Letter From the Editors

By Mary Pat Campbell and Jennifer Wang

Welcome to the all-digital version of *The Modeling Platform*! What has changed?

- New articles are published every other month instead of only twice a year.
- For each issue, two to three articles will be published.
- Without paper page limitations, article length is dependent on the specific content's needs.

What hasn't changed?

- Great content!

We are looking forward to working with authors to take advantage of this new format. We've already tried out the digital advantage in our November 2019 issue with the publication of "Confessions of an Efficiency Junkie" by Jeff Samu.[1]

In the print version of *The Modeling Platform*, we could include only half of Jeff's article because of page-count limitations. However, we did not have such restrictions in the digital format and were able to publish the entire piece.

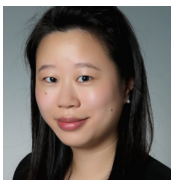
Separately, with the more frequent publication schedule, one author can respond to other authors' pieces without having to wait half a year to see print. The publication lag we had in our prior process made it difficult for us to keep up to date, and when authors were working on multipart articles, a missed deadline meant we could end up with a lag of six months or a year until we could publish the next installment.

If you're interested in writing for *The Modeling Platform*, please contact us, your friendly editors! With the more frequent publication, we are always taking submissions. We are interested in any modeling-related articles, whether it's getting into the technical nitty-gritty detail of structuring a database (as we see in Matthew Caseres' piece, "Tidy Data Formats and Cloud Storage, Part 1: Tidy Data") or discussing high-level issues in how best practices from IT can be considered for actuarial modeling (as in Bryon Robidoux's article, "The Importance of Centralization of Actuarial Modeling Functions, Part 2: DevOps—The Path to Actuarial Modernization and Consolidation").

From relative modeling beginners to grizzled modeling veterans, we're interested in hearing from all of you! Perhaps instead of answers, you have questions for the actuarial modeling community. Share your challenges, successes, not-so-successes, helpful tips or useful frameworks.

We hope to be hearing from you—yes, you, the person reading this right now. Email Mary Pat Campbell, and get started writing! ■

Mary Pat Campbell, FSA, MAAA, PRM, is vice president, insurance research, at Conning in Hartford, Connecticut. She can be reached at *marypat.campbell@gmail.com.*



Jennifer Wang, FSA, CERA, MAAA, is a consulting actuary at Milliman. She can be reached at *jennifer.wang@milliman.com.*

ENDNOTE

1 Samu, Jeff. Confessions of an Efficiency Junkie. *The Modeling Platform*, November 2019, *http://digitaleditions.walsworthprintgroup.com/display_article.php?id=3514494&view=629391* (accessed March 11, 2020).

Get Plugged in—New InsurTech Partnership

The SOA and Plug and Play relationship will allow InsurTech start-ups to validate their technology and modeling processes with actuaries. In turn, SOA members will have an exclusive look inside the world of emerging technologies. These efforts will help with the development of fair and financially sound insurance products to better serve consumers.

The strategic partnership with Plug and Play demonstrates the SOA's commitment to providing its members with dynamic learning experiences, rewarding volunteer opportunities, and collaborative events where they can learn from the experiences and ideas of peers around the world. Through this partnership SOA members and start-ups can share best practices and advance ideas for the benefit of the insurance industry, regulators and the public. The SOA and Plug and Play officially announce this partnership to support an exchange of knowledge between actuaries and start-ups. ■

SOCIETY OF ACTUARIES
MODELING SECTION

Tidy Data Formats and Cloud Storage, Part 1

Tidy Data

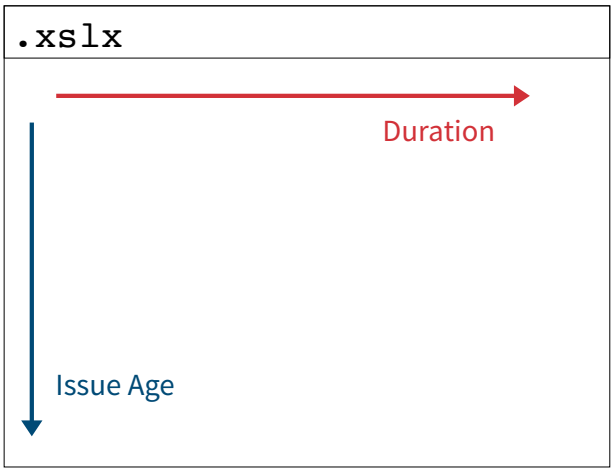
By Matthew Caseres

As of this writing, there are 3,010 Excel files in the mortality table repository. These files are in a format that does not play well with the databases R and Python. I converted all 2015 VBT and 2017 CSO tables (about 150 Excel files) into a format that can be easily joined to an experience study engine for A/E analysis. In this article, we discuss why we would want to reformat mortality tables and demonstrate how to reformat hundreds of select and ultimate tables quickly using R.

