books. Researchers associate prestige with leaders in the field. Establish contact with all leading scholars in your field and try to get them to publicly support your enterprise. Go for people the general public might recognize. Create a list of supporters and list those researchers at the very top.

Get people from abroad. Prestige grows with the square of the distance. Endorsements from another city are great, from another country even better, but you should really try to get endorsements from researchers from several continents.

For Language Science Press, we got support from Luc Steels and Steven Pinker, who have both been featured in the popular press. Their support was crucial in substantiating our claims to prestige.

3.1.4 **Number of supporters**

Next to the big names (quality), you can also signal prestige by sheer quantity. Nothing attracts the crowd like the crowd, so open up your supporters list for all interested people and show the large backing you have from the community. Interested authors will see how large a supporter base you have and conclude that those people cannot possibly all be wrong. This is again a virtuous cycle: if you have many supporters, people will be more likely to join the crowd, increasing your number of supporters even more, etc. Again, this means that you should make sure you have a sizable number of supporters to start.

3.1.5 **CI**

CI is a technical term from design and marketing and stands for ‘Corporate Identity’. This term is used even if the entity under discussion is not a corporation. Your project should have a professional and concerted appearance, online and offline. Hire a professional designer to assist you. Pay for a professional logo. Be strict. Do not allow deviations from the CI, especially in the beginning.

In the case of Language Science Press, the designer advised on the choice of fonts, colours and book layout. All books use the same layout and the same title page. The colours used for the books series are also used on the website and

in flyers. The font used is the same for books and flyers etc. This makes it easy to recognise some item as coming from Language Science Press. The uniform appearance across formats suggests stability and seriousness, even if no actual book has been published yet.

3.1.6 Innovation

As a new publishing initiative, you are not hampered by legacy technology and can easily propose innovative procedures which publishers with already established workflows find difficult to implement. Language Science Press has Open Peer Review for instance, was (to our knowledge) the first publisher to use hypothes.is, the first to incorporate PaperHive in the workflow, the first to provide all sources, graphics and figures in raw form on GitHub, and the first to provide various versions of the book (open review version, proofreading version, first edition, subsequent editions). Since you have to design your processes from scratch anyway, you might as well be innovative. Clearly highlight these innovations of yours and show that you are leading the way in the technological domain.

3.1.7 Content

The seventh component to signal prestige is the actual content of your first book. The six items discussed before are only promises. With the first book, you have to deliver on them.

There is trade-off between quality and speed of publication. In the beginning, you have less books, hence more time per book. Make sure your first book is picture perfect, and add some bells, whistles and gongs. These do not really add to the scientific quality of the book, but they underscore that you take your new publishing business seriously. At Language Science Press, we have clickable cross-references. For the string “as seen in Table 2”, you can click on the “2” and your PDF viewer will jump to Table 2. Normally, Table 2 is in the direct vicinity of the text passage, hence there is little need for the hyperlink, but this is some extra feature which we highlighted. Another item is that most of our geographical maps are searchable. This means if you search “Somalia” in the PDF, you will find the string “Somalia” in the prose text, but also in maps of Africa where it occurs. Finally, all of our books have three indexes (Subject index, name index, language index). Most people probably never use the index, but rather hit CTRL+F when they want to find a particular term in the pdf, but the creation of the index was a proxy for us to signal prestige. In order to make claims to prestige, you

must make absolutely clear that you do not simply provide printouts of random Microsoft Word documents as they are submitted.

3.2 Acquisition

In order to get the publishing platform rolling, you have to have a certain number of manuscripts in the pipeline. The manuscripts must provide you with work until the second wave submissions come. In the case of Language Science Press, we had 7 initial books.

For acquisition, we have to distinguish researchers by their seniority level. Types of work and motivation are different between younger and more advanced scholars.

Early career researchers are digital natives. They have been sharing things over the Internet for their whole life and are generally close to technological progress. Furthermore, they are often more idealistic than older people. This is good for open access platforms. On the down side, they are normally less experienced writers. When talking about books, the things which immediately comes to mind are theses. Most theses have to be revised in order to become good books, but early career researchers are normally happy to receive feedback on their work and adjust it accordingly. When using a series model (see below), the series editors should normally have access to some PhD students in their field eager to publish open access.

In case the authors worry about their career, you can refer to the OA citation advantage.¹

For senior researchers with tenure, it is less important to publish in venues traditionally seen as prestigious. Instead, they can show that they are a leader in the field by moving first and directing their disciplines towards novel and innovative ways of publishing. Series editors should be aware of who in their field has book publications projects. In the case of *Studies in Diversity Linguistics*, we were able to get a manuscripts from a senior scholar which was already finished, but had never been formally published (Dahl 2016).² Technically, that manuscript had been available from the document server of Stockholm university for some time. Since its formal publication with Language Science Press, that book has seen more than 3000 downloads. This is probably more than it would have gotten if it

¹<https://sparceurope.org/what-we-do/open-access/sparc-europe-open-access-resources/open-access-citation-advantage-service-oaca/>

²Östen Dahl. 2016. *Grammaticalization in the North: Noun phrase morphosyntax in Scandinavian vernaculars* (Studies in Diversity Linguistics 6). Berlin: Language Science Press

had only been available from that document server. For Language Science Press, however, more important than the downloads was the fact that we could count Östen Dahl as one of the leading Scandinavian linguists among our authors. This complemented the list of otherwise junior authors for our first books and sent out the message that Language Science Press was to be considered as a serious venue for senior researchers as well.

Typical careers in linguistics only give you the time during your PhD to complete a monograph. After that, the research becomes more granular and more focussed, and journal articles and papers in edited volumes take over. Edited volumes are great to build up a relation with more authors. At the same time, edited volumes are really tough and demand about 3-4 times more work than a monograph of the same length. The first edited volume we did was *The Alor-Pantar languages* (Klamer 2014),³ and it was definitely much more demanding than we had expected. We are pretty certain that the paper authors and the editors questioned the wisdom of their idea of publishing with Language Science Press more than once. However, in the end, they seem happy, and several of them have now become regular authors in edited volumes of ours (Klamer, Kratochvíl, Corbett). It might help that *The Alor-Pantar languages* was huge success as far as downloads are concerned, with more than 12 000 downloads at the time of writing.

Try to get subcommunities of linguistics to adopt your press as “their” publisher. For instance, the Association for Laboratory Phonology has the *Journal of Laboratory Phonology* for their articles with the Open Library of Humanities, and they have *Studies in Laboratory Phonology* as their series with Language Science Press, which is set to become their regular outlet for books. Similar approaches are being taken in the fields of Second Language Acquisition, Slavic Studies, and African Studies.

3.3 Discoverability: Getting readers and getting read

The example of Östen Dahl’s book sitting on the e-doc server of Stockholm University without being widely noticed leads to another important aspect of publishing, namely discoverability. It is not enough that the content is available somewhere on the Internet, people have to know that the content exists in the first place as well.

³Marian Klamer (ed.). 2014. *The Alor-Pantar languages: History and typology* (Studies in Diversity Linguistics 3). Berlin: Language Science Press

3.3.1 Website

The obvious first step is to put the PDF of a book on the website of your project. That website should be state-of-the-art. For us, this meant that we adapted ours (in 2014) to be responsive, meaning that the website can be viewed on mobile phones or tablets. We furthermore focussed on the main use case “researcher wants to download PDF” and tried to make this as easy and obvious as possible. It currently takes three clicks (“read books”, click on book, “PDF”).⁴ We use the OMP software to host our books, which provides an OAI-PMH interface for libraries to harvest the catalogue. That interface has a serious bug since 2014, which is apparently low priority for the developers to fix, but the Bielefeld Academic Search Engine has managed to get around that bug, and university libraries now can automatically harvest our catalogue.

3.3.2 Search engines

Open access publications have the advantage that the content of the books is immediately visible to search engines like Google. This means that interested researchers looking for “grammar Wersing language” for instance will be offered the chapter “Plural number words in the Alor-Pantar languages” on the first page of the result list, even if neither “grammar” nor “Wersing” are found in the title. Since Google was able to access and crawl the content, it is easy to return that chapter in response to such a query.

