



# Improving Medical Research Data Management

AN INTERACTIVE QUALIFYING PROJECT



Submitted by Vandana Anand, Carlisle Ferguson, Caitlin Kyaw, and Jeffrey Xiao



# WPI



# Universitätsspital Zürich

# Improving Medical Research Data Management

An Interactive Qualifying Project  
submitted to the Faculty of  
WORCESTER POLYTECHNIC INSTITUTE  
in partial fulfillment of the requirements for  
the degree of Bachelor of Science

**Submitted by:**

Vandana Anand  
Carlisle Ferguson  
Caitlin Kyaw  
Jeffrey Xiao

**11 October 2018**

**Report Submitted to:**

**Sponsor:**

Doctor Ivo Dobrev,  
UniversitätsSpital Zürich (USZ)

**Project Advisors:**

John Orr and Dirk Albrecht



**WPI**



**University Hospital  
Zurich**

Zurich Project Center

*This report represents work of WPI undergraduate students submitted to the faculty as evidence of completion of a degree requirement. WPI routinely publishes these reports on its website without editorial or peer review. For more information about the projects program at WPI, see <http://www.wpi.edu/Academics/Projects>*

# Abstract

Medical research is instrumental to the development of new treatments and procedures to remedy human illness. UniversitätsSpital Zürich (USZ), in Zürich, Switzerland, conducts multiple medical research projects concurrently, generating large amounts of data that must be managed well. Currently, there are two main obstacles to good data management at the hospital: the absence of standardized guidelines and researchers' lack of time to develop good practices. To improve data management at the hospital, the team evaluated the current practices in place, identified issues impeding effective organization, and developed solutions. The project resulted in two main deliverables: best practice guidelines for managing medical research data and an application that generates standardized file names.

# Acknowledgments

The team would like to thank the following individuals and organizations for their guidance on and cooperation with our project. Without them, our project would not have been possible.

- Our advisors, Professors John Orr and Dirk Albrecht, for providing guidance throughout the project.
- Our sponsor, Dr. Ivo Dobrev, for providing guidance, feedback, and resources throughout the project.
- The following researchers and staff members at UniversitätsSpital Zürich for providing their expertise and insights on data management systems:
  - Dr. Andreas Eckhard
  - Dr. Alex Huber
  - Dr. Tobias Kleinjung
  - Dr. Nicole Peter-Siegrist
  - Dr. Flurin Pfiffner
  - Dr. Rudolf Probst
  - Dr. Lukas Prochazka
  - Dr. Jae Hoon Sim
  - Dr. Laurent Simon
  - Dr. Dirk Smolinski
- The following individuals for explaining data management procedures at their respective facilities:
  - Dr. Mohammad Divandari, Dr. Raoul Hopf, and Dr. Kathryn Rosowski from Eidgenössische Technische Hochschule
  - Dr. Emine Cagin and Dr. Stefan Bertsch from NTB Buchs
  - Ms. Rita Fagan from the University of Massachusetts Medical School
- The various individuals who provided the team with feedback on the file naming application.
- Worcester Polytechnic Institute, for providing us with the opportunity to complete a social science project abroad.

## List of Abbreviations

CSS	Cascading Style Sheets
csv	comma separated values
DMP	Data Management Plan
ETH	Eidgenössische Technische Hochschule
EU	European Union
IDE	Integrated Development Environment
IQP	Interactive Qualifying Project
IT	Information Technology
JDBC	Java Database Connectivity driver
JVM	Java Virtual Machine
MQP	Major Qualifying Project
NTB	Neu-Technikum Buchs
SNSF	Swiss National Science Foundation
ORL	Otorhinolaryngology
UI	User Interface
UMMS	University of Massachusetts Medical School
USZ	UniversitätsSpital Zürich / University Hospital Zurich
WPI	Worcester Polytechnic Institute

## Authors



### **Vandana Anand**

Vandana Anand is a Computer Science major and a Data Science minor with an interest in cybersecurity and user experience design; she enjoys traveling, baking, and playing badminton.



### **Carlisle Ferguson**

Carlisle Ferguson is a Biomedical Engineering major and a Computer Science minor focusing on biomaterials and tissue engineering; she enjoys pole vault, is involved with the Debate Club on campus, and is a sister of Alpha Gamma Delta.



### **Caitlin Kyaw**

Caitlin Kyaw is a Biomedical Engineering major specializing in biomechanics; she is a sister of Alpha Gamma Delta and enjoys working out, baking, and cooking.



### **Jeffrey Xiao**

Jeffrey Xiao is a Computer Science major with an interest in architecture and user interface design; he is a member of the Cheese Club and the Robotics Club at WPI and enjoys hiking, playing the piano, woodworking, and blacksmithing.

# Authorship

Significant authorship of each section in the report is listed below. Editing was performed by all members of the team.

**Chapter 1:** Carlisle Ferguson, Caitlin Kyaw

**Chapter 2:**

**2.1:**

**2.1.1:** Caitlin Kyaw

**2.1.2:** Vandana Anand, Carlisle Ferguson

**2.2:**

**2.2.1:** Caitlin Kyaw

**2.2.2:** Caitlin Kyaw

**2.3:**

**2.3.1:** Carlisle Ferguson

**2.3.2:** Jeffrey Xiao

**2.3.3:** Vandana Anand, Jeffrey Xiao

**2.3.4:** Carlisle Ferguson

**2.3.5:** Vandana Anand

**2.3.6:** Jeffrey Xiao

**2.3.7:** Vandana Anand

**2.4:** Caitlin Kyaw

**Chapter 3:**

**3.1:** Caitlin Kyaw

**3.2:** Caitlin Kyaw

**3.3:** Caitlin Kyaw

**Chapter 4:**

**4.1:**

**4.1.1:** Carlisle Ferguson, Caitlin Kyaw

**4.1.2:** Carlisle Ferguson, Caitlin Kyaw

**4.2:**

**4.2.1:** Carlisle Ferguson

**4.2.2:** Vandana Anand

**4.2.3:** Vandana Anand

**4.2.4:** Carlisle Ferguson

**4.2.5:** Vandana Anand, Carlisle Ferguson

**4.2.6:** Caitlin Kyaw

**4.3:** Caitlin Kyaw

**4.4:** Caitlin Kyaw

**Chapter 5:**

**5.1:**

**5.1.1:** Caitlin Kyaw

**5.1.2:** Vandana Anand, Carlisle Ferguson, Jeffrey Xiao

**Chapter 6:** Carlisle Ferguson, Caitlin Kyaw

**Appendices:**

**A:** Carlisle Ferguson

**B:** Vandana Anand, Caitlin Kyaw

**C:** Carlisle Ferguson, Jeffrey Xiao

**D:** Carlisle Ferguson, Jeffrey Xiao

**E:** Vandana Anand

**F:** Carlisle Ferguson

**G:** Caitlin Kyaw

**H:** Caitlin Kyaw

**I:** Caitlin Kyaw

**J:** Vandana Anand

**K:** Carlisle Ferguson

**L:** Caitlin Kyaw

**M:** Carlisle Ferguson

**N:** Vandana Anand

**O:** Carlisle Ferguson

**P:** Caitlin Kyaw

**Q:** Vandana Anand

**R:** Caitlin Kyaw

**S:** Jeffrey Xiao

**T:** Vandana Anand, Carlisle Ferguson, Jeffrey Xiao

**U:** Vandana Anand, Carlisle Ferguson, Jeffrey Xiao

**V:** Vandana Anand, Carlisle Ferguson, Jeffrey Xiao

**W:** Carlisle Ferguson, Caitlin Kyaw

**X:** Vandana Anand, Carlisle Ferguson, Jeffrey Xiao

**Y:** Vandana Anand

**Z:** Vandana Anand, Carlisle Ferguson, Caitlin Kyaw

**AA:** Vandana Anand, Carlisle Ferguson, Caitlin Kyaw, Jeffrey Xiao



# Executive Summary

UniversitätsSpital Zürich (USZ) is a teaching hospital located in Zürich, Switzerland. As a top-tier hospital, USZ provides high-quality care to patients using cutting-edge technology (Ranking Web of Hospitals, n.d.). USZ also conducts both clinical and non-clinical medical research projects, which generate hundreds of gigabytes of data per day. The data must be efficiently processed and organized for subsequent analysis and publication. Due to the wide variety of research topics, USZ currently lacks universal guidelines documenting and organizing data. As a result, organization may be inconsistent across projects.

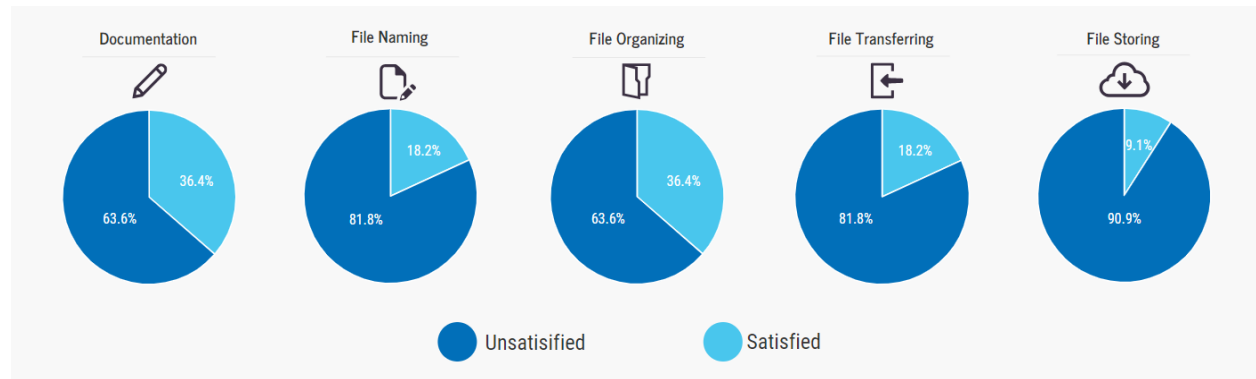
In order to receive funding, many researchers must abide by the guidelines set forth by the Swiss National Science Foundation (SNSF). One of the core aspects of a grant application to the foundation is the Data Management Plan (DMP), which requires researchers to detail how their data will be generated, collected, documented, shared, and preserved (Swiss National Science Foundation, n.d.a). The amount of front-end organization required to receive SNSF funding results in well-organized projects; however, many of USZ's non-clinical research projects are exploratory and will not be submitted for funding consideration.

In exploratory or pilot projects, researchers are uncertain if their idea will fully pan out or if the experiments planned will actually generate the data they need. The goal of an exploratory project is to quickly generate enough data to determine if the project merits further research. The data structure is often unclear at the onset of exploratory projects, and it can take multiple trials of an experiment for a researcher to determine how to organize the data in the best way. As a result, organization can vary widely among researchers and projects. Furthermore, some exploratory projects require researchers to collect large quantities of data over a short period of time. Because of this, researchers may not have enough time to document and organize their data well. This leads to inconsistent data management. Researchers at USZ acknowledge this issue but cannot sacrifice research time to formulate standardized data management protocols.

The goal of this project was to improve data management at USZ. In order to achieve this, the team established three objectives:

1. Evaluate the current research data management procedures at USZ
2. Identify key issues impeding efficient data management
3. Develop solutions for improving data management

To accomplish the first objective, the team interviewed researchers at USZ and other research facilities in order to determine which data management practices worked well and identify any gaps in the current systems. From these interviews, the team identified five key areas of data management: documentation, file naming, file organization, file storing, and file transferring. The interview results revealed that most researchers are unsatisfied with their current data management procedures in all of these areas, as shown by Figure ES1 below. Out of these, researchers were most unsatisfied with file naming, file transferring, and file storing.



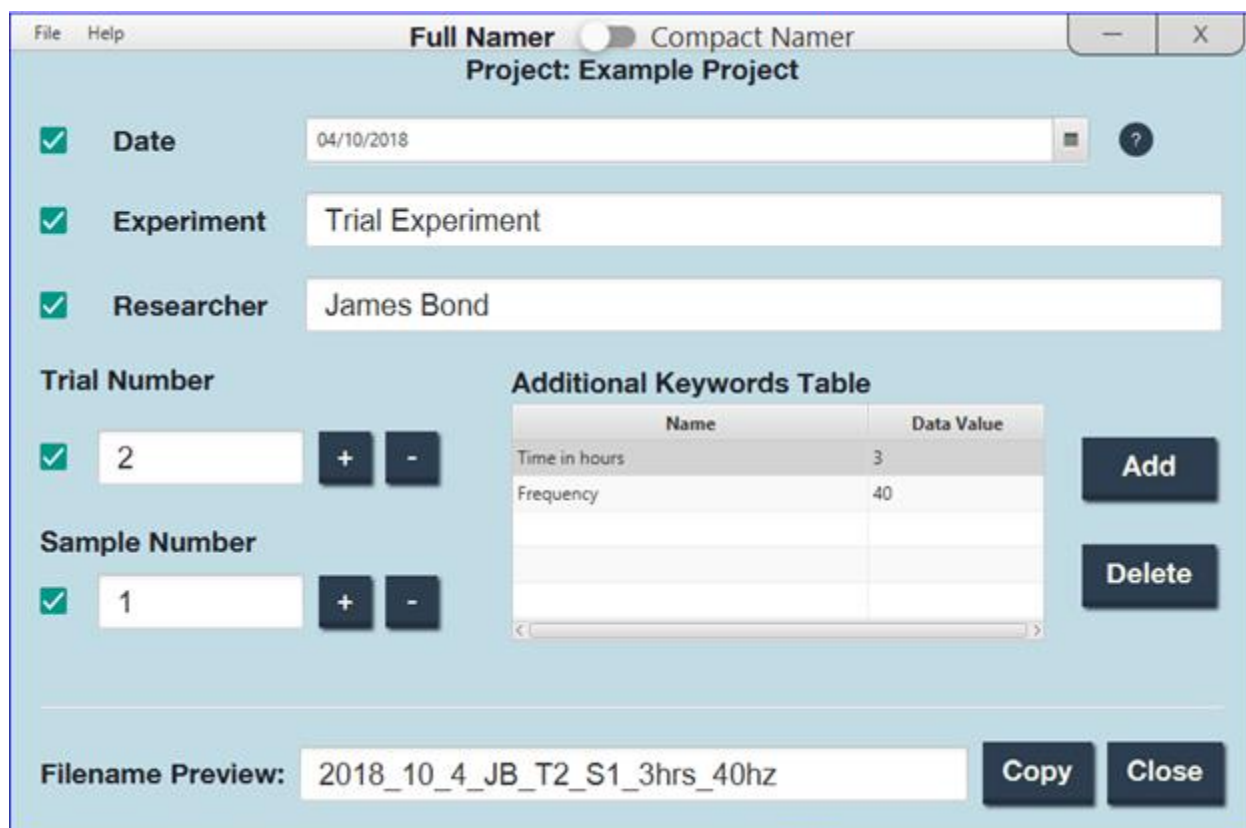
*Figure ES1. Researcher satisfaction rates based on interview results*

In order to best address researchers’ needs, the team developed two main deliverables: a Research Data Management Handbook that provides “good rules of thumb” for documenting, storing, and organizing data; and File Namer Pro, a computer application that generates standardized file names. The Research Data Management Handbook supplies broad recommendations for all five areas while File Namer Pro provides targeted improvements to file naming. The team chose file naming as the area of focus because improvements in file naming will have beneficial spillover effects in file organizing and documentation.

Both the Research Data Management Handbook and File Namer Pro were developed with feedback from researchers. After discussions with researchers about the handbook, the team recognized that researchers may not have the time to read the document in its entirety. As a result, checklists were created for each chapter of the handbook. The checklists have two main benefits: they serve as a condensed summary of the information and can be printed out and tangibly used in researchers’ day-to-day experimentation.

File Namer Pro, shown in Figure ES2 below, was developed via weekly sprints, which are part of the Agile Methodology used in software development. The team had four sprints, and

at the end of each sprint, a completed version of the program was presented for improvements. In order to develop the best product possible during the project term, the team collected feedback from a wide variety of groups, including the project advisors, the project sponsor, researchers at USZ, and software developers. By the end of the fourth sprint, the application included all the features deemed necessary by USZ researchers. The team also developed a user manual for the application that will help both new users and future software developers use and modify the application.



*Figure ES2. Final application*

Both of the team's deliverables aim to make data management easier and less time-consuming for researchers. The Research Data Management Handbook provides a clear and concise summary of good data management practices as well as checklists that can be posted in the lab and used daily. File Namer Pro allows researchers to quickly generate verbose file names which provide easy identification. By utilizing the Research Data Management Handbook and File Namer Pro, USZ will improve their medical research data management.

Furthermore, inefficient data management is not unique to USZ. Many medical research facilities lack standardized procedures for organizing and documenting research data. While the deliverables developed by this project were created in cooperation with researchers at USZ, the deliverables can be applied to and implemented at other research facilities. File Namer Pro and its user manual are available online via GitHub, allowing for easy access and further modification and development. The Research Data Management Handbook and other auxiliary materials developed by this project are available on the WPI eProjects website and are publicly available for interested parties.

Ultimately, this project will not only have positive impacts on USZ but also on any other research facilities who use the deliverables. By saving time during the research process while providing improvements for data organization and management, researchers can conduct more research with increased efficiency, thereby facilitating further advancements in medical science that will help better serve patients.

# Table of Contents

Title Page .....	ii
Abstract .....	ii
Acknowledgments .....	iii
List of Abbreviations .....	iv
Authors .....	v
Authorship .....	vi
Executive Summary .....	viii
Table of Contents .....	xii
Table of Figures .....	xv
Chapter 1: Introduction .....	1
Chapter 2: Background .....	4
2.1 Research Data Handling .....	4
2.1.1 Common Types of Medical Research Data .....	4
2.1.2 Good General Data Management Practices .....	5
2.1.2.1 Handwritten Data .....	6
2.1.2.2 Digital Data .....	6
2.2 Guidelines for Research Data Management .....	7
2.2.1 UniversitätsSpital Zürich .....	7
2.2.2 Swiss National Science Foundation .....	8
2.3 Software Development Tools .....	10
2.3.1 Java and JavaFX .....	10
2.3.2 IntelliJ .....	10
2.3.3 GitHub and Git .....	11
2.3.4 Scene Builder .....	12
2.3.5 JFoenix .....	13
2.3.6 Gradle .....	13
2.3.7 Apache Derby .....	13
2.4 Conclusions .....	14
Chapter 3: Methodology .....	15
3.1 Evaluate Current Research Data Management Procedures .....	15
3.2 Identify and Prioritize Key Issues .....	17
3.3 Develop Solutions for Improving Data Management .....	17

Chapter 4: Results .....	19
4.1 Interview Findings.....	19
4.1.1 UniversitätsSpital Zürich.....	19
4.1.1.1 Networked Computer Drives .....	19
4.1.1.2 Data Organization and File Naming .....	21
4.1.2 Other Facilities .....	24
4.1.2.1 University of Massachusetts Medical School.....	24
4.1.2.2 Worcester Polytechnic Institute .....	25
4.1.2.3 Eidgenössische Technische Hochschule (ETH) Zürich.....	26
4.1.2.4 Interstaatliche Hochschule für Technik NTB Buchs .....	27
4.2 Summary of Common Issues .....	29
4.2.1 Minimal or Inconsistent Documentation Standards .....	29
4.2.2 Lack of Standardized File Naming Procedures .....	29
4.2.3 Inefficient File and Code Structure Organization.....	30
4.2.4 Inadequate Data Transferring Methods .....	30
4.2.5 Inconsistent File Storage Mechanisms .....	31
4.2.6 Lack of Initiative for an Organizational Culture Change .....	32
4.3 Conclusions .....	32
4.4 Selection of Deliverables .....	33
Chapter 5: Implementation and Recommendations.....	35
5.1 Creation of Deliverables.....	35
5.1.1 Best Practice Guidelines.....	35
5.1.1.1 Chapter 2: Documenting Data .....	37
5.1.1.2 Chapter 3: Naming Data Files.....	38
5.1.1.3 Chapter 4: Organizing Data File Folders .....	39
5.1.1.4 Chapter 5: Storing and Transferring Data.....	39
5.1.1.5 Further Information.....	40
5.1.2 File Naming Application .....	40
5.1.2.1 Gathering Requirements .....	41
5.1.2.2 Software Setup and Initial Application Design .....	43
5.1.2.3 Bug Fixing and Improvements.....	45
5.1.2.4 Survey Results and Final Application Improvements .....	46
5.1.2.5 Final Iteration of the Application.....	47
5.1.2.6 Further Development .....	52

Chapter 6: Conclusion.....	53
References.....	54
Appendix A: Pre-Interview Team Introduction.....	56
Appendix B: Sponsor Description .....	57
Appendix C: Interview Template, Spring 2018.....	60
Appendix D: Interview Template, Fall 2018 .....	61
Appendix E: Dr. Dirk Smolinski Interview Summary.....	62
Appendix F: Rita Fagan Interview Summary .....	63
Appendix G: Prof. Dr. Rudolf Probst Interview Summary .....	65
Appendix H: Dr. Flurin Pfiffner Interview #1 Summary .....	66
Appendix I: Prof. Dr. Raoul Hopf Interview Summary.....	68
Appendix J: Prof. Dirk Albrecht Interview Summary .....	69
Appendix K: Dr. Andreas Eckhard Summary .....	71
Appendix L: Dr. Flurin Pfiffner Interview #2 Summary .....	73
Appendix M: Dr. Laurent Simon Interview Summary .....	75
Appendix N: Dr. Nicole Peter-Siegrist Interview Summary .....	77
Appendix O: Dr. Emine Cagin Interview Summary.....	79
Appendix P: Dr. Mohammad Divandari Interview Summary .....	81
Appendix Q: Dr. Prof. Thomas Kleinjung Interview Summary .....	83
Appendix R: Prof. Dr. Stefan Bertsch Interview Summary.....	85
Appendix S: Iteration One Product Backlog .....	87
Appendix T: Iteration Two Product Backlog.....	88
Appendix U: Iteration Three Product Backlog .....	90
Appendix V: Iteration Four Product Backlog.....	92
Appendix W: Application Feedback Survey .....	94
Appendix X: Code Authorship .....	101
Appendix Y: Code Location and Modification Steps.....	104
Appendix Z: Research Data Management Handbook .....	105
Appendix AA: User Manual for File Namer Pro.....	128

## Table of Figures

Figure 1. University Hospital Zurich grounds (University Hospital Zurich 2014) .....	1
Figure 2. USZ's file organization template .....	7
Figure 3. GitHub Project Board .....	11
Figure 4. An unmodified button in Scene Builder .....	12
Figure 5. A user interface in Scene Builder .....	12
Figure 6. Process workflow for improving research data management.....	15
Figure 7. File transferring practices .....	20
Figure 8. A ledger created by Ivo Dobrev to track data.....	21
Figure 9. Research workflow .....	23
Figure 10. Researcher satisfaction rates based on interview results.....	32
Figure 11. Process workflow for developing the best practice guidelines .....	35
Figure 12. A preliminary draft of the user interface .....	42
Figure 13. Process workflow for file naming application sprints.....	43
Figure 14. Version One.....	44
Figure 15. Excerpt of a product backlog.....	46
Figure 16. Expanded 'File' menu.....	47
Figure 17. Final Full Namer user interface.....	48
Figure 18. Autocomplete text fields in Full Namer .....	48
Figure 19. Add New Keyword to Database user interface .....	49
Figure 20. Compact Namer user interface .....	49
Figure 21. Log File user interface.....	50
Figure 22. Example Excel log file .....	50
Figure 23. Project Preferences user interface .....	51



## Chapter 1: Introduction

Medical research is instrumental to the discovery and development of new treatments and procedures to remedy human illness. Researchers produce a large amount of data through experiments and projects, which can be crucial for the eventual publication of the research and future development of new treatments and procedures.

UniversitätsSpital Zürich (USZ, Figure 1), a teaching hospital in Zürich, Switzerland, actively participates in medical research and clinical trials. The hospital also pioneers many of the latest treatment methods, seeking to provide therapies to patients from around the world with various complex health problems.



*Figure 1. University Hospital Zurich grounds (University Hospital Zurich 2014)*

Additionally, medical professionals at UniversitätsSpital Zürich collaborate with scientists and engineers who research and experiment with diagnostic and therapeutic approaches in order to improve areas such as patient lifespan and quality of life. The hospital also engages in other research fields, such as regenerative medicine, personalized medicine, highly specialized medicine, nursing care innovation, and health services innovation. For example, one research group is developing an inner-ear probe with the hope of developing a fully implantable cochlear

implant; another is hoping to non-invasively measure tumor oxygenation using near-infrared optical tomography.

As technology and medical research continue to evolve, new forms of measurement and experimentation develop. While the process of analyzing research data has greatly improved due to the growth of commercially-available computing power, the volume of data generated per experiment has grown enormously over time. Data types can range from digital files to verbal discussions to handwritten notes, and hundreds of gigabytes of digital data or multiple full lab notebooks can be generated by one project alone. The sheer size of the generated data and the wide variety of types that can be produced through experimentation makes it difficult for researchers to organize and document their data in an effective and efficient manner. Each data type is processed differently, and because the various research groups at the hospital run countless experiments and research projects concurrently, poor documentation and organization of the generated files and data are common. This often impedes efficient processing, understanding, and eventual publication of the research.

Researchers acknowledge this problem but cannot sacrifice research time to formulate a standardized system for organizing and documenting data. The Swiss National Science Foundation (SNSF) funds some of USZ's research projects, and these projects are held to a higher degree of organization due to the foundation's strict guidelines. Exploratory or pilot experiments tend to be less organized because researchers must quickly generate enough data to determine if a project merits further exploration. Research data documentation and organization is often inconsistent due to the rapid nature of the project, and as a result, researchers spend significant amounts of time reviewing and searching for old data files prior to publishing their findings.

The overall goal of this project is to improve research data management at UniversitätsSpital Zürich. The team accomplished the following objectives to complete the project:

- Evaluated the current research data management procedures at USZ
- Identified key issues impeding efficient data management
- Developed solutions for improving data management

In order to achieve the overall goal, the team interviewed members of the research staff at USZ to determine the scope and usefulness of pre-existing data management systems. The team

investigated and identified the most pressing issues and gaps in said systems, which were then ranked by priority based on feedback from researchers at USZ and other facilities. This work culminated in the creation of best practice guidelines for managing medical research data, a document that details how to properly annotate, document, store, and organize research data. Additionally, the team developed a computer application that generates standardized file names. These deliverables will help researchers UniversitätsSpital Zürich process and analyze their research data more efficiently and effectively, allowing them to continue pioneering the latest treatment methods.

## Chapter 2: Background

This chapter begins by discussing common types of medical research data and outlines good management practices for data in general. Next, current guidelines for managing research data in place at UniversitätsSpital Zürich are reviewed. Lastly, specific tools used for software development are described.

### 2.1 Research Data Handling

#### 2.1.1 Common Types of Medical Research Data

Medical research aims to investigate the causes of human illnesses and discover new treatments and procedures to remedy them. The types of data generated by medical research can fall under one or more of several categories:

- narrative data
- numerical measurements
- recorded signals
- pictures

Narrative data describes qualitative data that relies on “written or spoken words or visual representation of individuals” (“Narrative Research”, n.d.). This type of data describes narratives, stories, and textual analyses of research questions. As it pertains to medical research, it can often convey the history of present illness, social or family history, or failure to thrive (Szolovits, n.d.). Additionally, this data can be based on information derived from verbal or written patient complaints or a recounting of family medical history. Other types of narrative data include voice memos and oral notes.

Numerical measurements for use in medical research can include lab results, vital signs, and physical measurements. This type of data can be obtained from living and non-living subjects and is collected using instruments. For example, the ORL (Otorhinolaryngology) research team takes numerical measurements of promontory and intracranial pressure using hydrophones and the laser Doppler velocimetry technique. This type of quantitative data is measured objectively.

Recorded signal data, similarly to numerical measurements, is quantitative data obtained using instruments. These can include electrocardiogram tracings, sound files, and spectrograms

(Szolovits, n.d.). Recorded signals can translate into numerical data for numerical and statistical data analysis.

Image data is not limited to still images and can include radiologic pictures, sketches, x-rays, and microscope video and images. In some cases, this data can also be translated into numerical data for statistical analysis. Additionally, image data can indicate how certain experiments were set up and can be useful for any subsequent replication of the results.

Different experiments and projects will necessitate different types of data depending on the research question, but all data types are vital to the processing, analysis, and any future publication of the research.

### 2.1.2 Good General Data Management Practices

Several key characteristics remain constant across all good data management practices. One of these characteristics is thorough documentation. Clear and concise documentation allows researchers to quickly determine who did what, what was done, when it took place, what the results were, and what observations can be drawn from data.

Other significant characteristics are effective data storage and organization. When an experiment or research question is established, one should first consider what type of data is needed to fulfill the experiment or answer the research question and examine how the data should be stored and organized. Furthermore, all sources, files, and lab notebooks should be located in one easily accessible location so they can be retrieved for future analysis and interpretation (Inouye and Scheiner, 2009). The proper storage and organization of such data is often delayed until towards the end of the experiment or even viewed as an unnecessary task that detracts from scientists' valuable research time. However, without efficient data management, valuable information is often lost (Inouye and Scheiner, 2009). Ultimately, the time saved disregarding data management while an experiment is conducted pales in comparison to the time wasted trying to find certain data files later. During the publication phase, researchers must revisit their old data, and if the information is poorly organized, they lose valuable time attempting to decipher which file or lab notebook contains the data they are looking for.

Proper data management practices will save researchers valuable time in the long term, help improve data accessibility, aid the continuation of previous projects, and establish precedence if similar research is subsequently conducted or published (Coulehan and Wells, n.d.).

### *2.1.2.1 Handwritten Data*

In a paper-based system, well-documented and meticulously maintained lab notebooks can provide an excellent organizational system in the form of a comprehensive log of a researcher's activities. In theory, one researcher should be able to open another's notebook twenty years later, easily recreate the experiment, and understand where the previous researcher left off; however, this is often challenging because different researchers tend to have different standards and methods for note-taking. Searching for a specific fact or trying to compare observations from several sources can be difficult. Furthermore, researchers may tape graphs or images into their lab notebook, and these can easily be lost. Additionally, maintaining handwritten records is almost impossible for larger projects that result in significant amounts of data, such as clinical trials or epidemiological surveys (Coulehan and Wells, n.d.).

### *2.1.2.2 Digital Data*

Good data management practices also extend to electronic records. Digital data allows researchers to efficiently access and compare information from different sources and projects. Today, electronic data capture programs allow researchers to enter, store, and examine research data. Recently, cloud storage platforms have increased in popularity. These platforms allow users to access nearly unlimited amounts of storage and share files from any device that has an internet connection (Network Computing, 2016). Electronic data storage is steadily becoming the preferred method of data storage, but a large amount of data still remains handwritten.

