Guideline for Thesis Projects in the Research Group on Process Analytics

Han van der Aa

Junior Professor, School of Business Informatics, University of Mannheim
han.van.der.aa@uni-mannheim.de

Last updated: July 18, 2023

1 Introduction

This document provides information for students that are interested in, preparing, or are working on a bachelor or master thesis in the Research Group on Process Analytics at the University of Mannheim. Its aim is to answer the most common procedural concerns, as well as give helpful input for the various stages of a thesis project. The information in this document trumps the more generic DWS thesis guidelines.

The remainder of this document focuses on the preparation (Section 2), execution (Section 3), and completion of your project (Section 4). The templates and resources referred to in this document are available at https://bwsyncandshare.kit.edu/s/g9fs2yQB4xC6SfA.

2 Before You Start

This section provides information on eligibility criteria (Section 2.1), topics and project types (Section 2.2), how to apply (Section 2.3), the thesis proposal (Section 2.4), and thesis registration (Section 2.5).

2.1 Eligibility

Please consider the following requirements before applying. These prerequisites are necessary to ensure that you have sufficient experience with the topics of concern to be able to successfully conduct a thesis project.

Bachelor students. The following criteria apply:
- Enrolled in the Business Informatics program;
- Completed the Seminar on Process Analytics (SM 457) or other demonstrable experience with process analysis;
- Experience with Python (required for most projects).

Master students. The following criteria apply:
- Enrolled in the Business Informatics or Mannheim Master in Data Science program;

- Completed Process Management & Analytics (IS515) and/or Advanced Process Mining (IE 692);
- Preferably: Completed the Seminar on Process Analytics (CS 719);
- Experience with Python (required for most projects).

Note that we are a small research group and we receive way more applications than we can accept. The above criteria just reflect the minimal requirements; exceptions for excellent students may apply.

2.2   Topics & Project Types

Finding a topic. We generally do not have topics readily available to hand out to students (though there are exceptions); we typically define topics in collaboration with a student.

Thesis projects should fit the research topics that are of interest to our members. This primarily includes technical research conducted in the area of process analysis. Good starting points for finding a topic are:
- Considering future research directions identified during your seminar project;
- Reading papers published by our group;
- Reading papers in recent proceedings of the conferences on Business Process Management (BPM), Process Mining (ICPM), and Advanced Information System Engineering (CAISE).

Project types. A project typically fits one of the following categories:
- Conceptual development: These projects involve the development, implementation, and evaluation of an *artifact* (typically an algorithmic approach) that solves a particular problem;
- Experimental study: These projects aim to provide new insights into the capabilities of existing approaches through a structured comparison or by applying them in a new setting;
- Empirical study: These projects aim to provide new insights by studying a phenomenon in an industry, e.g., through case studies or interviews. This project type is less common, given our main focus on technical contributions.

Industry collaborations. Projects can be conducted in collaboration with an industrial partner, in which case the topic should be provided by them (or you). Note that your thesis project must still make a research contribution, i.e., the results should be of interest to a broader audience than just the industrial partner itself. Note that successful industry projects require careful alignment between the interests of the academic and industrial sides.

2.3   Application

If you meet the aforementioned eligibility criteria and have one or more topics of interest in mind, you can apply for a thesis project by sending an e-mail with the following contents to Prof. Van der Aa:
- A short motivation for writing a thesis at our chair;

– Your qualifications with respect to the eligibility criteria;
– One or more topics that are of interest to you (as concrete as possible, do not just write that you want "to do something with process mining");
– Intended start date.

Make sure to include the following attachments:

– CV;
– Transcript of records;
– Thesis candidate profile form;

Applications that do not meet these requirements might not receive a response.

When to apply. We receive more applications than we can handle per semester, which means that it is important that you apply on time (and in a convincing manner), i.e., several months in advance. Note that, depending on our project and application load, we may collect applications until a certain time moment and then make a selection.

2.4   Thesis Proposal

If your application is selected, we typically set up a meeting to discuss the potential topics, with the aim of establishing a concrete research direction. Once we have agreed on one (and not sooner), it is time for you to start working on a thesis proposal. This is a crucial step for the success of a thesis project!