3.3.3 Platforms

The traditional reader-pays model of scientific publishing relies on the gatekeeper function. In order for it to work, there must be exactly one way to access the content (the gate), and that gate has to be manned by someone who charges the fees for passage/access (the gatekeeper). As a result, traditional publishers have problems with their content being available at other places since this will allow readers to bypass the gate.⁵ For open access publications, this problem is moot. You can have as many gates as you want, and open them all up. Different readers may have different needs or preferences. We do not know who has which preferences, but we are happy to provide our books at as many places as

⁴With the number of books growing, we might need additional navigation for the catalogue, though.

⁵They could also set up additional gates for other platforms, but they would come with overhead and therefore costs as well.

we can. Currently, our books are available on our site, the e-doc server of Freie Universität Berlin, Google Books, Google Play Store, Directory of Open Access Books, OAPEN, Zenodo, Knowledge Unlatched, GitHub, PaperHive, university library catalogues, Amazon, and local book stores. We have access stats for most of these, but not for all (see the spreadsheets provided). The absence of access information for all platforms from where the books are available is unfortunate, but only a minor nuisance since we still get a pretty good overall picture (to be discussed in more detail in Section 3.8 Accountability below).

Most of the mentioned platforms are available free of charge. OAPEN charges membership fees, and Knowledge Unlatched will only host books which have been funded via their platform. Local book stores in Germany rely on the VLB, which charges yearly fees per book being listed there.

3.3.4 Blog and social media

Next to providing the books in a user-friendly way to researchers who know that the books exist, and to guiding researchers looking for information but ignoring the existence of the particular books, one can also try to spark interest among researchers previously not interested in the topics. We run a blog at <http://userblogs.fu-berlin.de/langsci-press/>, where we give information about Language Science Press as a community-driven enterprise. The blog covers aspects such as the business model, conversion of documents, or download statistics. Its goal is to be interesting in itself, and not to be redundant with the main website. Next to the intrinsic pleasure we derive from writing for the blog, it should also attract readers, which are then led to the book pages.

A similar function to the blog is fulfilled by our social media channels. We mainly use them to announce books as “forthcoming” and as “just published”, but we also contribute to discussions and share impressions from conferences or funny bits from our everyday work. We try to keep the persona of Language Science Press neutral. For more political discussions, an account from a real person will take over. Our impression is that librarians and OA activists prefer Twitter, while researchers in linguistics prefer Facebook. One could think about tailoring the respective contents offered to these target audiences. As a matter of fact, however, we use Twitter as our main social media channel, and only repost selected content on Facebook. This has mainly to do with the coordinator being familiar with Twitter, but reluctant to use Facebook. Another person in that position might have resulted in opposite preferences. We do not use Instagram or any other social media, mainly because this has never been requested or suggested.

Facebook and Twitter accounts are free for everybody to create. Setting up an organisation on Facebook is a bit more tricky, but not overly complicated. Any university should have some provisions to host a blog.

3.3.5 Identifiers

All our books have several ISBNs, one for the digital version and one for each of hardcover/softcover. All books have DOIs (Digital Object Identifiers), and we now also provide DOIs for chapters. ISBNs and DOIs for books are displayed on our website and included in the book. DOIs for chapters are only included in the chapters themselves, but not on the website.

The DOIs allow the documents to be uniquely identified, and they can be used to easily retrieve a well-defined version of a document. They also help establish relations such as “prior version”, “later version” or “part of”. This is not done a lot yet in linguistics; it is more common in the natural sciences. In any case, having a DOI will assure a document can be integrated in citation graphs, etc.

Our DOIs are provided by Zenodo. Our ISBNs are provided by the library of Freie Universität Berlin. Zenodo provides DOIs to anybody free of charge. Most university libraries should be able to get you ISBNs.

3.3.6 Keywords

Many of the various platforms where we make our books available have forms for keywords. These are intended to describe the content of the work. In the context of Language Science Press, we do not systematically use them, and we have not yet understood how to create good keyword lists. Should we use “Alor-Pantar”, “Alor” and “Pantar”, “Alor-Pantar languages”, “Languages of Alor and Pantar”, “Languages of the Alor archipelago”? Probably there is a way to match all of those, but we are not sure which it is and how we could facilitate that. For the time being, we assume that the full text being available will be enough for Google et al. to extract the relevant concepts via machine learning, and the Wersing example cited above seems to prove the point.

There has been a lot of discussion about metadata and discoverability at publishing conferences as of late, and it might very well be the case that we are missing important parts of the picture. In any case, we have managed to get where we are without paying too much attention to keywords.

3.3.7 Paper copies

Our books are listed on Amazon and are available via local book stores and the standard academic book distribution channels. As for presenting sample books to prospective authors, we use a novel concept. Traditional manned book stands at conferences are very expensive. You have to fly a representative there, pay for accommodation, and they can normally not do a lot of their ordinary work. Instead, we opt for unmanned book stands and conference ambassadors. We send free copies to conferences. These are displayed on an unmanned table. All books have a small sticker on them saying "Available for free download from langsci-press.org. Please leave physical copy in place". After the conference, we donate the book to the local organisers, their student assistants, or their libraries. For inquiries, we have so-called conference ambassadors. These are normally editors of one of our series who are present at the conference anyway. Interested authors can get in touch with them if they need information for a particular publishing project.

3.3.8 Book reviews

We do not send unsolicited copies to journals for review. If a journal asks for a copy, we send it to the reviewer, no questions asked. We do not track whether the review is actually completed, and we do not ask for the copies to be returned. The overhead would be too important.

Our books are posted on LinguistList, and thereby automatically included as "available for review". A list is also sent to *Language*, the leading linguistics journal, on a regular basis.

We encourage reviewers to use PaperHive for reviews of the electronic documents.

3.3.9 JSTOR and Project Muse

These two projects have been suggested to us, but we have not yet understood what the advantage would be for native open access books.

3.3.10 Readers become authors

One of the main advantages of being available on all these channels is that the books get wider exposure, and readers might consider Language Science Press as a venue for their next book to be written.

3.4 Quality control

In order to give weight to your claims about prestige, you must establish procedures of quality control, enforce them and communicate them clearly. Language Science Press uses a series model. Scientific quality is assured by the series editors, who are specialists in their field. Series are autonomous, as even the most erudite press directors will not have in-depth knowledge of all subfields, ranging from Phonetics over African Linguistics to Computational Linguistics.

Scientific quality control is complemented by formal quality control provided by the Language Science Press office. This regards references, typography, DOIs etc.

The viability of series and their setup underlies “meta quality control” by the advisory board. Only series proposals which can show that they have enough reputation, competence, clout, and backing from their respective subcommunities will be accepted. The first five volumes of a series undergo review by the press directors to assure that the series is line with the general expectations of quality.

3.5 Requirements

In order to run a publishing platform, you have to rely on technology. It is good practice to first take stock of what you want to achieve (the requirements), and then to see which technological solutions can help you. Normally only a subset of the requirements will be must-haves, so the best solution for your particular use case will vary.

3.5.1 Requirements for a web platform

Coming from the reader’s perspective, the most important thing for an open access book is a place on the Internet where the books can be downloaded. This is a very simple requirement. In principle, a static website or even an FTP server would do. Additional requirements are user authentication and access control. This is obviously not necessary for fully open access books: the very definition of open access precludes access control. Again, coming from a reader’s perspective, an open access publisher could easily do without access control, and therefore also without user authentication.

Things are different when we take into account the whole book production work flow. Before the final book, we have the initial submission, revised ver-

3 Tasks

sions, and final drafts. Having some control on who can access these documents (authors, editors, reviewers, but not random people) can be justified. Note, however, that simply sending these documents via email to the intended recipients would already solve that issue.

Finally, some series editors like to have a well-defined interface which models the progress of a manuscript through the stages of submission, review, revision, typesetting, and publication. For this, you need different roles: authors can submit; editors can assign reviewers, read reviews, and advance a manuscript to the next stage; reviewers can only write reviews. Authors or reviewers are barred from actions such as accepting a manuscript or publishing it, and reviewers might only be allowed to see their own review, but not the reviews of others.

An additional requirement might be reporting. You want to know how many submissions you have at what stage at a given point in time, or during a given interval (e.g. a year). In Section 3.7, we will discuss different solutions for (subsets of) these requirements.

