



ride. A review journal for digital editions and resources

published by the IDE

Litteraturbanken

Litteraturbanken, Mats Malm, Cai Alfredson, Dick Claésson, Paulina Helgesson, Anja Hellström, Carl-Johan Lind, Ellen Mattson, Ljubica Miočević, Therese Röök, Ilaria Tedde (ed.), 2004ff.. <http://litteraturbanken.se> (Last Accessed: 27.07.2017). Reviewed by Mats Dahlström (University of Borås), mats.dahlstrom (at) hb.se and Wout Dillen (University of Antwerp), wout.dillen (at)uantwerpen.be.

Abstract

Litteraturbanken (The Swedish Literature Bank) is a freely available digital collection of Swedish literary works, ranging from medieval to contemporary literature. It is the result of a cooperation between literary and linguistic scholars, research libraries, and editorial societies and academies. The collection consists not only of digital facsimiles, but of ocr'ed, proof-checked and TEI-encoded transcriptions as well, including EPUB and HTML versions of texts, and in addition scholarly presentations and didactic introductions to works and authors in the collection. It is also being used as a publishing platform for ongoing Swedish scholarly editing projects. *Litteraturbanken* currently comprises more than 2.000 works, mounting up to more than 100 million of machine-readable words. *Litteraturbanken*'s main weak spot is transparency; it does not openly provide satisfactory ways to ensure the editors accountability for the edited texts and images. As a whole, however, *Litteraturbanken* is an impressive endeavour and paves the way for fruitful cooperation and massive data exchange with e.g. computational linguistics and bibliographic databases.

1 This review evaluates *Litteraturbanken* (in English: the *Swedish Literature Bank*; <http://litteraturbanken.se>) – a digital text collection of Swedish literary works.¹ The review builds to some extent on an earlier evaluation report by Mats Dahlström and Johan

Eklund, commissioned by *Litteraturbanken* in 2011 (Dahlström and Eklund 2011), but primarily takes stock of *Litteraturbanken's* more recent developments, with respect to its contents, interface, and advanced functionalities.²

General Parameters

2 *Litteraturbanken* is an organization that aims to make the most important works of Swedish literature freely accessible to the general public through its website, litteraturbanken.se. It is gradually being recognized as the most prominent collection of digitized literary works in Sweden, and is the result of a co-operation between the *Swedish Academy*; the *Swedish National Library*; the *Royal Swedish Academy of Letters, History and Antiquities*; *Språkbanken* (the *Swedish Language Bank* of the University of Gothenburg); SVS (*Svenska Vitterhetssamfundet*, translated as the *Swedish Society for Belles-Lettres* in English); and the *Society of Swedish Literature in Finland*. This composition is also reflected in the *Bank's* board, which includes members of each of these institutions.³

3 Discussions on the possibility of developing *Litteraturbanken* started in April 2002, when the *Swedish Academy* hosted a seminar on the digital storage of literary works. At this meeting of representatives of various libraries, universities and literary societies and associations, the benefits of such a digital collection of Swedish literature were widely acknowledged, leading the *Academy* to initiate an official investigation of the matter, which in turn led to the successful application for a two-year pilot project (2004-2005) financed by the *Swedish Foundation for Humanities and Social Sciences*. In 2006, when this project was finished, *Litteraturbanken* took its present form as a non-profit organisation,⁴ the development of which was funded by the *Swedish Academy*. Since 2009, the *Royal Swedish Academy of Letters, History and Antiquities*, the National Library, and SVS have supported *Litteraturbanken* as well by making resources and materials available.

4 As a Swedish enterprise, backed by Swedish (or Swedish language) institutions, collecting a corpus of Swedish language texts, it is not surprising that litteraturbanken.se is almost exclusively written in Swedish as well. Where the website includes lengthy and informative discussions of the general parameters in Swedish,⁵ this information remains difficult to access to people who do not read the language. To find a short, summarizing introduction to *Litteraturbanken* in English, the user needs to navigate to a Swedish

language menu item ('om LB,' which translates as 'about *Litteraturbanken*'), and then to a sub-level menu item called 'In English'.

5 This is all quite understandable given *Litteraturbanken's* national context and the language of its corpus. A person who cannot even find the English introduction would be hard pressed to make good use of the text collection in the first place.⁶ But still, there is a practical concern here too. At present, including more English in a collection's interface is also a sure-fire way of improving its outreach. The Norwegian *eMunch* project for example noticed a considerable increase in their international website traffic when its interface extended its official languages to English and French – despite the fact that practically all of the content in its corpus is in Norwegian.⁷ For *Litteraturbanken*, such a strategy could become especially pertinent if the aim is to also include translations of Swedish literature in the future – which is one of the goals mentioned on its English introduction page.

Goals

6 *Litteraturbanken's* mission, as stated on the website,⁸ is to collect and digitize fiction and other works that are important in the humanities, and to make these materials available in a user-friendly way. As such, it allows the user to read and search text, to download (and print out) PDF files, and to view facsimile images of books and manuscripts that would otherwise have been difficult to access.⁹ The specific goal here is to develop litteraturbanken.se into a website that holds the most important Swedish literature. But rather than only *representing* a national corpus, it seems to have taken on the challenge of *extending* the canon as well. Exactly how the texts that make up this corpus are selected will be discussed in more detail below (see: 'Content').

7 But *Litteraturbanken* aims to be more than a mere archive of texts. Navigating through the website it becomes clear that it also wants to function as a resource for research and education;¹⁰ to develop and disseminate technology and competences; to serve as an archive and publishing platform for Swedish scholarly editions; and to make obscure or hard-to-access materials available that the staff deems to be culturally significant.¹¹ This way, it tries to be a resource for everyone: for researchers, for teachers, for students, as well as for a heterogeneous public that is just interested in (Swedish) literature in general.

8 This extensively comprehensive approach is certainly a strength: acknowledging the fact that its source materials have many different qualities and may serve many different purposes, a lot of time and effort is put into making sure that they cater to the widest possible audience. As we will explore further down in this review, however, this qualitative ambition will sometimes clash with *Litteraturbanken*'s quantitative ambitions – the diversity of its content and functionalities sometimes makes it difficult to fully grasp its aims and scope.

9 This is also what makes it difficult to put a label on *Litteraturbanken*. If we use Ulrike Henny and Frederike Neuber's definition of the term, *Litteraturbanken* may be regarded as a digital text collection: indeed, it is undoubtedly a 'digital resourc[e] that involve[s] the collection, structuring and enrichment of textual data' (Henny and Neuber 2017). But of course, as they acknowledge in their "Preliminary Remarks," this definition is meant to be a generic term that comprises a number of different types of collections like corpora, archives, libraries, repositories, databases, etc. If we take a look at the way *Litteraturbanken* self-identifies instead, we find terms like resource ('resurs')¹² and storage place ('lagringsplats'); in other words, *Litteraturbanken* is presented as a place where digitized versions of literary works are deposited for safekeeping and consultation.

10 This could also explain why *Litteraturbanken* calls itself a 'bank' for Swedish literature: there is a certain analogy with safes and safety deposit boxes to be found here. And with insurance policies too, as *Litteraturbanken* tries to ensure the long term sustainability of its literary canon. But still, a bank is usually a more interactive space: not just for deposits, but for withdrawals, loans and exchange as well. And as we will argue further down in this review, the interactive aspect of *Litteraturbanken* still leaves much to be desired. So at this point in time at least, perhaps *The Literature 'Trust'* would be a more appropriate translation than *The Literature 'Bank'* – especially also since deposits to the 'Bank' can only be made by vetted trustees at the moment.¹³

Litteraturbanken in Context

11 As a national collection of high-quality digitized classical literary works, *Litteraturbanken* has several Northern European counterparts, some of which served as inspiration during the conception phase (Svedjedal 2003, 24), such as the Danish *Arkiv for Dansk Litteratur* (ADL), the *Deutsches Textarchiv*, the Norwegian *Bokselskap* and the Swedish *Projekt Runeberg*. Further Nordic projects with a similar scope are planned or being launched, such as the *Klassikerbiblioteket* at the university of Helsinki.¹⁴

12 By and large, these collections share ambition and comprehensiveness with *Litteraturbanken*.¹⁵ ADL for instance houses classical Danish works that are out of copyright, based on scholarly edited transcriptions, and made available as both HTML and as digital facsimiles. Similar to *Litteraturbanken*, ADL is a cooperation between literary scholars, editorial societies and academies, and the National Library. *Bokselskap* is even more similar to *Litteraturbanken* in terms of partners, literary scope, degree of textual accuracy, introductory texts and other ancillary materials, type and degree of markup, and publishing formats.

13 There are marked differences as well. ADL has a much smaller collection and is not as technically sophisticated. *Bokselskap* has a more heterogeneous collection than *Litteraturbanken* – both in scope and technology – and could be said to be a bit more open in the sense that it is based on a blog platform, and makes available all its encoding principles (including sample TEI files), a publishing plan for the upcoming years, a facility to download entire works in different formats, and presentations of the project in several languages.

