

# The Integration of the Brain Bank Imaging Workflow into the Infrastructure of the Multiple Sclerosis Research Network

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades

"Doctor rerum naturalium"

der Georg-August-Universität Göttingen

im Promotionsprogramm in Computer Science (PCS)

der Georg-August University School of Science (GAUSS)

vorgelegt von

Lika Svanadze

aus Georgien



Göttingen, 2019

**Betreuungsausschuss:**

Prof. Dr. Otto Rienhoff, Institut für Medizinische Informatik, Universitätsmedizin Göttingen

Prof. Dr. Dieter Hogrefe, Institut für Informatik, Universität Göttingen

Prof. Dr. Ulrich Sax, Institut für Medizinische Informatik, Universitätsmedizin Göttingen

**Mitglieder der Prüfungskommission:***Referent:*

Prof. Dr. Otto Rienhoff, Institut für Medizinische Informatik, Universitätsmedizin Göttingen

*Korreferent:*

Prof. Dr. Wolfgang Brück, Institut für Neuropathologie, Universitätsmedizin Göttingen

*Weitere Mitglieder der Prüfungskommission:*

Prof. Dr. Dieter Hogrefe, Institut für Informatik, Universität Göttingen

Prof. Dr. Ulrich Sax, Institut für Medizinische Informatik, Universitätsmedizin Göttingen

Prof. Dr. Jens Grabowski, Institut für Informatik, Universität Göttingen

Prof. Dr. Wolfgang May, Institut für Informatik, Universität Göttingen

Tag der mündlichen Prüfung: 27.02.2019

*To my parents, sister, brother, and lovely grandmother...*

# Acknowledgements

*“If I have seen further,  
it is by standing on the shoulders of giants.”*

*- Sir Isaac Newton*

I wish to express my sincere gratitude to the following individuals for their invaluable support and guidance when I studied and lived in Göttingen. It is a pleasure to express my thanks to all those who contributed in many ways to the success of this study and made it an unforgettable experience for me.

First and foremost, I would like to thank my supervisor Prof. Dr. Otto Rienhoff for his continuous support, encouragement, and valuable care. I thank him for giving me the opportunity to work with the great people at the Department of Medical Informatics. I had absolute freedom to find scientific topics and research questions, which he always supported with his excellent tips and inspirational ideas. Thus, digital pathology became my favorite research area – one in which I have a serious desire to continue working in future. I am thankful for his support in expending my scientific knowledge through participation in international conferences. I will always remember his advices and if I stay in the academic field, I will share them with my young colleagues. Thank you for placing your trust and confidence in my abilities.

My special thanks go to my second supervisor Prof. Dr. Dieter Hogrefe for his comments and suggestions, which gave me the ability to always take my research in the right direction. I would like to sincerely thank Prof. Dr. Wolfgang Brück for his support and great cooperation within the KKNMS research project. My deepest gratitude also goes to the thesis committee member Prof. Dr. Ulrich Sax for his valuable time, constant readiness for discussion, suggestions for overcoming challenges during the research work, and sharing his vast knowledge in Medical Informatics, which helped me to perceive my research from different perspectives.

I can not express my gratitude in words to Thomas Franke, who supported me a lot during these years and provided constant scientific feedback throughout the whole project and research work. My sincere thanks go to Romanus Diekmann for his significant contribution in the development phase of the project. I also want to thank Karoline Buckow and Tina Meißner for their support to the research project in an overwhelming way. I want to acknowledge my colleagues Christoph Jensen, Andrea Mostarac, Sebastian Kunze, Sara Nußbeck, Mahsa Lee and Svenja Wolff for their help whenever needed. I am thankful to all members of the Department of Medical Informatics who gave me valuable advices, paved the way for my research, and always had time to respond to my questions. It was a huge pleasure working with all of you.

I want to thank Dr. Erik Bahn for close and successful collaboration on the very interesting research project and for valuable scientific input in my research work. I also would like to thank Dr. Gertrude Bunt for giving me the opportunity to operate microscope scanners very often and for being always ready to solve the problems related to the scanner. I thank Mr. Heimrich and all members of IT department, who always tried to solve all my problems related to the server, ports, firewall, etc. Also, my special thanks go to the Olympus project manager, Dr. Daniel Göttel, for his continuous and quick support, valuable advice, and suggestions for any problems that I had during the project.

My deepest gratitude also goes to Prof. Ralf Meyer, who visited my native country Georgia in 2011, met undergraduate students at Ivane Javakishvili Tbilisi State University (TSU), and advised us to continue our professional development at the University of Göttingen. I thank him and the professors from TSU who guided me and gave their recommendations for my admission at the University of Göttingen.

I want to acknowledge the following organizations: Alrakis project of Erasmus Mundus, University Medical Center Göttingen, and the International Education Center of Georgia for funding my study in Göttingen.

I want to thank my friends for supporting me and being a priceless gift in my life. Incomparable friends from TSU: Eka, Tamta, Lika, Qeti, David, and Lasha; and marvelous friends from Tbilisi State Medical University: Nincho, two Eka, Meko, Levan, Ani, and Giorgi. I want to thank my friends whom I met in Göttingen: Alex for his continuous support and being an example of a great friend and humaneness; Marika for her wonderful support from very first day of my stay in Göttingen; Giorgi, Gvantsa, Nina, and my international friends - Emma, Menu, and DeeJay for their support and wonderful entertainment. Thank you from the bottom of my heart my dear friends.

Finally, I want to express my gratitude to my parents, my beloved sister and brother for everything that they are doing for me throughout my entire life. I also want to thank my lovely aunts – Nino, Ketino, Tamar, Tina, and my cousins – Mariam, Sandro and Parna – for their unconditional love and support. At the end, I want to thank the oldest member of my family, modern 94-year-old grandmother Mimi, who always prays for me and is my guardian angel.

# Abstract

Digital pathology (DP) provides powerful tools for the acquisition, management, storage, viewing, and analysis of high-resolution digital microscope images. During the last decade, DP tools were increasingly used for a number of applications including biobank, research, education, remote consultations, etc. However, full implementation of the Digital Pathology System (DPS) into the laboratory workflow has remained still challenging in terms of regulatory, technical, financial, and interoperability obstacles. This thesis presents a tightly integrated image handling system for the biobank application within the complex IT infrastructure of the German Competence Network Multiple Sclerosis (KKNMS) research network. The developed concept addresses all steps started from a very early phase of planning and ends with the evaluation, operation, and maintenance of the image handling system for biorepository within the University Medical Center Göttingen (UMG).

In the beginning, the thesis represents the requirements engineering process to address the needs for setting up the complex IT infrastructure for biobank purposes. The requirements were mostly elicited from regular meetings with the direct stakeholders of the KKNMS comprising the main pathologist of the brain bank, research associates, and IT professionals. Additional features were also collected with regard to the scientific literature analysis in already existing DPS established by different pathology laboratories and the market analysis for available image handling systems. Multiple requirements were gathered, determined, and categorized into the functional, technical, quality, and environmental requirements that had been prioritized as essential, preferred, and desirable features of the system.

Once the requirements were determined, a digital workflow for the laboratory processes has been developed and an interoperable image handling system has been set up according to needs. Prior to the digitization of the glass slides, labeling and identification mechanism for more than 4,000 glass slides of the biorepository were determined. For the digitization of the glass slides, hardware and software components of the DPS were configured and customized in such a way that provides an effective approach for acquisition, management, and storage of digital images. Using the automated tools, the scanning procedures were accelerated and the quality of digital workflow was enhanced by avoiding manual steps.

Finally, to increase the efficiency of the separate systems within the biobank infrastructure, an interfacing solution was developed between the DPS and the Laboratory Information Management System (STARLIMS) where the specimens are registered. By using the automated and manual tools, multiple imaging data stored on the DPS can be linked to the corresponding specimen in STARLIMS.

Thus, the thesis provides an effective solution for the interoperable image handling system applied to the biobank research environment, which can be improved further in response to the needs of stakeholders and researchers of the KKNMS. Additionally, it can be adapted to different research needs within the UMG in the future.

# Contents

<b>Acknowledgements</b> .....	<b>i</b>
<b>Abstract</b> .....	<b>ii</b>
<b>1. Introduction</b> .....	<b>1</b>
1.1 Problem statement .....	1
1.1.1 Regulatory and standardization barriers .....	2
1.1.2 Information technology obstacles .....	4
1.1.3 Digital workflow challenges .....	4
1.1.4 The high costs of digital pathology infrastructure .....	5
1.1.5 Challenges of interoperability .....	5
1.1.6 Quality management of imaging systems .....	6
1.2 Research questions and objective of the thesis .....	6
1.3 Main methods and contributions of the thesis .....	8
1.4 Structure of the thesis .....	11
<b>2. Foundations</b> .....	<b>13</b>
2.1 Digital imaging informatics in medicine .....	13
2.1.1 Image acquisition .....	15
2.1.2 Image representation .....	16
2.1.3 Image management and storage .....	17
2.1.4 Image processing and interpretation .....	20
2.2 Digital pathology .....	21
2.2.1 Workflow of digital pathology .....	22
2.2.2 Benefits and limitations of digital pathology .....	23
2.2.3 Applications of digital pathology .....	24
2.3 Initial literature analysis in digital pathology .....	25
2.3.1 Historical review and current status of digital pathology .....	25
2.3.2 Regulations, guidelines, and standards in digital pathology .....	26
2.4 Research environment of the application .....	32
2.4.1 Common architecture framework for biobank research networks in Germany .....	32
2.4.2 Research infrastructure at UMG .....	33
2.4.3 Digital pathology system architecture at UMG .....	36
2.4.4 German Competence Network Multiple Sclerosis research consortia .....	37
2.5 Related work .....	38
<b>3. Requirements engineering for an integrated brain bank imaging system</b> .....	<b>40</b>
3.1 Requirements engineering framework .....	40
3.1.1 Planning of requirements analysis .....	42
3.1.2 Requirements elicitation sources and techniques .....	43
3.1.3 Interpretation of requirements .....	46
3.1.4 Negotiation and analysis of requirements .....	46
3.1.5 Requirement specification and modeling techniques .....	47
3.1.6 Requirements validation .....	47
3.1.7 Requirements management .....	48
3.2 Results of requirements engineering for the interoperable DPS .....	49
3.2.1 Requirements elicitation questions .....	49
3.2.2 General requirements of the DPS .....	51
3.2.3 Use case scenarios in the requirements engineering .....	52
3.2.4 Market analysis of existing digital microscopic image handling systems .....	53

<b>4.</b>	<b>Development of the digital workflow and the interoperable digital image handling system ..</b>	<b>56</b>
4.1	Preparation of glass slides .....	56
4.1.1	Multiple sclerosis tissue sample processing in a pathology laboratory .....	56
4.1.2	The information management system for a pathology laboratory .....	58
4.1.3	Hierarchical structure of pathology data.....	59
4.1.4	Identification mechanism and barcode labeling concept for glass slides .....	60
4.2	Digitization of glass slides .....	70
4.2.1	Digitization strategy for sample selection .....	70
4.2.2	Digital pathology workstation .....	70
4.2.3	Scanning settings .....	71
4.2.4	Batch scanning mode and metadata entry .....	73
4.2.5	Automatic naming and saving of images.....	75
4.2.6	Scanning results.....	75
4.3	Image server and database.....	77
4.3.1	Image server settings and network configuration parameters.....	77
4.3.2	Database and structure of image folders .....	79
4.3.3	Data storage estimation for large volumes of growing data.....	80
4.3.4	Overview of user management processes .....	82
4.4	Image web viewer for the researchers of the KKNMS.....	83
4.4.1	Requirements for the image web viewer .....	83
4.4.2	Installation and deployment of web application.....	84
<b>5.</b>	<b>Realization of the interfacing solution for the DPS and evaluation of working prototype.....</b>	<b>86</b>
5.1	Interoperability platform of targeted systems .....	86
5.2	Automated and manual methods for data exchange .....	89
5.2.1	General architecture of web services.....	89
5.2.2	SOAP-based web service middleware application .....	90
5.2.3	Image metadata viewing interface in the biobank management system .....	96
5.2.4	Manual method for the addition of imaging data.....	97
5.3	Evaluation of the prototype.....	100
5.4	Concluding literature analysis of the currently existing DP frameworks and interfacing standards in pathology .....	103
<b>6.</b>	<b>Discussion .....</b>	<b>107</b>
6.1	Contributions to the optimization of the Brain Bank workflow .....	107
6.1.1	Regulatory and standardization aspects .....	107
6.1.2	Requirements engineering for the microscopic imaging system .....	108
6.1.3	Development of the digital image handling system.....	109
6.1.4	Interfacing solution between the DPS and the LIMS.....	111
6.2	Advantages and limitations of the developed solution .....	113
6.3	Outlook and future development .....	114
	<b>References .....</b>	<b>116</b>
	<b>Appendices .....</b>	<b>123</b>
	Appendix A . Comparison of the DPA, CAP, Canadian, ATA, and RCP guidelines .....	124
	Appendix B . Use case scenarios for digital pathology workflow .....	125
	Appendix C . Result of the market analysis in digital image handling systems and viewers.....	135
	Appendix D . Results of test scanning process.....	139
	Appendix E . Metadata for multiple glass slides .....	143
	Appendix F . Accompanied DVD .....	145
	<b>List of Figures .....</b>	<b>146</b>
	<b>List of Tables .....</b>	<b>147</b>
	<b>Abbreviations.....</b>	<b>148</b>
	<b>Curriculum Vitae .....</b>	<b>150</b>

# 1. Introduction

Digital pathology (DP), also known as virtual microscopy, is a technology that creates high-resolution, whole slide digital images of thin tissue samples mounted on glass microscopic slides. DP uses software and hardware components to digitize, view, manage, store, analyze, and transmit microscopic images using a computer. Because microscopic images are likely to be used in a digital format, whole slide imaging tools are being used for different applications such as research, education, remote consultation, and clinical diagnosis.

## 1.1 Problem statement

The development and adoption of digital pathology systems (DPSs) within large healthcare organizations is a complex task and requires collaboration between pathologists, scientists, technologists, and IT specialists. This fact and the very high costs associated with the required infrastructure can be attributed to the slow progress of DP technology over the last two decades.

Because the topic is multidisciplinary, major challenges in the adoption and maintenance of tightly integrated DPS within a pathology laboratory environment should be enumerated and discussed from multiple perspectives (Figure 1.1).

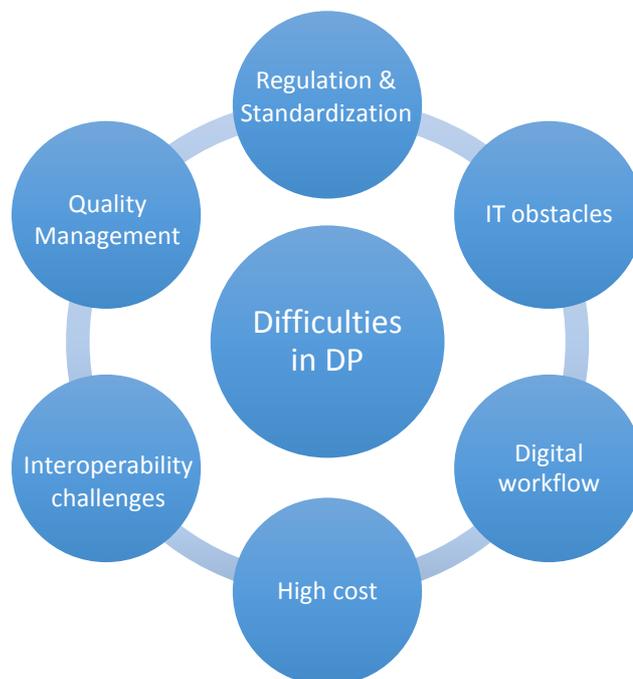


Figure 1.1. Difficulties in DP. The major challenges in DP are related to these complex interlinked components.

### 1.1.1 Regulatory and standardization barriers

Regardless of the benefits of DP, very few laboratories in the world have fully integrated it into their routine diagnostics up until now. DP is more widely used for small-scale research and educational purposes. This disparity is mainly caused by a lack of compliance with complex national regulatory environments [1]. For example, the Food and Drug Administration (FDA), which is responsible for protecting and promoting public health through regulation and supervision in the United States, considers whole slide imaging (WSI) scanners as Class III<sup>1</sup> devices and designates them as the highest risk devices. Therefore, Class III devices are labeled as unsatisfactory for diagnostic purposes in the United States and require safety and effectiveness evaluations by the FDA to be subject to premarket approval [3]. The Digital Pathology Association (DPA) collaborates with the FDA with regard to the shift of WSI devices from Class III to Class II, and later to Class I, which would allow them to be used for clinical (primary) diagnosis [4, 5]. Unlike the United States, several WSI devices have been certified by the European Conformity (CE) mark for clinical diagnosis under certain circumstances in the European Union [6, 7]. However, precise regulations for the validation of WSI systems are needed in Europe as well. These regulatory restrictions result in the slow development of DPSs and workflows.

DPS vendors install their products (e.g., scanners) worldwide, but guidance on their usage and harmonization of international standards is lacking [8]. Practical guidelines for WSI systems have been developed by various leading organizations in the United States, Canada, Europe, and Australasia [9–14]. The principal objective of the guidelines is to create a quality management system and to promote a standardization process for WSI systems. Thus, these guidelines may have a significant impact on the quality of complex research projects at pathology laboratories. However, their implementation remains a significant challenge to professional bodies [1].

Currently there is no standard approach for handling and transferring digital microscopic images. Since there are different scanner vendors, standardizing digital images remains challenging. Digital Imaging and Communications in Medicine (DICOM) is a de-facto standard

---

<sup>1</sup> The FDA classifies medical devices into three groups according to their risks. Class I devices are seen to be low-risk devices, and they require the least regulatory controls. Class II devices are seen as higher risk (called moderate risk) than Class I devices, and are subject to tighter regulatory controls. Class III medical devices are under the highest risk, and require regulatory controls that assure device safety and effectiveness [2]. As conventional microscopes are classified as Class I devices, they are applied for routine diagnostics.

for storage, transfer, and the manipulation of medical images. Initially, DICOM was developed for radiological images like magnetic resonance imaging (MRI), computed tomography, radiography, etc. In 2010, the DICOM Working Group 26 officially published supplement 145, which describes the workflow for digital microscopic images using a DICOM standard [15]. However, as pathology images significantly differ from any other medical images, several limitations in the DICOM standards have arisen for microscope images (such as file size restrictions, lack of specimen information, etc.). Very close collaboration between pathologists, scientists, IT technicians, and scanner vendors is necessary for the wide adoption of DICOM in the pathology field.

An anatomic pathology working group established by Health Level-7 (HL7)<sup>1</sup>, has tried to improve and customize the current HL7 standards for pathology laboratories. The main goal of the working group is to maintain interoperability between different information systems and define the relationship between specimens, observations, patient information, and images [16]. Because the specifications of these components differ from each other, they are thus typically managed and handled by different information systems, such as a patient management system, a laboratory information management system, an image handling system and so on. Therefore, data sharing between these systems is required to provide improved patient care and manage information more effectively. The Integrating the Healthcare Enterprise (IHE) provided the Anatomic Pathology Technical Frameworks that describe the integration of information systems and encourage pathology laboratories to use the existing standards (such as HL7, DICOM, ISO, etc.), rather than to create new standards [16–18]. However, this highly collaborative initiative requires intensive work to make the standards applicable to most of the microscopic scanners and anatomic pathology information systems.

Even though the above-mentioned regulations and guidelines are mostly intended for diagnostic purposes, some specifications of standards and guideline suggestions should also be taken into account during the adoption and implementation of a digital image processing system for different applications, including biobanks, research, or remote consultation. This type of approach can facilitate, standardize, and support the use of DP tools for various applications within the same medical environment.

---

<sup>1</sup> HL7 provides standards for the transfer of electronic health information.

### **1.1.2 Information technology obstacles**

Although rapidly developed information technologies (IT) have had a positive impact on laboratory workflow, significant IT difficulties exist in the field of DP [19]. The main challenges related to IT are large digital image file sizes and the growing number of images that require proper secured storage environments. The size of a single digital image file varies from hundreds of megabytes up to several gigabytes depending on the scanning settings and size of the scanning area. Images can be scanned with 20, 40, 60, or 100 times magnification, which leads to the large size of the file. Because there are multiple microscopic images that are much larger than any other medical images, a high throughput storage environment is required in order to ensure a high availability to respond to the request, as well as scalability to handle the growing amount of data. Some pathology laboratories prefer to save on costs and use cloud storage for their digital images. This allows the images to be available via the internet. The IT infrastructure can also be managed with virtualization models. According to the EU data protection regulations, organizations should determine the appropriate safety solution for their sensitive data [20, 21]. If the data is stored in the cloud, the service provider is required to ensure the data protection. Cloud storage services raise many questions regarding security and privacy issues, particularly for personal medical data [22].

As the IT obstacles vary based on special needs, the proper estimation of IT infrastructure is required for each individual application.

### **1.1.3 Digital workflow challenges**

Even though digital workflow provides improved ergonomics and efficient tools for pathologists, it also faces many difficulties. Pathologists need extra time to digitize glass slides and store images on the server or external hard disk with enriched metadata. The standardization of the digital workflow is therefore required for the effective management of the processes and efficient usage of DP tools [23]. Some pathologists have reported that viewing slides using a computer is aggravating and unreliable. Therefore, they prefer to use the conventional microscope viewing method [24]. In order to reduce the hassles and provide a better workplace for pathologists, it is important to implement automated and user-friendly tools in a digital workflow that optimize processes and improve the productivity of pathologists [25]. In addition, special training and instruction are essential to building pathologist confidence and to help them to better understand how to operate a DPS.

#### **1.1.4 The high costs of digital pathology infrastructure**

The purchase, setting up, and maintenance of complex DPS for routine diagnosis, education or research purposes is associated with very high costs [26, 27]. This is one of the significant reasons why DP is not widely used in most pathology laboratories. At the planning phase of the adoption and implementation of DPS, total cost should be considered and a focus needs to be placed on the most important project requirements [28]. Total cost includes the expenses for the scanning equipment, storage environment, training of personnel, technical support, user licenses, and system updates [29]. In addition, there are extra costs associated with frequently used procedures, such as purchasing durable labels or label printing services. The price of a digital scanner is rapidly increasing along with technological advances. Moreover, maintenance of an operable DPS is expensive since a WSI system requires continuous monitoring and improvement to keep the system up-to-date and relevant to the user's requirements. In order to reduce total cost for the adoption and operation of the system, it is important to evaluate the existing technological resources within the institution, which can be included into the DPS architecture during the development process. Oftentimes, the existing information systems and technologies found in healthcare organizations are commercially available and can be updated regularly and improved with the help of the manufactures.

#### **1.1.5 Challenges of interoperability**

As microscopic images are increasingly used in the digital format, their well-organized integration into another system, such as a Laboratory Information Management System (LIMS), becomes significantly more important. There are several types of data in pathology that are handled by distinct systems, such as microscopic images that contain significant information about a specimen. Regardless of the high demand of linkage between the various components of pathology data, a very limited number of information systems are capable of interfacing with each other. This lack of ability to exchange data is mainly caused by the different integration tools provided by manufacturers of targeted systems that vary from loosely coupled systems to tightly coupled systems [30]. Therefore, a comprehensive interoperability analysis of the functional units of targeted systems has to be performed to determine a method through which these systems can effectively be linked.

According to the annual reports and recommendations published by the IT review board of the Technology, Methods, and Infrastructure (TMF) for networked medical research in Germany, individually programmed applications for medical research can increasingly be substituted by the robust systems developed for a wide range of medical organizations [31]. Therefore, the report highlights the need for professional IT solutions that support complex customization and integration tools while handling large amounts of heterogeneous data. Additionally, a complex IT infrastructure that consists of multiple independent components, requires various levels of the integration of diverse applications.

#### **1.1.6 Quality management of imaging systems**

Quality management consists of quality control (QC) and quality assurance (QA). The main goal of QC is to fulfill the quality requirements of the system, while QA verifies the quality requirements fulfillment [32]. Both components are required to achieve the main objectives and deliver system operation in the right way. Quality management techniques are used to ensure the accuracy and quality of high-resolution scanned images. It also refers to the adequate usage of resources, such as information systems and technologies applied for operation of a microscope image handling system [33]. Validation and verification procedures should be performed for the successful delivery of an operable system. In addition, to reduce repetitive tasks and provide a better quality of service, standard operating procedures using automated and customized tools should be involved in the DP workflow. Currently, there is a lack of standard characteristics for estimating the quality of WSI systems. Therefore, each pathology laboratory has to identify its own quality management techniques for the individual imaging system used for different applications.

### **1.2 Research questions and objective of the thesis**

This PhD study is motivated by the German Competence Network Multiple Sclerosis (KKNMS) research project, in which histological samples are to be digitized and handled to facilitate the viewing of microscopic images from remote institutions. The aim of the study is to integrate new digital technologies for pathological practice into biobank research infrastructure, which also partly supports the future organization of digital microscopic images within the University Medical Center Göttingen (UMG).

The overall research question addressed in this study focuses on how to design and efficiently integrate a digital microscopic image handling system that enables the acquisition, management, sharing, viewing, analysis, and interpretation of digital slides into the laboratory management system in a research network environment.

This thesis emphasizes three main views: improving the workflow of the pathology laboratory, the efficient management of digital microscopic images, and the development of an interfacing solution for data sharing between heterogeneous systems within the complex biobank research environment. The general research question is addressed by considering the following sub-questions:

1. What kinds of regulatory and standardization aspects should be considered for the development and operation of a microscopic image handling system within the scope of the research environment?
2. What are the main requirements of the DPS in the environment of the collaborative research network?
3. How should the DPS be developed and deployed based on the requirements analysis?
4. How should a method be constructed for the seamless integration of the DPS into the existing collaborative IT infrastructure?

The major hypothesis in response to the problem stated above can be summarized as follows: **An operable DPS can be developed and effectively integrated into the complex IT infrastructure of the biobank research environment. The automated tools are able to be implemented in the pathology workflow that can improve and standardize the processes, as well as reduce repetitive tasks. Based on SOAP web service technology, the interfacing solution can lead to the efficient usage of DP tools for neuroscience researchers.**

### 1.3 Main methods and contributions of the thesis

The thesis workflow can be divided into the steps illustrated in Figure 1.2.



Figure 1.2. Thesis workflow. **Step 1** - An initial literature analysis provided to gain knowledge of the current state of DP; **Step 2** – A requirements analysis to collect and evaluate all features and characteristics of the DPS for biobank research; **Step 3** – The development and set up of the DPS to use digital tools in the pathology workflow; **Step 4** - The integration of the DPS into the laboratory management software; **Step 5** – The evaluation of the developed solution.

To answer the questions and find a solution to the problem, various methods have been used in each step of the thesis workflow:

#### **Step 1. Analyzing literature in the context of the validation and implementation of a DPS for different applications**

**Methods:** First, the most prevalent international regulations and guidelines centering on the validation and implementation of DP tools in clinical and nonclinical environments were analyzed. The literature this thesis employed was provided by regulatory bodies in Europe, the United States, Canada and Australasia. Furthermore, the existing standards in pathology, proposed by standardization bodies such as DICOM, HL7, and IHE, were evaluated for the biobank research network. In the end, IT frameworks for medical research environments suggested by TMF were explored in terms of image processing, data protection, and specimen identification.

#### **Step 2. Analyzing the requirements for the DPS in the biobank research environment**

**Methods:** At the beginning of this phase, a requirements engineering framework for the development of the integrated DP workflow was defined with two components – requirements analysis and requirements management. The complete process of requirements engineering was broken down into several sub-parts. First, the main stakeholders were considered and the system environment for which the DPS should be developed was defined. Second, requirements elicitation sources and techniques were determined using widely approved methods, such as documentation analysis, meetings with stakeholders, as well as process observation techniques [34]. Third, based on the questions list, requirements were gathered; prioritized as essential, preferred and desirable requirements; categorized into different groups like functional, technical, quality and environmental requirements.

Additionally, an investigation of microscopic image viewers and handling systems on the market was carried out. Based on this stage, supplementary features of the DPS were identified. In all, 19 web viewers for microscopic images based on various platforms were found and the main features, advantages and disadvantages of the most prevalent microscope imaging systems were assessed. Fourth, requirements specification and modeling techniques were determined. Therefore, five use case scenarios were designed to describe the digital workflow process from different points of view. To manage and balance different concerns effectively, the multiple viewpoints approach was used during the requirements engineering [35]. The following three viewpoints were identified: a process viewpoint to describe the processes of the whole workflow from various user perspectives, an information system viewpoint to define and characterize the appropriate information systems used for the management of processes, and the interfacing viewpoint to determine the interface solution between the different information systems [36]. Finally, the requirements validation was carried out using inspection and testing methods to ensure that the documented requirements met stakeholder needs. Together, these sub-steps were significantly important for requirements elicitation, analysis, and management. A detailed description of the methods used for the requirements engineering are described in Subchapter 3.1.

### **Step 3. Developing and starting operations for an interoperable solution of the digital image handling system for the biobank research network**

**Methods:** Based on the requirements analysis, fundamental modules of the DPS were evaluated. Prior to the digitization of the samples, methods for labeling and identification of glass slides were determined for the biorepository. Therefore, the key components of the labels were considered based on the current guidelines for the uniform labeling of slides [37]. In addition, label material and printing options were designed and the identification mechanism of specimen-derived assets in the current biospecimen management software was assessed. Most of the LIMS provide the identifiers (IDs) only for the case (brain) and its blocks (specimens) [38]. To improve the identification system for specimen-derived assets, a user-friendly interface was built in the local LIMS by an IT technician. This approach was implemented in the LIMS for two main reasons. First, each glass slide requires a unique ID that identifies not only the glass slide, but also the specimen from which the glass slide is derived. Therefore, the LIMS is the exact environment in which identifiers for specimen-derived assets should be generated. Second, it is highly important to export all components of the label in a

structured format for label printing directly from the laboratory software in which the specimens are registered.

Once the labeling concept for glass slides of the MS biorepository was determined, a digitization strategy was defined in collaboration with the Department of Neuropathology. Based on that, five glass slides colored with different standard staining types were selected for each specimen. Furthermore, scanning settings (such as magnification, sample detection sensitivity, size of scan area, etc.) were defined with regard to the normal-appearing, dim, and faint samples. To expedite the digitization of the glass slides, batch scanning modes and metadata entry via comma-separated values (CSV) files were used. Because the digital microscopic images are larger than any other medical images, a sustainable data storage infrastructure was needed to support them [39]. Therefore, the Net Image Server (NIS) was installed and updated to the latest version to support Simple Object Access Protocol (SOAP)-based web service technology. By estimating the storage requirements, an adequate amount of disk space to store multiple digital images on the server during the next five years was determined. A web application was published for the Brain Bank database to access, view, and annotate high-resolution digital microscopic images.

#### **Step 4. Integrating the DP module into the already existing IT-infrastructure of the KKNMS by defining the set of metadata for the scientific data collection process**

**Methods:** Once the DPS was developed, automated and manual tools for integration of the DP module into the LIMS were designed. At the beginning of the implementation phase, interoperability analysis between the DPS and the LIMS was performed to provide efficient mechanisms for interaction between the targeted systems. In addition, a communication scheme was determined between the LIMS and the DPS. In this context, an interfacing solution was developed in the LIMS to link digital images to the corresponding specimen. The SOAP-based web services were used to transfer imaging metadata (such as staining type, a thumbnail preview with the image URL<sup>1</sup>, etc.) from the DPS to the LIMS. The query structure and trigger settings were determined and considered during the development of the interfacing solution. Finally, a viewing interface for displaying the imported imaging metadata was developed and implemented in the LIMS. Users can view microscopic image thumbnails

---

<sup>1</sup> URL - A Uniform Resource Locator

for each specimen and launch a web viewer directly from the LIMS for further investigation. The viewing interface acts as a bridge between the DPS and the LIMS.

### **Step 5. Evaluating working prototypes in relation to the up-to-date literature analysis**

**Methods:** Finally, the pros and cons of tightly integrated DPS were evaluated. Additionally, the system development life cycle was tested using the W-model, which is a method used in the validation and verification of the system. Finally, an up-to-date scientific literature review was performed regarding the standards, current trends, and future perspectives of DPSs.

*Note: The literature analysis was performed two times – at the beginning to understand the existing DP state and at the end of the study in order to evaluate the developed system compared to current DP states. As such, these steps are stated in the thesis as initial and concluding literature analyses, respectively.*

In conclusion, the above five steps were significant in the successful development of an interoperable DPS that allows for the management of high resolution digital microscopic images using automated tools for the acquisition of images, and viewing and analyzing them using a fast viewer. In addition, this thesis introduces a new interfacing solution between the DPS and the pathology system that ensures automated links between the images and corresponding specimens.

## **1.4 Structure of the thesis**

This thesis is organized as follows: after introducing the problem and the main goal of the study, the related background knowledge and the current DP state are described in **Chapter 2**. This chapter also provides a literature analysis regarding the standardization and regulatory methods in DP. It ends with a basic description of the research infrastructure at UMG. This phase is required to realize and estimate all components related to similar research work.

**Chapter 3** describes the requirements analyses for developing the DPS and is divided into two parts. In the first section, the main techniques and methods used in requirements engineering are discussed. The second part of this chapter summarizes the results of the requirements analysis and represents the most important requirements of the DPS for biobank research.

**Chapter 4** turns to the main part of this thesis. The design of each functional component of the DP workflow is described together with the corresponding characteristics models.

Additionally, it introduces working steps for the successful operation of the DPS by providing an easy-to-understand overview of all its components.

**Chapter 5** is devoted to demonstrating how the interfacing solution is developed between the DPS and the laboratory management system. It provides the interoperability analysis of targeted units that leads to the right degree of integration of these systems. The architecture of the middleware solution and how the web services procedures are invoked are described schematically from different viewpoints. Additionally, this chapter provides a viewing interface implemented in the LIMS, in which users are able to view remote microscopic images. The chapter ends with the evaluation phase of the working prototype using the W-model and the concluding literature analysis in relation to the existing standards and approaches.

**Chapter 6** provides the overall conclusion of the results by summarizing the contributions of the thesis. The thesis is finalized with an outlook describing the limitations and possible refinements of the implemented DPS. In addition, it ends with recommendations for further improvement of the developed system.

## 2. Foundations

Since DP is a quite new discipline, there is a lack of scientific literature that describes detailed characteristics of WSI and standardization approaches for handling digital images from different perspectives. This chapter begins by introducing the characteristics of medical imaging informatics (**Subchapter 2.1**). Besides providing a general description of medical imaging, **Subchapter 2.2** focuses on the DP, highlighting the main challenges in this area and characterizing microscopic images in more detail. As there are regulatory and standardization barriers in DP, **Subchapter 2.3** analyzes most of the relevant and important regulations, guidelines, and standards in DP provided by different international organizations. **Subchapter 2.4** gives an introduction to the current state of the research infrastructure and describes the digital microscopic environment within the UMG. At the end of **Subchapter 2.4**, the main objectives and characteristics of the German research project, for which the developed DPS is used, are precisely described. This chapter concludes with **Subchapter 2.5**, which enumerates the examples of the work related to this thesis and describes different approaches for development and implementation of the imaging system in pathology laboratories.

### 2.1 Digital imaging informatics in medicine

Digital image processing is increasingly becoming used for numerous applications from microbiology to astronomy [40, 41]. These types of applications use the same principle of digital image processing, but with their own distinct methods and approaches for the development and operation of the imaging systems. Before the main characteristics of digital imaging are described, significant terms need to be defined.

**Non-digital image** is defined as a two-dimensional function with  $x$  and  $y$  spatial (plane) coordinates, where a  $f(x, y)$  value is the amplitude, also called the intensity of the image [42]. Value of the intensity at a specific point depends on the energy derived from physical processes.

**Digital image** is a binary representation of a picture consisting of multiple pixels (**P**icture **E**lements). More precisely, it is a mathematical function that transforms a continuous (non-digital) image into a discrete image, represented with a two-dimensional array (matrix) with multiple rows and columns (Figure 2.1).

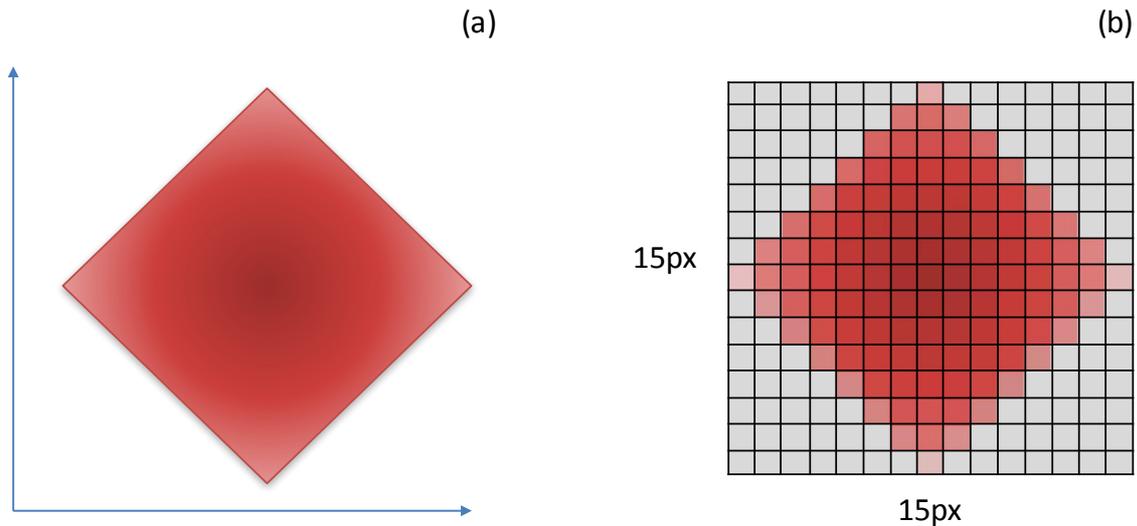


Figure 2.1. Non digital and digital images. (a) non digital (continuous) image represented with the coordinate system; (b) corresponding digital image represented with the matrix of the integer pixel values. Each small square corresponds to one pixel.

During the last decade, digital images have played an important role in various fields of medicine mostly in radiology where the clinical diagnosis and investigation are performed using digital imaging devices. According to the National Library of Medicine, **biomedical imaging** is described as “the science and the branch of medicine concerned with the development and use of imaging devices and techniques to obtain internal anatomic images and to provide biochemical and physiological analysis of tissues and organs” [43]. **Medical imaging** is a subset of biomedical imaging that uses technology to view the human body, mainly for diagnostic and therapeutic purposes. However, both terms, biomedical imaging and medical imaging, are sometimes applied interchangeably. As medical imaging is widely used in radiology, sometimes the term medical image is associated with a radiology image. However, medical images are also increasingly used in pathology, cardiology, oncology, ophthalmology, and other medical disciplines. There are different types of digital images and various techniques of image acquisition according to their application. For example, in radiology, in which the subject is a patient, a digital image is represented in real time using different technologies such as computerized tomography or MRI, while in pathology digital image is acquired from tissue on a glass slide using a virtual microscope slide scanner.

**Medical imaging informatics** is the science of information systems and technologies that provides acquisition, processing, distribution, analysis, management, and storage of digital images for medical applications. Medical imaging informatics, described as a multidisciplinary study, plays a leading role in the healthcare system, especially in medical treatment,

diagnostic, and therapeutic procedures, and in research and educational systems. Since image processing is technologically enhanced and widely adopted in medical applications, imaging informatics has become a significant part of medical informatics [40].