3.5.2 Requirements for book production

As detailed above in Section 3.1.5, your books should have a common appearance. They should be recognisable as coming from your press. The manuscripts authors submit will often not be very close to the intended final appearance. One of the main tasks of the press is to make sure that the raw manuscripts gets crafted into a worthy member of the prestigious set of books published by your press.

As discussed above, we distinguish the quality assurance of content from the quality assurance of form. The scientific merit of a manuscript can be established independent of details such as font size or page geometry. But once we know that a manuscript has sufficient quality to be published by our press, we have to make sure that the formal specifications are also respected. This means that at submission stage, Language Science Press accepts any format that allows the evaluation of scientific merit. There is no requirement to reformat the book according to our guidelines for the initial submission. Only when the manuscript gets accepted do we require that our guidelines be respected for the revised version.

Coming back to the requirements, you have to have an ingestion interface where revised manuscripts can be entered, a transformation pipeline where they will be conformed to your complete specifications, and an outlet interface, where the final document can be made available electronically or as print-on-demand.

Depending on the practices and tech-savvy of the particular discipline and the resources available, the ingestion can be more or less strict. Some fields are tech-savvy and are used to respecting detailed technical specifications, while others

have a more heterogeneous and inconsistent approach to text production. The more formats you accept for ingestion, and the higher the tolerance in deviation from the guidelines, the more effort will be required for the press to conform the manuscripts. On the other hand, shifting the burden completely to the authors might be a deterrent, so you have to strike a good balance. At Language Science Press, our impression is that we have demanded a lot from our authors. They have not always been happy about this, but our impression is that they feel that this is what you have to pay in order to get a high quality open access publication. Following the claims to prestige, we have probably also been stricter with regard to certain typographic subtleties than other publishers. This was surely annoying for the authors, but it also served to signal that we take publishing seriously and that we are not the right outlet for shoddy work.

3.5.3 Required elements

Books have a number of recurring elements. This may sound banal, but it is important to take stock of what is required here in order to be able to see which elements could easily be automated. For instance, the Language Science Press series all have an associated colour. We generate the title page based on the author name and the given title, use the background colour as specified by the series and add the series name on the title page as well. This process is completely automated. There is for instance no way how you could misspell the series name on the title page, or choose a wrong colour for a series, as both these elements depend on the series.

The list of elements to consider is:

- frontmatter
 - title page
 - imprint/colophon/masthead
 - frontispiece with series information
 - inner title page
 - table of contents
 - list of tables
 - list of figures
 - acknowledgements
 - list of abbreviations

3 Tasks

- dedication
- backmatter
 - references
 - indexes
 - ad
 - back title
 - blank pages
- mainmatter
 - text
 - headings
 - footnotes
 - crossrefs
 - bibliographic references
 - tables
 - figures
 - maps
 - charts
 - captions
 - page headers
 - epigrams
 - discipline-specific elements
 - * linguistic examples
 - * proofs
 - * theorems
 - * trees
 - * vowel charts

The more items from the list above can be handled in an automated way, the better. For instance, automated cross-references avoid manually updating the references if an item is added or removed. A citation manager allows to keep the text and the list of references at the end in sync etc.

3.5.4 Requirements for distribution

At the most basic level, we can distinguish digital distribution and paper distribution.

3.5.4.1 Requirements for digital distribution

For digital distribution, PDF is the format of choice. There should be a way for you to produce PDFs. For books in linguistics, the relative ordering of elements (horizontally and vertically) is often very important. This is why we require PDF. There are other digital formats like HTML or EPUB which might be suitable for other disciplines; for linguistics they are not suitable. You might also give some thought to XML as a specified generic backend format, or Markdown as an easy-to-use backend format with somewhat looser specifications.

PDF as a format is accepted by all major distribution platforms like DOAB, GoogleBooks or Zenodo.

The PDFs should be searchable, should not be encrypted and should be free of Digital Restrictions Management (DRM). Additional features like PDF bookmarks, hyperlinks and a clickable table of contents can be useful.

3.5.4.2 Requirements for analog distribution

There are a number of print-on-demand services available. They can be distinguished by

- geographical regions covered,
- shipping speed,
- quality,
- ease of use,
- setup costs,
- page costs,
- colour page costs,
- margin,
- available bindings (soft/hardcover),

3 Tasks

- available sizes (width/length),
- colour printing,
- page limits (min/max),
- distribution channels,
- own ISBN permitted.

As of 2018, Amazon CreateSpace has the best worldwide coverage and very fast shipping. The quality of the books is, however, rather low, and it has happened that we have returned more than 40% of a batch due to faulty quality. Depending on the number of books your press publishes, the time to upload a book for print-on-demand can become a factor, especially when it comes to filling in metadata. This can take up to half an hour per book. We now use BoD. BoD provides an FTP interface where we can automatically upload our books with the metadata in a specified XML-format.

BoD lets you choose your width and height freely. For CreateSpace, only a dozen predefined sizes exist.⁶ Other providers will only accept A5 or A4.

For print-on-demand, you can either bring your own ISBN, or you have to buy an ISBN from the service provider. Providers vary as to whether your book will be available only via them directly, via Amazon, or via local bookstores and standard distribution channels as well. In Germany, for instance, your book has to be listed in the *Verzeichnis lieferbarer Bücher*, which costs a nominal fee per book and year.

Note that most countries require you to provide some free copies to their national libraries. If you print in India for instance, you will have to pay shipping for these copies if your press happens to be elsewhere.

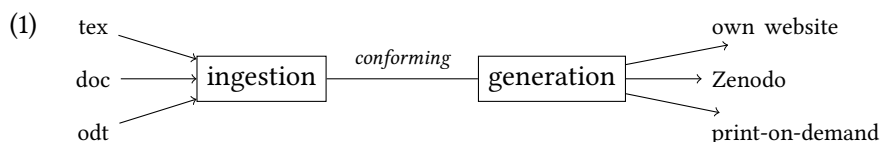
3.5.5 Requirements for archiving

The first requirement for your archive is that the copy be safe. You want to have trust that 100 years down the road, your copy will still be there. The second requirement is that the copy be accessible without too much hassle, ideally over the Internet, without restrictions. These are the central points. Further points are the richness of metadata you can provide, and the ease of uploading to the archive and updating metadata (e.g. further editions).

⁶12.7 x 20.32 cm; 13.34 x 20.32 cm; 13.97 x 21.59 cm; 15.24 x 22.86 cm; 12.85 x 19.84 cm; 15.6 x 23.39 cm; 16.99 x 24.41 cm; 17.78 x 25.4 cm; 18.9 x 24.61 cm; 19.05 x 23.5 cm; 20.32 x 25.4 cm; 21.59 x 27.94 cm.

3.6 Automation

Taking a mechanistic approach, we can see book production as a process which takes an input manuscript (ingestion), conforms it and produces well-defined output (generation), which is then made available at various outlets.



The more constrained your input is, the easier automation becomes.

3.6.1 Ingestion

It is a good idea to have clear criteria about the formats you accept. Will you only accept camera-ready copies (CRC) and offload the burden of book production completely on the authors? Or will you also accept less finished formats, which will then be conformed and finalised in-house? Or will you accept any kind of format and do the whole production from scratch in-house? The more restrictive, the less work you will have, but your number of submissions will also go down.

Potential formats which could be accepted at ingestion stage are: txt, docx, odt, tex, pdf, markdown. Graphics could be submitted as png, jpg, svg, idd, or pdf.

At Language Science Press, we stated that we would only accept \LaTeX submissions, which we would then finalise. This is thus not CRC, but very close to it. In general, we only do final typesetting with placement of figures, page breaks, and line breaks.

For the initial period, we proposed complimentary conversion of docx and odt submissions. In the beginning, this held for any document. As we progressed, we created templates which are tailored to our conversion pipeline. We now require that authors use these templates if they want us to convert their manuscript for them. As the number of submission progresses, the press can become stricter in enforcing the formal requirements.

3.6.2 Generation

This step describes the production of the various output format based on the input manuscript. Some background might be in order here: while word processors like Microsoft Word or LibreOffice use a WYSIWYG approach (“What you see is

3 Tasks

what you get”), this is not practical for multiple outlets. In those cases, it is better to use a separation of meaning and presentation, i.e. a WYSIWYM approach (“What you see is what you mean”). For instance, chapters in \LaTeX are marked with `\chapter`, but otherwise look like normal text. This information can be accessed for a variety of use cases. Most typically, one will use the chapter title to put the relevant words in a larger font on a particular page, but it is also used for the table of contents, the PDF bookmarks, and we also use it to extract a list of chapter titles when announcing a book to the community proofreaders.