Electronic data storage addresses several of the key issues that result from handwritten documents. First, it supplies an opportunity to provide thorough and legible documentation, allowing for improved comprehensibility of the data for future use. Additionally, an electronic storage format can be updated and adapted in parallel with evolving computer hardware and software. Furthermore, this storage mechanism provides quick access to the data, rapid reading and writing rates, the ability to archive the data, and a backup system, such as storing data on other devices such as compact discs or flash drives. While this storage mechanism requires additional protection and safeguards to prevent unauthorized access to sensitive medical data (Coulehan and Wells, n.d.), the benefits of digital data storage are numerous when files are well-documented and organized.

## 2.2 Guidelines for Research Data Management

### 2.2.1 UniversitätsSpital Zürich

As a teaching hospital, UniversitätsSpital Zürich conducts research in several different areas. Their research falls under two broad categories: clinical and non-clinical research. Clinical research investigates new ways of diagnosing, treating, and preventing disease. Because human patients are involved in these projects, the data generated from any tests, measurements, or experiments is held to a higher standard of security and confidentiality. The hospital has a plethora of guidelines that require researchers to state how data will be anonymized and labeled and describe what ethical considerations must be considered. Researchers must also provide documentation stating who owns the data, who will be viewing and using the data, and who holds the key for anonymizing the data. These guidelines for the required documentation, however, can be characterized as general and vague, leaving clinical researchers to decide how best to convey this information in order to meet the requirements.

For non-clinical research, hospital-wide guidelines for organizing and documenting research data are virtually non-existent. In the past, the hospital tried to implement a folder organization template, shown in Figure 2 below, but this did not gain traction. This template simply provided categories in which data and files could be stored, but did not provide information on how to organize subdirectories, name files, or take notes. Because many researchers found this template was irrelevant to or useless for their projects, it went widely unused.

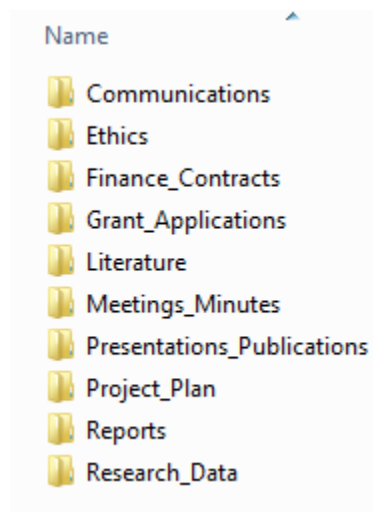


Figure 2. USZ's file organization template

The lack of detailed guidelines for organizing and documenting non-clinical data is one of the many obstacles hindering the efficient analysis of data. Due to the variability among research projects and groups, it would be difficult to reach a hospital-wide consensus on what the best methods for organizing and documenting data are; however, having basic guidelines has its benefits.

### 2.2.2 Swiss National Science Foundation

USZ and other Swiss research centers and laboratories currently follow guidelines on the management of research data established by the Swiss National Science Foundation (SNSF) in order to receive funding. Because good scientific practices necessitate the proper management and sharing of research data, the SNSF requires that its funded researchers submit a Data Management Plan (DMP).

The DMP helps researchers plan how their project data will be generated, collected, documented, shared, and preserved. It is also an integral part of the grant proposal (Swiss National Science Foundation, n.d.a). The SNSF has provided researchers with general DMP templates; however, the content of each DMP may differ from project-to-project depending on the specific standards of other applicable procedures. The DMP consists of four sections: how data will be collected and documented, issues concerning matters such as ethics and security, the plan for the storage and preservation of the data and its documentation, and the plan for any future sharing and reusing of the data.

The DMP must also comply with the SNSF's Open Research Data Policy, which agrees with the principle that "research data should be freely accessible to everyone – for scientists as well as for the general public" (Swiss National Science Foundation, n.d.b). The Open Research Data Policy also states that its funded researchers are expected to store and preserve any relevant research data that is produced, share the data unless bound by ethical, copyright, or other restrictions, and deposit their data on platforms in such a way that is accessible and re-usable by anyone with minimal restrictions. When submitted, the DMP must be a clear and accurate representation of the standards set by the SNSF and suit their project needs. Because USZ and other research facilities throughout Switzerland are often solely funded by the SNSF, the DMP is a key part of the grant proposal. It is essential that the DMP is well-thought-out and properly completed.



Additionally, the SNSF expects that research data will be collected and stored in accordance with the FAIR Data Principles, which “define a range of qualities a published dataset should have in order to be Findable, Accessible, Interoperable and Reusable” (Swiss National Science Foundation, n.d. a).

In order to qualify as findable, data sets must be “assigned by a globally unique and persistent identifier” (Wilkinson et al., 2016), and researchers must thoroughly describe the conditions under which the data was generated, provide directions for its reuse, and supply the proper context for future interpretation or replication. A data set’s accessibility is determined by its ease of retrievability through the unique identifier via a communications protocol. The protocol must be “open, free, and universally implementable” (Wilkinson et al., 2016) and allow for anyone with access to a computer and an internet connection to at least access the metadata (data that describes and gives information about other data). Interoperability describes the ability of computer systems or software to exchange and make use of information. In the case of medical data, interoperability requires that the data is in a format that is recognizable by different computers, uses a controlled vocabulary, and provides links between the data and any additional data sets that are needed. The last FAIR Data Principle, reusability, is determined by plurality, accuracy, and relevancy. Plurality, in this case, is the data’s ability to be described with a range of relevant attributes. Under this principle, it is also important that the “conditions under which the data can be used” (Wilkinson et al., 2016) are clear and that any relevant data is organized in a standardized fashion.

These four foundational principles for the management of research data are key conduits to the progression of innovation and outline the tenets of proper data management. The drawback with the SNSF guidelines is that each facility must determine how to implement them for itself. Furthermore, the guidelines can be characterized as vague and are applicable to a broad range of research experiments. Additionally, the guidelines can be interpreted in different ways and do not provide step-by-step procedures on how to handle data. Since researchers are often unable to sacrifice valuable research time to develop standardized guidelines, documentation can, at times, resemble a conglomeration of inconsistent notation. Research facilities have worked to dedicate specific individuals to improve data management, but the vagueness of the foundation’s requirements remains a persistent issue.

## 2.3 Software Development Tools

The following section describes the various tools the team used in order to develop a file naming application. Additional details about the application, File Namer Pro, can be found in subsequent chapters.

### 2.3.1 Java and JavaFX

Java is an object-oriented computer programming language. Programming in Java has several benefits, one of which is portability. Java code compiles inside a Java Virtual Machine (JVM), which can be installed on any computer regardless of the operating system. This means the code runs equally well on a computer with Windows as it would on a computer with Mac OS or Linux. This is highly beneficial for development because team members can use their preferred operating system, and for application use, since the application can function on different platforms. The second benefit of Java is the presence of a large web community of programmers. This online community provides an abundance of resources, ranging from clear documentation to guides for debugging common issues. Furthermore, Java is relatively easy to learn and understand, allowing beginner programmers to easily start making their own programs.

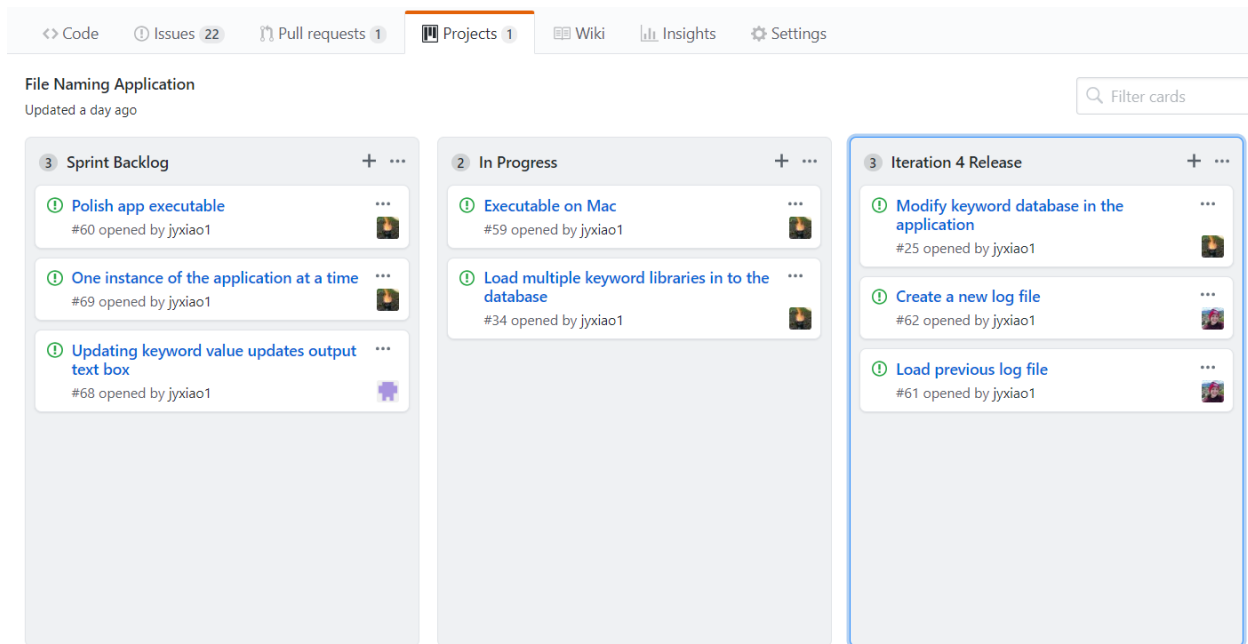
JavaFX is an extension of Java that serves as a software platform and allows users to create desktop applications. It has several key features, including Scene Builder, which is discussed in further depth in Section 2.3.2. Additionally, JavaFX has built-in user interface (UI) controls and CSS (Cascading Style Sheets). These allow the programmer to customize styles such as fonts and clickable buttons in the layout of the program. The JavaFX software, particularly Scene Builder, allows programmers to quickly and visually build UIs.

### 2.3.2 IntelliJ

IntelliJ is one of the most widely-used Java Integrated Development Environments, or IDEs. IDEs are primary platforms for writing code and include a source code editor, a debugger, and a build automation tool. IntelliJ also has a built-in compiler for compiling and running the code and can support a wide variety of third-party software. It is also compatible with Git and GitHub, which are explained in detail in the following section.

### 2.3.3 GitHub and Git

GitHub, an online repository platform used to store code online, offers native integration with IDEs such as IntelliJ. GitHub allows users to host, distribute, review, and collaborate on code. Any developer with sufficient permissions can collaborate by downloading a new version of the program, modifying it, then uploading the newest version. A user can create multiple repositories to store different projects that are accessible via a unique URL. Developers can further advance a project by forking a repository, allowing them to make changes without affecting the original project. The pull request feature enables the original author or project manager to accept or deny changes made to the program by other GitHub members. In a team setting, GitHub offers a project board feature where all tasks are placed in columns such as “Backlog”, “In Progress”, and “Release” (Figure 3). These tasks can be categorized and assigned to any project collaborator.

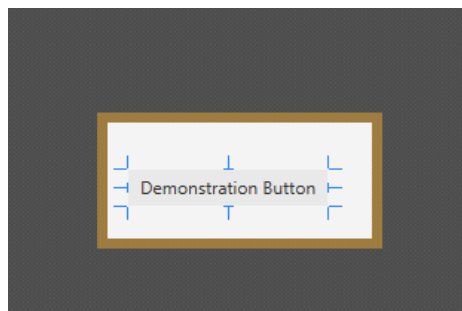


*Figure 3. GitHub Project Board*

GitHub uses Git to maintain and make changes to repositories. Git, a generic version-control system, allows multiple developers to contribute and combine different versions of code (Kanaparthi, 2018). It can be used to stay up-to-date on code changes from other developers, revert to certain stages in production, and keep track of who did certain work. These features make Git and GitHub an ideal location to host a software application.

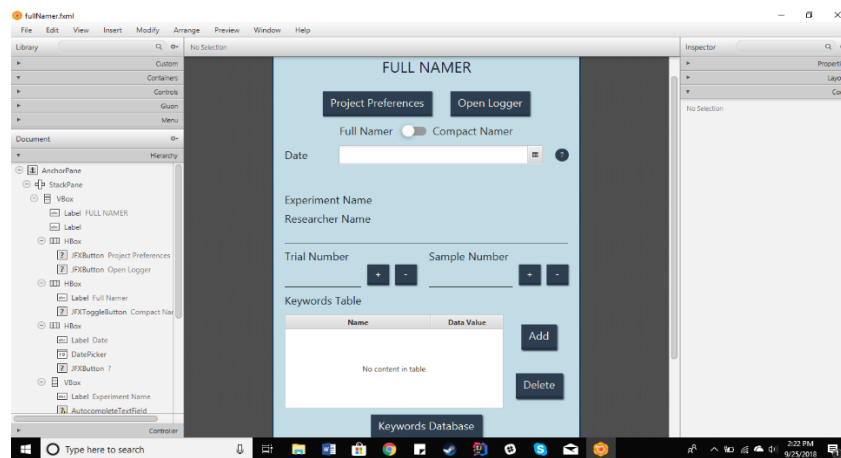
### 2.3.4 Scene Builder

Scene Builder, a tool within JavaFX, is used to create user interfaces. Each Scene Builder file has a corresponding FXML file in the project folder. FXML is an XML-based markup language that defines the user interface in JavaFX applications. The code in the FXML file updates automatically as the programmer creates the UI in Scene Builder. Scene Builder also has a plethora of premade UI elements, such as buttons, toggles, and text fields. These UI elements start off simplistically (see Figure 4) but are easily modifiable inside Scene Builder by changing the background color, setting an outline color, etcetera.



*Figure 4. An unmodified button in Scene Builder*

Moreover, the drag-and-drop interface in Scene Builder is easy to use. Programmers can also use style sheets, which are represented in JavaFX as CSS files. The CSS file contains the specific information about the visual appearance of the UI element, ranging from background color to font. The CSS files allow the programmer to store preferences, and these style sheets can be loaded for each button, saving time, effort, and improving visual appearances. A more developed UI is shown in Figure 5 below.



*Figure 5. A user interface in Scene Builder*

### 2.3.5 JFoenix

JFoenix is an open-source Java library containing pre-built user interface elements, adapted from Google Material Design (JFoenix, n.d). Users can download JFoenix from GitHub or import it from JavaFX Scene Builder. Open-source Java libraries such as JFoenix are particularly helpful when building web applications from scratch because users can easily drag-and-drop custom-made components to build a graphical user interface (GUI) instead of creating the components by hand. Some Java components included in JFoenix include different types of clickable buttons, menu icons, and input fields.

### 2.3.6 Gradle

Gradle, a build automation system designed for multi-project support, is used to generate various scripts. One of its most significant uses in a Java project is to support the automatic download and configuration of libraries. Libraries are sets of behavior implementations used by different, unrelated programs. This is especially useful when sharing repositories with large amounts of external dependencies. Rather than having to download various libraries individually, the Gradle build will automatically download any needed libraries. There are also various plugins available for Gradle that perform various tasks ranging from cleaning up and formatting code to generating JAR (Java ARchive) files for creating a stand-alone executable for the software program.

### 2.3.7 Apache Derby

Apache Derby is a relational database management system written in Java. Derby allows users to store relevant data and details in databases on the back-end of an application (Apache Derby, n.d.). It also has several key advantages. Firstly, Derby can easily be embedded in any Java-based solution due to its small file size. Secondly, it includes a Java Database Connectivity driver, a software component that allows Java applications to interact with databases (Apache Derby, n.d.). Thirdly, Derby is based on Java, JDBC, and SQL (Structured Query Language), a programming language commonly used to manage data within a database. This allows users familiar with Java, JDBC, or SQL to easily get started with Derby.

## 2.4 Conclusions

The aforementioned information provided the team with both the context of the problem of inefficient data management and the technical knowledge needed to mitigate it. The lack of clear and concise guidelines available for managing medical research data impedes effective and efficient data analysis at the hospital. Although there are no simple, universal solutions for these problems, having guidelines and tips for ameliorating even some of the data management issues that researchers at UniversitätsSpital Zürich face will be particularly useful. This project aims to provide such guidelines and recommendations, as well as a targeted method to standardize file names and improve data management and organization.

## Chapter 3: Methodology

The overall goal of this project was to improve research data management at UniversitätsSpital Zürich. The team established and accomplished the following three objectives in order to complete the project:

- Evaluate current research data management procedures at USZ
- Identify key issues impeding efficient data management
- Develop solutions for improving data management

These objectives, as well as the tasks completed to accomplish them, are described in further detail in subsequent sections of this chapter. Figure 6 below illustrates the process the team followed in order to achieve each of the objectives.

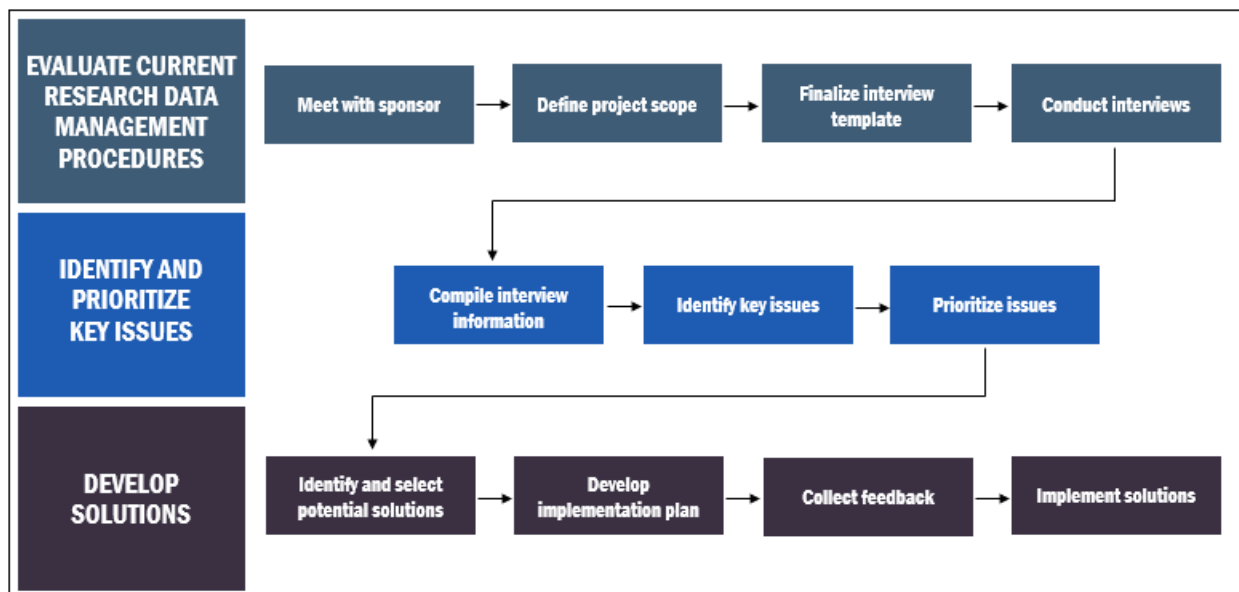


Figure 6. Process workflow for improving research data management

### 3.1 Evaluate Current Research Data Management Procedures

In order to begin evaluating USZ's current process for organizing and documenting research data, the team visited the hospital and research labs in-person. Within the first week of the team's arrival, the project sponsor, Doctor Ivo Dobrev, presented a sample of what his experimental process looked like, identified several instances of poor documentation and organization, and explained why these shortcomings could eventually lead to a loss of

information or waste valuable time. This allowed the team to properly visualize and understand many of the issues currently impeding researchers' ability to efficiently process and analyze their experiments.

Next, the team defined the scope of the project using the information gleaned from the initial meeting. The most efficient method of documenting and organizing data varies from project-to-project and even from person-to-person. Because of this, the team recognized it would not be feasible to try to implement highly specific, universally-applicable solutions for every project and area of research.

In order to further understand current data management practices in place at USZ and identify specific issues and problems impeding effective data analysis, the team chose to conduct interviews with several researchers from the hospital. Appendix A, the pre-interview team introduction, was written to introduce the team to interviewees and provide context for the project. Additionally, an interview template (Appendix D) was created that would guide the conversation towards a discussion about:

- What, if any, documentation and organization standards are used throughout the course of a project
- If any of the standards are uniform across projects and departments
- What kind of information is used to label and differentiate between data files generated during a project
- What the general workflow tends to be in terms of gathering data, organizing data, and preparing the data for review or publication

Using the team introduction and interview template, the team met with researchers to discuss their various data management procedures and asked for their opinions on areas they felt needed improvement. The team transcribed and documented these interviews and later followed up with additional questions as needed. Additionally, interviews were conducted concurrently with researchers and staff at other facilities, such as Eidgenössische Technische Hochschule (ETH) in Zürich and NTB Interstaatliche Hochschule für Technik Buchs in Buchs, to gain a better understanding of other examples of standards for managing research data and any other obstacles and issues that researchers face.



These interviews and visits allowed the team to collect information about current data management practices at UniversitätsSpital Zürich and other facilities and identify any shortcomings of their current processes.

## 3.2 Identify and Prioritize Key Issues

The team compiled and consolidated the information gathered from interviews that occurred during the first two weeks of the project. Follow-up interviews were conducted to obtain further information and address any confusion the team had. Summaries of all of the interviews can be found in the Appendices. Upon review of the information, the team developed a list of issues and obstacles that hinder the efficient analysis of data. In week three, the team presented this list to the project sponsor and other members of the ORL research team. Following this presentation, researchers were asked to identify key areas of data management they felt were top-priority areas needing improvement. Additionally, the team asked researchers to indicate whether they were satisfied or unsatisfied with current practices in each area. These results were then quantified and can be found in Chapter 4.

## 3.3 Develop Solutions for Improving Data Management

After identifying and prioritizing key issues impeding effective data analysis, the team brainstormed several solutions that could be developed in order to improve data management at UniversitätsSpital Zürich. During further discussions with the project sponsor, the team narrowed down a selection of potential project deliverables that would be both particularly impactful to the hospital and feasibly produced in the short amount of time available. Ultimately, the team chose to produce two deliverables: a set of best practice guidelines that would provide general recommendations for several aspects of data management and a computer application that would remedy the issue of inconsistent file names. Chapter 4 further discusses the motivation behind this decision.

Following this decision, the team set about developing a plan of implementation for these deliverables. The team's goal in selecting the deliverables was to provide solutions and tools that would actually be of use and practical for the researchers. In order to confirm that the deliverables would be helpful to the project sponsor, the concepts of the best practice guidelines and the file naming application were presented to Dr. Dobrev and other researchers, who then

provided their feedback on what they would like to see from the deliverables. Once the feedback and additional comments were collected and compiled, the team created the guidelines and the application. Information about the process the team followed in order to develop the deliverables is described in Chapter 5.

## Chapter 4: Results

As explained in Chapter 3, the team conducted interviews with various researchers in order to glean information about current data management practices. The information was taken into consideration during the creation of the final deliverables. This chapter discusses the findings from the interviews, summarizes the common issues mentioned during interviews, and describes the motivation behind the selection of the deliverables.

### 4.1 Interview Findings

#### 4.1.1 UniversitätsSpital Zürich

From interviews conducted at UniversitätsSpital Zürich, the team gained a deeper understanding of current data management practices in place at the facility and collected feedback about certain aspects of current practices that researchers felt needed improvement. This section briefly summarizes the various data naming, storing, and organizing practices that were described by interviewees, as well as any areas the interviewees felt could be improved. Further details from the interviews can be found in the Appendices.

##### *4.1.1.1 Networked Computer Drives*

Dr. Ivo Dobrev, the project sponsor, explained the new network drive system that was implemented at the hospital in February 2018. In this system, each research group has a folder that is accessible by all group members. This folder includes content such as administrative materials, manuals, and non-project-specific information. Other project-specific folders are accessible only to researchers working on the project, and access is granted on a person-by-person basis by the IT department. Furthermore, access is limited to certain computers. This system was implemented in order to prevent accidental data transfer, whether it be saving files to the wrong project folder or removing files from folders that need them. Unfortunately, the network drive is not always intuitive; folder headings are often generalizations or vague. Furthermore, there is no easy way to transfer data to collaborators outside of USZ. It is often common for researchers to have to put their information on a memory stick that must be physically plugged into a collaborator's computer.

Dr. Laurent Simon, a postdoctoral researcher in the experimental audiology lab at USZ, also provided his insight into the network drive system. Dr. Simon explained that the software

required for conducting tests is often unavailable on or missing from the researchers' computer, and lab computers that are equipped with the necessary software do not have access to the needed project folders. In order to circumvent this issue, Dr. Simon manually transfers files from the networked drive to his computer drive. Unfortunately, this creates another problem. Dr. Simon explained that when he transfers files, he usually does so by duplicating the file and moving the duplicate to another location, leaving the first copy in its original location. For example, if he needs a program from one computer, he will transfer it onto his computer drive via a USB, leaving a copy of the program on the original computer. He may make changes to the program file on his computer; however, those changes are not reflected in the original file. This practice results in multiple versions of files in multiple locations. The problem with this, Dr. Simon explained, is that he and his colleagues often have trouble determining which version is the most accurate or most recently updated. File explorers do indicate the date of last modification, but in some cases, the "modification" could simply be the date on which the file was last accessed. Having the ability to maintain and update one version from any authorized computer or by any authorized user would reduce the chance of confusion. Figure 7 visualizes this issue.

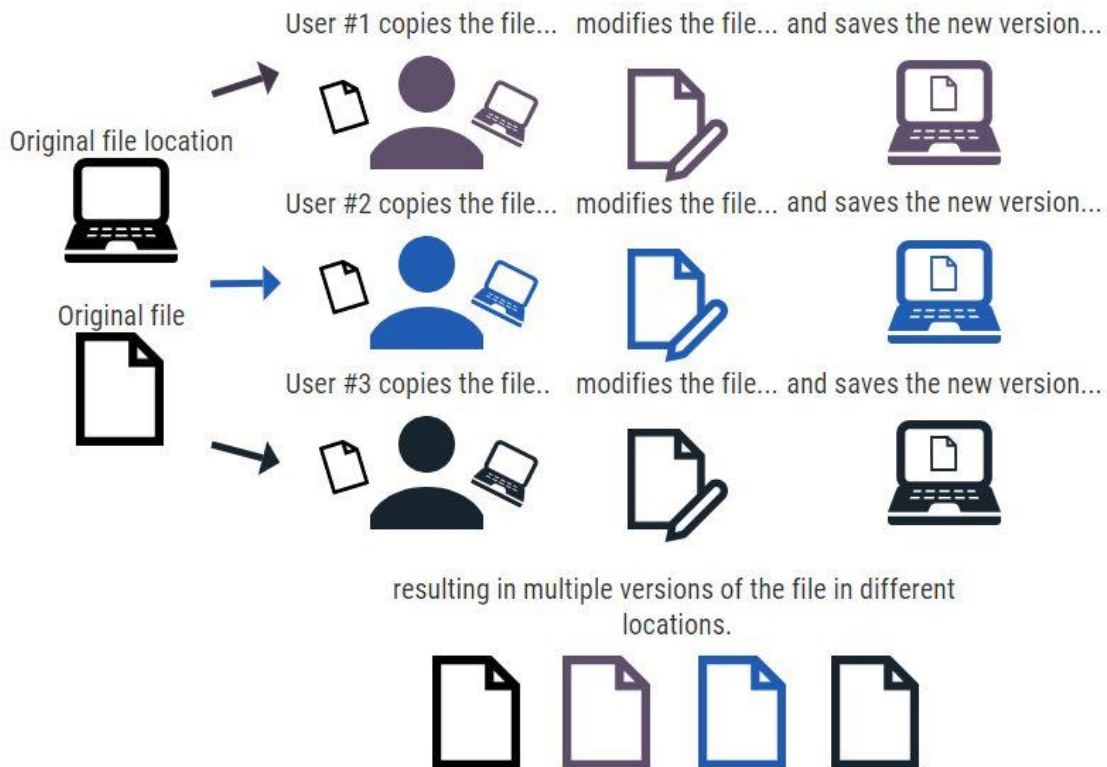


Figure 7. File transferring practices

#### 4.1.1.2 Data Organization and File Naming

Although USZ currently lacks a standardized naming and organization system, Dr. Dobrev has developed a system for his own use. He explained that projects conducted internally at USZ are typically exploratory and generally lack outside funding. During exploratory projects, researchers tend to pursue findings as quickly as possible in hopes of making enough progress to apply for a grant. Because of this, data for these exploratory projects can often be characterized as disorganized. In his file system, Dr. Dobrev labels his data files in terms of which experiments they pertain to and the date; for example, a file could be named ‘Test 1 01\_02\_2018’. Once the structure of the experiment becomes clearer, typically after seven or eight experiments, Dr. Dobrev creates a ledger in Microsoft Excel (Figure 8) that provides more descriptive information about each planned test.

Aim	Water level	Stimulator	Attach. type	Attach. Loc.	Scan area	Group	Meas #	CH1	CH2	CH3	CH4	CH5
2a	Full	Baha C2	5N steel band	Ipsi mastoid	Ipsi	1A	1	07.05.18	13.06.18			
2a	Full	Baha C2	5N steel band	Ipsi mastoid	Top		2					
2a	Full	Baha C2	5N steel band	Ipsi mastoid	Contra		3					
2a	Full	Baha C2	5N steel band	Forehead	Ipsi	2A	4	08.05.18	13.06.18			
2a	Full	Baha C2	5N steel band	Forehead	Top		5					
2a	Full	Baha C2	5N steel band	Forehead	Contra		6					
2a	Full	Baha C2	5N steel band	Eye	Ipsi	3A	7	08.05.18				
2a	Full	Baha C2	5N steel band	Eye	Top		8					
2a	Full	Baha C2	5N steel band	Eye	Contra		9					
2a	Full	Baha C2	5N steel band	Neck	Ipsi	4A	10	09.05.18				
2a	Full	Baha C2	5N steel band	Neck	Top		11					
2a	Full	Baha C2	5N steel band	Neck	Contra		12					
2b	Full	Baha C2	Baha Attract	Ipsi mastoid	Ipsi	5A	13	09.05.18				
2b	Full	Baha C2	Baha Attract	Ipsi mastoid	Top		14					
2b	Full	Baha C2	Baha Attract	Ipsi mastoid	Contra		15					
2b	Full	Baha C2	Implant screw	Ipsi mastoid	Ipsi	6A	16	11.05.18				
2b	Full	Baha C2	Implant screw	Ipsi mastoid	Top		17					
2b	Full	Baha C2	Implant screw	Ipsi mastoid	Contra		18					
2c	Full	BoneBridge	5N air cyl.	Dura	Ipsi	7A	19	14.05.18				
2c	Full	BoneBridge	5N air cyl.	Dura	Top		20					
2c	Full	BoneBridge	5N air cyl.	Dura	Contra		21					
2a-d	Empty	Baha C2	5N steel band	Ipsi mastoid	Ipsi	1B	22	18.05.18				
2a-d	Empty	Baha C2	5N steel band	Ipsi mastoid	Top		23					
2a-d	Empty	Baha C2	5N steel band	Ipsi mastoid	Contra		24					
2a-d	Empty	Baha C2	5N steel band	Forehead	Ipsi	2B	25	15.05.18				
2a-d	Empty	Baha C2	5N steel band	Forehead	Top		26					
2a-d	Empty	Baha C2	5N steel band	Forehead	Contra		27					
2a-d	Empty	Baha C2	5N steel band	Eye	Ipsi	3B	28	15.05.18				
2a-d	Empty	Baha C2	5N steel band	Eye	Top		29					
2a-d	Empty	Baha C2	5N steel band	Eye	Contra		30					
2a-d	Empty	Baha C2	5N steel band	Neck	Ipsi		31					

Figure 8. A ledger created by Ivo Dobrev to track data

Although the ledger details all of the experiments and measurements that researchers plan to run, not all tests are actually conducted. The rationale for not running a test can vary; for example, one common reason is to avoid destroying a sample. The ledger, however, does not record the reasoning as to why researchers decide not to run a test. If the project is revisited

several months later, researchers often do not remember why certain tests were not performed, leading to a loss of information.