Medical imaging informatics involves the following components (Figure 2.2): image acquisition, representation (visualization), management and storage, processing, and interpretation of images [44].



Figure 2.2. Components of medical imaging informatics: (1) image acquisition that is process to acquire digital signal of sample using information technologies, (2) image representation also known as visualization to represent digital signal into digital image, (3) management and storage of medical images to handle, transmit, and display images, (4) image processing to operate and analyze images, and (5) image interpretation also known as computer reasoning to support clinicians with additional information generated by automated reasoning algorithms. [44].

### 2.1.1 Image acquisition

Image acquisition is the first link of the chain of imaging informatics. It forms the digital signal using special equipment and converts it into a digital format. Over the last decade, an increasing number of new technologies have arisen to support image capturing and acquisition, starting with the high resolution fluorescence digital microscopic scanner and ending with a digital/mobile camera or office scanner. Regardless of the diversity of medical image acquisition technologies, most image forming methods are based on the standard optical camera model [45].

Contemporary medical imaging equipment consists of both hardware and software components. Using integrated digital and optical cameras, the hardware is capable of forming a digital image, and the software controls and manages the acquisition device from the computer. In order to convert a continuous image into a digital (discrete) image, both sampling and quantization of pixel size should be performed [45]. Sampling is the process that converts coordinate values of a continuous image into the digital format. Quantization is a process that digitizes the value (amplitude) of the image into a discrete function using mathematical methods [42]. As a result, digitizing the coordinate values and amplitude values make up the digital image represented as a matrix of numbers.

### 2.1.2 Image representation

Once the digital image has been composed, it is represented as an image file stored in the computer memory. Digital images have the following attributes: image size, resolution, and pixel value. **Image size** is measured with the width and height (width x height) of the image that is given in pixels (e.g., 20 x 20 px). **Image resolution** designates the intensity of the pixels. The units of resolution are pixels per inch (ppi) and dots per inch (dpi), which identify the number of square pixels spaced in a single inch<sup>1</sup>. While ppi is used for digital images represented on a computer screen, dpi is mainly used for hard copies of images, such as a photograph, poster, etc. [46]. Image resolution has a significant role in diagnostic procedures, as high resolution determines how high the pixel density is, which leads to a high quality of sharpness in the digital image [44]. **Pixel value** is a quantity that defines the color (brightness) of pixel and being dependent upon the image type, which could be either binary, greyscale, or color. For binary images, also called black and white images, pixel value can only return either 0 or 1 per bit. For greyscale images, each pixel is capable of taking a value within a set of integers between [0,..,255]. Therefore, 8 bits are required for each pixel value. Color images, based on the RGB<sup>2</sup> concept, contain three channels in which each pixel is represented as a triplet of red, green, and blue components. Therefore, RGB images are mostly represented with 24 bits (3 x 8 bits) of space for a single pixel value [42].

Digital images can be enriched with metadata information. Metadata is defined as the data about the data. According to this definition, medical image metadata refers to the information about the image that can comprise a wide range of attributes, including patient information or any other keywords regarding the condition recorded in the image. Digital images with embedded metadata merge visual and textual information to improve understanding of the imaging data [47]. In addition, sophisticated metadata information is structured in that way that facilitates the searching process, improves the workflow, and accelerates retrieving of digital images [48]. Benefits of metadata are much more visible when it is used for large collection or archiving of digital images rather than for individual images [49, 50]. Metadata is likely to be stored either in conjunction with images represented via standard metadata protocols or outside the image as an XML document. As there are multiple open and proprietary file formats based on the distinct acquisition and compression methods,

---

<sup>1</sup> 1 inch is equal to 2.54 cm

<sup>2</sup> RGB - Red, Green, Blue

management of image metadata is challenging in terms of professional applications. In radiology, in which the DICOM is a de-facto standard for viewing, managing, and transferring digital images, metadata is represented in the header of the image using the DICOM metadata standard [51]. DICOM metadata provides a rich description of images using multiple attributes.

### **2.1.3 Image management and storage**

Management and storage of the imaging data is one of the key components in imaging informatics, especially when it is related to the large amount of medical images that need to be effectively handled. This stage comprises methods for compression, storage, retrieval (access), transmission, display, and archiving of imaging data [52]. This subchapter reviews the most important aspects of the image handling process. Medical imaging has an issue with file size due to the enhanced image quality of the required images. To decrease the image sizes while maintaining the quality, different compression methods are applied. Two main principles, lossy and lossless compressions, are the most prevalent for imaging applications. Based on these techniques, the image data is rewritten in a form that removes image redundancy using mathematical methods, such as Huffman encoding [42]. **Lossy** compression uses the algorithms that ensure reducing of file size at the cost of image quality. As a result, a compressed image contains some artifacts due to losing image information. Because of the high degree of compression, the lossy method is suitable for data transfer and viewing of comparatively small images. Therefore, the lossy method is not recommended when high quality images are required. Under such circumstances when the image quality is critical, the **lossless** method is more suitable, as it creates a condensed image file and keeps exactly the same quality of the original image. Using the lossless compression technique, the original file is completely recovered (decoded) from the compressed image, which occupies less space on the disk rather than original file. Lossless compression is widely used for professional applications and has been labelled the ISO standard [53]. Both compression techniques are integrated into the contemporary file formats that are described below.

Image file formats are crucial in organizing, sharing, and archiving image data [51]. In the medical context, image file formats are likely to be divided into two groups: standard image file formats and medical image file formats. Most prevalent file formats applied in medical applications, mostly in radiology and pathology, are as follows:

## 1. Standard file formats:

- **TIFF** (Tagged Image File Format) - keeps the identical quality of the original image at the cost of large file size. TIFF is ideal for most medical images where high quality is required. The TIFF format is based on the lossless compression techniques that ensure a wide range of pixel depth and image size. TIFF files contain more than 70 public and private tags conveying information about the imaging data [51]. A number of alterations to the TIFF file format are being performed due to the growing demand for professional applications. For example, the Big-TIFF file format was created to store files more than 4GB in size, which is the maximum size of a TIFF file.
- **JPEG** (Joint Photographic Experts Group) – is the most prevalent file format for standard image size [51, 54]. By using the lossy compression method, file size and image quality are significantly decreased. However, it is a convenient file format for fast and easy image transfers [55]. Also, there is a lossless compression method defined for JPEG files, but it is not widely adopted in image acquisition software.
- **JPEG2000** - an alternative JPEG file format (file extension is .JP2) based on a sophisticated compression method that provides improved image quality with less artifacts [56]. JPEG2000 file format can be represented with both lossy or lossless compression methods.
- **PNG** (Portable Networks Graphics) – widely used to store and transfer images mostly in web-browsers and for transparent image manipulation using graphic software.
- **GIF** (Graphics Interchange Format) – file format based on the lossless compression method developed for storing and transferring image data. Currently, the GIF format is mostly used for its unique characteristics capable of small image animation using image frames [57].

## 2. Medical image file formats:

- **DICOM** (Digital Imaging and Communications in Medicine) – file format published by the American College of Radiology that became a standard file format and communication protocol for storing and transferring of radiology images, such as ultrasound, MRI, etc. File extensions can be .dcm or .dcm30. DICOM images contain imaging data and patient information packed into a single file. It is constructed of two main structural components: a header element that describes the file and the study for which the image is used, and a data set that represents an instance of the information object. The data

set consists of multiple data elements that describe the attributes of the object using encoded values [58]. DICOM supports a number of compression methods, such as the lossless JPEG Image Compression, lossless Run Length Encoding Compression (also known as RLE), JPEG-LS (lossless) Image Compression through the encapsulated format, lossy and lossless JPEG 2000 Image Compression, and others [59]. The maximum size of a DICOM file is 4GB as a result of 32-bit offsets.

- **Nifti** (Neuroimaging Informatics Technology Initiative) – developed for multidimensional functional neuroimaging to facilitate image analysis using information technology [60]. It is mostly applied to visualization and analysis of functional magnetic resonance imaging data and is widely used by the neuroimaging research community. The Nifti format specification consists of voxel data and a standard image header containing information about the image metadata. The header block, with 348 bytes in fixed length, is likely to be extended to keep additional metadata [61].
- **OME-TIFF** (The Open Microscopy Environment) – based on the standard TIFF format and is a widely accepted standard for biological microscopy data. OME-TIFF is mostly focused on fluorescence microscopic images [62]. In addition to the TIFF format, it contains an OME-XML metadata block to enrich microscopic images with a large amount of metadata, such as annotations, experimental results, and specific information representative of biological microscopy data [63]. Like the TIFF file format, an OME-TIFF file is restricted to 4GB file size.
- **Proprietary file format** (non-standard format) – designed by an organization or a company that provides image acquisition hardware and/or software components commercially available on the market. Organizations use their own acquisition scheme and compression methods that are coded and unknown to customers. As there are multiple image acquisition devices on the market, image acquisition and compression methods vary from organization to organization. Companies that provide proprietary file formats additionally have to provide converting tools to transform images with this proprietary file format into an open file format such as TIFF, JPEG2000, and others. However, during the conversion of digital images, imaging data or embedded metadata can be partially lost or modified [49].

Existence of the above-mentioned file formats indicates that there is growing demand of usage for medical images with associated image information. In order to choose the

appropriate image format, it is important to efficiently define the needs and main requirements of specific applications.

Storage and retrieval of digital images are crucial components of imaging informatics. In radiology, the commercial Picture Archiving and Communication System (PACS) is already widely used for storage, transmission, access, and archiving of digital images. It uses standard compression methods for transmission and display [64]. As DICOM is a standard in radiology, PACS systems mostly use the DICOM service to transfer data, and images are represented with open file formats such as TIFF, JPEG, JPEG2000, or DICOM itself [65]. Digital images that are stored as proprietary file formats can not be handled by PACS. Therefore, suitable storage solutions are required either to be provided by scanner manufacturers or to be developed by healthcare organization itself.

#### **2.1.4 Image processing and interpretation**

Digital image processing is a set of methods used to manipulate and analyze digital images through mathematical operations. The main purpose of medical image processing is to enhance digital images for human visualization, and to segment and analyze images by which the specific organs or tissues can be characterized. Automated image analysis algorithms are mostly applied for radiology images, but in the last decade an increasing amount of image analysis applications have arisen in pathology that can also extend the functionality of the DP workflow [66, 67]. For example, the FIJI open sourced platform is designed for microscopy image processing and used by biology research communities [68]. Most widely used methods of image processing in pathology are image segmentation and quantitative analysis to assess the histological samples. Image segmentation methods offer the possibility to detect cells through edge detection [69, 70]. Once the cells are detected, a quantitative estimation enables the pathologist to assess the image of a specific part of the desired organ through numerical measurement of shapes, sizes, and quantity of cells or subcellular components [71–73]. Furthermore, image transforming algorithms such as Fourier or wavelet transform methods are being used to enhance the appearance of an image or a selected region of the image by changing the color settings, including contrast, brightness, sharpening, or other features of digital image visualization [74].

Image interpretation is the final link of the chain of the medical imaging informatics and represents the process wherein the physician investigates the image to identify abnormalities

or main characteristics of the histological specimen. To improve the process of image interpretation, automated reasoning algorithms are developed that can detect the abnormality on the image or moreover, characterize the abnormality and provide possible diagnoses about the specimen. Using image processing algorithms like quantitative methods, diagnostic accuracy can be improved by searching and retrieving similar-appearing abnormalities (called content-based image retrieval) from already characterized images stored in the database. A physician may have to look at the results and make a diagnosis with more confidence [44]. Furthermore, professional opinion can be provided by computer assisted detection that finds the abnormality on the image, or computer assisted diagnosis (CAD) that offers suggestions to pathologists for the suspected diagnosis.

## 2.2 Digital pathology

DP provides novel ways for pathologists to view, store, transmit, manage, process, and analyze tissue samples. As microscopic images are differed from other medical images, better understanding of the main characteristics of the DP is an essential part for the successful development of the appropriate imaging system in this field.

According to the scientific publications on the PubMed, there has been a growing interest in DP among researchers over the last decade (Figure 2.3). Popularity of this field is following the progression of technological advancements.

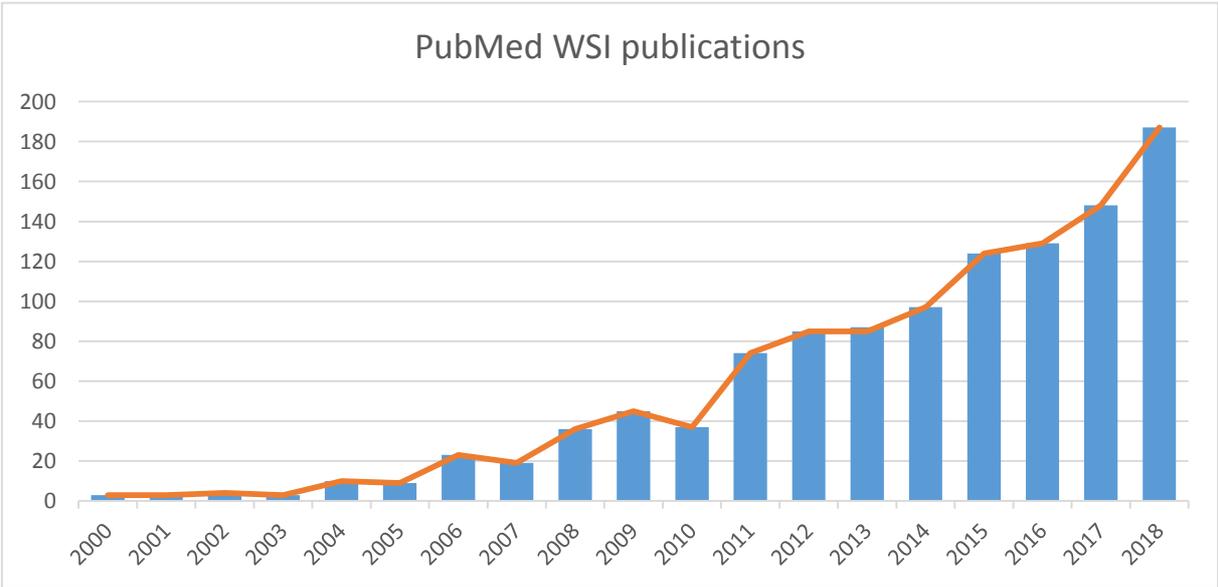


Figure 2.3. Publication in whole slide imaging in PubMed between 2000–2016. Searching keywords: whole slide imaging, digital pathology, or virtual microscopy in the title/abstract.

### 2.2.1 Workflow of digital pathology

Generally, a pathologist spends several hours per day viewing glass slides and making diagnoses using conventional microscope. DP will likely dramatically change the working process of pathologists and will provide additional tools that enhance the pathology workflow. Once a glass slide is digitized, the pathologists can easily retrieve, view, and analyze the image of the entire slide on a computer monitor.

At the beginning of the traditional pathology workflow, pathologists cut the organ or part of the organ into small pieces called tissue blocks<sup>1</sup>. Each tissue sample is sliced in very thin layers, called tissue slides, which are mounted on the glass slides. Afterwards, the glass slide with the attached tissue sample is processed and stained with various histopathological staining types. Finally, the glass slide is viewed by the pathologist under the microscope.

There is one extra step in the DP workflow in comparison with traditional methods. Glass slides are digitized using a high-resolution microscope slide scanner, and the digital images are likely to be displayed on the computer screen. Both workflows, traditional and DP methods, are illustrated in Figure 2.4.

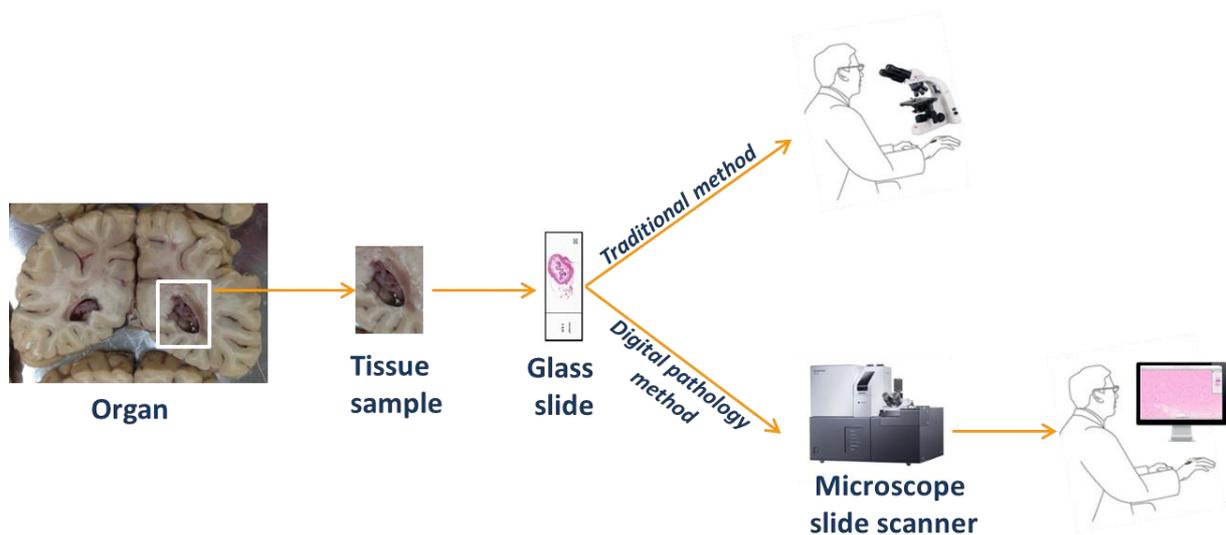


Figure 2.4. Traditional and digital pathology workflows. (1) Traditional method: once the tissue is mounted on the glass slide, a pathologist can inspect it under the microscope; (2) DP method: first the glass slide is digitized using a microscope slide scanner. Afterwards, a pathologist views the high-resolution image of the slide on the monitor.

<sup>1</sup> Sometimes the term “tissue block” is mentioned as “tissue sample” or “specimen”.

### **2.2.2 Benefits and limitations of digital pathology**

According to Sucaet's book on DP, there are two types of pathologists: "those who feel that DP is the future and are willing to start working with it today; and those who feel that DP is the future, but it will always remain so" [75].

Some pathologists see many advantages compared to conventional microscopy, including improvement in daily work. They believe that the implementation of the technology has great promise for the future of pathological practice. It would improve laboratory workflow, enhance QA programs, and facilitate communication between physicians. Pathologists could view slides anytime and anywhere. This capability would open up opportunities for collaborative work, research, and teaching. Also, pathologists could use automated image analysis and annotation tools.

Other pathologists are more skeptical as long as DP is not fully integrated into the routine diagnostic, research, and educational practice. Some pathologists think that in rare cases it is possible to lose particles of the image during the scanning process, which may cause a misdiagnosis if a pathologist views and makes a diagnosis only using digital images [75]. Additionally, pathologists need extra time to digitize microscope slides using special scanners and store them on an image server or external hard disk. Most pathologists find it more difficult to make a diagnosis on virtual slides rather than with traditional glass slides. Furthermore, proficient training is important for pathologists to become aware of proper scanner usage. Digital scanners are also much more expensive than conventional microscopes and require a high throughput technology. Last but not least, seamless integration of well characterized microscope images into other medical IT-systems for research and primary diagnosis is still challenging despite the fact that this link is essential. Furthermore, since there are multiple scanners vendors, standardization methods are challenging.

### 2.2.3 Applications of digital pathology

Digital microscopic images are increasingly being used for a number of applications (Figure 2.5) [75]:

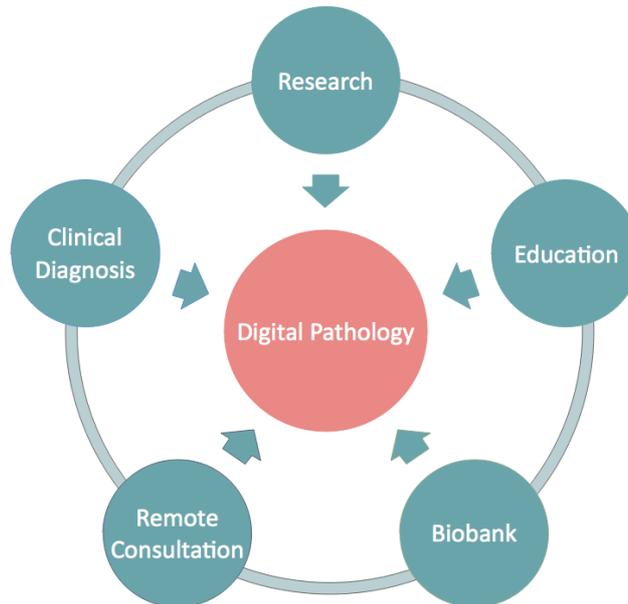


Figure 2.5. Applications of digital pathology from basic research to diagnostic purposes.

- **Research:** Whole slide images are likely to be routinely used for image analysis in research study. Increasingly, cell image analysis algorithms are being developed for digital microscopic images, such as segmentation and cell detection, image restoration, etc. In addition, digital representation of glass slides facilitates collaboration between distinct researchers.
- **Education system:** Medical students can easily access microscopic images from their own computer without the conventional microscope. Tests, quizzes, and exams are likely to be performed via digital systems at the university. A couple of companies on the market provide virtual slide educational management systems for creating and viewing course materials and also for managing trainees [76].
- **Biobank:** Within the context of biobanking, biological materials, including tissue samples, are stored for further investigation. Tissue banking is a collection of specimens obtained from a specific organ and which are focused on a specific disease. Using DP tools, glass slides of biorepositories can be digitized and stored in a digital format for relatively long time. In addition, digital images of histopathology samples

are extremely beneficial for researchers who need to view the digital microscopic images from a remote location.

- **Remote consultation and second opinions:** Instead of sending physical glass slides to remote medical centers, URLs of microscopic images can be sent for consultation. Therefore, a distant pathologist can view digital images on the computer screen. This solution saves money and time required when sending physical glass slides.
- **Clinical diagnosis:** Digital microscopic images are not routinely used in clinical diagnosis. It seems that validation and standardization of DP workflow is the most difficult part for the adoption of DP in clinical diagnosis. CAD<sup>1</sup> for automated detection and characteristics of lesions has potential in pathology, as it might be applied as a second opinion while making the diagnosis [77]. From a technical perspective, as dozens of glass slides are applied in everyday practice in a laboratory, a high throughput system and many terabytes (even petabytes) of disk space are required to manage all the digital images that will be made with DP. Therefore, several years are needed for full integration of DP tools in a routine diagnostic practice.

Most of the benefits of DP are achieved when whole slide images are applied for different purposes in the same institution, including research, education system, remote consultation, and clinical diagnosis.

## **2.3 Initial literature analysis in digital pathology**

At the beginning of the PhD research study, a scientific literature review was carried out within the publicly available and relevant research papers, regulations, and official guidelines focused on DP. The majority of those scientific papers mostly published in the last decade describe the current state, limitations, needs, and opportunities of DPS for various applications.

### **2.3.1 Historical review and current status of digital pathology**

The terms digital pathology and virtual microscopy were used in the scientific papers starting in 1997 [78–80]. However, before that, telepathology, first defined as “the work of a pathologist at a distance” [80], had a leading role in numerous pathology laboratories. Telepathology, an ancestor of DP, provides telecommunications between remote places by transferring of the microscope images or video data using technology [81]. In 1968, Weinstein

---

<sup>1</sup> CAD - Computer Aided Diagnosis

performed the first telecommunication via video between the Massachusetts General Hospital in Boston and the Logan Airport Medical Station to view surgical pathology cases [82]. However, a specific period of time was needed to enhance this technique parallel to technological advances. From the 1990's personal computers have been developed with opportunities to view images using integrated image boards and graphic interfaces [82]. This opportunity resulted in the quick utilization of telepathology techniques in more pathology laboratories. Since then, telepathology has been mostly used for remote consultations, consensus diagnosis, video conferencing, education, and research purposes. Telepathology could be represented in either static or dynamic form. Using a static telepathology, microscopic images are captured and sent to a remote place. A dynamic telepathology provides the robotic system, where a pathologist can operate and control microscope features from a remote place, such as a microscope stage, objective lens, magnification, and other motorized details [83]. The main limitation of both types of telepathology is the quality of the microscopic images [82]. In 2000, the first microscope slide scanner arose on the market with support for digitization and acquisition of high resolution images using a computer. This development was the start of a new era in pathology.

During the last decade, DP has comprised a number of subdomains, including data transfer, image analysis, etc. Currently, the term telepathology is rarely used and is mostly substituted with digital pathology.

### **2.3.2 Regulations, guidelines, and standards in digital pathology**

By investigating the widely known guidelines and regulations in DP, the main rules, recommendations, and suggestions of multiple international organizations were assessed. As DP workflow varies from laboratory to laboratory, implantation of the recommendations of guidelines is a complicated process [1].

The following scientific materials provided by different organizations were analyzed: The FDA<sup>1</sup> published "Technical Performance Assessment of Digital Pathology Whole Slide Imaging Devices" in 2015, in which it provides a regulatory evaluation of WSI components for the manufacturers of scanning systems [3]. Widely distributed guidelines for the validation of WSI were also produced by the College of American Pathologists (CAP) in a document titled "Validating Whole Slide Imaging for Diagnostic Purposes in Pathology" and by the DPA in

---

<sup>1</sup> FDA - Food and Drug Administration

documents titled “Validation of Digital Pathology in a Healthcare Environment” and “Validation of Digital Pathology Systems in the Regulated Nonclinical Environment” [9, 84]. The latter was endorsed by the Scientific and Regulatory Policy Committee and the Executive Committee of the Society of Toxicologic Pathology [10]. Additional guidelines mainly focused on the implementation of telepathology services for clinical applications are as follows: “Clinical Guidelines for Telepathology” provided by the American Telemedicine Association (ATA), “Guidelines from the Canadian Association of Pathologists for establishing a telepathology service for anatomic pathology using whole-slide imaging” established by the Canadian Association of Pathologists, and “Telepathology: Guidance from The Royal College of Pathologists” [11, 12, 85]. Furthermore, major efforts for validation of WSI and telepathology systems are being made by the Telepathology Network in Europe (EURO-TELEPATH) and various other pathology laboratories at the local and county level in Sweden, the Netherlands, and Germany [7, 86–88]. Regardless of the regulations and guidelines, the DICOM standard was also evaluated for pathology use cases [59].

#### 2.3.2.1 Assessment of regulatory standards

Anatomic pathology laboratories in the United States are accredited and regulated by regulatory bodies. Validation of WSI systems for primary diagnosis is required for certified laboratories that should be in full compliance with the regulation of the following governing agencies: Clinical Laboratory Improvement Amendments, The Joint commission, CAP accreditations, and state regulations [84]. Regulatory justification demonstrates safety and effectiveness of clinical and digital microscopic imaging data.

The FDA does not control the WSI systems at the laboratory level, but does provide regulations for the manufacturers of the WSI devices. All WSI devices used for clinical or nonclinical studies should be produced in compliance with the Code of Federal Regulations (CFR) Title 21 Part 58 (Good Laboratory Practice for Nonclinical Laboratory Studies) and Part 11 (Electronic Records; Electronic Signature). According to the regulations 21 CFR 58.3(k) and 21 CFR 58.63, an image is defined as raw data that is the result of original observations and activities of a nonclinical laboratory study [89]. Raw data also may refer to photographs or exact copy generated by automated instruments, including a digital representation of the microscopic image. Equipment used for the generation, measurement, or assessment of data should be adequately tested, calibrated, and standardized [90]. Generally, the FDA classifies medical devices into three groups based on their risks [2]. According to the classification, conventional

microscopes are included in a Class I as they are subject to the low risk, while WSI devices are classified under Class III. Therefore, the latter type of devices requires further assurance prior to marketing that they are both safe and effective for clinical diagnosis [91].

In the beginning of 2016, the DPA suggested that WSI manufacturers submit de novo premarket applications to the FDA for primary diagnosis [5]. While de novo applications receive marketing authorization by the FDA, WSI devices would shift accordingly into Class II. Afterwards, they still need a decisive step forward to show reasonable assurance of safety and effectiveness, and they would need to become a Class I medical device for primary diagnosis.

The European WSI device market is regulated by the Rules Governing Medicinal Products in the European Union and is subject to good manufacturing practice guidelines [92]. Medical devices should be approved by CE Mark for primary diagnosis in Europe. Over the last few years, several DP devices have received the CE Mark for diagnostic usage in Europe, for example the Aperio AT2 scanner for in vitro diagnosis. In addition, more and more vendors are willing to submit their product to receive the CE Mark that gives the possibility to the customers to use the devices for clinical diagnosis [93–95].

EURO-TELEPATH was run between 2007-2011 and was financed by the European Cooperation in Science and Technology, named the COST Action IC0604. The objective of the EURO-TELEPATH was to estimate and validate technological and communication standards between European collaborative research groups. Sixteen European countries have participated in EURO-TELEPATH and are distributed into four working groups: (1) Business modeling in pathology; (2) informatics standards in pathology; (3) image analysis, processing, retrieval, and management; and (4) technology and automation in pathology [86]. The core of this collaboration was dedicated for the standardization methods in pathology working with IHE, DICOM, HL7, and other standardization bodies [96].

#### 2.3.2.2 Main features and recommendations of DP guidelines

The guideline produced by the CAP Pathology and Laboratory Quality Center provides the most important requirements for validation of WSI for diagnostic purposes. The CAP guideline includes 12 statements and an accompanying grading system. They were formulated and analyzed by the expert panel consensus, a published literature review and open comment feedback. The strength of the recommendations is indicated by four categories: Grade A or B, termed as Recommendations, are dedicated to the statement that is derived from strong

evidence and can be used in all or most cases; Grade C, designated as Suggestion, matches uncertain situations; and finally, Grade D is based on weak evidence such as a personal opinions and signed as an Expert Consensus Opinion [9]. Table 2.1 lists recommendations of CAP guidelines with accompanying grades. The CAP provides significant statements and recommendations for validation, but the guideline lacks the step-by-step procedures for implementation. Till today, there is a lack of published studies that express the complete validation process of a specific laboratory prior to clinical use.

Based on the CAP guideline, all pathology laboratories should carry out their own validation studies when implementing the WSI system by considering its clinical use and settings. At least 60 cases for each application (e.g., Hematoxylin and eosin, frozen sections, etc.) are required to be included into the validation process. Digital and glass slides should be examined separately in both random and non-random order. They must then be compared with each other. Each laboratory should define the main scope of their validation project monitored by a team that consists of people from different backgrounds, such as pathologists, researchers, manufacturers, information technologists, programmers, and others. Defining responsibilities and roles of the team members is important for successful validation [84].

Other important features of the multiple guidelines provided by different organizations are compared and listed in Appendix A.

Regardless of the fact that the guideline is intended for diagnostic purposes, most of the recommendations should be taken into consideration (modified in response to the needs) during the development of the digital image system for the biobank research or other complex environments as well. Requirements for the successful implementation of the DPS into the collaborative research network shall be determined in compliance with the current recommendations provided by international organizations. In such case, if some of those recommendations are successfully implemented for research purposes within the UMG, then there is more potential that the same approach can be adapted according to needs and will work not only for other research projects related to DP, but also for clinical practice in the laboratory of the same medical center.

Thus, validation of WSI is a crucial component in the adoption of DPS. It is evident that the widely distributed and comprehensive guidelines could significantly assist in validation process of WSI systems. Moreover, comprehensive guidelines would seem to encourage manufacturers of scanners and regulatory bodies to make their standards more flexible.

Table 2.1. Statements of the guideline for the validating whole slide imaging for diagnostic purposes in pathology [9]

CAP Guideline Statements		
1.	All pathology laboratories implementing WSI technology for clinical diagnostic purposes should carry out their own validation studies.	Expert consensus opinion – Grade D
2.	Validation should be appropriate for and applicable to the intended clinical use and clinical setting of the application in which WSI will be employed. Validation of WSI systems should involve specimen preparation types relevant to intended use (e.g., formalin-fixed paraffin-embedded tissue, frozen tissue, immunohistochemical stains, cytology slides, hematology blood smears).	Recommendation – Grade A
3.	The validation study should closely emulate the real-world clinical environment in which the technology will be used.	Recommendation – Grade A
4.	The validation study should encompass the entire WSI system. <i>Note: It is not necessary to separately validate each individual component (e.g., computer hardware, monitor, network, scanner) of the system nor the individual steps of the digital imaging process.</i>	Recommendation – Grade B
5.	Revalidation is required whenever a significant change is made to any component of the WSI system.	Expert consensus opinion – Grade D
6.	A pathologist(s) adequately trained to use the WSI system must be involved in the validation process.	Recommendation – Grade B
7.	The validation process should include a sample set of at least 60 cases for one application (e.g., Hematoxylin and eosin stained sections of fixed tissue, frozen sections, cytology, hematology) that reflects the spectrum and complexity of specimen types and diagnoses likely to be encountered during routine practice.	Recommendation – Grade A
8.	The validation study should establish diagnostic concordance between digital and glass slides for the same observer (i.e., intraobserver variability).	Suggestion – Grade C
9.	Digital and glass slides can be evaluated in random or nonrandom order (as to which is examined first and second) during the validation process.	Recommendation – Grade A
10.	A washout period of at least two weeks should occur between viewing digital and glass slides.	Recommendation – Grade B
11.	The validation process should confirm that all of the material present on a glass slide to be scanned is included in the digital image.	Expert consensus opinion – Grade D
12.	Documentation should be maintained recording the method, measurements, and final approval of validation for the WSI system to be used in the clinical laboratory.	Expert consensus opinion – Grade D

### 2.3.2.3 The role of DICOM, IHE, and HL7 in pathology

Radiology regulations serve as good examples for clarifying several concerns for WSI and give guidance for better understanding of safety and effectiveness of WSI [97–99]. Additionally, many years of experience in radiology can serve as a successful road for the effective adoption of DP in clinical practice. The DICOM became the de-facto standard and is widely adopted in radiology for handling and transferring medical images. According to supplement 145 provided by the DICOM Working Group 26, the DICOM standard can be also used for the pathology workflow [59]. However, DP faces specific challenges to DICOM, such as: (1) image size is mostly limited to 4GB due to the underlying 32-bit architecture, while digital microscope images are much larger in comparison to radiology images [100]; (2) Even though DICOM suggested an efficient pyramid performance of image data and sub-region access for image

display, rapid movement and zoom control of digital images throughout the navigation still remain the main technical difficulties; (3) A hierarchy of pathology images derived from the specimen differs from the radiology images where the subject is a patient. The DICOM standard supports patient information with accompanying elements, such as study, series, and image (without specimen information); whereas, pathology data is represented with patient, case, specimen, and glass slide (hierarchical structure of pathology data is detailed described in Subchapter 4.1.3); (4) Furthermore, most DICOM services demand more resources to transfer pathologic image data over a network based on an entire image transfer protocol. In spite of these limitations, DICOM has a potential to become the imaging data handling standard for most medical disciplines, including pathology.

The anatomic pathology working group launched by IHE published technical frameworks for the implementation of information systems in pathology laboratories [17, 18]. The IHE provides the recommendations and defines the integrated pathology workflow by using the existing standards, such as DICOM and HL7. In parallel, the HL7, which provides set of standards for transfer of electronic health information between different applications, published the anatomic pathology working group to support pathology laboratories with enhanced HL7 standards [101]. Therefore, since 2005, several collaborative meetings of the working groups of these standards along with the vendors of scanners have been conducted every year to enhance the workflow, define challenges of the standards, and find effective solutions. The main objective of the IHE is to implement the technical framework for clinical diagnostics, but it can be also used for research or biobank purposes [102]. The IHE also tries to encourage the vendors of scanners and pathology information systems to start adopting their standard. However, as there are several limitations of DICOM standard for pathology, collaborative working groups are trying to improve the DICOM standards according to needs of pathology laboratories.

In the next 8 to 10 years, major changes in the automated processes in anatomical pathology and widespread adoption of WSIs are expected. Currently, most of the digital scanner vendors have their own file formats and provide proprietary web viewer and server infrastructure for their customers. However, simultaneous usage of scanners provided by different vendors requires standardization of these systems.

The estimation and consideration of the existing regulatory barriers, recommendations, standards and guideline statements are the initial steps in the research work to lead it to the successful implementation of DP tools for the specific application.

## **2.4 Research environment of the application**

The Department of Medical Informatics (MI) of the UMG provides methods and IT tools in medical care and biomedical research. In order to deal with important tasks at the department, there are several working groups, including CIOffice, that focus on the development and evaluation of IT solutions for the clinical or research environment [103]. In addition, research groups at the Department of MI analyze the information platform requirements for the competence networks in medicine and provide the IT infrastructure according to requests. Because of the interdisciplinary characteristics, the Department of MI has close collaboration with other departments of the University of Göttingen in clinical, research, and educational contexts. Additionally, the MI team works with external medical organizations and institutions, like the TMF<sup>1</sup> umbrella organization that provides basic concepts for IT solutions according to the ethical and technological problems of the networked medical research environment in Germany. This kind of large-scale collaboration provides a sustainable research environment at the local, national, and international levels.

### **2.4.1 Common architecture framework for biobank research networks in Germany**

The German Federal Ministry of Education and Research initiated TMF to improve architectural frameworks and infrastructural environments for networked medical researches in Germany. From the biobank point of view, the main topics of TMF are as follows [104, 105]:

- Data protection framework that is based on the three aspects [104]: (1) separation of distinct information, such as patient data, samples, or image data; (2) pseudonymization of patient identities using two-step ID encryption; and (3) identity management to protect the identity of the sample or sample-derived assets, such as digital images.
- Legal and ethical aspects for biobanks and for conducting medical studies.

---

<sup>1</sup> TMF - Technology, Methods, and Infrastructure for Networked Medical Research

- Supporting IT infrastructure to develop complex solutions for networked medical researches.
- Quality management in biobanking by providing the QA checklist to ensure the quality of data collections [106].

According to the TMF IT review board annual reports, there is no need to develop individual applications, especially for the complex research environments, as there are commercially available products that can be adapted according to their needs [31]. However, in that case, there is a need for complex tools to implement new modules and extend systems with regard to the application. Additionally, as several products can be used within individual research, proper interfacing solutions are required to connect different components of a whole system. With regard to medical image data, the report highlights the need for further development of imaging systems in terms of processing, management, and storage of imaging data with accompanying metadata. Additionally, the tight linkage between images and specimen and the need for remote access to the images are key factors for networked medical research [107]. Instead of programming their own image management solution, TMF suggests healthcare organizations deploy existing systems and customize them to provide high quality medical research.