We have a number of routines to create various outputs from the input files. These are bundled together in a Makefile. We can use commands like `make book`, `make index` or `make paperhive` to automatically run certain well-defined tasks. Among other things, the Makefile contains the following automated routines:

- `make pdf`: create the book
- `make index`: create the indexes
- `make complete`: create the book with the indexes
- `make proofreading`: add a watermark to all pages
- `make clean`: delete all unnecessary files
- `make chapterlist`: get a list of all chapters
- `make languagecandidates`: get a list of all words which look like language names

Additionally, there are a number of scripts which automate more complicated tasks:

- `doc2tex.py` convert a MS Word or LibreOffice document to \LaTeX
- `sanitycheck.py` check \LaTeX files for common violations of the guidelines
- `normalizebib.py` checks a bibliography for common mistakes
- `autoindex.py` marks up potential index terms
- `fixindex.py` repairs the author index for names with diacritics
- `chopchapters.sh` create PDF files for the individual chapters from a master file

- `linkchecker.sh` check all URLs in a document whether they resolve
- `checkppi.sh` check all graphics whether resolution is high enough
- `extractaw.py` extract download figures from server logs and create nice colourful charts per book and for all books

Finally, there are specialised commands for common types of charts and diagrams. Instead of creating them in a graphics program or a spreadsheet, we can directly use the data in the \LaTeX code to generate these elements. This also allows for easy extraction of the values for further computations if required.

Depending on workload, we also redraw graphics and maps in Inkscape. This is not automated, but there are routines for converting raster graphics to vector graphics. We have experimented with downloading geographical data from OpenStreetMap via Maperitive, export to svg and then edit the output. It has turned out, however, that simply tracing a raster image in Inkscape is normally faster than using the more granular OpenStreetMap data.

A desideratum would be a central metadata store which would feed the various outlets (website, PDF, GoogleBooks, Zenodo). Currently, metadata are stored in a variety of formats, and a change in the title of a book for instance must be replicated for all these formats, which is error-prone. It would be preferable if such a change could only be made once and then automatically propagate to all relevant outlets.

3.6.3 Outlets

Our main outlet is our website. There, we have to upload the PDF and the individual chapters manually. The same is true for Google Books, CreateSpace, and DOAB. A number of other tasks are automated:

- `make googlebooks:` prepare the content for upload to GoogleBooks
- `fetchompdata.py` fetches the metadata for a book from our website for cover creation
- `createbookblock.sh` remove cover and add blank pages to create a total number of pages which is a multiple of 4, required for BoD
- `podcovers.sh` counts the number of pages of the book and creates three covers with the correct measurements for CreateSpace, BoD hardcover and BoD softcover

3 Tasks

- `bodxml.py` creates the necessary PDF and XML files for upload to BoD and checks their validity
- `make bodrepo`: store all auxiliary files needed for print on demand via BoD to our private GitHub repository
- `make paperhive`: create the necessary files for PaperHive, upload them and tell PaperHive that this book is ready
- `amazonid.sh` creates referral links for Amazon
- `zenodo.py` create Zenodo records based on metadata extracted from the source files

In general, we are happy to add additional outlets as long as the incorporation of our content can be automated.

3.6.4 Archiving

Our books are deposited in the long-term archive of the Freie Universität Berlin library. Furthermore, they are archived at Zenodo. They are also deposited with OAPEN, which feeds them into DOAB and also serves as an additional archive. DOIs ensure that the documents can be matched.

3.7 Solutions

For all the technological tasks at hand, you have three basic options: use a computer at your institution to perform the task (in-house), use a webservice, or use an external service provider, often for a fee.

3.7.1 In-house

3.7.1.1 Presentation software

The software which takes care of the presentation of your books to the outside world will often be hosted by your institution itself. See Section 3.5.1. There is a path dependency: once you have settled for a solution, it becomes very costly to change course and choose another solution. You normally have to go down the path you first chose.

At Language Science Press, we use Open Monograph Press (OMP) by PKP.⁷ OMP is free software and available free of charge. It is closely related to Open Journal Systems (OJS), used for articles in Open Access journals. OMP offers a workflow with the phases “submission”, “internal review”, “external review”, “copy-editing” and “production”. As a matter of fact, the main functionality we use is the “catalogue” function. The workflow functions are used by some series, but in general, the interface is not intuitive enough for random users. As explained in Section 3.5.1, it might be the case that a detailed model of the phases of the workflow, and a software implementation thereof, is not part of your requirements. In that case, you could go for a simpler solution. If submission and acceptance are completely decoupled from the catalogue, a Content Management System like WordPress might actually be good enough to simply expose your books to the world.

Competitors to OMP include RUA (unclear status as of 2018)⁸ and Janeway⁹. If you go for a simpler solution without workflow management, you can also think about a static presentation, e.g. via WordPress¹⁰ or even plain Jekyll.¹¹

3.7.1.2 Generation software

In order to produce the books on our local computers, Language Science Press uses very powerful desktop computers with $\text{X}_{\text{Y}}\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ as the main workhorse. This is augmented by a number of scripts and Makefiles.

As a graphical user interface for editing the TeX files, we use Kile; for editing scripts, we use Kate. On operating systems other than Linux, TeXStudio would be a good replacement.

We use Inkscape to create vector graphics. An alternative would be Adobe Illustrator. We generally do not have the need to edit raster graphics. For most graphics files used in books, you should go for vector graphics. Only photographs should be edited in a raster-based program such as GIMP or Photoshop. Whenever possible, we create graphics directly in $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$, with pgfplots for charts and TikZ for diagrams and trees.

If you want to avoid $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$, you will probably have to use InDesign or QuarkXPress for qualitatively convincing typesetting. You could use MS Word or LibreOffice, but the results will not allow you to support your typographical claims to prestige.

⁷<https://pkp.sfu.ca/omp/>

⁸<https://github.com/ubiquitypress/rua>

⁹<https://github.com/BirkbeckCTP/janeway>

¹⁰<https://wordpress.org/>

¹¹<https://jekyllrb.com/>

3.7.1.3 Versioning software

Version control is a must for large projects such as the production of a book. If you want to run a versioning server yourself, GitLab seems to be the program of choice, although we have opted for a web service (GitHub) at Language Science Press.

3.7.2 Web services

For several years now, we are observing a move away from software running on the user's computer to software running on the computers of a specialised service provider, which is accessed via a web browser. This has the advantage that the users do not have to take care of the correct installation of their software, the dependencies it might have, and the security issues. It has the disadvantage that the users are no longer in control of what is happening. For critical processes, one should consider very carefully whether you want to entrust them to a computer you are not in control of.

At Language Science Press, we have opted for the following web services:

- Overleaf. This saves us from having to teach every author how to install \LaTeX on their operating system.
- Paperhive. This allows us to centralise all comments on a manuscript in one place, rather than sending around dozens of PDFs with commentary.
- GitHub. A reliable versioning server with much better performance or up-time than we could provide ourselves.
- Zenodo. A very user-friendly archiving service operated by CERN which is much easier to integrate with than with the archiving services of university libraries.
- Google Books. Allow people to read the books on the Internet. PaperHive seems more suitable for the task of reading online, but if there are people searching for us on GoogleBooks, they should find us.

Another web service which could be mentioned but which we do not use is GitBook.

One aspect which should always be considered when opting for a web service is the exit strategy: what will you do if the service shuts down, changes its product, or decides to charge astronomical amounts of money? Is there a way for you

to get your data back home? Are there competitors available which you could switch to? For instance, for our Overleaf project, we keep local copies of the books via the git bridge. If ever Overleaf shut down we would still have copies of our own available.

3.7.3 Outsourced

Next to doing the work on your own computers or on third parties' computers, you could also simply hire service providers to perform a specified task. This could be typesetting, copy-editing, indexing or graphics. We have occasionally had freelancers help us with conversion, but not on a regular basis. Major publishing houses routinely outsource their typesetting and other publishing services to Asia. At Language Science Press, we regularly receive unsolicited email from South Asian publishing service providers, but we have never been convinced by their offers.