14 *Projekt Runeberg* however differs in some important respects: a collection of digitized, machine-readable Nordic classical works launched already in 1992, this is a bottom-up community project based on voluntary work, where *Litteraturbanken* and the other projects mentioned above are top-down, expert-based, and professional. There is in fact a considerable number of voluntary crowd-based projects online, devoted to scanning printed editions of classical works and transcribing them using OCR. Given the ease with which this can be done, we can expect such work to continue or even increase in the future. Largely, however, these are plain image and text collections with minimised metadata, sporadic proof-reading (at best), little or no descriptive markup, and with little or no ancillary scholarly material. Between crowd-based projects on the one hand and generalized mass-digitization projects à la Google Books on the other, activities such as *Litteraturbanken* have found a niche as a collection with high quality facsimiles, carefully proof-read OCR transcriptions, rich metadata, advanced search and a scholarly subject expertise in e.g. literary studies, book history and textual criticism.

Content

15 Not as exhaustive in scope as the National Library of Norway's *Digital Library* programme,¹⁶ *Litteraturbanken* does not aim to be the publisher and storage space for *all* Swedish literature. In the first place, *Litteraturbanken* wants to make the canon of

Swedish literature accessible to the general public, but also to include a range of lesser known Swedish literary authors. In general, *Litteraturbanken* distinguishes between five different foci that help determine the selection of its materials: 1) central authors and editions; 2) thematic or genre specific collections of works; 3) individual works of significance; 4) historically significant Swedish translations of works in other languages; and 5) relevant modern non-fiction literature such as reference works.

16 In the ‘central authors and editions’ (1) section, we are presented with a majority of the works of the most renowned Swedish authors, often based on scholarly editions. The list of authors included is comprehensive, certainly up to the early 20th century, from which point onwards a smaller (but steadily increasing) selection of later authorships is represented, for reasons of copyright. *Litteraturbanken* emphasises that the selection of works is not made on the basis of an existing list of canonical authors or works, but that instead its corpus is expanded through a constant process of renegotiation where both *tradition* and *regeneration* play a principal role. From the start it was made clear that *Litteraturbanken* would become the home of at least two large-scale scholarly editing projects: those of the complete works of both August Strindberg and Carl Jonas Love Almqvist. And through collaborations with the *Swedish Society for Belles-Lettres* ¹⁷ and a number of editorial and other scholarly projects¹⁸ these efforts are now extended to include the works of a wide range of canonical authors. Interestingly, this means that *Litteraturbanken* has gradually become the *de facto* digital publishing and archiving platform for the major Swedish scholarly editing projects – ongoing as well as finished ones.

17 This section is different from *Litteraturbanken*’s self-proclaimed focus on ‘individual works of significance’ (3), where it aims to add a number of works to its collection that are not quite so well-known or critically acclaimed – but that nevertheless had a significant impact on, or signify a good representation of, the literature of the time.¹⁹ Rather than expanding the literary canon, this focus area instead aims to expand our understanding of the context in which more canonical works of literature could blossom. And more of such context is provided in *Litteraturbanken*’s last two sections that include important literary translations (4), and reference works (5).

18 While these goals and selection criteria are all elaborately explained and illustrated in *Litteraturbanken*’s ‘about’ page, it is more difficult to see how they are put into practice in the corpus – and navigable through the Graphic User Interface (GUI).

Browsing the website, it seems that at the moment many of these groupings are still underdeveloped and not yet visible enough. A good example here is *Litteraturbanken*'s focus on thematic and genre related collections (2). To find a collection of texts within a specific genre, the user needs to navigate to the *Litteraturskolan*, where texts are grouped together as being either prose or poetry.²⁰ And deep inside *Litteraturbanken*'s Selma Lagerlöf Archive ('Lagerlöfarkivet'), the user can find a presentation of different themes that are present in Lagerlöf's works. But these are all hard to find, and not living up to their full potential. A simple faceted genre classification should be relatively easy to accommodate in the metadata,²¹ which in turn would make the construction of clearer browsing pathways or genre based advanced search options feasible. This problem relates to some of the usability issues that will be discussed below. At this point, it suffices to say that these are the areas that *Litteraturbanken* is focusing on while it is still actively expanding its corpus, even if this is not always as clearly visible as such in its design.

19 More concretely, the reader of this review may be curious to know how large this corpus actually is. During the first few years of its existence online, the pace with which works were added to *Litteraturbanken* was quite modest, but once robust procedures and workflows were set in place from around 2010 onwards, *Litteraturbanken*'s publishing has acquired considerable momentum. Given this, chances are that the collection will have grown considerably by the time you are reading this review. Still, as a point of reference, at the moment of writing²² its collection counted 2143 works written by a total of 1326 different authors. These works are made available as online readable e-texts (217.754 pages), digital facsimiles (363.684 pages), and EPUBs (844 files). Combined, these works counted an impressive 117.425.313 words. And in addition to this corpus of literary texts, *Litteraturbanken* also includes a number of peripheral materials that were mentioned earlier, such as relevant papers and presentations, materials prepared for students, pupils and teachers (including a Swedish lexicon for literary terms), and a sound archive of texts that are read out loud by Swedish celebrities (these are also made available in the form of an iTunes podcast).²³ In other words, *Litteraturbanken* offers its materials in a wide range of formats, and individual items in the corpus are often (but not always) published in more than one of these.

Usability

20 This brings us to the discussion of the way in which the user may interact with the materials that *Litteraturbanken* has to offer. In this regard, it may be interesting to note that the initial project was created on the basis of an actual user need rather than a more general academic interest. In addition to the affiliated members that were mentioned in our introduction, *Litteraturbanken's* board also includes Swedish author Sven Lindqvist, who is credited with the idea behind the project. Missing ready access to Swedish literature on his travels, Lindqvist reportedly provided the *Swedish Academy* with the incentive to hold the aforementioned seminar that would get the project off the ground in 2002. On the one hand, this would suggest that *Litteraturbanken* has a strong incentive to put a lot of effort into usability and design issues. On the other hand, it may also explain why more advanced analytical and interactive functionalities are not yet fully developed at the moment (as will be more elaborately discussed below). In the terms of Krista Stinne Greven Rasmussen's recent typology of reader roles for scholarly editions, the graphical user interface of *Litteraturbanken* will cater to an audience of *readers*, rather than to one of *users* – let alone *co-workers* (Rasmussen 2016, 127).

Browsing the Archive

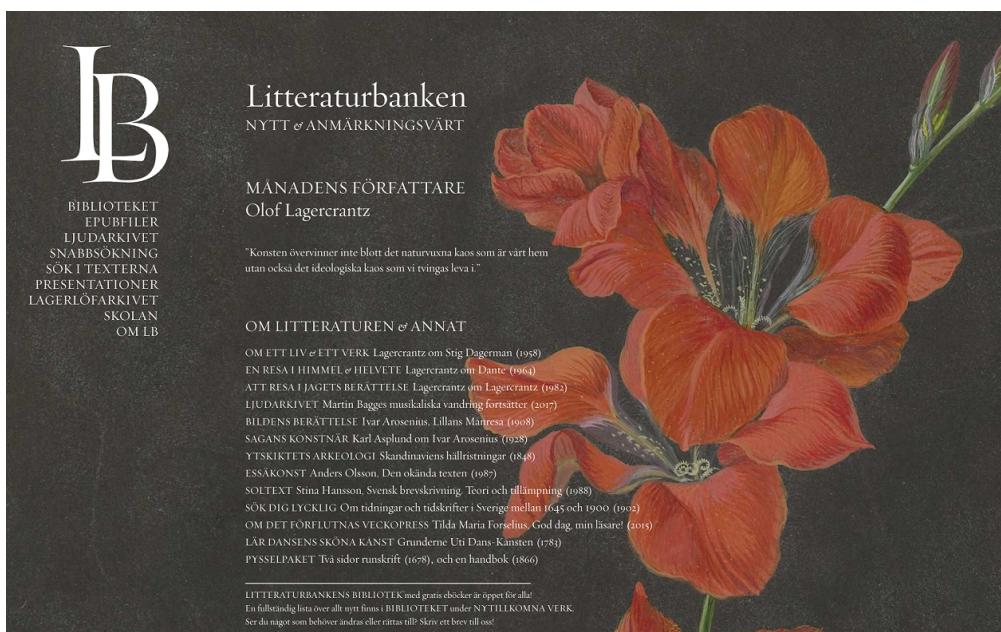


Fig. 1: *Litteraturbanken's* homepage. Its background image changes from time to time.

