

README and notes about the template

Gerald Q. Maguire Jr.

July 2025

This document is a work in progress.

This document describes some of the L^AT_EX thesis templates that I have developed for use at KTH Royal Institute of Technology (KTH) and provides some background about why they have the form that they do. It is important to note that these templates are **not prescriptive**, as not every thesis will have all the parts a given template shows. However, if there is something that you decide to leave out, you should make a conscious decision to do so, and you should consider the impact this may have on your thesis being approved*.

*By your examiner or your grading committee

2|

Contents

1	Introduction	5
2	Motivation	6
3	Deliminations	8
4	Structure of the files for the template	9
4.1	Essential directories	10
4.2	Optional directories	15
4.3	Informational directory	16
4.4	Directory for a compilation thesis	16
5	Expected users and their differences	17
6	Further information for authors(s)	17
7	Those working in parallel with the authors(s)	18
7.1	Supervisor	18
7.2	Creating the custom_configuration.tex file	19
7.3	Opponent(s) and oral presentation/defense	19
8	Administrative steps	20
8.1	Using the collected metadata	21
8.2	What is a TRITA number and why does each thesis need one?	24
8.3	Entering the metadata into DiVA	24
8.4	Make the MODS file	25
8.5	Other uses of the JSON file	25
9	(Human) Readers of the thesis	26
9.1	Machines reading the metadata or full text of the thesis	26
9.2	Template author and maintainers	27
10	While writing	28
10.1	Conventions for todo notes	28
10.2	Removing README_notes	28
10.3	Removing unused fonts	29

References **31**

1 Introduction

This template evolved (radically) from an earlier L^AT_EX thesis template that was widely used at KTH. The start of this evolution was a DOCX template I developed over many years for use with students for whom I was the examiner and/or supervisor. The suggested structure and contents of the thesis reflect my experience as an examiner for more than 600 degree projects and the experience I have had as a teacher and examiner for the course *II2202 Research Methodology and Scientific Writing*. The template also reflects my interest as a member of KTH's Language Committee in facilitating the parallel use of English and Swedish at KTH, as well as supporting other languages. The latter aspect reflects my experience with double-degree students, who often need to have at least the abstract of their thesis in their home university's language(s). The thesis template also reflects my experience in entering the metadata for hundreds of theses into Digitala Vetenskapliga Arkivet (DiVA) and announcing a very large number of degree project seminars (with a large fraction of them having abstracts and keywords in both English and Swedish).

Several major thoughts have influenced the design of my templates:

Thought 1 The template should help every student produce a high-quality thesis.
 The template should also help an undergraduate student be successful in their degree project and should help a graduate student be successful in their presentation or defense.

Thought 2 The template should help support all of the (relevant) phases of the degree project process or the third-cycle thesis process.

Thought 3 Redundant data entry should be minimized to increase consistency.

Thought 4 There are several thousand theses written each year at KTH. Theses are the second most common type of publication at KTH.

Thought 5 Every approved undergraduate thesis and all third cycle theses will have at least its metadata entered into DiVA. DiVA features multi-language support for title, subtitle, abstract, and keywords.

Section 5 describes several different groups of users and how the template is relevant to them.

2 Motivation

This section should be skipped, unless you really care about my motivation. In summary: There are a lot of theses at KTH each year, hence a lot of metadata to be entered. Additionally, there are people who use the metadata. Unfortunately, there are a lot of errors in the data. Therefore, in keeping with Thought 3, the data should be entered once and then mechanically replicated as needed.

Several thousand first-cycle and second-cycle theses are written each year by KTH students. For these, every approved thesis will be entered into DiVA*. Additionally, hundreds of third-cycle theses are written each year by KTH students. Every third-cycle thesis that is to be defended[†] will be entered into DiVA.

After having entered several hundred theses into DiVA myself, I realized the large amount of time needed to enter the metadata and upload each thesis[‡]. Much of this work involved either typing a lot of text or cutting and pasting text from the PDF of the thesis. Even worse was that some of these theses had been produced by print drivers that used special encodings of the glyphs and some drivers used composition to handle the characters 'å', 'ä', and 'ö'. These problem cases required even more time to enter the metadata. Additionally, some abstracts have equations that need to be turned into the appropriate HTML for DiVA. In a first-cycle thesis, two students estimated that it took nearly three full-time equivalents of employees to enter all of the undergraduate theses at KTH [1].

*Independent of whether the full text is made available via DiVA

[†]In this document, the licentiate seminars will also be included under the term *defense*; although more formally it is a seminar.

[‡]A DiVA administrator for ABE estimated that it took ~30 minutes per thesis.