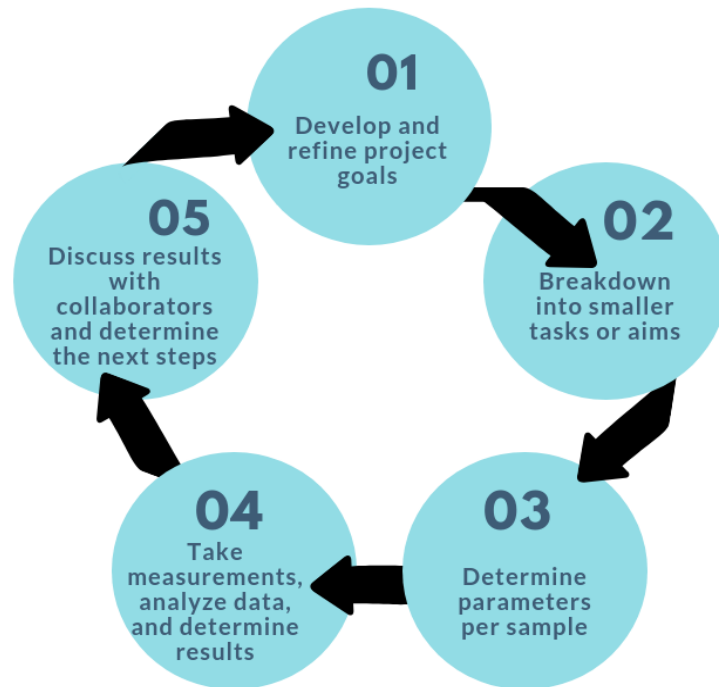
Photos are often used by researchers in the ORL clinic in order to facilitate documentation. These photos fall into one of two general categories: photos of notes, and photos of experimental setups. Dr. Dobrev explained that paper lab notebooks are becoming increasingly uncommon at UniversitätsSpital Zürich, but any notes written on spare paper are later photographed using a smartphone and uploaded onto a folder on the networked drive. In subsequent interviews, the team learned that although the ORL does have a camera designated for documentation purposes, it often goes unused because smartphones are more convenient. All photos captured on a smartphone from one experiment are saved in the same folder, but there is no standard procedure for labeling or naming them. While giving the team a tour of the lab space, Dr. Dobrev explained that his smartphone automatically generates a generic timestamp file name for each photo that does not include any additional descriptive information. This can cause issues later in the research process, as the lack of descriptive labels requires a researcher to click through all of the photos in the folder in order to find the desired information.

Dr. Andreas Eckhard, a clinical resident and research group leader specializing in neurotology and inner ear biology at UniversitätsSpital Zürich, relayed similar information about naming files. Dr. Eckhard's lab does not have a standardized procedure for organizing data, but he does have a consistent system for naming his own data files. By including the date and the code for the experiment series, Dr. Eckhard is easily able to identify his data files, which is a testament to how consistent file naming benefits researchers.

Researchers in the ORL clinic use MATLAB as the primary software for data processing. The software is relatively easy to use and includes built-in features that help researchers analyze their data. For those familiar with computer programming, organizing MATLAB code is a simple and useful task when done routinely. Well-organized code is characterized by the presence of a new code file for each project, the use of comments to explain how each function works, and the absence of unused or irrelevant code. Clearly-documented code is a valuable resource for those who are less familiar with MATLAB because the file is self-explanatory. Without standardized guidelines, researchers' code is often messy. Some researchers use the same MATLAB file across projects, leading to errors when the code is edited, saved, and reused repeatedly. This problem is compounded by the presence of thousands of lines of code, many of

which are commented out and therefore unnecessary. This inefficient setup makes the code cumbersome and more difficult to use.

SNSF-funded projects tend to be more organized, primarily because grants require detailed research proposals beforehand. The proposal forces researchers to think about what they want to prove, and how they will prove it prior to commencing an experiment. Furthermore, they must develop a specific list of tasks, which breaks down the project into manageable steps. From there, researchers determine which parameters they will need in their experiment, and how they will test for them. After taking measurements, they can process data and determine their results. This process is commonly referred to as the research workflow and is illustrated in Figure 9 below.



*Figure 9. Research workflow*

During the research and experimentation process, researchers tend to spend less time efficiently organizing their data. This results in a conglomeration of data files that is difficult to navigate when the data needs processing. Proper data organization provides traceability of files, allowing researchers to quickly determine where files came from and where they are currently located. When files lack traceability, data processing becomes substantially more difficult.

## 4.1.2 Other Facilities

The team also interviewed researchers at other facilities in order to obtain more information about current data management practices. This section summarizes the findings from interviews conducted with researchers at the University of Massachusetts Medical School, Worcester Polytechnic Institute, Eidgenössische Technische Hochschule (ETH) Zürich, and Interstaatliche Hochschule für Technik NTB Buchs. Summaries from these interviews can be found in the Appendices.

### *4.1.2.1 University of Massachusetts Medical School*

The University of Massachusetts Medical School (UMMS) is a public medical school located in Worcester, Massachusetts. Ms. Rita Fagan, a Ph.D. student studying neuroscience at UMMS, was interviewed about data organization practices in the small lab in which she works. Because the lab does not process large quantities of data and conducts small-scale research, digitization is regarded as a bonus feature rather than an absolute necessity. Numerical data is stored in a computer, typically in an Excel spreadsheet, and attributed to the researcher. Paper lab notebooks contain information about the processes that took place.

Ms. Fagan noted that their method of documenting and storing data is easy to use and that one of the strongest features of this method is the lab notebooks, which contain detailed information about negative or unsuccessful results. Although these additional details are typically not mentioned in research papers, they can be valuable for scientists. At the end of each day's notes, Ms. Fagan takes time to write down all of her results and logs what worked well. If she encountered any problems, she notes why the problem occurred and what she will do next in order to work around it. For Ms. Fagan, it is worth taking the extra time to record details, as this allows her to return to a project weeks later and determine why each decision was made.

Ms. Fagan noted that the paper lab notebook system at the small UMMS lab comes with two main drawbacks. First, if a researcher initially finds information about an experiment on a computer, they then must search through the collection of lab notebooks to find the corresponding notebook and experiment. The process of finding the exact lab notebook can be time-consuming, impeding the quick location of information. Secondly, note-taking practices can vary greatly among researchers. While some researchers fastidiously log each and every detail into their lab notebooks, others jot down vague ideas on scrap pieces of paper. These scrap



pieces often do not go into the notebook, and any valuable information that was written therefore becomes lost.

#### *4.1.2.2 Worcester Polytechnic Institute*

Worcester Polytechnic Institute (WPI) is a private research university located in Worcester, Massachusetts. Prof. Dr. Dirk Albrecht, a faculty member at WPI researching microtechnology and neuroscience, provided the team with insights on documentation and data management in his lab. The lab currently does not have a standardized procedure for documentation; however, all researchers are strongly encouraged to document their data. Dr. Albrecht explained that the data collection software used in the lab will automatically name files with the date, the time the data was captured, under which conditions the data was captured, and the data point in question. From there, the data files are stored in specific folders. Access to the folders is restricted to users working on the project in question, and the data is backed up in multiple off-site locations.

Like at USZ, exploratory experiments are typically not well-documented. These pilot experiments are generally used to determine if further experimentation will yield interesting or useful results, so documentation is suggested but not required. When notes are taken, the expectation is that researchers write a text file detailing the purpose of the experiment and any preparation steps they took. This practice is not yet consistent throughout the lab. Furthermore, trials are often conducted within a short time frame; as a result, researchers sometimes cannot take detailed notes. However, information from these experiments could prove to be helpful, so even brief documentation indicating the results and observations could be beneficial. While inconsistencies in documentation are not a major issue, Dr. Albrecht noted that implementing a documentation standard could save researchers time during the data processing phase.

In terms of data organization, Dr. Albrecht's lab has tried to implement a ledger system in Excel to keep track of their data files. Each file name had a specific numeric code, which allowed them to easily repeat experiments or analyze files against each other. Initially, the system worked fairly well, but it was difficult for subsequent researchers to come in after the ledger was established and discern what was going on. Ideally, researchers would have a system where they could easily compare data sets and quickly determine the goal or question that was tested.

#### *4.1.2.3 Eidgenössische Technische Hochschule (ETH) Zürich*

Eidgenössische Technische Hochschule (ETH) is a technical and scientific university located in Zürich, Switzerland. Two researchers, Dr. Kathryn Rosowski and Prof. Dr. Raoul Hopf, were interviewed to observe their methods of documenting and organizing data. This section summarizes the findings from their interviews and further details from these can be found in the Appendices.

Prof. Dr. Raoul Hopf, a member of the Department of Mechanical and Process Engineering, focuses on doing experimental work on soft materials. Any electronic data acquired in his labs is saved onto hard drive disks. When the data is needed, these disks are then loaded onto computers that are disconnected from any network drive. These computers are not connected to the network drives because the information technology (IT) specialists at ETH do not want to clutter online storage spaces. Necessary data from these drives is then transferred to temporary secure drives. Once the data is processed and a paper is written, any raw, processed, and simulation input/output data included in the publication is transferred onto a secure and backed-up online drive. This drive is designed to hold and protect the data long-term, as it is published work and needs to be accessible.

For Dr. Rosowski's lab, which primarily conducts research on soft matter, living matter, and the interacting forces between cells, image files are the most commonly generated data and are ideally labeled with the name and date of the experiment and some additional descriptive keywords. These image files often take the form of time lapses or Z stacks. Z stacks are images recorded at different focal planes so that the volume of a sample can be visualized and rendered. Time lapses and Z stacks tend to be large in number and collected on microscopes. These image files then get stored on an attached computer. Ideally, the necessary files should be transferred from the computer and backed up onto the ETH's cloud system, which has designated folders in which specific labs can store their files. Any unneeded files on the computer should be deleted. Unfortunately, although this is a suggested procedure, there is no lab-wide policy in place that states this. Computers are often regarded as "storage places", which can make it difficult for researchers to decide which files are needed so that unnecessary files can be deleted and more space can be made on the computer.

Writing notes in a lab notebook is the preferred method in Dr. Rosowski's small-scale lab. Although the general template of documentation varies from person-to-person, common

identifiers and labels include the title of the experiment and the date. Each project has its own notebook. The lab's system of keeping lab notebooks runs into many of the same obstacles that were described previously by Ms. Rita Fagan from the University of Massachusetts Medical School. Documentation and the level of detail dictated in lab notebooks is inconsistent from person-to-person, and sometimes even project-to-project. For Dr. Rosowski's lab in particular, which employs lab members who hail from several different countries, even the language in which notes are taken can vary between researchers. This issue, coupled with inconsistent documentation, can make it difficult for new members coming into the lab or onto the project to read the notebook and understand what was done during an experiment. In these scenarios, the common practice is to email or contact the person who originally conducted the experiment or project. Occasionally, researchers may start the project over again.

Although their current practices suit their needs at the moment, Dr. Rosowski and Prof. Dr. Raoul Hopf noted that guidelines or defined processes for documenting and storing data would be particularly useful and could resolve some of their occasional issues data management issues.

#### *4.1.2.4 Interstaatliche Hochschule für Technik NTB Buchs*

Interstaatliche Hochschule für Technik NTB Buchs, more commonly known as NTB (Neu-Technikum Buchs), is a technical school located in Buchs, Switzerland. NTB often works with industry partners, and much of the research is later used for patents and the subsequent development of products. Dr. Emine Cagin, a researcher at the Institute for Micro- and Nanotechnology at NTB, provided information about her current process for documenting and organizing data.

Although much of her work is done in interdisciplinary teams with differing documentation strategies, Dr. Cagin has found that her group is relatively well-organized and communicative. In the brainstorming phase, researchers collaborate via RealtimeBoard, an online whiteboarding platform where teams can share ideas, make sticky notes, drawing mockups, and more. This system allows team members to contribute to project efforts and help brainstorm, even from remote locations.

During projects, most notes are taken in paper lab notebooks. For every experiment, process-specific parameters such as temperature and humidity conditions must be well-documented. Dr. Cagin and her team work with highly sensitive nanotechnology and small

sample materials, so this information is of great importance. Additionally, the date and location must be noted, as well as who is conducting the experiment. Because their research is often shared with industry partners and subsequently used to make products, Dr. Cagin and her team must maintain “patent-compatible documentation”, an in-house term used to describe the level of detail needed for research data they hope to use for patents. Information must be clear and comprehensive, so team members often take a significant amount of time to record project notes. She and her team are currently exploring ways to make this documentation process faster and more efficient so that more time can be dedicated to other important tasks.

Electronic files are labeled with the date, the project name, a few short keyword descriptors, and the initials of the person who either created the file or last edited it. This method of labeling files is consistent across team members, resulting in standardized file names. Dr. Cagin stated that the amount of information included in these file names is sufficient enough to allow for the rapid identification of file contents.

In terms of file sharing and transferring, NTB has transferred from SWITCH, a data transfer system, to a secure-server system hosted by the university. This new system, similar to a networked drive, allows researchers to upload and view files related to solely the projects they are working on while maintaining a high level of security. In collaboration with their IT (Information Technology) department, NTB has developed a procedure for the accessing of files by outside collaborators. Because Dr. Cagin and her team often work with industry partners, the ability to allow outsiders to view files and documents is highly necessary. In this new system, after signing a document ensuring that the shared data will not be used for irrelevant or malicious purposes, outside collaborators are given access to a secure, predetermined folder path. This folder path is always labeled “Exchange”, and any outside collaborators with the correct level of access are able to view the folder’s contents as if they were working at NTB. Dr. Cagin noted that this new system works well for their purposes and has made collaboration much easier.

The methods of organizing and documenting research data described by Dr. Cagin have evolved over the course of several years, but are now solidified into lab-wide practices. Researchers in her teams generally work on multiple projects simultaneously and tend to be self-motivated to write detailed experiment notes. Because many teams work with industry partners, writing detailed notes prevents them from losing data, enables them to easily answer questions from their partners, and allows them to pick up where projects left off. Dr. Cagin’s lab does not

have guidelines detailing how data should be documented and organized, but new lab members are mentored closely, allowing for these methods to be passed along.

## 4.2 Summary of Common Issues

Overall, research facilities both inside and outside of Switzerland face similar challenges in data management. Minimal documentation standards, a lack of uniform file naming procedures, inefficient file folder and code structure organization, inadequate data transfer methods, inconsistent file storage mechanisms, and lack of initiative for an organizational culture change are among the common problems researchers commented on during interviews. In the next few sections, these major issues and their prevalence as widespread obstacles are discussed in greater detail.

### 4.2.1 Minimal or Inconsistent Documentation Standards

Due to the lack of defined, standardized protocols, the level of detail in documentation varies among researchers. Some researchers record procedures or conclusions in more detail than others, and as a result, experiments and projects are documented with varying levels of information. Experimental conditions or outcomes that may be clear or common knowledge for some researchers may not be written or documented with enough detail for those who are unfamiliar with certain concepts, leading to a loss of information.

This issue is further compounded by language differences. English is widely considered the lingua franca of science, but it is not an official language of Switzerland. Additionally, many researchers from outside of Switzerland collaborate on projects, bringing even more languages into question. This leads to disputes between researchers over which language to take notes in and is especially pertinent in labs that heavily depend on the use of lab notebooks. Some researchers argue that they are best suited to take notes in their native language, while others insist that notes must be written in English so that all can understand the research.

Regardless of the documentation language, insufficient documentation consistently leads to the poor repeatability of experiments and often impedes comprehension of the project.

### 4.2.2 Lack of Standardized File Naming Procedures

Although naming files is regarded as an important step in documenting and organizing data, most facilities and lab groups do not have a standardized file naming system. Researchers

often do not have the time to develop such a system, and as a result, different researchers use different naming conventions, leading to confusion.

There are common parameters that researchers use to name files, such as the date and experiment type, but the issue revolves around consistency among file names. For example, some researchers use additional keywords to provide details about the experiment. This can result in long file names. Others utilize the same keyword with different capitalizations or abbreviations, leading to inconsistent naming conventions. Consequently, files become harder to find, data could eventually get lost, and the experiment has to be repeated again, wasting time and resources.

#### 4.2.3 Inefficient File and Code Structure Organization

Developing an organized file structure to store various types of files has proven difficult in many facilities that the team came across. Several facilities use a file directory and set up a folder for each project or experiment. However, this structure becomes cluttered as more and more data is produced during experiments, and various types of files are used. As a result, it becomes difficult to manage and locate specific files. Additionally, as more files are generated in the file directory, more storage space is taken up, which poses as another issue. Other facilities use an external hard drive to store data. This can lead to loss of data if the hard drive becomes irretrievable.

Another concern is code structure and organization. Many facilities use MATLAB as their primary method of analyzing data, but the code is often not organized in an efficient manner. Particularly at USZ, the team observed that non-functional code is frequently commented out and unnecessarily included in the final program. This results in programs with lines of irrelevant code, resulting in a cluttered file that is difficult to read and understand.

#### 4.2.4 Inadequate Data Transferring Methods

Due to certain software and computer access restrictions, data is not always easily transferred. Many experiments that the team came across had large file sizes, and many labs generate hundreds of gigabytes of data per day. Such amounts of data can take a long time to transfer, even across drives that are connected directly to the facility's network.

If a computer is not connected to the network drive or lacks sufficient authorization, data must be physically transferred via a flash drive. For example, data from CT (computerized

tomography) scanners that is needed by researchers often must be copied onto a flash drive for use on their “unauthorized” computers. Researchers also use flash drives to transfer data to collaborators outside of USZ who implicitly cannot be connected to the USZ networked drive.

This system of using flash drives as a means of transferring files poses a key threat to researchers: the loss of data. If the flash drive is broken or misplaced, researchers may lose all of their data from that experiment, forcing them to spend time repeating it.

#### 4.2.5 Inconsistent File Storage Mechanisms

Although paper and electronic records each have their advantages, having mixed forms of storage mechanisms create three key issues: vague documentation, poor organization, and inconsistent conversion to a digitized format. When data exists both in paper and digital form, documentation becomes more difficult, resulting in problems determining who did what and when the data was collected. Because of this, returning to a project after a hiatus is often onerous because it is unclear as to where previous researchers left off.

Digitization provides a clear solution to this problem; however, the conversion from paper to digital files is not always perfect and can be time-consuming. There are two main ways of converting paper files into digital files: manually entering data and scanning data. Digital data is a powerful tool for researchers, but manual data entry methods can be a time commitment. Manual data entry essentially requires the researcher to take double the number of notes, as he or she must maintain both a physical and digital notebook.

Scanning provides a simpler and faster way to convert files, but this method also contains pitfalls. Even if the files are present in a digital format, the notes are not always readable, as some handwriting is illegible. Informal activities, such as brainstorming, are sometimes not written in lab notebooks. Because of this, the ideas that come from these activities are easily discarded or lost. Moreover, even if the brainstorming document is digitized, the order of information is not always clear. Multiple researchers may contribute to one handwritten document, and turning the page into a flowchart that may make sense at the moment may not be discernible several weeks later. Overall, the digitization of paper files causes two secondary issues in accessibility and analysis, as data may be lost in the paper shuffle during file conversion, thereby reducing the amount of data that is accessible and interpretable.

#### 4.2.6 Lack of Initiative for an Organizational Culture Change

Implementing better practices for managing research data can be difficult, particularly because the current mindset and culture of organizations and labs are deeply embedded. Changing the culture and mindset requires time and effort from all parties involved.

For many labs, adopting new practices will be a challenge. Researchers are comfortable in their current methods of documenting and organizing data, even if they are not conducive to efficient data processing. Asking researchers to deviate from their old practices will take time and persistence.

In the case of Dr. Cagin’s lab, described in 4.1.2.4, data management methods have been well-established and are passed on to new lab members. This has resulted in a culture of well-documented and organized research data. This required a marked effort by early members of the lab in order to solidify these practices as commonplace actions. Such determination to change a culture’s organization is rare, and many labs have not had the time to implement new practices. Committing to new ways of documenting and organizing data will be a challenge.

### 4.3 Conclusions

An analysis of the interview findings revealed high dissatisfaction rates among researchers at USZ. Many researchers were unsatisfied with the current state of affairs in all five of the areas the team identified, and graphs of researcher satisfaction rates can be found in Figure 10 below.

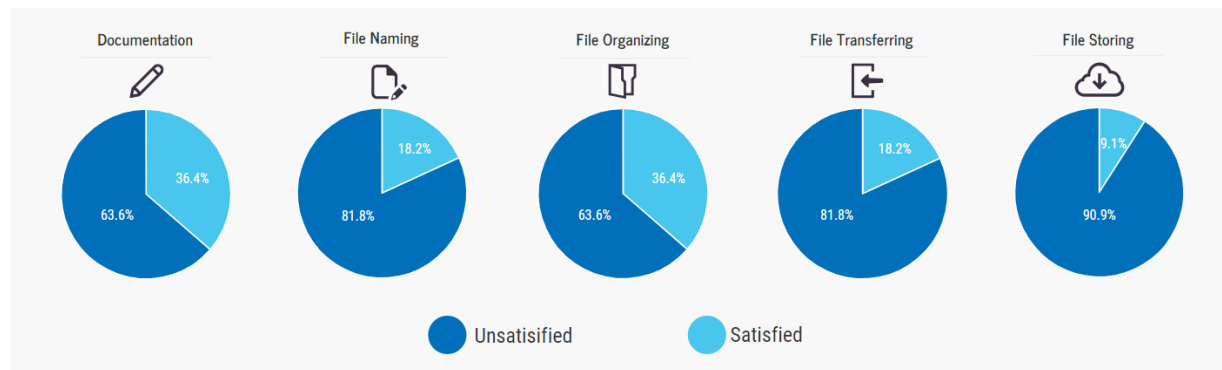


Figure 10. Researcher satisfaction rates based on interview results

Documentation and file organizing had the highest satisfaction rates at 36%; file naming and file transferring had the second highest satisfaction rate at 18%, and file storing had the



lowest satisfaction rate at 9%. The low satisfaction rates are largely due to the lack of guidelines on how to manage and organize data and are further exacerbated by the current lab culture at USZ. This project aims to increase satisfaction rates in all five key areas.

#### 4.4 Selection of Deliverables

A wide range of solutions exist that could be used to improve data management at USZ; however, the team chose to develop best practice guidelines to provide recommendations for major areas of data management and an application that will facilitate file naming for the following reasons.

In discussions with the project sponsor, it was apparent that research groups could greatly benefit from a document that outlines good practices for managing data. Efficient data management practices are rarely at the forefront of a researcher's mind while conducting an experiment, and lab-wide rules for managing data generally go unimplemented. In order to increase satisfaction rates in the five key areas of data management mentioned previously and partially fulfill the hospital's lack of instructions regarding data management procedures, the team decided to create best practice guidelines. Best practice guidelines have the flexibility to address documentation, file naming, file organizing, file transferring, and file storing on a general level and provide small recommendations that researchers can implement on their own without the help of external sources, such as their IT department or new software.

The decision to develop an application that generates standardized file names was also motivated by discussions with Dr. Dobrev. Issues that researchers have in regards to file organization, transfer, and storage are often traced back to inconsistent or non-descriptive file names. When files are named vaguely, from a quick glance at the label, researchers are cannot discern the file's contents. A file name such as 'Test 1' provides little to no information about details such as who conducted the test, when the test was conducted, and the project to which the test belongs are not easily known. Such ambiguous names may hold some meaning to the researcher who created the file, but a collaborator on the project will have a much harder time understanding the context of the file. Additionally, it is difficult for a researcher to organize and store data when file names show little indication of how it relates to other data.

A file naming application ensures that abbreviations and acronyms for certain keywords are used consistently across file names. This is also one of the easiest software solutions to

implement given that the application does not require any special interfaces or external software to ensure compatibility. On the other hand, file transferring and storage improvements require solutions that must be integrated on a per-department basis, and developing such solutions would be outside the scope of a student project, and more in the realm of the hospitals' IT department.

The development of a file naming application, in conjunction with the best practice guidelines, specifically addresses an aspect of data management that can greatly improve how data is managed as a whole. These solutions can be implemented by researchers themselves and aren't dependent on assistance from outside groups such as IT departments.

## Chapter 5: Implementation and Recommendations

This chapter summarizes the process the team followed in order to create and implement the project deliverables, discusses the results from the feedback sessions and surveys that were conducted, describes the team’s recommendations for the use of the deliverables, and provides information about opportunities for future development of the file naming application and other projects.

### 5.1 Creation of Deliverables

As discussed in Chapter 4, the team decided to create a set of best practice guidelines for managing medical research data and develop a computer application to generate standardized file names. This section discusses these deliverables in further detail and describes how they were created.

#### 5.1.1 Best Practice Guidelines

After conducting and analyzing interviews, the team identified five key areas that needed to be addressed by the project deliverables: poor documentation, inconsistent and ambiguous file names, poor file organization, inefficient data transfer methods, and insufficient data storage methods. Figure 11 below outlines the process the team followed to develop the best practice guidelines, called the Research Data Management Handbook. The handbook in its entirety can be found in Appendix Z, and its contents are described in detail in subsequent sections of this chapter.



*Figure 11. Process workflow for developing the best practice guidelines*

The first step in developing the best practice guidelines was synthesizing information about the obstacles researchers face in each of the five aforementioned areas of data management. By doing so, the team identified which specific aspects to address. For example, some researchers choose convenience over efficiency when storing and organizing electronic

data files and folders. Those researchers simply store files on their computer desktop for ease of access in lieu of properly filing them into a folder. While this practice can be convenient in the moment, it becomes difficult to keep track of files as more are added to the screen, resulting in poor file organization. Similar issues were identified as specific practices needing recommendations for improvement.

To brainstorm recommendations for the obstacles and issues mentioned above, the team identified successful data management practices that were mentioned by interviewees. For example, Dr. Dobrev spoke briefly about his personal file organization system: each project has its own overarching folder. Within that folder are subfolders labeled with a task or aim of the experiment, and each subfolder has separate folders for raw data, processed data, and supporting materials. The team noted similar data management practices while developing other recommendations and solutions. Additionally, the team consulted resources from other institutions, such as the University of Cambridge, to obtain further information about other successful data management practices. Using this information, the team developed “good rules of thumb” for documenting data, naming data files, organizing data file folders, and storing and transferring data. These provide specific instructions on how to manage data well.

A preliminary draft of these recommendations was written and presented to Dr. Dobrev and his research group for feedback. Overall, the team received positive feedback about the handbook, and the sponsor and his colleagues felt that the content was appropriate and useful. For the final document, the team was asked to incorporate requirements set by the Swiss National Science Foundation and USZ into the guidelines. Specifically, the team needed to integrate information about the SNSF Data Management Plan and FAIR Data Principles into the document, as these are the regulations and principles that researchers at USZ must abide by in order to receive funding from the foundation.

For example, the first section of the DMP requires researchers to explain what data will be collected and how they intend to collect it. To clearly convey this information to the SNSF, Chapter 2 of the Research Data Management Handbook recommends researchers do the following before or during the initial stages of the experiment or project:

- List the types of data that will be collected
- Describe, in detail, the formats of raw and processed data
- Estimate the volume of raw and processed data that will be generated

In addition to these recommendations, the handbook supplies specific examples of how researchers would write this information in an actual grant application. An example of this is:

*Western blot data is collected for the detection of proteins. Images are taken using a CCD camera. The image data is quantified and processed in Microsoft Excel. The data files for the images and Excel files total 2GB and 4GB, respectively.*

These examples were cross-checked with information provided by the SNSF in order to ensure the information is accurate. After collecting the feedback from Dr. Dobrev and his team, the document was refined and finalized. The Research Data Management Handbook includes the following chapters and sections:

- Chapter 1: Introduction
- Chapter 2: Documenting Data
- Chapter 3: Naming Data Files
- Chapter 4: Organizing Data File Folders
- Chapter 5: Storing and Transferring Data Files
- Additional References

Each chapter of the Research Data Management Handbook describes a certain aspect of data management, provides specific suggestions for executing each aspect effectively, and discusses how to apply these suggestions to projects that want to apply for funding from the SNSF. The following sections summarize chapters in the Research Data Management Handbook.

#### *5.1.1.1 Chapter 2: Documenting Data*

The level of detail in documentation varies among researchers. While some record experimental procedures, observations, and conclusions in thorough detail, others may not be as specific. When documentation is vague or general, researchers who later read the notes cannot comprehend the full details of the experiment.

Chapter 2: Documenting Data begins by explaining what research documentation is and details its importance. Subsequent sections provide recommendations on how to best document notes in a paper lab notebook, electronic notes, and audio notes. For these three mediums, the handbook recommends the following items are dictated or referenced:

- The date
- The name and aims of the experiment

- The experimental procedure and conditions
- Any observations and conclusions, including mistakes
- Any explanations of abbreviations

This chapter also provides recommendations on how to document code well. These include clearly naming variables, methods, and classes, and more.

The handbook supplies researchers with clear lists of what should be included in research documentation. Researchers are often too busy to remember every item that needs to be included in their documentation, so having the information written out will provide useful reminders.

#### *5.1.1.2 Chapter 3: Naming Data Files*

The manner in which a file is named has a substantial impact on a researcher's ability to locate it later and discern its contents. File names that have additional keywords that provide further details about the experiment are informative but occasionally may be too long. On the other hand, short file names often provide too few details, forcing researchers to spend time opening the file in order to determine its contents. The team sought to provide suggestions for finding a balance between these. Chapter 3: Naming Data Files describes the benefits of well-named files and recommends that the following components always be in a file name:

- The date
- The name of the researcher
- The type of the experiment
- The trial number, if applicable
- One or two keywords that describe additional details about the experiment

Furthermore, the handbook provides examples of well-named and poorly-named files in addition to recommendations for implementing these practices in labs, such as discussing standard conventions with the whole lab group and using the app the team developed, File Namer Pro, which is a computer application that generates standard, detailed file names. Further information about File Namer Pro can be found in Section 5.1.2.

These recommendations will enable researchers to create and implement standardized and detailed file naming conventions, remedying the issue of inconsistency.

### *5.1.1.3 Chapter 4: Organizing Data File Folders*

Being unable to maintain well-structured data file folders is a problem among labs. This is an issue because different projects will necessitate different organizational structures; however, the Research Data Management Handbook provides information that serve as general guidelines for organizing data file folders.