#### **2.4.2 Research infrastructure at UMG**

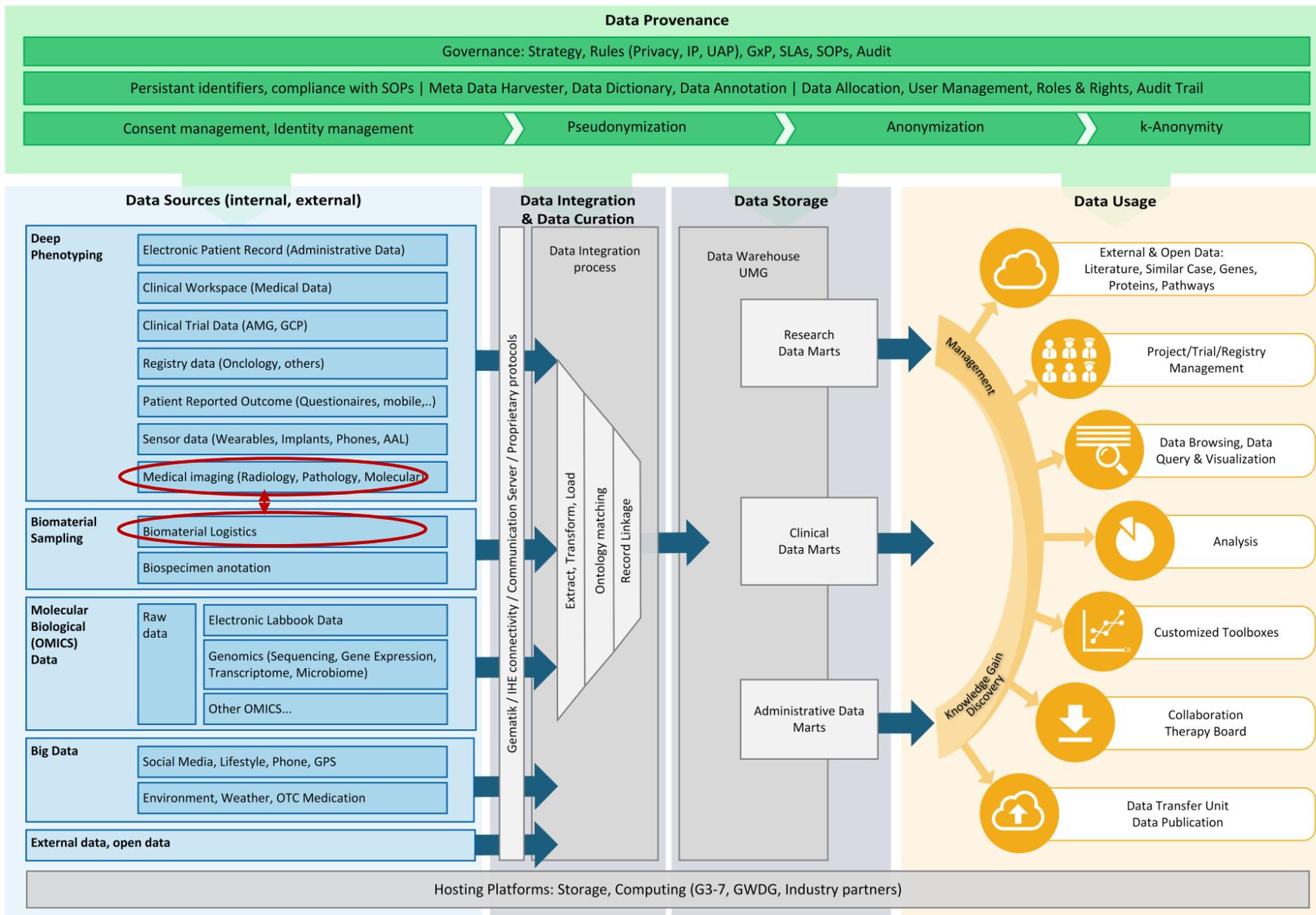
Like many other medical centers, the UMG possesses a huge amount of data derived from multiple data sources, including electronic patient records, medical images, biomaterial data, genotyping data, etc. Modern methods and tools are required to handle heterogeneous medical data using information technologies.

IT infrastructure for clinical research projects within the UMG are based on the “Data-Rules-Tools” method [108]. This approach consists of combinations of three main components: (1) heterogeneous medical data generated from different sources, (2) rules for the sufficient usage of data, and (3) tools to operate data for different applications.

The Department of MI provided the complex research architecture of the research platform of the UMG [108]. The MI research platform, illustrated in Figure 2.6, comprises three main pillars: (1) data sources depicted in the blue pillar, (2) data integration and storage highlighted with grey color, and (3) applications illustrated in yellow. All these pillars are covered by the rules for data provenance highlighted in green.

So far, radiology images, handled by the Chili system, are widely used by research networks at UMG, while the management of pathology images still remains a challenge. As the IT infrastructure of the UMG, the data, rules, and tools used for the DPS should be in compliance with the general principle of the research platform.

The data sources highlighted in blue in Figure 2.6 contain different types of data, including phenotypic, genomic, and biomaterial data. From a DP point of view, two main data types, medical imaging and biomaterial samples (signed with the red circles), are likely to be used in pathology laboratories for research purposes. However, they are located in the different data blocks, meaning those types of data are handled by separate information systems and might be stored within the distinct hosting systems. Methods and tools for close linkage of both types of data and communication mechanisms between the different information systems will be discussed in detail in the following chapters, in accordance with the MI research platform.



© 2016 UMG: Bauer, Buckow, Franke, Gusky, Quade, Rey, Umbach, Sax 04/01/2016 V3

Some icons created by flaticon.com under CC BY (<http://creativecommons.org/licenses/by/3.0>). The authors are: Freepik (<http://www.freepik.com>) and Simpleicon (<http://www.simpleicon.com>).

Figure 2.6. Research infrastructure at UMG [108]. The MI research platform is comprised of three pillars: (1) Data sources depicted in the blue pillar, (2) Data integration and storage in grey, and (3) applications illustrated in yellow. All these pillars are covered by the rules for data provenance highlighted in green. Two data types are signed with the red circles, namely medical imaging and biomaterial samples, that are used in the pathology laboratory derived from tissue samples.

### 2.4.3 Digital pathology system architecture at UMG

DP tools have been used at UMG since 2008. During that time, the initial version of the Olympus microscope slide scanner (dotSlide scanner) and the image server were set up at UMG for pathology educational purposes. Medical students could access the microscopic images during exams or quizzes in pathology using computers located in the auditorium. Additionally, DP tools were being used for image analysis and small-scale research purposes by individual researchers from various UMG departments. The state of DPS infrastructure in 2013 is illustrated in Figure 2.7. Despite the use of scanners for various applications, a very limited number of features and tools were used in a practical manner. Additionally, the scanner, along with the accompanying IT infrastructure, was not operating at full capacity. Since 2013, DP tools have been used within the scope of a complex biobank research network. Therefore, more advanced tools and automated procedures for the digitization and storage of images are required for effective pathology workflow. At the beginning of 2015, a contemporary digital microscope slide scanner – the Olympus VS120 – was installed and an image server was set up and updated to the latest version at UMG in order to satisfy research needs. Therefore, the IT infrastructure of the DPS was significantly modified. The new DP workflow currently used for the biobank research network and the accompanying IT architecture of the DPS are described in the following chapters (Chapters 4 and 5).

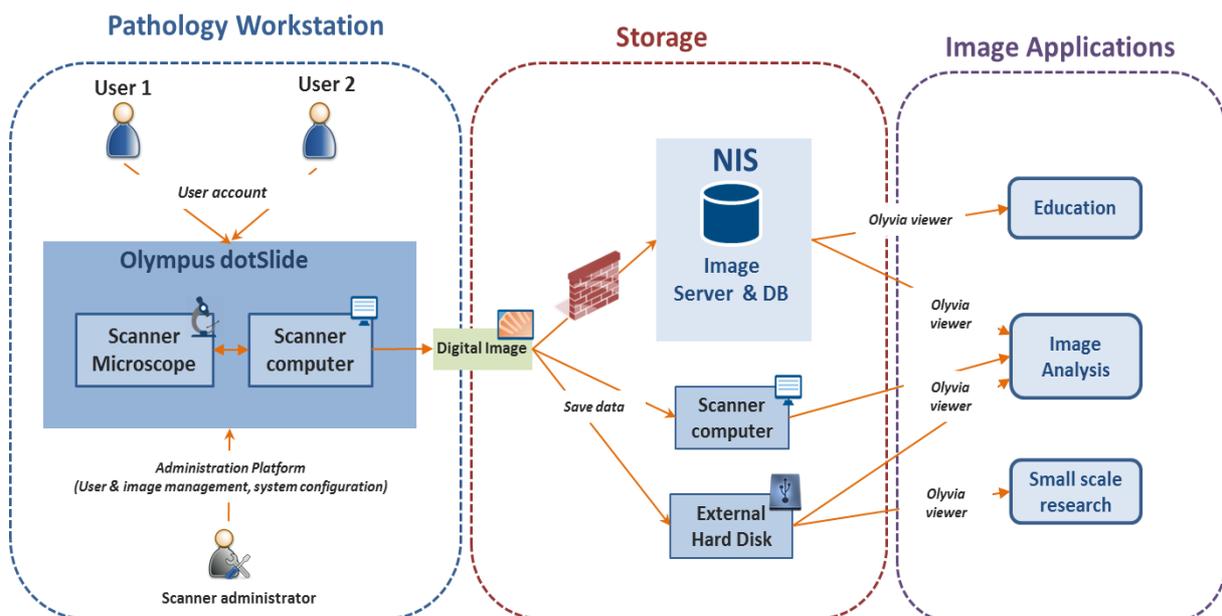


Figure 2.7. The initial state of digital pathology infrastructure at UMG through 2013. The DPS architecture contains three main parts: (1) A pathology workstation that includes a microscope slide scanner and acquisition software, (2) storage infrastructure, and (3) image applications. Digital images are stored either on the image server, scanner computer, or external hard disk. At that time, images were applied for educational purposes, image analysis, and small-scale research purposes.

In the future, DP is likely to be simultaneously used for a number of applications within the UMG. Moreover, it can later be successfully applied for use in clinical diagnosis. However, the implementation of DP for clinical diagnostic purposes is still a challenge, as a WSI system requires additional regulatory and standardization adjustments.

#### **2.4.4 German Competence Network Multiple Sclerosis research consortia**

The Department of MI is one of the main participants of the KKNMS German research project that establishes IT infrastructure for neuroscientists [109]. Multiple Sclerosis (MS) is a disease of the central nervous system that damages the flow of information between the brain and body. Currently, the cause of MS is unknown. Around 2,500 German citizens are diagnosed with MS each year. In total, the number of registered patients suffering from MS in Germany is estimated to be between 100,000 and 140,000 and the average age is 35 [110]. The disease is most frequent in women with 2.5:1 sex ratio [111]. The etiology and heterogeneity of the disease is still not completely understood. The main objective of the KKNMS is to change this and to improve diagnostic and therapeutic possibilities.

There are several working groups within the KKNMS to deal with important research topics, including the Multiple Sclerosis Brain Bank (MS-BB) unit. The MS-BB subproject aims to collect, store and characterize autopsy-derived tissue samples from patients registered as brain donors. From the perspective of medical informatics, the main objective of the MS-BB is to develop an interoperable microscopic image handling system that is efficiently integrated into the existing IT infrastructure used by the researchers of the KKNMS. Therefore, the CIOffice Research Network is responsible for determining, implementing and integrating a viable solution for the archiving and distribution of the MS Brain Bank's high-resolution scans of histopathological samples into the KKNMS IT infrastructure. The ultimate objective of the MS-BB is to increase knowledge of the etiology and pathogenesis of MS. Using the DP tools, researchers can be supplied with the high-resolution scans of the histopathological samples of brain tissue. Thus, MS-BB serves as a use case for the application of the DPS in a collaborative clinical research project. Digital representation of complete images can be provided to the researchers to assist them in obtaining a digital view of the sample for further investigation prior to shipping. This digitalization process is expected to optimize the exploitation of the Brain Bank's tissues. Access to sample images will not be limited to one researcher at a time. It may even be possible to avoid the physical delivery of a specimen given the quality and informative value of the image.

Because this PhD thesis is based on the KKNMS research project, it provides a description of the data, methods and tools for the development and integration of the complex DPS into the laboratory workflow used for the biobank application.

## **2.5 Related work**

Even though the DP tools are being used for an increasing number of applications, there are very few laboratories in the world that completely changed their workflow using the DP tools. In this subchapter, we discuss examples of the work that are related to the thesis and describe the DP workflows with the accompanying implementation approaches for pathology laboratories.

In 2010, the DPA published a white paper that provides the interoperability processes between Anatomical Pathology Laboratory Information Systems and DPSs [30]. Their initiative was to describe a detailed overview of the current state of the interoperability of those systems and define the future work that may improve the pathology workflow. According to the DPA, there are different degrees of integration between LIMS and DPS that vary from tightly coupled to loosely coupled integration techniques. Metadata sharing between different systems might be performed using standard communication protocols or interfaces.

Thorstenson et al. bring their experience and describe the implementation of a large-scale WSI system for routine diagnostics at the departments of pathology of the Kalmar County Hospital and at Linköping University Hospital, Sweden [7]. They began digitizing glass slides in 2006 to improve the working process for pathologists at Kalmar County Hospital. Since 2008, more than 500,000 glass slides have been digitized for primary diagnosis in both hospitals. According to the web-based questionnaire completed by pathologists in Kalmar and Linköping, the working ergonomics was sufficiently improved and diagnosis was performed mostly based on the digital images. However, during the validation of the WSI, each specimen was investigated with both techniques—on the monitor and with a conventional microscope. The digital images were scanned at x20 magnification using two scanners (Aperio Scanscope XT) and converted into a JPEG format (with 70% quality). The Imagescope web viewer was used to view the digital images. To manage the samples, the Sympathy (Tieto, Sweden) information system was used in both hospitals. As the providers of the Sympathy laboratory information system and the Imagescope scanner do not provide a standard mechanism for data sharing, integration of the scanner was performed using the Picsara middleware software

(Euromed Networks, Sweden), which was developed with local partner companies. This software links digital images to the corresponding specimen into the laboratory information system. Despite the fact that routine diagnostics were effectively performed by WSI systems in both hospitals, it was declared that integration of the image web viewer into the laboratory information system was not completely adapted to the whole workflow.

In Stathonikos' et al. paper, the implementation mechanism of a full digital workflow in the pathology department of the University Medical Center Utrecht (the Netherlands) is represented [87]. Digital images stored at x20 magnification are compressed using JPEG or JPEG2000 image standards. Despite the use of the lossy compression method, they still required the storage of nearly 5TB data each month in order to store the images. The digital images are stored as objects on a tiered storage system that contains various storage units for short-term, long-term, and permanent images. The images located in the short- and long-term storage are automatically deleted after a specific period of time. The Department of Pathology developed a connection that links the specimen from the specimen reporting system to the corresponding digital images in the image management system. Using the one-way communication mechanism, the required metadata (such as staining name) from the laboratory management system is sent to the image management system.

One of the greatest limitations of these reports is that there is a lack of explanation regarding the communication mechanisms between the LIMS and image management system. However, it is clear that each laboratory will have to determine an optimal level of interoperability and integration techniques to achieve their specific goals within the scope of the research project or clinical diagnosis. Additionally, digital images were stored with an open file format (such as JPEG, JPEG2000), sometimes compressed with lossy compression mechanisms, to minimize data volume used for storing the images.

The reports that currently exist describe digital workflow using the specific systems (scanning systems and specimen management systems) provided by various vendors that are not the same systems used within the UMG. Therefore, questions remain on how to determine the optimal level of the integrated DP workflow used for the complex research environment within the UMG.

### **3. Requirements engineering for an integrated brain bank imaging system**

This chapter provides a description of the requirements for engineering processes performed for a biobank research environment. In the first part of this chapter, the main methods, principles, and techniques used in the requirements engineering process are presented. In the second part, results of the requirements analysis are described. In addition, market analysis of existing DPSs that can meet the research requirements are presented at the end of the chapter. As the requirements for the DPS have a leading role throughout the development process of the microscopic image handling system, specific features and accompanying UML diagrams are given in the following chapters (Chapter 4 and Chapter 5).

#### **3.1 Requirements engineering framework**

According to the Klaus Pohl's book, the term "requirements engineering" is defined as a systematic approach for the efficient detection, documentation, and management of relevant requirements using a conceptual model [34].

Precise planning and requirements analysis have a leading role in the successful development of an IT system. Requirements engineering provides an effective way to describe the tasks and requirements, avoid expected failures, and define the obstructive factors that could occur during the project development. The main goal of the requirements analysis is to ensure successful development of the comprehensive system in response to the project needs.

Within the scope of the KKNMS research project, requirements analysis for the development of the interoperable DPS were developed according to the handbooks of requirements engineering being based on the curriculum of the international requirements engineering board [34, 112].

As the interdisciplinary subject—digital pathology—is used for the biobank application within the complex research environment, there are a wide range of features from multiple perspectives that should be taken into consideration during the system development. Firstly, multiple stakeholders are involved in the process, including pathologists, IT technologists, developers, biobank technicians, and researchers with different responsibilities, skills, and knowledge. Therefore, all their concerns and requirements should be registered and closely examined. Secondly, DPS generally involves complex and distinct technologies and

information systems, such as a WSI scanner, acquisition software, image server, database, and image viewer. Multiple requirements should be considered to fully digitize pathology workflows and effectively use the above-mentioned technologies. Third, existing IT infrastructure for the KKNMS research project already involves multiple and various components, into which the developed DPS should be integrated. Therefore, different approaches are required to be taken into consideration to find effective interface solution. The multiple viewpoints approach ensures that requirements and processes of the system are categorized and managed according to the specific viewpoints by using structured, visual, and formal representations. In the beginning, viewpoints should be clearly identified. According to the three main perspectives mentioned above, the following viewpoints were detected: process viewpoint, information system viewpoint, and interfacing viewpoint [36, 113]. Process and interfacing viewpoints are represented using the UML models and diagrams, whereas the information system viewpoint is represented by describing each component of the complete DP infrastructure. Describing the processes from different perspectives provides a bigger picture of the needs that should be discussed during the requirements engineering to provide an effective solution at the end.

According to the scientific literature, the requirements engineering process involves the following steps: planning, requirements elicitation, interpretation, negotiation and analysis, specification, and validation crossed by requirements management [34]. General overview of the requirements engineering framework are illustrated in Figure 3.1. Below, each facet will be described in greater details within the context of a biobank research environment.

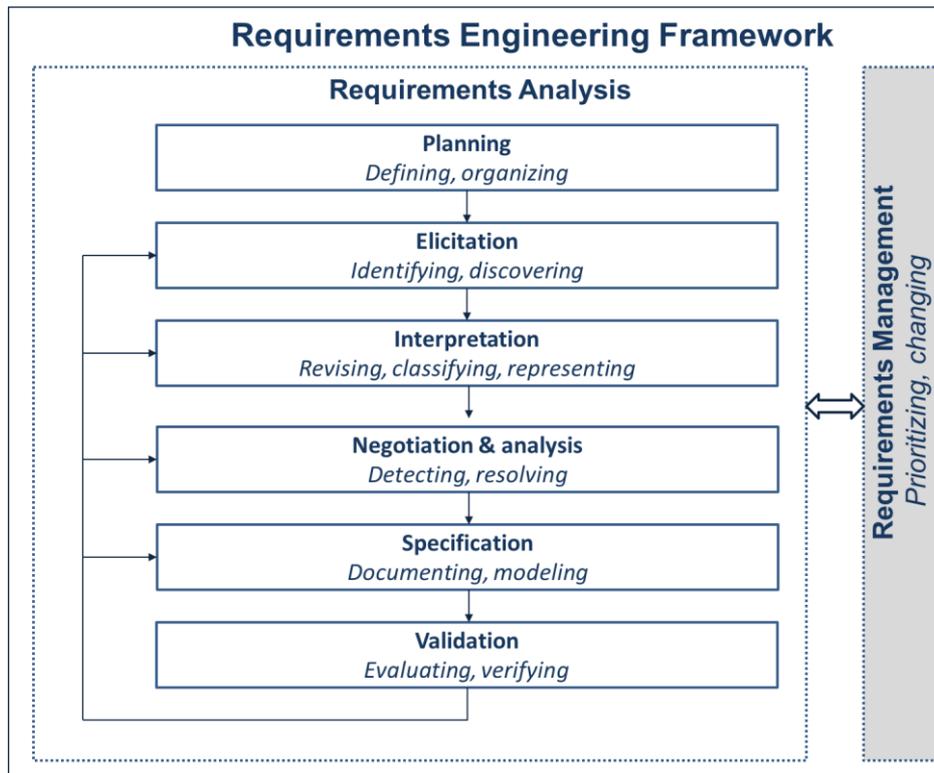


Figure 3.1. Requirements engineering framework. It contains two main building blocks: requirements analysis with several steps and requirements management [34].

### 3.1.1 Planning of requirements analysis

The earlier stage of the requirements analysis is to identify and define the system boundaries within the current operational environment in which the potential system should be developed [114]. Once the system boundaries are known, requirements for the development of the DPS for the research purposes can be identified more precisely. In addition, it helps to define the problems and possible solutions. If the planning phase is performed incorrectly, then the result of the requirements engineering can be ineffective and the final DPS can be incompatible with the existing research environment.

Complex IT infrastructure for the collaborative research network contains multiple components at the local and global levels. Within the scope of the KKNMS research project, this phase identified the following aspects:

1. Responsible stakeholders of the KKNMS research network who may have an interest in or impact on the DPS were considered.
2. Current system environment and system boundaries in which the DPS should be developed were identified and defined (described in Chapter 2.4).

3. Documentation of the KKNMS, regulations, and guidelines related to the subject were reviewed.
4. Hardware and software components of the existing complex IT infrastructure of the KKNMS, mainly regarding the biomaterial processes, were analyzed. The potential DPS should be in compliance with the current technical systems at UMG, such as microscopic scanner, laboratory management system, label printers, etc. If the technical systems do not satisfy the requirements, presumably the software will be updated or new systems will be purchased.

By identifying these issues, the on-going processes and the environment of the research project can be considered for further development. Thus, significant components at the planning phase are identified that may lead to a smooth course of the requirements analysis.

### **3.1.2 Requirements elicitation sources and techniques**

Requirements elicitation is the next step and the core part of the requirements analysis. Requirements for the development of microscopic image handling system within the biobank research project were collected from different sources. The main sources of requirements are as follows:

1. **Stakeholders** – Stakeholders are people or a group of people who are directly or indirectly related to the system to be developed. Identification of the stakeholders and their roles is an important task for the elicitation process [34]. The stakeholders describe their requirements from diverse perspectives. Based on those requirements, the needs of the whole system can be demonstrated. If some stakeholders are not included in the elicitation process, then the final system could be incomplete and may need additional work and costs to fill the gap later. The direct and indirect stakeholders are pathologists, IT technicians, medical informaticians, scanner manufacturers and developers of the information systems used in a laboratory. Within the KKNMS research project, most of the requirements were collected from the direct stakeholders that are the potential end-users (main pathologists) of the system from the Department of Neuropathology, and responsible people (medical informaticians, project manager, interfacing technician, research association) from the Department of MI. Furthermore, there were indirect stakeholders, such as technicians from the IT department at UMG, the responsible person from the Molecular and Optical Live Cell Imaging (MOLCI) unit where the scanners are located, and manufacturers of the microscopic scanner and the LIMS. The stakeholders'

roles are crucial and therefore the needs of all stakeholders should be taken into account during the requirements elicitation process.

2. **Environmental standards and regulations** – There are multiple regulatory as well as additional barriers that may influence the development and implementation of DPS (described in Chapter 2). The requirements elicited from the regulations, standards, and guidelines specifically dedicated for the research environment are included into the MS-BB research project.
3. **Pre-existing solution systems** – Functional requirements are likely to be collected from the existing pathology systems, such as scanners, viewers, and image analysis tools running in the same environment. In addition, requirements can be collected from the existing DPS or relevant products that are available on the market.

The following techniques were used for requirements elicitation:

1. **Regular meetings with stakeholders** – Most of the requirements were collected and discussed at the meeting with the direct stakeholders, i.e., pathologists from the Department of Neuropathology and medical informaticians from the Department of MI. After several brainstorming sessions, all ideas were collected and possible negative and positive sides were identified. Different working scenarios were discussed and analyzed with the main pathologist. Additionally, there were regular meetings and telephone conferences with other stakeholders. The main advantage of the brainstorming session was that requirements were examined from different perspectives. At the end of the meetings, the next steps were defined and the corresponding tasks were distributed to the responsible people.
2. **Observation** – Observation is the method used to acquire information by monitoring the processes. This method leads to enhancing the current workflow. Direct and ethnographic observation methods were used for the requirements elicitation to describe the activities and systems at the pathology laboratory [34]. During the direct observation, an observer watches the pathologist who is performing activities at the pathology lab, such as preparing and processing the histopathological samples at the laboratory. An observer is describing the processes, gathering the information and addressing them for better description of workflow. During the ethnographic observation, the observer is involved into the working process and continues their same activities as in normal situations in the lab. The ethnographic method was used for system observation to describe and test the

information systems used in the pathology lab, such as digital scanning systems and the LIMS. By observing the systems, sample glass slides were digitized using the digital scanners to analyze the DP workflow. In addition, the observer used the LIMS system with a test user account and performed the same procedures as the pathologist did. By using the described observation methods, a profound understanding of the working processes and IT systems used in the lab were gained. One disadvantage of the latter method is that the external person's observation might not be as proficient as that of the stakeholders'. Thereby, we used the ethnographic method only for the system observation, as it is more technical, rather than the laboratory activities, such as tissue and slide processing, for which direct observation was used.

3. **Documentation-based requirements elicitation technique** – Some of the requirements were collected from the documentation of the KKNMS project, which described the general needs and objectives of the project. Furthermore, additional requirements provided by different associations and organizations (such as TMF, FDA, European Union, CAP, etc.) were considered for the research environment. Crucial components of these requirements were discussed with the stakeholders.
4. **Hardware and software analysis** - This step determines whether targeted systems are able to reach the research goals. During this study, two digital scanners were tested for digitization of glass slides. At the beginning, the Olympus dotSlide scanner was estimated, and later, the Olympus VS120 became available at UMG. The Olympus VS120 was subsequently used for the scanning processes. Several glass slides were scanned to estimate technical details like file size, scanning duration, conversion tools, and batch scanning modes. Based on that information, the input and output components of the DPS were determined. While the project is sufficiently broad and different working groups are involved, the pathology information system used for the KKNMS needed to be considered, and the interfacing requirements were elicited during the examination of this system. Within the KKNMS project, the STARLIMS application is used for the management of laboratory processes. Documentation and sample registration, as well as the tracking of tissue samples, were analyzed in STARLIMS. A key component of this step was a feasibility study to determine the ability of integration between targeted systems. In addition, the main restrictions and barriers of the systems were identified and discussed with the system manufacturers.

5. **Market analysis** - A wide range of requirements was collected and the significant needs were clarified from the market analysis. The most prevalent web viewers for microscopic digital images on the market were investigated and the main characteristics, advantages, and disadvantages of each viewer were evaluated. Market analysis was conducted to identify supplementary features of the DPS and to subsequently choose the best solution being compatible with the needs of the research project.
6. **Interviews** – Several official meetings were conducted with the manufacturers of the image handling systems, such as PathXL, ImageScope, VMScope, Chili, and Zeiss [115–119]. Using the interviews, potential image handling systems were estimated for our research needs. Furthermore, additional requirements and desirable features were gathered from this technique.
7. **Support techniques** – Information regarding the supplementary features of the DPS was gathered from international conferences, exhibitions, workshops, webinars, etc.

Requirements were mainly elicited using predefined questionnaires that were focused toward the stakeholders in the elicitation procedures (during the observation, interviews, and regular meetings).

### **3.1.3 Interpretation of requirements**

Once the raw requirements were collected, they needed to be revised for the selection of the relevant requirements that satisfy the main goals and fit the final solution. During the interpretation phase, solution-oriented requirements were sorted according to their importance and classified into the following groups: functional, technical, quality, and environmental requirements. Regardless of the classification, requirements mostly are delivered with several use case scenarios breaking down into different activities. For this reason, gathering requirements were represented by several scenarios in which the stakeholders could more easily understand the processes of the pathology workflow. The scenarios were represented in such a way that answers the following questions: how can researchers use the system, how should the scanning process be performed, how are digital images stored and handled, etc. (Chapter 3.2.3).

### **3.1.4 Negotiation and analysis of requirements**

After representing the necessary requirements that fit the needs of research project, some features are likely to be in conflict with each other or with existing systems. Conflicts in the

requirements can arise especially when there are multiple stakeholders and complex infrastructure. We faced such problems not only in the requirements analysis phase, but also during the integration of the DPS and the LIMS related to the data exchange. As the old version of the imaging system was not able to interact with the external system, the image server was updated to the latest version. Our goal is to identify such kind of conflicts as soon as possible, analyze them, and try to find an optimal resolution.

### **3.1.5 Requirement specification and modeling techniques**

After requirements negotiation, all requirements should be documented and corresponding models need to be constructed. Requirements models, such as class or data flow diagrams, are widely used in requirements analysis to represent and describe the features of the system that is planned to be built [120]. In software engineering, a Unified Modeling Language (UML) provides a standard way to construct the models and depict the features of the system using visual components. Additionally, the UML diagrams represent system design in such a way that people from different disciplines can quickly understand the main aspects of the system. Based on various types of UML models, we used activity diagrams, use case specifications, process and sequence diagrams (in Chapter 4 and Chapter 5).

Model-based requirements documentation ensures better understanding compared to the natural language. In addition, each model describes the specific aspects of the whole system that reduce the complexity of discourse and emphasize the most significant details of the processes.

### **3.1.6 Requirements validation**

The main goal of this stage is to validate the identified requirements and the extent to which the current documented requirements are deviated from the wishes of stakeholders [34]. During this facet, requirements artifacts were evaluated and specific errors were discovered, such as incompleteness and inconsistency of the requirements. The sooner the deviations between the requirements were identified, the fewer problems and changes will occur during the development of the DPS. Leading methods used in the validation stage were requirements inspection with internal and external stakeholders and testing the functional requirements [121]. For example, ambiguous requirements and processes were discussed with pathologists and the IT technicians. Also, some requirements regarding the integration processes were discussed with the manufacturers of the LIMS and the WSI devices. Besides this technique,

trial cases were determined to identify the deviations between the requirements, e.g., technical requirements, such as file size limitations and barcode scanning failures, were revealed during the testing the scanning process. According to the discovered errors and deviations in this facet, specific changes can occur in the previous steps of the requirements analysis, and repeated validation is likely to be performed at any time during the whole process.

### **3.1.7 Requirements management**

Within the scope of a biobank research environment, requirements management comprises a determination of the method for the requirements prioritization and management of requirements changes. As requirements management is likely to be a continuous process, it goes in parallel to the requirements analysis from beginning to the end. Throughout the requirements analysis process, and mostly in the interpretation facet, the requirements need to be prioritized with regard to the needs of the stakeholders. Therefore, three types of requirements were determined to identify the value of the gathered requirements: essential (must-have) requirements, preferred (optional) requirements when they are achievable within the system, and desirable (nice-to-have) requirements that could be considered if they are reasonable. All the requirements not included in these groups were defined as unnecessary features or features that are desired but not possible to implement in the current system. Therefore, such requirements were ignored during the requirements analysis and development process, but they were postponed for future enhancement of the system.

During the life cycle of the system, some of the requirements that require changes in requirements analysis and system development processes might be added, modified, or removed. Within the biobank environment, the main changes in requirements were caused by installing a new scanning system within the UMG that was purchased to fulfill the needs of a biobank research project and the needs of the other applications in the future. These changes significantly improved the pathology workflow, enhanced scanning settings, and improved the quality of the digital images. Additionally, the viewing application and server infrastructure were updated and other details of the system were configured in the way that satisfied the essential requirements of the project (described in Chapter 4 and Chapter 5). Some of the requirements were changed based on errors (or difficulties) that were encountered during the trial cases. At the beginning of the requirements analysis, it was declared that images should be stored as open file formats such as TIFF, BigTIFF, or JPEG2000.

However, when the test samples were digitized and conversion tools were verified, it was discovered that converting images into open file formats is a very time consuming task and was mostly completed unsuccessfully (with errors). While the proprietary file format keeps all imaging information and metadata, requirements for image formats were modified accordingly and images will be stored as .vsi files (proprietary file format provided by the scanner) on the server. They can later be converted into another file format if needed.

Over the course of the system life cycle, all potential changes were discussed with the appropriate stakeholders prior to implementation.

## **3.2 Results of requirements engineering for the interoperable DPS**

Based on the methods described in previous subchapters, requirements analysis was carried out for the biobank research network. The most important results of the requirements engineering are represented below.

### **3.2.1 Requirements elicitation questions**

Asking the right questions leads to a system development process and puts it on the right track [122]. In addition, well and correctly organized questions can reveal problematic issues. Questions were categorized into several sub groups that provide better understanding of the problems and needs of the target user groups [123]. The questions asked during the requirements elicitation are listed in Table 3.1.

The developed DP workflow integrated into the laboratory information system is described in the following chapters (Chapter 3.2.2, Chapter 4, and Chapter 5) that answer all questions listed below and describe the processes in detail.

Table 3.1. Requirements elicitation questions. The questions are categorized into the following groups: general questions, questions for functional and non-functional requirements, and interfacing questions.

General question	
<ul style="list-style-type: none"> <li>▪ What is the primary objective of the DPS?</li> <li>▪ What is the current state of the DP within the UMG?</li> <li>▪ What are the benefits of developing the DPS?</li> </ul>	
Questions for functional requirements	
<b>Preparation</b>	<ul style="list-style-type: none"> <li>▪ How are the glass slides identified?</li> <li>▪ Who should prepare the glass slides for digitization?</li> <li>▪ What kind of label is essential for glass slides?</li> </ul>
<b>Input &amp; output</b>	<ul style="list-style-type: none"> <li>▪ What kind of data should be input along with the glass slide during the scanning?</li> <li>▪ What kind of scanning settings should be configured?</li> <li>▪ What kind of scanning mode should be chosen for effective use of the system?</li> <li>▪ What kind of data must be received after scanning?</li> <li>▪ What kind of images should be acquired?</li> <li>▪ What imaging metadata should be stored along with the digital image?</li> </ul>
<b>Management</b>	<ul style="list-style-type: none"> <li>▪ What kind of imaging data should be stored?</li> <li>▪ Where should the digital images be stored?</li> <li>▪ How should the folders and images be structured and categorized?</li> </ul>
<b>Viewing</b>	<ul style="list-style-type: none"> <li>▪ What will the main functionality of the image viewer need to be?</li> <li>▪ What kind of viewing platforms are desirable?</li> </ul>
<b>User</b>	<ul style="list-style-type: none"> <li>▪ Who will use the DPS?</li> <li>▪ What kind of users should be registered in the DPS?</li> <li>▪ What kinds of roles and permissions are necessary for each kind of user?</li> <li>▪ What is the objective of creating different user roles?</li> <li>▪ How should the user accounts be managed?</li> </ul>
Questions for nonfunctional requirements	
<b>Security</b>	<ul style="list-style-type: none"> <li>▪ How well is the system secured from unauthorized users?</li> <li>▪ How well are the digital images secured on the server?</li> <li>▪ How well is the metadata exchange guarded between different applications?</li> <li>▪ Will the LIMS and DPS have separate authorization systems?</li> <li>▪ Does the security provide the only read/read&amp;write permissions for individual images or folders?</li> <li>▪ Does the system provide the error messages?</li> </ul>
<b>Access</b>	<ul style="list-style-type: none"> <li>▪ Does a user have access to only specific images (folder) or the whole system?</li> <li>▪ What should the average time be for loading images into the viewer?</li> </ul>
<b>Storage</b>	<ul style="list-style-type: none"> <li>▪ How much storage capacity is necessary to handle a large amount of imaging data for a specified period of time?</li> <li>▪ How easy is it to add data volume once the system is already operating?</li> </ul>
<b>others</b>	<ul style="list-style-type: none"> <li>▪ How easy is it to learn and operate the system?</li> <li>▪ How easy is it to convert/modify the system (or part of the system) for usage of another application in the same or different institution?</li> <li>▪ How easy is it to transfer the system?</li> <li>▪ Does the system need to be applicable with the batch scanning mode?</li> <li>▪ How easy is it to customize the operable system with regard to the needs in future?</li> <li>▪ How easy is it to update the operable system in the future?</li> <li>▪ How it should be shown that system performs its functions?</li> <li>▪ How should the system be maintained?</li> </ul>
<b>Quality</b>	<ul style="list-style-type: none"> <li>▪ How should the accuracy and quality of scanned images be checked?</li> <li>▪ What kind of scanning settings should be set to ensure high quality images?</li> <li>▪ How fast and how well does the system (viewer) respond?</li> </ul>
Interfacing questions	
<ul style="list-style-type: none"> <li>▪ How should systems be interlinked?</li> <li>▪ To what extent are the LIMS and DPS interoperable to each other?</li> <li>▪ What kind of triggers should be implemented in the interface?</li> <li>▪ How should trigger events be executed?</li> <li>▪ How can the DPS and LIMS interface with other applications (systems)?</li> </ul>	

### 3.2.2 General requirements of the DPS

According to the listed questions, all requirements were collected and categorized into the following groups: functional, technical, quality, and environmental requirements. As the DPS contains several sub components, there are specific requirements determined for sub-processes, such as requirements for labeling the glass slides or the web service interface. Only the general requirements of the DPS are listed in Table 3.2.

Table 3.2. List of the requirements for the DPS categorized into the following groups: functional requirements, technical, quality, and environmental requirements.

Functional requirements
<b>Data input:</b> <ul style="list-style-type: none"><li>• Batch scanning mode is required for scanning glass slides</li><li>• Tools are required for uploading the images to the system:<ul style="list-style-type: none"><li>▪ Storing images either automatically on the server during the scanning process (preferable) or later from a local computer to the server by operators</li><li>▪ Entry of accompanying metadata for a large number of glass slides using an excel file is preferable</li></ul></li><li>• A tool for adding annotations on a specific area of an image and registering them into the image database is desired</li></ul>
<b>Data management:</b> <ul style="list-style-type: none"><li>• Data needs to be organized hierarchically in the following manner: Subject (Brain) -&gt; Sample/Specimen(s) -&gt; Image(s)</li><li>• Accompanying metadata needs to be stored along with the images</li><li>• Brains, specimens, and images need to be identified with their own ID within the DPS</li><li>• Glass slides should be labelled and uniquely identified with a Data Matrix barcode</li></ul>
<b>Access control:</b> <ul style="list-style-type: none"><li>• Access to the system/images needs to be restricted to registered users with configurable roles</li><li>• Operators could be able to grant temporary access to certain images to unregistered users (desirable)</li></ul>
<b>Viewing/Display:</b> <ul style="list-style-type: none"><li>• A web viewer for viewing images in a browser is required</li><li>• The viewer has to support at least zooming, panning, and navigating the images. Additional image viewing and processing features, such as changing color settings and image rotation, are preferable.</li><li>• Side-by-side image viewing and annotation tools are desired</li><li>• The accompanying metadata need to be displayed alongside preview images of the samples on an overview page</li><li>• A web viewer has to provide the viewing tools for accompanying annotations on a specific area of the image</li></ul>
Technical requirements
<ul style="list-style-type: none"><li>• The system needs to be web-based</li><li>• All connections to the system must use protected/encrypted channels (e.g., TLS)</li><li>• Compatibility with the output file formats of the Olympus VS120 system (e.g., VSI, JPEG2000, or TIFF) is required</li><li>• Support for large file sizes (up to several GB for a single file) is required</li><li>• On-demand image streaming is required for the viewer</li></ul>
Quality requirements
<ul style="list-style-type: none"><li>• Rapid scanning and fast image transfer to the server are required</li><li>• Quick response time for image loading on the web viewer is required</li><li>• System needs to be modifiable and usable</li><li>• Security and reliability of data are required</li><li>• Data storage system with expandable storage capacity is required</li></ul>
Environmental requirements
<ul style="list-style-type: none"><li>• System needs to be integrated into the LIMS</li><li>• An interface needs to be built within STARLIMS that interacts with the DPS and exchanges the imaging metadata</li><li>• Triggers should be executed in the interface and its settings (such as frequency and interval) should be set according to needs</li></ul>

As the next chapters (Chapter 4 and Chapter 5) describe the processes of the components of DPS in detail, other requirements for specific modules are also represented along with the proper processes.