21 These readers are able to interact with the texts in *Litteraturbanken's* collection through its Graphical User Interface – which has recently been extensively updated. As a

vehicle for human-computer interaction, the first aspect of this interface that we should discuss is probably the website's navigation structure – which is organized in a menu on the left hand side of the website, below its logo (see Fig. 1). This menu can be a bit confusing for the first-time visitor, because it includes different ways to search different parts (or aspects) of *Litteraturbanken*'s archive, and there is some redundancy here as well. Thankfully, however, the best place to start browsing the archive is rather self-explanatorily named *biblioteket* (or: 'the library'); a link that is conveniently located at the top of the menu. If you are looking for a specific author or work, you can find any item in the archive through this page's search bar.²⁴ And if you just want to explore the text collection in general, there are plenty of lists with suggested items (such as *Litteraturbanken*'s most popular authors or works, or the newest additions to the archive). You can also refine your search by (de)selecting specific text formats, be they e-texts, facsimiles, EPUBs, or PDF files.²⁵

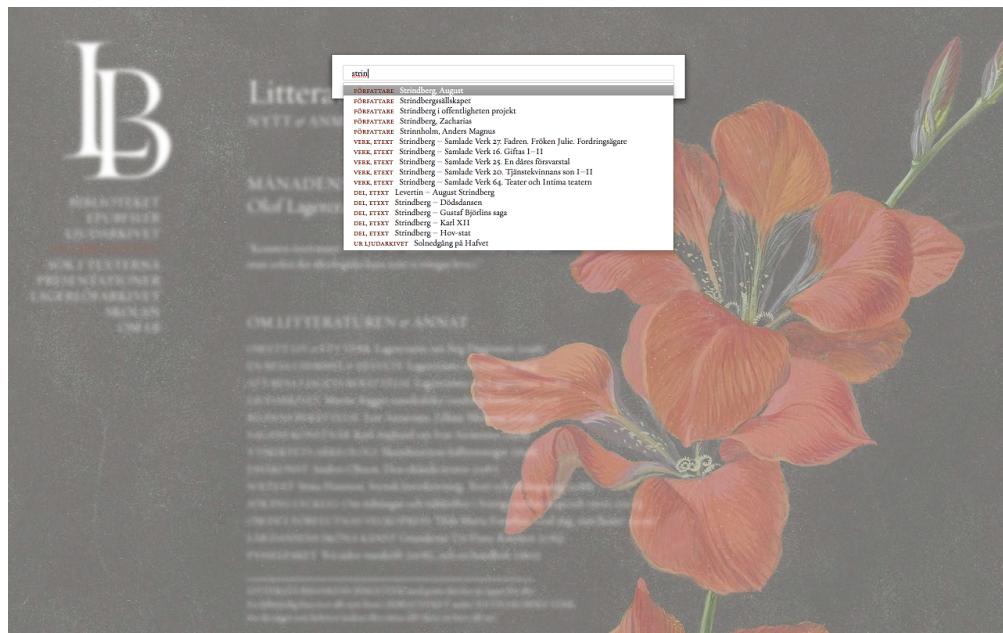


Fig. 2: *Snabbsökning* – *Litteraturbanken*'s quick search feature.

²² Instead of going to the *biblioteket* page, the user can also try the *snabbsökning* link (or: 'quick search') a few menu-items down the list. Except for the list with suggested items and the option to include or exclude certain formats (neither of which is accessible through *snabbsökning*), the functionality of this page is almost identical to that of *biblioteket*: it summons a search bar that allows the user to look for specific authors or titles in the archive. The difference here is that by clicking on *snabbsökning* the user does not navigate away from the page she is currently watching (see Fig. 2). Instead, that page is blurred, and the quick search bar is projected as a pop-up on top of that page –

which the user can quickly return to by clicking outside of the search bar (and in the ‘blurred’ area) like she never left. *Snabbsökning*’s pop-up search bar is a welcome feature on litteraturbanken.se, because it functions as the website’s fixed search bar.²⁶ Still, offering two almost identical search bars on different locations in the website feels somewhat redundant. While there are small differences between the two search bars, these are not advertised.²⁷ This is confusing for the user, who is left to find this difference out on her own, and (failing that) may unwittingly miss out on some interesting search results. Choosing one of these search options and abandoning the other would perhaps be the more user-friendly option.²⁸

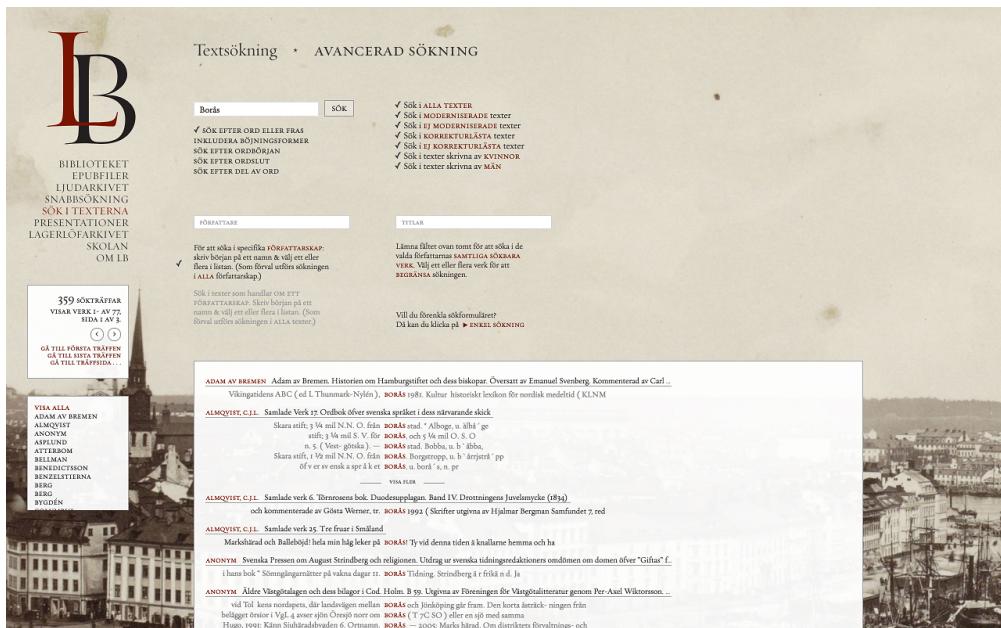


Fig. 3: *Sök i externa* – an advanced search engine for browsing *Litteraturbanken*’s content.

23 To search *inside* the texts, the user needs to navigate to the menu-item *sök i externa* (‘search in the texts’). Here, she can look up specific words or phrases in the corpus, accompanied by some useful advanced search options to narrow down the results (see Fig. 3).²⁹ This is definitely a useful feature that helps the user grasp and navigate *Litteraturbanken*’s content. It should be noted however, that this is again in the first place a search engine that is designed to help the user find specific texts – not to analyse the corpus textually. As of yet, *Litteraturbanken* does not offer an API that would allow researchers to perform text analysis on the works in the text collection. Instead, *Litteraturbanken* provides a corpus of all of its texts to *Språkbanken*, which allows some basic text analysis of the corpus through its *Korp* service;³⁰ and is currently collaborating with Chalmers University of Technology to develop an experimental ‘lab’ website that will allow the user to explore its corpus in various ways. *Litteraturbanken*’s corpus as a

whole, however, will never be downloadable by individuals, because of copyright issues.³¹

Rights and Reuse

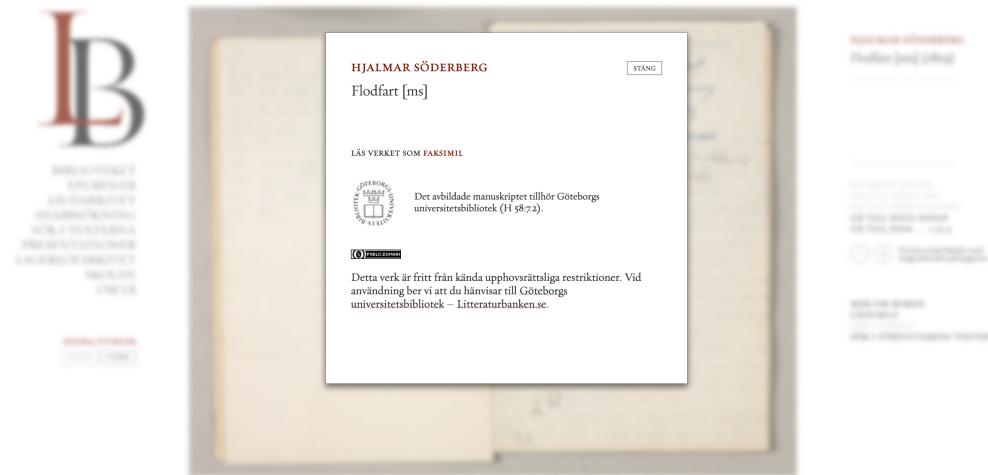


Fig. 4: *Litteraturbanken*'s copyright pop-up.

24 Despite the size and diversity of its corpus and the inevitably limiting effect this has on copyright restrictions, *Litteraturbanken* makes a commendable effort to make its materials freely available in a number of different formats. Alongside the aforementioned collaborations with *Språkbanken* and Chalmers, all the works in *Litteraturbanken*'s corpus can be freely consulted online in the form of so-called e-texts, that are accompanied (when availability and copyright issues allow it) by facsimile images, PDFs and EPUB files – all of which (unless explicitly stated otherwise) are downloadable and by default reusable under a CC BY-NC-SA license.³² However, since the non-commercial restriction has recently come into some disrepute in academic circles,³³ it is encouraging to see that *Litteraturbanken* actually employs a much more flexible approach to copyright licensing. Many of the collection's facsimile images, for instance, are released under a CC0 license: effectively dedicating their rights to the public domain, proper authorship attribution is encouraged but not required.³⁴ These licenses are presented to the reader in a handy pop-up message that appears whenever the user opens a document (see Fig. 4), and can be consulted again afterwards by clicking on the everpresent *mer om boken* ('more information about this book') link.