In Chapter 4: Organizing Data File Folders, effective folder organization requires that:

- Files are easily filed
- Files are easily found
- File folder structures are standardized

In order for researchers to create and maintain file structures with these characteristics, the following methods are recommended:

- Maintain logical folder hierarchies, with more important parameters located in higher levels of the hierarchy
- Create uniform folder structures for different stages, tasks, or aims of the experiment
- Group folders by date
- Minimize the number of folders by only creating new folders when there is a logical grouping of similar files
- Limit the number of files saved to the desktop
- Name files and folders logically
- Discuss folder structure standards with collaborators and come to a consensus

By implementing these practices and consulting the examples included in the chapter, researchers will file, find, and store data more effectively.

### *5.1.1.4 Chapter 5: Storing and Transferring Data*

Due to certain software and computer access restrictions, researchers often have problems storing and transferring data. This is an issue because data can be lost or inaccessible by parties that need it for analysis or processing. Although providing substantial solutions to this is outside the realm of the project, the team has supplied recommendations for improving data storage and transfer that researchers can implement on their own.

Chapter 5: Storing and Transferring Data describes ways that handwritten and electronic

data can be stored and transferred well. For example, it is suggested that any handwritten notes be kept in a secure location in an organized manner and digitized as an extra precaution. For electronic data, one recommendation is that computer drives be routinely backed up and maintained, data is not transferred via flash drive or other physical means that could be lost or misplaced, and that researchers consult with their IT department to ensure their data storage and transfer needs are met. With these recommendations in mind, researchers can better ensure their data will not be lost.

#### *5.1.1.5 Further Information*

Additionally, the team recognized that researchers may not have enough time to read the entirety of the Research Data Management Handbook, negating the document's purpose. As a result, the team decided to include checklists in the final document. Checklists serve to solve this issue by providing short and concise information. Checklists also have the added benefits of:

- Promoting completeness by ensuring the steps in a process are not skipped
- Ensuring consistency
- Compensating for memory and attention shortcomings
- Preventing mix-ups

In order to improve satisfaction with the five areas of data management identified above, the team recommends that the lab or department head at the very least encourages usage of the checklists. These can be printed out and posted on walls or lab benches to serve as visual reminders to researchers to manage their data well. Although reading the Research Data Management Handbook would be highly beneficial to the researchers, the team recognizes that they may not have enough time to do so.

Lastly, although the handbook was written with the needs of USZ in mind, the document can be used by other researchers and facilities. The information included in each chapter is general and broad enough for wider use in areas beyond medical data, but specific enough to provide targeted recommendations for improving medical data management.

#### **5.1.2 File Naming Application**

To specifically improve an aspect of data organization at USZ, the team decided to develop an application that helps researchers name their files in a standardized manner. While the best practice guidelines serve as broad coverage for multiple aspects of data management, the



file naming application serves as a functional solution. Standardized file names help researchers locate specific documents quickly and efficiently. In order to successfully develop this application, the team used an Agile Methodology, which is further explained in Section 5.1.2.2.

#### *5.1.2.1 Gathering Requirements*

The first step in the application development process was establishing the overall purpose of the program and its outputs and define certain design considerations. A key design consideration was intuitiveness; while a user manual was created for the application, it should not need to be frequently referenced. To establish the overall purpose of the program and its outputs, the team started the development process by crafting user stories. User stories detail specific features that the user wants to have. The exact user stories from each iteration of the application can be found in Appendices S, T, U, and V.

The following are several examples of user stories:

- As a researcher, I want to name my files by multiple parameter types so I can easily understand the general contents of a file at a glance.
- As a researcher or administrator, I want to be able to modify or add to the parameter types so I can select for parameters that have not been added using the original program.
- As a researcher, I want to save my project preferences between different experiments so I can quickly save and load preferred file name types
- As a researcher, I want to quickly increment file names when I am recording data consecutively, perhaps with a little button in a corner that would allow file names to be quickly generated.

From the user stories, the team developed an initial draft of the application's primary function, which is to automatically provide the user with a file name that can be copied.

The next step was to finalize which parameters that the user would input into the program. Based on interviews with researchers, the team decided that necessary parameters to include were the date, experiment type, trial number, and optional keyword descriptors in the file name. The user can type these parameters into the program, and the program will automatically consolidate them into a clear and detailed file name abbreviation. Additionally, the user can store

names and abbreviations for the experiment type and keyword descriptor fields in a comma-separated values (csv) database file. In the User Interface (UI), the names are displayed in a drop-down menu.

After outlining the key functionality requirements, the team designed a UI mock-up to visualize the application. Figure 12 shows the first draft of the UI design.

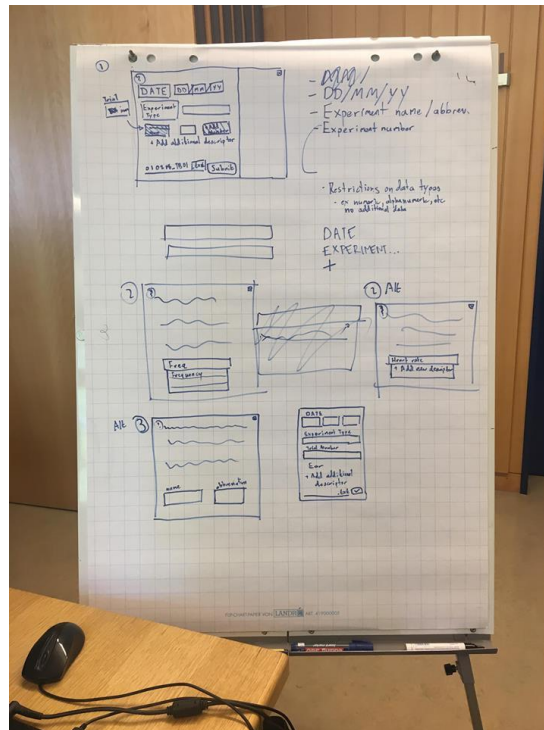


Figure 12. A preliminary draft of the user interface

The initial design of the application featured a rectangular window with fields to type in the trial number, select the experiment type from a drop-down list, and optionally add in additional descriptors using a plus (“+”) icon. The drop-down menus included an autofill feature that displayed words or phrases in the keyword that contain the characters typed into the field in order to help the user find their desired parameters more quickly. Additionally, the application automatically sets the date output as the current date; however, the date could be changed by the researcher if necessary.

After the user had entered the necessary information, the final standardized file name appeared at the bottom as a preview. From there, the user would be able to copy the name to their clipboard and paste it into their file directory.

### 5.1.2.2 Software Setup and Initial Application Design

After brainstorming and planning out the procedure to develop the file naming application, the team started the development process. In order to better organize this process, the team decided to work on the program in one-week sprints, which are detailed in Figure 13.

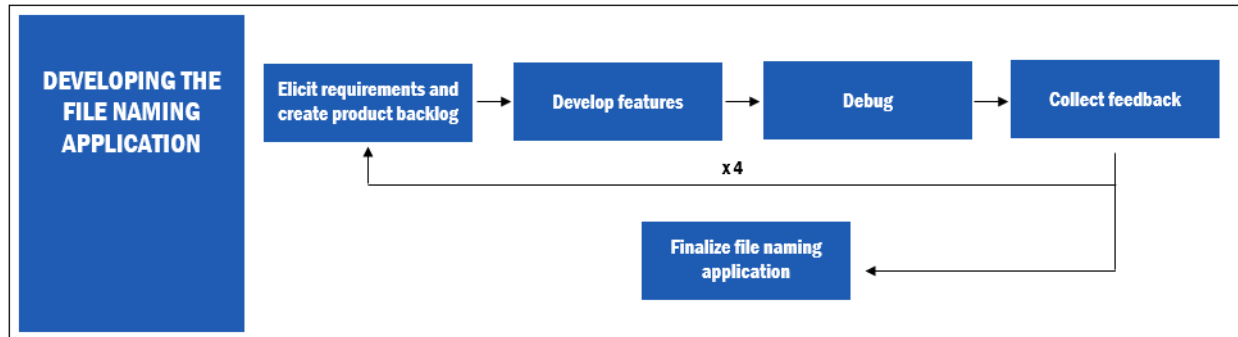


Figure 13. Process workflow for file naming application sprints

Sprints are part of the Agile Methodology in which a product follows the entire software development life cycle in a single iteration. A completed version of this program is then presented for feedback at the end of the time allocated for each sprint. Suggestions are subsequently collected and improvements are made to the existing program, resulting in improved versions of the product. The process repeats until a satisfactory application is delivered.

The first step in developing File Namer Pro was downloading and setting up software. The team used Java and JavaFX to create the application. In order to build the UI, the team utilized two different software development environments: IntelliJ, the primary IDE for code development and Scene Builder, a JavaFX tool that allows for more visual UI development and can be integrated directly into IntelliJ. We also used Gradle to download and build all the required libraries and JFoenix, an open source JavaFX library, which contains pre-built, visually appealing UI elements. Java DB, a distribution of the Apache Derby database, was used in order to build a persistent glossary of projects and additional descriptors with their corresponding appreciations and definitions. Additional information concerning the software used for the development of the application can be found in Chapter 2.

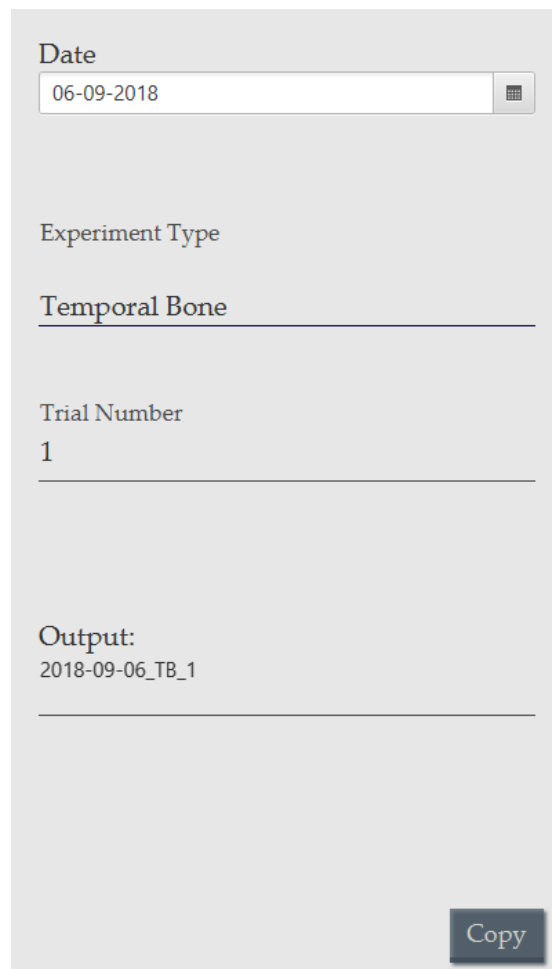
The next step was to establish the file structure to store and write the code. The two main folders were the java and resource folders as is typical of a Java project directory structure. Then, work on the database began by developing functions to store words that the user inputs into the

fields. The basic UI was developed in Scene Builder, and functionality was added via functions to input the date, format per European standards, and automatically show the current date in the field.

The following additional functions were added:

- Retrieve the experimental type and additional keyword parameters and return the corresponding abbreviation
- Obtain the trial number and automatically increment it
- Show a preview of the file name
- Copy the file name to the clipboard

The first working version of the application is shown in Figure 14.



The screenshot displays a user interface with the following elements:

- Date:** A text input field containing "06-09-2018" and a calendar icon to its right.
- Experiment Type:** A text input field that is currently empty.
- Temporal Bone:** A text input field containing the text "Temporal Bone".
- Trial Number:** A text input field containing the number "1".
- Output:** A text area displaying the generated file name "2018-09-06\_TB\_1".
- Copy:** A button located at the bottom right of the interface.

*Figure 14. Version One*

### *5.1.2.3 Bug Fixing and Improvements*

In this step of development, the team addressed functionality issues within the application, sought feedback from researchers, and made improvements based on the feedback collected.

The team hosted feedback sessions with the advisors, the project sponsor, and researchers from the ORL lab. The team held two feedback sessions for the first iteration of the application. One was held with the project advisors, Professors Albrecht and Orr, during the team's weekly meeting, and the other was held with the project sponsor, Dr. Dobrev. Both groups requested several new features for the application, which are reflected by the Iteration Two Product Backlog (Appendix T).

Some of the new features included:

- Options to expand and collapse the window
- Ensure the same word with different capitalizations or abbreviations was not stored as a separate entry in the keyword library
- Display the file name in real time
- Add an abbreviation of the researcher's name in the file name.

The team made major changes to both the user interface and the functionality of the application during the second iteration. A compact version of File Namer Pro was developed as well as a screen for the user to input and save project preferences. The compact version provides a smaller screen as well as fields for the user to update the date, the trial number, and the sample number. The team also held a larger feedback session with Dr. Dobrev, Dr. Pfiffner, Dr. Lukas Prochazka, and Dr. Jae Hoon Sim.

One of the most requested features by Drs. Dobrev, Pfiffner, Prochazka, and Sim was an Excel log file feature that would aid researchers in creating a timeline of the experiment as well as demonstrate the relationship between various project files. The overall tone of the feedback was positive, and the researchers appeared genuinely excited about the application.

The team demonstrated the third iteration of the application during the weekly ORL clinic meeting to a group of ten researchers in order to obtain further feedback. At this meeting, the researchers did not request any new major features but asked clarifying questions.

In order to prioritize the development of each feature, the team used a three-part ranking system. The product backlog (see Figure 15 below) contains a list of user stories followed by

three ranking columns. In the first column, the user story is assigned a t-shirt size based on the estimated difficulty, ranging from small (S) to extra-large (XL). Next, the user stories are ranked numerically by difficulty. The easiest small task is given the rank of one, and the rank doubles for a task that is twice as hard to complete. Once all the user stories were given t-shirt sizes and difficulty ranks, the team collectively decided on a priority number, which ranged from low priority (one) to high priority (five). From there, tasks were assigned to team members and development continued on the application. This ranking process was repeated for each iteration of the application and allowed for efficient and targeted improvements to the application.

I want to be able to modify and add new keywords	M	8	4
I want to be able to input a value in addition to a keyword if necessary	S	2	3
I want the software to automatically keep track of what trial number/sample number I'm on even after the application is closed	S	4	2
I want the researcher name to appear in the file name	S	2	4
I want to select the order in which information appears in my file name	S	2	2
I want to save my project preferences between different experiments	L	10	3

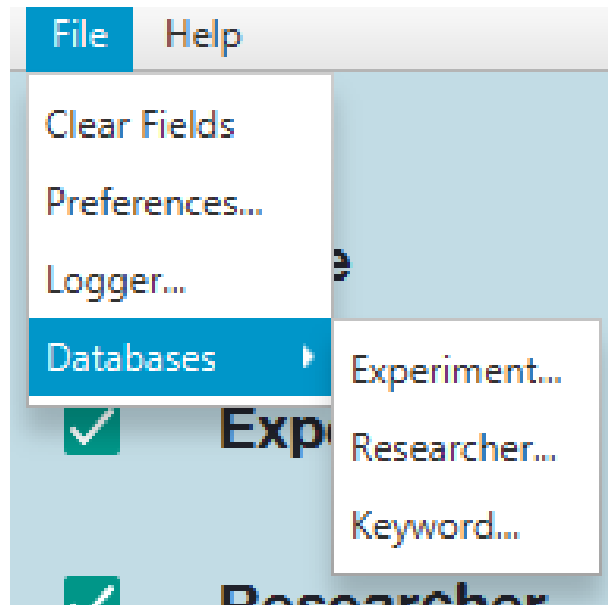
*Figure 15. Excerpt of a product backlog*

#### *5.1.2.4 Survey Results and Final Application Improvements*

Once the application was portable, the team sent the application along with a survey (Appendix W) to researchers. The survey was also sent to a software tester and a software architect who had no prior experience or involvement with the application. The feedback was positive overall, and researchers stated they would be likely to use the application in their daily research. Most of the features of the application were intuitive, although the logger feature was slightly confusing to several of the testers. As a result, the team modified the logger user interface to make it more intuitive, which included adding a preview of the log file.

The team also improved the UI of Full Namer by adding a file menu, reducing clutter. The file menu, pictured in Figure 16 below, includes Logger and Project Preferences, which previously were buttons on the Full Namer UI. It also includes several new features such as a

button to clear all the fields in Full Namer and table views of all three databases. A help menu was added as well. This menu includes an “About” page as well as a “Getting Started” page that contains detailed instructions on how to use the application.



*Figure 16. Expanded 'File' menu*

#### *5.1.2.5 Final Iteration of the Application*

The final iteration of the application generates file names containing the following identifiers:

- The date, in the format YYYY\_MM\_DD
- The abbreviation for the experiment
- The trial number
- The sample number
- The researcher's initials
- The abbreviations of keywords.

These components of the file name allow the researcher to easily identify the file's contents, and the data format enables them to easily sort their files chronologically. When using the

application, the researcher inputs their desired information for each field into an expanded version of the application (Figure 17).

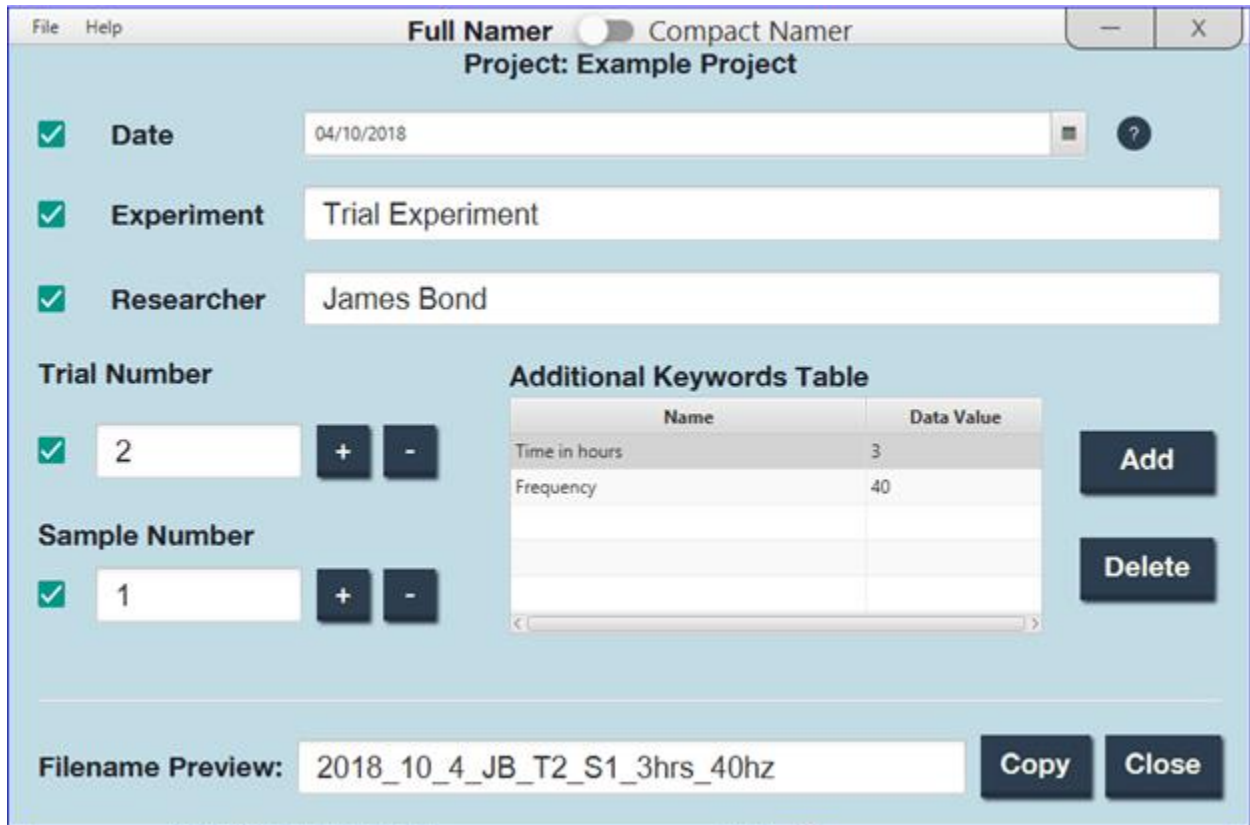


Figure 17. Final Full Namer user interface

Both the experiment type and the keyword entry fields include autocomplete text fields (Figure 18) where the user can search for their desired parameters in a database. If the researcher's name, experiment type or a keyword is not included in the database, the user can easily add it via a popup menu (Figure 19).

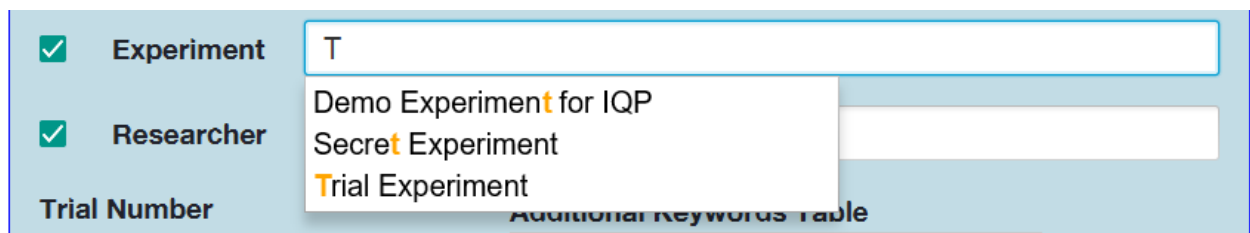


Figure 18. Autocomplete text fields in Full Namer



Figure 19. Add New Keyword to Database user interface

The application also includes a compact version (Figure 20) that stores the parameters inputted in the expanded version and allows the user to update the date, the trial number, and the sample number.

Figure 20. Compact Namer user interface

Each time a file name is generated, it is saved in a log file, which contains a more detailed timestamp, the full file name, and comments about that particular trial (Figure 21).

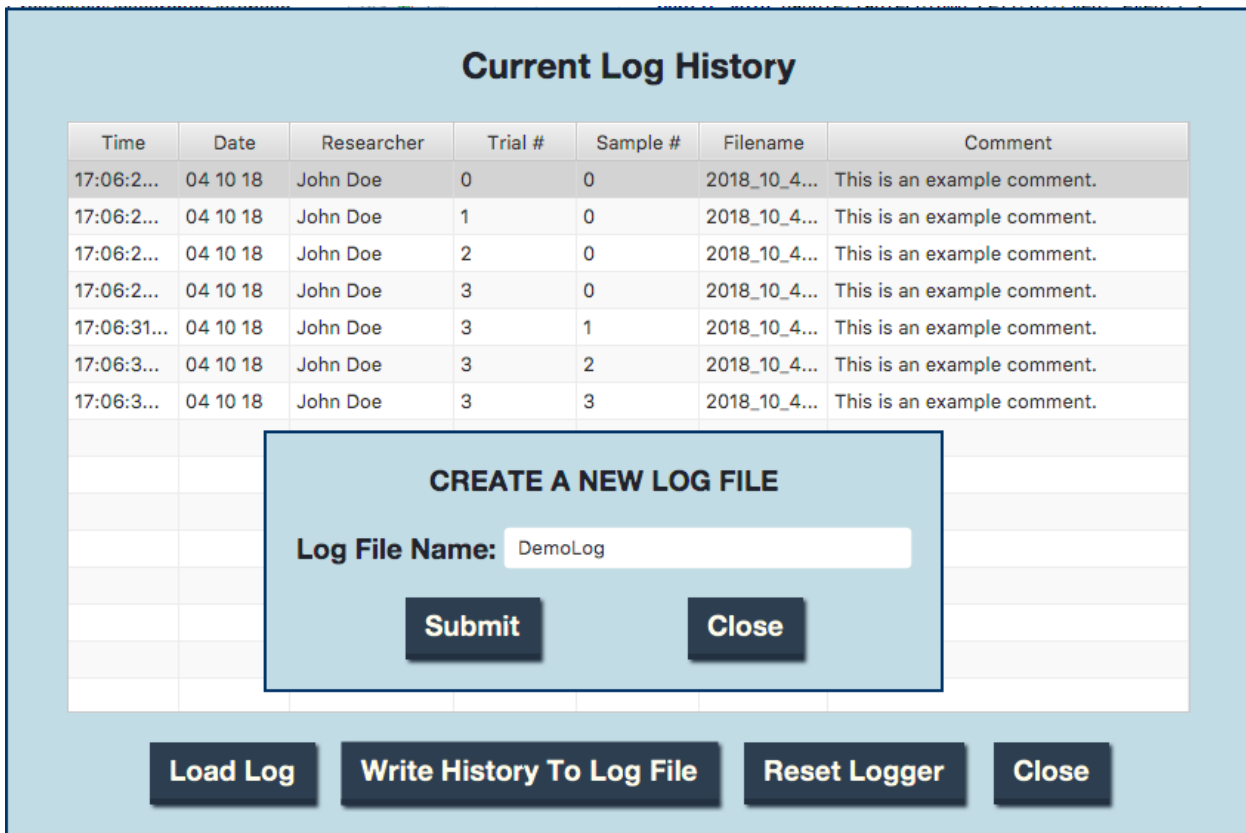


Figure 21. Log File user interface

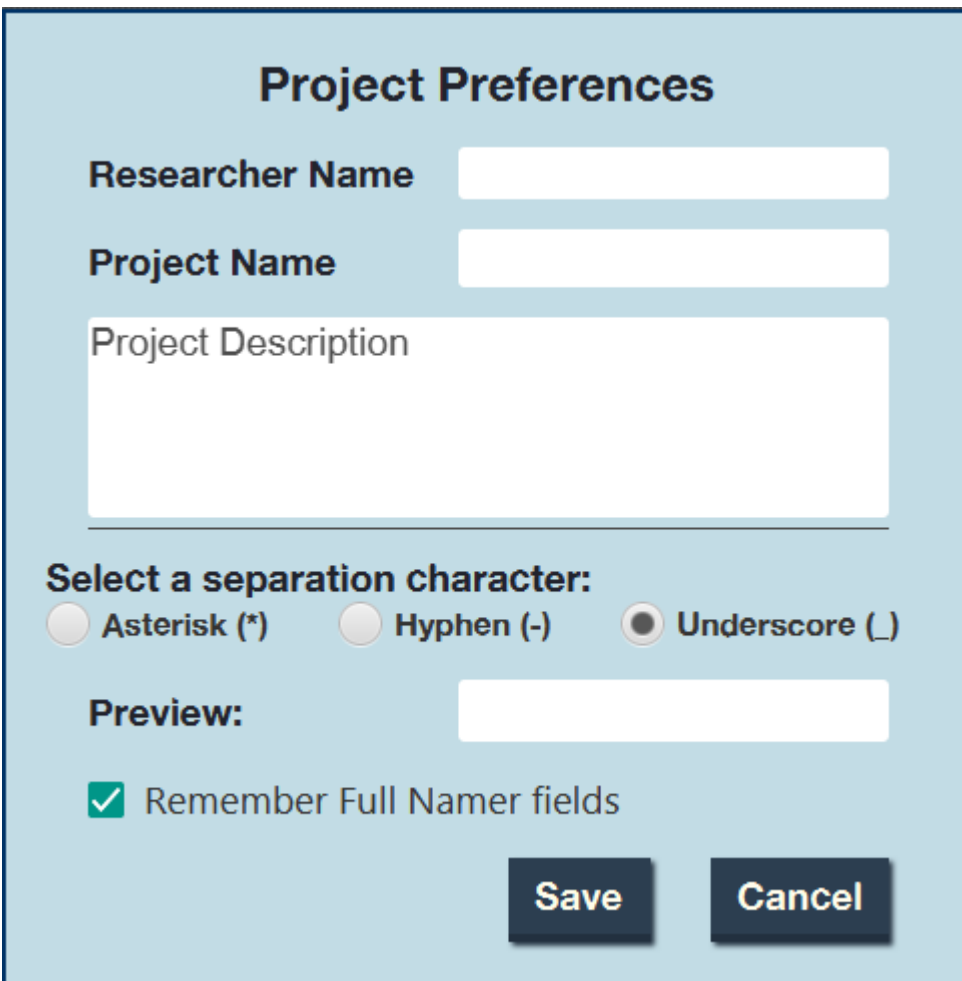
The log file was a frequently requested feature by the researchers because it enables them to quickly determine the project timeline. By creating a new log file, researchers can export the log to Excel, allowing for easy storage and reference (Figure 22).

DATE	TIME	RESEARCHER NAME	EXPERIMENT TYPE	TRIAL NUMBER	SAMPLE NUMBER	FILE NAME	COMMENTS
04 10 18	16:42:31	James Bond	Generic Experiment	1	1	2018_10_4_JB_T1_S1_2hrs	
04 10 18	16:42:33	James Bond	Generic Experiment	2	1	2018_10_4_JB_T2_S1_2hrs	
04 10 18	16:42:34	James Bond	Generic Experiment	3	1	2018_10_4_JB_T3_S1_2hrs	
04 10 18	16:42:36	James Bond	Generic Experiment	4	1	2018_10_4_JB_T4_S1_2hrs	
04 10 18	16:42:38	James Bond	Generic Experiment	5	1	2018_10_4_JB_T5_S1_2hrs	Error in this trial, needs to be re-done
04 10 18	16:42:44	James Bond	Generic Experiment	1	2	2018_10_4_JB_T1_S2_2hrs	
04 10 18	16:42:46	James Bond	Generic Experiment	2	2	2018_10_4_JB_T2_S2_2hrs	
04 10 18	16:42:48	James Bond	Generic Experiment	3	2	2018_10_4_JB_T3_S2_2hrs	
04 10 18	16:42:50	James Bond	Generic Experiment	4	2	2018_10_4_JB_T4_S2_2hrs	
04 10 18	16:42:52	James Bond	Generic Experiment	5	2	2018_10_4_JB_T5_S2_2hrs	

Figure 22. Example Excel log file

Additional application functionality includes the ability to add project preferences (Figure 23). The project preferences option allows users to choose which delimiter they would like to use in the file name as well as save the following parameters:

- Researcher name
- Project name
- Project description



*Figure 23. Project Preferences user interface*

Additionally, researchers can select whether the information inputted into the Full Namer fields should be stored when the application is closed. The information saved in project preferences will automatically populate the fields in the file naming screens, saving the researcher time and effort since they only need to input all the information once per project.

Additionally, while the file naming application, File Namer Pro, has an intuitive design, the team acknowledged that researchers may have difficulty downloading, installing, and using it

without any guidance, and therefore may not use it. To reduce the chance of this happening, the team developed a user manual for the application, which can be found in Appendix AA. Any confusion or questions that researchers had while testing File Namer Pro were taken into consideration during the development of the user manual to best address any confusion that may arise in the future. The user manual includes the following chapters, which contain detailed instructions and screenshots to aid the user:

- Chapter 1: Introduction
- Chapter 2: Features
- Chapter 3: Download and Launch Instructions
- Chapter 4: Quick Start Guide
- Chapter 5: Adding to the Databases
- Chapter 6: Modifying the Code

The team recommends that the application is used uniformly throughout research groups to increase the level of consistency among file names.