### 3.2.3 Use case scenarios in the requirements engineering

Processes within the DPS are being represented by five scenarios that are listed in Table 3.3. All scenarios are described with accompanying textual specification in Appendix B.

Table 3.3. List of DPS scenarios.

No	Scenario name
S1	Preparation of glass slides
S2	Digitization of glass slides
S3	Managing user roles and administrating images within the DPS
S4	Viewing of digital images by researchers of the KKNMS or by a pathologist of the MS-BB
S5	Data sharing between the LIMS and the DPS using middleware interface

The S1 describes the preparation of the glass slides for digitization that includes several activities for generation of unique IDs and label components for the slides, as well as label printing and attaching procedures (scenario is described in Appendix B, Table B.1). The S2 provides standard operating procedures for scanning the glass slides with their accompanying metadata. It also includes procedures for automatically saving the images on the server (scenario is represented in Appendix B, Table B.2). The S3 defines the management of user accounts and administrative procedures for effective storage of imaging data on the server (scenario is described in Appendix B, Table B.3). The S4 provides procedures for viewing digital images using a web viewer. In addition, it describes the main characteristics and customization tools of the web viewer (described in Appendix B, Table B.4). The S5 defines manual and automated processes for the effective linkage between the DPS and the LIMS through a SOAP-based web service interface. In addition, it describes the trigger activities and configuration steps for middleware application (scenario is described in Appendix B, Table B.5). Technical details, setup of the system, development of the interfacing solution, configuration, and customization details with regard to each scenario are described in Chapter 4 and Chapter 5.

### 3.2.4 Market analysis of existing digital microscopic image handling systems

After defining essential requirements of the research project, market analysis with regard to the digital image handling systems was conducted. The main goal of the comparative market analysis was to ascertain that the image handling system is likely to meet the needs of the biobank research project and to gather supplementary requirements that were later taken into account during the development of the imaging system.

In total, 18 image handling systems based on the different platforms and developed for different purposes were investigated. Main components, features, advantages, and disadvantages of these systems were addressed. Supplementary requirements were discovered through market analysis.

Some of the examined viewers were dedicated for different applications, such as education, research, and clinical practice. However, a general overview of these products was important in developing a robust and capable system for a research-specific network. The main image handling systems that were mostly applicable to our requirements are illustrated in

Table 3.4. All details of the other image handling systems are presented in Appendix C (Table C.1).

One of the most important features of the image handling systems is the supported file formats. Because most whole slide imaging scanners have their own proprietary file formats, tools for converting these formats into open file formats are extremely essential. Almost all image handling systems support standard file formats. The ImageScope and Simagis additionally work with the BigTIFF file format. The OMERO web viewer supports the OME-TIFF file format, which was specifically created for microscopic images and handles rich metadata along with the TIFF image file. As Olympus provides a proprietary file format (.vsi), a very limited number of viewing applications, such as NIS, PathXL, Simagis, and mScope, support it.

According to the results of the market analysis, potential products were selected from the list and individual meetings were scheduled with the system manufacturers for further investigation. The best suited image handling systems for the biobank research environment were provided by Olympus NIS, PathXL, ImageScope, OMERO, and Zeiss. However, significant financial or technical limitations were identified during the meeting with some of the manufacturers, including high price, incompatibility with the LIMS, insufficient tools for image viewing in the browser, etc.

Based on the market analysis, the most suitable solution was chosen according to the following three criteria: a solution (1) meets all essential and preferred requirements of the MS-BB imaging system; (2) is interoperable with different applications via the most usable methods such as web services, file transfers, or direct database communications; and (3) has the potential for future improvement. Thus, an effective solution was determined that led to the successful implementation of a DP workflow in the laboratory for biobank processes. As a result, the Olympus NIS was chosen for handling of digital images as it provides almost all essential and additional features that meet the needs of the biobank research network. Once the decision was made, NIS v.2.9 was installed and the web viewer was updated to the latest version.

Table 3.4. List of the image management systems and viewers in digital pathology (based on the market analysis performed in 2013).

N	Name	Developed by	Supported file formats	Navigation tools	Based on	Advantages	Disadvantages	Commercial/open source
1	Net Image Server (v2.8)	Olympus	.vsi, TIFF, JPEG2000 and all standard file formats	STANDARD <sup>1</sup> + Annotation, Joystick, take snapshots, send region of interest via email, side-by-side viewing tools, color settings	HTML5	<ul style="list-style-type: none"> <li>• Support for vsi and all standard image formats</li> <li>• Viewing all metadata</li> <li>• Support for SOAP-based web services</li> <li>• Access by web-viewer, Olyvia software, or mobile applications</li> <li>• Rapid access of virtual slide image</li> <li>• Security with password and user login</li> <li>• Support for adding advanced annotations</li> <li>• Automated data storage from VS120 scanner to NIS</li> </ul>	<ul style="list-style-type: none"> <li>• Connection with only Olympus scanners</li> <li>• Old scanner (dotSlide) is unlikely to be connected directly to the updated NIS</li> </ul>	Commercial
2	PathXL	PathXL	Olympus, Zeiss, Aperio, Hamamatsu, leica, Philips, Ariol, Microscan, 3Dhistech	STANDARD + Annotation, upload media, take snapshots	HTML5 + JavaScript	<ul style="list-style-type: none"> <li>• Support for modules for biobank, research, education, and clinical usage</li> <li>• User management system</li> <li>• Image analysis framework</li> <li>• Support for annotations, multiple windows, TMA toolbox, and case modules</li> <li>• Support for API integration into the LIMS</li> </ul>	<ul style="list-style-type: none"> <li>• High cost</li> <li>• Complexity of the consolidation methods for the multiple scanners provided by different vendors</li> </ul>	Commercial
3	Open Seadragon	Silverlight	.dzi, .xml (.jpg)	STANDARD+ Full screen	HTML, AJAX	<ul style="list-style-type: none"> <li>• Switching images</li> <li>• Adding custom controls</li> </ul>	Viewing images only by deep zoom composer	Open source
4	Digital Image Hub	Slidepath, Leica Biosystems	Standard file format, Aperio, Hamamatsu, Leica	STANDARD + Joystick, Free rotate, brightness panel	Flash player	<ul style="list-style-type: none"> <li>• Support for annotations</li> <li>• Support for multiple windows</li> <li>• Well-organized panel</li> </ul>	<ul style="list-style-type: none"> <li>• No support for other proprietary file formats</li> <li>• Lack of the user management system</li> </ul>	Commercial
5	Webscope/ ImageScope	Aperio	.svs, .TIFF (BigTiff file) Aperio - svx	STANDARD	Flash player	<ul style="list-style-type: none"> <li>• Add and view annotations</li> <li>• Adjust brightness and contrast</li> </ul>	<ul style="list-style-type: none"> <li>• No support for other proprietary file formats</li> <li>• Lack of the user management system</li> </ul>	Commercial
6	ZEN browser (ZEISS Axio Scanner Z1)	Zeiss	Zeiss (Mirax), and all standard file formats	STANDARD+ Annotation, side-by-side viewing, color settings	Flash player	<ul style="list-style-type: none"> <li>• Saving metadata and supporting documents</li> <li>• User management system</li> <li>• Remote access for mobile devices</li> </ul>	<ul style="list-style-type: none"> <li>• Connection with only the Zeiss scanners</li> </ul>	Commercial

<sup>1</sup> STANDARD (navigation tools) include the following tools: zoom in, zoom out, pan, reset view, and navigator.

## 4. Development of the digital workflow and the interoperable digital image handing system

A DP workflow for the biobank environment consists of three main components: preparation, digitization, and provision (Figure 4.1). The preparation stage is comprised of methods, tools, and procedures for the characterization of the MS lesions and collection of the tissue samples and glass slides. Furthermore, it involves techniques for preparing glass slides for digitization, such as label printing and attaching procedures. The digitization stage contains hardware and software tools for the acquisition, management, and storage of digital images. The provision step consists of applications for viewing digital images and interfacing solutions between the LIMS and the DPS. The features and processes of each step require detailed descriptions. The following subchapters successively represent the main technical aspects of the DPS with accompanying requirements models.



Figure 4.1. Elements of the DPS. The workflow of the DPS contains three main components: (1) preparation (pre-digitization) for preparing glass slides for scanning, (2) digitization for scanning and management of digital images, and (3) provision (post-digitization) for viewing and analysis of digital images.

### 4.1 Preparation of glass slides

Prior to digitizing the glass slides, the identification mechanisms and barcode labeling techniques for already processed glass slides should be determined. However, even prior to that, histological and immunohistochemical processing of MS tissue samples are to be considered in order to successfully determine the next steps forward to digitization.

#### 4.1.1 Multiple sclerosis tissue sample processing in a pathology laboratory

At the beginning of the requirements analysis, the histological and immunohistochemical processes for tissue sample and glass slides preparation were analyzed. This process is comprised of the standard operating procedures for autopsies of the MS patients, histological staining procedures, and characterization of the MS lesions with regard to the localization and lesion activity. In this subchapter, the significant histopathological features of the tissue

samples that need to be taken in consideration during the development of the DP workflow and operation of the image handling system are described.

MS is an auto immune disease of the central neural system in which the cover of the cell, myelin, is damaged. Demyelinated zones delay the transmission of nerve impulses between the brain and the body [125]. Figure 4.2a shows a schematically represented damaged neural cell compared with a normal nerve. Figure 4.2b macroscopically illustrates the brain section in which the damaged area appears with the white plaques [124].

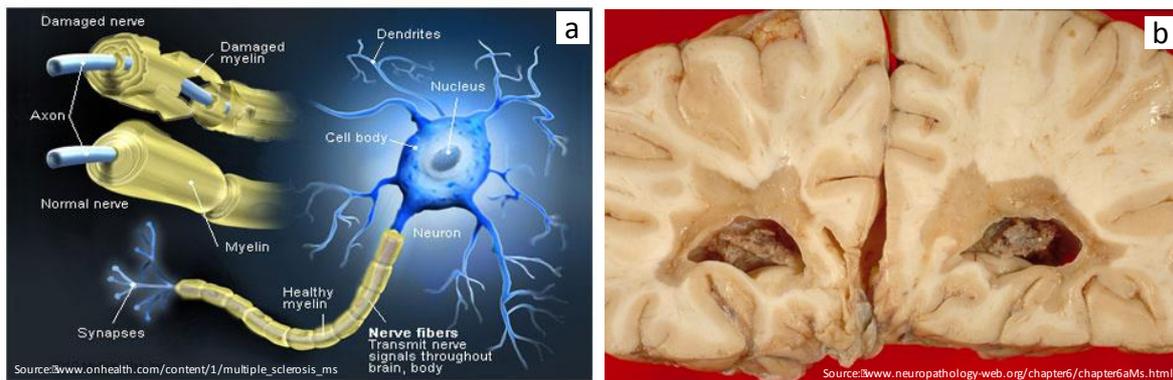


Figure 4.2. Myelin damage in MS. (a) Schematic representation of normal and damaged nerves; (b) Macroscopic image of a brain section. MS plaques extend into the grey matter [124].

The location of the lesion of MS defines the nature of the disease [126]. There are likely to be multiple inflammatory areas that extend to a larger area. Therefore, MS plaques could be located in the grey matter, white matter, or deep grey matter. Lesion activity is likely to be active, chronic active, inactive, or shadow plaque. In order to identify the characterization of the MS lesion, tissues are colored with different standard staining procedures such as H&E (Hematoxylin and eosin), LFB-PAS (which stands for Luxol fast blue-periodic acid), Schiff, Bielschowsky, Kim1P macrophage marker, PLP (Proteolipid protein), etc. For example, MS plaques appear with more fade colors compared to the undamaged areas on the tissue sample stained with the H&E. Histopathological stains are characterized by a wide range of colors. They may begin as a very dark blue that turns to pink, and end as a very faded grey color (Figure 4.3).

Characteristics of MS lesions are identified with the following histological staining methods:

1. Staining standards for myelin and axonal damage: H&E, LFB-PAS, and Bielschowsky
2. Inflammatory activity: B- and T-cells (CD20, CD3, CD8), macrophages (KiM1P, MRP14)

3. Demyelination and remyelination: PLP, MBP, MAG, CNP, MOG
4. Oligodendrocytic marker (Nogo A)
5. Markers for the axonal damage (SMI 31, SMI 32, APP)

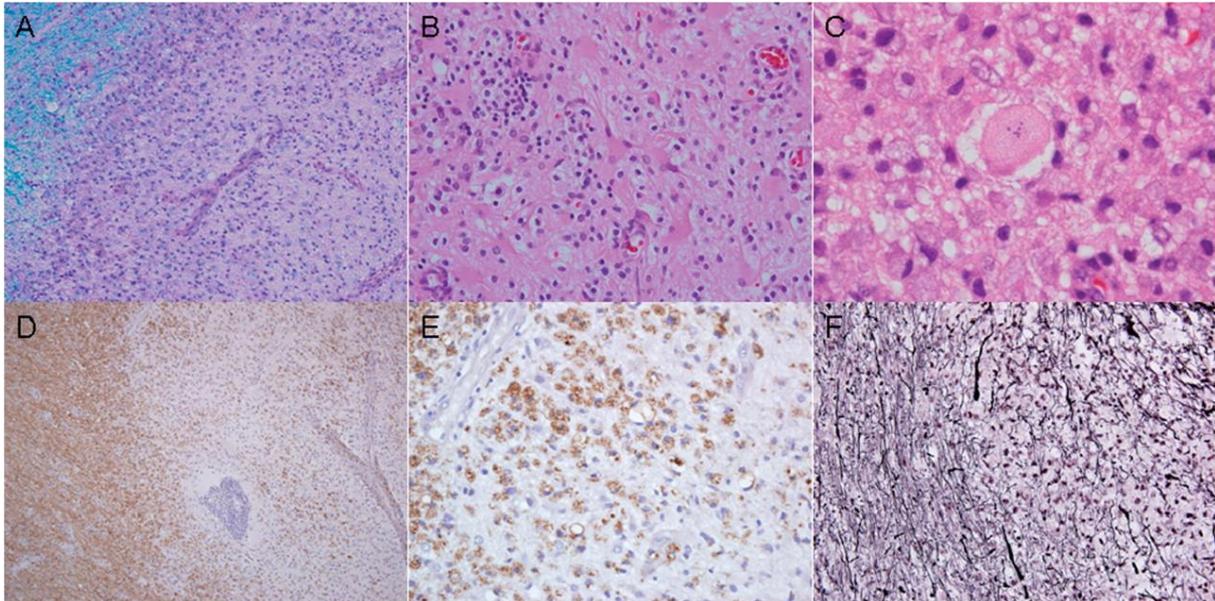


Figure 4.3. MS lesion stained with different colors. A - Luxol-fast blue and periodic acid Schiff; B – Hematoxylin and eosin; C - Hematoxylin and eosin; D and E - immunocytochemistry for proteolipid protein; F - Bielschowsky silver impregnation [127].

Within the scope of the KKNMS, the tissue samples are processed according to the standards of BrainNet Europe [128]. While the tissue blocks are cut from the brain, they are either wax embedded or frozen as soon as possible to preserve the structure of the tissue. Tissues are processed into paraffin and stored in a cassette, and they are sectioned and mounted on glass slides that are later stained with H&E, LFB-PAS, Bielschowsky, Kim1P, or PLP. If the lesion is active, then the tissue is examined with additional staining methods. If the lesion is inactive, the macrophage markers (Kim1P, MRP14) are used to detect microglia.

Staining colors, characteristics, and localization of the lesions are the main features that should be taken in account during the digitization of glass slides to acquire high-quality digital images based on the proper calibration and configuration of the scanner.

#### 4.1.2 The information management system for a pathology laboratory

There is a wide range of information systems to effectively manage laboratory processes that are used for clinical research, biorepository, or other purposes. LIMS provides powerful tools to improve laboratory workflow and facilitate the specimen handling mechanisms. Even though a LIMS is being rapidly developed and an increasing number of tools are being

implemented according to laboratory needs, there are still several limitations. For example, most of the LIMS face problems regarding the handling of specimen-derived assets, such as microscopic glass slides. As there are several new autopsy cases every year, biomaterial collections grow accordingly. Therefore, robust systems with advanced features are required for the effective management of the biomaterial data.

Within the scope of KKNMS, the STARLIMS web-based application offered by Abbott Informatics is being used for the management of the laboratory processes for the MS brain bank [129]. Using the web-based platform, the main pathologist registers all specimens into a local STARLIMS being already established within the scope of the KKNMS. Once the tissue sample is registered and characterized into STARLIMS, it is linked to the specific brain area and a unique ID is generated for each sample. For effective handling and management of biomaterial data, a pathologist can use comprehensive features of the LIMS, such as tracking management, quality control management, reporting, and others. Within the LIMS system, specimens of the biorepository are easily collected, tracked, and effectively processed throughout the entire specimen lifecycle.

As STARLIMS, like any other LIMS, cannot handle a huge number of high resolution digital images within the system, a DP workflow needs to be developed and separately managed by various information systems that will be later integrated into STARLIMS. According to a statement from the STARLIMS corporation, the LIMS applications can effectively work and exchange data with various information systems, such as audit systems or enterprise resource planning, known as ERP. Until now, however, the STARLIMS system has not been interfaced with any kind of digital microscopic image handling systems. To exchange metadata between the LIMS and the image handling system, an interfacing solution needs to be developed.

#### **4.1.3 Hierarchical structure of pathology data**

Hierarchy of information entities in pathology, depicted in Figure 4.4, contains four levels: patient, case (Kit in STARLIMS), block (same as tissue sample/specimen), and glass slide. Clinical data, like patient information, is handled by separate information systems. For the KKNMS research network, secuTrial, a web-based electronic data capture system, is being used to collect and manage patient data. Only two components of hierarchy—a case and a block—are managed by a LIMS. In general, glass slides are represented either as a digital

format handled by a separate system or they are not stored digitally at all (glass slides are only inspected using the traditional microscope).



Figure 4.4. Hierarchy of information entities in pathology. Patient is illustrated in green as it is managed by a separate system – secuTrial; a case and a sample are illustrated in blue, as they are handled by a LIMS. Glass slides illustrated in red are not able to be handled by a LIMS.

According to the data protection framework, Patient ID and Case ID are separated from each other. They are managed by the various information systems within the already established KKNMS infrastructure. The secuTrial provides pseudonymized registration of patient data. The pseudonym, generated by secuTrial, is indicated for each case (Kit) into STARLIMS, by which each case can be linked to the corresponding patient. Thus, two-step encryption of Patient ID is used to provide a secured system and protect patient information. Case ID (brain ID), also known as Kit ID in STARLIMS, is represented with a 17-digit sequence of numbers and letters separated by an underscore (e.g., Kit\_BB\_0123456789). Unique ID of the sample is generated from the Kit ID, which is followed by the suffix “TIS\_” and a 2-digit identification number (e.g., Kit\_BB\_0123456789\_TIS\_01).

As was defined within the scope of the project, approximately 60 specimens might be registered for each brain case in a local LIMS. Hence, a 2-digit identification system allows users to add a maximum of 99 specimens for each brain, which was defined sufficiently enough at the beginning of the project. As there are specimen-derived assets like glass slides, configuring the LIMS in that way was required in order to identify glass slides along with their specimen information to effectively manage biomaterials. For this purpose, the identification mechanisms for the glass slides were determined and implemented in the biomaterial management software. The main aspects of the identification and barcode labeling concept implemented into STARLIMS are described below.

#### 4.1.4 Identification mechanism and barcode labeling concept for glass slides

Using unique IDs, management of glass slides is improved and the handling of digital images is significantly facilitated. Identifiers and other characteristics of the glass slide are printed on a label that is later attached to the label area of the glass slide. Specific requirements for the

identification and labeling techniques, such as the material of a label and essential elements printed on it, should be considered for effective identification of slides.

#### 4.1.4.1 Requirements for identification of glass slides

According to the guideline for the uniform labeling of slides provided by the College of American Pathologists, there is a lack of regulations for the identification of specimen-derived assets [37]. Labeling and identification mechanisms for glass slides vary from laboratory to laboratory, and should be managed by specimen management software. Within the biobank research environment, we collected all requirements for the identification technique and implemented it in the biomaterial management software. The main requirements for the identifiers of the glass slides are as follows:

1. Unique identifier is required for the glass slides (Slide ID) to effectively track slides.
2. The ID of a glass slide should be constructed with a similar structure as that of the identifiers of the case and sample. It should contain the Sample ID in order to identify from which specimen it was derived (Figure 4.5).
3. Identification number for the glass slides is required to be represented as a suffix that contains the identification name as “IMG” and a 3-digit number (e.g., Kit\_BB\_0123456789\_TIS\_01\_IMG\_001). Each specimen placed in the paraffin cassette is sliced into very thin layers that are then mounted on glass slides. Using a 3-digit number in the suffix of the ID, multiple slides are likely to be identified within the single paraffin block. In most cases, this feature could be effectively used in the future for the three-dimensional (3D) representation of specimens. Therefore, multiple images (more than 100 slides) of each block should be efficiently identified to construct a space representation of each tissue sample.
4. In an ethical context, the identifier of the glass slides should not involve any patient IDs that are considered confidential information. If there is a patient identifier already handwritten on the glass slide, it should be removed.
5. Unique IDs for glass slides should be generated using the user-friendly interface built into the laboratory management software.

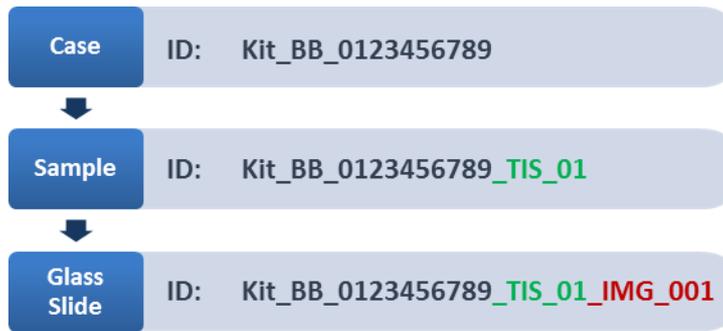


Figure 4.5. Identification of pathology information in STARLIMS. A Case ID is represented with a 17-digit sequence of numbers and letters separated by an underscore. The Sample ID consists of Case ID and a suffix “\_TIS\_” along with an auto incremented 2-digit number (suffix is written in green). A Glass ID is represented with the Sample ID and the suffix highlighted in red.

#### 4.1.4.2 Requirements for barcode labels of glass slides

Glass slides of the MS biorepository have standard dimensions of 1 x 3 inch (25 mm x 75 mm). A slide is divided into two parts: flat surface area for labeling (also called a label area) and tissue area that is covered with a coverslip (also called a scanning area). Figure 4.6 illustrates the dimensions of each part of a glass slide.

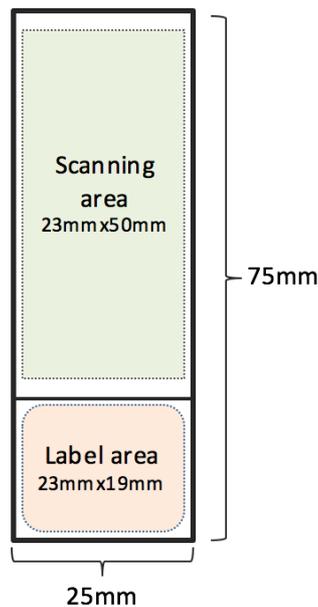


Figure 4.6. Dimensions of scanning area and label area of glass slides.

Labels for the biomaterial equipment, such as tissue cassettes, frozen containers, and glass slides are widely used in pathology laboratories. Using barcode labels, specimens or specimen-derived assets are identified simply, which facilitates the handling of glass slides. Each laboratory defines its own strategy for labeling glass slides.

Tissue biobank specimens and glass slides are regularly collected and preserved for long-term periods. In the MS-BB repository, a huge number of glass slides were being stored without

barcode labels (the glass slides were only identified with handwritten labels). For long-term preservation and effective handling, it was decided that the labeling strategy for the current glass slides kept in the repository (approximately 4,000 glass slides) should be defined.

First, requirements for the labels were collected and analyzed with stakeholder input. Additionally, the recommendations of the current guidelines for uniform labeling of glass slides were considered [37]. Table 4.1 presents the questions and corresponding requirements for the uniform labeling of glass slides. This information is categorized into two groups — content and technical features.

Table 4.1. Requirements for uniform labeling of glass slides divided into two subgroups — content and technical details.

<b>Questions and appropriate requirements for uniform labeling</b>	
<b>Content</b>	<p>Q1. What elements should be placed on the label?</p> <ul style="list-style-type: none"> <li>Barcode, staining type, Sample ID, date of autopsy, project name, and institution name. All components except the barcode are represented as text.</li> </ul>
	<p>Q2. What is (are) the unique identifier(s) of the glass slides?</p> <ul style="list-style-type: none"> <li>A 32-digit DataMatrix barcode uniquely identifies each glass slide. Additionally, human-readable Sample ID and the staining type identify the glass slide.</li> </ul>
	<p>Q3. How should the elements be placed and prioritized on the label?</p> <ul style="list-style-type: none"> <li>Barcode, staining type, and Sample ID are the highest priority components. They should appear prominently with a large font size.</li> <li>Barcodes should be printed in the middle of the label in a size of at least 5 mm × 5 mm so that they can be easily recognized by the microscopic scanner.</li> <li>Date of autopsy, project name (MS-BB), and institution name (UMG Neuropathologie) have to be placed near the margin of the label as non-priority components. They are printed with relatively a smaller font size.</li> </ul>
<b>Technical details</b>	<p>Q4. What type of label material is preferred for the glass slides?</p> <ul style="list-style-type: none"> <li>Chemical- and solvent-resistant labels that are suitable for long-term storage should be used.</li> </ul>
	<p>Q5. What type of mechanism should be used to export data from the LIMS?</p> <ul style="list-style-type: none"> <li>The LIMS has to be customized to export all components in the standard template in a CSV/XLS format that is compatible with the label printing system.</li> </ul>
	<p>Q6. What types of printing techniques should be used for the labels?</p> <ul style="list-style-type: none"> <li>Durable labels should be printed via a thermal printer (used regularly).</li> <li>Barcode label printing software should be used to design the label and manage the printing process (used only once to design a uniform label).</li> </ul>

4.1.4.3 Interface for the identification of glass slides and label data generation

Based on the requirements for the identification mechanism and labeling techniques as described above, a new interface was implemented in the laboratory management application

with the help of a biobank interface technician. Using the interface, a user can indicate a specific brain and generate a list of data that contains label components ready for the label printing service. The interface provides the opportunity to export the required features of selected specimens (or brains) into the CSV or Microsoft Excel spreadsheet (XLS) files. Additionally, it also ensures changes at the database level by storing the generated barcode information. Figure 4.7 illustrates the interface that exports labeling data for multiple glass slides into the CSV/XLS format. The exported data contains the following components of each glass slide: Sample ID, Glass ID (for the barcode of slide), staining type, date of autopsy, project name, and organization (Table 4.2).

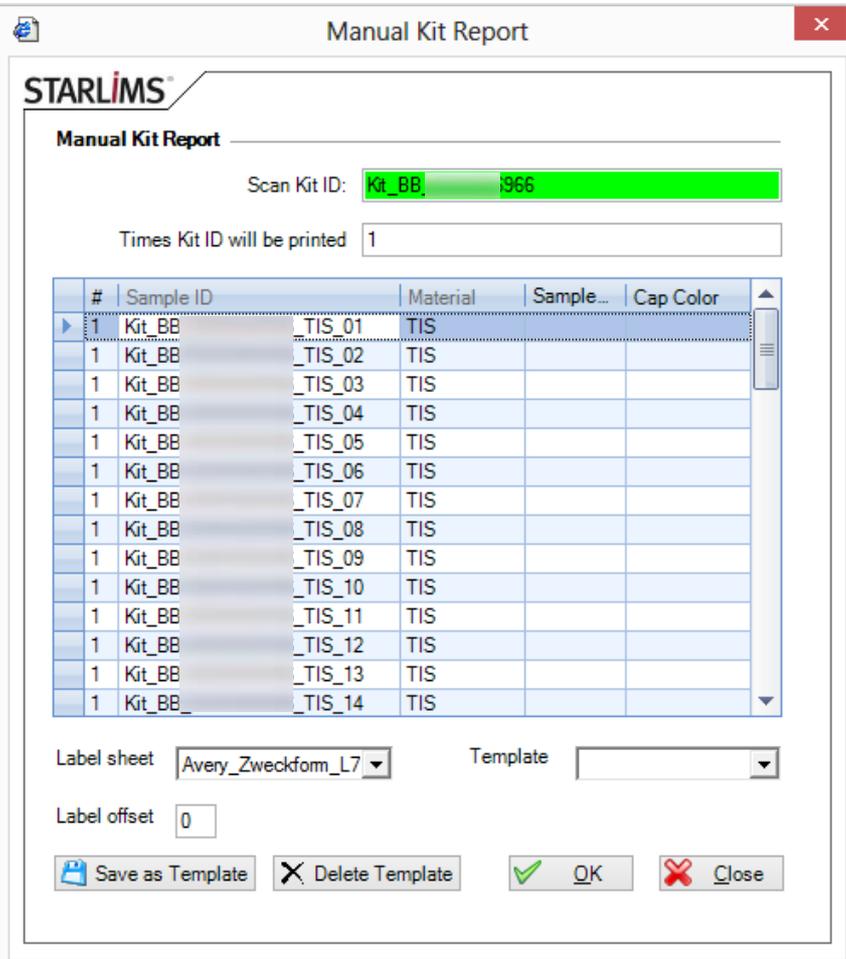


Figure 4.7. Interface for identification of specimen-derived assets. Actor activates the “kit reports” module and enters the Kit ID (Brain ID) in STARLIMS; The STARLIMS system generates unique IDs and label components (Sample ID, Glass ID, staining, project name, and institution name) for all specimens corresponding to the selected Case (brain) ID; Reporting system saves data as an XLS/CSV file that is compatible with the label printing software.

Table 4.2. Extract from an XLS exported file. It contains the following information regarding the glass slides: Sample ID, Glass ID, staining type, date of autopsy, project name, institution name.

Sample_id	Glass_id for barcode	Staining type	Date of autopsy	Project name	Organisation
Kit_BB_0123456789_TIS_01	Kit_BB_0123456789_TIS_01_IMG_01	HE	05/13	MS-BB	UMG Neuropathologie
Kit_BB_0123456789_TIS_01	Kit_BB_0123456789_TIS_01_IMG_02	LFB-PAS	05/13	MS-BB	UMG Neuropathologie
Kit_BB_0123456789_TIS_01	Kit_BB_0123456789_TIS_01_IMG_03	Biel	05/13	MS-BB	UMG Neuropathologie
Kit_BB_0123456789_TIS_01	Kit_BB_0123456789_TIS_01_IMG_04	KiM1P	05/13	MS-BB	UMG Neuropathologie
Kit_BB_0123456789_TIS_01	Kit_BB_0123456789_TIS_01_IMG_05	PLP	05/13	MS-BB	UMG Neuropathologie
Kit_BB_0123456789_TIS_02	Kit_BB_0123456789_TIS_02_IMG_01	HE	05/13	MS-BB	UMG Neuropathologie
Kit_BB_0123456789_TIS_02	Kit_BB_0123456789_TIS_02_IMG_02	LFB-PAS	05/13	MS-BB	UMG Neuropathologie
Kit_BB_0123456789_TIS_02	Kit_BB_0123456789_TIS_02_IMG_03	Biel	05/13	MS-BB	UMG Neuropathologie
Kit_BB_0123456789_TIS_02	Kit_BB_0123456789_TIS_02_IMG_04	KiM1P	05/13	MS-BB	UMG Neuropathologie
Kit_BB_0123456789_TIS_02	Kit_BB_0123456789_TIS_02_IMG_05	PLP	05/13	MS-BB	UMG Neuropathologie

#### 4.1.4.4 Barcode labels and printing options

The format of the label is designed according to the guidelines for uniform labeling of slides. Xylene- and solvent-resistant durable labels provided by GA International were chosen for the MS repository glass slides. These barcode labels are ideal for long-term storage and bio-reservation. By importing the XLS file, a barcode label printer managed by the BarTender software prints the corresponding data on the labels. All components imported from XLS file (Table 4.2) are placed on the label as textual data, with the exception of the Glass ID (string data), which is converted into a corresponding DataMatrix barcode printed in the middle of the label. The 2D DataMatrix barcode uniquely identifies the corresponding glass slide and specimen which it is derived from. The design of the glass slide label is illustrated in Figure 4.8.



Figure 4.8. Design of the barcode label for glass slides. It contains the following components: Sample ID at the top, barcode in the middle that corresponds to the Glass ID, staining type beneath the barcode, project name at the left margin, institution name at the bottom margin, and date of autopsy at the right margin.

There are several advantages of the described labeling concept:

1. Labeling with two identifiers — barcode of glass slide on the one hand, and Sample ID and staining type on the other — can reduce errors during the tracking process.
2. The human-readable Sample ID, written at the top of the label, and the staining type convey the most important information for pathologists.

3. The long barcode can be split into several components and added into the slide properties automatically during the scanning process in the Olympus VS120 software.
4. The project name and asset owner organization is beneficial information when identifying the origin of the glass slide.
5. The date of autopsy facilitates gathering of materials according to specific requests in the MS-BB biorepository. Furthermore, it provides pathologists with more specific information regarding the specimen.
6. Developing a uniform standard for labeling a large number of glass slides is a vital component to the organizational process and achieving handling efficiency in the biorepository.
7. Components of the label that are designed in an effective format assist with accurate reading of the labels' contents.

There are two kinds of glass slides in the biorepository of the MS-BB:

- a. **Histopathologically processed glass slides already preserved in the MS biorepository:**  
More than 4,000 glass slides had already been processed and stored in the repository for further investigation. All these glass slides had handwritten labels. Therefore, these slides required a label printed using a thermal transfer printer. Most of the glass slides of the biorepository were processed in the last decade before setting up the complex IT infrastructure for the KKNMS research and preceding the adoption of the data protection framework for research purposes. Therefore, the glass slides of the repository contain patient information (such as Patient ID) handwritten on the label area. From an ethical context, personal information needed to be removed from the glass slides before the new labels were attached to them. Therefore, before attaching the new labels, any personal information already written on the slides was covered using the black color marker. Using a predesigned label template and a CSV-exported file containing the labeling data, the printing system printed data on the labels that were placed on the individual ribbon corresponding to each brain. The labels were ordered in increasing order by sample ID. For each sample, five labels were printed and ordered by suffix of the Glass ID: IMG\_001, IMG\_002, IMG\_003, IMG\_004, IMG\_005. In total, 4,375 labels were printed via a thermal transfer printer and manually attached to the glass slides with the help of a research assistant. Figure 4.9 depicts the result of the attaching process for a single specimen.

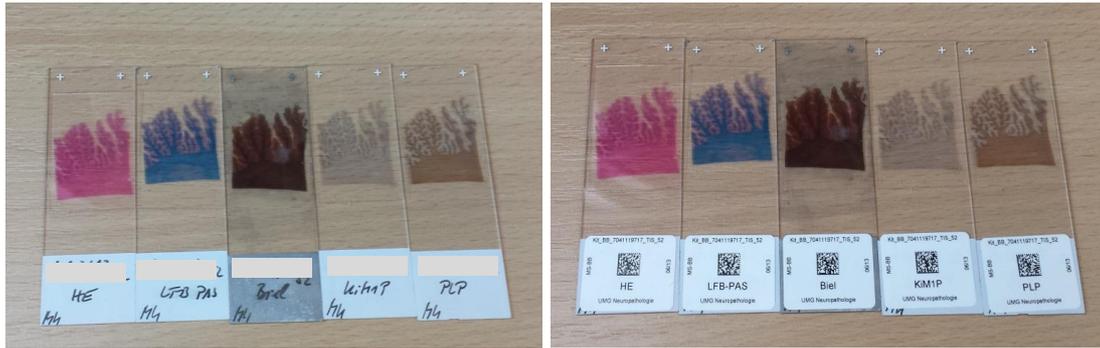


Figure 4.9. Attaching the labels on the glass slides. The left picture depicts glass slides with handwritten labels. Five glass slides are selected for each specimen with regard to the standard staining types; Patient ID is written on the top of the label is hidden with the marker; printed labels are carefully attached on the glass slides (the right-side picture).

- b. **New empty glass slides for new cases in the future:** To facilitate the labeling technique and enhance the workflow, the barcode labels might be printed directly onto the label of the glass slide using a slide printer (such as the SlideMate slide printer that was already being used in the pathology laboratories at UMG) that eliminates the need for manually attaching the labels onto the slides. However, the slide printer only works for empty glass slides intended for new histological cases.

Although only the thermal transfer printer was used for existing glass slides, it is highly recommended to define uniform labeling techniques for both new and already processed slides. So, Figure 4.10 schematically illustrates the workflow for the uniform labeling for both kinds of glass slides. On the one hand, the thermal transfer printer is used for already processed glass slides. On the other hand, the SlideMate printer will be being used for the new glass slides in the future.

Processes for preparing the glass slides can be also effectively displayed using the UML activity diagram in Figure 4.11. The activity diagram is represented for the existing glass slides preserved in the biorepository. If various label printing systems are likely to be used for the MS-BB in the future (e.g., SlideMate printer), then the workflow will have to be slightly modified according to the workflow displayed in Figure 4.10. Textual specifications for the same scenario are represented in Table B.1 (shown in Appendix B).

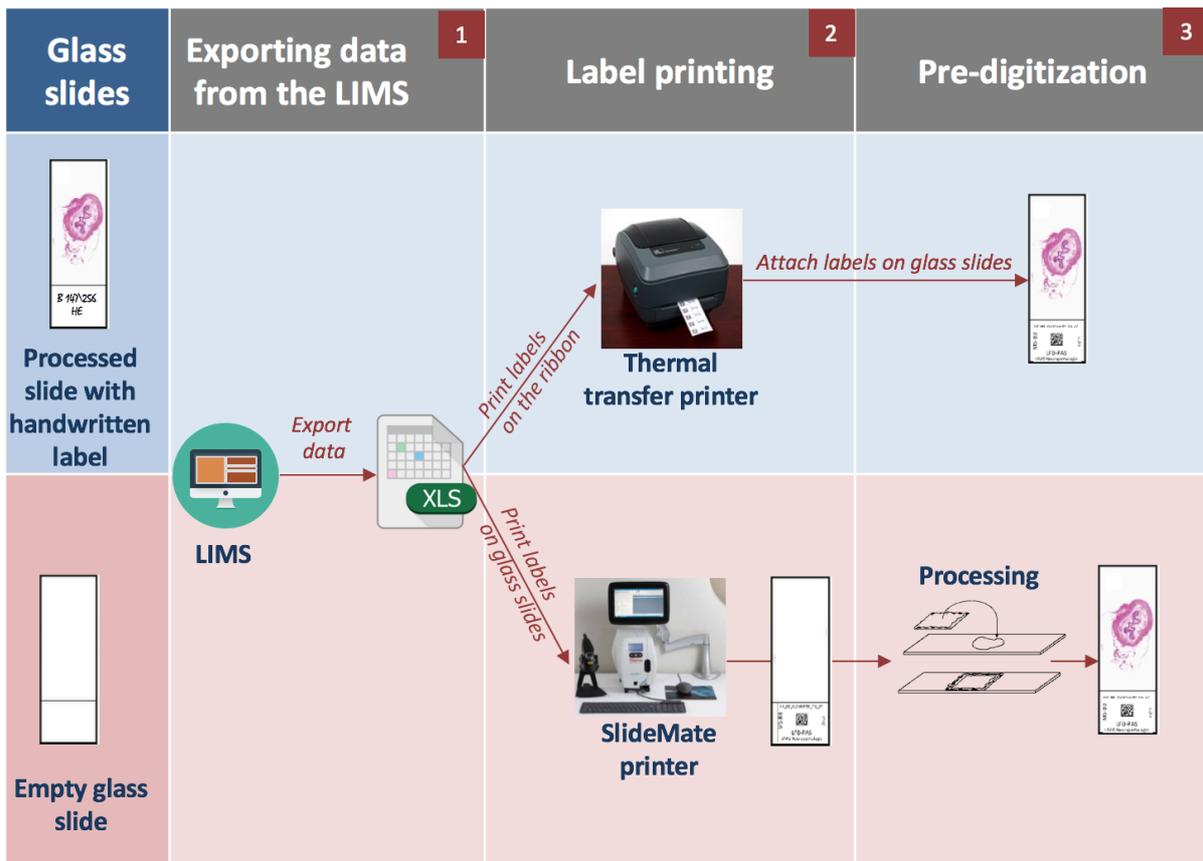


Figure 4.10. Uniform labeling workflow for two types of glass slides. The first row (colored in blue) describes the workflow for already processed glass slides, whereas the second row (in pink) represents the workflow for empty glass slides. There are three steps in the workflow (respectively three columns). (1) Exporting data from the LIMS: labeling data is exported from STARLIMS as an XLS/CSV file; (2) Label printing: the exported file is imported into label printing software; for processed glass slides, label data is printed on the label material, and the printing service is managed by a thermal transfer printer; empty glass slides have the label data directly printed on the label area of the glass slide via a SlideMate printer; (3) Pre-digitization: labels printed for first type of glass slides are manually attached on the slides; glass slides labeled via the SlideMate printer are histopathologically processed; afterwards, both types of glass slides are ready for digitization.