Reading the (e-)Texts or Facsimiles

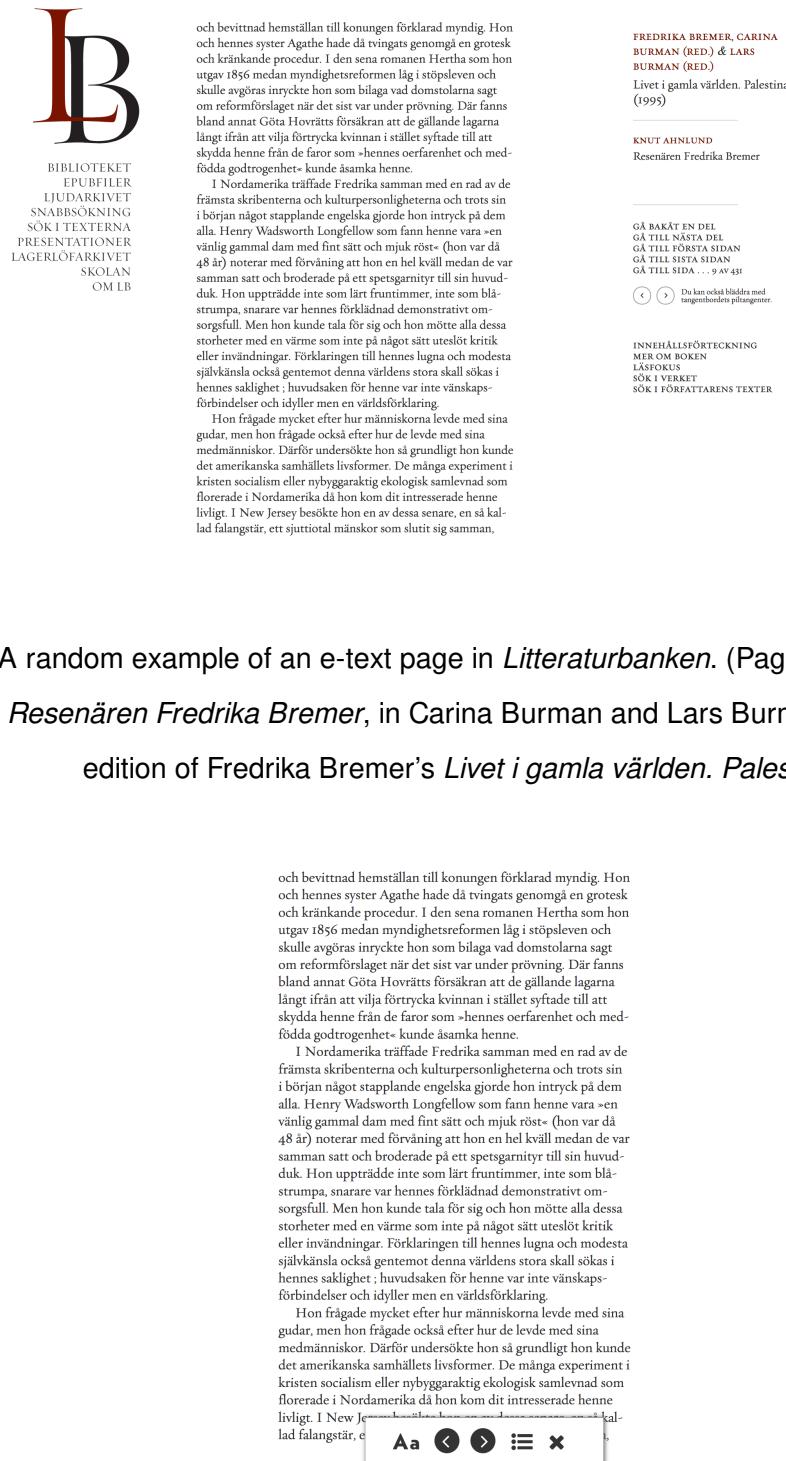


Fig. 6: The same page from Fig. 2, displayed in the *läsfokus* view. See: <http://litteraturbanken.se/forfattare/BremerF/titlar/LivetIGamla/sida/9/etext?fokus>.

25 The core of *Litteraturbanken*'s text collection, as mentioned before, is its large archive of e-texts. These are almost documentary editions of the Swedish texts, where much attention went into letting the e-text resemble the primary source as much as

possible (see ‘Methods’). As goes for *Litteraturbanken*’s entire website, the e-texts look aesthetically pleasing, with a minimal design, nice fonts, etc. (see Fig. 5). In the sidebar on the righthand side, the user finds some basic navigational and search functionalities.³⁶ There is also a link to the *läsfokus* view (‘focussed reading’) that removes the clutter (i.e. the two sidebars) and gives the user some limited options for manipulating the way the text is visualized: using a larger or smaller font, switching between night-mode or day-mode, and minimal navigation controls (see Fig. 6).³⁷

26 Almost as an easter-egg feature for the e-texts, the user can also look up the meaning of individual words in the text in SAOB, the *Svenska Akademins Ordbok* (i.e. the Swedish equivalent of the *Oxford English Dictionary*) by double-clicking on that word.³⁸

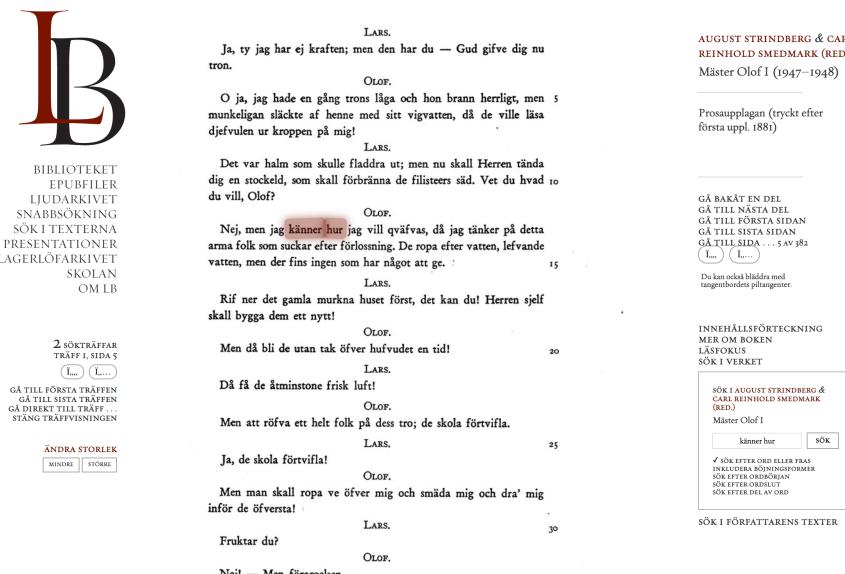


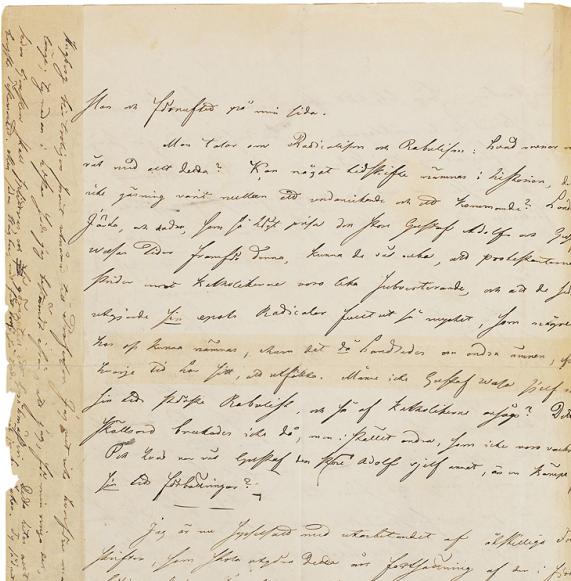
Fig. 7: *Litteraturbanken*’s facsimile view. Here displaying page 5 of part 1 of Strindberg’s *Mäster Olof* (ed. by Reinhold Smedmark, 1947–1948) with the highlighted results of the query ‘känner hur’.³⁹

27 Alongside these e-texts and (when available) their EPUB equivalents, *Litteraturbanken* also makes many of the original source documents available on its website in the form of facsimile images (manuscripts, printed books, etc. – see examples in Fig. 7 and 8). These high-quality images with usually progressively free copyright licenses (such as CC0) are a welcome addition to *Litteraturbanken*. The user’s options for viewing these facsimiles are the same as for those of the e-texts – including even (in the case of OCR’d printed text) the *sök i verket* (‘search the work’) option that finds and highlights queried phrases in the facsimile’s text layer (see Fig. 7).