Proposal guidelines. Your thesis proposal describes the plans for your thesis project. It should clearly specify the problem that your thesis aims to address, the research goals or questions that will provide a solution to the raised problem, and outline a high-level work plan to achieve this. Keep the proposal concise (max. two pages, excluding literature). Make sure to use and follow our proposal template and its instructions.

The contents of a proposal are not binding. Deviations may happen during your project, which is fine when done in coordination with your supervisor.

Proposal procedure. The typical procedure for writing your proposal is as follows. After agreeing to write a proposal on a certain research direction, you will start working towards a first draft version. This involves familiarizing yourself with the main literature that is relevant to your topic (do not just start reading literature after registration).

Once you feel that you have a version suitable for discussion or you could really use input, you can send your proposal to your supervisor for feedback. Generally, there will then be a few iterations to reach a final version, which may be discussed in a meeting before the final decision to get started.

IMPORTANT. Note that writing a proposal is not yet a commitment from either side to conduct or supervise a project! There is always a chance (for you or us) to decide against a thesis project during its preparation, e.g., if we cannot agree on the research objectives or we question the project's feasibility.

Supervision. Most projects in the Process Analytics group are supervised by Prof. Han van der Aa. However, for certain topics, a doctoral candidate from

the PA group may serve as your main supervisor, with Prof. Van der Aa serving as a secondary supervisor and the formal examiner.

2.5   Thesis Registration

Once we have agreed on a final version of your proposal, we can officially register your thesis project.

Registration form. Fill out the thesis registration form and send it to Prof. Van der Aa, who will then sign and submit it.

Start and submission date. You can freely choose the official start date and the corresponding final submission date based on your planning, though the start should not be unnecessarily long after finalizing the proposal. The final submission date is 3 (for bachelor) or 6 (for master) months after the start date, or the earliest working day afterwards (in case of a weekend or holiday).

IMPORTANT#1. The final submission date is binding. It can only be extended in exceptional circumstances (e.g., for medical reasons), in which case you must approach the education administration.

IMPORTANT#2. Because of this strict deadline, make sure that you are able to run all software tools or libraries that are crucial to your project *before* registration. The same applies to projects that strongly rely on the availability of certain data (e.g., in industry projects).

3   During Your Project

This section provides information on the execution of your project itself (Section 3.1), project meetings with your supervisor (Section 3.2), and the writing of your thesis (Section 3.3).

3.1   Project Execution

The execution of your project should roughly follow the steps you outlined in your thesis proposal. What these steps are, how you should perform them, and how you progress over time will differ a lot per project. As indicated before, it is fine to deviate from the plans made in your proposal, and it is natural that steps are not conducted in a purely sequential fashion or follow the schedule you initially had in mind. However, make sure to discuss significant deviations with your supervisor.

Make sure that you follow an appropriate research methodology and best practices throughout your project. This means that, for example, all choices you make during the project should be justifiable in a way (e.g., related to approaches to select or compare to, how to design your approach, which parameters to use during experiments). A great recommendation here is to draw inspiration from high-quality research papers that relate to your topic.

Common datasets. The most common resources for data in the area of process analytics are:

- **Event logs:** Public real-world and synthetic event logs can be found here. A small set of object-centric event logs is available here.
- **Process models:** A large and a huge collection of process models are available, stemming from users of the academic version of the SAP Signavio process modeling tool. Note that these models differ wildly in terms of characteristics and quality.
- **Generating synthetic models and logs:** There are various options to generate synthetic data yourself. This includes but is not limited to the PTandLogGenerator for process models and logs, the BIMP online simulator, and CDLG for the generation of event logs with concept drifts.

Compute resources. Check out the DWS recommendations for resources to run large computations, e.g., using GPUs. For medium-sized computations or quick prototyping, you may also try options such as Google Colab.

3.2 Project Meetings

Once your project has officially started, we will schedule regular progress meetings between you and your supervisor. Such meetings are scheduled to take 45 minutes and typically occur once every 2 weeks (4 for industry projects).