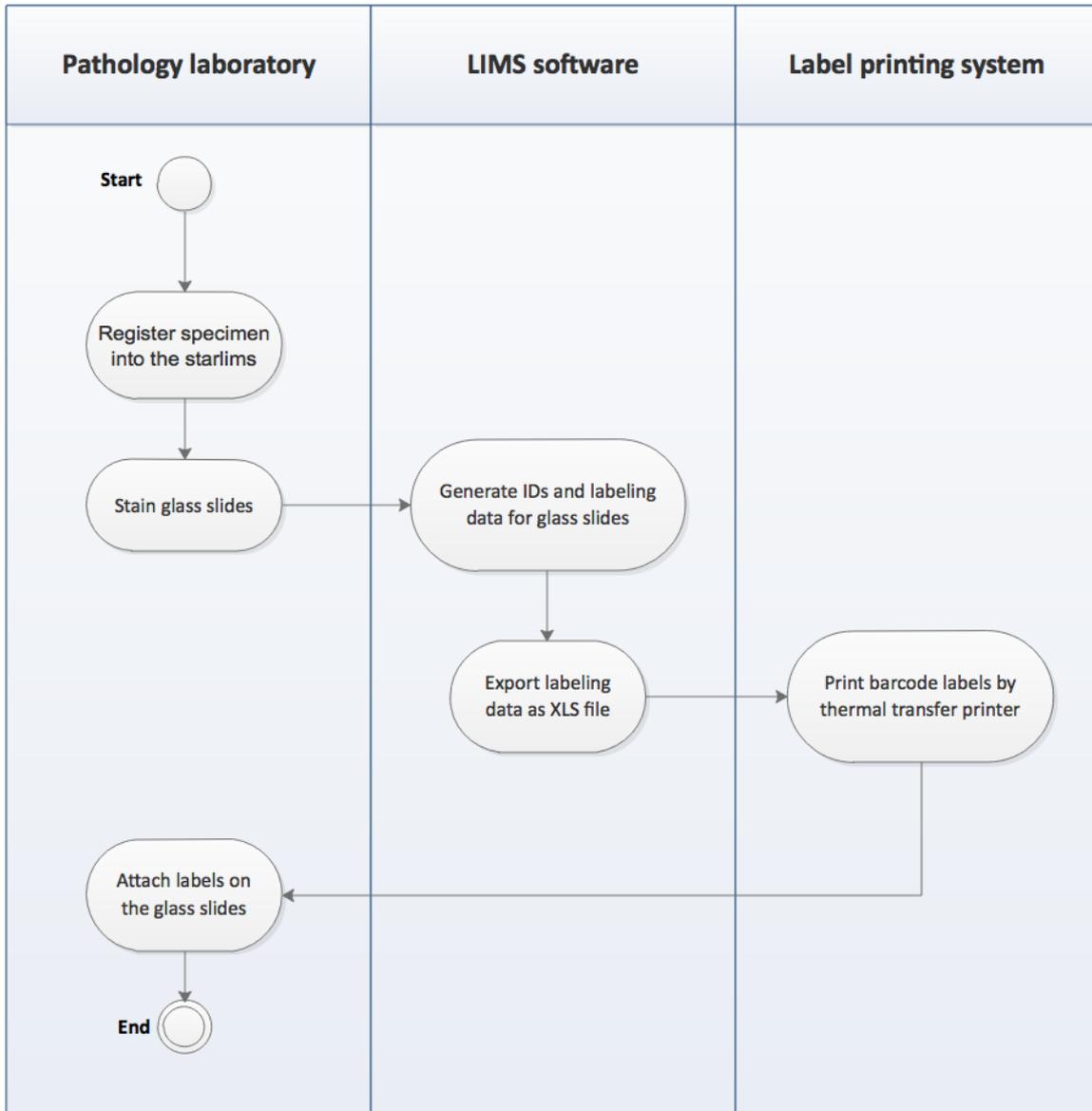


Figure 4.11. UML activity diagram for the preparation of glass slides

## **4.2 Digitization of glass slides**

Digitization of glass slides is the most crucial stage of image processing. Once the identification and barcode labeling mechanism for the glass slides was determined, the digitization strategy was defined.

### **4.2.1 Digitization strategy for sample selection**

With an effective digitization strategy, a large number of glass slides in the biorepository can be scanned in an organized way. The scanning strategy was determined through discussions and input from responsible stakeholders. At that time of strategy determination, there were tissue samples of 14 brain donors already processed and registered into the LIMS. On average, 60-80 samples are cut from each brain. At least five glass slides are kept in the biorepository for each tissue sample. In total, approximately 5,000 glass slides being in the biorepository need to be scanned, stored, viewed, and linked to the corresponding specimen into the LIMS.

To start with, all samples for the two brains considered as the most interesting cases were selected for scanning. In total, there were up to 160 specimens. For the initial scanning step, five glass slides stained using the various staining methods were selected for each specimen. In total, 800 glass slides that should be digitized using the WSI scanner were counted. Since the scanning process is evaluated, the remaining glass slides in the biorepository will be scanned according to the same selection method.

As biobanking is an ongoing process dedicated for the long-term, new cases are regularly added into the repository (about 3-4 brains each year). According to the digitization strategy, first, all existing and processed glass slides should be scanned and glass slides that are newly added in the biorepository should be digitized later.

### **4.2.2 Digital pathology workstation**

The DPS workstation, also called the cockpit, contains the Olympus VS120 microscopic scanner and the scanner PC with the acquisition software that manages the scanning process and provides a powerful toolset to acquire high-resolution images along with their associated metadata (Figure 4.12). The microscopic system supports a wide range of magnification (from 2x to 100x), batch scanning mode, and advanced metadata management tools. Although the Olympus VS120 generates high-resolution images in the proprietary file format “.vsi”, it offers

tools for batch conversion of images into standard image formats, such as TIFF, JPEG, and JPEG2000.

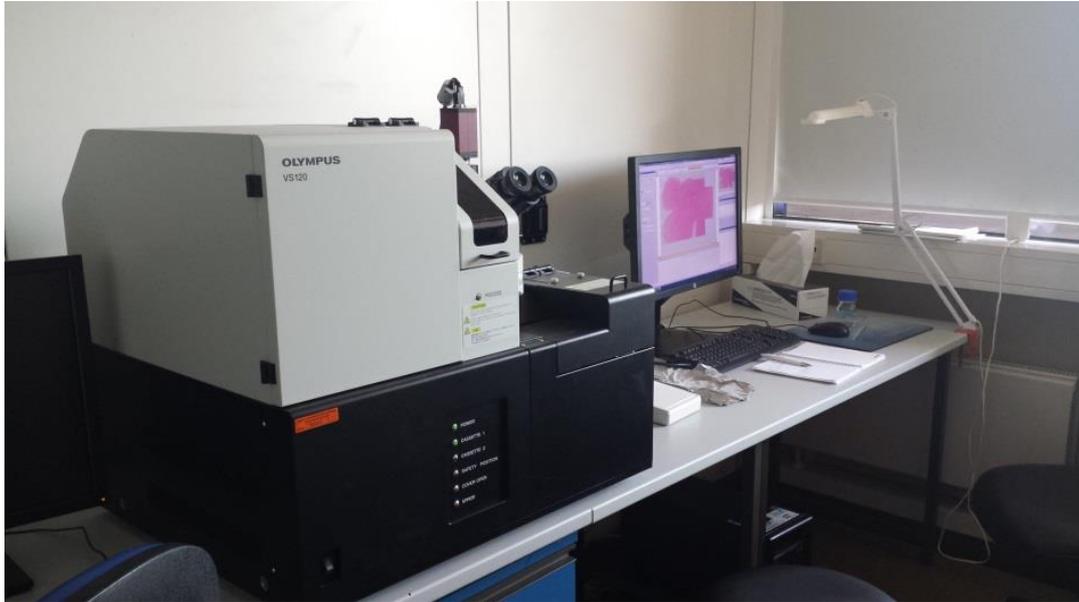


Figure 4.12. Olympus VS120 workstation at UMG

#### 4.2.3 Scanning settings

Before the glass slides are digitized, scanning settings should be determined and configured in the microscope slide scanner and tested specifically for the glass slides of the MS-BB repository. Scanning parameters should be customized according to the nature of the glass slides and staining types. More specifically, different glass slides may have different thicknesses up to 1mm. For example, a bone tissue can be thicker than a brain tissue on the glass slide that needs the z-stack scanning mode. Several glass slides were digitized to determine the best scanning options for the biorepository pathology materials. Here we enumerate the main scanning parameters that are characteristic of the MS-BB glass slides:

1. Scanning area of the glass slide (about 23 mm x 50 mm) should be calibrated so that the entire tissue on the glass slide is digitized. Also, label area of the glass slide (about 23 mm x 19 mm) needs to be calibrated so that the whole area of the label is captured.
2. As there are at least five types of staining, the following parameters are chosen for each:
  - a. Glass slides stained with HE, LFB-PAS, and Biel are considered normal-appearing samples. Accordingly, sample detection sensitivity is set to 0 (normal) for this kind of glass slide. Because the tissues on these glass slides are sharp and noticeable, the

scan area is decreased and only the tissue area of glass slide is scanned that is easily detected by the scanner.

- b. Glass slides stained with PLP are considered dim samples, as the staining color is grey and is difficult to successfully digitize using standard settings. Therefore, sample detection sensitivity is set to 5 (higher value). While it is difficult to notice sharp edges of tissue on the scanning area, the whole area of the glass slides is decided to be scanned. Focus map density is set to the high value.
  - c. Glass slides stained with the Kim1P are considered faint samples because the staining color is a very light grey. Correspondingly, sample detection sensitivity is set at 10 (highest value) and the focus map density is set at the extra high value. The complete area of the glass slide is selected for scanning.
3. Except the automated settings, there are exceptional cases that need to be configured manually during scanning. For example, if the tissue is smaller than a quarter of the whole area of the glass slide and the whole glass slide is automatically selected for scanning, the scanning area needs to be reduced from the whole area of the glass slide to the tissue area, thus accelerating the scanning process.
  4. Metadata of each glass slide should be entered along with the scanning process via the external CSV file. Thus, manual data entry for each slide is eliminated.
  5. Digital images should be stored directly on the server in a well-structured form. For this reason, image names and storage locations (corresponding folders) are specified automatically using metadata information. Storage location on the server is specified with metadata information that is described in the following subchapter in detail.
  6. Automated digitization techniques are pre-loaded in the software for rapid and effective use for situations that are likely to be used multiple times using the batch scanning mode.
  7. 20x magnification is selected for scanning all glass slides of the MS-BB repository. If there is an exceptional case that needs different magnification, it can be manually defined in the scanning settings.

The main scanning settings from the list are illustrated in Figure 4.13 and categorized for different staining types.

Before the final determination of scanning settings, multiple glass slides (about 30-40) were scanned using two various scanners. During the tests, the scanning settings and technical

details such as file size, scanning duration, batch scanning mode, barcode processing, and conversion tools were evaluated (technical details are presented in Appendix D). Afterwards, the final scanning settings were determined with the assistance of the main pathologist by evaluating different digital images.

All these final settings were selected in order to facilitate a rapid scanning process. In addition, using these settings, the scanner provides high quality digital images suited for the MS-BB.

Scanning Settings for standard staining types			
Settings	HE; LFB-PAS; Biel	PLP	Kim1P
Magnification	20x		
Sample detection sensitivity	0 (normal)	5 (higher)	10 (highest)
Scan area	Scan whole area and focus sample only 	Scan and focus whole area 	Scan and focus whole area 
Focus map density	Normal	High	Extra High
Camera settings	Normal samples	Dim samples	Faint samples

Figure 4.13. Scanning settings for the glass slides of the MS-BB repository specified in the acquisition software

**4.2.4 Batch scanning mode and metadata entry**

Batch scanning features let a pathologist digitize multiple glass slides without human intervention. As there is a large number of histology slides in the biorepository, 100 glass slides held in two cassettes are likely to be scanned in one batch. More specifically, by choosing five slides for each specimen, the glass slides of 20 specimens can be scanned in one batch scanning process. While acquiring digital images, a barcode for the glass slide is automatically read by a 2D barcode scanner integrated in the image acquisition system. If a barcode is located in the middle of the label, it is more convenient for system to recognize it. Reading permission for the DataMatrix ECC000-140 was set up on the barcode reader of the scanner

so that it could recognize a barcode and register for a long ID (more than 30 letters) within the slide properties.

Regardless of the acquisition of digital images, metadata needs to be entered for each slide. To facilitate data entry, metadata are assigned to the corresponding slides using CSV importing tools. The CSV metadata file contains multiple rows corresponding to the values of slide properties of multiple slides (one row corresponds to a single slide). The structure of the template file is likely to be designed once and used in multiple cases for metadata entry. For the digital images of the MS-BB repository, the following metadata is required to be stored along with digital images: slide name, Glass ID, slide information, location of lesion, Sample ID, staining type, date of autopsy, project name, and institution name. The last two fields, project name and institution name, are constant for all slides, while the other features are variable according to the glass slides. The slideName column is the same as the barcode of the corresponding glass slide. Table 4.3 shows an example of a CSV metadata file that is filled with the corresponding information. The metadata file can be filled in two ways: manually or automatically using the metadata generator template (in the XLS document that later is converted into CSV file). The user can enter only four parameters (Case ID, initial and ending specimen index, and date of autopsy) in the generator template. Based on that, metadata table for multiple glass slides will be generated. The automated method is useful when a user wants to fill the metadata file for multiple numbers of slides. The metadata file with accompanying generator template is represented in Appendix E (Figure E.1 and Figure E.2).

Table 4.3. Examples of metadata for multiple glass slides. The list contains the following information about each slide: Slide name same as a barcode, slide info, location of specimen, folder name to create folder on the server automatically, staining type, Sample ID, date of autopsy, institution name, and project name.

TrayNo	SlideNo	SlideName	SlideInfo	Specimen	FolderName	Staining	SampleID	DateofAutopsy	Institution	ProjectName
1	1	Kit_BB_1234567890_TIS_01_IMG_001	1234567890_TIS_01_HE	Gyrus frontalis medialis L	1234567890_TIS_01	HE	Kit_BB_1234567890_TIS_01	01/13	UMGNeuropathologie	MS-BB
1	2	Kit_BB_1234567890_TIS_01_IMG_002	1234567890_TIS_01_LFB-PAS	Gyrus frontalis medialis L	1234567890_TIS_01	LFB-PAS	Kit_BB_1234567890_TIS_01	01/13	UMGNeuropathologie	MS-BB
1	3	Kit_BB_1234567890_TIS_01_IMG_003	1234567890_TIS_01_Biel	Gyrus frontalis medialis L	1234567890_TIS_01	Biel	Kit_BB_1234567890_TIS_01	01/13	UMGNeuropathologie	MS-BB
1	4	Kit_BB_1234567890_TIS_01_IMG_004	1234567890_TIS_01_KIM1P	Gyrus frontalis medialis L	1234567890_TIS_01	KIM1P	Kit_BB_1234567890_TIS_01	01/13	UMGNeuropathologie	MS-BB
1	5	Kit_BB_1234567890_TIS_01_IMG_005	1234567890_TIS_01_PLP	Gyrus frontalis medialis L	1234567890_TIS_01	PLP	Kit_BB_1234567890_TIS_01	01/13	UMGNeuropathologie	MS-BB
1	6	Kit_BB_1234567890_TIS_02_IMG_001	1234567890_TIS_02_HE	Gyrus frontalis medialis R	1234567890_TIS_02	HE	Kit_BB_1234567890_TIS_02	01/13	UMGNeuropathologie	MS-BB
1	7	Kit_BB_1234567890_TIS_02_IMG_002	1234567890_TIS_02_LFB-PAS	Gyrus frontalis medialis R	1234567890_TIS_02	LFB-PAS	Kit_BB_1234567890_TIS_02	01/13	UMGNeuropathologie	MS-BB
1	8	Kit_BB_1234567890_TIS_02_IMG_003	1234567890_TIS_02_Biel	Gyrus frontalis medialis R	1234567890_TIS_02	Biel	Kit_BB_1234567890_TIS_02	01/13	UMGNeuropathologie	MS-BB
1	9	Kit_BB_1234567890_TIS_02_IMG_004	1234567890_TIS_02_KIM1P	Gyrus frontalis medialis R	1234567890_TIS_02	KIM1P	Kit_BB_1234567890_TIS_02	01/13	UMGNeuropathologie	MS-BB
1	10	Kit_BB_1234567890_TIS_02_IMG_005	1234567890_TIS_02_PLP	Gyrus frontalis medialis R	1234567890_TIS_02	PLP	Kit_BB_1234567890_TIS_02	01/13	UMGNeuropathologie	MS-BB
1	11	Kit_BB_1234567890_TIS_03_IMG_001	1234567890_TIS_03_HE	Gyrus cinguli L	1234567890_TIS_03	HE	Kit_BB_1234567890_TIS_03	01/13	UMGNeuropathologie	MS-BB
1	12	Kit_BB_1234567890_TIS_03_IMG_002	1234567890_TIS_03_LFB-PAS	Gyrus cinguli L	1234567890_TIS_03	LFB-PAS	Kit_BB_1234567890_TIS_03	01/13	UMGNeuropathologie	MS-BB
1	13	Kit_BB_1234567890_TIS_03_IMG_003	1234567890_TIS_03_Biel	Gyrus cinguli L	1234567890_TIS_03	Biel	Kit_BB_1234567890_TIS_03	01/13	UMGNeuropathologie	MS-BB
1	14	Kit_BB_1234567890_TIS_03_IMG_004	1234567890_TIS_03_KIM1P	Gyrus cinguli L	1234567890_TIS_03	KIM1P	Kit_BB_1234567890_TIS_03	01/13	UMGNeuropathologie	MS-BB
1	15	Kit_BB_1234567890_TIS_03_IMG_005	1234567890_TIS_03_PLP	Gyrus cinguli L	1234567890_TIS_03	PLP	Kit_BB_1234567890_TIS_03	01/13	UMGNeuropathologie	MS-BB
1	16	Kit_BB_1234567890_TIS_04_IMG_001	1234567890_TIS_04_HE	Gyrus cinguli R	1234567890_TIS_04	HE	Kit_BB_1234567890_TIS_04	01/13	UMGNeuropathologie	MS-BB
1	17	Kit_BB_1234567890_TIS_04_IMG_002	1234567890_TIS_04_LFB-PAS	Gyrus cinguli R	1234567890_TIS_04	LFB-PAS	Kit_BB_1234567890_TIS_04	01/13	UMGNeuropathologie	MS-BB
1	18	Kit_BB_1234567890_TIS_04_IMG_003	1234567890_TIS_04_Biel	Gyrus cinguli R	1234567890_TIS_04	Biel	Kit_BB_1234567890_TIS_04	01/13	UMGNeuropathologie	MS-BB
1	19	Kit_BB_1234567890_TIS_04_IMG_004	1234567890_TIS_04_KIM1P	Gyrus cinguli R	1234567890_TIS_04	KIM1P	Kit_BB_1234567890_TIS_04	01/13	UMGNeuropathologie	MS-BB
1	20	Kit_BB_1234567890_TIS_04_IMG_005	1234567890_TIS_04_PLP	Gyrus cinguli R	1234567890_TIS_04	PLP	Kit_BB_1234567890_TIS_04	01/13	UMGNeuropathologie	MS-BB
1	21	Kit_BB_1234567890_TIS_05_IMG_001	1234567890_TIS_05_HE	Gyrus parietalis inferior L	1234567890_TIS_05	HE	Kit_BB_1234567890_TIS_05	01/13	UMGNeuropathologie	MS-BB
1	22	Kit_BB_1234567890_TIS_05_IMG_002	1234567890_TIS_05_LFB-PAS	Gyrus parietalis inferior L	1234567890_TIS_05	LFB-PAS	Kit_BB_1234567890_TIS_05	01/13	UMGNeuropathologie	MS-BB

IDs of the glass slides can be entered using two different sources. On one hand, a barcode is recognized by the acquisition scanner and is assigned to the relevant field in the slide property called "Barcode." On the other hand, IDs of the glass slides (elements of the column

SlideName from the CSV file) are assigned in succession to the corresponding glass slides (same element of the tray), and they are automatically registered in the field "SlideName." If both values, barcode and SlideName, are the same, it means that the metadata is correctly assigned to the corresponding digital image. Thus, we can avoid erroneous linkages between metadata and digital images. In addition, at the end of scanning, it is important to analyze the log file that describes the results of batch scanning process for each slide. By analyzing log file, several errors that can occur during scanning can be found. For example, errors could occur if a glass slide is not scanned properly or if a barcode is not recognized by the scanner.

Using the metadata entry and batch scanning mode, the scanning duration is significantly reduced and human errors in assigning metadata to the corresponding slide are eliminated.

#### **4.2.5 Automatic naming and saving of images**

A pathologist or research assistant who is going to scan glass slides specifies the structure of the image name using three placeholders: <SlideName>\_<Staining>\_<Magnification>. The first two parameters are generated from the CSV metadata file, while the third parameter, magnification, is addressed automatically through the scanning settings. Once the structure of the image name is defined, the software automatically generates an image name for all slides that are going to be scanned. In addition, the pathologist will specify the destination location on the image server as follows: digital images should be stored in a subfolder named with a corresponding Sample ID. This subfolder is automatically created on the server when the corresponding images are generated by the scanner. For automatic saving, the user specifies a predefined placeholder <Folder Name> for the names of subfolders, which is extracted from the CSV metadata file by the acquisition software. Therefore, multiple subfolders will be created automatically on the server. The structure of the subfolders stored on the server is described in detail in Subchapter 4.3.2.

#### **4.2.6 Scanning results**

According to our scanning strategy, during the initial phase, 800 glass slides coming from two brains (designated as the most interesting cases of biorepository) were digitized using the VS120 scanner. Once the entire glass slides were scanned and stored on the server using a proprietary file format, technical features and scanning results were estimated. A total amount of data volume for 800 digital images comes to 1.8TB. So, an average file size is about

2.3GB. The duration of scanning for 100 glass slides as a batch process was about 15 hours. According to a rough calculation, the average scanning time for each slide is approximately 9 minutes (scanned with 20x magnification). Results of the scanning process are calculated in Table 4.4.

Table 4.4. Scanning results

Results of scanning process	
Number of brains	2
Number of blocks	162
Number of glass slides(digital images)	800
Missing glass slides	10
Data volume (for 800 slides)	1.8TB
Average file size	2.3GB
Duration of scanning (for 100 slides)	15 h
Image error	19
Barcode error	1

Because the metadata is entered from an external file and automatically assigned to the digital image according to their turn in the list, verification is required to determine whether the metadata was correctly linked to the corresponding image. For this reason, after the scanning process, the digital image barcode was compared to the slide name for all glass slides (both values should be the same). This required a couple of minutes in the acquisition software to compare these two values for 100 images after they were scanned. As a result, there was no erroneous linkage between metadata and digital images. However, there was only one case when the barcode value was empty as a result of bad quality barcode icon that was not recognizable by the scanner. The glass slide barcode icon was repaired and the barcode value was manually entered for the corresponding digital slide in the slide properties.

Image quality has a crucial role in the successful implementation of image applications. All digital images were viewed using the imaging software and the quality of each image was estimated manually. During the assessments of image quality, two kinds of image distortion were detected. Due to the scanning mistakes, in total, 19 digital images were damaged to the extent that could not be repaired. Different types of damaged images are illustrated in Figure 4.14. All these corrupted images were rescanned according to the same scanning settings without image damage.

To summarize the digitization step of the DP workflow, like the preparation stage, the processes for scanning the glass slides as described above need to be effectively characterized using use case specification from the actors' perspectives and are succinctly represented in Table B.2 in Appendix B.

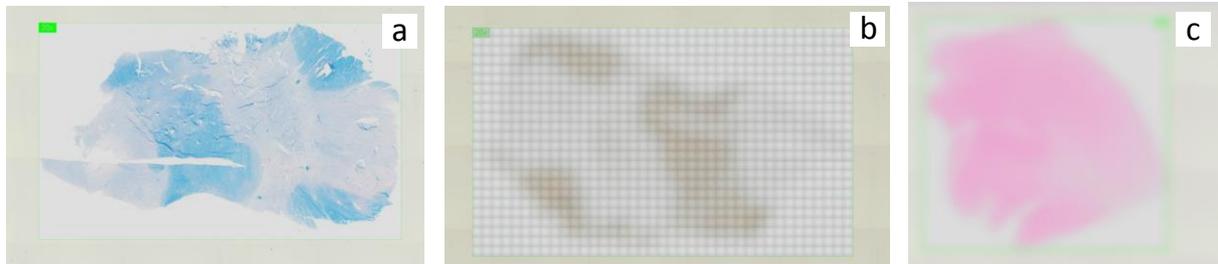


Figure 4.14. Types of the damage image revealed after the scanning process: (a) normal image; (b) scanning area is split down with the multiple blur rectangles; (c) digital image is blurry due to incorrect focus.

### 4.3 Image server and database

As digital microscopic images are quite big, a separate server with high throughput is required to handle them appropriately. For the MS-BB project, different kinds of storage environments were evaluated. As a result, an image server (i.e., NIS) was installed that meets the demands of the research project. All features regarding the image server and database for the biobank system are described below.

#### 4.3.1 Image server settings and network configuration parameters

Digital images need to be automatically stored on an image server that ensures versatile image access and provides a storage solution for digital virtual microscopy. Before the image server is established, the following storage requirements need to be assessed specifically for a biobank research project. The server should provide rapid access to the image and should automatically save images to the database. Another alternative storage solution could be an external hard disk that is may be used in urgent cases if the image server is not available. The image server should ensure that pathologists have web access to the digital images so they can view, navigate, and annotate the digital slide. Furthermore, the system needs to be accessible by multiple users from both internal and remote places.

According to these requirements, the image server for this case was established and configured within the UMG. As a result, the OLYMPUS Net Image Server-SQL (NIS-SQL) v.2.9 and Microsoft SQL 2008 Express database were installed on the PC server within the UMG.

The supporting operational system was Microsoft Server 2012 R2, 64bit. An MS SQL database and the NIS-service are running on the same hardware. The Microsoft Internet Information Service (IIS) was required to be installed on the same server in order to set up the web interface for access to the digital images.

Net Image Server was installed in the Demilitarized Zone (DMZ) accessible from both the Local Area Network (LAN) and the internet via a viewer software or web browser. In order to secure the image server from attacks and allow access to the NIS-SQL database for both internal and external communications, a firewall system was configured accordingly via network address translation. The imaging server is available for internal users only from predefined IP addresses via the image viewing software installed on the pathologist's workstation. In addition, the image server is accessible via a web viewer for external users with credentials. At least a 500 Mbit LAN connection is required for the VS120 scanner and image server to store digital images without delay. The installed NIS supports SOAP-based web services technologies that are crucial components for data exchange with the external system. The NIS can handle images in either the proprietary file format or open file formats, such as JPEG2000 or TIFF. In addition, Microsoft Word, Excel, or PowerPoint documents can be stored on the server.

There are various ways to design a network between various systems. Although the DPS server and the LIMS are both running at the same medical center, they were set up and configured on separate servers (i.e., virtual machines). Also, for security reasons, these systems should be running in a DMZ that provides an isolated and secured network with access from both inside and outside the LAN. The internal firewall should be configured in such a way that provides server-to-server communication for data exchange using web services. The implementation of the DPS into a LIMS requires further description from the perspective of the network communication that defines more specifications about two servers wherein these systems are running. Within the consideration of security issues, all details regarding the servers such as ports, IP addresses, firewall settings and other valuable information are not revealed here.

Communication between the image server, DPS components, and the LIMS server is depicted in Figure 4.15. This setup was designed and configured with support of the IT department at UMG.

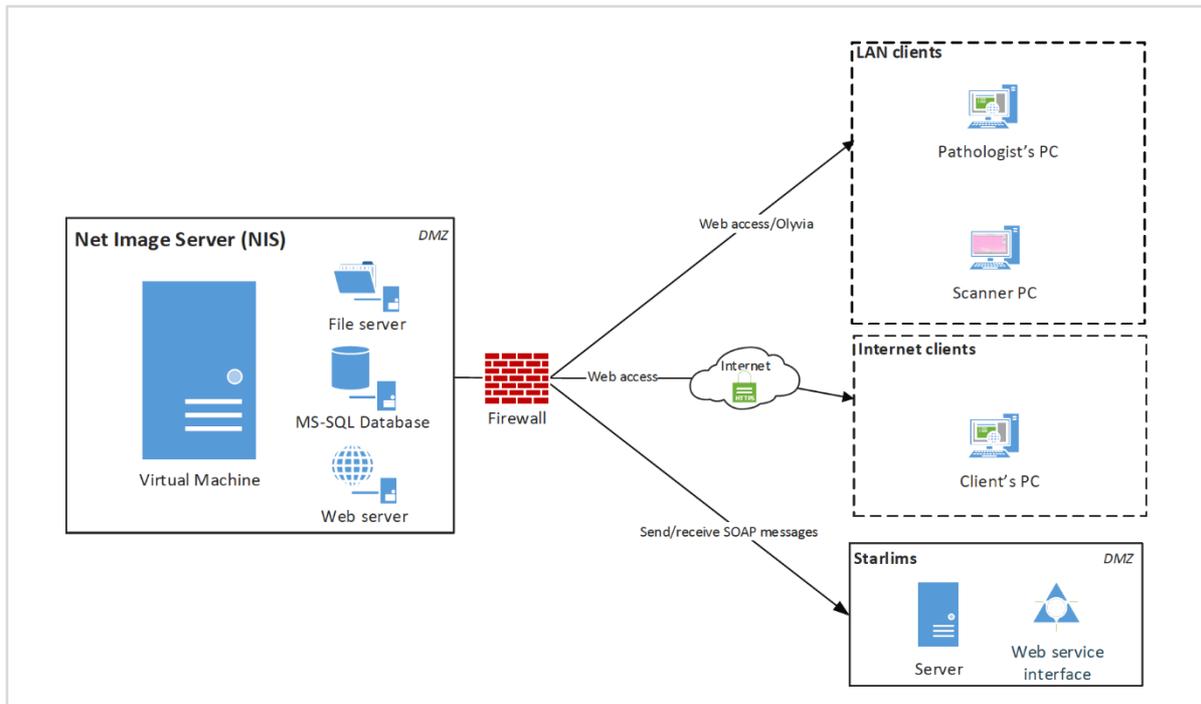


Figure 4.15. Communication scheme between the image server and the other components of the DPS. The NIS is installed on the virtual machine. Digital images stored on the NIS are available either for internal users via LAN or external authorized users (KKNMS researchers) via the Internet using a web viewer. The web interface connects to the image server via webaccess using an https secure connection. The scanner workstation located at UMG is connected to the NIS via LAN. Firewall between NIS and other components are configured according to need to provide secure access to the server. Using the firewall, the image server PC is secured from attack. In addition, incoming and outgoing network traffic is controlled. The NIS is also connected to another application – STARLIMS (details are described in Chapter 5).

### 4.3.2 Database and structure of image folders

Using the scanner software, digital images are likely to be automatically stored on the server once they are scanned. However, further configuration is needed to save the files in an organized way.

A separate database is dedicated for the MS-BB project on the NIS. Once the database is created, the basic structure of the image folders is defined. In the main subfolder there is a Case Folder for each histological case (brain). In each Case Folder there should be multiple subfolders that correspond to the proper specimens. In each subfolder are digital images of the corresponding specimen. Figure 4.16 illustrates the database structure described above:

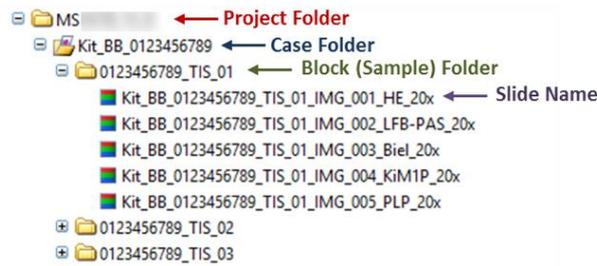


Figure 4.16. Folder hierarchy on the database. The main database folder (marked with the red arrow) is dedicated for the project. This folder contains case folders for each individual brain (marked with the blue arrow). In the case folder there are multiple subfolders (marked with the green arrow) that consist of digital images derived from the corresponding specimen (marked with the purple arrow). Digital image name consists of the Glass ID, staining type and magnification.

As there are multiple specimens and corresponding images that should be scanned, folders for each specimen must be created automatically in the corresponding Case Folder. Using automatic saving tools, the name of the subfolder is created with the predefined placeholders. The image metadata is stored along with the corresponding images.

The MS-BB database can be controlled either from the scanner workstation PC or directly on the server using database administration software. The administrator needs to be signed in the system via the SQL authentication procedure to access the MS-BB database.

To automate image saving on the server, the project administrator must define the folder structure and the folder naming convention on the server (department folder, case folder, block folder). As there are numerous folders and images, it is very important to maintain a well-organized tree structure to allow the administrator to effectively manage the large image dataset.

### 4.3.3 Data storage estimation for large volumes of growing data

Storage of a large volume of high-resolution microscopic images is one of the biggest challenges of DPSs. As the MS-BB is an ongoing research project, the data storage system must have the potential to expand storage capacity in response to the needs of the biorepository.

The term “big data” refers to large volumes of data that are not easy to manage using standard methodologies [130]. This term can be literally applied to pathology images with double meaning for a biobank application: because a biobank is a collection of data, it is already associated with a large volume of growing data, with a single data item taking up an extremely large number of storage space (average size of an image is 1.2GB).

In order to specify the adequate volume of disk space on the server for on-going research projects, storage capacity needs to be estimated for the next few years. Specifically, the data storage estimation for the MS-BB project has to be performed for both short-term and long-term periods within the scope of KKNMS3 (the third funding period from March of 2016 till September of 2019) and the next five years.

To define a total amount of storage resources as accurately as possible, different characteristics need to be assessed. For the MS-BB project, the following criteria were taken into account during the estimation of storage capacity:

1. Scanning settings (such as magnification of images, focus settings, etc.);
2. Average number of autopsy cases and glass slides per year;
3. Average file size based on the results of the scans performed in 2016.

Based on the storage estimation calculated in Figure 4.17, 10TB was provided, 1.8TB of which is already occupied by 800 digital images. According to the calculation, another 10TB is likely to be required for the next period between 2019 and 2024. Server storage capacity can be extended by an additional disk enclosure at any time according to the demand. When the existing volume is full, a second hard drive/network drive can be added to the server using data volume management tools.

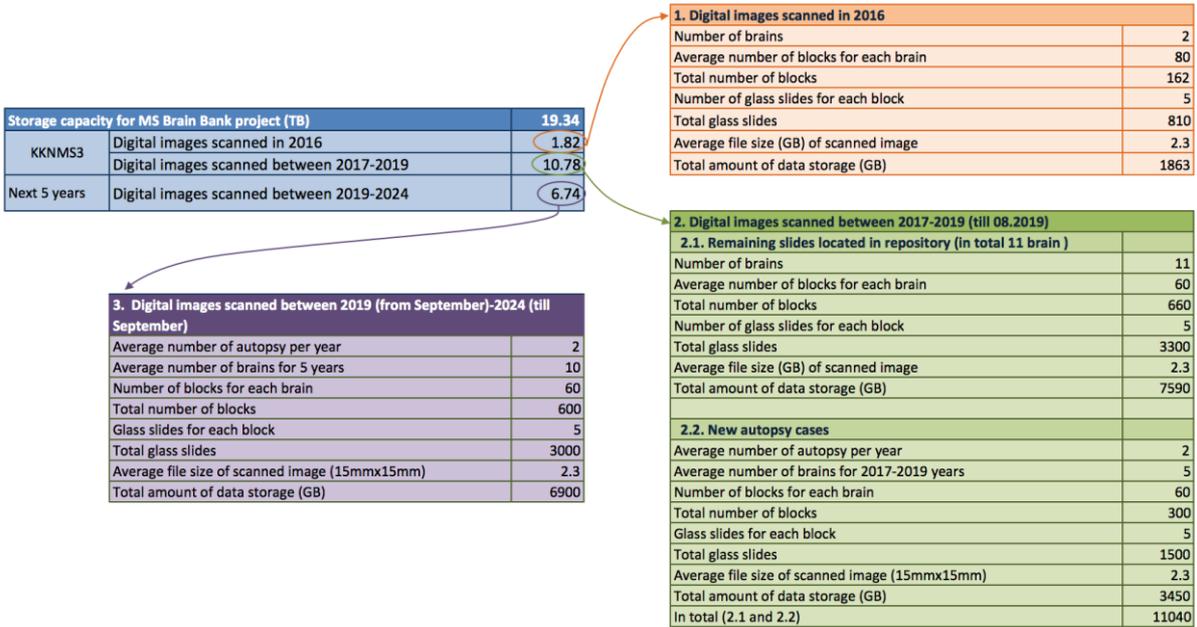


Figure 4.17. Estimation of Storage needs for the digital microscopic images of the MS Brain Bank project. Storage needs for the digital images scanned in 2016 is colored in orange; storage estimation for the digital images that will be scanned between 2017-2019 is colored in green; data capacity calculated for the digital images calculated for the next five years (2019-2024) is colored in purple; finally, all results are summarized in the table colored in blue.

#### 4.3.4 Overview of user management processes

DPS allows users to access specific digital images via different types of user accounts. Within the MS-BB research project, users should be differentiated according to their roles in the research network. As digital images are likely to be viewed by multiple researchers, separate user accounts need to be created. Therefore, a role-based access control (RBAC) mechanism should be used to manage access and permissions. In the MS-BB project, the following user types are applied:

**Administrator** – has full access to the image database and imaging software to manage images on the server and change the viewing settings in the web application. In addition, the administrator can manage user accounts and define the roles and rights.