BIBLIOTEKET
EPUBFILER
LJUDARKIVET
SNABBSÖKNING
SÖK I TEXTERNA
PRESENTATIONER
LAGERLOFARKIVET
SKOLAN
OM LB

ÄNDRÅ STORLEK
MINIMERA
STÖRRE



CARL JONAS LOVE ALMQVIST
Brev till Atterbom den 10
mars 1840

GÅ BAKÅT EN DEL
GÅ TILL FÖRSTA SIDAN
GÅ TILL NÄSTA SIDAN
GÅ TILL SIDA... 4 av 4

Du kan också bläddra med
tangenterna på tangentbordet.

MER OM BOGEN
LÄSFokus
...

SÖK I FÖRFATTARENS TEXTER

Fig. 8: *Litteraturbanken's* facsimile view. Here displaying page 4 of Carl Jonas Love Almqvist's letter to author P. D. A. Atterbom, dated March 10 1840. A feature such as the left-hand margin text points to the need of an image rotation function.⁴⁰

28 What becomes perhaps even clearer in *Litteraturbanken's* facsimile view, is that its developers seem to make an effort to try and keep the user on their website – instead of harvesting its data and manipulating it on their own computer, for instance. While the facsimile images can be freely downloaded (and usually reused) through the regular browser functionalities⁴¹ this has to be done page by page, and is not facilitated for the user by means of a download link. Similarly, for the e-texts, it is impossible to print the whole text of a work for off-screen reading, for instance. While the latter may be a limitation imposed by the copyright holders, the former should be an easy feature to add to the website (especially in the case of the CC0 images) that could drastically improve the user's experience. Other ways in which the website's user-friendliness could be improved would be by allowing the e-text view and the facsimile view to interact with each other instead of being two completely separate views (see also 'Methods');⁴² and (especially for the facsimile view) by adding some kind of scroller or thumbnail mode to allow the user to quickly find a specific page in the work. Finally, some way for the user to manipulate the images in the browser (e.g. rotating images, which would be especially useful in cases like Figure 8) would also be a great addition to the facsimile view.

29 To conclude we can say that – given the inevitable copyright issues involved when dealing with relatively young materials – *Litteraturbanken* provides an impressive (and still growing) collection of Swedish texts that are presented in a slick interface that

provides a pleasant reading experience (especially in the case of the e-texts). But at the same time we find the interface lacking when it comes to the integration of and communication between *Litteraturbanken*'s many different aspects. In each view, simple links to related documents and views (e-texts, facsimiles, EPUBs,⁴³ PDFs, *Litteraturskolan*, or *Litteraturbanken*'s scholarly essays on particular works or authors etc.) would already make a great difference. This is another example of where *Litteraturbanken*'s quantitative aspirations clash with its qualitative aspirations: in the choice to let each of these aspects develop on their own as individual projects for a quick quantitative growth of qualitative materials, some of the *Litteraturbanken*'s potential as a whole (e.g. as linked data) remains untapped. Less urgently, perhaps, it is also a pity that *Litteraturbanken* does not make more use of web 2.0 features:⁴⁴ the user cannot edit, contribute, comment, or annotate any of the texts in the collection, nor can she download TEI files (see 'Methods') for editing and re-use. With the addition of such features and the development of an API, for instance, *Litteraturbanken* could truly fulfil its aim to be a 'bank': allowing for its data to be used and reused by external projects through aggregation, e.g. by building new and better visualization tools for the data. It is not hard to imagine such services emerging as coral reefs in symbiosis with litteraturbanken.se. To reach its full potential, we believe it should be possible to easily export data from and import data into *Litteraturbanken*.⁴⁵

Methods

30 In the preceding section, we had a look at some usability features of litteraturbanken.se and concluded that it offers a variety of formats and view modes in an aesthetically pleasing interface. But how are these views and formats actually produced? In this section, we will take a peek behind the scenes and look at workflows and methods. The information in this chapter on Methods draws partly on the research that Mats Dahlström and Johan Eklund conducted in 2011 as part of their evaluation of the project in Swedish, but has been supplemented during subsequent updates and communication with staff of *Litteraturbanken*. In particular, private email communication with the staff during this review study supplied additional information while also confirming that the workflow is still in place.⁴⁶

Workflow

31 The workflow of *Litteraturbanken* is pretty straightforward. Once an adequate physical copy for a selected work has been chosen as a source (see ‘Contents’), it is subjected to image capture through scanning or photographing. The image files are then run through a transcription process, either by using OCR or by transcribing the text manually. The transcription text is subjected to careful proofreading and then imported into a database and encoded in TEI-XML, from which presentation files in HTML (i.e. the e-text) or EPUB are generated using XSL pre-transformation. In addition, some of the image files are published as digital facsimiles, and *Litteraturbanken* also publishes some material as PDF files (with searchable OCR’ed text; so-called image front transcriptions). Finally, litteraturbanken.se is also being used as a publishing and archiving platform for smaller collections digitized and encoded elsewhere, in particular from ongoing Swedish scholarly editing projects.

32 To many of the authorships, works and to some extent genres, *Litteraturbanken* also writes or commissions expert introductions, comments and didactic resources. As for selecting the primary source (or copy-text) to digitize and edit, the main principle is to choose a copy from the first printed edition, and not to modernize the spelling. Further, the aim is to have the transcription replicate the graphic form of the source as much as possible, thus preserving line breaks, margins, initials, blank pages etc. However, some source document’s paratextual elements like blurbs are discarded.

33 Essentially, this workflow is similar to the way in which the average digital scholarly edition operates – and thus probably quite familiar to the RIDE reader. It is also akin to the production principles of most digital collections housed by memory institutions and universities, although *Litteraturbanken* is considerably more text-oriented (where the standard digital collection of classical works produced by a library is more image-oriented), both for the production and the publication of its materials.

Image Capture

34 As for image capture, *Litteraturbanken* mainly commissions digital facsimiles from the major Swedish research libraries, in particular the University Library in Gothenburg and the National Library in Stockholm. Efforts are made to set uniform quality level benchmarks with the libraries. The image quality is very high, and the commissioned materials are delivered in several versions varying in resolution and size.

These versions are published on the website as facsimiles, and enable users to zoom in on the object (not dynamically, but by selecting one of several predefined versions). For the facsimiles, *Litteraturbanken* provides information about the specific physical copy that was used as source for the digital reproduction, along with full bibliographical information and library shelfmark.

35 The *Litteraturbanken* staff also performs in-house image capture. These internally produced image files (usually scans) are of lower quality than the commissioned images and are primarily used for pre-publication work, but rarely for direct publication together with the encoded transcriptions. Because of their lower quality they do not meet the quality requirements set by *Litteraturbanken* to be published as digital facsimiles. To an outsider, however, it would not seem a far stretch to either put these scans online anyway with a quality caveat, or to raise their quality level only slightly in order to be able to publish them as facsimiles – this would be of great value to the users, not in the least if *Litteraturbanken* would consider integrating a synoptic view where facsimiles and transcriptions are presented side by side,⁴⁷ a feature so far missing in litteraturbanken.se.

Transcription, encoding and publishing

36 For a majority of the works, and in particular printed sources, text is captured using OCR, a process that is often outsourced to external partners. The transcription texts are then meticulously proof-read by the *Litteraturbanken* staff, most of them being literary, linguistic and textual scholars themselves with an expertise in the kind of literary works and genres favoured by *Litteraturbanken*. Given the editorial choice to let the final e-text reflect the source text both graphically and textually, special attention is paid in the transcription phase to cases where e.g. authorially intentional line breaks have been dismissed by the OCR software, or where the OCR dictionaries have modernized old spellings.⁴⁸

37 Text encoding is performed in TEI.⁴⁹ The markup is relatively fine-grained, certainly when compared to similar digital collections elsewhere (but about the same level as that of *boksekskap*). Each word occurrence is assigned a unique ID, and the metadata is extensive, with rich information about the encoding work, the source copy and its provenance, including a reference to the equivalent record in LIBRIS, the joint catalogue of the Swedish academic and research libraries.⁵⁰ This part of the work is excellent, and lays the groundwork for various kinds of data transformation, export and

aggregation possibilities. The encoding is a combination of machine-based (using genre or author specific templates) and manual encoding. The templates (called text ‘factories’) automatically mark up a selected section of text with genre specific pre-settings. This markup is then checked manually and edited extensively. According to staff this procedure works well, and requires only a few hours of manual work for the encoding of a novel of average length.⁵¹ Since the settings, principles and workflow routines are well documented and described in internal manuals, *Litteraturbanken* is in this sense well prepared for staff changes or for data export. As of yet, the TEI-XML files are only used for pre-publication work, i.e. as the basis for the HTML and EPUB formats that are published on the website. The XML files themselves are however not offered either as downloadable files or to be viewed in markup mode on the screen. Likewise, very little information about the editing and encoding principles is made available to the general public on litteraturbanken.se (and hard to find when it is). Since *Litteraturbanken* is gradually becoming the *de facto* publication forum for major scholarly editing projects, and since the documentation of such principles is one of the cornerstones of scholarly editing, this lack of transparent documentation must be considered a shortcoming.