To streamline the meetings and general communication, please consider the following points carefully:

Meeting agenda. Please send your supervisor a brief agenda the day before a meeting with the main points you would like to discuss. Also indicate if you would like to do the meeting in person or on Zoom (if not yet clear). The idea behind this agenda is two-fold: 1) it's beneficial for you to think ahead about what you want to get out of a meeting, 2) it gives your supervisor an idea of what to expect during the meeting and prepare for it where applicable.

A meeting agenda should consist of a few bullet points, where each point should be something that will help you further in your project and should be as concrete as possible, e.g., not "discuss approach" but rather "discuss how to overcome issues faced during the transformation from BPMN to RDF" or "discuss which settings to test in the evaluation experiments". Smaller questions can be grouped into a larger point, e.g., "miscellaneous organizational matters". Make sure that it is realistic to discuss all your points in a single, 45-minute meeting.

If you do not have any points to discuss, you can just cancel a meeting via e-mail. However, it is best not to do this too often, to make sure that you regularly receive input and that your supervisor is aware of the project's progress.

During the meeting. A thesis meeting should be a conversation between you and your supervisor to help you further in your project. This means that it should not be a one-sided presentation from your side.

Meetings typically start with a short update on your progress, but this should be kept brief. The bulk of the meeting should relate to the points listed in the meeting agenda, though allowing for flexibility where beneficial. You can use slides to show specific things, such as examples, results, or overview figures.

Please don't establish an entire slide deck for that (with a title, outline, etc.). Finally, showing code or raw data is generally not useful during a meeting since it's really difficult to help you with that on the spot.

Meetings are scheduled to take up to 45 minutes. You are responsible for managing the time that you have. Do not just assume that the meeting can run longer, though you can, of course, ask.

IMPORTANT. We reserve the option to reduce the frequency of meetings if the above aspects are repeatedly disregarded or if the meetings are generally unproductive, e.g., when provided input is ignored. Furthermore, meeting schedules are subject to deviations and irregularities, especially outside of the lecture periods.

Additional communication. If you have any urgent concerns or small questions outside of the regularly scheduled meetings, just write an e-mail to your supervisor, either to discuss things directly or to schedule an additional meeting.

3.3   Writing Your Thesis

Formalities. Your thesis should be written in English using our LaTeX template (a slightly modified version of the general DWS template). The target audience for your thesis consists of people with a similar background as you, e.g., your fellow students.

Thesis length. The page count of your thesis should be appropriate for its contents, which depends heavily on the type of research that you conducted. Nevertheless, bachelor theses generally span about 40 to 50 pages of content (from Chapter 1 up through the end of the conclusion), and master theses about 50 to 65 pages. Do not exceed these limits without explicit permission; writing involves separating important aspects from less relevant details.

Writing process. Although everyone has different preferences when it comes to writing, consider the following recommendations:
 – **Start early.** Writing down your ideas or describing what you have done in a research project is a great way to get a different perspective, e.g., allowing you to recognize if things make sense and if your methodological choices are justifiable. Therefore, start writing (or rather sketching) early in your thesis project, rather than waiting until the end.
 – **Sketch first.** Focus on *what* to write before thinking about *how* to write it. This means that you should start by sketching bullet points, rather than writing full-fledged sentences.[1] Keep revising these until you feel that the storyline makes sense. Only then start investing time into *how* to write things down and by filling in additional details. This is much more efficient since the sketching approach allows you to quickly reiterate. Furthermore, this procedure can be applied at any level of detail, i.e., you can sketch bullet points at a chapter, section, and paragraph level (see next point).

---

[1] You don't need Latex for this part, so using pen-and-paper, a whiteboard, post-its, etc. can be helpful here.

– **Write in a top-down manner.** If the high-level flow in your thesis does not work, any effort invested into the more fine-granular details will not save you from that. Therefore, work in a top-down manner: first settle on the chapters that your thesis should comprise, then split a chapter into individual sections, then split a section into individual paragraphs, and finally think about the individual sentences that should go into a paragraph.[2] Do not start from the low level, e.g., by starting to write a bunch of a text and trying to figure out later how to group them into different paragraphs or sections.