**Power users** – can insert, update, and delete image records and modify database structure. However, they cannot manage user rights. Power users might be the pathologists or student assistants at UMG who are directly or indirectly involved in the project. The main responsibility of these users is to use their accounts for scanning, saving, and/or viewing images on the server.

**Guests** – can view images with accompanying metadata only via a web viewer, add annotations, but are not able to modify or insert images. This type of user account is assigned to KKNMS researchers who are external users within the system. The administrator should define the roles and permissions for specific guests. External users (i.e., KKNMS researchers) may have access to individual images (or folders) according to their needs. User permissions of the DPS are illustrated in Figure 4.18.

Permissions	Administrator	Power User	Guest
Read records (access to the specific folder)	✓	✓	✓
Add annotations	✓	✓	✓
Modify records	✓	✓	—
Create new user accounts & manage user rights	✓	—	—
Administer user groups	✓	—	—
Full access to the database	✓	—	—

Figure 4.18. Different types of user accounts and corresponding roles in the Olympus system used for the MS-BB project. The following three types of user accounts are used: (1) administrator account for the main pathologists who manage the project, (2) power user account for the users within the UMG, and (3) guest account for the remote KKNMS researchers.

Administrators can create user groups and assign users to the specific user group. By setting the permission, a user group (also individual user) can be assigned to the folder or its child records (folder or image) located on the database. Using this feature, users added in the user group have the same permissions to access specified image records. Thus, the user groups provide the opportunity to categorize users who might have the same access and rights to the database. Examples of the user groups are illustrated in Figure 4.19. Users can access the database to view and analyze images either from the Olympus acquisition software (i.e., VS-ASW), from OlyVIA installed on the pathologist’s PC within the UMG, or via the web viewer (for internal and external users).

As user roles and administration management are part of the DP workflow, a corresponding use case scenario (S3) is represented in Table B.3 (Appendix B).

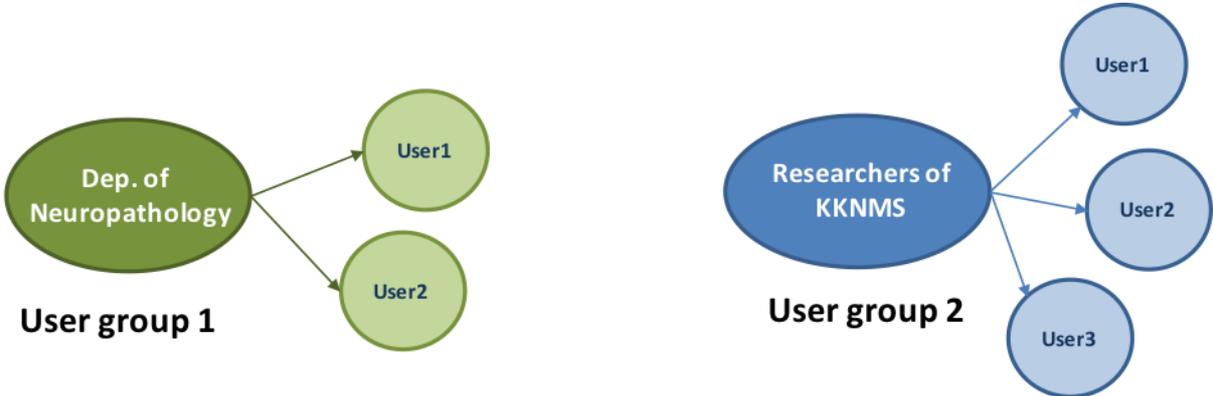


Figure 4.19. Example of user groups. When there are multiple researchers, who need the same permissions to use the database, the administrator can create the user group, define the group permission for all members of this group only once and assign multiple user accounts to the group.

**4.4 Image web viewer for the researchers of the KKNMS**

A microscopic image web viewer gives the researchers access to view digital images from any location. Here, we describe the main requirements for the KKNMS research network to view digital images and the technical aspects of the existing image web viewer.

**4.4.1 Requirements for the image web viewer**

As the researchers of the MS brain bank are from remote clinical centers, one of the main requirements is to publish a web platform and provide secure access to the microscopic images. Furthermore, from an ethical point of view, patient information must not be published in the web viewer.

The DPS must support images of different file formats. Tools for adjusting color settings, such as brightness and contrast, should be included in the image viewer. Furthermore, the DPS web viewer should be separated from the local biobank system. Since the DPS is able to operate independently, some KKNMS researchers may have access only to the digital images without access to the specimen information. In the viewing application, pathologists need to access digital images rapidly, view them with different magnifications, mark and annotate regions of interest, compare different stains side by side, and view metadata information along the image. In addition, pathologists may need to create snapshots of regions of interest and save them with a standard image format on the computer.

#### **4.4.2 Installation and deployment of web application**

Based on the requirements analysis for the web viewer and our investigation of existing web viewers on the market (described in Subchapter 3.2), the Olympus web interface was set up on the image server PC. The web viewer is based on HTML5 technology which is platform independent and compatible with the most prevalent internet browsers. No additional software installation is required at the client side to access images on the Olympus NIS. The web platform can be publicly available, and therefore it should be secured by secure socket layers using Hypertext Transfer Protocol Security (HTTPS).

The web application meets the requirements described above and provides a metadata view, navigation, and annotation tools in conjunction with digital images. In addition, only authorized users are able to access the web platform. As the users of the KKNMS are predefined, the main pathologist is responsible to supply the researchers with the user accounts and define the user roles and permissions. The researchers only need to view images and add annotations on the region of interest. Therefore, they should not be able to modify any records or the tree structure of the database.

The main specification where the MS-BB may differ from other tissue biobanks is that the MS-BB researcher first contacts the main pathologist and asks about the desired specimen. Once the main pathologist creates a guest account with specific roles and permissions for the researcher, he supplies the researcher with user credentials and an image URL if needed. This technique is required, as there are multiple images with the accompanying metadata. It is not

recommended to open the whole database for any researcher at the beginning of the operation of the system.

The image web viewer supports the viewing tools with different magnifications and color settings such as brightness, contrast, and saturation (Figure 4.20). To access the database, a user should authorize via a SQL server authentication method at the first page of the web viewer. The web viewer is designed to balance the speed of the stream, efficiency of the network bandwidth usage, and image quality. The administrator can customize the view in the web viewer and define the metadata fields that will be visible or hidden in the web viewer, and assign the configuration for the specific users or user groups. For example, if the project administrator does not want to show the Sample ID to specific researchers of the KKNMS, the viewer can be customized accordingly in the configuration software.

A use case scenario for viewing glass slides via the web viewer is represented in Path A of Table B.4 (Appendix B).

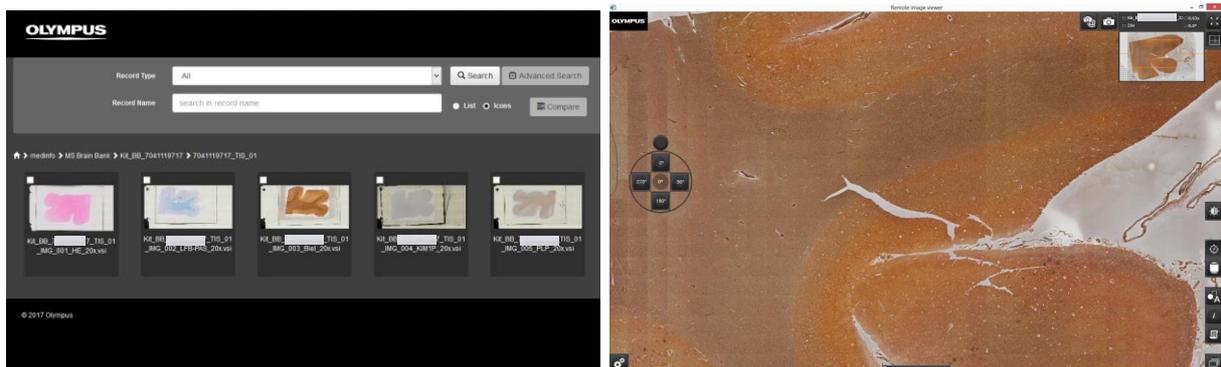


Figure 4.20. Web interface for digital microscopic images

## **5. Realization of the interfacing solution for the digital pathology system and evaluation of working prototype**

Since digital microscopic images are handled by the DPS, and specimen are registered in a separate biobank management system, data exchange between these information systems is essential to improve the laboratory workflow. For this purpose, significant steps were taken and the interfacing solution was developed within the biobank environment. In this chapter we describe the main components and processes by which the targeted systems, DPS and LIMS, interact with each other. Moreover, at the end of this chapter, a developed working prototype is evaluated by means of a W-model. A final (up-to-date) literature analysis about existing integrated DPSs within the different pathology laboratories is presented at the end of this chapter.

### **5.1 Interoperability platform of targeted systems**

Interoperability is the ability of a system to ensure effective interaction with another system providing a service through which information is sent to and received from another application in a secured way [131]. It also means that targeted systems can read, understand, and process exchanged data in response to some unambiguous needs. In the healthcare system, there are different degrees of interoperability that mainly depend on medical appliances. Healthcare system applications interact with each other through standard communication protocols, such as the HL7 transfer protocol. However, in digital pathology, since there is a lack of standardization methods (described in Subchapter 2.3.2), most current digital microscopic systems face interoperability problems with different medical devices [132].

Before defining the road map for the interfacing solution, the degree of interoperability between the targeted systems should be evaluated. Based on the assessment of the requirements, a smart, secured, and safe interfacing solution should be developed.

Medical device manufacturers and information systems developers try to include services in their products that allow their customers to implement an interfacing solution for different purposes. Over the last few years, very few WSI and LIMS companies have developed interfacing solutions in collaboration with an individual laboratory. Since there is no standard

solution used for any kind of microscopic systems, different degrees of interfacing solutions can be developed by each laboratory according to their own needs.

The interoperability mainly depends on the information that is exchanged among the systems. Information can be different types used in a variety of ways, such as textual data for storage in another application, numerical measurements for analysis, medical images for display, etc. As explained in Chapter 1, an interfacing solution is essential to linking microscopic images from the DPS to the corresponding specimen in the LIMS. For this reason, imaging metadata is required to be sent and registered into the laboratory software.

At the beginning of the interoperability analysis of the targeted system, it was revealed that both systems, the Olympus NIS and the STARLIMS, support web service technologies that can be effective methods for implementing the DP workflow in the biobank research system. The STARLIMS system can use the Representational State Transfer (REST) or SOAP-based web services to operate with external application. As it was already mentioned, the STARLIMS system does not have an in-built interface for a DPS. On the other hand, the Olympus supports web services for connection to external applications, but it also has no experience working with any LIMS system via web services. Currently, integrators of the Olympus NIS use the web-services mostly for the integration of the proprietary Olympus file format in their application. These kinds of integrators are mostly the companies that provide the image handling systems or image viewers on the market that support most of the file formats of the digital microscopic images established by different scanner manufacturers. Although web service technology is widely used for different purposes, up until now, neither customers nor the manufacturers of these systems (STARLIMS and Olympus) have used web services for the DP workflow implementation. Mostly, customers of WSI devices try to find a simple solution (such as direct file transfer or SQL stored procedures) that needs very limited human or time resources for development. However, these kinds of solutions are partly automated and are not able to provide a fully integrated digital workflow. Because of scarcity of the experience of web services in DP, interoperability requirements and an architectural model of the interfacing solution needs to be designed.

According to the definition of the Healthcare Information Management Systems Society, there are three levels of health information technology interoperability: technical, syntactic, and semantic interoperability (Figure 5.1) [133]. Technical interoperability is a basic level of

interoperability that provides technical ability to send/receive information between the systems by means of correct transmission protocols. This involves technical aspects such as data integration services, network specifications between connected systems, data transferring protocols, etc. Syntactic interoperability layer covers a well-defined format and encoding for data exchange, such as HL7, XML, SQL standards [134]. Using the defined syntax, systems can read the message and understand the type of information that was sent in an unambiguous manner, but it is unable to recognize the meaning of the information. Semantic interoperability, known as high level of interoperability, refers to the ability to rigorously understand the meaning of the exchanged information. Using the semantic layer, a system that receives the information, processes and interprets exchange information respectively. For example, in the healthcare system, using a common vocabulary of terminology enables semantic understanding of information during system-to-system communication. From the biobank point of view, in digital pathology, semantic interoperability is achieved when the system can read, understand, process, and interpret the complex retrieved data.



Figure 5.1. Three levels of interoperability

### **Interoperability requirements:**

Interoperability requirements are requirements that ensure the optimal levels of interoperability platforms so that data can be shared between the targeted systems in a secure and effective manner. In conjunction with technical and syntactic interoperability, a small degree of semantic interoperability could be concerned with providing an advanced level of information sharing. Furthermore, one of the main requirements is to use the services provided by the manufacturers of the systems and develop the interfacing solution without changing the system. As the DPS and the LIMS are located in the same medical center, sharing information across the systems should be performed via a local network, which will provide secure and safe data transmissions.

## 5.2 Automated and manual methods for data exchange

To fulfill the MS-BB requirements, two methods, an automated technique for the SOAP-based web service interface and a manual method for an end-user to exchange data manually between the systems, were designed. The determined interfacing solution and accompanying manual uploading techniques were implemented into STARLIMS with the help of a biobank interface technician. Additionally, the viewing interface for microscopic imaging metadata within STARLIMS was implemented. The design of these components and their accompanying development tools are described below.

### 5.2.1 General architecture of web services

Currently, the term “web services” is used with different meanings starting from basic concepts to very specific ones. From a very generic point of view, it is described as application services provided for web-based programs [135]. It acts as a layer between the application client and application code (Figure 5.2). Based on this definition, an HTML website is likely to be a web service.

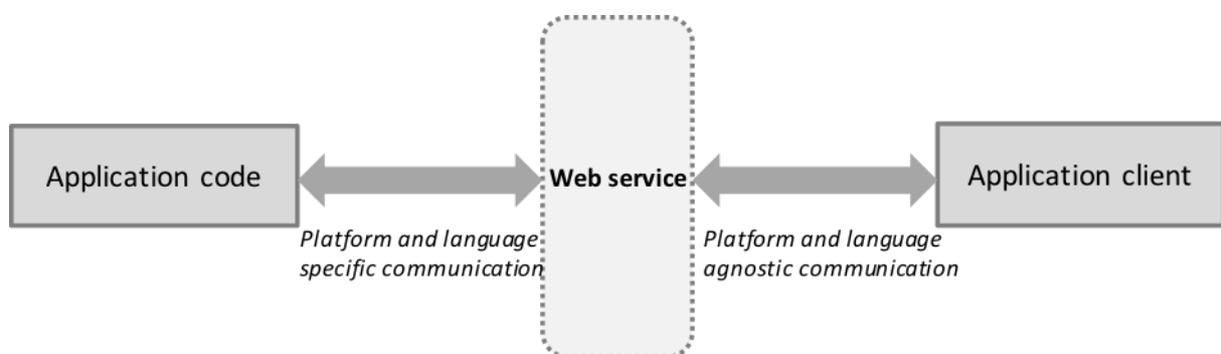


Figure 5.2. Web service as a layer between the application code and application client [135]

According to the definition of the World Wide Web Consortium, web services allow heterogeneous software systems to come together and interact with each other via a network [136]. The web service interface, described in the Web Services Description Language (WSDL) format, sends the request via the network and receives the information using the standard Extensible Markup Language (XML) format.

Web service architecture contains three main components: provider, consumer, and service registry. The web services provider publishes a description of services using the service registry, while the consumer uses the services and implements it for specific purposes [135]. A simple web service architecture is illustrated in Figure 5.3.

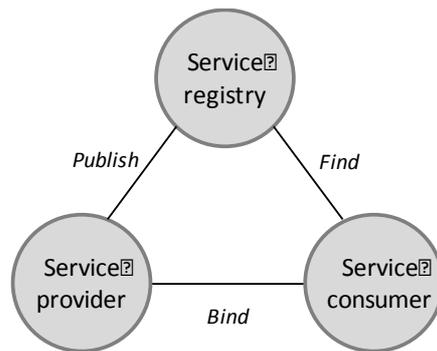


Figure 5.3. Architecture of web services. It contains three main components: (1) service provider, (2) service registry or service description published by provider, and (3) service provider that consumes the web services. Using the web services, provider and consumer are linked each other [135].

To design and build the web service application, multiple tools and technologies might be used. The main technological components of web service architecture that play crucial roles in the development of the interfacing solution require more detailed definitions:

**XML** – a standard and platform-independent markup language that encodes information in a textual format. An XML schema specification provides the rules to define the structure of XML files [137, 138]. XML or XML-based languages are increasingly used to represent the web services (called XML web services). Mostly, XML web services are used for data exchange by different applications.

**WSDL** – an agreed-upon XML-based descriptive language that defines information sending and receiving between provider and consumer. WSDL describes how the web services work, what kind of operations can be used, and how to use these operations [135].

**SOAP** – a communication protocol for data exchange between the provider and consumer. Using SOAP messages (SOAP request and SOAP response) remote procedures are called from the web services provider, and the corresponding responses are retrieved for the web services consumer. A SOAP message is based on the XML format [135].

## 5.2.2 SOAP-based web service middleware application

Even though the web services architecture for different applications and systems may contain the same technology, there are many ways to use the web services and various specifications of the web service interface to link remote systems together. Both applications — the DPS and STARLIMS — are installed within the UMG and support web service technologies, but the linkage between them may vary from loosely coupled type to tightly coupled type. Olympus NIS, an application that implements web services, is called a provider, where STARLIMS, an

application that invokes the services offered by the provider, is called a consumer. Below, we describe the interfacing solution that meets the requirements of the MS-BB and provides an effective link between the specimen and specimen-derived assets. Therefore, the developed interface provides the mechanism that uses the interoperable system services. Because the developed interface acts as a layer between the different applications, it is also called a middleware application and is likely to work either as a separate application run between the systems or implemented as an interface into the system that consumes the web services. Therefore, the terms – interface and middleware application – might be used interchangeably in the thesis. Before describing the web service architecture, the middleware application requirements need to be identified.

#### 5.2.2.1 Interface requirements

The interface that will interact between the LIMS and the DPS should satisfy the following functional requirements:

- Digital microscopic images regularly added on the image server should be automatically linked to the specimen using the barcode information.
- The middleware application should send the request to the server and get the following parameters for multiple images at the same time using triggers: thumbnails, URLs, barcodes, staining types, and annotations. The retrieved parameters should be registered into the LIMS database.
- The responsible pathologist should be able to update the information regarding the microscopic images for specific specimens (manually by selection) directly from the LIMS.
- In exceptional cases, if the automated metadata entry from the remote application does not work properly, the pathologist should have an opportunity to perform the same procedures manually within the LIMS.
- Settings for the trigger activities, such as frequency and time range of repeated actions, should be able to be configured according to needs of the LIMS user.
- The pathologist should be able to run image web viewer for selected images directly from the LIMS.

### 5.2.2.2 The web service interface architecture with accompanying processes

In Subchapter 5.2.1, the basic web service approaches were defined. However, as we are discussing the specific applications and their interfacing solutions, a more precise description of the web service architecture from the biobank's point of view is required. The corresponding web service architecture is illustrated in Figure 5.4.

The web services are represented by svc files that are responsible to run services using Microsoft IIS in response to the request. The svc file reference is indicated in the client application and activates the hosted services directly from STARLIMS.

The web service registry contains the methods for accessing and retrieving the image records from the DPS database. The web service methods are accessed via a local proxy class from the consumer. The proxy class acts as a real remote object and makes the methods and properties of the web services available to the application. Once the proxy class is declared, it should be defined as an object that refers to web services. The web service interface developed by the STARLIMS Scripting Language sends an XML-based SOAP request to the web service provider. As a result, provider of web services returns the response to the LIMS.

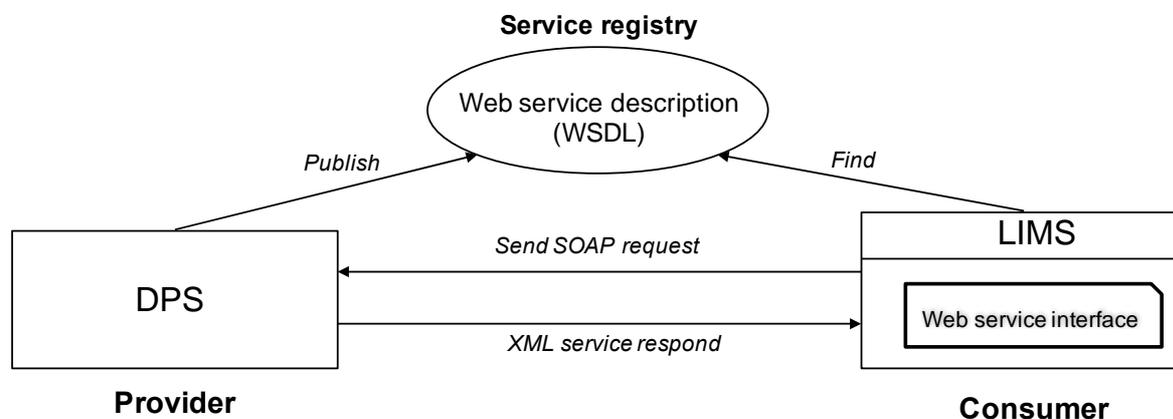


Figure 5.4. Web service architecture for the DPS and the LIMS

Before sending a query to the server to access the database records and imaging data, the database accessing methods and additional operations are required to be called. For example, steps involved include getting server information, returning parameters of the published database, connecting to the suitable database using user credentials, retrieving fields of the database to reconstruct the query, and finally sending the database query to the remote application to retrieve imaging parameters.

Key components of the web service interface are the messages sent back and forth between the consumer and provider systems. SOAP messages come in two types: request and response messages. The crucial element of the message sent to the image server is a query that is defined with an XML query scheme and has a standard XML document tree structure. This contains several components: header, selection fields (attributes), filter block analogues to the “where” statement in the SQL query, and sorting parameters. The interface solution implemented in STARLIMS uses web service methods from the Olympus system. Both systems are commercially available and therefore the whole code implemented for the interfacing solution is not included in the thesis. However, some part of the analog code written in C# programming language is represented in Appendix F. To cover all the pertinent issues, the interface solution details are represented using flowcharts and are described here in the text.

A query is structured in a way that selects the following components of specific digital images: thumbnail, image name, barcode, staining type, Sample ID, Glass ID, and URL (annotations in case of need). The returned value is always represented with an XML file and contains information about multiple images. It is decoded using the STARLIMS Scripting Language in the web service interface. Finally, the result is instantiated as an array, where each element corresponds to a single imaging metadata. Despite the fact that staining type is returned from the server and is read using the STARLIMS Scripting Language, the meaning of the staining type can be additionally recognized using the Glass ID (specifically suffix such as IMG\_001 or IMG\_002, etc.). Thus, some degree of the semantic interoperability can be included in the interface solution, which can be used for different purposes in the future.

From a procedural perspective, accessing the web services and the accompanying processes are illustrated by the sequence diagram in Figure 5.5.

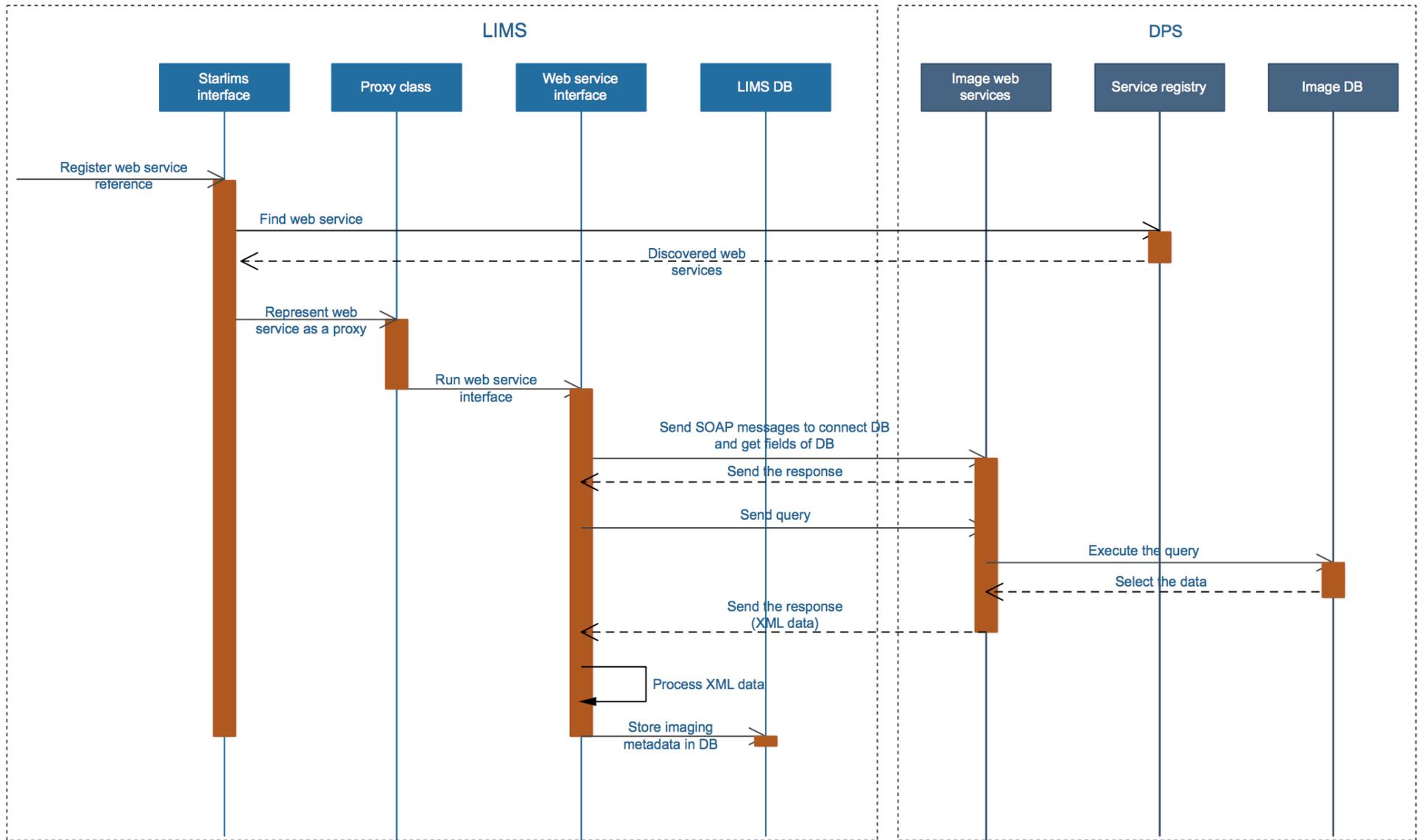


Figure 5.5. Sequence diagram for the data exchange from the DPS to the LIMS

### 5.2.2.3 Saving the imaging metadata in the biobank management system

Once the imaging metadata is retrieved and stored as an array of complex elements using the interface, the individual digital image information is required to be registered and assigned to the corresponding specimen in the LIMS database. Each value of the array element corresponds to the metadata of an individual image. One of the main components of the image metadata is the barcode of the glass slide that was initially generated from the LIMS and printed on the glass slide label before scanning (Subchapter 4.1.4). Since the barcode consists of a Sample ID and a suffix, the middleware interface can easily correlate it to the corresponding specimen registered in the LIMS. As a result, the whole imaging data including the barcode are registered into the table dedicated for the specimen-derived assets in the LIMS database. The database table is made up of the following fields represented in Table 5.1.

Table 5.1. Fields of the table for the microscopic images in the LIMS database.

Fields	Type
ID	Int64
Image ID	Int64
URL	String
Thumbnail	Base64 (LongBinary)
Sample ID	String
Staining type	String
Barcode	String
Insertion date	dateTime
Modification date	dateTime

### 5.2.2.4 Data synchronization and communication settings

Multiple digital slides have already been scanned and images were stored on the server. Using the middleware application, the already existing digital images should be linked to the corresponding samples in the LIMS. There are new glass slides that will be regularly scanned. Therefore, the imaging metadata for the newly scanned glass slides needs to be automatically added to the LIMS. To perform this function effectively, the optimal frequency of data exchange from the DPS to the LIMS needs to be determined. Execution of the web service interface can occur either automatically or using the manual tools inside the LIMS system.

There are three types of methods for invoking of the web service interface implemented in the LIMS according to the interface requirements:

1. Run the web service interface for the images that were already scanned and stored on the server before the interface was developed: For the 800 digital images already stored on the server, the web service interface should be run only once and the imaging data should be retrieved for the associated brain IDs. As there are multiple digital images on the server corresponding to two brains, the web service interface needs to be executed once for each Case ID (Brain ID) by a biobank interface technician to avoid overloading the system.
2. Call the web services through triggers: Using a standard method, the web service interface should be automatically executed at a certain frequency in response to new or modified imaging data stored on the NIS. Therefore, the web service interface is configured in such a way that it is launched once every 24 hours. A query sent to the image server ensures selection of the images inserted or modified during the previous 24 hours by using two fields - modification time and creation time of images, respectively.
3. Run the web service interface by user (pathologist) manually: In an exceptional case, the pathologist can execute the web service application directly from the “remote images” tab of STARLIMS that retrieves imaging data for selected sample or brain IDs. New, updated, or deleted digital images are retrieved for selected items (sample or brain) using this method. This option is also applicable when a pathologist needs to link remote images to the specimen once they are stored on the server (manual method is described in detail in Subchapter 5.2.4).

Settings of trigger activity, such as frequency and time range of repeated actions, are likely to be easily configured according to needs of users in the future.

### **5.2.3 Image metadata viewing interface in the biobank management system**

Most LIMS systems, including STARLIMS, allow users to customize and configure the system according to their needs. In order to view the microscopic imaging information in STARLIMS, a new interface (the tab is called “Remote Images”) was configured and implemented (Figure 5.6). On the one hand, using this tab, the pathologist can view imaging metadata of each specimen, such as a thumbnail of digital images, staining type, image URL, and barcode. In addition, the user can manage the list of images and manually edit or delete specific images.

On the other hand, this interface acts as a bridge between the LIMS and the DPS. By clicking the image thumbnail, a new tab in the web browser is opened and the DPS web viewer is loaded, which allows users to view and investigate the whole digital image with the highest magnification.

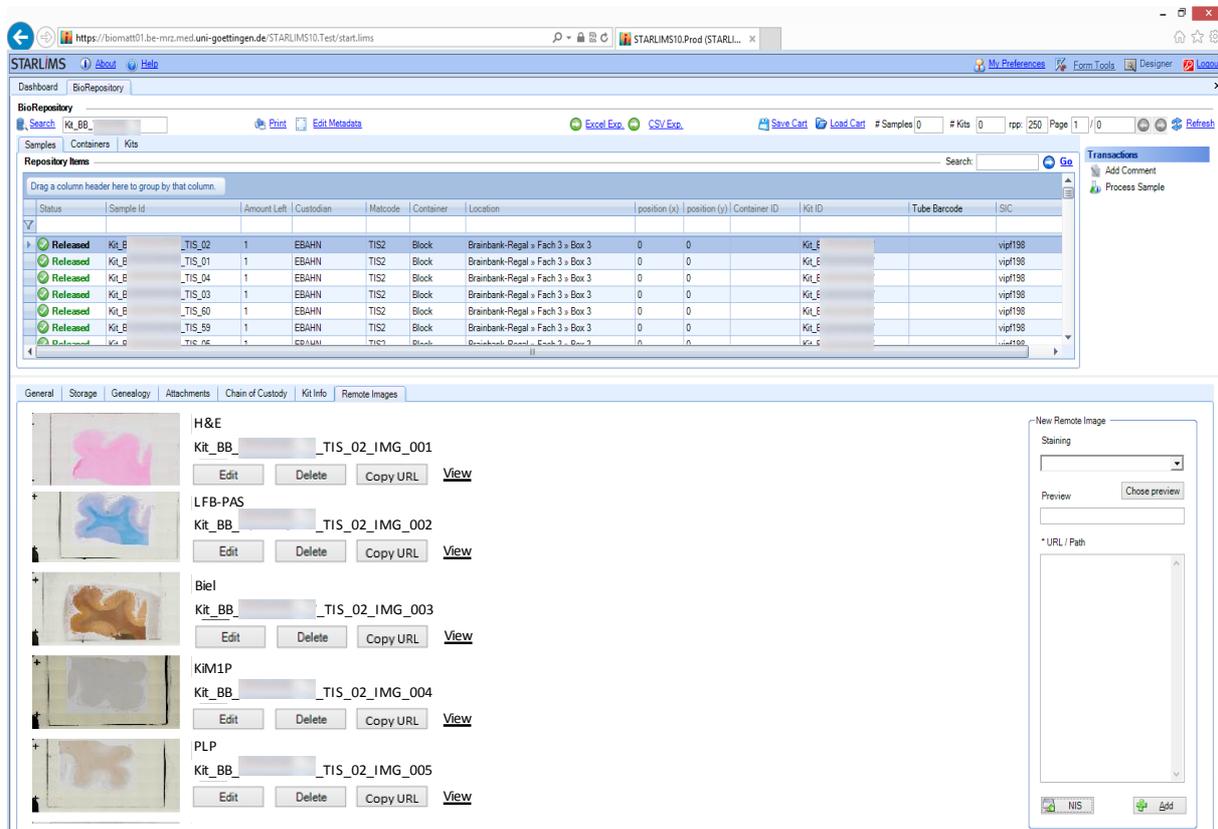


Figure 5.6. Mockup of “Remote images” tab in STARLIMS

## 5.2.4 Manual method for the addition of imaging data

Besides the automatic mechanism described above, manual tools for the data exchange from the DPS to the LIMS are already implemented in the laboratory software (in the tab “Remote images”). A user can manually execute the middleware application for the selected specimen that works similarly to automated tools. Thus, the imaging information for specific specimens can be updated, added, or deleted.

Additionally, in the “Remote images” tab (on the right side in Figure 5.6), there is a remote image uploading tool that provides the opportunity to link an image from another remote application or folder. Using this interface, it is possible to add a thumbnail, staining type, and a URL of image. Once the imaging metadata is uploaded, they are displayed in the same tab on the left side just as any other digital image thumbnails. Because a specimen may have multiple corresponding images, the manual uploading interface supports multiple-image

uploading tools. Thus, manual techniques give the pathologist the opportunity to add references for images that are not located on the NIS. This will facilitate the navigation and viewing process when the digital images are located on different applications.

At the moment, manual tools are not frequently used because the images are only scanned with the Olympus scanner and there are no other types of images that need to be cited into the LIMS. However, if there is the exceptional case when a pathologist needs to link microscopic or any other type of image, such as radiology images stored on a different server, the interface can be used appropriately.

A use case scenario for the data sharing between the LIMS and DPS using automated and manual tools is represented in Table B.5 in Appendix B. Additionally, the scenario for viewing images from the LIMS is represented in Path B of Table B.4 (Appendix B).

To summarize the complete DP workflow, all steps were significant to provide standardized processes starting from the preparation and ending with the application. The flow of the main operations between different components of the DPS and the LIMS is outlined in the process diagram in Figure 5.7.

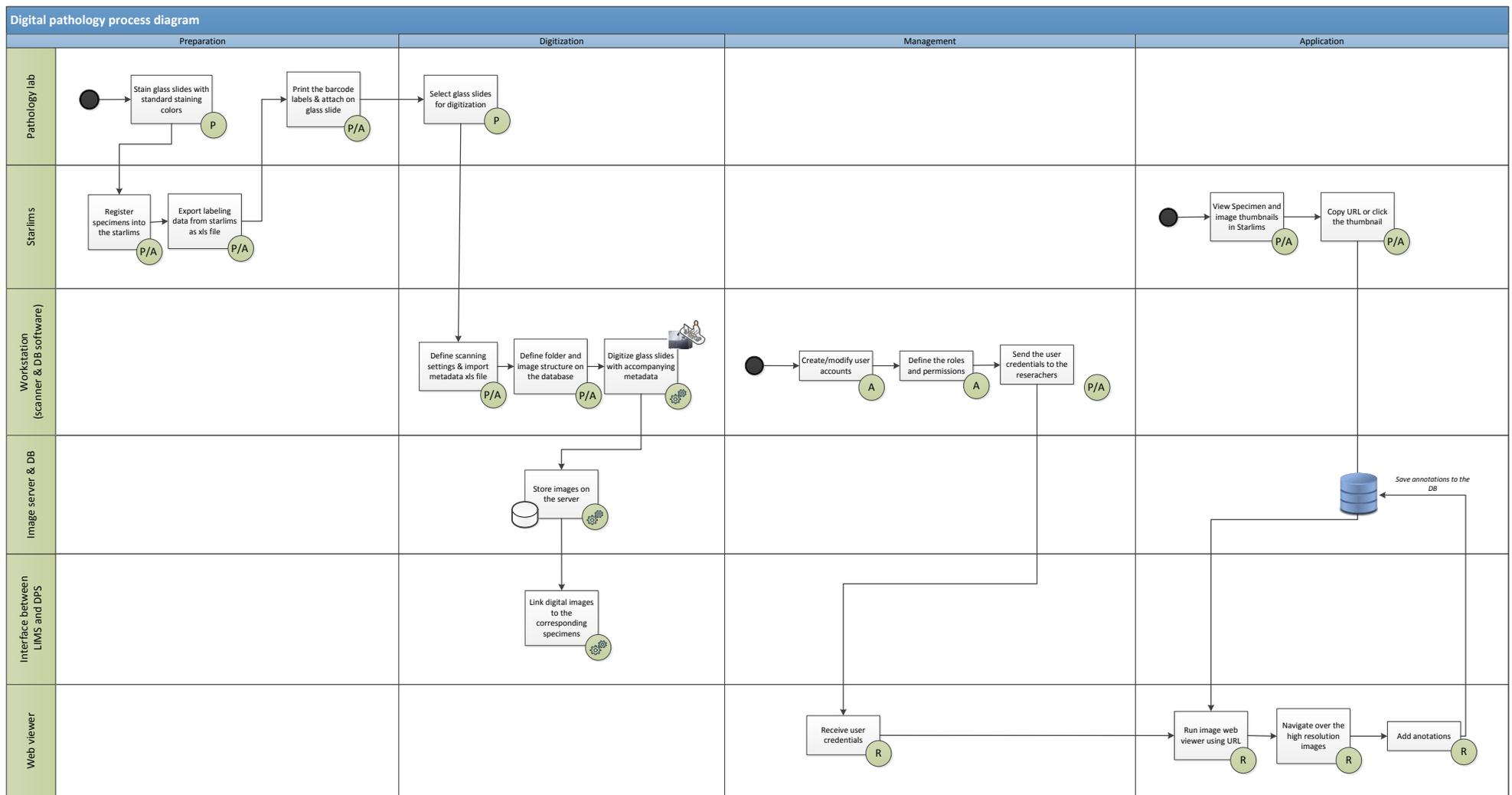


Figure 5.7. Digital pathology workflow. The tasks are performed by different users that are indicated in the green circle: Administrator (A), Researcher of KKNMS (R), Pathologist (P), Pathologist or research assistant (P/A), and automated process (depicted with the icon ).