Conclusion

38 To conclude, we can say that *Litteraturbanken* is an ambitious endeavour, whose principal goal to make the most important works of Swedish literature widely and freely available (accompanied by commissioned presentations and comments on authors and works by subject experts) can only be applauded. Having gained considerable momentum in its publishing pace, it is well on its way to reach this goal: an absolute majority of the canonical Swedish authorships and works in the public domain are now adequately represented in *Litteraturbanken* in the form of quality sealed versions, subjected to scholarly scrutiny. And each month that goes by, yet another important authorship seems to have been added to the corpus and presented to *Litteraturbanken’s* readers.

39 It is a well-networked enterprise that collaborates with other literary, linguistic, and editorial projects across Sweden; and especially its extensive cooperation with *Språkbanken* has paved the way for a high degree of competence with regard to computational linguistics, database and corpora work, and a high standard for deep encoding.⁵² The transcriptions are meticulously proof-read, and rendered with a high level of concern for textual and graphical authenticity with regard to the source

documents. To accomplish this (and to maintain the high level of quality for the published facsimiles as well), *Litteraturbanken* has an extensive array of in-house documentation detailing their well established routines for the project's workflow, encoding standards, publication practices, (sub-)versioning and backups with regular log data collections etc. In short, *Litteraturbanken* is a well-oiled machine that provides a large and qualitative collection of Swedish texts.

40 As we see it, however, the project's main problem is that it tries to do too many things at once to be able to do all of them right. Of course, since it is still under development it would be unfair to treat it as a finished project, and we are confident that many of our concerns will be addressed in the near future. Still, as it stands at the moment of writing, by attempting to satisfy its qualitative and quantitative needs simultaneously, *Litteraturbanken* does not fully meet one of the field's most fundamental requirements: providing ways to ensure the editors' accountability for the edited texts (and images) that it has to offer.⁵³ This is an issue that could be remedied by, for instance, making a detailed general account of editorial methods, principles and digitization practices more readily available; by providing more tools for textual scholars to compare the edited texts to their original source materials; and by allowing researchers to read or download individual TEI-XML files. Due to the high quality of the rigorous editorial work and the well-considered construction of the infrastructure and workflows, most of this documentation is already present inside of *Litteraturbanken*, even if it is not yet made available to the users. If only the team would take this one step further, they could transform what is already an impressive, vast, open access collection of high quality digital texts into the genuine digital scholarly editing platform it seems to aspire to become.

Notes

1. The authors' initial research leading up to this publication was conducted as part of the DiXiT Network, a Marie Curie ITN which has received funding from the People Programme (Marie Skłodowska-Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA grant agreement n°317436.

2. Thanks are due to Mats Malm, Cai Alfredsson and Dick Claésson from *Litteraturbanken* and to Hilde Bøe from the Norwegian *eMunch* project for taking the time to answer various questions that came up during the study.

3. *Litteraturbanken*'s board exists of: Gunnel Engwall (President; Stockholm University and *The Royal Swedish Academy of Letters, History and Antiquities*), Lars Borin (University of Gothenburg), Sara Danius (*Swedish Academy*), Pia Forssell (the *Society of Swedish Literature in Finland*), Ingrid Svensson (the National Library), Johan Svedjedal (Uppsala University and SVS), and Sven Lindqvist (a Swedish author). *Litteraturbanken* is directed by Mats Malm of the University of Gothenburg, and its editorial staff includes Cai Alfredson, Dick Claesson, Paulina Helgeson, Anja Hellström, Carl-Johan Lind, Ellen Mattson, Ljubica Miočević, Therese Röök, and Ilaria Tedde.

4. As such, *Litteraturbanken* differs from initiatives similar in scope but with different publication models, such as private-public partnerships and resources funded by subscription or advertisements.

5. I.e. an introduction to the academics and institutions that are involved; its history and financial resources; a clear account of how the texts in the corpus are to be selected; detailed descriptions to help the user manoeuvre *Litteraturbanken*'s many different types of content and functionalities; etc.

6. Not to mention the irony that comes into play when English speakers complain when not everything is written or translated into their own language – thus perpetuating the use of English as a lingua franca.

7. Based on a personal IM communication on May 5 2017 between Mats Dahlström and Hilde Bøe, head of the *eMunch* project.

8. <http://litteraturbanken.se/#!/om/ide>.

9. In addition, the website also hosts a number of secondary materials on both *Litteraturbanken* itself and on the works in the corpus – such as in-depth reviews and presentations on authorships and related subject areas.

10. A whole section of litteraturbanken.se is devoted to its 'school' – a platform aimed at students and teachers at both secondary school and university levels (see: <http://litteraturbanken.se/skola>).

11. Rather than summarized in a list as they are presented here, these diverse points of interest are mentioned individually in various places across *Litteraturbanken*'s introductory pages, and (to some extent) become apparent browsing the project's catalogue and content types.

12. On the website, however, this term usually relates to individual objects in the collection.

13. The name *Litteraturbanken* also contains a reference to the computational linguistics resource *Språkbanken*, hosted by the University of Gothenburg. *Språkbanken* served as both model and partner for the *Swedish Literature Bank* already from the start (Malm, 2010, 4). As such, the name implies that *Litteraturbanken* aims to do for Swedish literature what *Språkbanken* does for Swedish linguistics.

14. ADL: <https://web.archive.org/web/20170727081400/http://adl.dk/>; Deutsches Textarchiv: <https://web.archive.org/web/20170727081501/http://www.deutsches-textarchiv.de/>; Bokselkap: <https://web.archive.org/web/20170727081639/http://www.bokselkap.no/>; Projekt Runeberg: <https://web.archive.org/web/20170727081717/http://runeberg.org/>. Information about *Klassikerbiblioteket* can be found at: <https://web.archive.org/web/20170727081747/https://oa.doria.fi/handle/10024/91449>.

15. Indeed, the fact that these are produced and maintained at a national rather than regional or institutional level, is something Espen Ore (2015, 62) refers to as a possible ‘Nordic twist’ of digital text collections.

16. See: <https://web.archive.org/web/20170727081842/http://www.nb.no/English/The-Digital-Library/What-is-being-digitized>.

17. This society, presented at <https://web.archive.org/web/20170727081932/https://svenskavitterhetssamfundet.wordpress.com/english/>, is the major producer and publisher of printed scholarly editions of classical, post-Reformation Swedish authors.

18. E.g. the ongoing scholarly edition of Selma Lagerlöf’s collected works. Note that the ADL and *Bokselkap* collections (see ‘*Litteraturbanken* in Context’) work with similar arrangements.

19. ‘För varje August Strindberg går det, enkelt uttryckt, ett stort antal John Personne,’ *Litteraturbanken*’s ‘about’ section reads. John Personne was a Swedish author who criticized Strindberg’s work, but whose own works did not yield any critical acclaim. In other words, this quote suggests that for every important author there are a great number of second-rate authors. Arguably, such authors (and their works and criticisms) may

provide crucial insights for contextualizing the more canonical authors in *Litteraturbanken*'s corpus.

20. On the 'about' page, *Litteraturbanken* boasts that there will soon (2017–2018) also be a section for drama texts available.

21. Behind the scenes, there is already the beginning of such a classification in *Litteraturbanken*'s own database files. This was confirmed in an e-mail communication with staff, May 16 2017.

22. These figures were available at litteraturbanken.se on May 9 2017. For current data about the corpus's current size (in Swedish), please visit: <http://litteraturbanken.se/#!/om/statistik>. This page also includes a list of the 30 works that are the most often read (counting the number of times the work's e-text version was consulted online), and the 30 works of which the EPUB files were downloaded the most often. We suspect that the number of words presented on the website combines those of the e-texts with those in the EPUB files – not counting the (untranscribed) words written on the digital facsimile images, nor the OCR'd words in the text layer of the PDFs.

23. See: <https://web.archive.org/web/20170727082110/https://itunes.apple.com/se/podcast/litteraturbankens-upplasningar/id1180639320?l=sv>.

24. This includes all the audio files in the sound archive ('ljudarkivet'), and any work used in *Litteraturbanken*'s school ('litteraturskolan'). It does not, however, include the introductions etc. listed in the website's *presentationer* section.

25. It is not possible to refine the search results by (de)selecting the audio files mentioned in the previous note. Instead, these are always included in the search results.

26. Such a fixed search bar would indeed be difficult to integrate into the website, given that its layout changes considerably depending on the type of content the user is reading (see below).

27. These differences include 1) that *snabbsökning* has a dropdown autocomplete feature where *biblioteket* does not; and, more importantly, 2) that the search results of *biblioteket* seem to be more thorough than those of *snabbsökning*. It appears that the search engine behind the *biblioteket* search bar also takes section titles inside text-items into account, where the *snabbsökning* search bar does not. In our opinion, *Litteraturbanken*'s ideal search bar would combine features of both: the dropdown

autocomplete function from *snabbsökning*, and the more thorough search (and refinement) settings from *biblioteket*.