In addition to the time taken to process each thesis, there was another problem: there were errors in the entered data. These errors include both errors of omission and errors when the data entered does not match the data in the thesis. Examples of errors of omission include not entering into DiVA alternate titles (and subtitles), abstracts, and keywords in additional languages – despite the fact that this content was in the thesis. The data entry errors (particularly from cut and paste) lead to thousands of titles and abstracts being incorrect. Worse is that in many cases, the title(s) in DiVA and the title(s) in Lokalt adb–baserat dokumentationssystem (LADOK) did not match. Some of the errors were quite small, such as a small number of missing spaces, but other errors were much larger. These larger errors ranged from abstracts where a large fraction of the words were merged together* to incorrect or missing characters (especially in equations). Part of the problem is that people are going to make mistakes, while another part of the problem is due to the tools that people were using.

As I am very lazy, but willing to work hard to be lazy, I generally try to automate tasks that are likely to be done again and again. Additionally, I have found that one way to avoid problems is to only enter data once – then let tools replicate it as needed. For example, consider the title of a thesis in some language X. How many times is this title going to be used? Typically, this title will appear on the cover, on the title page, in the announcement for the presentation or defense, in the title field in the DiVA metadata, and in a title entry in LADOK. Now consider what happens when this is scaled up by a factor of more than 3 000 per year. What do you expect the error rate to be?

Raymond R. Panko of the Shidler College of Business at the University of Hawai‘i has studied human error rates, especially with regard to data entry in spreadsheets. He states that “base error rates (BERs), which are long-term average error rates quantified as the number of errors per 100 actions. BERs obviously depend on complexity. For simple but nontrivial cognitive actions such as writing, calculating, and writing program statements, BERs are usually in the range of 1% to 5.”[2]. He also states that “Rushing, inexperience, code complexity, stress, and other factors can exacerbate error rates. However, as Figure 2 illustrates, reducing exacerbators does not eliminate errors. There usually is a floor beyond which error reduction is minimal.”

So even at a 2% error rate, there are going to be hundreds of errors just in the titles. Consider that commonly there is an alternative title (in another language),

*Perhaps due to the fact that PDF does not output ‘space’ characters but rather positions glyphs on the page at some coordinates.

one or more abstracts (each of which can be a hundred or more words long), keywords, author names, ... – the total number of errors is large*. I learned that in some cases, the differences in titles were **intentional** – as the person entering the title(s) into LADOK wanted the students to look “less stupid” - as the title(s) are printed on the student’s diploma.

It is important to note that I have made **no** attempt to correlate errors with *who* entered the data. Moreover, much of the evidence (at least with respect to the DiVA entries) has been lost, as I have sent a very large number of corrections to KTH Biblioteket (KTHB).

Another factor that motivated me was that not only was I concerned with putting data into DiVA and LADOK, but *I have used data from both of these systems*. For example, one of the uses of data from LADOK was to see if I could match theses up between LADOK and DiVA. Because of accidental or intentional differences, it is difficult to match the theses up; moreover, there is not field in LADOK to record a references to the DiVA identifier. An example of one of the uses of data from DiVA was to collect vocabulary terms and to collect paired keywords in English and Swedish. Another use of data from DiVA was to understand what languages had been used in titles, subtitles, keywords, and abstracts. In some cases, the language of the content as indicated in DiVA did not match the language of the content!

As a result, collecting the data necessary for DiVA[†] and repeatedly making use of this data was a major driving force in the design of my L^AT_EX templates. This data is also useful for many phases of an undergraduate degree project, and also useful for third-cycle theses defenses, such as announcing the oral presentation or defense.

3 Delimitations

I was told by one of my colleagues in applied mathematics that theses in this area generally do not follow the Introduction, Methods, Results, and Discussion (IMRAD) structure. Instead, the main body typically follows the Definition-Theorem-Proof pattern.

*For one school, I found that the total error rate when considering both DiVA and LADOK was $\approx 10\%$.

[†]Also, to a lesser extent, for LADOK. I say to a lesser extent, as the possibility to automatically insert the titles in LADOK had not been taken into production when I retired – despite my demonstration of how it could be done and a IT developer having introduced it into a test system.

Also, as noted in a footnote, in the `README_3rd_cycle_author.tex`: While the template includes support for textual & image artifacts and references to other types of artifacts, I have not fully considered additional types of media, such as video, audio, physical artifacts, ..., and artistic output. I apologize in advance that the template might not suit a student in a doctoral program, such as: Architecture, “Art, Technology and Design”; and others. A practice-based thesis or thesis by project frequently features creative work (*i.e.*, the *praxis*) with a written document (*i.e.*, the *exegesis*). Constructive suggestions as to how the template can be improved are welcome.

Some parts of the template are conditional based on the value of a switch: `\ifinswedish`. The idea is to easily have a single template that supports theses written in English or Swedish. However, in many places, the conditional has not been used but could be. Examples of this include the Swedish names for chapters and sections. Generally, this information is in a note after the English chapter or section name. More complete implementation of the use of this condition remains as future work.

The template does not actually support the G5 paper format. In particular, the KTH cover (produced with `\kthcover`) and back cover (produced with `\kthbackcover`) have only been adapted for A4 paper.

Support for G5 paper remains as (potential) future work. However, it is unlikely to be done, as this paper size basically results from the printing service trimming A4 pages - as opposed to directly producing G5 pages.