CURRENT MORTALITY TABLE FORMATS

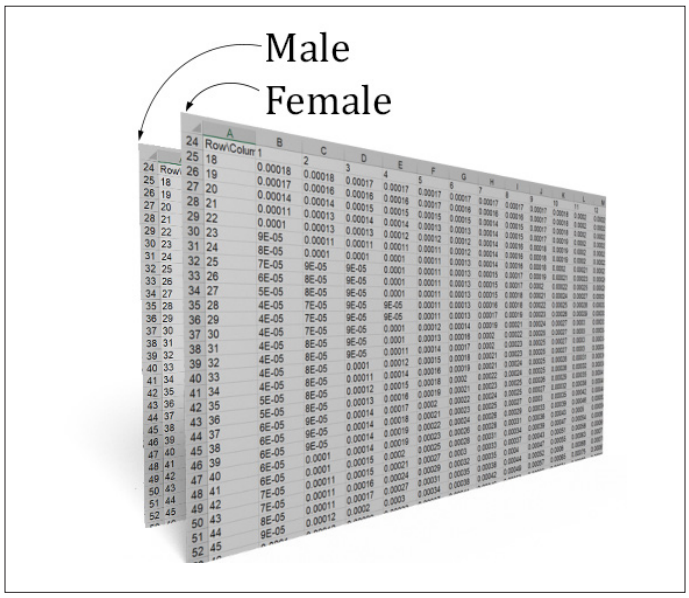
Let us consider a situation where the existing mortality tables are stored in Excel. For each demographic split, we have a single Excel file that contains the select and ultimate tables. In the select table worksheets, rows represent issue ages and columns represent durations. The ultimate table is stored as a single column with mortality rates for each attained age, as shown in Figure 1.

Figure 1
Structure of Mortality Table



Some mortality table sets have many demographics, which leads to many separate files, as shown in Figure 2. The 2017 CSO tables with preferred structure have 10 Excel files associated with them.

Figure 2
Structure of Mortality Table Set With Multiple Demographics

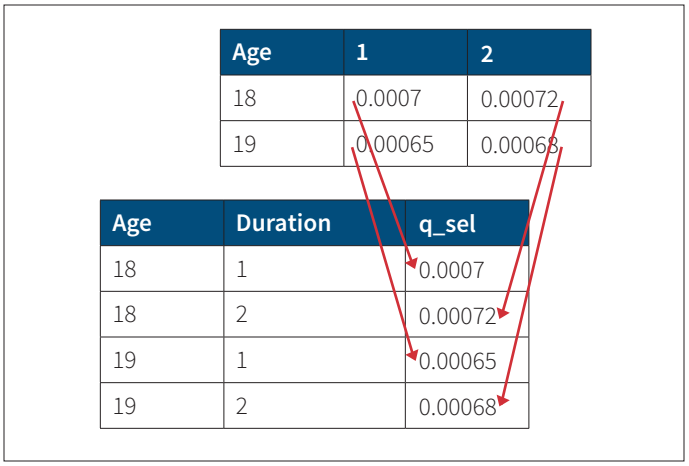


My view is that we should stack all the tables on top of each other and add new columns to serve as identifiers. In this way, we can include an arbitrary number of tables within a single rectangle of data that has columns to identify things like the mortality basis. This is one of the steps in making the dataset "tidy."

WHAT IS TIDY DATA?

Tidy data is a concept that was introduced in a paper by Hadley Wickham. Put simply, each column represents a variable and each row represents an observation. The mortality table Excel files are not tidy because the mortality variable belongs to many columns instead of a single column. Figure 3 demonstrates how one could reformat these tables into a tidy format. See that there is now a single column representing the mortality rate variable.