3.8 Accountability

If you are spending public money, the public should know where that money went. If you are using crowdfunding, people will be even more eager to know what happened with their money. A community-based publisher should try to give back to the community as much as possible. The most obvious return would be books, but other by-products, like software, workflows or business models should also be made available.

3.8.1 Numbers

Numbers should never be the driving force behind a project, but it helps to have some quantifiable idea of where you are now and where you would like to be in the future, and to evaluate that idea periodically. We can distinguish internal monitoring from external reporting. Not all things you monitor are worthy to be reported. For instance, if your monitoring reveals that your tweets perform 6% better on Tuesdays than on Mondays, that is probably within the margin of error anyway and there is no need to spend resources on making that information available in a user-friendly way. This being said, you should think about what kind of reports you want to publish in what intervals. At Language Science Press, we publish download figures on a monthly basis, and we provide other data (number of series, number of expressions of interest, number of submissions, acceptance rates, followers etc) in January for the past year. Starting with this

OpenAire project, we will also report on our finances. See the spreadsheet and business model provided together with this cookbook.

Other projects might have different setups and hence will have other priorities in what they give back to the community. This is perfectly OK, but in any case, there should be a clear plan of what to give back when and how. Other topics which might be worthy of reporting are, for instance, the geographical or gender balance of authors/submitters/editors/series editors.

3.8.2 Technological

All software you produce should be made available for reuse via a common platform. GitHub would be one, but there are others. Choose an open license which allows for free use, modification and redistribution of your code.

Ideally, all software should be documented, but do make the software available even if the documentation is wanting. Access to badly documented code is better than no access to code. Do not wait for your code to be perfect; that will never happen. If your code is faulty and other people find out and tell you, the better.

Particularly useful or novel pieces of software should be covered on your blog. Invite others to contribute towards the development. More often than not, there will be no response, but it shows your commitment to openness and underscores the innovative aspect of your project.

3.8.3 Organisational

Setting up a community-based publisher is no easy task, and it is a rather novel concept. On the one hand, there is the repartition of roles: who does what; what are the tasks; who is responsible of which aspect; who relays information to whom. On the other hand, there are processes: what input is being transformed to what output via which pipeline?

3.8.3.1 Roles

Try your best to explicate these relations, even if they are somewhat murky. This is not about creating hierarchies or chains of commands. It is rather about making sure that everybody is on the same page about what they will be doing within the project. For instance, do authors communicate directly with the press directors, or is the communication mediated via the series editors? In Figure 3.2 on page 37, you find an organigram of the Language Science Press structure.

3.8.3.2 Workflows

The organisational structure is by definition static. The dynamic part is the transformation of the input to the output. The tool chains, pipelines or workflows should be made available as far as possible. The barest version of this would be Makefiles and scripts, where the succession of steps is cast into a succession of instructions to the computer. More explicit would be guidelines, HowTos and screencasts for different steps within the process. Note that software becomes outdated quickly. It is a good idea to have a periodical check of all your documentation material to see whether the information is still current (e.g. every year).

3.9 Money

One of the most pertinent questions of all time: Where does your money come from, and what are you spending it on? Is what your spending it on really worth the expense?

3.9.1 Revenue

For a community-based publisher, we can identify five major revenue streams:

1. processing charges
2. print margin
3. institutional memberships
4. individual memberships, and
5. donation.

Minor revenue streams are

6. affiliate/referral programs
7. prizes and awards
8. flattr/Patreon
9. grant money for research and further development

3 Tasks

10. and payout from copyright collecting societies such as VG Wort in Germany.

For more information, refer to the business model and the spreadsheets distributed together with this cookbook.

3.9.2 Expenses

The single most important cost factor is personnel. This is which will make or break your project. Other costs, such as IT, rent, or legal advice do not even come close. For more details, see the business model and the spreadsheets distributed with this cookbook.

3.9.3 Liquidity

If your project enters into relations with suppliers and service providers, or if it employs staff, you have to make sure to have the necessary money available when payment is due, so you should have some buffer if some revenue is late to come in for some reason. Print margins are paid monthly or quarterly, memberships are paid yearly, and donations cannot be predicted. Processing fees are normally paid by universities, which can take considerable time. You should have a financial buffer of 6–12 months to be on the safe side.

If ever you have liquidity problems, you could consider a loan, but probably some natural person will have to vouch for the loan. Since no one derives profits from the operation of a community-based publisher, there will probably be no one to vouch for a loan, so you should avoid the necessity of loans at all costs. This is different than for-profit companies, where loans are a common form of assuring liquidity.

3.9.4 Taxation

Your revenue streams might be subjected to taxes, especially VAT. In most countries, there is a standard rate (19% in Germany) and a reduced rate (7% in Germany). Depending on the goods you sell, one or the other will apply. Cars for instance get the standard rate in Germany, while books get the reduced rate. Your legal form can also have an influence. If you are a charity and your product is in line with your charitable goals (e.g. furthering of science), the reduced rate is used for that product as well. There are some other taxes, too, which can be waived for charities. If you employ staff, your jurisdiction might require that you levy their income taxes and forward them to the State.

Your company must be registered with the tax office (Finanzamt, IRS, HMRC).

There are all kinds of thresholds and exemptions, so get in touch with a tax advisor about these matters.

3.9.5 Bank accounts

A bank account is a very useful thing to have, and getting one can turn out surprisingly difficult. You want to be able to receive money, to send money to arbitrary accounts worldwide, and you want to have a credit card. Depending on the legal form you choose (see below, Section 3.10.1), this can be more or less difficult.

If you stay with your university, you might be denied a credit card. For example, this is the policy for both Freie Universität and Humboldt-Universität in Berlin. This means that you cannot buy flight tickets for instance, and it also means that you cannot use certain web services at all. CreateSpace e.g. requires you to have a credit card. The only solution in those cases is to use a personal credit card and demand reimbursement. This can take a long time depending on the speed of the university administration, and it can easily go into the four or even five-digit range.

Receiving money on university accounts has turned out to be difficult at Freie Universität Berlin as well. The university has a central account and asks you who you think has sent what amount of money with what stated purpose to that account. E.g. “Amazon Germany should have sent 123.34 EUR between May 8 and May 13 with the subject ‘Print margin ABCDEF12345’. Please check.” Especially for smaller sums, this is hardly worth the effort. Things get even more complicated when other currencies are used. In that case, you can only state “some amount in the 120 EUR range”, and a clerk will have to sieve the account history to retrieve it.

If you launch a company, you can open a bank account at a financial institution of your choice. There are some catch-22s (you have to have a bank account to launch a company, but you have to be registered as a company to create a bank account), but there are usually ways to solve that. At least in Germany, accounts for companies are far more expensive than standard personal accounts. Depending on your legal form, you might also have difficulties in finding institutions who want to do business with you. As a rule of thumb, the more of a stated charity you are, the more difficult will it be for you to get a bank account.

You should pay attention to whether you can name several persons with access to the account. Furthermore, some accounts come with an obligatory debit card for a fee, which you might not need. Related to this is the question whether you

3 Tasks

will need cash. Some banks are exclusively online, and are very good with all things electronic, but it can be difficult to get cash from them, and even worse, to deposit cash *into* your account.

3.10 Legal

3.10.1 Legal forms

In order to enter into contractual obligations with authors, institutions or service providers, you have to have a legal form, i.e. a juridical person of your own. We can distinguish the broad types of companies, foundations, and charities. Depending on your jurisdiction, the types might be clearly distinct or mixed forms might be possible (e.g. charitable company). As long as you have not created such an entity, you might be able to operate under the umbrella of your home institution or your institute. Bear in mind, however, that academic institutions are normally equipped to spend money, but that they might not be used to receiving money. Finding someone who can produce an invoice can be exceptionally tough as there are legal, financial, budgetary and academic questions involved, and buck-passing between the departments involved is a common risk.

In the German context, launching a company costs 1 EUR in capital and about 300 EUR in legal fees for the registration, or 1000 EUR in legal fees if your company is not run-of-the-mill (e.g. a charity). Creating an association (*Verein*) takes 7 founding members and costs about 200 EUR. Creating a foundation requires substantial initial capital, is rather complicated and involves a lot of red tape. Things are completely different in the Netherlands, where foundations are comparatively simple to set up. If your project involves several countries, it might be worthwhile to shop around.