#### *5.1.2.6 Further Development*

The current functionality of the application aims to satisfy the basic needs of the researchers and provides opportunities for further development. The ORL research group identified the following features as desired, but not necessary:

- Users can select the order in which information appears in the file name
- Extra characters in the file name truncate or turn grey
- Comments on each experiment or project can be edited in a full text editor
- The application includes an integrated file manager which can directly read, lookup, and write files
- Additional keywords and descriptions are included in the file details rather than the file name.

In order to enable further development, the team provided clearly documented code and the user manual. Appendix Y provides information about the location of the code and indicates the different programs a user would need to download in order to make modifications.

## Chapter 6: Conclusion

Both the best practice guidelines and the file naming application work to improve data management and organization at USZ. The best practice guidelines provide general recommendations for improving practices in the five key areas of data management, and the file naming application provides targeted improvements to file naming. By using the guidelines and the application in conjunction with each other, data organization at the hospital can be improved, further facilitating the research process.

Because data management is such a broad field, this project provided targeted improvements to file naming well as general guidelines covering the five key areas identified: documentation, file naming, file organization, data transfer, and data storage. Based on the analysis of the interview findings, future WPI Interactive Qualifying Projects (IQPs) can take several forms. First, the team would recommend an IQP that would take a closer look at the relationship between the IT department and researchers. Miscommunication between these two groups was a common problem identified during interviews, but fully addressing the issue was outside the scope of the team's project. Secondly, data management challenges are not unique to USZ. The team interviewed researchers from several other institutions who faced similar difficulties, and future IQPs could perform a similar project premise at these facilities. Thirdly, improvements on the application can either be made by engineers at USZ or by future WPI-USZ collaborations. Specifically, a Computer Science Major Qualifying Project (MQP) team could partner with USZ to further develop the application.

In conclusion, the team hopes that the Research Data Management Handbook and File Namer Pro supply USZ with solutions for improving research data management and provide useful information for potential future projects.

## References

- About us (2014, October 28). Retrieved from <http://www.en.usz.ch/about-us/Pages/default.aspx>
- Apache Derby. (n.d.). Retrieved from <https://db.apache.org/derby/>
- Coulehan, M. B., & Well, J. F. (2006). *Guidelines for Responsible Data Management in Scientific Research*. Chapel Hill, NC: Clinical Tools
- Data Storage: An Evolution. (2016, November 18). Retrieved from <https://www.networkcomputing.com/data-centers/data-storage-evolution/218664820>
- Facts and figures (2017, May 1). Retrieved from <http://www.en.usz.ch/about-us/pages/facts-and-figures.aspx>
- History (29, May 2017). Retrieved from <http://www.en.usz.ch/about-us/pages/history.aspx>
- History of the University of Zurich (2017, December 11). Retrieved from <http://www.uzh.ch/en/about/portrait/history.html>
- Inouye, D., & Scheiner, S. (2009). *Some Simple Guidelines for Data Management*. Retrieved from <https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.1890/0012-9623-90.2.205>.
- JFoenix. (n.d.). Retrieved from <http://www.jfoenix.com/>
- Kanaparthi, V. (2018, February 14). Git and GitHub in a NutShell. – codeburst. Retrieved from <https://codeburst.io/git-and-github-in-a-nutshell-b0a3cc06458f>
- Narrative Research - Analysis of Qualitative Data - Design & Method. (n.d.). Retrieved from <https://atlasti.com/narrative-research/>
- Ranking Web of Hospitals. (n.d.). Retrieved from <http://hospitals.webometrics.info/en/europe/switzerland%20>

Swiss National Science Foundation. (n.d. a). Data Management Plan (DMP) - Guidelines for researchers. Retrieved from

[http://www.snf.ch/en/theSNSF/research-policies/open\\_research\\_data/Pages/data-management-plan-dmp-guidelines-for-researchers.aspx](http://www.snf.ch/en/theSNSF/research-policies/open_research_data/Pages/data-management-plan-dmp-guidelines-for-researchers.aspx)

Swiss National Science Foundation. (n.d.b). Open Research Data. Retrieved from

[http://www.snf.ch/en/theSNSF/research-policies/open\\_research\\_data/Pages/default.aspx](http://www.snf.ch/en/theSNSF/research-policies/open_research_data/Pages/default.aspx)

Szolovits, P. (n.d.). Medical Informatics: Computer Applications in Health Care.

Retrieved from <https://groups.csail.mit.edu/medg/courses/6872/96/notes/Tsien/index.htm>

University Hospital Zurich. (2013, March 6). Retrieved from

<http://www.swisshealth.ch/en/patienten/spitaeler/usz.php>

University Hospital Zurich (2014) *University Hospital Zurich grounds* [Photograph] Retrieved

from <http://www.en.usz.ch/media/pages/downloads.aspx>

Wilkinson et al. (2016), The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data* 3, doi:10.1038/sdata.2016.18

Willkommen am UniversitätsSpital Zürich (2015, October 13). Retrieved from

<http://www.vis.usz.ch/fachwissen/morbide-adipositas/Seiten/geschichte.aspx>

Zurich is not all about Money – here are 9 Biotechs proving it! (2017, May 18).

<https://labiotech.eu/zurich-money-9-biotechs-ecosystem/>

## Appendix A: Pre-Interview Team Introduction

*The purpose of this document was to introduce the team and the project to interviewees.*

Gruezi!

Our names are Vandana Anand, Carlisle Ferguson, Caitlin Kyaw, and Jeffrey Xiao. We are a project team from Worcester Polytechnic Institute, a university in the United States. We are working in partnership with USZ to evaluate current data management practices at USZ with the goal of establishing good practice guidelines for data management.

Thank you for volunteering your time to speak with us. If you wish to remain anonymous, we would be happy to accommodate you and will redact any identifying information from this interview in our records and reports. Before we begin, do you have any questions for us?



## Appendix B: Sponsor Description

*This document provides additional information about the project sponsor.*

UniversitätsSpital Zürich (University Hospital Zurich or USZ), is a teaching hospital located in Zürich, Switzerland. The mission of UniversitätsSpital Zürich, in collaboration with the University of Zurich medical school, is to give excellent service to patients by providing fundamental medical care using cutting-edge medicine while ensuring that the hospital remains at the forefront of medical research in Switzerland and continues to maintain an outstanding international reputation (“About us”, 2014).

As one of five university hospitals in Switzerland and one of the largest and most important teaching hospitals in Europe, UniversitätsSpital Zürich has made great strides in health care, education, and research (“University Hospital Zurich”, 2013). USZ traces its impressive origins to the first hospital in the canton of Zürich, the Heiliggeist-Spital, which was founded in the 1200s by Duke Berthold. In 1833, the canton combined several of its colleges to form the University of Zürich, where 98 medical students were enrolled in the first year of operation. The Heiliggeist-Spital soon moved to the present-day USZ site in 1842, and in 1953, the main hospital building opened (“History”, 2017).

Today, USZ employs more than 7,800 employees from 87 different countries. This number includes approximately 1,400 doctors and scientists, 2,500 nursing professionals, 800 medico-therapeutic and medico-technical employees, and 2,000 employees in non-medical and support areas, which include finance, IT, facility management, and more (“Facts and figures”, 2017). Each year, the hospital has about 500,000 outpatients and more than 41,000 inpatients between over 40 departments and institutes. USZ is one of the largest hospitals in Switzerland with approximately 300,000 inpatient days, 950 hospital beds, and an average of nearly 3,000 births a year. Furthermore, in 2016, hospital collected 1,326 million Swiss francs (CHF) in revenue with a profit of CHF 86.2 million.

As a university hospital, USZ actively participates in medical research and clinical trials. In 2016, The University of Zurich contributed CHF 72 million towards research and teaching programs at the hospital to support current and future endeavors. USZ has pioneered many of the latest treatment methods and seeks to provide therapies to patients from around the world with complex health problems. Moreover, in 2008, the Institute of Primary Care opened to educate

students in medical school about the work and responsibilities of general practitioners (“History of the University”, 2017). In addition to these duties, medical professionals at USZ are tasked with researching and experimenting with diagnostic and therapeutic approaches to improve patient lifespan and quality of life. The hospital also engages in other research in regenerative medicine, personalized medicine, highly specialized medicine, nursing care innovation, and health services innovation. Because USZ conducts research in so many different areas, vast quantities of data are created that must be stored, organized, and processed.

### **What is the concern?**

The development of new technology has allowed new and novel medical research and experiments to be conducted, not only at USZ but around the world as well. At the same time, collecting and processing of experimental research data has been improved and simplified due to the growth of commercially available computing power. As a result, researchers are conducting more and more experiments. However, these increases in experimentation have resulted in the generation of growing amounts of data ranging from tens to hundreds of gigabytes of data per experiment. Digital files in these experiments include, but are not limited to, pictures and videos from microscopes, data from micro-computed tomography scans, and x-rays. Other data types, such as verbal discussions and handwritten notes, are highly valuable to researchers as well.

Since the collected data holds such importance, proper storage, management, and organization of these materials are crucial to the processing, understanding, and eventual publication of the research. USZ conducts countless experiments and research projects concurrently, meaning this large amount of data is often not stored, organized, or managed efficiently. This lack of organization can make it difficult to process and interpret the data.

### **Where does this problem fit into the purpose of the organization?**

USZ strives to be at the forefront of medical research, making their ability to process data vital to maintaining their reputation as an elite university hospital. The creation of a standardized procedure or set of guidelines detailing a set process for handling data for a range of applications would greatly benefit the hospital. Therefore, creating an efficient way to manage data is paramount to aiding USZ in advancing and achieving its goals and sustaining its stellar reputation.

**Analogous efforts:**

Organizations embarking on similar research efforts as USZ include other university hospitals in Switzerland, which include Geneva University Hospital, University Hospital Inselspital Bern, Lausanne University Hospital, and University Hospital of Basel (“Zurich is not all about Money, 2017”). Large biotech and pharmaceutical companies, such as Octapharma in Zürich, likely face similar challenges and presumably have established mechanisms for data storage, organization, and management.

## Appendix C: Interview Template, Spring 2018

*This interview template was used for interviews conducted in the spring of 2018.*

**Interviewer:**

**Interviewee:**

**Interviewee's Title and Organization:**

**What field and/or what type of research?**

**What type of data (paper, digital images, numerical data sets, etc.) and how is it collected?**

**How is the data organized/stored after it is collected?**

**How is the data shared with other individuals in the organization?**

**What type of analysis is done on the data, and how does that type of analysis work?**

**Is the current organization system easy to use? Is it efficient?**

**What parts of the system do you like the most? The least?**

**In terms of time, how much effort are you willing to put in to organize the data?**

**Are you able to pick up the project several weeks/months later and be able to figure out where you left off? Are outsiders to the project able to? How easy, or how difficult?**

**In an ideal world, how could the organization system be improved?**

**Can you refer us to any other researchers who may offer additional insight?**

## Appendix D: Interview Template, Fall 2018

*This interview template was used for interviews conducted in the fall of 2018.*

**Interviewer:**

**Interviewee:**

**Interviewee's Title and Organization:**

**What programming languages/software do you use in daily work, if any?**

**What are the documentation standards for the current software/organization system?**

**Do you typically take notes while doing research? Are they on paper/computer/other? How are these notes organized?**

**Is there a standardized operating procedure (if so, what is it) or does each researcher use their own organization method?**

**How much time are you willing to spend documenting/organizing your research?**

**What information do you include in file names? What sort of information helps you easily identify a file's contents?**

**What parameters do you use to differentiate between experiments?**

**In performing a single instance of an experiment, what tends to be the general workflow in terms of gathering data, organizing data, and doing write-ups?**

## Appendix E: Dr. Dirk Smolinski Interview Summary

From 13 April 2018

*Dr. Dirk Smolinski is the Head of Services for Clinical Research at the UniversitätsSpital Zürich (USZ). He currently is on the Data Management Team focusing on data handling procedures and managing data structure software. The following questions were used to guide the interview, and our findings from the interview are summarized below.*

1. What type of data (paper, digital images, numerical data sets, etc.) is stored and how is it collected?
2. How is the data stored after it is collected?
3. What type of analysis is done on the data, and how does that type of analysis work?
4. Is the current organization system easy to use? Is it efficient?
5. What parts of the system do you like the most? The least?
6. How is the data shared with other individuals in the organization?
7. Does the software you're using allow users to pick up the project several weeks/months later and be able to figure out where they left off? Can users that were not part of the original project do so as well?
8. In an ideal world, how could the organization system be improved?

### Findings:

Dr. Dirk Smolinski explained that his team only uses electronic data capture systems in order to collect data. Once data is collected, it is stored in one of two softwares. SecuTrial is used for more complicated clinical trials while REDcap is used for simple trials. Software such as SPSS and R are used to analyze the data. Dr. Smolinski stated that sharing and transferring data is complicated because of data protection laws, but data can be exported easily using SPSS or R. Exported data is then stored in a different location.

The current data organization system has fifteen standard operating procedures which describe different aspects of the software. Dr. Smolinski mentioned that the main issue with this system is the layout. It is outdated, but is still functional and meets the requirements. In an ideal environment, it would be useful to have more tools to capture deviations in the software and display or visualize data through graphs and statistics during the trial.

## Appendix F: Rita Fagan Interview Summary

From 14 April 2018

*Ms. Rita Fagan is a Ph.D. student at the University of Massachusetts Medical School in Biomedical Sciences. Ms. Fagan works in a small research lab at UMMS, and her research focuses on neurobiology. The following questions were used to guide the interview, and our findings from the interview are summarized below.*

1. What type of data do you collect, and how is it collected?
2. How is the data stored and organized after it is collected?
3. How is the data shared with other individuals in your organization?
4. Is the current data organization system easy to use? Is it efficient?
5. Can you go back to a research project several weeks or months later and easily determine where you left off? Can others do this as well?
6. In an ideal world, how could the data organization system improve?

### Findings:

Ms. Fagan typically handles numerical data sets and uses software such as MATLAB and Excel to sort through the data. Within the lab she works in, all researchers can access the network where the data is stored. Researchers from other labs are able to request access to view the files, but this is a rare occurrence. In addition to recording their data in the computer system, researchers maintain paper lab notebooks which provide further details on their research.

Ms. Fagan explained that the current organization and storage system is easy to use, but accessibility could be improved. The system is efficient overall since data can be imported directly, and the researcher does not have to “copy data into the computer.”

One of the biggest strengths of the system is the presence of the paper lab notebooks as they provide information on negative or unsuccessful trials that often do not make it into the final report. A secondary benefit of the system is that it is theoretically easy to come back to a project after a hiatus. The ease of access can be hindered, however, by messy or incomplete notes.

Ms. Fagan indicated two weaknesses in the system. Firstly, to look up specific details from an experiment one must look up the notebook’s location in the computer system, then

locate the physical copy of the notebook. After the notebook is located, the researcher needs to search through the notebook to find the specific experiment. The second drawback of the system is that brainstorming and scratch notes can easily become lost. There is no current procedure to document notes taken outside of lab notebooks, so they often end up recycled.

The primary area of improvement mentioned by Ms. Fagan was digitization. Some digitization softwares create a large time burden, which disincentives their usage. Moving forward, electronic lab notebook software or tablets may be an option, but currently, many lab leaders tend to favor the traditional paper-based methods.



## Appendix G: Prof. Dr. Rudolf Probst Interview Summary

From 17 April 2018

*Emeritus Prof. Dr. Rudolf Probst is the former Chairman of the Department of Otorhinolaryngology (ORL) at UniversitätsSpital Zürich and is currently acting as an advisor as the research section of the ORL department undergoes reorganization. He also conducts his own research through clinical trials. Due to time constraints, the team could not conduct an interview, however, Prof. Dr. Probst was sent a list of questions, to which he responded to by email. The following summarizes his responses.*

Through his clinical trials, Prof. Dr. Probst generates many types of data, all of which is stored electronically. Any paper data must be transferred to electronic storage, usually by scanning.

Patient data is stored in the patient-related database of the hospital through electronic patient records. Access to this data is controlled by the hospital, and the database records any changes made to files and documents who is viewing them.

Research data is stored in a database assigned to the department. Generally, it is organized by the research group and research project. Projects will organize their database according to their specific needs, and access is limited to a group of people defined by the group or project leader. Any changes made to files and viewers are also tracked.

Prof. Dr. Probst notes that the individual systems and methods of organizing and storing data can be made easier and more efficient; however, the current focus of the hospital and departments as a whole is on security. This requires any definitions and permissions of access to go through the administration, which often can lead to unnecessary delays.

Ideally, during the experimentation phase, every effort is taken to organize the data. This should mean that minimal effort will be necessary for the processing phase. The current aim is that projects can easily be picked up after a hiatus and that a standardized system to organize data is established.

## Appendix H: Dr. Flurin Pfiffner Interview #1 Summary

From 20 April 2018

*Dr. Flurin Pfiffner is a researcher in the ORL department of USZ. He is also an audiologist, and his focus is on research with patients for clinical studies. The following questions were used to guide the interview, and the findings from the interview are summarized below.*

1. Which types of data (paper, digital images, numerical data sets, etc.) do you collect and how is it collected?
2. How is the data organized/stored after it is collected?
3. Which aspects of the current data management system or process do you like the most? The least?
4. Are you able to pick up project several weeks/months later and be able to figure out where you left off? Are outsiders to the project able to? How easy or difficult is this?
5. In an ideal world, how could the organization system be improved?

### Findings:

Dr. Pfiffner collects several different types of data, including postoperative data, audiological data, questionnaires for patients, and CT scans. This data is stored on a program/system provided by the hospital called KISIM and is shared with all of the people involved with the project.

Noted benefits of this system, KISIM, include the following:

- All records from as far as forty years ago are accessible
- A lot of data can be stored
- Documentation is uniform

Noted drawbacks of this system include the following:

- Not designed for research, and was mainly developed for work with patients
- Data cannot be exported easily or automatically (if you want to have the information from an Excel sheet on another computer or external folder, you need to manually type in the information)
- Extremely well-protected, difficult to give additional access

At the time, it is difficult to pick up projects after a hiatus. Understandably, a small amount of communication is required in order to be fully caught up on the progress of a project, however, in many cases, documentation and organization are often not even consistent within a project, much less across projects. In these situations, because there is no general or standardized method, a new member of the project would require frequent follow-up to understand the organization and documentation system.

Currently, each project has its own folder, which is already a step in the right direction. To improve the organizational system, there needs to be a consensus on the best way to organize the data within the folder. Additionally, a way to obtain an overview or short report of the details and parameters of an experiment or measurement, the number of experiments that were conducted, and whether or not an experiment worked and why would be greatly beneficial.

## Appendix I: Prof. Dr. Raoul Hopf Interview Summary

From 24 August 2018

*Raoul Hopf currently works in the Department of Mechanical and Process Engineering at Eidgenössische Technische Hochschule (ETH) Zürich, focusing on modeling and experimental work on soft materials, including human tissue. The team corresponded with him over email, and the following is a summary of the information he provided about how his team manages experimental data.*

The data acquired in the labs is saved on large disks, which are then loaded onto computers that are disconnected from any network drives. These computers are not connected to the network drives because the information technology (IT) specialists at ETH do not want to clutter online storage spaces.

Necessary data from these drives is then transferred to temporary secure drives where everything is processed. Once the data is processed and a paper is written, any of the raw data, processed data, and simulation input/output data that is included in the publication is transferred onto a secure and backed-up online drive. This drive is designed to hold and protect the data long-term, as it is published work and needs to be accessible.

In the rare case that a project would concern patient data, the necessary steps are taken to ensure the data is anonymized and secure in accordance with the standards of the law and guidelines set by the collaborating hospital.

## Appendix J: Prof. Dirk Albrecht Interview Summary

From 31 August 2018

*Prof. Dr. Dirk Albrecht is a Biomedical Engineering professor at Worcester Polytechnic Institute (WPI). He also runs a research lab based on microtechnology and neuroscience and organizes data management procedures for various experiments. The following questions were used to guide the interview, and our findings from the interview are summarized below.*

1. What programming languages/software do you use in daily work, if any?
2. What are the documentation standards for the current software/organization system?
3. Do you typically take notes while doing research? Are they on paper/computer/ other?  
How are these notes organized?
4. How much time are you willing to spend documenting/organizing your research?
5. What information do you include in file names? What sort of information helps you easily identify a file's contents?
6. What parameters do you use to differentiate between experiments?
7. If you had to identify the two or three biggest obstacles impeding efficient data processing, what would they be?

### Findings:

Prof. Dr. Albrecht mentioned that his lab mostly focuses on imaging and video data that is typically analyzed by MATLAB. Currently, the documentation standards in the organization revolve around automatically naming experiment data files with usually only the date, but could also sometimes have the time, conditions, and data point. The most common practice is to try to get the user to document the experiment although there is no official requirement to do this.

Prof. Dr. Albrecht said that the notes taken for research are mostly electronic. The main focus of these notes is to document the purpose of the experiment and the preparation needed to run it. The data is organized in file structures in a specific network drive where every lab and researcher have a folder and permissions and access are managed. There is a new implementation of a structure that acts more as a cloud storage system, but the data is stored on a local drive so there is a risk of losing it. In the lab, Prof. Dr. Albrecht and his team tried to

implement a structure in which they could compare repeated experiments that use different parameters each time, but soon found it would be difficult for someone new to the project to catch up.

When discussing what kind of information would be beneficial to have in a file name to identify its contents more easily, Prof. Dr. Albrecht explained that he prefers to have a numerical identifier that can be referenced for each experiment. In addition, he prefers adding the date in the file name by year, month, then day in order to be able to sort chronologically. In addition, he also adds the genotype, numerical time point number, and simulation number if applicable.

## Appendix K: Dr. Andreas Eckhard Summary

From 1 September 2018

*Dr. Andreas Eckhard is a clinical resident and research group leader at USZ focusing on neurotology and inner ear biology. The following questions were used to guide the interview, and our findings from the interview are summarized below.*

1. What programming languages/software do you use in daily work, if any? What are the documentation standards for that system?
2. Do you typically take notes while doing research? Are they on paper/computer / other? How are these notes organized?
3. Is there a standard operating procedure (if so, what is it) or does each researcher use their own organization method?
4. How much time are you willing to spend documenting/organizing your research?
5. What information do you include in file names? What sort of information helps you easily identify the file's contents?
6. What parameters do you use to differentiate between experiments?
7. In performing a single instance of an experiment, what tends to be the general workflow in terms of gathering data, organizing data, and doing write-ups?

### Findings:

Dr. Eckhard does not use software or programming languages in his daily work. He typically takes notes on paper, and a lab notebook is dedicated to each project. Any digital data is stored in project-specific files on hospital-owned servers. There are currently no standard operating procedures for documenting data in his lab. When naming files, Dr. Eckhard uses three main parameters: codes for experimental series, project names, and dates which refer to entries in his physical lab notebook. The workflow of his experiments is as follows:

- Planning phase, which includes defining experimental groups and coding them for the documentation
- Working on documentation sheets which are adapted to the experiment
- Performing the experiment

- Documenting the experiment

Dr. Eckhard stated that he is willing to spend as much time as is needed to properly document his work.



## Appendix L: Dr. Flurin Pfiffner Interview #2 Summary

From 4 September 2018

*Dr. Flurin Pfiffner is a researcher in the ORL department of USZ. He is also an audiologist, and his focus is on research with patients for clinical studies. The following questions were used to guide the interview, and the findings from the interview are summarized below.*

1. What programming languages/software do you use in daily work, if any?
2. What are the documentation standards for the current software/organization system?
3. Do you typically take notes while doing research? Are they on paper/computer/ other?  
How are these notes organized?
4. How much time do you spend documenting/organizing your research?
5. What information do you include in file names? What sort of information helps you easily identify a file's contents?
6. What parameters do you use to differentiate between experiments?
7. In performing a single instance of an experiment, what tends to be the general workflow in terms of gathering data, organizing data, and doing write-ups?
8. What are the current guidelines for documentation or organization of data at USZ?

### Findings:

Dr. Flurin Pfiffner primarily conducts clinical research, which uses KISIM, ENTstatistics, Excel, MATLAB, and Microsoft Access. For any non-clinical research that he conducts, text files and Excel spreadsheets are used.

In terms of documentation, researchers record the ethical considerations are taken, who owns the data, who has the key which anonymizes the data, and what roles people are designated must be made clear in a report. Additionally, the uses for the data must be documented by the research team and approved by the ethical committee at the hospital.

For clinical studies, all notes must be taken and kept on a clinical reporting form. This ensures that the data is only seen by authorized individuals, and that patient identities are kept safe.

For non-clinical studies, data is collected exclusively on a computer. Only those with pre-approved access may view the data folder and its contents. These folders are not organized in a uniform way, and the actual structure is dependent on the project and its needs. In the past, a template was created to try to standardize the way folders are organized. Folder labels included “Ethics”, “Budget”, “Data”, and “Presentation”. Unfortunately, this template was deemed inapplicable to many projects and eventually became unused.

In terms of file naming, Dr. Pfiffner generally tries to label data with a project acronym (generally around six letters long), the exact time an experiment was conducted, and the identification number assigned to a patient in the clinical study whose name is anonymized. Additionally, if there are measurements done, incremental numbers and the type of measurement done is included in the file name.

Dr. Pfiffner’s workflow depends on the project but usually begins with pilot studies. These are data-gathering experiments that are conducted simply to determine if certain procedures work and if subsequent testing should be performed. Once a pilot study begins to look promising, the procedure to be followed is planned out and written down so that a standardized protocol can be followed.

Although there are not any guidelines currently in place that dictate the documentation and organization of research data, Dr. Pfiffner has tried to maintain good practices when generating his data.

## Appendix M: Dr. Laurent Simon Interview Summary

From 5 September 2018

*Dr. Laurent Simon is a postdoctoral researcher working in the experimental audiology lab at USZ. His research centers around finding new localization methods for hearing aids. The following questions were used to guide the interview, and our findings from the interview are summarized below.*

1. What programming languages/software do you use in daily work, if any? What are the documentation standards for that system?
2. Do you typically take notes while doing research? Are they on paper/computer / other? How are these notes organized?
3. Is there a standard operating procedure (if so, what is it) or does each researcher use their own organization method?
4. How much time are you willing to spend documenting/organizing your research?
5. What information do you include in file names? What sort of information helps you easily identify the file's contents?
6. What parameters do you use to differentiate between experiments?
7. In performing a single instance of an experiment, what tends to be the general workflow in terms of gathering data, organizing data, and doing write-ups?

### Findings:

Dr. Laurent Simon uses MATLAB frequently in his daily work, and he documents his code every line or every few lines. His documentation is clear enough that colleagues unfamiliar with MATLAB can easily figure out how each script works. There are no set rules in his lab for documenting code.

One of the problems Dr. Simon has experienced in his work is data transfer. In USZ's new networked drive system, access to project folders is restricted to certain researchers' computer accounts. The issue arises when the researchers' computers lack the software they need to conduct their tests, and the computer with the software doesn't have their user access. Dr.

Simon has tried to email the IT department, but communication is slow and they have misunderstood his request in the past. He has also tried to use GitHub, but downloading the necessary software has proven complicated, and getting help from the IT department is often a time-consuming process.

In terms of file organization, Dr. Simon uses the file directory extensively. He creates a folder for each experiment, then creates subfolders to further organize his work. This works well for smaller projects, but for bigger projects, the number of subfolders can become overwhelming. Dr. Simon's files are automatically named by MATLAB, and he will also include a small description of what the file contains. He stated that is easy for him to determine what is in the file just by looking at the file name.

# Appendix N: Dr. Nicole Peter-Siegrist Interview

## Summary

From 10 September 2018

*Dr. Nicole Peter-Siegrist is a medical doctor who conducts clinical research on single-sided deafness and tinnitus. The following questions were used to guide the interview, and our findings from the interview are summarized below.*

1. What programming languages/software do you use in daily work, if any?
2. What are the documentation standards for the current software/organization system?
3. Do you typically take notes while doing research? Are they on paper/computer / other?  
How are these notes organized?
4. Is there a standard operating procedure (if so, what is it) or does each researcher use their own organization method?
5. How much time are you willing to spend documenting/organizing your research?
6. What information do you include in file names? What sort of information helps you easily identify the file's contents?
7. In performing a single instance of an experiment, what tends to be the general workflow in terms of gathering data, organizing data, and doing write-ups?

### Findings:

Dr. Nicole Peter-Siegrist takes notes on the computer when conducting research. Because she works primarily on clinical research data, she must document everything in the patient's file, and overall there is more of a set procedure than non-clinical data because of patient protection laws. She also must fill out a case report form (CRF) which collects data from patients. This is usually done on paper, but there are electronic versions (ECRF) available. The paper version is typically used since the electronic version is more expensive and she needs to take time to program it. The documentation process for clinical data is very time consuming, and Dr. Peter-Siegrist tries to spend the minimum amount of time on it.

In order to organize files within the lab in her department, Dr. Peter-Siegrist uses the network drive as the primary file organization system. Because the networked drive is only accessible to USZ researchers, she has often had difficulties collaborating with outside research groups. Recently, a new system called CampusCloud was implemented for collaborative projects and it allows researchers to share files across institutions.

In regards to file structure, each project has a folder and similar analyses that are conducted on the data are stored in sub-folders. For example, as part of their current file structure, they have a trial master file which has information on how to legally handle data, a folder for data sets, a folder for analysis, and a folder for publication.

For file naming purposes, Dr. Peter-Siegrist prefers to have the date, a descriptive name describing the experiment, and an indication if the file is the most up-to-date version. Moreover, the lab group has a list of glossary terms with corresponding abbreviations for each possible experiment name. The glossary is not standardized throughout the hospital but it is standardized throughout the lab.

The general workflow that is followed in the clinical lab is to collect questionnaires via a software called Unipark that is completed by patients. Their data is then anonymized to maintain patient privacy. Then, this data is processed using software such as SPSS, Graphpad, or R. There is also a new software called ENTstatistics, which extracts all clinical data.

# Appendix O: Dr. Emine Cagin Interview Summary

From 10 September 2018

*Dr. Emine Cagin works at the Institute for Micro- and Nanotechnology at NTB Buchs. She works on interdisciplinary teams during her research. The following questions were used to guide the interview, and our findings from the interview are summarized below.*

1. Do you typically take notes while doing research? Are they on paper/computer / other?  
How are these notes organized?
2. Is there a standard operating procedure (if so, what is it) or does each researcher use their own organization method?
3. How much time are you willing to spend documenting/organizing your research?
4. What information do you include in file names? What sort of information helps you easily identify the file's contents?
5. What parameters do you use to differentiate between experiments?
6. Are you able to easily share data files with collaborators from other institutions?

## Findings:

Dr. Cagin's lab has a standardized system both for organizing and naming files, as well as for documenting data. The documentation of research starts at the brainstorming stage, and Dr. Cagin's team uses RealtimeBoard, which is an online whiteboard and collaboration website, to plan out projects. This system has worked well for them as it allows them to edit flow charts and templates while live on a call with industry partners.

Most of the notes taken in the lab are on paper, and they always include the date, the location, and the name of the person who ran the experiment. Once a day, they record process-specific parameters, such as temperature, humidity, and the start condition of the sample. Dr. Cagin and her team also create PowerPoint presentations as they go, which ensures they are always ready for a presentation with their project partners.