### 5.3 Evaluation of the prototype

Using the standard methodologies, a developed DPS needs to be evaluated and each step of the process should be summarized. The W-model, also called the validation and verification model, has been used to evaluate the system and represents the structured test of the system development life cycle (SDLC). Validation of the system or of part of the system means that the final result fulfills the user's expectations (it answers the question "did we build the right system?"), whereas the verification checks that the result meets requirements, regulations, or existing standards (it answers the question "did we build the system right?") [34]. The W-model, which is an extension of the widely used V-model, shows the parallelism of development and testing activities more clearly. Figure 5.8 illustrates the W-model, and contains two "V"s depicted with blue and red lines representing the SDLC and testing procedures, respectively. The W-model covers system testing to discover and debug system failures [139]. Testing or review of the system starts once the SDLC is initiated. Thus, possible failures at each stage of the SDLC can be detected and analyzed in the earlier stages of development.

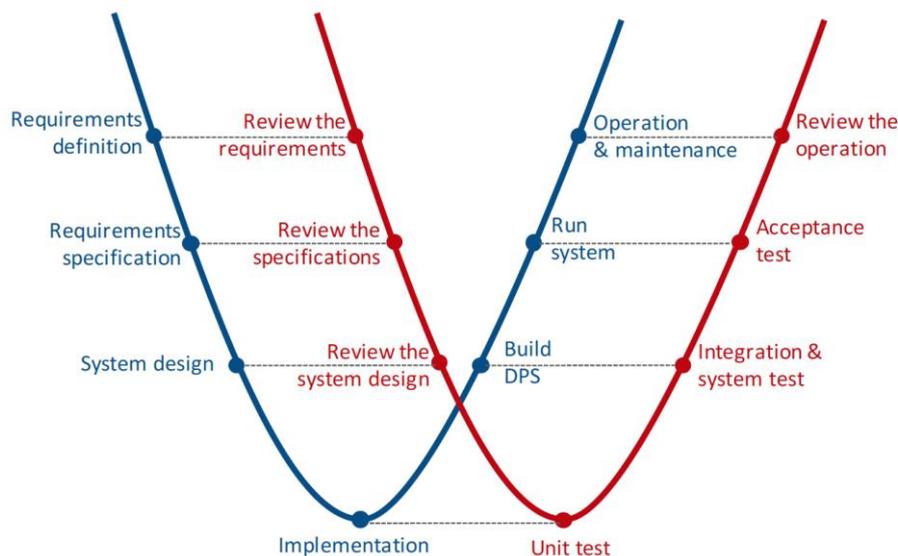


Figure 5.8. W-model for the evaluation of the DPS. The blue line represents the SDLC, whereas the red line represents the corresponding testing procedures.

The W-model is a widely used standard system development process and describes all steps, including the system design and coding. As the developed DPS solution applies to the commercial products that have already been developed, tested, and operated within multiple clinical environments, the W-model has been adapted to the specification of MS-BB. The W-model is used for the following purposes: (1) to determine that final DPS meets the

requirements defined at the beginning of the study, (2) to ensure that DPS is correctly set up and the interfacing solution is effectively developed, and (3) to ensure that the developed system can be operated successfully.

The W-model of the DPS is made up of the following phases, which are validated with the corresponding test activity:

1. **Requirements definition** – The first step of the SDLC is shadowed by a **review of the requirements**. Once the requirements were collected, the fulfillment of the requirements according to the wishes of the stakeholders was determined. This means that the requirements should be clearly defined so that users (stakeholders) with different backgrounds can understand them. Then, the ambiguous requirements regarding the preparation of glass slides, barcode processing, and viewing interface were considered with the pathologists and IT technicians in detail to avoid any contradictions about these features. The requirements were reviewed during regular meetings with stakeholders. As a result, they were categorized into different groups and prioritized according to need (described in Chapter 3).
2. **Requirements specification** – This is shadowed by the **review/test of the specifications**. Multiple glass slides have been scanned several times using two different scanners to test the functional and non-functional requirements and ensure to that the chosen product is able to fulfill the requirements. During the scanning process, technical details like file size, file formats and structure, image metadata, barcode processing, scanning time, conversion tools, and batch processing functionality were evaluated. Furthermore, the requirements for viewing the digital images were checked in various viewing applications, such as OMERO, Zeiss, PathXL, Imagescope, etc. Interfacing requirements were reviewed with the manufacturers of the system. Test results were evaluated along with the responsible stakeholders and the next steps were determined according to the matches and failures detected during the testing process. Some additional missing features, like the automated entry of metadata, were discovered during testing of the scanning process. Thus, not all but most of the requirements were verified and validated during the requirements engineering phase. Test results are presented in Appendix D.
3. **System design** – This is the phase in which the architecture, DP workflow, and interfaces were designed. To validate this stage, **review/test of the system design** were performed, by which the expected failures were detected and discussed with IT

technicians. For example, there was an incompatibility in the web services of the existing systems that required modification on the server side. Furthermore, the structure of image folders in the database for digital images was determined incorrectly and the images could not to be automatically stored. Once the faults were discovered, the corresponding parts of the system design were modified and tested again.

4. **Implementation (set up & coding)** – This is the base of the W-model where the system design is transformed into the implementation of the system and converted into codes for the web service interface. During this phase, individual components of the DP workflow and web service interface were tested.
5. **Build the integrated DPS** – This is the phase where the DPS is set up as a complete system and the internal modules have the ability to communicate with each other. In addition, it ensures that the whole system can interact with external systems, such as LIMS. To validate this step, **integration and system tests** were performed. Using the integration test, components of the DPS, such as scanner, image server, database, and viewer, were combined and tested as a complete system. Using the system test, the complete functionality of the system and communication with the LIMS were tested for different cases.
6. **Run system** – In this phase, the developed DPS is ready to use in real time within the biobank environment. To detect faults during the running process, **user acceptance tests** were performed. Five main use case scenarios that described the whole DP workflow from various user perspectives were carried out several times. The user acceptance tests verified that the DPS meets to the requirements and can be used as a working prototype. User acceptance tests also covered load and performance defects. Several cases were tested to measure the data transfer rate and the speed of image loading on the viewing interface.
7. **Operation and maintenance** – This is the last step of the W-model. Once the system was set up, the main pathologist began a live operation of the system. Initial training was provided for the pathologist and the assistant who will work with the system. All technical components such as preparation, scanning, storage, viewing, and interpretation of digital glass slides were explained and discussed during the training. Therefore, 800 glass slides were prepared, digitized, and stored and are available via web viewer directly from the LIMS. Because it was the beginning phase of the operation

of the system for the biobank project, each step of the operation was inspected very carefully. There were some failures detected during operation that were related to the connection between the scanner and the server, firewall settings, and other minor incidents. Once the system was launched, the operational concept for the users of the DPS was provided. This concept involves operational workflow including all use cases, personnel, and IT resources used in the process, as well as the estimation of the storage environment for the following years. In addition, for the successful operation of the DPS, it is vital to manage the information system risks and to deal with challenging events. Therefore, the concept also describes the most likely adverse events that already occurred or that might occur in the future during system operation. The problems and feasible risks were analyzed and effective solutions and controlling mechanisms were identified. Thus, the risks to the system will be mitigated to an acceptable level and the DPS will work effectively. As the operation and maintenance of the system is an ongoing process after the development of the DPS, this step is not explained in detail in this thesis. On the other hand, because the DPS is dedicated for the KKNMS research network, a detailed operational concept is available only for the the KKNMS researchers and is not involved in this thesis. The most important part of the concept, such as use cases, are presented in Appendix B.

All these steps were used in the validation and verification of the working system and led to the successful operation and maintenance of the sustainable system for the coming years.

#### **5.4 Concluding literature analysis of the currently existing DP frameworks and interfacing standards in pathology**

Since 2013, multiple pathology laboratories have implemented DP tools to a certain degree in their workflow – mostly for research purposes. However, until now, very few laboratories have used a fully digital workflow for clinical diagnosis.

In the middle of 2016, the Department of Pathology at the University of Pittsburgh Medical Center integrated DP into its anatomical pathology laboratory information system for diagnostic purposes with the help of vendors (Omnyx Integrated Digital Pathology and CoPathPlus) [140]. This integration solution provides unidirectional data exchange from the pathology laboratory information system to the DPS using web services. Therefore, a

pathologist can access patient identifiers, histology, and case data directly in the DPS. Regardless of the fact that they use the same technology (SOAP web services), the specifications and application of the system differ from the approach used for biobank purposes. However, this example provides stronger assumptions that our integrated DPS developed using web services, has practical uses for a variety of potential purposes, including clinical diagnosis.

From 2010-2016, LabPON, a pathology laboratory located in the Netherlands, which handles more than 55,000 histological cases per year, fully digitized its diagnostic workflow [141]. The entire process of implementing DP in clinical diagnostics required a significant investment of time and money. There are still challenges that create some obstacles in the process. For example, the digital images require about 450TB of storage capacity per year, but the server capacity is currently limited to 89TB. Therefore, at LabPON, the digital images are deleted after 6-8 weeks. Making a diagnosis based only on digital images is allowed in the Netherlands, but there is no regulation regarding the storing and archiving of digital microscopy images. Digital images are managed by a case viewer that is connected to the LIMS. While images are used for diagnosis, cases are being moved from the folder "In Preparation" to a "For Review" folder, and finally into the "Finished" folder to sign out the case [141].

According to these examples, it seems that the final implemented solution is mostly based on the needs of applications and the organizational environment. Additionally, as there are multiple distinct WSI technologies and LIMS systems on the market, and the reality that standardization of DP is still challenging, implementation mechanisms should be determined for specific systems that are used in a laboratory.

The HL7 Anatomic Pathology Working Group tries to enhance HL7 standards for pathology use cases. In order to facilitate interoperability of laboratory information, they have covered various aspects of pathology, such as improving the tracking of anatomic pathology specimens, integrating pathology data into the medical records, ensuring consistency of the image association, and improving the usage of tissue collection [101]. Additionally, the HL7 working group has tried to improve the transfer of data between pathology labs and clinical settings, which requires filling the gap around vocabulary (such as LOINC and ClinGen codes) for pathology laboratory results [142]. Because there are multiple components that need to be taken into account to enhance existing HL7 standards, progress in pathology units has been quite slow.

During the last few years, several imaging applications have emerged on the market to handle WSIs, autopsy images, electron microscopic images, and other pathology images [143, 144]. In the most popular of these applications, the DICOM standard is used either for handling digital images as a file format or for image transmission to integrate digital images into a hospital's infrastructure [145]. In the second case, when the DICOM is used for data transmission between different applications, digital image file formats should be JPEG, JPEG2000, TIFF, or DICOM itself. If the images are presented in JPEG (using the lossy compression method) format, the image quality is not sufficient enough. TIFF, JPEG2000, or DICOM file formats (despite the fact that these formats are based on the lossless methods) have several limitations for pathology images. For example, the maximum file size is limited to 4GB. The DICOM file format is still very rarely used for pathology images, as it has not been integrated into microscope image acquisition systems. There are several applications (e.g., JP2 WSI Converter) that convert some of the proprietary file formats into the standard file format. However, the main limitation of these applications is that the metadata information is likely to be lost during the image conversion from their original formats (mostly proprietary formats) to standardized image formats (JPEG or TIFF).

Even though the DICOM standard and the PACS storage solution are not used for our integrated DPS, some specifications of the DICOM standard were taken into account. Since there was a growing interest in radiology to access DICOM objects from HTML pages via HTTP/HTTPS, the DICOM working group 27 (web technology for DICOM) developed the Web Access to DICOM Persistent Objects (WADO) standard as a web extension to DICOM. Because the PACS system is generally used as a storage solution for DICOM objects, the WADO Server is connected to the PACS system to retrieve the imaging information. According to the DICOM specification provided by working group 20 (integration of the imaging and information systems) and working group 27, data exchange involves transferring two different kinds of data – pixel data and metadata. WADO uses the web service technology to facilitate access to the images. Pixel data is retrieved by the RESTful web services and metadata is exchanged based on the SOAP-based web services [146]. From a technological point of view, the WADO standard utilizes technology similar to our developed system, but uses a different approach for various applications. Additionally, to keep the enriched metadata of the microscopic images, it is better to store them in the proprietary file format so that information is not lost.

Later, it can be freely converted into the DICOM standard when it becomes integrated into the WSI systems.

Despite the fact that the DICOM standard has significant technical limitations in pathology (discussed in Subchapter 2.3.2.3), it is still possible to roughly define the future directions of the developed system in terms of the DICOM standard. As our interfacing solution is based on the same technology (SOAP web services using XML) for metadata exchange, it allows us to suppose that our solution will be able to easily adapt to the new DICOM standard, since the imaging or metadata information is kept without any loss. Moreover, the interfacing solution can be improved along with the development of the DICOM standard in the ensuing years.

## **6. Discussion**

This final chapter summarizes the most important results of the research carried out within the scope of the PhD study. The results are categorized in the coming subsections, which respond to the research questions described at the beginning of the thesis. Additionally, the main advantages, impediments, and limitations of the developed approach are considered. The chapter is finalized with the future directions and recommendations for the maintenance, deployment, and improvement of the developed solution.

### **6.1 Contributions to the optimization of the Brain Bank workflow**

The developed IT infrastructure for pathology workflows plays a crucial role in the successful use of tissue banks. The main purpose of the study was to determine and implement into the existing biobank IT infrastructure, a microscopic imaging system that improves the pathology workflow by integrating automated tools, facilitating the handling of digital and glass slides, and enabling the effective management of the digital images.

#### **6.1.1 Regulatory and standardization aspects**

By investigating the current regulations and guidelines, it seemed that regulatory impediments have a leading role in the wide adoption of digital pathology in primary diagnostics. Despite the fact that standards for storing and transferring microscopic images have not yet been declared, pathology laboratories are capable of using a WSI system for various purposes, including research, education, or remote consultation. The key recommendations of the DPA and CAP guidelines for WSI system validation for clinical and nonclinical environments were considered and adapted for the biobank research environment. Additionally, the TMF recommendations regarding the medical imaging system were used, by which existing infrastructure at UMG was deployed and improved to provide high quality medical research. Therefore, a general roadmap for the development of the interoperable DPS described in the thesis follows the main principle of the existing approaches and involves additional aspects that are specific to a biobank application. Thus, the developed system has more potential to be easily adapted and modified when the existing standard is improved. Consideration of the regulatory or standardization factors has led to the successful setup, deployment, and maintenance of the system.

### **6.1.2 Requirements engineering for the microscopic imaging system**

The requirements engineering process was the initial development stage for the interoperable DP workflow. By using the predefined requirements engineering framework, which involves progressive procedures for requirements analysis and management, all requirements were collected through various elicitation techniques, such as regular meetings with stakeholders, observations, documentation-based, and others. Using a question list, the requirements were elicited and effectively categorized and prioritized according to their needs. The requirements engineering described in Chapter 3 summarizes the functional, technical, system, and environmental requirements for the microscopic image handling system for biobank applications. In addition, requirements were prioritized according to their necessity – namely essential, preferred, and desirable.

To manage and organize the various perspectives in an effective way, three viewpoints were identified: (1) the process viewpoint, which analyzes processes from different user perspectives and describes process models; (2) the information technology viewpoint, which represents the techniques for setup, configuration, and management of WSI technologies and related components such as network communication, storage environment, and user management and (3) the interfacing viewpoint, which provides the concerns for both, i.e., developers to develop the interfacing solution for data sharing between two independent systems and the system users who navigate within these systems. Using the dissenting viewpoints, a more structured and modular approach was used in the requirements engineering that differentiates multiple concerns, reduces the complexity of understanding by focusing each aspect of the system separately, and improves the communication between various stakeholders [113]. Using the UML specification and use case models, a DP workflow was broken down into several scenarios that describe the processes and behavior of the system from different user perspectives.

During the precise requirements analysis, the needs and existing state of the biobank were estimated. Additionally, potential products on the market that possibly matched the requirements were thoroughly investigated. Through the market analysis of DP viewing systems, supplementary requirements such as color features, advanced navigation tools, and the support of different file formats were identified. There were several options for image handling systems on the market at that time, and the most suitable solution was chosen.

### **6.1.3 Development of the digital image handling system**

Based on the requirements analysis, a viable solution for handling microscopic images was determined, designed, and configured according to the project's needs.

In the beginning, the identification mechanism and uniform labeling concept were determined for the glass slides that needed to be managed effectively and handled through the entire workflow. Using the barcode labels, all glass slides were identified with a unique ID that improves the laboratory workflow and provides the link to the corresponding specimen. Labeling with at least two IDs can reduce handling errors and provides an effective assessment of samples [147]. The glass slide labels contain no personal information that could identify a patient (such as patient ID or name). Thus, in terms of ethical and privacy issues, the DPS maintains a data protection strategy (suggested by TMF) via separating distinct information within various information systems. A barcode solution diametrically changes the pathology workflow. A two-dimensional matrix barcode that can be scanned using special equipment will enhance the automated tools. On the other hand, human readable text provides a quick and simple way to identify the main features of a tissue slide. In addition, textual components are convenient for the identification of glass slides when the barcode scanner is not available. Studies have shown that barcode labels have huge benefits in terms of patient safety, automation, error reduction, and the velocity of tracking [148].

According to the digitization strategy, the most interesting cases from the MS biorepository were being scanned regularly. To supply the researchers with sufficient information, at least five glass slides for each specimen stained with standard staining colors were selected for scanning. Using the high resolution scans of glass slides, a researcher has the ability to investigate the specimen, compare corresponding digital images colored with different stains, make a very detailed visualization of lesions and characterize lesion activity. Because the scanning process is an essential part of the workflow, several automated tools were integrated into this stage. First, multiple glass slides were digitized using a batch scanning mode, since the general scanning settings had been determined for the glass slides of the biorepository. Second, during digitization, a huge amount of metadata for the glass slides was imported into the system using an XLS/CSV file. The file was designed and configured in such a way that pathologists could easily fill it in by specifying a couple of parameters for multiple glass slides, such as Brain ID, starting and ending index of the specimens that are going to be scanned, and date of autopsy. Finally, automatic naming and saving of digital image files on the image server

was applied during the scanning process. The names and storage location of the images were automatically determined according to the slide property values. Thus, using the automated tools, the scanning workflow was improved and significantly accelerated.

The metadata of the glass slides plays a crucial role in data collection and analysis within a biobank environment. For this reason, extensive metadata was entered into the WSI system during scanning and was stored along with the digital images. Metadata contains information about the glass slides, including corresponding specimens, location of lesion, staining type, owner, scanning settings, and other specifications of the digital file. Using the extensive metadata, digital images are characterized in detail, which facilitates image analysis and the search process. In addition, it is likely to be effectively used to share data between different systems via the query tools. From the biobank's point of view, the metadata of digitized slides stored along with the images acts as a descriptor of microscopic images. Because metadata is stored along with the images, a separate IT infrastructure for metadata management is not provided. However, that might be useful in the future.

Scan settings were initially assessed for 800 glass slides of the MS biorepository. Even though scan settings were inspected and modified according to needs during the scanning process, approximately 2% of all the digital images were damaged and required rescanning. To locate the corrupted images or errors, a pathologist should review the scan results. This might be an additional task, but it is a significant step in controlling image quality. Since a biobanking is dedicated to the long-term, digitizing glass slides is an ongoing process that can be successfully continued with the same scanning settings.

The digital images of the repository need to be available from remote locations, so an image server, database, and web viewer were set up and configured according to the identified requirements. After scanning, the digital images were automatically stored on the server in a well-structured form. By using a hierarchical structure of folders, such as folders and subfolders for projects, cases, specimens, and images, a large dataset can be organized and grouped effectively. Because of the large size of the digital images stored in proprietary file formats, sustainable storage solution that can handle thousands of images plays a crucial role in the successful operation of the system. Based on the estimation of the storage capacity for handling digital images for the next five years, 10TB of data volume is available on the image server, with the potential to add additional volumes of disk space in the future. Thus, problems related to data redundancy are avoided. Because implementing an operable image handling

system for the ongoing research project was the chief goal, long-term protection of imaging data (such as archiving techniques for the retention of digital images) are not included in this thesis. Further investigation is necessary in this direction.

User accounts with accompanying roles and permissions are managed using the RBAC mechanism (three different types of users: administrator, power user, and guest). This provides the advantage of differentiating KKNMS researchers who can view images from remote places via the web viewer, from the users who need to access the images from the UMG. The scanning system is configured in a way that an administrator can manage the permissions needed for accessing specific images or image folders and assign them to an individual or group users.

An instant digital representation of glass slides is available to researchers using the Olympus web application. Once the digital slides are viewed remotely, the KKNMS researchers can easily annotate or capture a region of interest, determine the desirable MS tissue samples for further investigation and request physical delivery of the specimen. Thus, the viewing of digital images facilitates and improves the biobank workflow.

#### **6.1.4 Interfacing solution between the DPS and the LIMS**

The proposed manual and automated methods provide an effective prototype for the integration of digital microscopic images into the laboratory information system. This prototype improves virtual access to the microscopic images and enhances the management of tissue samples for the biobank research network. The proposed solution is ready to be used in practice for the KKNMS research environment. Using the middleware application, several problems were solved:

(1) Data sharing between disparate systems is performed using SOAP-based web service technologies. Automated data exchange between the DPS and the LIMS can optimize DP workflow, decrease turnaround time, and reduce the linkage errors of the specimens. For efficient communication, the middleware application uses a query-based mechanism to select the desired information from the image server and retrieves data with an XML-based standard format. One of the important benefits of the interfacing solution is its flexibility, which enables the simple modification of any parameter according to need (for example, parameters for the data synchronization could be easily customized). Before the DICOM standard is widely

adopted in pathology, web services will have a leading role in the communication between the different systems [30]. However, the primary limitation is that scanner vendors provide varying degrees of integration tools that lead to the interoperability challenges of the targeted systems. A very limited number of scanners support web service technologies. It would be more beneficial if vendors of the LIMS and microscopic scanners implement the web services in their systems and customers use them according to their needs.

(2) There are two types of manual tools integrated in the LIMS. First, a user can execute the interface solution manually for each specimen, which updates its imaging metadata in LIMS (connects to the NIS). Second, an image metadata uploading tool is integrated in the LIMS that enables it to add the image references (such as a URL, thumbnail, and staining type) for each specimen. The described approach is applicable in very exceptional cases either when the automated method is not working properly, or when other types of digital images (such as macro images or radiology images) stored at a different location (not only on the NIS) need to be linked to the specimen. The main limitation of the latter technique is that it requires extra time while specifying the multiple image references for numerous specimens. However, this method efficiently fits not only the DPS, but also the other project's needs in which the reference to the remote images in STARLIMS is necessary.

(3) From the user's perspective, the linkage between the specimen and the microscopic images provides fast and simple navigation within the different applications. While viewing the specimen information within STARLIMS, a pathologist can view microscopic images with the accompanying metadata that corresponds to the specimen, and load the microscope image viewing application directly from the LIMS. Until now, no other solution has been described for the integration of a WSI system into STARLIMS.

(4) The developed interfacing solution provides unidirectional data exchange from DPS to LIMS. Because the researchers only need access to the digital images and accompanying metadata, they do not need more information regarding the specimen. For this reason, a bidirectional data exchange mechanism is not integrated into the digital workflow. As the DPS operates independently and simultaneously interacts with the LIMS, it provides effective tools for the biobank researchers who view images without access to the LIMS, as well as the users who manage the biobank (e.g., pathologists) via simple navigation from the LIMS to the DPS viewer. Bidirectional data exchange is more useful for diagnostic purposes when the

pathologist needs to receive specific information regarding the specimen from the LIMS while viewing the microscopic images in the DP viewer.

By developing the middleware application, the quality of the targeted systems and services are improved. The interfacing solution provides a fast and reliable method for data exchange and works as a bridge between the DPS and laboratory software. Even though both systems are developed by different programming languages, the middleware application works successfully with the web services. Pathologists can save time and effort by running the DP viewer directly from the LIMS. In addition, because the interfacing solution can link images to the specimens automatically, it increases both the productivity and effectiveness of these systems.

## **6.2 Advantages and limitations of the developed solution**

The implementation of the digital workflow within the biobank research network will improve laboratory processes and facilitate communication between researchers. Additionally, virtual access to the microscopic images of specimens will clearly improve ergonomics as well. In an economic context, remote access to the images will reduce the cost of shipping slides. By integrating a digital workflow into the biobank research project, standard and repetitive tasks are automated to a great extent. The described approach can significantly reduce workload and eliminate specimen handling errors, which in turn increases the quality of the collaborative research project.

A uniform labeling practice is vital for the biorepository, and it will likely enhance data management and tracking for tissue samples. On the other hand, the barcode label plays a crucial role for data exchange between the various systems via unique identifiers of pathology materials. Additionally, via the use of the DP tools, specimen information is enriched with imaging data that increases biobank efficiency.

Even though several commercial LIMS solutions and the image handling systems with their accompanying interfacing solutions are available on the market, the ultimate goal of this research was to utilize the existing scanning equipment at UMG and find an interfacing solution for the already published IT infrastructure of the KKNMS. Thus, by using the advanced features of the systems and implementing new requirements, IT resources were effectively used and adapted to meet research needs. The high cost related to the DP was significantly

reduced and the developed solution was operated successfully without fundamental changes of information systems used for the KKNMS.

The limitations of this study need to be considered. First, the solution presented here did not evaluate the use of the digital scanners of different vendors. However, the current DPS supports open file formats, such as TIFF, JPEG, and others. Second, the underlying workflow resemble the MS-BB project. It is expected that a web-based DPS will probably become a useful product to facilitate collaboration among the KKNMS researchers. Different projects may have distinct workflows and thus different needs, but the goal was to implement a working prototype that can later be extended to other projects.

### **6.3 Outlook and future development**

As the MS-BB is one of the main parts of the ongoing KKNMS research project, additional tools and features will be required to integrate into the DP workflow in the following years. Within the scope of the KKNMS collaborative research project, more image metadata, such as annotations or any other characteristic of images, could be retrieved from the DPS and integrated into the LIMS using advanced queries. Currently, data sharing is performed only in one direction (from the DPS to the LIMS). The described DPS has the potential to improve the data sharing process and integrate a bidirectional data transfer method. Furthermore, since the interfacing framework is already developed for the DPS and the LIMS, a microscopic image handling system is likely to interact with various information systems within KKNMS's complex IT infrastructure, such as the radiology image management system. A collaboration of researchers from various disciplines will enhance the quality of services.

Currently, the digital images are remotely accessed by KKNMS researchers, and therefore it would be beneficial to request a desired specimen directly from the image viewer while viewing the digital images. Thus, the workflow of the biobank will be improved, increasing the quality of communication between the researchers and administrators of the biobank. Furthermore, a glass slide identifier was purposely designed with a 3-digit suffix (from IMG\_001 to IMG\_999), which could be effectively used for the special construction of tissue samples. As the 3D visualization of the tissue sample is generally constructed with a stack of layers (glass slides), it can be carried out easily, since the glass slides are already correctly organized and identified. Thus, the identification mechanism, which cannot be easily changed

in the future, was prepared in advance to meet the most likely features (e.g., 3D visualization) that might be needed.

The digital microscope imaging infrastructure contains interlinked components: (1) a scanning system managed with effective automated processes, (2) an image storage environment for saving multiple images with the accompanying metadata, (3) a web viewer to access, view, analyze, and annotate high-resolution images from remote locations, and (4) the middleware application that automatically links digital images to the corresponding specimens into the LIMS, by which a pathologist can view both the specimen information and related imaging data in STARLIMS. Thus, the developed solution describes the complete workflow chain starting from the slide preparation for digitization and ending with the application using the advanced features. As the DPS also operates independently from the LIMS, the described solution already applied to the KKNMS research project can be used for different purposes in the same clinical environment. It well suits a standard pathology laboratory workflow that requires the scanning and saving of multiple glass slides as proprietary or standard file formats on the image server and view them via the web viewer. In addition, if it is required that the digital images be integrated into any other LIMS solution that supports web service technologies such as SOAP or REST, the middleware application is likely to be modified and adapted according to the requirements. Thus, the integrated DPS described above has the potential for further development both in the existing biobank research environment and in other research projects within the UMG as well.

The described concept provides a better architecture for the digital pathology system at UMG compared to the initial state several years ago. In order to improve the DP environment in the UMG, more and more digital tools (e.g., digital image measurement and analysis tools, viewing of digital images on the screen, annotation tools, etc.) should be involved in the various applications such as research projects, education, and remote consultations. The multiple use of digital tools in the same clinical environment leads to the successful implementation of the DP in primary diagnosis, which can change the future of pathology.

## References

1. Parwani AV, Hassell L, Glassy E, Pantanowitz L. Regulatory barriers surrounding the use of whole slide imaging in the United States of America. *Journal of Pathology Informatics* 2014; 5(1):38.
2. US Food and Drug Administration (FDA) - Center for Devices and Radiological Health. Medical devices - "What does it mean for FDA to classify a medical device?" [cited 2018 May 11]. Available from: URL:<https://www.fda.gov/AboutFDA/Transparency/Basics/ucm194438.htm>.
3. US Food and Drug Administration (FDA) - Center for Devices and Radiological Health. Technical performance assessment of digital pathology whole slide imaging devices: draft guidance for industry and Food and Drug Administration staff: Guidance for Industry and Food and Drug Administration Staff; 2015.
4. Abels E, Pantanowitz L. Current State of the Regulatory Trajectory for Whole Slide Imaging Devices in the USA. *Journal of Pathology Informatics* 2017; 8:23.
5. Press Release. DPA Recommends Whole Slide Imaging Manufacturers Submit De Novo Applications to the FDA for Primary Diagnosis in the United States: Collaboration with FDA Results in Opportunity for Streamlined Approval Process; 2017 [cited 2018 May 11]. Available from: URL:<http://www.prweb.com/releases/2016/01/prweb13163307.htm>.
6. Tetu B, Evans A. Canadian licensure for the use of digital pathology for routine diagnoses: one more step toward a new era of pathology practice without borders. *Archives of Pathology and Laboratory Medicine* 2014; 138(3):302–4.
7. Thorstenson S, Molin J, Lundstrom C. Implementation of large-scale routine diagnostics using whole slide imaging in Sweden: Digital pathology experiences 2006–2013. *Journal of Pathology Informatics* 2014; 5(1):14.
8. Allen TC. Digital pathology and federalism. *Archives of Pathology and Laboratory Medicine* 2014; 138(2):162–5.
9. Pantanowitz L, Sinard JH, Henricks WH, Fatheree LA, Carter AB, Contis L et al. Validating whole slide imaging for diagnostic purposes in pathology: guideline from the College of American Pathologists Pathology and Laboratory Quality Center. *Archives of Pathology and Laboratory Medicine* 2013; 137(12):1710–22.
10. Long RE, Smith A, Machotka SV, Chlipala E, Cann J, Knight B et al. Scientific and Regulatory Policy Committee (SRPC) paper: validation of digital pathology systems in the regulated nonclinical environment. *Toxicologic Pathology* 2013; 41(1):115–24.
11. Bernard C, Chandrakanth SA, Cornell IS, Dalton J, Evans A, Garcia BM et al. Guidelines from the Canadian Association of Pathologists for establishing a telepathology service for anatomic pathology using whole-slide imaging. *Journal of Pathology Informatics* 2014; 5(1):15.
12. Lowe J. Telepathology: Guidance from The Royal College of Pathologists 2013 [cited 2018 May 11]. Available from: URL:<https://www.rcpath.org/resourceLibrary/telepathology-guidance-from-the-royal-college-of-pathologists-pdf.html>.
13. Garcia-Rojo M, Goncalves L, Blobel B. The COST Action IC0604 "Telepathology Network in Europe" (EURO-TELEPATH). *Studies in Health Technology and Informatics* 2012; 179:3–12.
14. The Royal College of Pathologists of Australasia (RCPA). Guidelines for Digital Microscopy in Anatomical Pathology and Cytology 2015 [cited 2018 May 11]; version 1.0. Available from: URL:<https://www.rcpa.edu.au/getattachment/57bcae98-8d47-4589-99e0-d061c7445c8f/Guidelines-for-Digital-Microscopy-in-Anatomical-Pa.aspx>.
15. Singh R, Chubb L, Pantanowitz L, Parwani A. Standardization in digital pathology: Supplement 145 of the DICOM standards. *Journal of Pathology Informatics* 2011; 2:23.
16. Daniel C, Booker D, Beckwith B, Della Mea V, Garcia-Rojo M, Havener L et al. Standards and specifications in pathology: image management, report management and terminology. *Studies in Health Technology and Informatics* 2012; 179:105–22.
17. Integrating the Healthcare Enterprise - Anatomic Pathology Committee. Anatomic Pathology Technical Framework, Volume 1: Profiles [cited 2018 May 11]. Available from: URL:[http://www.ihe.net/Technical\\_Framework/upload/IHE\\_PAT\\_TF\\_Rev2-0\\_Vol1\\_TI\\_2010-07-23.pdf](http://www.ihe.net/Technical_Framework/upload/IHE_PAT_TF_Rev2-0_Vol1_TI_2010-07-23.pdf).
18. Integrating the Healthcare Enterprise - Anatomic Pathology Committee. Anatomic Pathology Technical Framework, Volume 2: Transactions [cited 2018 May 11]. Available from: URL:[https://www.ihe.net/Technical\\_Framework/upload/IHE\\_PAT\\_TF\\_Rev2-0\\_Vol2\\_TI\\_2010-07-23.pdf](https://www.ihe.net/Technical_Framework/upload/IHE_PAT_TF_Rev2-0_Vol2_TI_2010-07-23.pdf).
19. Hedvat CV. Digital microscopy: past, present, and future. *Archives of Pathology and Laboratory Medicine* 2010; 134(11):1666–70.
20. European Parliament. General Data Protection Regulation; 2016 [cited 2018 May 11]. Available from: URL:<http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN>.

21. European Commission. Protection of personal data; 2011 [cited 2018 May 11]. Available from: URL:<http://ec.europa.eu/justice/data-protection/>.
22. Bundesdatenschutzgesetz (BDSG): Germany's federal data protection act [cited 2018 May 11]. Available from: URL:[https://www.gesetze-im-internet.de/bundesrecht/bdsg\\_1990/gesamt.pdf](https://www.gesetze-im-internet.de/bundesrecht/bdsg_1990/gesamt.pdf).
23. Sucaet Y, Waelput W. Digital Pathology's Past to Present. In: Sucaet Y, Waelput W, editors. Digital pathology. New York: Springer; 2014. p. 1–13 [SpringerBriefs in Computer Science].
24. Hahn HK, Harz MT, Seyffarth H, Zöhrer F, Böhler T, Filippatos K et al. Concepts for Efficient and Reliable Multimodal Breast Image Reading. In: Martí J, Oliver A, Freixenet J, Martí R, editors. Digital Mammography: 10th International Workshop, IWDM 2010, Girona, Catalonia, Spain, June 16-18, 2010. Proceedings. Berlin, Heidelberg: Springer Berlin Heidelberg; 2010. p. 121–8 Available from: URL:[https://doi.org/10.1007/978-3-642-13666-5\\_17](https://doi.org/10.1007/978-3-642-13666-5_17).
25. Krupinski EA. Optimizing the pathology workstation "cockpit": Challenges and solutions. *Journal of Pathology Informatics* 2010; 1(1):19.
26. Rojo MG, García GB, Mateos CP, García JG, Vicente MC. Critical comparison of 31 commercially available digital slide systems in pathology. *International Journal of Surgical Pathology* 2006; 14(4):285–305.
27. Pantanowitz L, Valenstein PN, Evans AJ, Kaplan KJ, Pfeifer JD, Wilbur DC et al. Review of the current state of whole slide imaging in pathology. *Journal of Pathology Informatics* 2011; 2(1):36.
28. Ho J, Ahlers SM, Stratman C, Aridor O, Pantanowitz L, Fine JL et al. Can digital pathology result in cost savings? A financial projection for digital pathology implementation at a large integrated health care organization. *Journal of Pathology Informatics* 2014; 5(1):33.
29. Gabril MY, Yousef GM. Informatics for practicing anatomical pathologists: marking a new era in pathology practice. *Modern Pathology* 2010; 23(3):349–58.
30. Ellin J, Haskvitz A, Premraj P, Shields K, Smith M, Stratman C et al. Interoperability between Anatomic pathology laboratory information systems and digital pathology systems; Digital Pathology Association; 2010. Available from: URL:[https://digitalpathologyassociation.org/\\_data/files/Interoperability\\_Between\\_Anatomic\\_Pathology\\_Laboratory\\_Information\\_Systems\\_and\\_Digital\\_Pathology\\_Systems.pdf](https://digitalpathologyassociation.org/_data/files/Interoperability_Between_Anatomic_Pathology_Laboratory_Information_Systems_and_Digital_Pathology_Systems.pdf).
31. Technology, Methods, and Infrastructure (TMF). IT in medical research: There are already a number of effective solutions for data capture – and no need to develop new applications in-house: IT Review Board's 2012/2013 report [cited 2018 May 11]. Available from: URL:<http://www.tmf-ev.de/Produkte/ITReport.aspx>.
32. Manghani K. Quality assurance: Importance of systems and standard operating procedures. *Perspectives in clinical research* 2011; 2(1):34–7.
33. Surić Mihić M, Meštrović T, Prlić I, Surić D. Importance of quality assurance program implementation in conventional diagnostic radiology. *Collegium Antropologicum* 2008; 32(2):181–4.
34. Pohl K. Requirements engineering: fundamentals, principles, and techniques. Berlin, Heidelberg: Springer Publishing Company, Incorporated; 2010.
35. Finkelstein A, Kramer J, Nuseibeh B, Finkelstein L, Goedicke M. Viewpoints: A framework for integrating multiple perspectives in system development. *International Journal of Software Engineering and Knowledge Engineering* 1992; 2(01):31–57.
36. Akkermans JM, Gordijn J. Value-based requirements engineering: Exploring innovative e-commerce ideas. *Requirements Engineering* 2003; 8(2):114–34.
37. Brown RW, Della Speranza V, Alvarez JO, Eisen RN, Frishberg DP, Rosai J et al. Uniform labeling of blocks and slides in surgical pathology: guideline from the College of American Pathologists Pathology and Laboratory Quality Center and the National Society for Histotechnology. *Archives of Pathology and Laboratory Medicine* 2015; 139(12):1515–24.
38. Nussbeck SY, Skrownny D, O'Donoghue S, Schulze TG, Helbing K. How to design biospecimen identifiers and integrate relevant functionalities into your biospecimen management system. *Biopreservation and Biobanking* 2014; 12(3):199–205.
39. Guzman M, Judkins AR. Digital pathology: A tool for 21st century neuropathology. *Brain Pathology* 2009; 19(2):305–16.
40. Kulikowski C, Ammenwerth E, Bohne A, Ganser K, Haux R, Knoup P et al. Medical imaging informatics and medical informatics: Opportunities and constraints. *Methods of Information in Medicine* 2002; 41(2):183–9.
41. Zhu HJ, Han BC, Qiu B. Survey of astronomical image processing methods. In: Zhang Y-J, editor. Image and Graphics: 8th International Conference, ICG 2015, Tianjin, China, August 13-16, 2015, Proceedings, Part III. Cham: Springer International Publishing; 2015. p. 420–9 Available from: URL:[http://dx.doi.org/10.1007/978-3-319-21969-1\\_37](http://dx.doi.org/10.1007/978-3-319-21969-1_37).
42. Gonzalez RC, Woods RE, Eddins SL. Digital image processing using MATLAB. New Jersey: Pearson Prentice Hall; 2003.