28. For example by treating the *biblioteket* page purely as an index page; and/or by merging all the different search bars (*biblioteket*, *snabbsökning*, and *sök i texterna*) in a single search function.

29. E.g. the option to include or exclude modernised texts; to only search texts written by either female or male authors; to narrow down the search to one or more specific authors; or even to search within one or more specific works.

30. See: <https://spraakbanken.gu.se/korp/>, and click on ‘Litteraturbanken’ (top left) to select the corpus. In a private email communication (May 16 2017) Dick Claésson confirmed that this corpus is updated regularly.

31. Because *Litteraturbanken* includes a lot of different types of materials, many of which are still protected by copyright, a considerable portion of its materials is not allowed to leave the website (with the exception of the collaborations with *Korp* etc. for corpus analysis). For a full explanation of the different rights that pertain to different materials (in Swedish), see: <http://litteraturbanken.se/om/rattigheter>. In a private email communication (May 16 2017), however, Mats Malm confirmed that *Litteraturbanken* is working on solutions for making as much of its individual materials (that are not restricted by copyright) available for research use in the form of .txt files.

32. This default setting is explained on *Litteraturbanken’s rättigheter* (‘rights’) page: <http://litteraturbanken.se/om/rattigheter>.

33. The ‘non-commercial’ appendix to Creative Commons licenses has been heavily criticized for being too restrictive, considerably limiting the source’s potential for reusability (see for instance Robinson 2013). While this is true, we appreciate that such a CC BY-NC-SA license can already be difficult to obtain for works that are not yet in the public domain (see also Dillen and Neyt 2016). That is why we think *Litteraturbanken’s* flexible approach to licensing makes for a good compromise between accessibility and reusability.

34. The Swedish text clarifying the copyright license at the bottom of the copyright pop-up in [Figure 4](#) can be translated as: ‘This work is free from known copyright restrictions. In case of use, we ask you to refer to Gothenburg University – Litteraturbanken.se’. It is

unclear if the em-dash here means ‘and’ or ‘signed’ – but since authorship attribution is requested rather than required this makes no legal difference, and the (re)user may be encouraged to refer to both.

35. See: <http://litteraturbanken.se/forfattare/BremerF/titlar/LivetIGamla/sida/9/etext>. As this link already shows, another useful feature of both the e-text viewer and the facsimile viewer is that each single digitized page has its own unique breadcrumb URL – which offers a great help for navigation and citation.

36. The navigational options are *gå bakåt en del* (previous section); *gå till nästa del* (next section); *gå till första sidan* (first page); *gå till sista sidan* (last page); *gå till sida...* (move to page by number); or *Innehållsförteckning* (a linked table of contents). The user can also move to immediately preceding or following pages using the arrows in the sidebar, or (as explained alongside those arrows) the arrows on her keyboard. The search options are *sök i verket* (search the work), or *sök i författarens texter* (search the author’s works). The former option opens up a mini search bar on the same page; the latter option transports the user to the advanced search page.

37. The word *limited* really is key here: as may become even more clear when we discuss the facsimile view, *Litteraturbanken* goes to considerable length in controlling the visual display of its materials. In the case of the e-texts there are, for example, no options to change the text’s typeface, or to re-flow the text for optimising the screen real estate on the user’s side. Of course, these options may become accessible through the user’s EPUB reader – if copyright restrictions allow for the work to be published in this format in the first place.

38. More precisely, double-clicking any word in the e-text conjures up a looking-glass icon, which in turn conjures up a pop-up window (on click) that displays the entry in SAOB. The only place where the existence of this feature is hinted at is halfway through *Litteraturbanken*’s ‘help’ section – see <http://litteraturbanken.se/om/hjalp>, specifically its section *ordböcker* (‘dictionaries’) section. This seems too obscure a place to describe such a useful feature that the user is unlikely to stumble upon accidentally – or would be hard-pressed to replicate if she did.

39. See: <http://litteraturbanken.se/forfattare/StrindbergA/titlar/MasterOlof1/sida/4/faksimil>. The *sök i verket* option (bottom of the right sidebar) allows the user to find all of the query’s results (bottom of the left sidebar) and highlight the matches (red).

40. See: <http://litteraturbanken.se/forfattare/AlmqvistCJL/titlar/BrevTillAtterbom10Mar1840/sida/4/faksimil>.

41. For example by right-clicking on the image and selecting ‘Save Image’ (or an equivalent option); or by drag-and-dropping the image onto the user’s desktop.

42. Once inside a work, the user now has to leave the page, go to the introductory page for the work or the author, choose the different view (facsimile or e-text) and go down to the relevant page from there. This is of course cumbersome when a user wants to see what a particular text passage looked like in the facsimile of a manuscript or printed edition. Links between equivalent pages in both views, or (perhaps even better) a side-by-side view that integrates both the e-text and the facsimile image would go a long way to resolve this issue (see also ‘Methods’).

43. When an e-text also has an equivalent EPUB, a download link *is* provided – but only in the popup window that appears when the user clicks on the *mer om boken* link (right sidebar in [fig. 5](#), [7](#) and [8](#)). Since not all e-texts have equivalent EPUBS, we think it would make much more sense to provide a more striking Call-To-Action button with a download link in the sidebar whenever available.

44. Given the low degree of user interaction possibilities, the fact that the website is completely JavaScript dependent seems like an unnecessary technical requirement on the user(‘s browser) as well.

45. While we appreciate that the variety of copyright licenses may make it impossible for *Litteraturbanken* to make *all* of its data exportable, it should be possible to at least provide this option for those materials that are not subjected to those restrictions.

46. We are grateful that the *Litteraturbanken* staff was so helpful and generous in supplying this kind of information. Still, as we will come back to in our conclusion, more of this information should be readily made available on the website for the general public, since a better understanding of these practices would help the user understand and gauge the relation between the source documents and their digital surrogates on litteraturbanken.se.

47. Such synoptic views have become fashionable in many digital scholarly editions, e.g. the editions of the van Gogh correspondence (<https://web.archive.org/web/20170727082400/http://vangoghletters.org/vgl/>); the works of Zacharias Topelius (<https://>

[/web.archive.org/web/20170727082430/http://www.topelius.fi/](http://web.archive.org/web/20170727082430/http://www.topelius.fi/)) or the works of Grundtvig (https://web.archive.org/save/_embed/http://www.xn--grundtvigsverker-7lb.dk/).

48. For instance, by using a dictionary an OCR engine may falsely correct the old spelling of the word *försiktig* ('careful') into the modern form *försiktig*.

49. In the strictest sense, the files do not follow TEI by the book. Thus, for instance, there is no root element <tei> declared, and there are instances of home-brewed elements.

50. See <https://web.archive.org/web/20170727082633/http://libris.kb.se/>.

51. E-mail communication with staff members of *Litteraturbanken*, May 16 2017.

52. Another example of the conscientiousness of *Litteraturbanken*'s collaboration is the fact that the file names and URLs for each page on litteraturbanken.se are linked to their respective records in LIBRIS (see 'Methods').

53. As Patrick Sahle recently argued in his contribution to Driscoll and Pierazzo's *Digital Scholarly Editing: Theories and practices*: '[t]he most basic exigency in traditional editing – *State your rules and follow them!* – is as well the central law and starting point of all digital editing (Sahle 2016, 36; emphasis in original).

References

Dahlström, Mats and Johan Eklund. 2011. *Litteraturbanken: utvärderingsrapport*. (Unpublished). 78 p. Borås.

Dillen, Wout and Vincent Neyt. 2016. 'Digital Scholarly Editing Within the Boundaries of Copyright Restrictions'. *Digital Scholarship in the Humanities* 31 (4): 785-796.

Henny, Ulrike & Frederike Neuber. 2017. 'Criteria for Reviewing Digital Text Collections, version 1.0.' Köln: Institut für Dokumentologie und Editorik.

<https://web.archive.org/web/20170727080330/http://www.i-d-e.de/criteria-text-collections-version-1-0/>.

Malm, Mats. 2010. *Fasta och flyktiga texter: Litteraturbankens digitala tillgångar*. Stockholm: Svenska Vitterhetssamfundet.

Ore, Espen. 2015. 'A Nordic Tradition for Digital Scholarly Editions?' *Journal of the Japanese Association for Digital Humanities* 1 (1): 58-67.

Rasmussen, Krista Stinne Greve. 2016. 'Reading or Using a Digital Edition? Reader Roles in Scholarly Editions'. In *Digital Scholarly Editing: Theories and Practices*, edited by Matthew J. Driscoll and Elena Pierazzo, 119–33. Cambridge: Open Book Publishers.

Robinson, Peter. 2013. 'Five desiderata for scholarly editions in digital form'. Paper presented at 'Digital Humanities 2013', University of Lincoln, Nebraska, July 16–19, 2013.

<https://web.archive.org/web/20170727080500/http://dh2013.unl.edu/abstracts/ab-314.html>.

Sahle, Patrick. 2016. 'What Is a Scholarly Digital Edition?' In *Digital Scholarly Editing: Theories and Practices*, edited by Elena Pierazzo and Matthew J. Driscoll, 19–40. Cambridge: Open Book Publishers.