To support Doctor of Philosophy and Licentiate of Philosophy degrees in addition to Tekn. Dr. and Tekn. Lic. in the defense description on the title page, I have added a `\degreeModifier{Philosophy}` macro. This is used when processing `\fullDegreeName`. Hopefully, this correctly formulates the defense description on the title page.

4 Structure of the files for the template

Table 1 shows the structure of the files for the template. These files are generally taken from an existing Overleaf project or a ZIP file.

One hope is that by automatically extracting information from various sources, this information is more likely to be *correct* and *consistent* (supporting Thought 3). This approach has been used to design two of the files used for the template. These files are:

1. The file `custom_configuration.tex` contains macros and values for configuring a project. These values are generally expected to be known at the start of the project, *e.g.*, author(s), supervisor(s), course cycle, program code, degree name, subject area, *etc.* or adjusted later in the project.*
2. The file `lib/schools_and_programs_3rd_cycle.ins` contains the English and Swedish names of schools and programs†. This information provides a consistent mapping from school code and program code to their English or Swedish expansions.

We will assume that these files have been generated by someone. Later, we will examine who this someone might be for each of these files.

4.1 Essential directories

A number of directories contain information that is necessary when processing your document.

- Table 3 shows the structure of files under **kth**. These files are KTH specific files that process the `documentclass` options, set up the fonts, define the layout, deal with the metadata, *etc.* These files are included by the `kththesis.cls` class file to realize the template.
- Table 4 shows the structure of files under **lib**. These files are used by L^AT_EX when processing the document.
- The directories **Figtree** and **Noto_Emoji** contain font information for fonts that are not already available in Overleaf projects. They also provide some examples of how you can add additional fonts, should you need to, and how to finely control these fonts.

*While this file can be manually edited, it was originally designed with the idea that much of its contents could be generated by a program that I have written that extracts data from a Canvas course used in conjunction with degree projects. One of the goals of using such a program is to extract data automatically from data sources, such as Canvas, the KTH profile Application Programming Interface (API), and Kurs- och programplaneringssystemet (KOPPS) – thus avoiding the need for anyone to manual enter much of this data. The macros for defining this information are described in the file `README_3rd_cycle_author.tex`.

†A program extracted this information from KOPPS. However, KOPPS was shut down in June 2025; hence, a new program will be needed, or the file will need to be manually updated. In addition, the file was manually augmented with fake entries to be used as examples in the template.

- The **unicode_blocks** directory contains files that establish a mapping between unicode code points to codepoints in specific fonts. Depending upon what languages you are writing in and what symbols you are using, you can include the necessary files via `lib/defines.tex`. Note that some of these files are rather complete, while others only have some entries. Thus, you may need to add entries to some of them for your specific thesis.

If you are only using characters associated with a given language when writing your abstract, you do not need to include the related unicode block, as babel will utilize the relevant fonts when processing these abstracts.

It is likely that at some point many of the files in the **unicode_blocks** directory will be removed since babel's font mechanism obviates the need for them.

Table 1: Structure of files for the template

Additional_Abstracts	directory containing abstracts in various languages
bibstyle	directory containing files related to the style of the bibliography
Figtree	directory containing Figtree font files
figures	directory containing files for figures in the <code>examplethesis.tex</code>
Included_publications	directory containing included publications
kth	directory containing KTH specific files
lib	directory containing various library files
Noto_Emjoii	directory containing Noto Emoji font files
README_notes	directory containing README files for various purposes and some useful files
unicode_blocks	directory containing files to map unicode code points to codepoints in specific fonts
<code>custom_configuration.tex</code>	macros and values for configuring a project
<code>examplethesis.tex</code>	an example of the thesis itself
<code>Figtree.fontspec</code>	Figtree fontspec specifies the files to be used
<code>hyperxmp.sty</code>	a modified hyperxmp package to store XMP information in the PDF file
<code>kththesis.cls</code>	the kththesis class file
<code>Quick_Start_Guide.tex</code>	a 2 page quick start guide
<code>README_3rd_cycle_author.tex</code>	notes for third-cycle authors
<code>references.bib</code>	references that may be cited in the thesis

The directory and files shown in Table 2 are optional but highly useful for improving Overleaf's performance. `Saving_and_restoring_font_cache.tex` explains how these files and the `saved_font_cache` directory are used.