Documentation encompasses a significant portion of Dr. Cagin's work. While she only spends a few minutes at a time documenting her work, this is done multiple times a day, resulting in over an hour of time spent on documentation. Dr. Cagin explained that her lab

documents the right amount of information, and she wishes her team would spend less time documenting but still record the same amount of information. Her team is currently working on shortcuts, such as a print out of the procedures. If the procedures go exactly as planned, the researcher can put a checkmark instead of writing out in words what happened.

In terms of naming files, researchers include the date, the project name, some additional descriptors, and the initials of the person who last changed the file. This system works well and allows the researchers to quickly and easily identify what each file contains. Files used to be stored on SWITCH, a cloud storage system similar to Dropbox, but more heavily encrypted. This system did not work well since editing files were cumbersome. SWITCH only had the capability to upload and delete files, and researchers could not edit files on the server. It was also difficult to collaborate with industry partners or team members from other institutions.

Recently, they changed to a file storage system hosted by NTB Buchs. The new system provides substantial improvements from the old system, especially in terms of collaboration. Industry partners can have access to a predetermined path into the NTB server to a folder called ‘exchange’ which allows the outside user to work on files as if they are at NTB, provided that they sign a standard IT agreement. Cagin did not have any complaints about the new system and spoke highly of all the current systems in place at NTB Buchs.



# Appendix P: Dr. Mohammad Divandari Interview

## Summary

From 11 September 2018

*Dr. Mohammad Divandari is a researcher in the Department of Materials at Eidgenössische Technische Hochschule (ETH) in Zurich. His research is focused on reducing friction between surfaces as well as tailoring surfaces for biological purposes. The following questions were used to guide the interview, and our findings from the interview are summarized below.*

1. What programming languages/software do you use in daily work, if any? What are the documentation standards for that system?
2. Do you typically take notes while doing research? Are they on paper/computer / other? How are these notes organized?
3. Is there a standard operating procedure (if so, what is it) or does each researcher use their own organization method?
4. What information do you include in file names? What sort of information helps you easily identify the file's contents?
5. What parameters do you use to differentiate between experiments?
6. In performing a single instance of an experiment, what tends to be the general workflow in terms of gathering data, organizing data, and doing write-ups?

### Findings:

Dr. Divandari primarily uses MATLAB and Excel to process his data. As he writes the code, he tries to leave as many comments as possible so that he can remember what happens. However, there are no rules or guidelines on how to document code.

During experiments, Dr. Divandari takes notes in a paper lab notebook. Over the course of his career, his method of documentation has evolved. Now, he tries to document the name, date, and specific aims of the experiment. Additionally, because much of his work is based on chemical reactions, Dr. Divandari tries to include information on the ratios of the reactants, the

molecular weight of the compounds, their exact names, and structures. Occasionally, if necessary, he also documents the duration and temperature conditions of the experiment.

Dr. Divandari admitted that it is sometimes difficult to decipher his and his colleagues' lab notebooks. Handwriting is not always legible, notes are not written in a standardized, systematic way, and the level of detail is often inconsistent.

For electronic data, Dr. Divandari likes to organize folders and files by the date, the instrument used to acquire the data and the project to which it belongs. This is his personal preference and is not consistent with the practices of other lab members. When electronic data needs to be shared, Dr. Divandari often runs into the issue of not being sure through which medium data should be sent. Researchers have their own personal preferences, ranging from data transfer through email, Dropbox, and other third-party providers. Although he is flexible and adaptable when it comes to transferring data, he wishes there was a standardized method of doing so.

Additionally, Dr. Divandari is particularly interested in what the outcomes will be of this project, as he would like the university to adopt some sort of standardized method of documenting and organizing data.

# Appendix Q: Dr. Prof. Thomas Kleinjung Interview

## Summary

From 11 September 2018

*Dr. Prof. Thomas Kleinjung is a medical doctor, professor, and head of the outpatient and emergency department. He designs and supervises studies while conducting mostly clinical research. The following questions were used to guide the interview, and our findings from the interview are summarized below.*

1. What programming languages/software do you use in daily work, if any?
2. What are the documentation standards for the current software/organization system?
3. Do you typically take notes while doing research? Are they on paper/computer / other?  
How are these notes organized?
4. Is there a standard operating procedure (if so, what is it) or does each researcher use their own organization method?
5. How much time are you willing to spend documenting/organizing your research?
6. What information do you include in file names? What sort of information helps you easily identify the file's contents?
7. In performing a single instance of an experiment, what tends to be the general workflow in terms of gathering data, organizing data, and doing write-ups?

### Findings:

In clinical research studies, there are strict guidelines on how to ethically document data. Dr. Kleinjung does not personally take notes while during research, but he supervises Ph.D. students who do. He stated that the students tend to take notes electronically.

Dr. Kleinjung's lab has a standardized operating procedure for certain measurement types. For example, the lab has flow chart templates for EEGs, PET scans, or MRI scans. The data is documented immediately after the completion of the procedure and tends to be difficult and time-consuming.

While the file name tends to vary between experiments, Dr. Kleinjung always includes the date and an abbreviation of the researcher's name. Additionally, he finds it beneficial to include general patient information such as age or gender. In terms of other abbreviations, the research group has a standardized glossary of each term and its respective abbreviation, which improves communication within the group.

In terms of his general workflow, Dr. Kleinjung collects data from patients that are fit for the experiment. The data is anonymized and subsequently analyzed via software such as SPSS, R, and Excel. The results and statistics are retrieved, compared, and discussed among the team.

Data collection tends to be organized due to the restrictions and regulations on clinical research, but data processing is unorganized, especially by Ph.D. students.

## Appendix R: Prof. Dr. Stefan Bertsch Interview Summary

From 19 September 2018

*Prof. Dr. Stefan Bertsch is the Head of the Institute for Energy Systems IES and a professor specializing in thermodynamics and refrigeration at NTB Buchs. The following questions were used to guide the interview, and our findings from the interview are summarized below.*

1. What programming languages/software do you use in daily work, if any? What are the documentation standards for that system?
2. Do you typically take notes while doing research? Are they on paper/computer / other? How are these notes organized?
3. Is there a standard operating procedure (if so, what is it) or does each researcher use their own organization method?
4. How much time are you willing to spend documenting/organizing your research?
5. What information do you include in file names? What sort of information helps you easily identify the file's contents?
6. What parameters do you use to differentiate between experiments?
7. In performing a single instance of an experiment, what tends to be the general workflow in terms of gathering data, organizing data, and doing write-ups?

### Findings:

In terms of software and code, Prof. Dr. Stefan Bertsch's lab has no real documentation standards in place, however, they do have a system where new modules that are programmed are documented and put in a centralized code library for others to use. Although documentation exists, Dr. Bertsch commented that it is occasionally difficult to understand the code completely.

Notes are taken in both paper and digital form. Day-to-day paper notes are often written in a notebook that is thrown away. Dr. Bertsch stated that throwing away these notes has not been an issue, as those are usually short-term. Any notes that are written in lab notebooks are kept, but there is no centralized storage location for them.

Electronic notes are located in a “Notes” folder in each project. The contents of these notes are not searchable unless the file is open, so anyone new to a project must open the file itself to understand what is happening.

Dr. Bertsch’s lab has a standardized folder organization. Each project has separate folders for the following: legal documents; documentation, protocols, and intermediate reports; ordering; simulation and measurement results; literature; pictures; and email correspondence. This file structure is consistent across projects. To maintain flexibility, folders can be added, however, the group tries to keep to this structure as much as possible.

In terms of file names, Dr. Bertsch and his team have different criteria for different file types. Literature files tend to have the year of publication and three of the most important keywords in the label. For notes, the file name generally includes the date and the protocol number. The way data files are named varies greatly between people, but Dr. Bertsch tends to include the date, a test matrix, and any test matrix data points he feels are relevant.

## Appendix S: Iteration One Product Backlog

<b>Iteration 1</b>			
<i>User Stories</i>	<i>Size</i>	<i>Diff.</i>	<i>Prior.</i>
As a researcher, I want to name my files by multiple parameter types	S	2	5
As a researcher or administrator, I want to be able to modify or add keyword types	M	6	5
As a researcher, I want the timestamp to be automatically included in my file name	S	3	5
As a researcher, I want the software to automatically keep track of what test I'm on even after the application is closed	S	3	4
As a researcher, I want to select a destination to write my file	S	2	2
As a researcher, I want to select the order in which information appears in my file name	S	2	3
As a researcher, I want to save my project preferences between different experiments	M	4	3
As a researcher, I want to quickly increment file names when I'm recording data consecutively	M	4	5
As a researcher, I want a readme glossary of all the terms and their descriptions	S	3	3
As a researcher, I want the program's name display to change while I'm making changes to the box's names	M	7	5

## Appendix T: Iteration Two Product Backlog

<b>Iteration 2</b>			
<i>User Stories</i>	<i>Size</i>	<i>Diff.</i>	<i>Prior.</i>
I want to be able to modify and add new keywords	M	8	4
I want to be able to input a value in addition to a keyword if necessary	S	2	3
I want the software to automatically keep track of what trial number/sample number I'm on even after the application is closed	S	4	2
I want the researcher name to appear in the file name	S	2	4
I want to select a destination to write my file	S?	2	1
I want to select the order in which information appears in my file name	S	2	2
I want to save my project preferences between different experiments	L	10	3
I want a readme glossary of all the experiment types and their descriptions (toggle)	M?	?	2
I want the file name preview to display in real time (observer pattern)	S	4	5
I want to choose what sort of character separates the date, choosing between underscore, asterisk, or dash.	S	2	4
I want to generate a log file with comments and detailed timestamps whenever a file name is created and copied	M	8	4
I want to have a maximum length for my file name and have it visually demonstrated via truncation.	M	5	3
I want extra characters in my file name to turn grey instead of truncating.	L?	?	1
I want to translate a file name back into the original parameters	M	8	3



I want a text box with a scroll bar where I can input a description of the experiment.	S	1	4
I want to select specific options from a drop-down menu on an additional keyword	L	10	2
I want to edit my comments in a full-text editor	L	?	2
I want to have an integrated file manager to directly read, write, and lookup files	XL	20	3
I want an additional input field where I can record which sample I'm working on.	S	1	4
I want plus and minus buttons on both the trial number and the sample number so I can easily increment and decrement it in simple namer program.	S	2	3
I want to easily toggle between simple namer and full namer.	S	1	2
I want to put additional keywords and descriptions into the file details rather than the file name	XL?	???	2

## Appendix U: Iteration Three Product Backlog

<b>Iteration 3</b>			
<i>User Stories</i>	<i>Size</i>	<i>Diff.</i>	<i>Prior.</i>
I want to be able to modify and add new keywords	M	8	4
I want the software to automatically keep track of what trial number/sample number I'm on even after the application is closed	S	4	2
I want to select a destination to write my file	S?	2	1
I want to select the order in which information appears in my file name*	S	2	2
I want to save my project preferences between different experiments	L	10	3
I want a readme glossary of all the experiment types and their descriptions (toggle)	M?	?	2
I want to generate a log file with comments and detailed timestamps whenever a file name is created and copied	M	8	4
I want to have a maximum length for my file name and have it visually demonstrated via truncation.	M	5	3
I want extra characters in my file name to turn grey instead of truncating.	L?	?	1
I want to translate a file name back into the original parameters	M	8	3
I want to select specific options from a drop-down menu on an additional keyword	L	10	2
I want to edit my comments in a full-text editor	L	?	2
I want to have an integrated file manager to directly read, write, and lookup files	XL	20	3
I want to put additional keywords and descriptions into the file details rather than the file name	XL?	???	2

I want to only have one instance of the application open at a time. Trying to start the application again should focus on the current application window.	M	6	3
I want to use T and S in my file name to differentiate between trial and sample number so I don't get confused	S	1	4
I want simple namer to act as a more compact version of full namer and include all the same information in the file name	S	3	5
I want all my keywords to appear in a scrollable box.	M	5	5
I want to select from multiple libraries of predefined keywords	M	6	3
I want to have a deployable stand-alone application/installer	M	5	5

## Appendix V: Iteration Four Product Backlog

<b>Iteration 4</b>			
<i>User Stories</i>	<i>Size</i>	<i>Diff.</i>	<i>Prior.</i>
I want to select the order in which information appears in my file name	S	2	2
I want a readme glossary of all the experiment types and their descriptions (toggle)	S	3	2
I want to have a maximum length for my file name and have it visually demonstrated via truncation.	M	5	3
I want a database for researcher names	M	4	3
I want to translate a file name back into the original parameters	M	8	3
I want to select specific options from a drop-down menu on an additional keyword	L	10	2
I want to edit my comments in a full-text editor	L	?	2
I want to have an integrated file manager to directly read, write, and lookup files	XL	20	3
I want to put additional keywords and descriptions into the file details rather than the file name	XL	???	2
I want the app to work on windows and on mac	L	10	4
I want the app not to require any special permissions to launch, download, or install	M	6	4
I want to be able to load a previous log file via a pop-up window	S	2	5
I want to be able to create a new log file via a pop-up window	S	2	5
I want to be able to add a keyword to the database inside the app	M	4	5

I want to be able to load a keyword library via project preferences	S	2	4
I want to be able to add an experiment type to the database inside the app	M	4	5
I want to only ever run one instance of the application at a time	M	4	3
I want a preview of my new keyword to show in a little box at the bottom of the add new keywords screen	S	2	5
I want to see project preferences, experiment, and keyword databases in a hamburger menu in full namer	M	6	5
I want test cases for everything	XL	19	4

# Appendix W: Application Feedback Survey

*This survey was used to gather feedback on the fourth version of the file naming application.*

## Application Feedback

Thank you for testing our application! Your answers to the following questions will be used to make improvements to our application. Optionally, you may provide your name and email address so that we may collect your feedback in further detail.

Optional - What is your name and email address?

Your answer

### Basic Functionality

Overall, was the app easy to use? \*

	1	2	3	4	5	
No, it was confusing	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Yes, it was intuitive

Were you able to navigate to the project preferences screen? \*

- Yes
- No

Were you able to set project preferences? \*

- Yes
- No

Were you able to navigate to the compact namer screen? \*

- Yes
- No

Do you prefer the compact file namer or the full file namer? \*

- Compact File Namer
- Full File Namer
- I have no preference.

Did you use the help icons? \*

- Yes
- No

If you used the help menus, were they useful?

	1	2	3	4	5	
Not helpful	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very helpful

Were you able to add keywords? \*

- Yes
- No

How likely are you to use this app in your daily research? \*

	1	2	3	4	5	
Not likely	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very likely

## Logger

Were you able to create a log file? \*

- Yes
- No

Were you able to open and view your log file? \*

- Yes
- No

Were you able to print to a previously saved log file? \*

- Yes
- No, I didn't try that feature.
- No, I couldn't figure out how to do that.

Does the Excel file format for the time logger work well for you?

\*

- Yes
- No

Please note any specific feedback on the logger feature.

Your answer \_\_\_\_\_

### Added Features

Would you like to be able to change the order in which information appears in the file name? (For example, put the researcher's initials first instead of the date) \*

- Yes
- No

Would you like to be able to click a button that would clear all the fields? \*

- Yes
- No

Are there any other features you would like in the application?

Your answer \_\_\_\_\_



## Additional Feedback

Were you confused about any of the following features of Full Namer?

### FULL NAMER

Project: My First Experiment

**1** Project Preferences **2** Open Logger

**3** Full Namer  Compact Namer

Date  **?**

Experiment Name  
Trial Experiment **4**

Researcher Name  
John Doe **5**

Trial Number  **6**   Sample Number  **7**

Keywords Table **8**

Name	Data Value
Depth in millimeters	30

Output:  
2018-9-25-TE-T1-51-JD-30mm **9** **?**

- No
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9

If so, why?

Your answer

Were you confused about any of the following features of Compact Namer?

The image shows a dialog box titled "COMPACT NAMER". It has a header with "FULL" and "COMPACT" options, where "COMPACT" is selected. Below the header is a "DATE:" field with the value "25/09/2018". There are two rows of controls: "SAMP..." and "TRIAL #", each with a text input field containing "1" and two buttons labeled "+" and "-". At the bottom are two buttons labeled "COPY" and "CLOSE". Red circles with numbers 1 through 4 point to the "FULL" label, the "DATE:" field, the "SAMP..." field, and the "COPY" button respectively.

- No
- 1
- 2
- 3
- 4

If so, why?

Your answer

---

Were you confused about any of the following features of the Add New Experiment pop-up?

The image shows a dialog box titled "ADD NEW EXPERIMENT TO DATABASE". It has three input fields: "Full Name" with the value "Trial Experiment", "Abbreviation" with the value "TE", and a larger text area containing "This is a Trial Experiment". At the bottom are two buttons labeled "Submit" and "Cancel". Red circles with numbers 1 through 4 point to the "Full Name" label, the "Abbreviation" label, the text area, and the "Submit" button respectively.

- No
- 1
- 2
- 3
- 4

If so, why?

Your answer

Were you confused about any of the following features of the Add New Keyword pop-up?

Choose a keyword from the database:

1 Name

2 Data Value

3

No

1

2

3

If so, why?

Your answer

Were you confused about any of the following features of the Project Preferences pop-up?

Project Preferences

1 Researcher Name

2 Project Name

3

4 Select a separation character:  
 Asterix (\*)  Hyphen (-)  Underscore (\_)

5 Preview:

6

No

1

2

3

4

5

6

If so, why?

Your answer

Were you confused about any of the following features of the Generate Log File pop-up?



- No
- 1
- 2
- 3
- 4
- 5

If so, why?

Your answer

---

Please enter any additional feedback in the box below.

Your answer

---

## Appendix X: Code Authorship

Feature	Subfeature	More details	Contributor
Full Namer	User interface design		Jeffrey Xiao
	Initial functionality	Set up the databases	Jeffrey Xiao
		Autocomplete text field to display experiments in database	Jeffrey Xiao
		Output box	Jeffrey Xiao
		Copy button that copies file name to clipboard	Jeffrey Xiao
	File name output	Generate file name from parameters	Jeffrey Xiao
		Preview file name in real time	Jeffrey Xiao
	Differentiate between trial and sample number		Vandana Anand
	Researcher name field	Appears in the file name	Vandana Anand
		Autocomplete text field to display researchers in database	Vandana Anand
	Navigate to menu bar options		Jeffrey Xiao
	Add keywords	User interface design	Vandana Anand
		Keywords appear in a table view	Vandana Anand
		Input parameters carry over from initial input screen to table	Vandana Anand
		Add/modify/delete keywords functionality	Vandana Anand
		Ability to add keyword into the database from the application	Jeffrey Xiao
		Keywords appear in the file name	Jeffrey Xiao
		Input numeric or alphanumeric values with keywords	Vandana Anand and Jeffrey Xiao
		Preview box	Carlisle Ferguson
	Autocomplete text field to display keywords in the database	Vandana Anand	
	Additional features	Toggle to Compact Namer	Carlisle Ferguson
		Help icons and hover menus	Carlisle Ferguson
		Clear fields	Carlisle Ferguson
Persistent trial number and sample number		Jeffrey Xiao	
Checkboxes		Jeffrey Xiao	
User interface re-design		Jeffrey Xiao	

<b>Compact Namer</b>	User interface design		Carlisle Ferguson
	Initial Functionality		Carlisle Ferguson and Jeffrey Xiao
	Additional Features	Toggle button	Carlisle Ferguson
		Increment and decrement buttons and functionality	Carlisle Ferguson
		Share information between full namer and compact namer	Jeffrey Xiao
		Checkboxes	Carlisle Ferguson and Jeffrey Xiao
User interface re-design		Jeffrey Xiao	
<b>Project Preferences</b>	User interface design		Carlisle Ferguson
	Functionality	Fields for researcher name, project name, and project description translate over	Jeffrey Xiao
		Fields in Full Namer automatically populate	Jeffrey Xiao
		Delimiter updates	Jeffrey Xiao
		Option to remember data	Carlisle Ferguson
	Preview delimiters		Carlisle Ferguson
Help icons and floating menus		Carlisle Ferguson	
<b>Logger</b>	User interface design		Carlisle Ferguson
	Help icons and hover menus		Carlisle Ferguson
	Excel file design		Jeffrey Xiao
	Functionality	Generate log entry whenever the copy button is clicked	Jeffrey Xiao
		File Explorer for loading and opening logs	Carlisle Ferguson
		Write log to Excel file	Jeffrey Xiao
		Reset logger	Carlisle Ferguson
Display current log entries in a table in the application		Jeffrey Xiao	
Log table is editable		Vandana Anand	
<b>Installer</b>	Deployable application (.exe)		Jeffrey Xiao
	Functionality on Windows		Jeffrey Xiao
<b>Getting Started</b>	User interface design		Carlisle Ferguson
	Functionality		Carlisle Ferguson
<b>About</b>			Carlisle Ferguson

<b>Databases</b>	User interface design	All inputs and additional parameters appear in a table view	Vandana Anand
		Enter parameters for new keywords, experiment, researcher	Vandana Anand
	User interface functionality	Parameters carry over from initial input screen to table	Vandana Anand
		Add/delete/modify in keyword, experiment, researcher tables	Vandana Anand
	Database functionality	Changes made in the UI table is also made in the .csv database	Jeffrey Xiao
<b>General Code</b>	Resources	Cascading style sheet design	Jeffrey Xiao
		Custom font	Jeffrey Xiao
		Icon images	Jeffrey Xiao
		Animations	Jeffrey Xiao
	Bug fixing		Vandana Anand
	Drag windows		Carlisle Ferguson

## Appendix Y: Code Location and Modification Steps

The code for the file naming application can be modified and adapted to fit the standards of a specific department. A working knowledge of Java and Git is required to make modifications. The program is currently hosted on GitHub, a repository that can be used to organize and manage code files of a software program. GitHub uses a tool called Git, which is a version control system that helps maintain the most current version of the program, especially when collaborating with other members.

The file naming application can be found at this link: <https://github.com/jyxiao1/filenamerpro>

A list of software needed to modify the program is listed below:

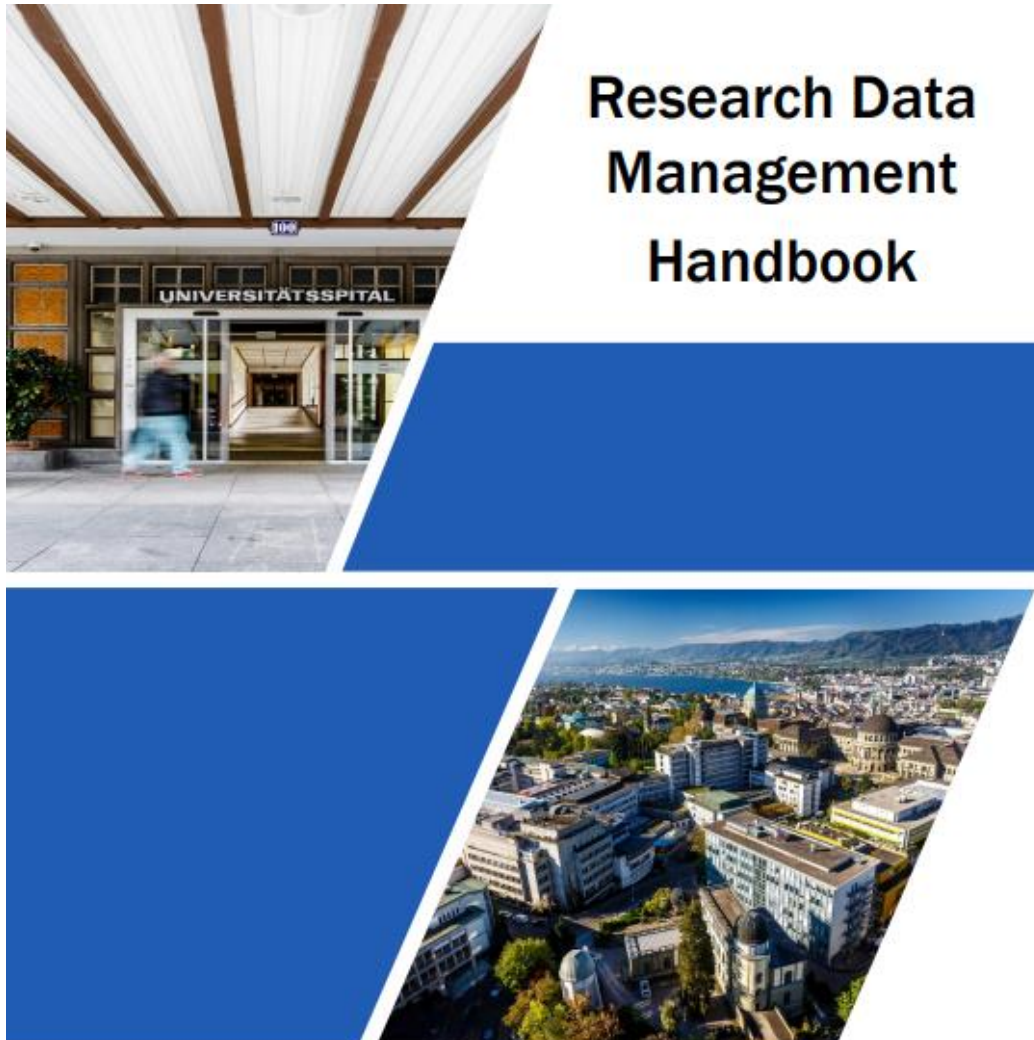
1. An Integrated Development Environment (IDE). We recommend using IntelliJ: <https://www.jetbrains.com/idea/>
2. Java SE Development Kit 8: <https://www.oracle.com/technetwork/java/javase/downloads/jdk8-downloads-2133151.html>
3. Scene Builder: <https://gluonhq.com/products/scene-builder/>
4. Gradle: <https://gradle.org/>
5. JFoenix: <https://github.com/jfoenixadmin/JFoenix>

If you would like to contact the team members with specific questions on the software please email us at [mdo-zurich@wpi.edu](mailto:mdo-zurich@wpi.edu).



## Appendix Z: Research Data Management Handbook

*The Research Data Management Handbook was written to provide the project sponsor, Dr. Ivo Dobrev, and the hospital with guidelines for managing research data. The document in its entirety is included on the following pages.*



Written in collaboration with Worcester Polytechnic Institute



**WPI**



**UniversitätsSpital  
Zürich**

# Research Data Management Handbook



Written in collaboration with Worcester Polytechnic Institute  
Authored by Vandana Anand, Carlisle Ferguson, Caitlin Kyaw, and Jeffrey Xiao  
11 October 2018



**WPI**



**UniversitätsSpital  
Zürich**

## Table of Contents

Chapter 1: Introduction .....	2
Purpose of this Handbook.....	2
Chapter 2: Documenting Data .....	3
Paper Lab Notebooks.....	3
Electronic Note-Taking.....	3
Audio Note-Taking .....	5
Code .....	5
Applications to SNSF Requirements.....	5
Conclusions .....	6
Chapter 3: Naming Data Files .....	7
Necessary Components .....	7
Examples of File Names .....	7
Common Mistakes .....	8
Applications to SNSF Requirements.....	8
Conclusions .....	8
Chapter 4: Organizing Data File Folders.....	9
Three Characteristics of Effective Organization.....	9
Methods.....	9
Applications to SNSF Requirements.....	10
Conclusions .....	11
Chapter 5: Storing and Transferring Data Files.....	12
Handwritten Data .....	12
Electronic Data .....	12
Applications to SNSF Requirements.....	13
Conclusions .....	14
Additional References.....	15
Paper Lab Notebook Checklist .....	16
Electronic Note-Taking Checklist .....	17
Audio Note-Taking Checklist.....	18
Naming Files Checklist .....	19
Organizing Data Files Checklist .....	20
Storing and Transferring Physical Data Checklist.....	21

# Chapter 1: Introduction

## Purpose of this Handbook

Managing your research data is a highly important task. Considering and planning how data should be created, named, stored, and organized at the onset of an experiment or project will save you both time and effort when trying to analyze and process results at the end. Doing so will also prevent any later confusion and allow you and others to quickly identify and locate information.

The purpose of this document is to provide recommendations for efficiently and effectively documenting and organizing research data. In the following chapters, you will be able to find guidance on:

- Documenting Data
- Naming Data Files
- Organizing Data File Folders
- Storing and Transferring Data Files

Additionally, for projects wishing to apply for grants from the Swiss National Science Foundation (SNSF), specific guidance on how to best manage your data in compliance with FAIR Data Principles and the Data Management Plan (DMP) can be found in each chapter.

# Chapter 2: Documenting Data

*In this section, best practices for documentation in both electronic and paper mediums are described.*

Documentation, as it relates to research data, refers to information that aids in the interpreting and understanding of data. The act of documenting data entails specifying what, when, how, and where an experiment was conducted. Comprehensive documentation not only enables others to understand the work that was conducted but also serves as clear reminders to the person who conducted the work if the data should need revisiting.

## Paper Lab Notebooks

The following are good rules of thumb for documenting information in paper lab notebooks, and an example of these practices can be found on the next page:

- **Create and update a table of contents.** This allows you to quickly find experiments in lab notebooks that have many entries.
- **Date every page.**
- **Write clearly and neatly.** Legible handwriting benefits yourself and others when trying to read and understand the information later.
- **Record the name and aims of each experiment or project.** This information is ideally located at the top of each page and allows you to remember what the goal of the experiment or project was at a later time.
- **Record the procedure in detail.** Doing allows experiments to be easily reproducible. Be specific when describing details such as techniques, amounts, machines used, and materials.
- **Record experimental conditions in detail.** Similarly to recording the procedure, this is important because it makes experiments easily reproducible. Conditions can include temperature, brand and lot numbers of chemicals, the duration of the experiment, instrument calibration numbers, etc.

- **Record observations and results in detail.** Details of the experiment or project can be clear to you at the moment, but you may forget certain aspects years, months, days, or even hours after completing an experiment. Similarly, things that may be clear to you may not be clear to others. Additionally, if there were mistakes or failures relating to the experimental results, explain why.
- **Do not “obliterate” mistakes.** To correct them, cross them out with a single line, and briefly explain why the crossed-out information should be disregarded.
- **Explain and expand abbreviations.**
- **Attach any external graphs or computer printouts using glue or other permanent binding materials.**

## Electronic Note-Taking

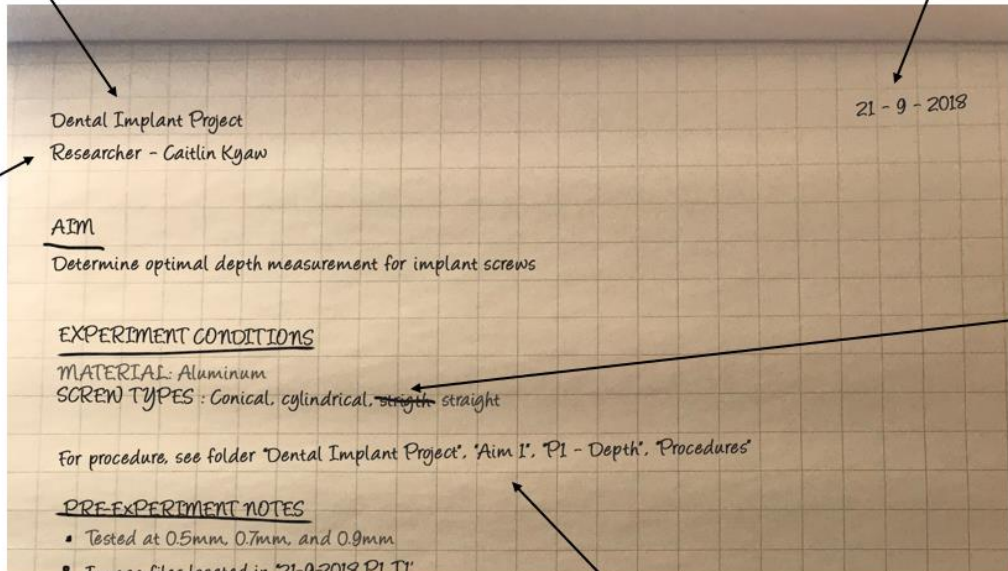
The following are good rules of thumb for documenting information when taking notes electronically:

- **Document the date.**
- **Explain and expand abbreviations.**
- **Record the name and aims of each experiment.**
- **Record experimental conditions, observations, and results in detail.**
- **Import any relevant images and add annotations.** This allows any related observations to be found in one place.
- **Create note-taking templates for common projects and experiments.** Doing so saves you time when repeating common experiments. An example can be found on the next page.
- **Share the above templates with colleagues.** This standardizes the way projects and experiments are documented, allowing for greater consistency.