Figure 3
Tidying Process



There are functions to perform this operation in R, Python, SQL and Power Query.

WHY TIDY DATA?

A single table can be created from multiple files, and tidy data also allows for join operations.

> The tidy data format allows for join operations.

Fewer Tables

There are 28 separate Excel files for the 2015 VBT ALB Relative Risk tables. Figure 4 shows what the Excel files for the 2015 VBT ALB Relative Risk tables look like after being converted to a single data frame in a tidy format.

Instead of having a separate block of data for separate genders, we include an additional column that distinguishes the gender. Instead of adding a new table for each demographic split included in the table structure, we add a new column. This allows for a single table to be made from the original 28 separate blocks of data.

We don't have to put the data in a tidy format to condense the Excel files into a single dataset. We could add more columns to the dataset and stack the data without moving the durations into their own column. However, this would make it more difficult to do lookups. The tidy data format allows for join operations instead of forcing us into the INDEX-MATCH-MATCH Excel pattern.

Join Operations

Suppose we are building a database application that analyzes company mortality experience. Figure 5 shows what a section of that data might look like.

Figure 4
Tidy Select Mortality Format

gender	tobacco	relative_risk	issue_age	duration	q_sel
Female	Non-Smoker	RR50	18	1	0.00018
Female	Non-Smoker	RR50	18	2	0.00018
Female	Non-Smoker	RR50	18	3	0.00017
Female	Non-Smoker	RR50	18	4	0.00017
Female	Non-Smoker	RR50	18	5	0.00017
Female	Non-Smoker	RR50	18	6	0.00017
Female	Non-Smoker	RR50	18	7	0.00017
Female	Non-Smoker	RR50	18	8	0.00017
Female	Non-Smoker	RR50	18	9	0.00017
Female	Non-Smoker	RR50	18	10	0.00018
Female	Non-Smoker	RR50	18	11	0.00020

Figure 5
Example Experience Database Format

Gender	Duration	Issue Age	Exposures	Death
F	12	30	1	0
F	13	30	1	0
F	14	30	1	0
F	15	30	1	1

To do an A/E analysis, we must add a new column representing the expected number of deaths. This information would be filled in using mortality tables. An INDEX-MATCH-MATCH looking up the mortality rate for issue age 30 and duration 15 from the non-tidy Excel tables looks like this.

```
"=INDEX(table,MATCH(30,issue_age),
MATCH(15,duration))"
```

Because we have different tables for each demographic, the actual formula would involve some conditional logic that changes the mortality table on the basis of the demographic. If you have 28 tables to choose from, this becomes unworkable in Excel without using VBA.

Changing to a tidy data format will help, because each column specifies a demographic variable and we can create a unique lookup identifier for each combination. In Excel, we concatenate the demographic columns to create an identifier column in our tidy mortality table and in our mortality experience.

```
"gender&risk&tobacco&issue_age&duration"
```

These new columns can be used as keys in a VLOOKUP to assign the appropriate rates to our experience intervals. In R/Python/SQL, this sort of operation is common and is said to be a "join on multiple keys." Here is how we could attach the 2017 CSO Preferred Structure to our experience data using R.

```
left_join(experience, CSO2017_preferred_
structure, by = c(gender, risk, tobacco,
issue_age, duration))
```

DATA CONVERSION PROCESS

The approximately 200 Excel files associated with the 1983 GAM, 2001 CSO, 2015 VBT, 2017 CSO, AM92 and UP1984 tables have been converted into a tidy format. We used R and interacted with Excel using the package "readxl."

We placed the files for a mortality basis in a folder and iterated over them, allowing us to convert the 28 files associated with the 2015 VBT Relative Risk ALB tables much quicker than converting one at a time. Identification columns were created from the cell containing the description of the table for each Excel file. Separate tables are created for each Excel file, which are then stacked on top of each other to create a single table.

The scripts that perform the conversions are in the subfolders of data-raw in the repository for the code. The converted data can be downloaded in a CSV format from OneDrive.

WHAT'S NEXT?