Table 2: Structure of files for the template

saved_font_cache	a directory to save fonts, so that they can quickly be restored to the working font cache
<code>restore_font_cache.tex</code>	compile this document to prefill cache with saved fonts
<code>Saving_and_restoring_font_cache.tex</code>	Guide to how to save and restore fonts
<code>warmup.tex</code>	compile this document to warmup the font cache (<i>i.e.</i> , load the fonts used in <code>example.tex</code>) and save them to a working directory

14 | Structure of the files for the template

Table 3: Structure of files for the template under kth

dkjson.lua	a Lua module to write JSON files
KTH_logo_RGB_bla.png	the KTH logo for use on the cover
kth-commands.tex	a collection of commands that are defined for the template
kth-fonts.tex	set up the fonts for the template
kth-layout.tex	the layout specific parts of the template
kth-metadata.tex	L <small>A</small> T <small>E</small> X related to the metadata
kth-metadata.lua	Lua code related to the metadata
kth-options.tex	define all of the different options
packages.tex	the package for the template itself
kth-seanity-checks.tex	sanity checking code
kthcolors.tex	defines a number of colors from the KTH palette
schools_and_programs_3rd_cycle	English and Swedish names of schools and the programs

Table 4: Structure of files for the template under lib

acronyms.tex	a place to define the acronyms that will or might be used
defines.tex	some generally useful defines
filter-warnings.tex	a place to put filters for silencing warnings
includes-after-hyperref.tex	a special include file for packages that have to be included after the hyperref package
includes.tex	a centralized place to include packages that might be useful
pdf_related_includes.tex	includes to be able to add the title and other information to the PDF file
placeHolder_lbx_files.tex	a file to provide placeholders (stubs) via a series of <code>filecontents</code> environments for use with biblatex for various languages, when one has not loaded the actual biblatex .lbx files

4.2 Optional directories

Some directories are strictly optional depending upon whether you have one or more abstracts in addition to the English and Swedish abstracts and whether you are using bibtex or biblatex.

- The **Additional_Abstracts** directory contains skeleton abstracts in a number of languages. Feel free to utilize those that you wish and add a new language if you want or need to.
- Table 5 shows the one file, a bibtex style file (`.bst`) file. This directory is *only* necessary if you are using `bibtex`. If so, you may need to replace it with a suitable `.bst` file for your thesis. The file that is there is based on the IEEE Transactions style but augmented to include International Standard Book Numbers (ISBNs), Digital Object Identifiers (DOIs), and Uniform Resource Locators (URLs).

Table 5: Structure of files for the template under bibstyle

myIEEETran.bst	a bibtex style file
----------------	---------------------

4.3 Informational directory

Table 6 shows the structure of files under **README_notes**. The files are purely to provide information to different target audiences and are *unnecessary* when processing your document. Most of these are standalone documents or files to be included when making these documents (such as `characters_used_reduced.tex` and `characters_used_reduced.tex`).

Table 6: Files for the template under README_notes directory. The .tex extension has been elided.

characters_used_reduced	a shorter table of the characters that have been used in theses
characters_used_reduced	a long table of the characters that have previously been used in theses
National_Subject_Categories	a table of the National Subject Categories and their codes
README_for_3rd_cycle_version	note about the template relative to the AW-template
README_for_administrators	a README targeted at administrators
README_notes_reduced	these notes
README_programmer_notes	notes for a programmer or maintainer of the template
README_supervisor_notes	a README targeted at supervisors

4.4 Directory for a compilation thesis

Additionally, for a compilation thesis, there is a folder **Included_publications** that contains .bib files and .pdf files for each of the included publications. There should be a pair of files: one a .bib file and the other a file containing

the publication to be included as .pdf, .png, ..., . See the section “Including publications” in README_3rd_cycle_author.tex.

5 Expected users and their differences

This template is relevant to several different sets of users:

Users 1 Author(s) should see Section 6.

Users 2 Those working together with the author(s) during the degree project or thesis process (see Section 7).

Users 3 Administrative staff working with the document after it has been approved for printing (see Section 8).

Users 4 The (hopefully) many (human) readers of the final document (see Section 9).

Users 5 The (hopefully) many computers reading the metadata and the full text of the final document (see Section 9.1).

Users 6 Those who are maintaining or updating this template (see Section 9.2).

Each of these different sets of users has different needs and perspectives. The following sections describe these needs and perspectives.

6 Further information for authors(s)

Authors will primarily be focused on learning (just) enough to effectively use the template to write their thesis. Authors will primarily want to read:

1. Quick_Start_Guide.tex,
2. README_3rd_cycle_author.tex,
3. this file (*i.e.*, README_notes_reduced.tex), and
4. the notes in the examplethesis.tex files.

7 Those working in parallel with the authors(s)

Those working together with the author(s) during the project include supervisors and may include the opponent(s). Potentially, this might also include the internal reviewer, who could be invited to an Overleaf project to both read and comment on the manuscript using comments and change tracking.

7.1 Supervisor

If a project is done in industry, there is generally an industrial supervisor in addition to the academic supervisors*. Although many students might have two academic supervisors, an industrial supervisor and/or another academic supervisor. The template supports up to five supervisors. The choice of up to five reflects my experience and observation of prior theses in DiVA; along with a recent request by a student to have four supervisors. The supervisors are enumerated as A, B, C, D, and E. For each of these, replace the "X" in the following macros:

`\supervisorXsLastname{}`the last name of the supervisor

`\supervisorXsFirstname{}`the first name of the supervisor

`\supervisorXsEmail{}`e-mail address of the supervisor

If the supervisor is from within KTH, then add their KTHID, School, and Department info:

`\supervisorXsKTHID{}`the supervisor's kthid

`\supervisorXsSchool{}`the school of the supervisor

`\supervisorXsDepartment{}`the department of the supervisor

If the supervisor is from outside of KTH, then add their organization with:

`\supervisorXsOrganization{}`the supervisor's organization

*Note that there is expected to be at least **two** supervisors (as there is a legal requirement for this).