The following is an example of a well-documented paper lab notebook:

Name of the project is indicated

Date the experiment was conducted is recorded



Researcher name is recorded

Mistake was not "obliterated", instead crossed out with a single line

Related external files are referenced

The following is an example of an electronic note file template:

<b>Date:</b>
<b>Project/Experiment Name:</b>
<b>Conducted by:</b>
<b>Aims:</b>
<b>Procedure:</b>
<b>Notes and Observations:</b>

## Audio Note-Taking

The following are good rules of thumb for documenting information when taking notes electronically:

- **Speak clearly.** It is a good idea to test if you can be heard clearly on the recording beforehand.
- **State the date.**
- **State who is taking the recording.** If it is just you, state your own name. If several people will be included in the recording, be sure that they also state their names.
- **State the name and aims of each experiment or project.**
- **State experimental conditions, observations, and results in detail.**
- **Verbally relate the recording to any relevant notes (handwritten or electronic) that are taken concurrently.** This allows you to group observations from the same project together.

## Code

Clearly and concisely documenting code used for processing and analyzing data is also highly important. The following are good rules of thumb for documenting code:

- **Name variables, methods, and classes clearly and consistently.** This allows viewers to quickly understand how certain items are related and used.
- **Only annotate your code if necessary.** If many comments are needed to clarify your code, you should probably write the code differently. However, comments are acceptable in moderation.
- **Be concise.** Remove any irrelevant lines of code. If any code has been commented out, you probably don't need it.

---

## Applications to SNSF Requirements

### *Data Management Plan (DMP)*

Documentation of research data is highly important when it comes to the submission of the DMP. The first section of the DMP will ask you to explain what data will be collected, and how you intend to collect it, and what metadata, or data that describes data, will be included. To be able to accurately detail such information, consider documenting the following during an experiment or project:

- Data type
- Format of raw data
- Format of processed data
- Estimated volume of raw and processed data

An example of this might look like:

*Western blot data is collected for the detection of proteins. Images are taken using a CCD camera. The image data is quantified and processed in Microsoft Excel. The data files for the images and Excel files total 2GB and 4GB, respectively.*

To be able to describe the metadata and how it will be collected and provided, consider documenting the following during an experiment or project:

- Content
- Responsibilities
- Software
- Organization

An example of this might look like:

*Documentation for this data describes experimental conditions such as temperature (Celsius) and humidity, as well as the experimental protocol used. It is the responsibility of the researcher to detail the conditions of the experiment daily. Documentation is in a text file format, and the file is named and organized in the pre-agreed format of "YYYYMMDD\_Researcher\_ProjectName.txt".*

### *FAIR Data Principles*

Documentation, as it pertains to the FAIR Data Principles, is specifically related to findability and reusability. In order to be findable, the data should be described clearly and thoroughly by metadata. The context, conditions, and characteristics of the data are all important aspects of metadata, and other researchers should be able to easily understand the nature of the data. Similarly, to be reusable, it is your responsibility to ensure that information such as the scope of the data, any limitations or failures of the experiment, the date the experiment or project was conducted, and other additional details are included in your documentation.

## Conclusions

Documenting your data thoroughly will allow you and other researchers to quickly understand the aims, conditions, observations, and results from a project or experiment. Although it may take longer to include more details, clear documentation will save time during the processing and publication phases.



# Chapter 3: Naming Data Files

*This section describes how to name data files in a descriptive and effective manner.*

Descriptive file naming is highly important when conducting research. A clear and detailed file name enables you and your collaborators to quickly and easily identify files. Furthermore, having standardized file names across projects, research groups, and departments allows for additional clarity in research.

## Necessary Components

The following components should always be included in your file name:

- **Date.** The recommended date format is YYYY\_MM\_DD. This will allow you to easily sort your file names by date in a file manager or explorer.
- **Researcher Name.** Incorporating your name or your initials into the file name allows collaborators and future researchers to determine who collected the data.
- **Experiment Type.** Listing the experiment type allows you to quickly identify the appropriate data files within each project subfolder.
- **Trial Number.** This allows you to identify exactly which iteration of an experiment the file describes and differentiates it between similar experiments.
- **Keywords.** These can either be included as full words or as abbreviations and describe additional parameters or conditions of an experiment. If you choose to use an abbreviation, be certain that it is used in the same way throughout your research group to avoid any confusion.

In addition to being descriptive, file names should be standardized across projects, research groups, and collaborations. Standardization can be accomplished in one of two ways:

- **Discussion.** Discuss and define how you and your project team would like to name files. Have this decision recorded somewhere and hold each other accountable for using it.
- **File Naming Application.** Utilize the file naming application, File Namer Pro, which automatically generates a standardized file name for you. This is the preferred method.

When file names are standardized, researchers do not have to spend time trying to discern what a files' contents are, and can easily glean information by looking at the file name. This also allows files to be easily located in a file manager. If you need to look for a file created on 10 September 2018, you simply need to look up files with "2018\_09\_10".

## Examples of File Names

The following is an example of a file that has been **poorly** named, based on the information provided below:

<p><b>Date:</b> 10 September 2018</p> <p><b>Researcher Name:</b> James Bond</p> <p><b>Experiment Type:</b> Sound Check</p> <p><b>Additional Keywords:</b> 40hz, Stealth Car</p> <p><b>File Name:</b> SoundCheck1</p>
--

This is a poor file name because it does not provide very descriptive information. "SoundCheck1" could refer to any soundcheck experiment done by any person, on any date, with any parameters.

The following is an example of a file that has been named **well**, based on the information provided below:

**Date:** 10 September 2018

**Researcher Name:** James Bond

**Experiment Type:** Sound Check

**Additional Keywords:** 40hz, Stealth Car

**File Name:** 2018\_09\_10\_JB\_SC\_40hz\_StealthCar

This is a good file name because it specifies who conducted the soundcheck experiment, when, and uses specific and descriptive parameters.

## Common Mistakes

- **Making long file names.** While including many keywords is useful, be certain to keep your file name within a reasonable character limit.
- **Using inconsistent abbreviations.** Your abbreviations may make sense at the moment, but might not several weeks or months later. Be certain to create and maintain a glossary of keywords so you can easily decipher your file names.
- **Using special characters.** Avoid using special characters such as “%” and “&”, as these symbols make reading file names difficult and are often not program-friendly.

---

## Applications to SNSF Requirements

### Data Management Plan

The Data Management Plan also requires information concerning how files will be named. As stated above, you and any collaborators should reach a consensus on how files will be labeled, and ensure that the agreed-upon convention is consistently used. In your DMP, explain clearly what the convention is. An example of this might look like:

*All raw data will be named according to the pre-agreed format of “YYYYMMDD\_Researcher\_ProjectName\_ExperimentName\_Condition1\_Condition2”. A concrete example of this is “20180903\_ER\_UJI\_DM\_7mm\_Al”, where the date is 3 September 2018, the experiment was conducted by “Example Researcher”, the project name is “Upper Jaw Implant”, the experiment name is “Depth Measurement”, and the conditions of the experiment are a depth of 7 millimeters and aluminum material. Information detailing any non-standard acronyms can be found in the project text file.*

### FAIR Data Principles

Standardized, specific file naming allows data to be findable, in accordance with FAIR Data Principles. To be findable, each dataset should be designated with a unique identifier that clearly differentiates it from other datasets. This identifier should be consistent through all mentions and references to the dataset and be descriptive. A good file name will allow you and others to quickly identify a file and its contents.

## Conclusions

Naming your files in a detailed, standardized manner will allow you and others to quickly and efficiently discern a file’s contents without needing to open it. This saves time during the processing and publication phases, allowing you to focus on your research instead of looking for files.

# Chapter 4: Organizing Data File Folders

*In this section, best practices for organizing file structures are discussed.*

Maintaining a proper file structure system over the course of an experiment allows you to quickly and efficiently locate and identify research data files. The time that would have been spent conducting a search similar to an archeological dig could be better spent on other research projects when files are properly organized.

## Three Characteristics of Effective Organization

There are several characteristics of a proper file organization system:

- **Data is easily filed.** The file organization system should have a logical file structure rather than a maze of folders and files. This will create an easier process to store files for both the user and the machine.
- **Files are easily found.** The organization structure should allow you to easily locate files via a search function or by eye. If it is confusing to figure out where to store a file, it'll probably be just as confusing to find it later.
- **File folder structures are standardized.** File organization methods should be consistent across projects and departments, only differing slightly depending on the project.
- **Create folder structures for stages, tasks, or aims of the experiment.** Creating subfolders within the main experiment folder for specific stages, aims, or tasks of an experiment or project can help track the progress of an experiment. Examples of folders in this structure could be Planning, Data Collection, Results, Documentation, Future, Objective 1.
- **Group folders by date.** A common and important experiment parameter is the date. Grouping projects or experimental data by date can simplify the process of searching for the most up-to-date files.
- **Minimize the number of folders.** Having a large number of folders can make it overwhelming and difficult to locate files in the future. Have fewer folders and more descriptive file names to easily locate files.
- **Limit the number of files and folders on the desktop.** Although storing files and folders on the desktop can seem like an easy way access them at the time, it becomes difficult to track as more and more files and folders are added.
- **Name files and folders logically.** Refer to the Naming Files chapter for more information.
- **Discuss folder structure standards with collaborators.** The whole group should come to a consensus on organizing data that can consistently be followed on a day-to-day basis with minimal changes. In addition, these templates and standards should be reused for other projects to enforce that folders be named a specific way each time.
- **Create shortcuts to frequently accessed folders.** If you find yourself returning to a certain folder often, it is a good idea to make a shortcut in file explorer or at the bottom of the menu bar. This way, you eliminate the need to shuffle through folders every time.

## Methods

The following are good rules of thumb for organizing data files:

- **Maintain folder hierarchies.** Designating a folder for the project as a whole will differentiate between projects. Within the main project folder, subdirectories can be created. An example hierarchy can be: Project -> Experiment -> Sample. In this way, general information about the project and data files from experiments will cause confusion.

- ▼ Dental Implant Project
  - ▼ Aim 1 - Optimize Measurements
    - ▼ P1 - Depth
      - Procedures
      - Processed Data
      - Raw Data
    - ▼ P2 - Width
      - Procedures
      - Processed Data
      - Raw Data
  - ▼ Aim 2 - Optimize Materials
    - ▼ P1 - Aluminum
      - Procedures
      - Processed Data
      - Raw Data
    - ▼ P2 - Steel
      - Procedures
      - Processed Data
      - Raw Data

The folder structure shown on the left is an example of a well-organized file structure. The project itself has a main folder, and subfolders are in a logical hierarchy. Additionally, the folders are named clearly.

- ▼ Project 2
  - 2018-09-04 Files
  - Head Data
  - Images
  - New folder
  - Other Files
  - Other Images
  - Project 3
  - Test 4 Data

The folder structure shown on the right is an example of a poorly-organized file structure. There is no logical hierarchy of folders, and file folders are named ambiguously.

## Applications to SNSF Requirements

### *Data Management Plan (DMP)*

Determining how files will be organized is also an important aspect of the DMP. You should indicate the structures of folders clearly in your DMP. An example of this might look like:

*The project has a main, overarching folder that includes all related files and data. Within the main folder, each aim of the experiment has its own subfolder. Within an aim subfolder, data and related files are categorized by experiment type. An example hierarchy of folders is:*

- *Dental Implant Project*
  - *Aim 1: Optimize Measurements*
    - *Depth*
      - *Procedure*
      - *Raw Data*
      - *Processed Data*
    - *Width*
      - *Procedure*
      - *Raw Data*
      - *Processed Data*
  - *Aim 2: Optimize Materials*
    - *Aluminum*
      - *Procedure*
      - *Raw Data*
      - *Processed Data*
    - *Steel*
      - *Procedure*
      - *Raw Data*
      - *Processed Data*

### *FAIR Data Principles*

File structure organization is also important when ensuring your data and your project adheres to the FAIR Data Principle of being findable. Metadata and the data it describes should be easily searchable by both humans and computer systems in large indices of data. As stated previously, your files and folders should be structured simply and logically with a certain degree of specificity so that the data is easily differentiated. Well-organized file structures will allow you and others to quickly discover and identify interesting sets of data.

## Conclusions

In summary, keeping a well-maintained file folder structure will allow you and other researchers to quickly and easily locate and identify files. With the proper organization, time spent looking for certain files and folders can instead be spent on research.

# Chapter 5: Storing and Transferring Data Files

*This section provides recommendations for storing and transferring data.*

Any data that is generated from experiments or projects is highly valuable. Because of this, effectively storing and transferring the data is crucial.

## Handwritten Data

The following are good rules of thumb for storing and transferring handwritten data:

- **Digitize any handwritten notes or hand-drawn pictures or diagrams.** Scanning or typing any handwritten notes and saving them onto a computer allows the information to be kept in two places in case the physical copy is lost. Typing handwritten notes will also improve comprehensibility, as handwriting can occasionally be illegible.
- **Have a safe, centralized, organized location for physical copies of notebooks or diagrams.** This allows the information to be found easily and efficiently. Additionally, this protects the physical copies from damage.
- **Keep physical copies in the lab.** Occasionally, experiments are conducted in sterile, controlled environments. Keeping the physical copies of notes or diagrams in the lab will decrease the risk of contamination. If the information is needed outside the lab, copy the page or diagram onto another sheet of paper or print out a digitized version.

## Electronic Data

The following are good rules of thumb for storing and transferring electronic data:

- **Perform regularly scheduled back-ups.** This can be nightly, weekly, or monthly. Consistently backing up your data will prevent accidental loss.
- **Clear out unnecessary data.** Computers attached to instruments are not meant to be storage systems. Back-up any needed data onto the cloud or to your own computers and delete any files that are no longer needed. If you are a non-permanent lab member, delete any of your unnecessary project files, and make it clear to your colleagues as to which files are not being deleted and why.
- **Avoid transferring data via external drives, such as flash drives.** If external drives are broken or misplaced, whole sets of data from experiments or projects can be lost.
- **Communicate with your IT department.** Occasionally, the biggest obstacle to efficient data storage and transfer is a lack of communication between labs and IT. Talk to your IT department to determine the easiest way for you to transfer and store data.
- **Create duplicates if you plan to edit or manipulate a file.** Maintaining an original version of a file can be useful if you erroneously change or manipulate a file.

## Applications to SNSF Requirements

### *Data Management Plan*

Data Management Plans specifically require information on how the data will be stored, preserved, shared, and reused. In terms of completing the DMP, there are specific things you should think about and describe when conducting your project.

- **Access rights**  
At the time of the project, data should only be accessible by authorized persons. Ideally, any time a file or dataset is accessed, viewed, or changed during a project should be logged somewhere.
- **Storage and backup procedures**  
The frequency of data backups, physical or digital location of the storage medium, the network capabilities of the storage location, and related information should be mentioned.
- **Preservation procedures**  
The selection criteria for data preservation, length of preservation, and other relevant information as it pertains to preservation procedures should be detailed.
- **Sharing repositories and methods**  
The repositories on or methods through which data will be shared after publication should be well-documented, if applicable.

An example description for this information might look like the following:

*Data access will be defined before starting the project, and the list of authorized peoples will be managed by the project head. Data is backed up nightly onto a restricted, centralized system provided by the university and managed by the IT department. Data will not leave the lab premises physically nor virtually unless authorized by the project head. Any data used for publication will be preserved for 10 years on the university server.*

*Research data is anonymized in accordance with institutional policy. Users are granted access with permission from the trial manager, access to any files will be logged and traceable. Data is backed up weekly onto a data center located in the facility, and after completion of the trial, will be stored for 20 years.*

### *FAIR Data Principles*

Your data's accessibility, interoperability, and reusability in accordance with the FAIR Data Principles will depend heavily on how you choose to transfer and store your data.

For accessibility and reusability, the SNSF at the very least expects you to share the metadata needed to make any published results reproducible. If you choose to share data, it can be in raw or processed form but should be accompanied with relevant metadata. If there are any particular tools or software needed to reuse, reproduce, or access the data, this information should be documented, and the tools or software used should be made available. Ideally, at least the metadata should be accessible by anyone with a computer. If necessary, the data can be restricted by authentication or authorization procedures.

For interoperability, your data and metadata should be machine-readable, or in a form that a computer can process. Additionally, you should properly link and cite any relevant files or datasets and describe their relevance.

If these requirements cannot be met, you should explain the reasoning in detail in your DMP.

## **Conclusions**

Properly storing and transferring research data is important to its eventual analysis and publication. By adhering to the practices stated, you can be assured your data is not lost.



# Additional References

Further information concerning SNSF guidelines and data management can be found through the following websites:

**The DCLM Project** – The Swiss “Data Life-Cycle Management” (DCLM) project aims to provide researchers with guidelines on research data management. An example DMP template and additional information concerning these guidelines can be found at <https://www.dlcm.ch/>

**Swiss National Science Foundation (SNSF)** – The Swiss National Science Foundation has provided a document explaining the contents of the mySNF form and the Data Management Plan, which can be found here: [http://www.snf.ch/SiteCollectionDocuments/DMP\\_content\\_mySNF-form\\_en.pdf](http://www.snf.ch/SiteCollectionDocuments/DMP_content_mySNF-form_en.pdf)

Additionally, a thorough explanation of the FAIR Data Principles can be found at [http://www.snf.ch/SiteCollectionDocuments/FAIR\\_principles\\_translation\\_SNSF\\_logo.pdf](http://www.snf.ch/SiteCollectionDocuments/FAIR_principles_translation_SNSF_logo.pdf)

Lastly, the language policy of the Swiss National Science Foundation can be found here: [http://www.snf.ch/SiteCollectionDocuments/Dossiers/dos\\_sprachenpolitik\\_snf\\_e.pdf](http://www.snf.ch/SiteCollectionDocuments/Dossiers/dos_sprachenpolitik_snf_e.pdf)

**National Institutes of Health (NIH)** – The National Institutes of Health has provided guidelines for scientific record keeping, which can be found here: [https://oir.nih.gov/sites/default/files/uploads/sourcebook/documents/ethical\\_conduct/guidelines-scientific\\_recordkeeping.pdf](https://oir.nih.gov/sites/default/files/uploads/sourcebook/documents/ethical_conduct/guidelines-scientific_recordkeeping.pdf)

**The University of Cambridge** – The University of Cambridge has provided useful information on general data management, which can be found here: <https://www.data.cam.ac.uk/data-management-guide>

**Stem Cell Technologies** – Stem Cell Technologies has provided useful information on how to organizing and document lab notebooks, which can be found here: <https://www.stemcell.com/efficient-research/lab-notebooks-references-protocols>

# Paper Lab Notebook Checklist

Have you...	✓, X, or N/A
... started on a new page?	
... made an associated table of contents entry?	
... dated the page?	
... recorded the name of the experiment or project?	
... recorded the aim or purpose of the experiment or project?	
... recorded the procedure in detail?	
... recorded experimental conditions in detail?	
... recorded observations and results in detail?	
... explained and expanded any abbreviations?	
... explained any mistakes or failed results?	
... attached any external graphs or computer printouts?	

# Electronic Note-Taking Checklist

Have you...	✓, X, or N/A
... created a note-taking template if this a common experiment or project?	
... shared the template with colleagues for widespread use, if applicable?	
... typed in the date?	
... recorded the name of the experiment or project?	
... recorded the aim or purpose of the experiment or project?	
... recorded the procedure in detail?	
... recorded experimental conditions in detail?	
... recorded observations and results in detail?	
... explained and expanded any abbreviations?	
... explained any mistakes or failed results?	
... imported any relevant images and made annotations?	
... named the document file with descriptive keywords?	

# Audio Note-Taking Checklist

Have you...	✓, X, or N/A
... tested the recording device to ensure you can be heard?	
... stated the date?	
... stated the name of the experiment or project?	
... stated the aim or purpose of the experiment or project?	
... stated the procedure in detail?	
... stated experimental conditions in detail?	
... stated observations and results in detail?	
... explained any abbreviations?	
... explained any mistakes or failed results?	
... verbally related the audio recording to any relevant notes (handwritten or electronic) that are taken concurrently?	

# Naming Files Checklist

Have you...	✓, X, or N/A
... included the date?	
... included your name and the name of any collaborators?	
... included the experiment type?	
... included the trial number?	
... included specific, descriptive keywords?	
... communicated the standards for labeling files with the project team?	
... created or updated a glossary of abbreviations and their expanded forms?	

# Organizing Data Files Checklist

Have you...	✓, X, or N/A
... created a main folder for the project?	
... created subfolders for the different stages, tasks, and aims of the experiment?	
... identified what additional subdirectories you will need?	
... grouped folders by date?	
... discussed folder structure standards with any collaborators?	
... created shortcuts to frequently accessed folders?	
... moved unnecessary folders out of the desktop?	

# Storing and Transferring Physical Data Checklist

Have you...	✓, X, or N/A
... digitized the data?	
... identified where the data will be stored in a central location?	
... communicated the storage location to any collaborators?	
... communicated whether or not the data can be taken out of the lab space?	

## Appendix AA: User Manual for File Namer Pro

*The User Manual for File Namer Pro written to provide the project sponsor, Dr. Ivo Dobrev, and the hospital with instructions on how to download, launch, and use the file naming application, File Namer Pro. The document in its entirety is included on the following pages.*



Written in collaboration with Worcester Polytechnic Institute



**WPI**



**UniversitätsSpital  
Zürich**





# File Namer Pro

User Manual



Written in collaboration with Worcester Polytechnic Institute  
Authored by Vandana Anand, Carlisle Ferguson, Caitlin Kyaw, and Jeffrey Xiao  
11 October 2018



# WPI



# UniversitätsSpital Zürich

# Table of Contents

Chapter 1: Introduction .....	2
Chapter 2: Features .....	3
Chapter 3: Download and Launch Instructions .....	5
Chapter 4: Quick Start Guide .....	6
Using Full Namer .....	6
Using Compact Namer .....	8
Setting Project Preferences .....	9
Generating Log Files.....	10
Chapter 5: Adding to the Databases .....	13
Adding to the Experiment Name Database .....	13
Adding to the Researcher Name Database.....	15
Adding to the Keywords Database .....	17
Chapter 6: Modifying the Code.....	19

# Chapter 1: Introduction

*File Namer Pro* is a Java application that generates standardized file names. Descriptive file naming is highly important when conducting research, and clear and detailed file names allow for the quick and easy identification of data files. Furthermore, file names that are standardized across projects and research groups can improve communication.

The purpose of this manual is to provide step-by-step instructions on how to download, launch, and use *File Namer Pro*.

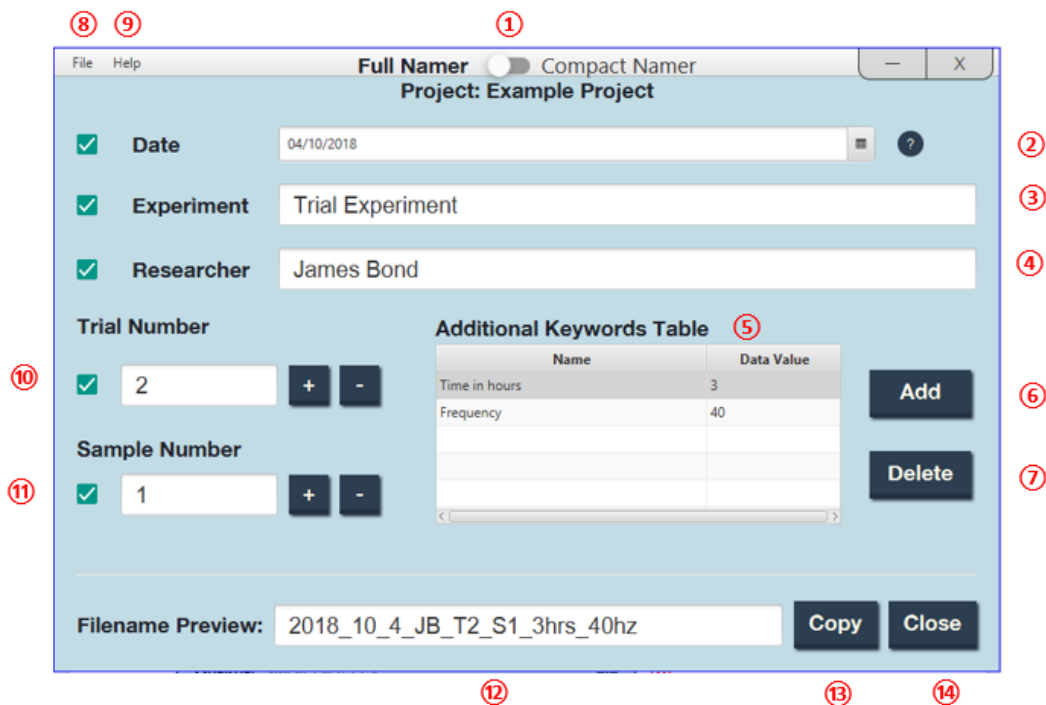
## Chapter 2: Features

This chapter provides an overview of the features of *File Namer Pro*. See Chapter 3 for download and launch instructions and Chapter 4 for the Quick Start Guide.

### Full Namer

As the main window of *File Namer Pro*, Full Namer allows you to fill out a few fields and generate detailed and standardized file names that can be copied onto your clipboard.

*Full Namer Feature Breakdown:*

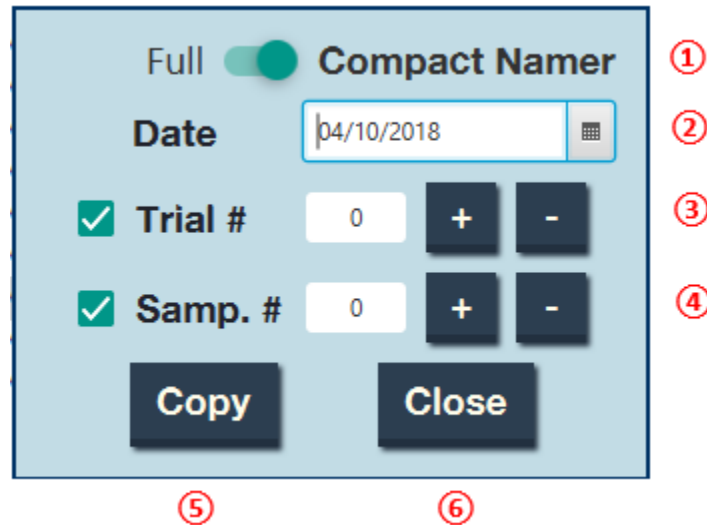


1. Toggle to compact namer
2. Date
3. Experiment name
4. Researcher's name
5. Table of keywords
6. Add a keyword button
7. Delete a keyword button
8. File menu to clear fields, access project preferences, access logger, and access the databases
9. Help menu to access the about page and the getting started page
10. Trial number
11. Sample number
12. Filename preview
13. Copy to clipboard button
14. Close button

## Compact Namer

In a hurry to conduct multiple trials or tests on different samples? Compact Namer allows you to quickly change the trial or sample number for file names while living unobtrusively in the corner of your desktop.

### Feature Breakdown



1. Toggle to Full Namer
2. Date
3. Sample Number
4. Trial Number
5. Copy Button
6. Close Button

## Logger

Want a summary of all of the files you've named? Logger records each file name you've copied and provides a timestamp of each entry.

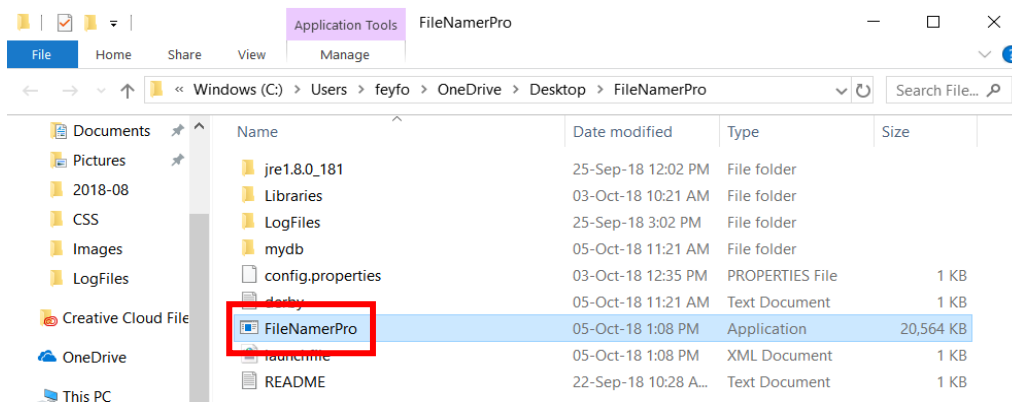
Project Name: Example Project							
Project Description: example project description							
DATE	TIME	RESEARCHER NAME	EXPERIMENT TYPE	TRIAL NUMBER	SAMPLE NUMBER	FILE NAME	COMMENT
05 10 18	11:30:13.4	John Orr	Generic Experiment	1	1	2018*10*5*GE*JO*T1*S1*3hrs	
05 10 18	11:30:15.797	John Orr	Generic Experiment	2	1	2018*10*5*GE*JO*T2*S1*3hrs	
05 10 18	11:30:17.553	John Orr	Generic Experiment	3	1	2018*10*5*GE*JO*T3*S1*3hrs	This trial had errors.
05 10 18	11:30:19.085	John Orr	Generic Experiment	4	1	2018*10*5*GE*JO*T4*S1*3hrs	
05 10 18	11:30:20.795	John Orr	Generic Experiment	5	1	2018*10*5*GE*JO*T5*S1*3hrs	
05 10 18	11:30:24.235	John Orr	Generic Experiment	1	2	2018*10*5*GE*JO*T1*S2*3hrs	
05 10 18	11:30:28.958	John Orr	Generic Experiment	2	2	2018*10*5*GE*JO*T2*S2*3hrs	
05 10 18	11:30:30.614	John Orr	Generic Experiment	3	2	2018*10*5*GE*JO*T3*S2*3hrs	
05 10 18	11:30:32.18	John Orr	Generic Experiment	4	2	2018*10*5*GE*JO*T4*S2*3hrs	
05 10 18	11:30:34.202	John Orr	Generic Experiment	5	2	2018*10*5*GE*JO*T5*S2*3hrs	

# Chapter 3: Download and Launch Instructions

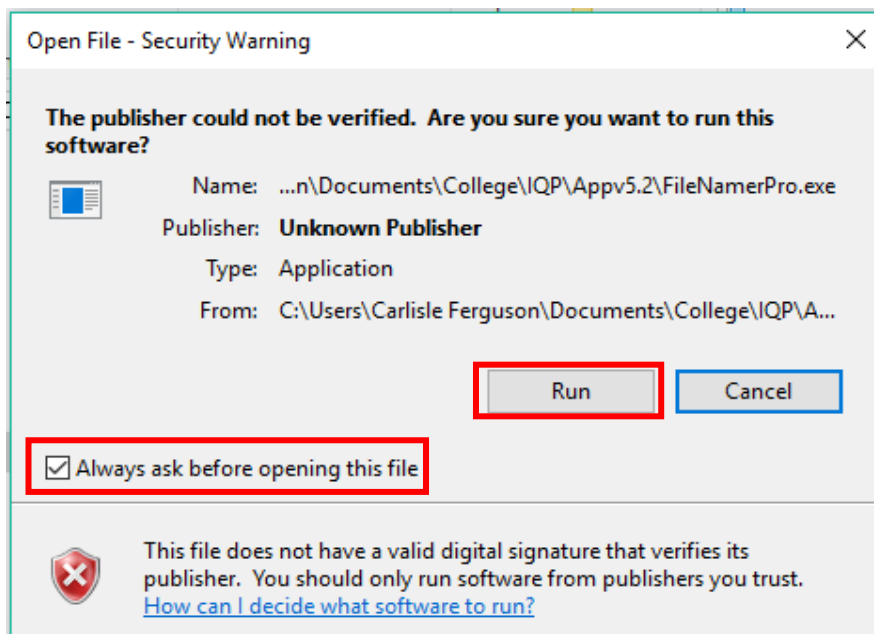
- 1 Download the zip file by clicking on the link here: [Link to zip file](#)
- 2 Unzip the file by right-clicking on the folder and click “Extract All...”. This may take several minutes.