3.10.2 Brands

Scholars should be able to recognise your publishing platform. They should associate it with high quality, prestige, innovation, and community. This is equivalent to saying it should be a **BRAND**. In academia, there is often some reluctance to talk about brands. Multinational corporations come to mind. Small-scale publishing initiatives normally have no such aspirations. In the context which matters here, “brand” simply means that a name is associated with certain desirable features, e.g. prestige or innovation. This is neither good nor bad, but simply a shorthand for referring to this particular association.

You can register a brand. This means that you receive the sole right to use this brand for a specific sector (e.g. publishing) and a specific geographical region (Germany, Europe, the World). The registration is for a fee, which increases with the number of sectors and geographical coverage. Count a range between 200 and 1200 EUR.

Brands can be used offensively or defensively. When used offensively, you actively want to make sure no one enters your sector with a similar name or logo. You are prepared to sue them if they do. Defensive use would be more common in academia. You register the brand simply to be sure that new entrants cannot sue *you*. We registered the brand “Language Science Press” in 2015. When another publishing house complained about our operations in 2018, we are able to remain calm since we had the registered brand and they did not, hence their complaint did not go anywhere.

It is your duty to check whether similar brands exist. You could always try and register “Elsuvier” as a brand, but Elsevier can have your brand deleted then, at your cost, so you better do your homework before. The legal department at your university might help you with the research for conflicting registered brands. Otherwise, there a lot of lawyers offering package deals for brand registration.

In the context of brands, it is also important to note that the FAIR Open Access Principles has at #1 “The journal has a transparent ownership structure, and is controlled by and responsive to the scholarly community”¹² with the added explanation “Key points are that the controlling organization, not a commercial publisher, must own the journal title”. In our case, that would not be the journal title of course, but the name of the publishing platform, and registered ownership means a registered brand.

3.10.3 Intellectual property rights (IPR)

3.10.3.1 Author contracts

Under the current copyright laws, authors own the rights to their creations. Parties which want to make available an author’s work have to get the explicit permission to do so. Traditional publishers often solve this via “copyright transfer agreements”. All rights pertaining to a work are transferred to the publisher, who can then do with the work as they see fit. Other parties, including the author themselves, are barred from using the work as a consequence, unless they obtain explicit permission from the new rights holder, i.e. the publisher.

¹²<https://www.fairopenaccess.org/>

However, such copyright transfer is not strictly necessary. A more granular approach, only granting the rights necessary for the particular service a publisher provides, is sufficient as well. In this case, the author would retain their copyright. This granting of rights can either be specific (“publisher X is allowed to do this and that”) or global (“anybody is allowed to do this and that”). The latter approach is the one espoused by Creative Commons.¹³ An author who licenses their work under a Creative Commons license grants everybody the right to use the work under the terms specified by the license (e.g. attribution of the source). This also means that a particular contract between the author and the publisher is no longer required if all services the publisher performs are covered by the Creative Commons license.

The Fair Open Access principles state as #2 and #3:

2. Authors of articles in the journal retain copyright.
3. All articles are published open access and an explicit open access licence is used.

The author contract used by Language Science Press is very short. It is given in a translated and slightly abridged version in Figure 3.1. This contract is provided for illustration and should not be taken over without consulting a legal expert.

The author contract must allow hosting the book on your servers and to have it printed via print-on-demand. Print-on-demand service providers ask you to confirm that you have the necessary rights to get a document printed. All print-on-demand service providers are commercial entities. Note that Creative Commons non-commercial licenses will effectively ban you from offering print-on-demand, so they should not be used.

You grant the rights to reproduce the work in print to the print-on-demand service providers. Read their terms carefully and see if they want exclusive or non-exclusive rights. You are generally not in a position to provide exclusive rights since, under the Creative Commons license, anybody can use the work and you do not have the power to restrict these usages. If the service provider asks for exclusive rights, get in touch with them and explain the situation for Open Access. Chances are good that they will grant you an exemption.

3.10.4 Hosting

In order to make the content available on your own website, you have to have the necessary rights as discussed above (Section 3.10.3.1). The Creative Commons

¹³<https://creativecommons.org/>

**Author's agreement
between**

(in the following "author")

and

Language Science Press

The goal of Language Science Press is to make results of top level linguistic research available to all readers without cost to these readers.

All publications on the Language Science Press platform are published under the Creative Commons CC-BY 4.0 Germany licence (or any newer version). The complete text of this license can be found at <http://creativecommons.org/licenses/by/4.0/>. This licence gives everyone (including the author) the right to copy the work, distribute it and make it available online without a fee and without individual consent by the creator. To this end, the users of the publication are granted non-exclusive, temporally and geographically unrestricted rights of usage under the conditions of this licence (proper attribution of the author). The licence chosen allows third parties to use the work, including commercial use, to modify it and to change it (e.g. translate it into another language).

§ 1 Object of this agreement

This agreement covers the work entitled

(referred to as "work" in the following)

§ 2 Grant of rights to use

1. The author grants the non-exclusive, temporally, geographically and content-wise unrestricted right of usage to Language Science Press allowing Language Science Press to replicate the work in all known or hitherto unknown kinds of usage, to distribute and to show it in public, especially to make it available publicly (online). Language Science Press may further transfer or sublicense these rights to third parties, particularly the right to make it available under the chosen CC-licence.

§ 3 Warranty of necessary rights

The author guarantees that they are in a position to transfer the copyright concerned, that the usage mentioned above does not infringe third parties' rights and that the work and all its components do not infringe third parties' rights, in particular in what concerns intellectual property rights or personal rights. This includes all materials included in the work concerned (e.g. photographs, graphics). In case Language Science Press is sued for an infringement by the author's work, the author shall fully indemnify Language Science Press from any liability and any costs incurred, including legal costs.

§ 4 Royalties

In consideration of Language Science Press being a not-for-profit enterprise, and considering that the work is made available without a fee under the CC-BY licence mentioned above, no royalties will be paid.

§ 5 Final clauses

1. Changes or amendments to this text must be done in written. 2. German law applies. 3. If any part of this agreement is declared unenforceable or invalid, the remainder will continue to be valid and enforceable.

Figure 3.1: Language Science Press author contract

license generally solves this. If you have user authentication, or if you use other means to identify your users personally, you will have to make sure you comply with the necessary privacy legislation in your jurisdiction. Note that Google Analytics and similar tools can be considered as identifying users. The easiest solution is always to simply not collect any user information (privacy by design). The less personal information you collect, the less legal requirements you have to respect.

There are a number of additional legal requirements which vary per jurisdiction. This concerns the contact details of the person responsible for the site, contact of the person responsible for the technical functioning of the site, and some information relating to taxation and information about privacy laws. The rules evolve quickly in this domain, so you should get qualified advice in this domain.

3.11 Governance

3.11.1 Internal

Your publishing project should have a structure which allows you to ensure that scientific, financial, organisational and ethical aspects are all well handled. The organigram of Language Science Press is given in Figure 3.2. The main roles are the series editors, who take care of scientific aspects; the coordinator, who takes care of financial aspects and day-to-day organisation; and the press directors, who oversee scientific quality and organisational matters. General advice and ethical aspects are the domain of the advisory board.

It pays to have a well-established structure with well-defined roles. In our case, the coordinator is free to enter into any contracts with service providers, attend events, organise promotion, etc; the series editors are free to accept manuscripts according to the standards of their discipline without interference from elsewhere; etc. This autonomy reduces friction and increases speed. Multiple roles (e.g. a press director who is also an author) cannot always be avoided but are a potential source of conflict: a series editor might find it hard to objectively assess the quality of an author's manuscript if that author happens to be the press director.