The next article will be on creating a single data table from the select and ultimate tables for real-world use cases. This format is available on OneDrive as the "combined" format. We include a demonstration of Google BigQuery and use it to store an alternative format in which there is a single table that represents what was once 130 Excel files. ■

Matthew Caseres, ASA, is working on starting an actuarial nonprofit. He can be reached at *matthewcaseres@outlook.com* and on GitHub at *github.com/ActuarialAnalyst.*

The Importance of Centralization of Actuarial Modeling Functions, Part 2

DevOps—The Path to Actuarial Modernization and Consolidation

By Bryon Robidoux

The first article in this series stressed that consolidating the actuarial modeling department was an important and worthwhile initiative. But consolidation doesn't solve much if the redundancies and complexities of the modeling department are not reduced in the process.

It was suggested that software engineering practices have many answers to our modeling problems, especially the monolithic system[1] issue, but there was no mention of what concepts were important or how to get started. This article will fill that gap by introducing IT DevOps and other software engineering principles and their application to the current actuarial modernization and modeling department consolidation.

CHANGE, CHANGE, CHANGE AND MORE CHANGE

Let's pause for a second and look at the actuarial modeling and processes landscape. The amount of change required of the insurance organization depends on the type of business written and where it is sold. In 2017, there was VM-20; in 2019, there was VM-21; and in 2021, there will be LDTI (long duration targeted improvements), and IFRS 17 is coming. Not to mention, interest rates have been excruciatingly low and the S&P has been on the rise for the last 11 or 12 years, so the products are getting more equity features to stay competitive.

Auditors and regulators are mandating that senior managers be able to attribute and explain changes to demonstrate confidence that their organization can properly manage its risk. Given the speed with which the playing field is changing, senior managers need to quickly and confidently do what-if analyses to make more informed decisions to stay ahead of their competitors. They need information quickly, which requires processes to be accurate, streamlined and efficient. They can't wait weeks or months for actuaries to update their spreadsheet processes for a decision that needs to be made in a week, or a day or less. As changes accelerate, change management and handling complexity become paramount.

DECENTRALIZATION—THE NASTY TRUTH

The previous article mentioned that decentralizing models is a bad practice and should be avoided, but it failed to recognize the driver behind the behavior. The real motivation for decentralization is to reduce the complexity of models so they are more maintainable and easier to understand.

The desire to keep things simple is a worthy cause, but decentralization is trading model complexity for operational complexity. There is a great book on this topic originally written in 1975

called *The Mythical Man-Month*.[2] The premise of the book is that it is a mistake to think that if one developer can do the job in one year, then hiring 12 developers will get the job done within a month. It explains that this is not possible because people can't learn complex systems instantaneously. Even if this were the case, communication among people and teams causes the development to slow to a crawl because everyone needs to coordinate. The book recommends coming up with a team of specialists who work together to accomplish an overall goal. These specialists should complement each other in such a way that they can maximize the ability to work independently, reducing operational complexity.

KEEPING MODEL COMPLEXITY UNDER CONTROL

In software engineering, refactoring and unit testing are performed together to mitigate model complexity. Refactoring is the practice of cleaning up the models to make them easier to maintain without changing their behavior. Unit testing is writing small, fast and single-purpose tests to verify the software is working as expected. Great books on these topics are *Refactoring: Improving the Design of Existing Code*[3] and *Working Effectively With Legacy Code*.[4]

The less frequently refactoring is performed, the faster the model's complexity will get out of control. If anyone mentions that there needs to be a project to refactor the code base, then this is a good sign that the development practices and standards of the organization should be revisited. Refactoring should be akin to cleaning up the woodshop at the end of each day's shift so that everything is clean and organized for the next day.

HANDLING CHANGE OUTSIDE THE ACTUARIAL PROFESSION

How do Facebook, Amazon, Netflix and Google (FANG)—along with other large technology organizations—handle hundreds of developers confidently, making many changes to their code on a daily basis and not suffering from the same model and operational complexities that actuaries suffer from? The answer is DevOps. It is a framework and guidelines on how to efficiently handle rapid change with confidence and reliability. It allows developer teams and operation teams to work closely together building robust processes and software systems.

DEVOPS

One main goal of DevOps is to shorten the deployment time of fixes and enhancements for complex software systems. It borrows a lot of its methodologies from lean manufacturing.[5] Even though it was conceived from manufacturing circles, there is no reason that actuaries should not exploit it for their needs! *The DevOps Handbook*[6] is a great book to get up to speed on the topic. To promote speed and reproducibility, automation is at the heart of DevOps, but it is bigger than that.