7.2 Creating the custom_configuration.tex file

I assume that someone (such as the main supervisor or the student) will generate the file: `custom_configuration.tex`. This assumption is based upon the fact that these two people know who the student is, who the supervisors are, what program the student is in, Third-cycle students should simply edit the `custom_configuration.tex` file in this Overleaf project.

Ideally, this file should be generated automatically by a computer program so that each student gets a customized template automatically *

7.3 Opponent(s) and oral presentation/defense

The macros related to the opponent and oral presentation are in the file `custom_configuration.tex`.

The macro for the opponent(s) is:

<code>\opponentsNames { }</code>	the names (in normal name order) of the opponent or opponents
----------------------------------	---

When there are multiple opponents, separate their names with '`\&`'; for example, A. B. Normal `\&` A. X. E. Normalè.

This should probably be expanded to include organization information for the opponent(s). This could be used when generating the announcement or calendar event and when entering metadata about the opponent(s) in DiVA.

*Assuming that a Canvas course room exists for the third-cycle students, this file could be generated automatically. For first- and second-cycle students who are in a Canvas course room for a degree project, a command line program (`create_customized_JSON_file.py`) can be used. Subsequently, a separate program (`customize_LaTeX_project.py`) takes this JavaScript Object Notation (JSON) data and creates the appropriate `LATEX` commands and inserts this information into the file and then inserts this file into a ZIP file, either replacing or augmenting the `custom_configuration.tex` within this ZIP file (if one exists). There is an option for this second program –`initialize` that causes the program to simply replace the file rather than appending the new information to the end of the file.

The above programs are available from <https://github.com/gqmaguirejr/E-learning>. The README file for this GitHub contains information about how to run the programs, their options, and gives examples.

For the oral presentation, the following macros are filled in once the defense has been scheduled:

\presentationDateAndTimeISO{} date and time of the presentation is ISO format, for example: 2022-03-15 13:00

\presentationLanguage{} three letter abbreviation for the language of the presentation according to three letter ISO 639-2 Code – specifically the "B" (bibliographic) variant of these codes (note that this is the same language code used in DiVA), generally eng or swe

\presentationRoom{} a room name and/or
“via Zoom [https://kth-se.zoom.us/j/d-
ddddd](https://kth-se.zoom.us/j/dddddddd)”

\presentationAddress{} location of the room, for example:
Isafjordsgatan 22 (Kistagången 16)

\presentationCity{} city where the presentation occurs,
generally: Stockholm

8 Administrative steps

Once the internal reviewer approves a thesis to go forward to a defense, the author needs to perform a number of administrative steps:

1. Request and add the assigned TRITA number to your thesis.

The TRITA number is assigned by the student affairs office from an annual series of numbers. Section 8.2) describes what a TRITA number is and where it appears. The student must request the TRITA number and ISBN *before* the thesis is printed, so that both numbers will appear in the printed thesis. Both will appear on the book information page and the TRITA number will also appear on the back cover of the thesis. See the instructions at <https://kth.diva-portal.org/dream/info.jsf>.

2. Request an ISBN - see Create an ISBN number. Note that this form now requires the TRITA series and number, as well (*i*), as whether it is for a Doctoral/licentiate thesis or Report, (*ii*) Title of publication, and (*iii*) first authors: (a) Last name, (b) first name, (c) KTH id, and (d) e-mail address.

As all the required information is in the `fordiva.json` file, the program `request-ISBN-with-JSON.py` at <https://github.com/gqm/aguirejr/E-learning> can be used to automate the process of making a request for an ISBN. The program generates a `assigned-ISBN.tex` file containing the assigned ISBN. This file can be uploaded to the project and will automatically be used.

3. The author will also have to arrange for the printing of the thesis by US-AB. See the instructions at Printing Thesis.
4. The author has to enter the metadatadata and thesis into DiVA. See Section 8.3.
5. Additionally, if the thesis is a compilation thesis, the author has to make sure that all of the publications (including manuscripts) that are in the “List of Publications” have a record in DiVA. The DiVA entry for the thesis will have cross references to these publications. You might find the `citedtags.bib` file attached to the PDF file to be helpful in this.

The Jupyter notebook `citations-to-DiVA-Notebook-20250811-all.ipynb` takes the `citedtags.bib` and `fordiva.json` files and looks up DiVA IDs for these publications. The author can use this information to (*i*) know which already have DiVA entries and which do not, and (*ii*) have the DiVA IDs to use when entering the metadata for the thesis. Hopefully, this should facilitate entering the thesis into DiVA. In the future, a program utilizing the Cora-based DiVA API could automatically do all of this and then enter the thesis metadata and upload the PDF.