– **Revise, revise, revise.** Revising is one of the key parts of the writing process. You cannot expect that any first version of what you write will be the final version. This is why we recommend you to start by sketching (less time spent if you decide to change something later). Also, you should not get too attached to any particular part; be prepared to remove or completely rewrite certain parts (whether it is a sentence, paragraph, or even an entire section).

Structure. Make sure to follow a well-established structure for your thesis. This saves your work (no need to reinvent the wheel) and makes things easier for readers. Again, it is a good idea to draw inspiration from high-quality research papers that present similar research.

For conceptual-development projects, a typical structure is as follows:

1. Introduction: Describes the research context (before Section 1.1), problem statement (Section 1.1), research goals and contributions (Section 1.2), and briefly discuss the structure of the remainder (Section 1.3).
2. Background: Describes all relevant concepts, definitions, and related research. Make sure that it is clear how each part of this chapter relates to your research goals. Introducing a running example is often very helpful here. Only use formalisms (like definitions, equations, or algorithms) if these are important later on.
3. Conceptual approach: Presents the approach you have developed to solve a particular problem. Each section in this chapter should represent a logical component, e.g., a single step of your approach.
4. Evaluation: Describes the evaluation experiments you conducted. Generally, this section should consist of a description of the employed data collection (Section 4.1), evaluation setup, incl. implementation details (Section 4.2), results (Section 4.3), and an optional discussion (Section 4.4).
5. Conclusion: Concludes the thesis with a summary of what has been done, a discussion of limitations, and ideas for future research.

Deviations from this structure are common. For example, you can have explicit chapters or sections for things like Preliminaries or a Problem Illustration. However, it is then still best to try to follow best practices from research papers. The structure of Chapter 1 should in principle not deviate (and follow ideas from your project proposal).

---

[2] You don't have to work in a breadth-first manner, e.g., you can start filling in details in Chapter 2, without having settled on the section structure of Chapter 3.

IMPORTANT. Section 1.3 (or its counterpart in your thesis) should describe the high-level structure of your thesis (i.e., what is the purpose of each upcoming chapter), whereas each subsequent chapter should start with a description of what the chapter does, followed by an explanation of the chapter structure (e.g., "Section X.1 does this, Section X.2 does that..".

If you struggle to explain the chapter structure of your thesis or of the sections within a chapter, e.g., because it is not clear which parts belong in Chapter 2 and which parts belong in Chapter 3, or because it is not clear why Section 3.2 is followed by Section 3.3, there is most likely something wrong with your structure! This means you need to go back to the drawing board.

Within a section, each paragraph should have a single clear topic/purpose (just like each chapter and each section should) and the paragraphs within a section should also form a logical story, i.e., also at this granularity, the relation between one paragraph and its predecessor and successor should be obvious.

Helpful resources and tips. Aside from learning from papers that you like, useful resources for writing include (all pdfs are also in the shared folder):
- LaTeX Do's and Don'ts by myself. **Follow these rules!**
- *How to Write Beautiful Process-and-Data-Science Papers?* by Wil van der Aalst. Very insightful!
- *Some Frequent Writing Tips I Give* by Philipp Leitner. A bunch of concrete tips!
- For those who want to go the extra mile, there are also tons of books on writing, some specific to academic research and some are more general purpose. Examples include Writing for Computer Science, How to Not Write Bad, Style: Towards Clarity and Grace.[3]
- AI Tools for Research Workflow in Academia by Niels van Quaquebeke. A great list of tools for all kinds of tasks in research, including but not limited to writing.

ChatGPT etc. Generative AI tools like ChatGPT are great tools to support you in your writing process. However, they should be used with care. The most important thing to consider is that you are 100% responsible for the contents of your thesis, whether you wrote them from scratch, used writing tips from tools like Grammarly, or if you used ChatGPT for certain parts. It is crucial that you recognize that ChatGPT does not actually understand the answers that it provides, can reply with unwarranted confidence, may provide non-existing references, and it does not know the rest of your thesis, e.g., which terminology you have defined.