- 3 In the unzipped folder, double-click on the FileNamerPro application.



- 4 If the following screen appears, uncheck “Always ask before opening this file” and click “Run”.



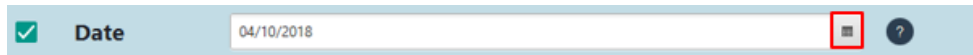
## Chapter 4: Quick Start Guide

This section contains general use instructions for File Namer Pro. These instructions can also be found in the “Getting Started” section of the application, which is located under the “Help” menu at the top left corner of Full Namer.

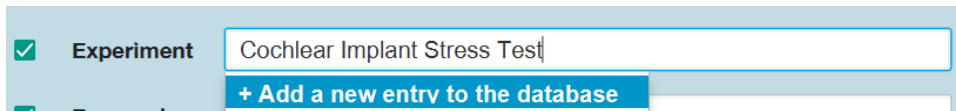
**Note: For more information on adding new entries to the Experiment, Researcher, and Keyword databases, please refer to pages 13-18 of the manual.**

### Using Full Namer

- 1 The “DATE” field will automatically be set to today’s date. To change the date, click on the calendar icon to the right of the date display field. If you do not want to include the date in the filename, uncheck the checkbox to the left.



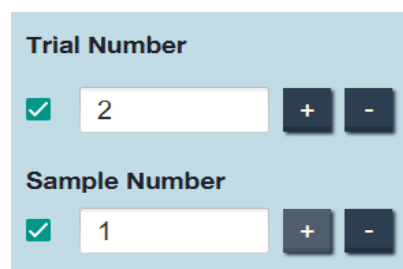
- 2 In the “EXPERIMENT NAME” text field, begin to type. Your entry should autofill. If your entry is not yet in the database, click on the “Add a new entry to the database” button and add the new experiment. If you do not want to include the experiment name in the file name, uncheck the checkbox to the left.



- 3 In the “RESEARCHER NAME” text field, type in your first and last name, and your name should autofill. If your name is not yet in the database, click on the “Add a new entry to the database” button and add the new researcher name. If you do not want to include the experiment name in the file name, uncheck the checkbox to the left.



- 4 In the “TRIAL NUMBER” and “SAMPLE NUMBER” fields, type in your desired trial number and sample numbers. Use the “+” and “-” buttons to increase and decrease the numbers. If you do not want to include the trial number or the sample number in the file name, uncheck the checkbox to the left.



- 5 To include additional keywords in your file name, click the “ADD” button.

**Additional Keywords Table**

Name	Data Value
Time in hours	3

**Add**

**Delete**

Begin typing in the “NAME” field, and your keyword should autofill. If your keyword is not yet in the database, click on the “Add a new entry to the database” button and add the new keyword. Fill in the “DATA VALUE” field if necessary, then click save.

**Choose a keyword from the database:**

**Name**

**Data Value**

**+ Add a new entry to the database**

**Save** **Cancel**

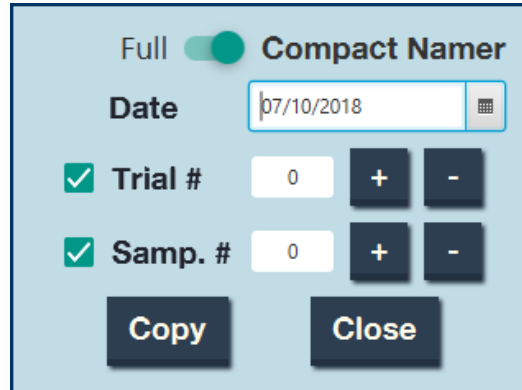
- 6 Click the “COPY” button to copy your generated file name to your clipboard, then paste the file name where necessary. Repeat steps 1-6 and modify text fields as needed.

**Filename Preview:**  **Copy** **Close**



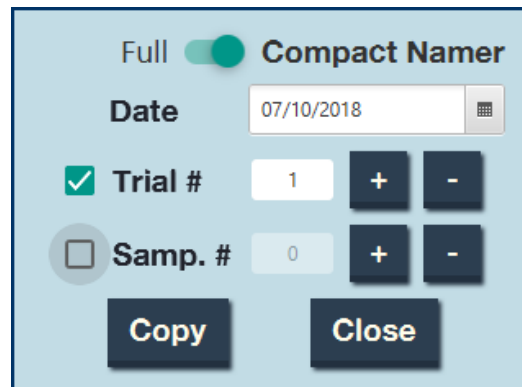
## Using Compact Namer

- 1 Click the toggle button at the top of Full Namer to navigate to Compact Namer.

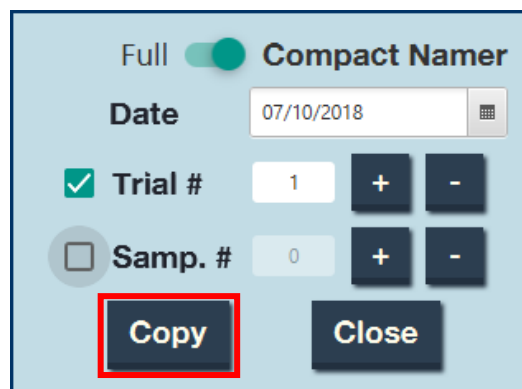
A screenshot of the 'Compact Namer' dialog box. At the top, there is a 'Full' toggle switch (turned off) and the text 'Compact Namer'. Below this is a 'Date' field with the value '07/10/2018'. There are two checked checkboxes: 'Trial #' with a value of '0' and 'Samp. #' with a value of '0'. Each checkbox has '+' and '-' buttons. At the bottom are 'Copy' and 'Close' buttons.

*Note: The experiment name, researcher name, any keywords you entered in Full Namer will be included in the file names created by Compact Namer unless you remove them from Full Namer.*

- 2 Modify the date, trial number, and sample number as necessary. Uncheck undesired fields.

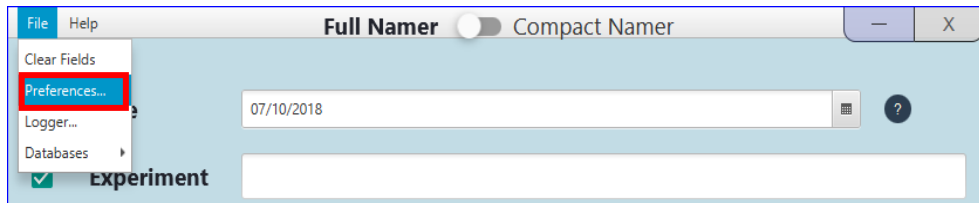
A screenshot of the 'Compact Namer' dialog box. The 'Date' field now shows '07/10/2018'. The 'Trial #' checkbox is checked and its value is '1'. The 'Samp. #' checkbox is unchecked and its value is '0'. The '+' and '-' buttons are present for both. 'Copy' and 'Close' buttons are at the bottom.

- 3 Click the "COPY" button to copy the file name to your clipboard and paste the file where necessary. Repeat steps 1-3 and modify fields as needed.

A screenshot of the 'Compact Namer' dialog box, identical to the previous one, but with the 'Copy' button highlighted by a red box.

## Setting Project Preferences

1 To save your name, the project name, the project description, and your preferred delimiter, hover over “FILE” in the top left corner of Full Namer, then click “Preferences...”.



2 In the next pop-up, type the researcher name in the “Researcher Name” text field, the project name in the “Project name” text field, and a description of the project in the “Project Description” text field. To change your preferred separation character, or delimiter, select one of the radio buttons. If you would like Full Namer to remember keywords associated with this project, check the “Remember Full Namer Fields” check box. Click “Save”.

### Project Preferences

**Researcher Name**

**Project Name**

**Project Description**

**Select a separation character:**

Asterisk (\*)     Hyphen (-)     Underscore ( \_ )

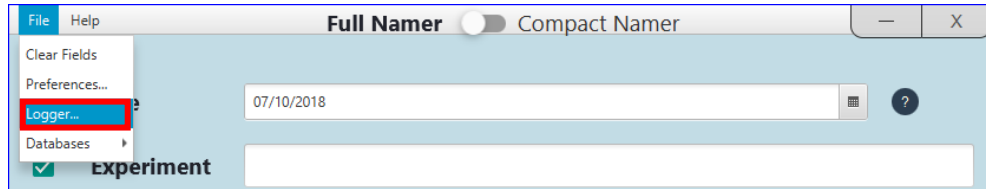
**Preview:**

Remember Full Namer fields

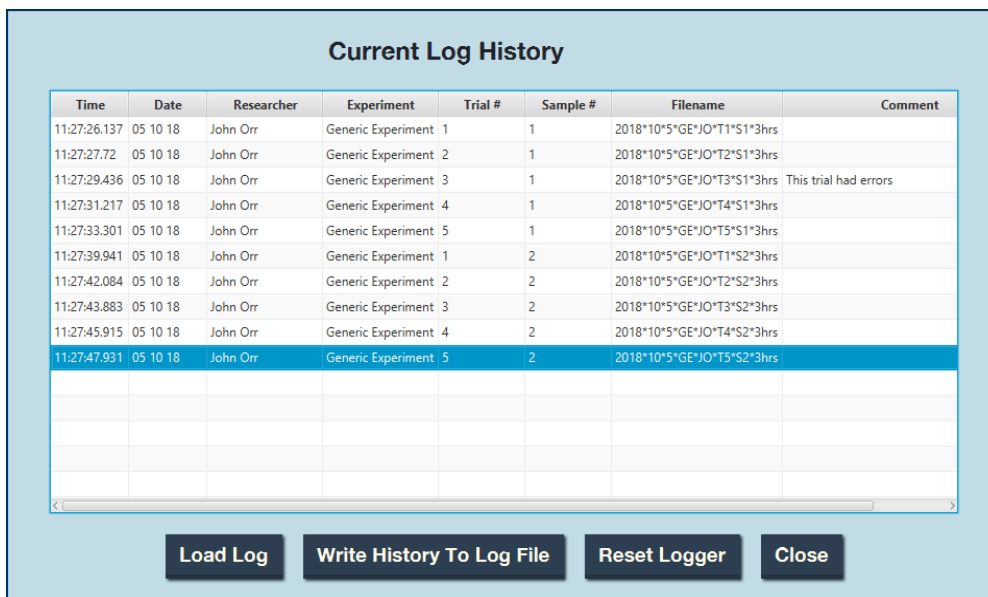
**Save**    **Cancel**

## Generating Log Files

- 1 Hover over "FILE" in the top left corner of Full Namer, then click "Logger...".



The following window will appear:



**Current Log History**

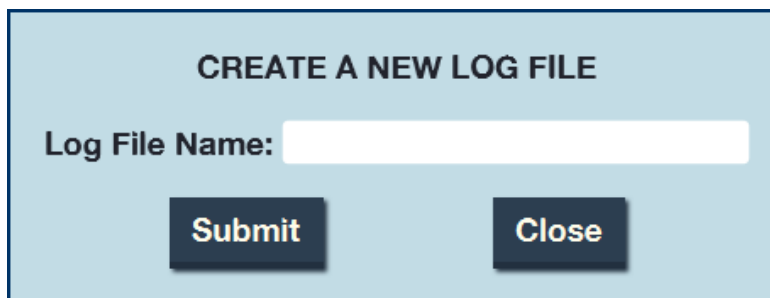
Time	Date	Researcher	Experiment	Trial #	Sample #	Filename	Comment
11:27:26.137	05 10 18	John Orr	Generic Experiment	1	1	2018*10*5*GE*JO*T1*S1*3hrs	
11:27:27.72	05 10 18	John Orr	Generic Experiment	2	1	2018*10*5*GE*JO*T2*S1*3hrs	
11:27:29.436	05 10 18	John Orr	Generic Experiment	3	1	2018*10*5*GE*JO*T3*S1*3hrs	This trial had errors
11:27:31.217	05 10 18	John Orr	Generic Experiment	4	1	2018*10*5*GE*JO*T4*S1*3hrs	
11:27:33.301	05 10 18	John Orr	Generic Experiment	5	1	2018*10*5*GE*JO*T5*S1*3hrs	
11:27:39.941	05 10 18	John Orr	Generic Experiment	1	2	2018*10*5*GE*JO*T2*S2*3hrs	
11:27:42.084	05 10 18	John Orr	Generic Experiment	2	2	2018*10*5*GE*JO*T2*S2*3hrs	
11:27:43.883	05 10 18	John Orr	Generic Experiment	3	2	2018*10*5*GE*JO*T3*S2*3hrs	
11:27:45.915	05 10 18	John Orr	Generic Experiment	4	2	2018*10*5*GE*JO*T4*S2*3hrs	
11:27:47.931	05 10 18	John Orr	Generic Experiment	5	2	2018*10*5*GE*JO*T5*S2*3hrs	

Buttons: Load Log, Write History To Log File, Reset Logger, Close

*Note: The log history will show all file names you have copied to your clipboard until it is reset by clicking "Reset Logger".*

- 2 This log history can be printed to a new or existing Excel log file.

To print to a new Excel file, click on the "Write History to Log File". The following pop-up will appear:



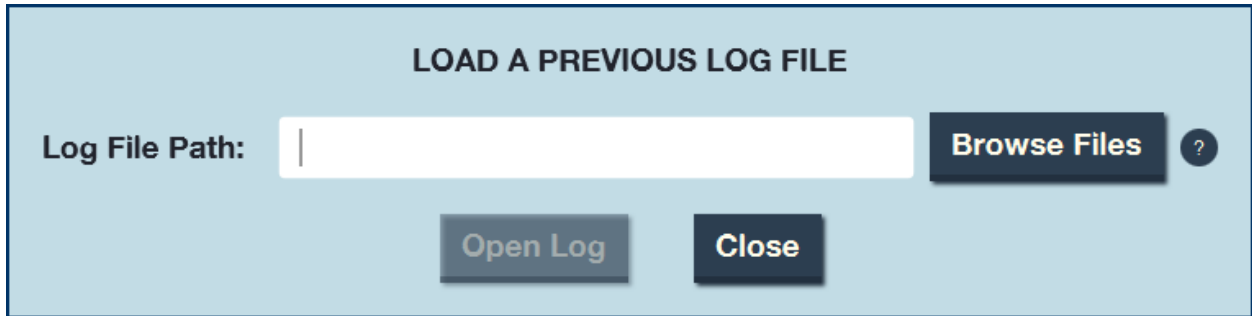
**CREATE A NEW LOG FILE**

Log File Name:

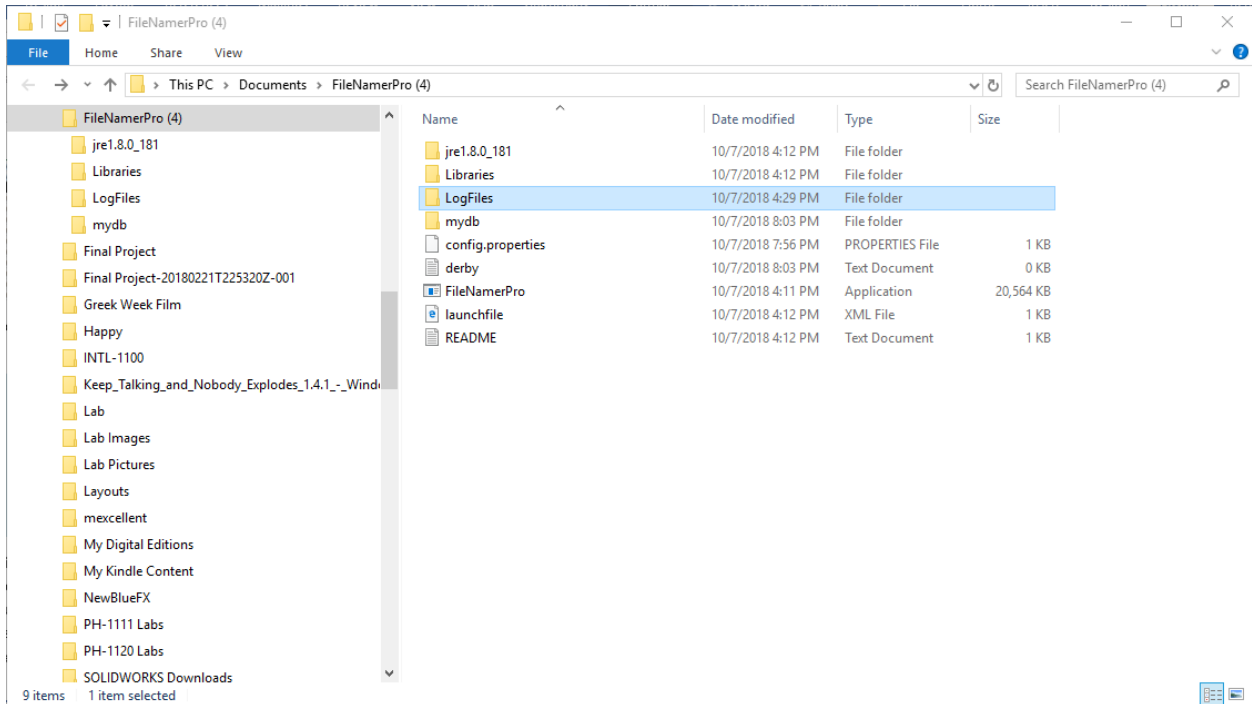
Buttons: Submit, Close

Type in your desired log name file, then click "Submit". To view your log file, navigate to the application file folder and click on "LogFiles". Your log file will be listed in that folder.

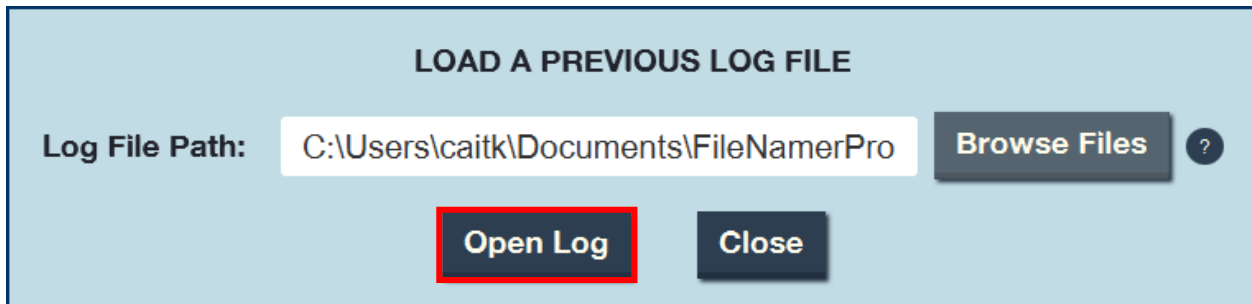
To print to an existing Excel file, click on the “Load Log” button. The following pop-up will appear:



Click “Browse Files” to browse for your log file. This will be located in the application folder, within the subfolder “LogFiles”.



Double-click your desired log file, then click “Open Log”.



Click on “Write History To Log File”.

**Current Log History**

Time	Date	Researcher	Experiment	Trial #	Sample #	Filename	Comment
11:27:26.137	05 10 18	John Orr	Generic Experiment 1	1	1	2018*10*5*GE*JO*T1*S1*3hrs	
11:27:27.72	05 10 18	John Orr	Generic Experiment 2	2	1	2018*10*5*GE*JO*T2*S1*3hrs	
11:27:29.436	05 10 18	John Orr	Generic Experiment 3	3	1	2018*10*5*GE*JO*T3*S1*3hrs	This trial had errors
11:27:31.217	05 10 18	John Orr	Generic Experiment 4	4	1	2018*10*5*GE*JO*T4*S1*3hrs	
11:27:33.301	05 10 18	John Orr	Generic Experiment 5	5	1	2018*10*5*GE*JO*T5*S1*3hrs	
11:27:39.941	05 10 18	John Orr	Generic Experiment 1	1	2	2018*10*5*GE*JO*T1*S2*3hrs	
11:27:42.084	05 10 18	John Orr	Generic Experiment 2	2	2	2018*10*5*GE*JO*T2*S2*3hrs	
11:27:43.883	05 10 18	John Orr	Generic Experiment 3	3	2	2018*10*5*GE*JO*T3*S2*3hrs	
11:27:45.915	05 10 18	John Orr	Generic Experiment 4	4	2	2018*10*5*GE*JO*T4*S2*3hrs	
11:27:47.931	05 10 18	John Orr	Generic Experiment 5	5	2	2018*10*5*GE*JO*T5*S2*3hrs	

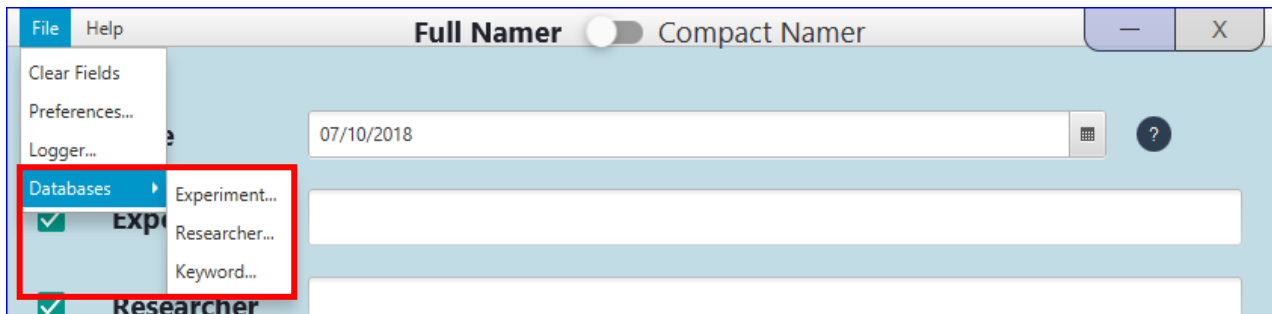
**Load Log**   **Write History To Log File**   **Reset Logger**   **Close**

In the pop-up window, type in the name of the existing log file you are trying to print to, then click “Submit”. To view your log file, navigate to the application file folder and click on “LogFiles”. Your log file will be listed in that folder.

# Chapter 5: Adding to the Databases

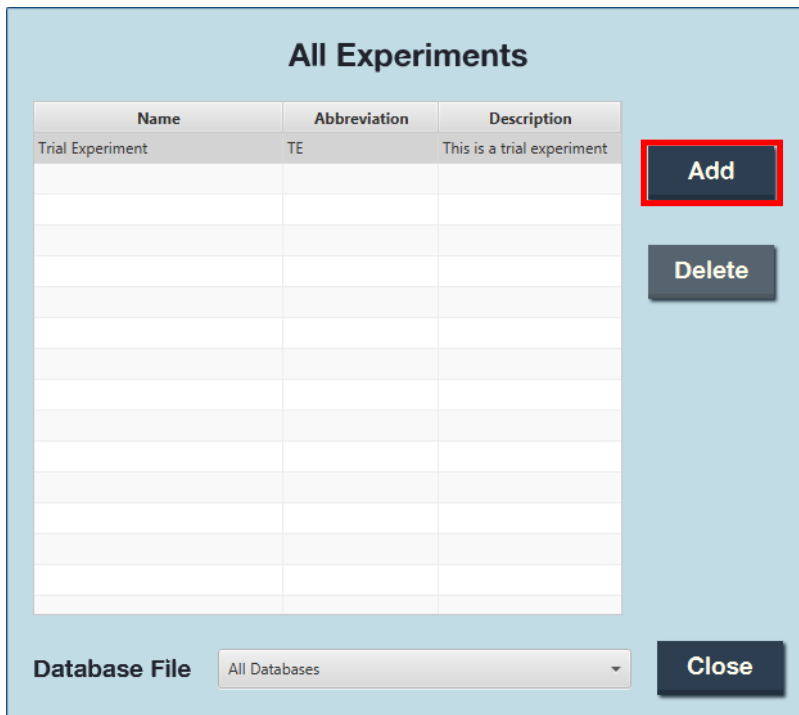
There are three databases in File Namer Pro: experiment names, researcher names, and keywords. This section provides instructions on how to add entries to these databases.

The databases can be accessed via the “FILE” menu in the top left corner of Full Namer or by clicking the “Add a new entry to the database” option in the drop-down menu of the text fields.



## Adding to the Experiment Name Database

- 1 In the “FILE” menu in the top left corner of Full Namer, hover over “Databases” and click “Experiment...”. In the pop-up shown below, click “Add”.



- 2 In the next pop-up, enter in the full name of the experiment in the “Full Name” text field, your desired abbreviation for the experiment name in the “Abbreviation” text field, and a description or any comments in the “Experiment Description” text field. Click “Save”.

**ADD NEW EXPERIMENT TO DATABASE**

**Full Name**

**Abbreviation**

**Database**

**Save** **Cancel**

- 3 Your new experiment entry will now appear in the “All Experiments” screen and will be usable in Full Namer when you start typing in the “Experiment” text field.

**All Experiments**

Name	Abbreviation	Description
Trial Experiment	TE	This is a trial experiment
Depth Measurement	DM	This experiment will tes...

**Add**

**Delete**

**Database File**

**Close**

## Adding to the Researcher Name Database

- 1 In the “FILE” menu in the top left corner of Full Namer, hover over “Databases” and click “Researcher...”. In the pop-up shown below, click “Add”.

Name	Abbreviation
James Bond	JB
John Orr	JO
Caitlin Kyaw	CK

Database File: All Databases

- 2 In the next pop-up, enter the researcher’s full name in the “Full Name” text field and the desired abbreviation in the “Abbreviation” text field. Then, click “Save”.

ADD NEW RESEARCHER TO DATABASE

Full Name: John A Smith

Abbreviation: JAS

Database: defaultResearchers.csv

Save Cancel



- 3 Your new researcher name entry will now appear in the “All Researchers” screen and will be usable in Full Namer when you start typing in the “Researcher” text field.

The screenshot displays a window titled "All Researchers" with a light blue background. At the top center is the title "All Researchers". Below it is a table with two columns: "Name" and "Abbreviation". The table contains four rows of data:

Name	Abbreviation
James Bond	JB
John A Smith	JAS
John Orr	JO
Caitlin Kyaw	CK

To the right of the table are two dark blue buttons: "Add" and "Delete". At the bottom left, there is a label "Database File" and a dropdown menu currently showing "All Databases". To the right of the dropdown is a dark blue button labeled "Close".

## Adding to the Keywords Database

- 1 In the “FILE” menu in the top left corner of Full Namer, hover over “Databases” and click “Keyword...”. In the pop-up shown below, click “Add”.

Name	Abbreviation	Affix	Data Type
Well number, 364-well plate	364WP	suffix	alphanumeric
Well number, 96-well plate	96WP	suffix	alphanumeric
Time in hours	hrs	suffix	numeric
Daytime	day		no data
Time in minutes	mins	suffix	numeric
Frequency	Hz	suffix	numeric
Nighttime	night		no data

Buttons: Add, Delete, Close

Database File: All Databases

- 2 In the next pop-up, enter the keyword’s full name in the “Full Name” text field and the desired abbreviation in the “Abbreviation” text field. Select the data type and indicate whether the abbreviation is a prefix or suffix, if applicable. Then, click “Save”.

*Note: Here, “Data Type” indicates the type of values that accompany the keyword. For example, frequency would need a numeric data type, such as “40”, to be 40Hz. To indicate a well position, you may select an alphanumeric data type, like “A3”.*

**ADD NEW KEYWORD TO DATABASE**

Full Name: Depth (millimeters, mm)

Abbreviation: mmD

Data Type:  numeric  alphanumeric  none

Abbreviation Affix:  prefix  suffix

Preview: 2mmD ?

Database: defaultKeywords.csv

Buttons: Save, Cancel

*Note: Here, “Abbreviation Affix” indicates the position of the keyword in relation to the data type. For example, a dollar sign, “\$” would be positioned before the numeric data type, and would therefore be a prefix. Hertz, “Hz”, would be positioned after the numeric data type, and would therefore be a suffix.*

- 3 Your new keyword entry will now appear in the “All Keywords” screen and will be usable in Full Namer when you add keywords to the table.

### All Keywords

Name	Abbreviation	Affix	Data Type
Well number, 364-well plate	364WP	suffix	alphanumeric
Well number, 96-well plate	96WP	suffix	alphanumeric
Time in hours	hrs	suffix	numeric
Daytime	day		no data
Time in minutes	mins	suffix	numeric
Depth (millimeters, mm)	mmD	suffix	numeric
Frequency	Hz	suffix	numeric
Nighttime	night		no data

**Add**

**Delete**

**Database File** All Databases **Close**

## Chapter 6: Modifying the Code

The code repository can be found on GitHub at <https://github.com/jyxiao1/filenamerpro>. In order to successfully modify the code, a working knowledge of Git, Java, and Gradle is recommended.

To make modifications, you will need the following software:

- Git
- Gradle
- SceneBuilder
- Java SE Development Kit 8 (JDK 1.8.0)



**WPI**



**UniversitätsSpital  
Zürich**