Fortunately, the template collects metadata that you will need, hence making the administrative steps easier for you (and others).

8.1 Using the collected metadata

Figure 1 shows the overall flow of data used in the process of getting a TRITA number assigned, adding this TRITA number to your L^AT_EX project, creating a Metadata Object Description Schema (MODS) file for import into DiVA, and uploading your thesis into DiVA.

22 | Administrative steps

Note that a DiVA administrator only needs to receive a `fordiva.json` file from a student and then can use this information to assign a TRITA number. The assigned TRITA number is communicated to the student.

Once the student has the TRITA number and ISBN, they should add this to their thesis as described in Section 8.2.

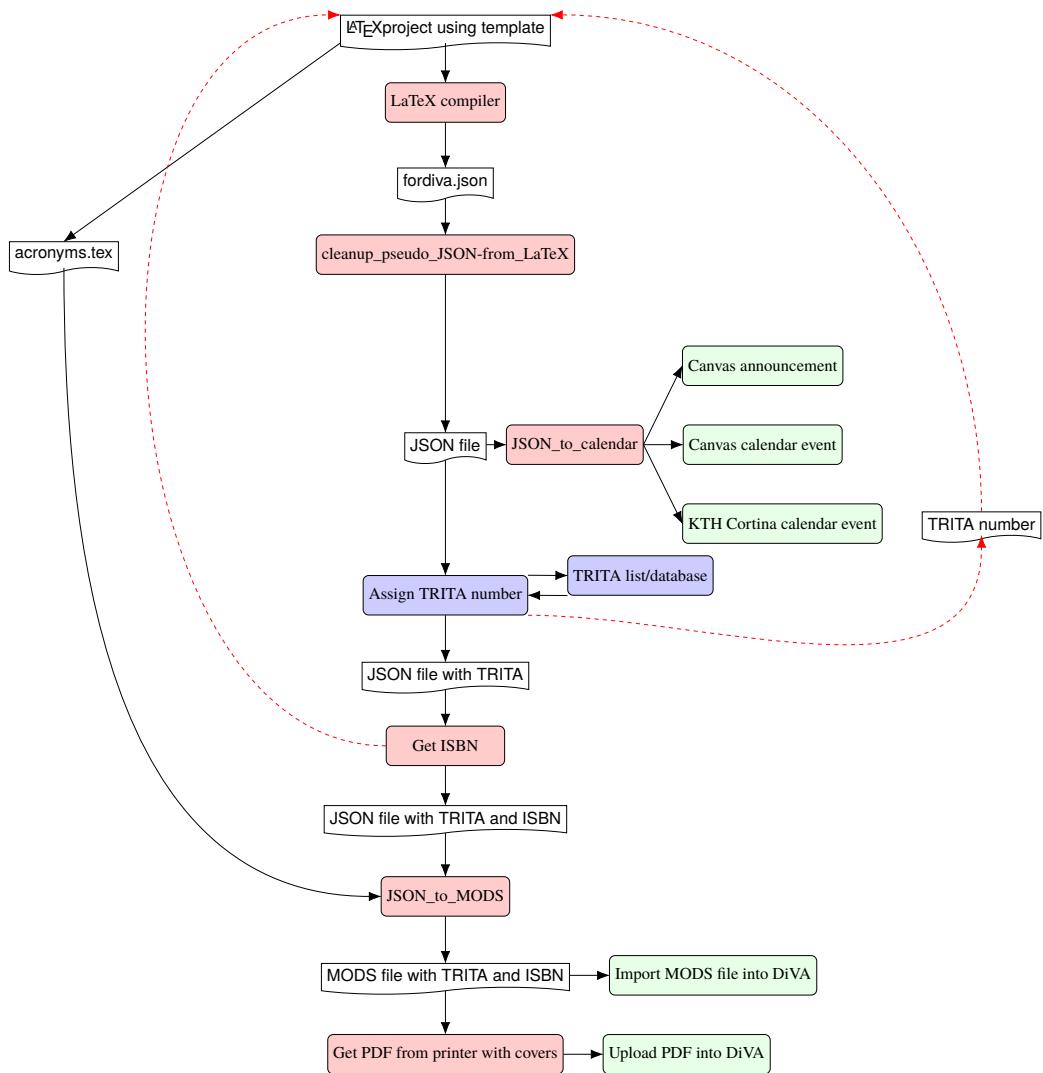


Figure 1: Overall data flow from a `LaTeX` project using the templates. The DiVA administrators' activities are illustrated with the blue boxes: Assign TRITA number and the TRITA list/database. The dashed red line shows the flow of the assigned TRITA number or ISBN back to the author to input into their `LaTeX` - project, while the black lines show other types of data.

8.2 What is a TRITA number and why does each thesis need one?

TRITA stands for Transactions for the Royal Institute of Technology, with the letter "A" appended to it. The definition is given in the 1971 report, "Mall för publikationsserier vid Kungl. Tekniska högskolan i Stockholm", TRITA-LIB-1001, <http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-127656>.