Therefore, you should only use this technology for tasks for which you can (and will) verify the correctness of the output. This means that with respect to your thesis, you should primarily use it as a tool for writing support, not to come up with argumentation and other actual content. For example, I like to use ChatGPT to get quick suggestions on how to turn a set of bullet points into a nicely flowing paragraph, or how to improve the textual flow of an initial

---

[3] For obvious reasons, we cannot just share pdfs of these books, but I am sure you are able to obtain access to them in some manner.

paragraph I drafted. I do not blindly copy the given answer, but pick and choose the parts of its answer that I like, while also making sure that the meaning and terminology of the final paragraph remain correct.

IMPORTANT. Make sure that you use generative AI tools in such a way that it does not violate plagiarism and other rules of the University of Mannheim. At the time of this writing (24-04-2023), no official university policy has been released, though.

Input on your thesis. While writing your thesis, you can ask your supervisor for feedback, although we will not read entire chapters in detail prior to submission. However, we are happy to provide input on specific aspects, such as your general thesis structure, the storyline in the introduction, or whether or not the level of detail of certain parts appears to be appropriate. Keep in mind here that any feedback that we provide will be based on incomplete information from our side. For instance, whether a thesis structure is correct or not strongly depends on the actual chapter contents. Everything in the thesis is in the end your responsibility!

4 Project Completion

This section provides information on things to consider when submitting your thesis (Section 4.1) and what happens afterwards (Section 4.2)

4.1 Submitting Your Thesis

Project repository. Create a single repository in which we can find any additional materials that are relevant to your thesis, such as source code, collected data, detailed analysis results, etc, and **provide a link to this repository in your thesis**. This link typically appears at the start of the evaluation chapter.

Generally the preferred option for this is to create a Github repository (especially when code is involved). Your repository should either be public or, if private, shared with @hanvanderaa (and your main supervisor, if applicable). To share larger files, other solutions, like Google Drive, Dropbox, or equivalent work as well. Only provide a USB stick instead of a repository link if necessary. Do not provide a CD-ROM.

Source code does not have to be perfectly clean and completely documented; research code rarely is. However, there should be instructions on how to get your code running and to obtain the results that you reported in your thesis.

Note that you can keep editing your repository for up to a week after your digital thesis submission, so that you can use your final days in the project for writing, and clean up/document code later.

Submission procedure. You need to submit your thesis in two ways:
- *Digital submission:* Before or on the final submission date for your thesis project, you have to submit a digital version (in pdf) to Prof. Van der Aa via e-mail. This can be as late as 23:59CET of the final day.

– *Hardcopy submission:* Within two weeks after your deadline, you need to submit **two** printed copies of your thesis. We prefer copies on A4 with an adhesive binding; they can be either soft or hardcover. For this submission, either make an appointment with your supervisor or send them via post to:

University of Mannheim
c/o Prof. Han van der Aa (WIM)
L1, 1
68161 Mannheim

IMPORTANT. Make sure to sign the *Ehrenwörtliche Erklärung* in the harcopies of your thesis using pen (no electronic signature)!

Note. There is no presentation or defense of your thesis.

4.2 After Submission

Confirmation of passing. After submitting your thesis, we can provide you with an official confirmation that states that your thesis project will receive a passing grade; you can take this to the administration for a certificate of completion for your degree. Request this confirmation when you submit your thesis digitally. We can typically provide this on rather short notice.

Thesis assessment. When grading a thesis project, we consider the project's scientific contribution, the thesis itself, and the project execution:
- *Contribution:* Foremost, we consider the quality of the scientific contribution(s) that your thesis makes. This considers how much your work advances the state of the art in research, but also relates to factors such as methodological soundness and quality of the evaluation.
- *Thesis itself:* We consider the quality in which the project is presented in the thesis, which ranges from factors such as the appropriateness of the thesis structure, to having a clear problem and research goals, the coverage of relevant literature, to the quality of the writing.
- *Project execution:* Finally, we also consider how well you worked during the project itself, which includes, e.g., your problem solving ability, project management, and communication with your supervisor.

Once we have finished your assessment, we will submit it to the examination administration and also send you our assessment via e-mail. If you have lost access to your student e-mail address in the meantime, let us know how to reach you. If desired, we are happy to have a final meeting to discuss the assessment and close our collaboration properly.

Publishing your work. If you conducted excellent work during your project, it may be possible to try to have it published in the form of a conference paper or a journal article. Let us know if this is something that interests you!