In the structure chosen by Language Science Press, the coordinator's role is central. They serve as a dispatcher and as a buffer for incoming demands. Ideally, they can solve as many requests as possible without resorting to other organs. One thing which Language Science Press treats explicitly as out of bounds for the coordinator is scientific quality. All questions pertaining to scientific quality will

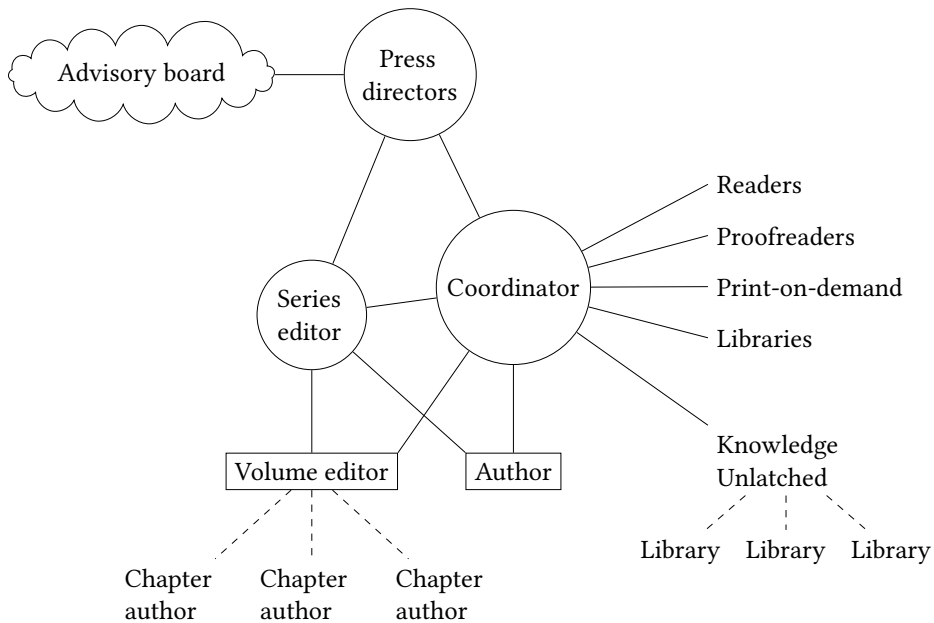


Figure 3.2: Language Science Press organigram. The main triangle is formed by the coordinator, the press directors and the series editors, who constantly are in close contact. The advisory board is consulted on special occasions. Authors and editors are only in contact with their specific series editors and the coordinator. Chapter authors liaise with their volume editor. On the right hand side, there are various stakeholders more peripheral to the book production process. Solid lines symbolise direct connections where Language Science Press can influence; dashed lines stand for indirect, mediated influence.

be relayed to series editors or press directors; the coordinator will never respond on their own.

The main tasks of the coordinator are day-to-day management, responding to inquiries, strategic development, liaison, and advocacy. These functions could possibly be split into two positions with one more inward-facing while the other one faces the outside world.

3.11.2 Relations to home institutions

The relations of the project to its home university can become a matter of concern as the publishing platform evolves. In the beginning, it will be only a project among many others, but as it progresses, it might outgrow the frame of an institute of a faculty. For instance, it would be strange if a smallish institute for

3 Tasks

Scandinavian Linguistics had 3 staff for publishing next to only 1 post-doc and 1 PhD student. In those cases, a transfer of the project to a larger structure (the university library or IT service), or to an own company should be considered. See Section 3.10.1 on legal forms above.

3.11.3 Organisational matters

The project will end up with lots of access tokens/passwords for different sites (versioning, hosting, archiving, print-on-demand). Make sure that more than one person can gain access to each site, e.g. by keeping an offline log of passwords or by having several accounts for different people for each service.

Your project will acquire institutional knowledge, i.e. know-how. This will be implicit at first: some employee will have understood how some aspect works best. Try to transform this implicit knowledge into explicit knowledge by maintaining a collection of recipes and HowTos, e.g. in a wiki. Make selected recipes available to the outside world in the form of manuals and guidelines to share your insights with the community. Establish a regular update interval, or a particular day every year where you make sure that your recipes and guidelines are still up-to-date.

Next to ensuring that your access credentials will survive the dropping out of one person, you should also define a data backup procedure which specifies what to back up at what intervals.

3.12 Community involvement

3.12.1 Overview

If the goal is to set up a community-based publisher, getting the community involved is obviously a prime task. But “community” does not have a clear definition as such. Basically, everybody who feels concerned is part of the community. That being said, we can obviously distinguish members based on seniority and academic position. On the one hand we have junior community members such as students, on the other hand, we have tenured professors or even emeriti and emeritae. For a community-based publisher, it is important to give all of them a sense of belonging. People want to be part of groups, and they want to contribute to the well-being of their group. How they can contribute varies with academic experience. An undergraduate cannot review a handbook, and a professor will probably not correct bibliographical entries. But the book production process is full of different tasks, and there is something in it for everybody. The challenge

is now to find a setup where everybody can contribute. It is good to have some sympathetic bystanders, but in order to really form a movement and make the community their own cause, they must invest some time, however small.

Expertise is not mono-dimensional. You could be a very good theoretician, but have horrible writing; or you could be a didactical genius, but have absolutely no graphic skills. Glossing over these differences for a moment, we can still say that the most experienced community members will be in the advisory board which vets new series. They have to have a very good overview of the field of your discipline as a whole, and they have to have sufficient publication experience to be able to see whether a new series proposal holds water. Next would be series editors, then editorial board members and reviewers. The least specialization is required in community proofreading. Basic familiarity with scientific writing is enough to see whether all tables are properly referenced and whether the citation guidelines are adhered to.

At Language Science Press, there are 24 people on the advisory board, about 200 editorial board members, and over 300 community proofreaders, mirroring the specialization and education required to perform the different tasks.

Somewhat orthogonal to that would be community typesetters and community illustrators, but we have not really had a lot of success in recruiting those. The same is true of volunteers to improve software. We can establish that in book production proper, it is possible to involve the community at various stages of their career, but for more peripheral aspects, like illustrations, this seems to be less straightforward.

3.12.2 **Recognition**

It is important that people receive some formal recognition of their being part of the community. We list all advisory board members, all editorial board members, and all community proofreaders who have worked on at least one published book. This performative act of formally including them is very important to tighten the bonds. For community proofreading, we also have a Hall of Fame, which ranks the proofreaders according to how many books they have contributed to. This is clearly a game and not very serious, but still, people like to see whether they are bronze, silver, or gold proofreaders, next to their intrinsic motivation of reading new works in linguistics and providing service to the discipline. The coordinator thanks all proofreaders personally when they finish their task, and some series editors send extra notes of gratitude in addition. All proofreaders and typesetters are additionally mentioned in the colophons of the books they contributed to. Being listed among many other linguists, some of them very famous, also creates

3 Tasks

a strong link with the community. Finally, formalised recognition makes it easy to provide service to the discipline, for instance when applying for a job.

3.12.3 Social media

Social media channels like Twitter or Facebook are a good means to stay in touch with community members. Do not use them as a one-way road to simply broadcast your new publications, but be receptive of input and participate in discussions. Do not let your social media presence grow stale. Do not open accounts you do not intend to use. If you use them, make sure that you use them for true interaction with the community. Both Twitter and Facebook provide some metrics about your interactions. We set our target for Twitter to 1000 profile visits a month, 30k impressions and 42 new followers. These are no hard targets, but if ever we were significantly below for several months in a row, we would increase our presence. Facebook is clearly our secondary social media channel and we devote a lot less attention to it. The main reason for this is that the coordinator has no experience with Facebook. If another person was hired, that might change.

3.12.4 Inquiries

All inquiries from the community reaching the office (book proposals, questions, suggestions) should be handled as speedily as possible. An acknowledgement of receipt should be sent within one working day, ideally one hour. This shows that you care about input. The actual full response can be sent later. This is OK if people know that you are working on the issue.

If you have more than one person in charge of answering to inquiries, you might want to set up a ticketing system like OTRS to make sure nothing gets overlooked.

3.12.5 Newsletters

Do inform the community about current events like new publications, new series or awards. Scientists are used to fact-based communications. In our experience, there is no point in glossy newsletters. We have monthly newsletters. If we have nothing to report, we skip a newsletter. We do not try to artificially generate content. Still, it is important that supporters know that things are happening in their community.

3.12.6 Merchandise

There are easy and cheap ways to have your logo printed on T-shirts, caps, mugs and the like via some Internet services. This gives people a way to show their support. For Language Science Press, the uptake has been close to nil, but it has not really cost a lot of effort either.

3.12.7 Series editors meeting

Between 2014 and 2017, we had one series editors' meeting a year. For three of these occasions, travel and accommodation were covered. This was very good to form bonds between the different series, and to see what problems they shared and what solutions one series had found could be adopted by other series. It is also a very good occasion for bootstrapping a newly accepted series and introduce them to the wider context of community-based publishing.