The format for TRITA numbers for third-cycle theses is TRITA-(school acronym)-XXX-YYYY:nnnn, where DLT for ABE, FOU for CBH & SCI, and AVL for EECS & ITM; and nnnn is a sequential number starting from 1 each year with the numbers assigned in chronological order to theses*. Note that the list of assigned TRITA numbers is archived each year†. The year, YYYY, is based on the year that the thesis was *published*.

The TRITA number value can be set with a macro that takes two arguments: series and year:number as shown below‡:

```
% for entering the TRITA number
\trita{TRITA-EECS-AVL}{2025:00}
```

Ideally, there should be a means to automatically get a TRITA number – similar to how one can get an ISBN.

8.3 Entering the metadata into DiVA

If a thesis has used this template, the "For DIVA" page contains the metadata for DiVA and the student could cut and paste this data into DiVA. Alternatively, this metadata can be extracted with a program from the PDF file to produce a JSON file that can subsequently be used to create a MODS file for import into DiVA. The L^AT_EX compiler produced a file called `fordiva.json` that contains the metadata and attaches this file to the thesis (to facilitate it being available).

Programs that can be used to extract data and to take a JSON file and create a MODS file are available from <https://github.com/gqmaguirejr/E-earning>.

*according to one of KTH's archivists

†It seems that this archiving is done twice a year.

‡The number '00' has been used in the example, as this number will **never** occur in a valid TRITA number.

Note that the import of the MODS file does **not import the collaboration data**, even though this is in the file. This is a limitation of the DiVA import function. Therefore, this information has to be manually entered along with uploading of the PDF file itself.

As mentioned earlier, there is also a `citedtags.bib` file that can be helpful when entering the included publications into DiVA.

With the new Cora-based DiVA system (expected in Spring 2026), it should be possible to avoid making a MODS file and instead use the Cora API to directly insert the metadata and thesis into DiVA.

8.4 Make the MODS file

If you save the `fordiva.json` file locally, then you can clean it up / transform it and make a MODS file with the following commands:

Listing 1: Cleanup pseudo JSON produced by the L^AT_EX compiler and then make a MODS file

```
./cleanup_pseudo_JSON-from_LaTeX.py --json fordiva.json --acronyms acronyms.tex
./JSON_to_MODS.py --json fordiva-cleaned.json
```

Now all you have to do is rename the XML file (`modsXML.xml`) that was produced to `xxx.mods` and you are all set to upload the MODS file into DiVA!

To find the `fordiva.json` file in the L^AT_EX project in Overleaf, look for the “Other logs and files” button shown at the bottom of the log output window. After clicking this button, you will see a list of logs and other files, and can now download them. Alternatively, you can get it from the PDF file, as described in Section 8.3.

Note that if the student has used the `glossaries` package to use acronyms in the abstract(s) they also need to provide an `acronyms.tex` file to the script.

8.5 Other uses of the JSON file

Not only is the JSON data useful with respect to DiVA and LADOK, but it can also be used for other purposes. For example, as shown in Figure 1, a supervisor or other person can use it for announcing the student’s oral presentation or defense via a Canvas course room and calendar and potentially in the KTH Cortina Calendar.

9 (Human) Readers of the thesis

Some theses have very few downloads from DiVA, while some have had hundreds of thousands of downloads. Therefore, you should remember that you have a wide range of (potential) human readers of your thesis. These readers include other students looking for information related to their own thesis or because they are interested in the future work that you have suggested, as potential topics for their own projects. Additionally, researchers who are looking at your results may find your thesis relevant to them. In many cases, companies will look at theses for ideas about what the state of the art is - in several cases, theses have been important as “prior art” and this invalidated patents that had been issued when the patent was submitted after the thesis became public (hence it pays to have theses publicly available as soon as possible). Other human readers are the Universitetskanslersämbetet (UKÄ) review teams that examine the degree programs offered at KTH. Finally, as KTH is a public agency, it is important that the general public know what is done at KTH*.

9.1 Machines reading the metadata or full text of the thesis

The file `pdf_related_includes.tex` contains L^AT_EX code that stores the title, author(s), and keyword information into the PDF document in such a way that if you ask for the properties of the PDF file you will get this data[†]. This information makes it easier for machines to get this information from the PDF file.

The thesis also uses features of Adobe’s Extensible Metadata Platform (XMP) and the `hyperxmp` package to embed data into the thesis. This data can be accessed by programs without parsing the PDF.

Additionally, many search engines (such as Google’s search engine) mine DiVA for the metadata and if the full text of the thesis is published via DiVA then they also process the full text of the thesis. The result is that search engines can find the content in these theses. This is likely to increase the probability that someone will download your thesis if they think it is relevant to them – increasing the number of your human readers (see Section 9). It also increases the probability that someone might find your work worth citing in their own publications.

*This is an important part of the Swedish Offentlighetsprincipen.

[†]This does not work correctly in Microsoft’s Edge: it gives the wrong paper size, does not show the author(s), keywords, and PDF version. Okular shows only some of the data. In contrast, Adobe Acrobat shows all of the basic document properties and enables you to view “Additional metadata”.