There are several components to DevOps, such as microservices, continuous testing, continuous integration, continuous delivery, continuous deployment, infrastructure as code, telemetry and continuous feedback, which will be discussed next.

Microservices

With DevOps, the collaboration can happen at such a fast pace because each team works to build microservices. (I will take a little liberty in describing microservices.[7]) For actuaries, a microservice can be thought of as just a single-purpose library. Microservices allow developers to work independently without trampling on each other. They contain application user interfaces (API), which are interfaces that encapsulate the details of the implementation behind a barrier.

The interfaces have contracts, which are called preconditions and postconditions, that describe the output of the services based upon the domain of the inputs. As long as everyone writes codes based upon these contracts, there is no reason to worry about the details of implementation. This greatly speeds up development, because it reduces dependencies among components in the model.

Microservices should be loosely coupled but have a tight cohesion, which means they should be able to communicate with each other, work independently and be singularly focused. The problem with monolithic systems is they have tons of dependencies that lead to tight coupling and loose cohesion of all their components. This leads directly to a system's complexity and the desire to decentralize it.

Continuous Testing

For each unit of work within a microservice, a unit test is made to verify that it operates as expected. These automated tests are small, fast, singularly focused and should run in milliseconds. They should not consume external resources or write to external locations, such as files or databases, so that they run very efficiently. They should be able to run locally on the modeler's local machine or on a server. This allows the developer to continuously run thousands of tests to get immediate feedback and quickly diagnose problems. Running a few sample policies is too slow and too little coverage. Running all policies on the grid doesn't give immediate feedback or good diagnostics on potential issues.

Once the enhancement passes all the unit tests, the changes should go through automated user acceptance testing (UAT). These should also be fast and plentiful, but they are usually larger, less granular tests. They would be designed to test the microservice API and its larger logical units. As stated in test-driven development (TDD), all the unit tests and UATs should be created before a line of code is ever written or modified so that the design of the tests is part of the design of the model. It is only after hundreds or thousands of the very fast automated tests

Monolithic systems naturally lead to Waterfall project management no matter how good the intentions are to go Agile.

have been run that more manual exploratory tests should even be considered.

Continuous Integration, Delivery and Deployment

As many developers are making changes throughout the day, the changes need to be continuously integrated into the master branch. If there are too many changes or the changes are too big, merging them can be time-consuming and difficult and can potentially produce instability. Therefore, each modeling task should be small and singularly focused to provide continuous delivery of new features multiple times per day.

Once the code is delivered, it can go through its last round of reviews and approvals. According to *The DevOps Handbook*, the reviews and approvals should not be delegated away to outside committees. The greater the distance between the committee and where the work is actually performed, the less familiarity there is with the changes and the slower the approval process will be. It is actually recommended to follow extreme-programming practices, which advocate for dual modelers working together on each task. This method has been shown to be quicker and more thorough than a committee approach, because the modelers help each other arrive at a better solution and spot potential issues faster. Once the approvals are passed, then the code can be automatically deployed into production.

INFRASTRUCTURE AS CODE

Infrastructure as code is the concept that all aspects of the model and its configuration are in source control, such as GitHub. This gives the ability for anyone to download the model and all its dependencies and quickly get any deployed version running and reproduce results. If things do go awry, the previous version can be brought back quickly with no manual intervention or setup time. For actuaries, this would include all the work products, such as spreadsheets and other items required to feed the model. This allows any part of the production environment or processes to be reproduced from beginning to end.

Spreadsheets are just ad hoc little programs that are mainly doing calculations and data transformations. They are manual process touch points that are cumbersome, error-prone and a major source of technical debt.[8] It would be much more robust to replace these with more traditional software applications so that the production processes can better follow DevOps principles.

TELEMETRY

Telemetry is monitoring and logging the model by recording data on all mission-critical aspects of its behavior. This allows problems to be addressed quickly with little or no downtime. Items to monitor are run times of all the intermediate processes and distributions of different input variables, crucial intermediate variables and output variables. By keeping the statistics, everyone can receive continuous feedback and learn ways to improve the processes and models. Machine learning and reinforcement learning can be used to monitor logs and detect errors faster, which will speed up response time of dealing with issues.

CONTINUOUS FEEDBACK AND LEARNING

In order for organizations to improve their models and operations, they need to be constantly learning from both their past successes and their past failures. This is not possible without continuously monitoring the health of the models and the supporting processes.