3.13 Branding and promotion

As stated in Section 3.10.2, the association of the name of your project with prestige, quality, etc. gives birth to a brand. This happens without your even wanting it. The effects and relevance of brands as a means to structure our perception of different goods or services must be recognised, even if one fundamentally disagrees with many strategies used by large corporations.

Next to identifying these values in the first place, they should also be transported to the outside world, via media and face-to-face interactions.

3.13.1 Values and claims

It is a good idea to identify some central values your projects stands for, e.g. “quality”, “openness”, “community”, and “innovation”, and make a conscious choice of highlighting those in your outward communication.

As argued in Section 3.1, a top-down approach starting with high-quality publications is most promising. Therefore, the claim of quality should feature very prominently in your communication. Think of a catchy phrase you can use repeatedly. For Language Science Press, this was “We want to play in the Champions League of publishing”.

3.13.2 Design

A uniform appearance of your platform across different media types (books, website, flyers) supports the entrenchment of your brand. Many projects have giveaways like notepads, bags, or pens. Since “no frills” is also a main tenet of Language Science Press, we had none of those, and we do not feel that a Language Science Press notepad would really sway people to publish with us.

3.13.3 Social media

Your social media presence should be in line with your branding and mirror the overall style.

3.13.4 Conferences

Conferences are a good way to make your project known. On the one hand, there are discipline-specific conferences, where you can showcase your books and get in touch with prospective authors. Very often, your series editors can serve as “conference ambassadors”. On the other hand, there are general (Open Access) publishing conferences. You will not get any authors or readers from there, but you can require a reputation as an innovative publisher among librarians, archives, aggregators, service providers, etc. Representation in those venues has to be done by press directors or the coordinator. Finally, there are policy conferences at state, national, or European level. If you have performed well in the publishing conferences, you might be invited as a stakeholder to these workshops, which are generally by invitation only.

3.14 Network

Obviously, you should get in touch with the national and international bodies representing your discipline and try to get their official endorsement. Also try to reach out to subcommunities you are normally less in touch with. So, if you are in “Applied X”, reach out to “theoretical Xers” and vice versa. Your network should also cover community-based publishers in other fields and other countries, libraries, archiving initiatives and the relevant legislative bodies. Language Science Press has loose contacts to practitioners in the fields of Archeology, Theatre Studies, and Law, for instance. Furthermore, broader initiatives like the Open Knowledge Foundation, Wikimedia, Chaos Computer Club, the Free Soft-

ware Foundation or the Electronic Frontier Foundation are good to be in touch with.

You can also grow your network by teaching about Open Access publishing, for instance in graduate schools, or by providing input for courses in Business Administration, IT, or Library Science.

3.15 Competition

In our experience, traditional publishing houses are not much more knowledgeable with regard to Open Access publishing than you. They are also trying to find out and are experimenting. There is no reason to hate them or badmouth them. Be nice, and you can learn from each other's experiences.

4 Readings

There are a number of reports dealing with the economics of open access publishing. The following list makes no claims of exhaustivity:

- Martin Paul Eve. 2014. *Open access and the humanities: contexts, controversies and the future*. Cambridge: Cambridge University Press DOI:10.1017/CBO9781316161012
- Nancy L. Maron et al. 2016. *The costs of publishing monographs: Toward a transparent methodology* DOI:10.18665/sr.276785
- Janneke Adema & Graham Stone. 2017. *Changing publishing ecologies: A landscape study of new university presses and academic-led publishing* <http://repository.jisc.ac.uk/6666/1/Changing-publishing-ecologies-report.pdf>.
- Eelco Ferwerda et al. 2017. *A landscape study on open access and monographs: Policies, funding and publishing in eight European countries*. Knowledge Exchange DOI:<https://doi.org/10.5281/zenodo.815932>
- Rob Johnson et al. 2017. *Towards a competitive and sustainable OA market in Europe: A study of the open access market and policy environment*. Research Consulting <https://blogs.openaire.eu/wp-content/uploads/2017/03/OA-market-report-28Final-13-March-201729-1.pdf>.

5 Timeline

The following is a rough timeline which has the official launch as reference point. There are a number of task which precede the official launch while others follow

- 2y form a core network of about 4-7 seven people from different subdisciplines. They will be responsible for building up the crowd
- 2y start promotion. Set up a website, social media channels, and send personal mails to colleagues. Start with your immediate network and broaden the scope to include more remote connections as you go along. All mails should be personal and highlight your individual motivations to start this enterprise.
- 1.5y You should now have a dedicated core group and several dozen interested researchers. Have a virtual meeting to see where everyone stands and what their interests and motivations are
- 1.5y set up a web page where people can become supporters and state that they want to publish with you. Collect manuscript titles from interested authors.
- 1.5y start collecting volunteers for other roles, e.g. typography, illustration, community proofreaders.
- 1.5y Research funding opportunities for the setup phase
- 1.5y Get in touch with designers, typographers, legal experts, IT people.
- 1y Form 5 initial series with editors, editorial boards, and description of aims and scope.
- 1y Set up your work flow.
- 1y Invite first manuscripts from among your network. These are invited manuscripts, so they will not be rejected. Still, reviews should be commissioned in order to identify potential weaknesses

5 *Timeline*

- 1y Submit grant proposal
- 6m Receive grant approval (hopefully)
- 6m Look out for staff to fill your grant positions
- 6m Revision of first titles starts
- 3m Typesetting of first title starts
- 3m identify potential revenue streams
- 1d First book is typeset
 - 0 launch at some larger event, e.g. annual meeting of relevant learned society.
 - 0 publish first book at the launch. It is sufficient that the book be available somehow afterwards.
- +3m get external help for working out your funding model
- +1y publish the majority of your invited volumes.
- +1y bang your own drum. Try to acquire more expressions of interest for books and series. Publicize the interest you receive.
 - After the first year, paths will diverge. Take an open and agile approach and follow impulses from the community. You are on uncharted terrain anyway, so it is easy for you to change course if this means that you get some more dedicated people on board. Community-based publishers should be run by the community, so try to involve them as much as you can.

6 Time sinks and lessons learnt

There are two single items which take much much longer than expected. Try to enforce standards here as rigidly as you can, and try to automate as much as possible. These two items are bibliographies and edited volumes.

Manuscripts written with a citation manager like BibTeX or Zotero are generally fine. But of all other authors, 95% will have mismatches between their prose text and their list of references. Different spellings of names, different year in the text and at the end, or works cited in the text but missing in the list of references or the opposite are all common. This permeates all career levels. Full professors provide sloppy references just the same as graduate students. Chasing them down and trying to establish what “Jones (2000)” was actually intended to mean, choosing between “Jones (1999)”, “Jones and Doe (2000)” and “Johns (2000)” is incredibly time consuming. This is even worse for edited volumes, where contact with the authors is only indirect, via the volume editors. Unless you want to take over the complete editing of the volume, we advise against sorting this out directly with the chapter authors.

Edited volumes are also much more time consuming than one should expect due to the communication overhead. Double the chapters and time will increase fourfold. Our experience is that up to 12 chapters or so, things go smoothly, and beyond that number, things quickly get unmanageable. Authors are on leave and do not respond, everybody expects everybody else to be late and responds late as well, etc.

7 Conclusion

This book has presented a *tour d'horizon* of our experiences in setting up a community-based publisher, Language Science Press. We have tried to cover all the different domains we touched and the things we learned in order for other people to be able to copy our setup. There might be remaining implicit assumptions, things which were obvious to us, but which cannot be taken for granted. If you find some of those, please let us know and we will update the document. The same is true if you would like to share some experiences from other fields, where the culture might be different and where the recipes from linguistics cannot be applied.

This document is probably most useful when accompanied by the business model, the business data and the 5-year spreadsheet to calculate your own project, which are all distributed together with this book.

Cookbook for Open Access books

This book describes the experiences of setting up a community-based publisher, Language Science Press. It discusses the main principles of community-based publishing and gives a very granular breakdown of the different tasks. The discussion of the different tasks is complemented by readings, time lines, and a list of time sinks.

This book is complemented by the business model, open business data, and a spreadsheet for drafting and calculating own business models.

ISBN 978-3-96110-089-7