To facilitate machines reading your thesis, you should **not** print your thesis as an image (*i.e.*, with each page being a bitmapped image of the page). This will *not* necessarily keep machines from reading your content, as they will typically apply optical character recognition (OCR) to the file. Additionally, doing so would make your thesis content much harder for humans who use screen readers and other accessibility technology to help them access material.

9.2 Template author and maintainers

KTH periodically changes the cover design for theses, introduces new programs of study, eliminates programs of study, reorganizes administratively, and faculty move between schools, departments, and divisions. It can be expected that this template will need to evolve with these changes.

For example, if there is a change in schools or programs, then there needs to be changes made to the file `schools_and_programs_3rd_cycle.ins`. While the current file was extracted from KOPPS, the program that does this will need to be replaced because further development of KOPPS has been terminated by KTH's central IT unit which plans to transition all of this information to LADOK.

As another example, on 13 December 2021, there was a change in the KTH cover for 1st and 2nd theses, and the cover generator web service was shutdown. The initial draft version of the cover used a proprietary font (TheSans B4 SemiLight and TheSans B6 SemiBold). The version that was publicly introduced uses another proprietary font (Arial) and officially only existed as a DOCX file for a thesis in Swedish. The result is that I had to make my own version in \LaTeX to try to emulate the DOCX cover. This led to a lot of effort, but one can get a reasonable cover with the correct font. There was another change in covers introduced on 2024-06-05. Once again, a new \LaTeX cover needed to be generated, along with a set of DOCX files that provide pull-down menus for configuring the cover. Recently, a new thesis template "Avhandlingsmall Word" from <https://www.kth.se/student/studier/examensarbete/avhandlingarochexamensarbeten/mall-for-avhandling-1.458236> was introduced. Unfortunately, this template did not provide a specification of the covers, title page, or book information page – since the authors of this new template simply assumed that US-AB would make these.

10 While writing

The thesis template contains lots of examples, notes, and comments. One method to provide additional information is the use of `\todo`. Several different types of `\todo` notes have been used in the thesis. These are described in Section 10.1.

10.1 Conventions for todo notes

The example thesis text includes extensive comments, directions, and warnings. These follow the form shown below:

```
\generalExpl{Comments/directions/... in English}
\sweExpl{Text på svenska}
\engExpl{English descriptions about formatting}
\warningExpl{warning}
```

and appear as:

Comments/directions/... in English

Text på svenska

English descriptions about formatting

Warning

Each of the above is a macro, so as usual in L^AT_EX you can redefine it - even defining it to produce nothing! Several previous students have placed these re-definitions in their `custom_configuration.tex` file.

10.2 Removing README_notes

As the various README notes are targeted at different readers, you may or not want to see them. I would suggest keeping the README files around (at least for a little while) as a source of examples of how to do things. Despite having spent a very large number of hours working on the template and drafts of students' theses, I find some of the README files very helpful as a reminder of how to do things.

At some point, you will no longer want this README information. You can remove the individual files or even remove the `README_notes` directory.

10.3 Removing unused fonts

This version of the template has some font information, in the form of Opentype Font files (with the extension “.otf”) and TrueType font files (with the extension “.ttf”). If you are not using these fonts (and no longer using any of the README files), then you can delete these font files.

Note that the Figtree font is needed to make the covers, titlepage, and some preface pages. It is also used for all of the headings! So you should **not** delete this font.

Acronyms

API	Application Programming Interface
DiVA	Digitala Vetenskapliga Arkivet
DOI	Digital Object Identifier
IMRAD	Introduction, Methods, Results, and Discussion
ISBN	International Standard Book Number
JSON	JavaScript Object Notation
KOPPS	Kurs- och programplaneringssystemet
KTH	KTH Royal Institute of Technology
KTHB	KTH Biblioteket
LADOK	Lokalt adb–baserat dokumentationssystem
MODS	Metadata Object Description Schema
UKÄ	Universitetskanslersämbetet
URL	Uniform Resource Locator

References

- [1] Shiva Besharat Pour and Qi Li, “Connecting Silos: Automation system for thesis processing in Canvas and DiVA,” Bachelor’s Thesis, KTH Royal Institute of Technology, School of Electrical Engineering and Computer Science, 2018, pp. xiii, 101. [Online]. Available: <https://urn.kb.se/resolve?urn=urn%3Anbn%3Ase%3Akth%3Adiva-230996>.
- [2] Raymond R. Panko, “What we don’t know about spreadsheet errors today: The facts, why we don’t believe them, and what we need to do,” in *Proceedings of the EuSpRIG 2015 Conference “Spreadsheet Risk Management”*, European Spreadsheet Risks Interest Group (www.eusprig.org), 2015, pp. 79–93, ISBN: 978-1-905404-52-0. doi: <https://doi.org/10.48550/arXiv.1602.02601>. [Online]. Available: <https://arxiv.org/pdf/1602.02601.pdf>.