The problem with a monolithic model is that all the pieces have to come together in order to get a functionality to work. It might take days, weeks or months to get all the pieces assembled depending on the size of the enhancement. The feedback on all the issues does not come until late in the development cycle. At this point, the enhancement is promised to senior management, and herculean efforts are required to get it all done. The enhancements are often brittle to boot.

This is why it is important to create tasks that are small and singularly focused—so that the feedback on potential issues comes as early as possible in the development cycle. The later the problems are realized, the more expensive they are to fix. This is why there has been a strong movement of Agile project management over Waterfall so that everyone can get immediate feedback and fix problems sooner. Monolithic systems naturally lead to Waterfall project management no matter how good the intentions are to go Agile.

FUTURE ARTICLES

Now that DevOps has been introduced, the following two articles in the series will get away from theory and get to the practice of implementing DevOps using Moody's Axis. Part 3 will address continuous integration, continuous delivery and infrastructure as code by creating a data-driven dataset that can be generated on the fly. Last, Part 4 will implement DevOps in code using Axis's formula link, formula tables and third-party DevOps tools to showcase all the principles in this article. With these detailed case studies, it will give actuaries the ability to start implementing DevOps in their organizations.

CONCLUSION

Given all the regulatory and accounting changes, such as LDTI and IFRS 17, the actuary has been asked to make a lot of changes in recent years. Even though the intentions of the regulators and accounting standards are to help produce stronger insurance companies and to better track changes, the current practices of actuarial modeling and processes have had a hard time coping with the tidal wave of change. The changes are too fast and the complexity too large for actuaries to brute force their way through them anymore. DevOps is the paradigm shift needed to better cope with change and change management.

The main aspects of DevOps are microservices, continuous testing, continuous integration, continuous delivery, continuous deployment, infrastructure as code, telemetry and continuous feedback. Each one of these concepts plays a crucial role in improving the actuaries' change management capabilities. By following the DevOps best practices, actuaries will be able to create smaller, better, faster and cheaper modeling and valuation departments. The herculean efforts required to get through production cycles and do what-if analysis will be greatly reduced.

Actuarial modernization should be more than moving to a new modeling vendor or software package. The support from our vendors is critical, but modernization is bigger than them. It is about changing how actuaries work and their culture by embracing DevOps and making the practices commonplace. Actuaries are not really modernizing if they are not incorporating DevOps practices in all their work. Replacing spreadsheets should be the first focus of all modernization efforts, because there is so much to gain. Spreadsheets in processes are like cockroaches. There is never just one, and it is expensive and difficult to get rid of the infestation!

With all the potential changes and unknowns on the horizon, it is important that actuaries incorporate DevOps practices sooner rather than later. ■

Bryon Robidoux, FSA, CERA, is an actuary ALM at Reinsurance Group of America. He can be reached at *bryon.robidoux@rgare.com.*

ENDNOTES

1 The monolithic system problem is when all the work products needed for the model are modified and/or stored in the model such that the model will not run or produce accurate results if all the components do not exist.

2 Brooks, F.P. 1995. *The Mythical Man-Month*. Boston, MA: Addison-Wesley.

3 Fowler, M. 1999. *Refactoring: Improving the Design of Existing Code*. Boston, MA, USA: Addison-Wesley.

4 Feathers, M.C. 2005. *Working Effectively With Legacy Code*. Prentice Hall.

5 From Wikipedia, lean manufacturing or lean production is a systematic method originating in the Japanese manufacturing industry for the minimization of waste within a manufacturing system without sacrificing productivity, which can cause problems. *https://en.wikipedia.org/wiki/Lean_manufacturing*

6 Kim, G., J. Willis, J. Humble, and P. Debois. 2016. *The DevOps Handbook: How to Create World-Class Agility, Reliability and Security in Technology Organizations.* 2nd ed. Vol. 2. T Revolution.

7 Library vs microservice. static void, March 7, 2017, *https://blog.staticvoid.co.nz/2017/library_vs_microservice/* (accessed March 12, 2020).

8 From Wikipedia, technical debt is a concept in software development that reflects the implied cost of additional rework caused by choosing an easy (limited) solution now instead of using a better approach that would take longer. *https://en.wikipedia.org/wiki/Technical_debt*