Svedjedal, Johan. 2003. *En svensk Litteraturbank?* Stockholm: The Swedish Academy. Accessed July 27 2017.

<http://litteraturbanken.se/red/om/ide/Litteraturbanken.pdf>.

Factsheet

| Resource reviewed | |
|----------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------|
| Title | Litteraturbanken |
| Editors | Mats Malm, Cai Alfredson, Dick Claésson, Paulina Helgesson, Anja Hellström, Carl-Johan Lind, Ellen Mattson, Ljubica Miočević, Therese Röök, Ilaria Tedde |
| URI | http://litteraturbanken.se |
| Publication Date | 2004ff. |
| Date of last access | 27.07.2017 |

| Reviewer | |
|---------------------|---------------------------|
| Surname | Dahlström |
| First Name | Mats |
| Organization | University of Borås |
| Place | Borås, Sweden |
| Email | mats.dahlstrom (at) hb.se |

| Reviewer | |
|---------------------|--------------------------------|
| Surname | Dillen |
| First Name | Wout |
| Organization | University of Antwerp |
| Place | Antwerp, Belgium |
| Email | wout.dillen (at) uantwerpen.be |

| General Information | | |
|----------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Bibliographic description | Can the text collection be identified in terms similar to traditional bibliographic descriptions (title, responsible editors, institution, date(s) of publication, identifier/address)? (cf. Catalogue 1.1) | yes |
| Contributors | Are the contributors (editors, institutions, associates) of the project documented? (cf. Catalogue 1.3) | yes |

| | | |
|-------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------|
| Contacts | Is contact information given? (cf. Catalogue 1.4) | yes |
| Aims | | |
| Documentation | Is there a description of the aims and contents of the text collection? (cf. Catalogue 2.1) | yes |
| Purpose | What is the purpose of the text collection? (cf. Catalogue 2.2) | Research, Teaching, General purpose |
| Kind of research | What kind of research does the collection allow to conduct primarily? (cf. Catalogue 3.1.8) | Qualitative research |
| Self-classification | How does the text collection classify itself (e.g. in its title or documentation)? (cf. Catalogue 2.3) | Collection, other: resource, bank |
| Field of research | To which field(s) of research does the text collection contribute? (cf. Catalogue 2.2) | Literary studies, Linguistics |
| Content | | |
| Era | What era(s) do the texts belong to? (cf. Catalogue 2.5) | Medieval, Early Modern, Modern, Contemporary |
| Language | What languages are the texts in? (cf. Catalogue 2.5) | English, French, German, Latin, Swedish |
| Types of text | What kind of texts are in the collection? (cf. Catalogue 2.5) | Literary works, Essays, Scientific papers |
| Additional information | What kind of information is published in addition to the texts? (cf. Catalogue 2.5) | Introduction, Commentary, Context material, Bibliography, Facsimile, other: |
| Composition | | |
| Documentation | Are the principles and decisions regarding the design of the text collection, its composition and the selection of texts documented? (cf. Catalogue 3.1.1-3.1.3) | yes |
| Selection | What selection criteria have been chosen for the text collection? (cf. Catalogue 3.1) | Language, Author, Country, Epoch, Genre |
| Size | | |
| Texts/records | How large is the text collection in number of texts/records? (cf. Catalogue 3.1.4) | > 1000 |

| | | |
|-----------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------|
| Tokens | How large is the text collection in number of tokens? (cf. Catalogue 3.1.4) | > 10 Mio. |
| Structure | Does the text collection have identifiable sub-collections or components? (cf. Catalogue 3.1.5) | yes |
| Data acquisition and integration | | |
| Text recording | Does the text collection record or transcribe the textual data for the first time? (cf. Catalogue 3.1.6) | yes |
| Text integration | What kind of material has been taken over from other sources? (cf. Catalogue 3.1.6) | Full texts, Metadata, Annotations, other: |
| Quality assurance | Has the quality of the data (transcriptions, metadata, annotations, etc.) been checked? (cf. Catalogue 3.1.7) | yes |
| Typology | Considering aims and methods of the text collection, how would you classify it further? For definitions please consider the help-texts. (cf. Catalogue 3.1.8) | General purpose collection, Corpus, Canon, Complete works/œuvre |
| Data Modelling | | |
| Text treatment | How are the textual sources represented in the digital collection? (cf. Catalogue 3.2.1) | Orthographic transcription, Diplomatic transcription, Edited text, Translated text |
| Basic format | In which basic format are the texts encoded? (cf. Catalogue 3.2.4) | XML |
| Annotations | | |
| Annotation type | With what information are the texts further enriched? (cf. Catalogue 3.2.2) | Semantic annotations, Linguistic annotations, Editorial annotations, Structural information |
| Annotation integration | How are the annotations linked to the texts themselves? (cf. Catalogue 3.2.2) | Embedded, Stand-off |
| Metadata | | |
| Metadata type | What kind of metadata are included in the text collection? (cf. Catalogue 3.2.3) | Descriptive, Structural, Administrative |
| Metadata level | On which level are the metadata included? (cf. Catalogue 3.2.2) | Individual texts |

| Data schemas and standards | | |
|----------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------|
| Schemas | What kind of data/metadata/annotation schemas are used for the text collection? (cf. Catalogue 3.2.4) | General standardized schema |
| Standards | Which standards for text encoding, metadata and annotation are used in the text collection? (cf. Catalogue 3.2.4) | TEI |
| Provision | | |
| Accessibility of the basic data | Is the textual data accessible in a source format (e.g. XML, TXT)? (cf. Catalogue 4.1) | no |
| Download | Can the entire raw data of the project be downloaded (as a whole)? (cf. Catalogue 4.2) | no |
| Technical interfaces | Are there technical interfaces which allow the reuse of the data of the text collection in other contexts? (cf. Catalogue 4.2) | other: non-copyrighted parts available as corpus through Språkbanken |
| Analytical data | Besides the textual data, does the project provide analytical data (e.g. statistics) to download or harvest? (cf. Catalogue 4.3) | no |
| Reuse | Can you use the data with other tools useful for this kind of content? (cf. Catalogue 4.4) | no |
| User Interface | | |
| Interface provision | Does the text collection have a dedicated user interface designed for the collection at hand in which the texts of the collection are represented and/or in which the data is analyzable? (cf. Catalogue 5.1) | yes |
| User Interface questions | | |
| Usability | From your point of view, is the interface of the text collection clearly arranged and easy to navigate so that the user can quickly identify the purpose, the content and the main access methods of the resource? (cf. Catalogue 5.3) | yes |
| Access modes | | |

| | | |
|------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------|
| Browsing | Does the project offer the possibility to browse the contents by simple browsing options or advanced structured access via indices (e.g. by author, year, genre)? (cf. Catalogue 5.4) | yes |
| Fulltext search | Does the project offer a fulltext search? (cf. Catalogue 5.4) | yes |
| Advanced search | Does the project offer an advanced search? (cf. Catalogue 5.4) | yes |
| Analysis | | |
| Tools | Does the text collection integrate tools for analyses of the data? (cf. Catalogue 5.5) | no |
| Customization | Can the user alter the interface in order to affect the outcomes of representation and analysis of the text collection (besides basic search functionalities), e.g. by applying his or her own queries or by choosing analysis parameters? (cf. Catalogue 5.5) | no |
| Visualization | Does the text collection provide particular visualizations of the data? (cf. Catalogue 5.6) | no visualization |
| Personalization | Is there a personalisation mode that enables the users e.g. to create their own sub-collections of the existing text collection? (cf. Catalogue 5.7) | no |
| Preservation | | |
| Documentation | Does the text collection provide sufficient documentation about the project in general as well as about the aims, contents and methods of the text collection? (cf. Catalogue 6.1) | yes |
| Open Access | Is the text collection Open Access? (cf. Catalogue 6.2) | yes |
| Rights | | |
| Declared | Are the rights to (re)use the content declared? (cf. Catalogue 6.2) | yes |
| License | Under what license are the contents released? (cf. Catalogue 6.2) | CC0, CC-BY-NC, CC-BY-NC-SA, PDM |

| | | |
|-------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Persistent identification and addressing | Are there persistent identifiers and an addressing system for the text collection and/or parts/objects of it and which mechanism is used to that end? (cf. Catalogue 6.3) | URN |
| Citation | Does the text collection supply citation guidelines? (cf. Catalogue 6.3) | no |
| Archiving of the data | Does the documentation include information about the long term sustainability of the basic data (archiving of the data)? (cf. Catalogue 6.4) | no |
| Institutional curation | Does the project provide information about institutional support for the curation and sustainability of the project? (cf. Catalogue 6.4) | yes |
| Completion | Is the text collection completed? (cf. Catalogue 6.4) | no |
| Personnel | | |
| Editors | Mats Malm Cai Alfredson Dick Claésson Paulina Helgeson Anja Hellström Carl-Johan Lind Ellen Mattson Ljubica Miočević Therese Röök Ilaria Tedde | |