43. U.S. National Library of Medicine, National Institutes of Health. Definitions: Biomedical Imaging; 2018 [cited 2018 May 11]. Available from: URL:<https://www.nlm.nih.gov/tsd/acquisitions/cdm/subjects15.html#1076792>.
44. Rubin DL, Greenspan H, Brinkley JF. Biomedical Imaging Informatics. In: Shortliffe EH, Cimino JJ, editors. Biomedical informatics: Computer Applications in Health Care and Biomedicine. 4th edition. London: Springer London; 2014. p. 285–327.
45. Burger W, Burge MJ. Digital Images. In: Burger W, Burge M, editors. Principles of digital image processing: Fundamental techniques. New York, London: Springer; 2009. p. 1–24 [Undergraduate topics in computer science, 1863-7310; vol. 7592].
46. ImageJ. Image analysis - What is a digital image; 2017 [cited 2018 May 11]. Available from: URL:<https://sites.google.com/site/learnimagej/image-processing/what-is-a-digital-image>.
47. Johnson J, Ballan L, Fei-Fei L. Love thy neighbors: image annotation by exploiting image metadata. In: 2015 IEEE International Conference on Computer Vision: 11-18 December 2015, Santiago, Chile:proceedings. Piscataway, NJ: IEEE; 2015. p. 4624–32.
48. Korenblum D, Rubin D, Napel S, Rodriguez C, Beaulieu C. Managing biomedical image metadata for search and retrieval of similar images. *Journal of Digital Imaging* 2011; 24(4):739–48.
49. Kroon-Batenburg LMJ, Helliwell JR. Experiences with making diffraction image data available: what metadata do we need to archive? *Acta Crystallographica. Section D, Biological Crystallography* 2014; 70(Pt 10):2502–9.
50. Linkert M, Rueden CT, Allan C, Burel J-M, Moore W, Patterson A et al. Metadata matters: access to image data in the real world. *The Journal of Cell Biology* 2010; 189(5):777–82.
51. Varma DR. Managing DICOM images: Tips and tricks for the radiologist. *The Indian Journal of Radiology and Imaging* 2012; 22(1):4–13.
52. Deserno TM. Fundamentals of Biomedical Image Processing. In: Deserno TM, editor. Biomedical image processing. Heidelberg: Springer; 2011. p. 1–51 [Biological and Medical Physics, Biomedical Engineering].
53. Barucci E, Brlek S, Brocchi S. PCIF: An algorithm for lossless true color image compression. In: Barneva RP, Wiederhold P, editors. Combinatorial image analysis: 13th international workshop, IWCIA 2009, Playa Del Carmen, Mexico, November 24-27, 2009, proceedings / Petra Wiederhold, Reneta P. Barneva, eds. Berlin: Springer-Verlag; 2009. p. 224–37 [Lecture notes in computer science, 0302-9743; vol. 5852].
54. Wallace GK. The JPEG still picture compression standard. *IEEE transactions on Consumer Electronics* 1992; 38(1):xviii–xxxiv.
55. Wiggins RH3, Davidson HC, Harnsberger HR, Lauman JR, Goede PA. Image file formats: past, present, and future. *Radiographics: a review publication of the Radiological Society of North America, Inc* 2001; 21(3):789–98.
56. World Wide Web Consortium (W3C). Portable Network Graphics (PNG) functional specification (Second Edition); 2003 [cited 2018 May 11]. Available from: URL:<https://www.w3.org/TR/PNG/>.
57. CompuServe Incorporated. Graphics Interchange Format (GIF) Specification: Version 89a; 1990 [cited 2018 May 11]. Available from: URL:<https://www.w3.org/Graphics/GIF/spec-gif89a.txt>.
58. Digital Imaging and Communications in Medicine (DICOM) Standards Committee. Data structures and encoding - Chapter 7: The data set; 2013 [cited 2018 May 11]. Available from: URL:[ftp://dicom.nema.org/medical/DICOM/2013/output/chtml/part05/chapter\\_7.html](ftp://dicom.nema.org/medical/DICOM/2013/output/chtml/part05/chapter_7.html).
59. DICOM Standards Committee, Working Group 26, Pathology. Supplement 145: Whole Slide Microscopic Image IOD and SOP Classes; 2006. Available from: URL:[ftp://medical.nema.org/medical/dicom/final/sup145\\_ft.pdf](ftp://medical.nema.org/medical/dicom/final/sup145_ft.pdf).
60. Neuroimaging Informatics Technology Initiative (Nifti). Background Information [cited 2018 May 11]. Available from: URL:<https://nifti.nimh.nih.gov/background>.
61. Larobina M, Murino L. Medical image file formats. *Journal of Digital Imaging* 2014; 27(2):200–6.
62. Goldberg IG, Allan C, Burel J-M, Creager D, Falconi A, Hochheiser H et al. The Open Microscopy Environment (OME) Data Model and XML file: open tools for informatics and quantitative analysis in biological imaging. *Genome Biology* 2005; 6(5):R47.
63. The Open Microscopy Environment (OME). The OME-TIFF format specification; 2017 [cited 2018 May 11]. Available from: URL:<https://www.openmicroscopy.org/site/support/ome-model/ome-tiff/>.
64. Erickson B, Greenes RA. Imaging systems in radiology. In: Shortliffe EH, Cimino JJ, editors. Biomedical informatics: Computer Applications in Health Care and Biomedicine. 4th edition. London: Springer London; 2014. p. 593–611.
65. Punys V, Laurinavicius A, Puniene J. A data model for handling whole slide microscopy images in picture archiving and communications systems. *Studies in Health Technology and Informatics* 2009; 150:856–60.
66. Collins TJ. ImageJ for microscopy. *Biotechniques* 2007; 43(1 Suppl):25–30.

67. Allan C, Burel J-M, Moore J, Blackburn C, Linkert M, Loynton S et al. OMERO: Flexible, model-driven data management for experimental biology. *Nature Methods* 2012; 9(3):245–53.
68. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T et al. Fiji: An open-source platform for biological-image analysis. *Nature Methods* 2012; 9(7):676–82.
69. Wienert S, Heim D, Saeger K, Stenzinger A, Beil M, Hufnagl P et al. Detection and segmentation of cell nuclei in virtual microscopy images: A minimum-model approach. *Scientific reports* 2012; 2:503.
70. Latson L, Sebek B, Powell KA. Automated cell nuclear segmentation in color images of hematoxylin and eosin-stained breast biopsy. *Analytical and Quantitative Cytology and Histology/the International Academy of Cytology [and] American Society of Cytology* 2003; 25(6):321–31.
71. Oberholzer M, Ostreicher M, Christen H, Brühlmann M. *Methods in quantitative image analysis. Histochemistry and Cell Biology* 1996; 105(5):333–55.
72. Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O et al. CellProfiler: Image analysis software for identifying and quantifying cell phenotypes. *Genome Biology* 2006; 7(10):R100.
73. Esgiar AN, Naguib RNG, Sharif BS, Bennett MK, Murray A. Microscopic image analysis for quantitative measurement and feature identification of normal and cancerous colonic mucosa. *IEEE Transactions on Information Technology in Biomedicine* 1998; 2(3):197–203.
74. Choi H, Bovik AC. *Wavelet Image Processing*. In: Wu Q, Merchant FA, Castleman KR, editors. *Microscope image processing*. London: Academic; 2008. p. 79–111.
75. Sucaet Y, Waelpuut W. *Use Cases*. In: Sucaet Y, Waelpuut W, editors. *Digital pathology*. New York: Springer; 2014. p. 57–69 [SpringerBriefs in Computer Science].
76. Hamilton PW, Wang Y, McCullough SJ. Virtual microscopy and digital pathology in training and education. *APMIS : acta pathologica, microbiologica, et immunologica Scandinavica* 2012; 120(4):305–15.
77. Al-Janabi S, Huisman A, van Diest PJ. Digital pathology: Current status and future perspectives. *Histopathology* 2012; 61(1):1–9.
78. Ferreira R, Moon B, Humphries J, Sussman A, Saltz J, Miller R et al. The virtual microscope. In: *Proceedings of the AMIA Annual Fall Symposium*; 1997. p. 449.
79. Felten CL, Strauss JS, Okada DH, Marchevsky AM. Virtual microscopy: High resolution digital photomicrography as a tool for light microscopy simulation. *Human Pathology* 1999; 30(4):477–83.
80. Kayser K, Kayser G, Radziszowski D, Oehmann A. From telepathology to virtual pathology institution: the new world of digital pathology. *Rom J Morphol Embryol* 1999; 45:3–9.
81. Kayser K. History of the European Conference Series on digital pathology: memories and perspectives. *Diagnostic Pathology* 2016; 2(1).
82. Pantanowitz L, J Mark Tuthill MD, Balis U. Telepathology. In: Pantanowitz L, Tuthill JM, Balis UJ, editors. *Pathology informatics: Theory & practice*. Chicago: American Society for Clinical Pathology Press; 2012. p. 261–72.
83. Maiolino P, Vico GD. Telepathology in veterinary diagnostic cytology. In: Kumar S, Dunn BE, editors. *Telepathology*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2009. p. 63–9 Available from: URL:[http://dx.doi.org/10.1007/978-3-540-85786-0\\_6](http://dx.doi.org/10.1007/978-3-540-85786-0_6).
84. Lowe A, Chlipala E, Elin J, Kawano Y, Long RE, Tillman D. Validation of digital pathology in a healthcare environment. San Diego, CA: Digital Pathology Association 2011.
85. Evans AJ, Krupinski EA, Weinstein RS, Pantanowitz L. 2014 American Telemedicine Association clinical guidelines for telepathology: Another important step in support of increased adoption of telepathology for patient care. *Journal of Pathology Informatics* 2015; 6:13.
86. García-Rojo M, Blobel B, Laurinavicius A. Perspectives on digital pathology: results of the COST Action IC0604 EURO-TELEPATH. Amsterdam: IOS Press; 2012.
87. Stathonikos N, Veta M, Huisman A, van Diest PJ. Going fully digital: Perspective of a Dutch academic pathology lab. *Journal of Pathology Informatics* 2013; 4:15.
88. Kayser K, Kayser G, Radziszowski D, Oehmann A. From telepathology to virtual pathology institution: the new world of digital pathology. *Romanian Journal of Morphology and Embryology/Revue roumaine de morphologie et embryologie* 1999-2004; 45:3–9.
89. Tuomari DL, Kemp RK, Sellers R, Yarrington JT, Geoly FJ, Fouillet XLM et al. Society of Toxicologic Pathology position paper on pathology image data: compliance with 21 CFR Parts 58 and 11. *Toxicologic Pathology* 2007; 35(3):450–5.
90. Food and Drug Administration (FDA) - Department of Health and Human Services. Good laboratory practice for nonclinical laboratory studies. Available from: URL:<https://www.fda.gov/downloads/AboutFDA/ReportsManualsForms/Reports/EconomicAnalyses/UCM518224.pdf> [cited 2018 May 11].

91. The Food and Drug Administration. Hematology and Pathology Devices; Subpart: Pathology Instrumentation and Accessories; Sec. 864.3600; Microscopes and accessories; Code of Federal Regulations (CFR); 21 CFR 864.3600 [cited 2018 May 11]. Available from:  
URL:<https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=864.3600>.
92. European Commission. The rules governing medicinal products for human and veterinary use in the European Union; Volume 4: Good manufacturing practice guidelines; 2016 [cited 2018 May 11]. Available from:  
URL:[https://ec.europa.eu/health/documents/eudralex/vol-4\\_en](https://ec.europa.eu/health/documents/eudralex/vol-4_en).
93. Royal Philips. Philips digital pathology solution receives CE Mark for diagnosis to help advance oncology diagnosis [cited 2018 May 11]. Available from: URL:<http://www.philips.com/a-w/about/news/archive/standard/news/press/2014/20141016-Philips-Digital-Pathology-Solution-Receives-CE-Mark-for-Diagnosis-to-Help-Advance-Oncology-Diagnosis.html#.VsGXOubrrOB>.
94. GE Healthcare. Omnyx Integrated Digital Pathology new high volume scanner achieves CE Mark in the European Union; 2018 [cited 2018 May 11]. Available from: URL:<http://www.genewsroom.com/Press-Releases/Omnyx-Integrated-Digital-Pathology-new-high-volume-scanner-achieves-CE-Mark-in-the-European-Union-276222>.
95. Sectra. Sectra Digital Pathology Solution CE marked [cited 2018 May 11]. Available from:  
URL:[http://www.sectra.com/press\\_releases/all/en/20150101-20151231/item/\(1922934\)](http://www.sectra.com/press_releases/all/en/20150101-20151231/item/(1922934)).
96. Rojo MG, Daniel C, Schrader T. Standardization efforts of digital pathology in Europe. *Analytical Cellular Pathology (Amsterdam)* 2012; 35(1):19–23.
97. FDA - Food and Drug Administration. Picture archiving and communications system; Radiology Devices; Code of Federal Regulations (CFR) Title 21; 21CFR892.2050;CFR [cited 2018 May 11]. Available from:  
URL:<https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/cfrsearch.cfm?fr=892.2050>.
98. FDA - Food and Drug Administration. Medical Image Communications Device; Radiology Devices; Code of Federal Regulations (CFR); 21CFR892.2020; [cited 2018 May 11]. Available from:  
URL:<https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=892.2020>.
99. FDA - Food and Drug Administration. Medical Image Hardcopy Device; Radiology Devices; Code of Federal Regulations (CFR) Title 21; 21CFR892.2040 [cited 2018 May 11]. Available from:  
URL:<https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=892.2040>.
100. Svanadze L, Franke T, Buckow K, Bahn E, Rienhoff O. Requirements for integration of microscope images into the laboratory information management system of a research network. *Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (GMD5)*; German Medical Science GMS Publishing House 2014.
101. HL7Wiki. Anatomic Pathology Work Group; 2018 [cited 2018 May 11]. Available from:  
URL:[http://wiki.hl7.org/index.php?title=Anatomic\\_Pathology\\_Work\\_Group](http://wiki.hl7.org/index.php?title=Anatomic_Pathology_Work_Group).
102. Daniel C, Macary F, Rojo MG, Klossa J, Laurinavičius A, Beckwith BA et al. Recent advances in standards for collaborative Digital Anatomic Pathology. *Diagnostic Pathology* 2011; 6(1):S17.
103. Universitätsmedizin Göttingen. Institut für Medizinische Informatik [cited 2018 May 11]. Available from:  
URL:<http://www.mi.med.uni-goettingen.de/>.
104. Prokosch HU, Beck A, Ganslandt T, Hummel M, Kiehntopf M, Sax U et al. IT Infrastructure Components for Biobanking. *Applied Clinical Informatics* 2010; 1(4):419–29.
105. Technology, Methods, and Infrastructure (TMF). Topics and Projects: Improving the infrastructure for medical research [cited 2018 May 11]. Available from: URL:<http://www.tmf-ev.de/EnglishSite/Topics.aspx>.
106. Kiehntopf M, Böer KW. Biomaterialbanken – Checkliste zur Qualitätssicherung. Berlin: Medizinisch-Wissenschaftliche Verl.-Ges; 2008. (Schriftenreihe der Telematikplattform für Medizinische Forschungsnetze (TMF); vol 5).
107. IT-Reviewing-Board der TMF – Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V. IT-Infrastrukturen in der patientenorientierten Forschung: Aktueller Stand und Handlungsbedarf - 2016 [Jährliche Berichte des IT-Reviewing Board der TMF]. 1. Auflage. Berlin: Akademische Verlagsgesellschaft AKA; 2016.
108. Bauer CR, Umbach N, Baum B, Buckow K, Franke T, Grütz R et al. Architecture of a biomedical informatics research data management pipeline. *Studies in Health Technology and Informatics* 2016; 228:262.
109. Competence network MS: Common IT structure for MS research [cited 2018 May 11]. Available from:  
URL:<http://www.kompetenznetz-multiplesklerose.de/en/research/infrastructure/it/>.
110. Höer A, Schiffhorst G, Zimmermann A, Fischaleck J, Gehrmann L, Ahrens H et al. Multiple sclerosis in Germany: Data analysis of administrative prevalence and healthcare delivery in the statutory health system. *BMC Health Services Research* 2014; 14(1):381.
111. Kingwell E, Marriott JJ, Jetté N, Pringsheim T, Makhani N, Morrow SA et al. Incidence and prevalence of multiple sclerosis in Europe: A systematic review. *BMC neurology* 2013; 13(1):128.

112. Kotonya G, Sommerville I. Requirements engineering: processes and techniques. New York: John Wiley & Sons LTD; 1998.
113. Rozanski N, Woods E. A viewpoint catalog. In: Software systems architecture: Working with stakeholders using viewpoints and perspectives. Upper Saddle River, New Jersey: Addison-Wesley; 2011.
114. Nuseibeh B, Easterbrook S. Requirements engineering: a roadmap. In: Proceedings of the Conference on the Future of Software Engineering; 2000. p. 35–46.
115. PathXL. Digital Pathology [cited 2018 May 11]. Available from: URL:<http://www.pathxl.com/>.
116. Imagescope: Leica Biosystems [cited 2018 May 11]. Available from: URL:<http://www.leicabiosystems.com/digital-pathology/digital-pathology-management/imagescope/>.
117. VMscope GmbH - Startseite; 2018 [cited 2018 May 11]. Available from: URL:<http://www.vmscope.de/>.
118. CHILI Solutions. Software components for radiology and other fields of medicine involving image processing [cited 2018 May 11]. Available from: URL:<http://www.chili-radiology.com/en/>.
119. ZEISS Group. Slide Scanner Axio Scan.Z1 for Brightfield and Fluorescence [cited 2018 May 11]. Available from: URL:<https://www.zeiss.com/microscopy/int/products/imaging-systems/axio-scan-z1.html>.
120. Thomas Kühne. What is a Model? In: Language Engineering for Model-Driven Software Development. Dagstuhl, Germany: Internationales Begegnungs- und Forschungszentrum für Informatik (IBFI), Schloss Dagstuhl, Germany; 2005 Available from: URL:<http://drops.dagstuhl.de/opus/volltexte/2005/23>.
121. Pohl K, Rupp C. Requirements validation and negotiation. In: Pohl K, Rupp C, editors. Requirements engineering fundamentals: A study guide for the certified professional for requirements engineering exam-foundation level-IREB compliant. 1st ed. Santa Barbara, Calif.: Rocky Nook; 2011 [Rocky Nook computing].
122. Wang M, Zeng Y. Asking the right questions to elicit product requirements. International Journal of Computer Integrated Manufacturing 2009; 22(4):283–98.
123. Miller RE. The Quest for Software Requirements. Milwaukee: MavenMark Books; 2009.
124. Agamanolis D. Neuropathology - illustrated, interactive course for medical students and residents: Chapter 6 - Demyelinative diseases [cited 2018 May 11]. Available from: URL:<http://neuropathology-web.org/chapter6/chapter6aMs.html>.
125. Ferguson B, Matyszak MK, Esiri MM, Perry VH. Axonal damage in acute multiple sclerosis lesions. Brain: A Journal of Neurology 1997; 120 (Pt 3):393–9.
126. Eloyan A, Shou H, Shinohara RT, Sweeney EM, Nebel MB, Cuzzocreo JL et al. Health effects of lesion localization in multiple sclerosis: spatial registration and confounding adjustment. PLoS one 2014; 9(9):e107263.
127. Lucchinetti CF, Gavrilova RH, Metz I, Parisi JE, Scheithauer BW, Weigand S et al. Clinical and radiographic spectrum of pathologically confirmed tumefactive multiple sclerosis. Brain: A Journal of Neurology 2008; 131(Pt 7):1759–75.
128. The BrainNet Europe consortium. Scientific Community; 2016 [cited 2018 May 11]. Available from: URL:<http://www.brainnet-europe.org/>.
129. Abbott Informatics. STARLIMS - LIMS Solution [cited 2018 May 11]. Available from: URL:<https://www.informatics.abbott/int/en/home>.
130. Fan W, Bifet A. Mining big data: Current status, and forecast to the future. SIGKDD Explor. Newsl. 2013; 14(2):1–5.
131. Real-Time Innovations (RTI). Interoperable Open Architecture (IOA): MOD and DoD – Architecting for Interoperability [cited 2018 May 11]. Available from: URL:[https://info.rti.com/hubfs/whitepapers/Interoperable\\_Open\\_Architecture.pdf](https://info.rti.com/hubfs/whitepapers/Interoperable_Open_Architecture.pdf).
132. Center for Devices and Radiological Health. Digital Health - Medical Device Interoperability: Center for Devices and Radiological Health [cited 2018 May 11]. Available from: URL:<https://www.fda.gov/MedicalDevices/DigitalHealth/ucm512245.htm>.
133. The Healthcare Information and Management Systems Society (HIMSS). Health Interoperability Ecosystem [cited 2018 May 11]. Available from: URL:<http://www.himss.org/library/interoperability-standards/what-is-interoperability>.
134. Kubicek H, Cimander R, Scholl HJ. Layers of Interoperability. In: Kubicek H, Cimander R, Scholl J, editors. Organizational interoperability in e-government: Lessons from 77 European good-practice cases. Heidelberg, New York: Springer; 2011. p. 85–96.
135. Snell J, Tidwell D, Kulchenko P. Programming web services with SOAP: building distributed applications. Sebastopol, CA: O'Reilly & Associates, Inc.; 2001.
136. World Wide Web Consortium (W3C). Web Services Architecture; 2004 [cited 2018 May 11]. Available from: URL:<https://www.w3.org/TR/ws-arch/>.
137. World Wide Web Consortium (W3C). W3C XML Schema Definition (XSD) Language 1.1; Part 1: Structures; 2012 [cited 2018 May 11]. Available from: URL:<https://www.w3.org/TR/xmlschema11-1/>.

138. World Wide Web Consortium (W3C). W3C XML Schema; 2016 [cited 2018 May 11]. Available from: URL:<https://www.w3.org/XML/Schema>.
139. Spillner A, Bremenn H. The W-MODEL. Strengthening the bond between development and test. In: International Conference on Software Testing 2002. p. 15–7.
140. Guo H, Birsa J, Farahani N, Hartman DJ, Piccoli A, O'Leary M et al. Digital pathology and anatomic pathology laboratory information system integration to support digital pathology sign-out. *Journal of Pathology Informatics* 2016; 7:23.
141. LABPON Whitepaper - Pathology Laboratory in the Netherlands. How to Go Digital in Pathology [cited 2018 May 11]. Available from: URL:[https://thepathologist.com/fileadmin/issues/App\\_Notes/0016-022-app-note-How\\_to\\_Go\\_Digital\\_in\\_Pathology.pdf](https://thepathologist.com/fileadmin/issues/App_Notes/0016-022-app-note-How_to_Go_Digital_in_Pathology.pdf).
142. Health level 7 (HL7). Anatomic Pathology WG: 2015 January WGM minutes [cited 2018 May 11]. Available from: URL:<http://www.hl7.org/Special/committees/anatomicpath/index.cfm>.
143. Tuominen VJ, Isola J. Linking whole-slide microscope images with DICOM by using JPEG2000 interactive protocol. *Journal of Digital Imaging* 2010; 23(4):454–62.
144. Amin M, Sharma G, Parwani AV, Anderson R, Kolowitz BJ, Piccoli A et al. Integration of digital gross pathology images for enterprise-wide access. *Journal of Pathology Informatics* 2012; 3.
145. CliniSys Group - Diagnostics Intelligence. Integrated Image Management for Pathology [cited 2018 May 11]. Available from: URL:<http://www.clinisysgroup.com/international/solutions/nhs-solutions/image-management/>.
146. DICOM Part 18, Supplement 148. Web access to DICOM persistent objects by means of web Services [cited 2018 May 11]. Available from: URL:[ftp://medical.nema.org/medical/dicom/final/sup148\\_ft.pdf](ftp://medical.nema.org/medical/dicom/final/sup148_ft.pdf).
147. Nakhleh RE, Idowu MO, Souers RJ, Meier FA, Bekeris LG. Mislabeling of cases, specimens, blocks, and slides: a college of american pathologists study of 136 institutions. *Archives of Pathology and Laboratory Medicine* 2011; 135(8):969–74.
148. Balis UJ, Pantanowitz L. Specimen tracking and identification systems. In: Pantanowitz L, Tuthill JM, Balis UJ, editors. *Pathology informatics: Theory & practice*. Chicago: American Society for Clinical Pathology Press; 2012.

*“Perfection is achieved, not when there is nothing more to add,  
but when there is nothing left to take away.”*

*- Antoine de Saint Exupéry*

## **Appendices**

Appendix A. Comparison of the DPA, CAP, Canadian, ATA, and RCP guidelines

Comparison of the DPA, CAP, Canadian, ATA, and RCP guidelines						
Organization	DPA - Digital Pathology Association		CAP - College of American Pathologists	Canadian Association of Pathologists	ATA - American Telemedicine Association	RCP - The Royal College of Pathologists (in UK)
<b>Title</b>						
<b>General</b>						
<b>Active release date</b>	2013	2011	2013	2014	2014	2013
<b>Previous version</b>	No previous version	No previous version	2011 - draft recommendations	No previous version	1999	2005
<b>Guideline title</b>	Validation of Digital Pathology Systems in the Regulated Nonclinical Environment	Validation of Digital Pathology In a Healthcare Environment	Validating Whole Slide Imaging for Diagnostic Purposes in Pathology	Guidelines from the Canadian Association of Pathologists for establishing a telepathology service for anatomic pathology using whole-slide imaging	American Telemedicine Association clinical guidelines for telepathology	Telepathology: Guidance from The Royal College of Pathologists
<b>Authorship</b>	Committee		Expert panel members	Telepathology Guidelines Committee	Expert panel members	Single author
<b>Responsible organizations</b>	Members of the Scientific and Regulatory Policy Committee, Society of Toxicologic Pathology		The College of American Pathologists Pathology and Laboratory Quality Center	The Canadian Association of Pathologists	Work Group of the American Telemedicine Association	Chair of the Specialty Advisory Committee on Cellular Pathology
<b>Definition of key terms</b>	Yes (with comprehensive glossary)		Yes (basic terminology only)	Yes (basic terminology only)	Yes (basic terminology only)	Yes (basic terminology only)
<b>Scope</b>						
<b>Guideline category</b>	Nonclinical applications (e.g. visual evaluation or image analysis)	Clinical applications	Clinical applications	Clinical applications	Clinical applications of telepathology	Clinical applications
<b>Telepathology modalities</b>	All types	All types	Whole slide imaging only	Whole slide imaging only	All types	All types
<b>Guideline objectives</b>	To provide technical aspects for validation of the WSI systems	Quality assurance (QA) and quality control (QC) programs for validation of WSI systems	To recommend validation requirements for WSI systems to be used for diagnostic purposes	To provide Canadian pathologist with the information how to use WSI system for different applications	To help pathologists and assure the uniform quality of service to patients	To use and practice telepathology systems and support clinical practice
<b>List of applications</b>	Nonclinical laboratory (in vivo or in vitro experiments)	Education, research, image analysis, archival and retrieval, LIS/LIMS integration, secondary consultations, and virtual slide sharing	Applications in clinical practice for diagnostic purposes(primary diagnosis, frozen section and consultation)	Intraoperative pathology consultation (frozen section), primary diagnosis, expert or second opinions and quality assurance activities	Primary diagnosis, intraoperative consultation, rapid cytology, secondary consultation, special studies, archival review, quality activities	Expert opinion, case conferences, quality assurance, education and assessment of competence
<b>Excluded issues (Is not provided)</b>	A step by step approach to validation		Education and image analysis applications	Hematopathology, microbiology, tumor boards, education, research applications	Use-cases for research or education purposes	Not mentioned
<b>Methodology</b>						
<b>Methods used to collect/select the evidence</b>	Not mentioned	Based on general guidelines and regulations in DP for validation of WSI systems	Systematic literature review and analysis in electronic databases. Searching terms: whole slide imaging OR virtual or digital microscopy OR digital pathology OR teleconsulation OR telemicroscopy AND validation; 767 international publications met the requirements	Based on available medical literature and the clinical experience of early adopters of telepathology in Canada	Not mentioned	Evidence was collected from the experience in pathology and additionally from the literature in radiology applications (e.g. "A Guidance Document for Clinical Radiologists")
<b>Methods used to formulate the recommendations</b>	System Life Cycle approach for validation	Main recommendations provided by regulatory agencies and vendors	Convening a nonvendor panel (teleconference and face-to-face meeting) 19 times. Central question - "What should be done to validate a whole slide digital imaging system for diagnostic purposes before it is placed in clinical service?"	Recommendations are formulated by working group consisting of multiple pathologists, technologists, and healthcare administrators from across Canada	Recommendations are developed with thorough consensus and rigorous review by experts and strategic stakeholders from different fields. Finally approved by the ATA board of directors	By describing potential benefits and real problems of telepathology
<b>Methods used to analyze the evidence</b>	Phases: Planning, design, system development and testing, system qualification and commissioning, system operation, system retirement	Validation program for each component of WSI system for specific applications (such as primary diagnosis, image analysis)	published studies were extracted, open comment feedback and expert panel consensus were graded for evidence evaluation	not mentioned in detail	The guideline statements are categorised as requirements, recommendations, or actions	-
<b>Methods used to assess the quality of the validation</b>	Quality management system and documentation	Checklists and draft guidelines provided by CAP	Grading system for evidence evaluation, evaluating the correlation between digitized slides and glass slides; At least 60 cases per application used for validation study	Quality management system; documentation and administration of procedures using WSI systems	Not described in detail (Only technical aspects of validation are covered)	-
<b>Recommendations</b>						
<b>Guideline statements</b>	Explanation of validation life cycle	13 guideline statements adopted from the CAP is described in detail	12 guideline statements described in detail	Adopted from the CAP	not listed as in CAPs guideline, but described for each application	Yes (only suggestions)
<b>Grading scheme for the strength of the statements</b>	No	No	Grade A,B,C,D; A,B-'Recommendations'; C-'Suggestions'; D-'Expert Consensus Opinion'	No	Recommendations are identified by the keywords "shall", "should", or "may"	No
<b>Benefits of implementing recommendations</b>	Yes	Yes	Yes	Yes	No	Yes
<b>Challenges during implementing the recommendations</b>	Yes	Yes	No	Yes (no widely accepted best practice guidelines, lack of multiplanar focus adjustments, unfamiliarity with scanners, low level of availability and ineffectiveness, transferring large images)	Yes (cost, legal and regulatory issues, resistance from pathologists, lack of standards, technology barriers such as limited resolution and large image files)	Yes (legal, quality, and communications issues)
<b>Implementation plan developed</b>	Yes	Yes	No	Yes (5 recommendations for on-site; 5 recommendations for remote site)	No	No
<b>Technology</b>						
<b>Details on technical specification</b>	Yes ( hardware, software, server, image acquisition)	Yes ( hardware, software, viewer, LIS, server, image acquisition)	Purely technical components	Yes (scanners, viewers, image management software, image analysis tools)	Yes (image acquisition, displays, transmission, storage)	No
<b>Overview of facility responsibilities</b>	Yes - installation, operational, and performance qualification	Standard operating procedures	Yes	Yes	Yes (technical support, functional verification of equipment, licensure)	Yes
<b>Communication responsibilities</b>	No	Metadata sharing, validation of LIS integrated DP	No	Yes	No	Yes (Among distant pathologist & clinical staff & local laboratory staff; linguistic problems)
<b>Legal liability issues</b>	Yes	Yes	No	Yes	Yes	Yes (confidentiality, jurisdiction, revalidation, liability)
<b>Security/privacy issues</b>	Yes	Yes	Yes	Yes	Yes	Yes
<b>Expressed need for user training</b>	Yes	Yes	Yes	Yes	Yes	Yes
<b>Documentation/Archiving/ Quality management</b>	Yes	Yes	Yes	Yes	Yes	Yes
<b>Summary</b>						
<b>Planned guidelines update</b>	Yes (no specific date)	Yes (no specific date)	Yes (in every 4 years)	Yes (no specific date)	Yes (no specific date)	Yes
<b>Compliance with regulatory bodies</b>	FDA, CMS, CLIA, CAP, state regulations, the Joint Commission	Anatomic Pathology laboratory, FDA,CAP, the Joint Commission, Health and human services, state law, CMS	Not mentioned in detail (only FDA)	CAP	FDA, CMS/CLIA, CAP	No
<b>Guideline availability (links)</b>	<a href="https://digitalpathologyassociation.org/_data/files/DPA_White_Paper_Final_-_2011-11-17.pdf">https://digitalpathologyassociation.org/_data/files/DPA_White_Paper_Final_-_2011-11-17.pdf</a>	<a href="https://digitalpathologyassociation.org/_data/files/DPA-Healthcare-White-Paper-FINAL_v1.0.pdf">https://digitalpathologyassociation.org/_data/files/DPA-Healthcare-White-Paper-FINAL_v1.0.pdf</a>	<a href="http://www.ncbi.nlm.nih.gov/pubmed/23634907">http://www.ncbi.nlm.nih.gov/pubmed/23634907</a>	<a href="http://www.ncbi.nlm.nih.gov/pubmed/24843826">http://www.ncbi.nlm.nih.gov/pubmed/24843826</a>	<a href="http://www.jpathinformatics.org/text.asp?2014/5/1/39/143329">http://www.jpathinformatics.org/text.asp?2014/5/1/39/143329</a>	<a href="https://www.rcpath.org/resource/library/telepathology-guidance-from-the-royal-college-of-pathologists.pdf.html">https://www.rcpath.org/resource/library/telepathology-guidance-from-the-royal-college-of-pathologists.pdf.html</a>

## Appendix B. Use case scenarios for digital pathology workflow

Table B.1. Use case specification for the preparation of glass slides

<b>Scenario 1 (S1) - “Preparation of glass slides”</b>	
<b>Short description</b>	Step-by-step procedures for preparation of glass slides for digitization
<b>Responsible stakeholders</b>	<ul style="list-style-type: none"> <li>• Pathologist (responsible for the preparation of glass slides)</li> <li>• LIMS technician (responsible for managing the interface that exports metadata for the labels of glass slides from the Starlims system)</li> <li>• Label printing company (responsible for label printing via thermal transfer printer)</li> </ul>
<b>Actors</b>	Pathologist and/or research assistant
<b>Goal(s)</b>	To prepare glass slides for digitization: <ul style="list-style-type: none"> <li>• Generate unique IDs and label components for glass slides</li> <li>• Print the labels and attach them on the glass slides</li> </ul>
<b>Pre-conditions</b>	<ul style="list-style-type: none"> <li>• Specimen should be registered in Starlims;</li> <li>• Glass slides should be stained with standard staining techniques</li> </ul>
<b>Main path</b>	<ol style="list-style-type: none"> <li> <b>Export label components from Starlims:</b> <ol style="list-style-type: none"> <li>Actor activates “kit reports” module and enters the Kit ID (Brain ID) in Starlims;</li> <li>Starlims system generates unique IDs and label components (Specimen ID, Glass ID same as a barcode, staining, project name, institution name, etc.) for all specimens corresponding to selected Brain ID;</li> <li>Reporting system saves data as an XLS/CSV file that is compatible with the label printing service.</li> </ol> </li> <li> <b>Print data on the labels via the thermal transfer printer:</b> <ol style="list-style-type: none"> <li>An XLS file (generated in the Step 1c) is imported for label printing in the label printer software (Labtag printing company provides this step);</li> <li>Printing software converts a barcode string to the corresponding barcode image;</li> <li>Using the predesigned label template, the printing system prints the labels for each brain on the ribbons by increasing order of Specimen IDs and Glass IDs (for each specimen 5 labels ordered by IMG_001, IMG_002, IMG_003, etc.).</li> </ol> </li> <li> <b>Attach barcode labels on the glass slides:</b> <ol style="list-style-type: none"> <li>Actor takes 5 labels from the ribbon and the corresponding five glass slides from the slide box in which microscopic pieces are placed (five labels for each sample);</li> <li>Actor checks whether Specimen IDs on the selected labels are the same as the ID on the glass slides. Only if they are the same:               <ol style="list-style-type: none"> <li>Patient ID on the glass slide is hidden with a black marker</li> <li>each label is attached onto the glass slide according to the same staining type (staining name on the label should be the same as on the glass slide).</li> </ol> </li> </ol> </li> </ol>
<b>Post-conditions</b>	Glass slides are labeled with durable barcode labels

<b>Alternative path(s)</b>	3a	If a glass slide stained with the specific histopathological staining is missing: <ul style="list-style-type: none"> <li>• Actor prepares the required glass slide from the corresponding specimen</li> <li>• If the label for this glass slide is already printed, proceed with Step 3b, otherwise proceed with Step 1a.</li> </ul>
<b>Exceptional scenario(s)</b>	3a	If a label slide is missing or damaged, it should be reprinted
<b>Result</b>	The glass slides are prepared for digitization. They are marked with durable barcode labels for identification	
<b>Relationship to other scenarios</b>	This is a pre-step of the digitization of glass slides (S2)	

Table B.2. Use case specification for the digitization of glass slides

<b>Scenario 2 (S2) - “Digitization of glass slides”</b>	
<b>Short description</b>	Step-by-step operating procedures for scanning the glass slides
<b>Responsible stakeholders</b>	<ul style="list-style-type: none"> <li>• Pathologist (responsible for the digitization of glass slides);</li> <li>• Scanner administrator (responsible for the calibration of the scanner).</li> </ul>
<b>Actors</b>	Pathologist and/or research assistant
<b>Goal(s)</b>	<ul style="list-style-type: none"> <li>• To digitize glass slides using a microscopic scanner</li> <li>• To check the quality of scanned images and precision of image metadata</li> </ul>
<b>Pre-conditions</b>	<ul style="list-style-type: none"> <li>• Glass slide is prepared for the digitization</li> <li>• Actors are required to book the scanner at the booking system of the MOLCI administrative program to use the system on a specific date.</li> <li>• All users should have the user credentials to access the scanning software.</li> </ul>
<b>Main path</b>	<p><b>1</b></p> <p><b>Select glass slides from biorepository according to the scanning strategy and prepare the metadata file:</b></p> <ul style="list-style-type: none"> <li>a. Actor defines the most interesting case (brain) to be scanned;</li> <li>b. Actor chooses glass slides ordered by increasing number of Specimen ID (for each sample the glass slides are ordered in the in the following sequence: H&amp;E, LFB-PAS, Bielschowsky, KiM1P, and PLP);</li> <li>c. Actor prepares the metadata XLS file for the slides that should be scanned according to the standard template and converts it into the CSV file. The XLS file contains the following metadata information of multiple glass slides: slide name, slide info, specimen, tissue, preparation (used for the names of subfolders), staining, Specimen ID, date of autopsy, institution, and project name. The XLS file can be filled in two ways: manually or automatically using metadata generator template (in the same XLS document - sheet “Generator”). User can fill 4 parameters (Brain ID, initial and ending specimen index, and date of autopsy) in the generator template, and metadata table will be generated accordingly. Automated method is useful when user wants to fill the metadata file for multiple numbers of slides.</li> </ul>
	<p><b>2</b></p> <p><b>Define the scan mode in the scanning software:</b></p> <ul style="list-style-type: none"> <li>a. Actor activates the Olympus VS120 scanner and places the glass slide in the slide loader’s trays of the scanner in the same sequence as in the CSV metadata file (the glass slides are ordered in the same sequence in the slide box);</li> <li>b. If the system needs calibration of a specific setting, a MOLCI administrator is responsible for making the scanner operable with those specific conditions and calibrate the system according to the needs;</li> <li>c. User selects the scan mode called “expert batch scan mode”.</li> </ul>
	<p><b>3</b></p> <p><b>Import metadata via a CSV file:</b></p> <ul style="list-style-type: none"> <li>a. Actor fills the slide property values in the scanning software by importing the metadata file (generated in step 1c);</li> <li>b. All image properties from the CSV file are automatically assigned to the corresponding slide according to the sequence of the slides located in the cassette;</li> <li>c. The barcode is registered by the barcode reader integrated into the scanner hardware and is automatically entered in the “barcode” field in the slide properties module.</li> </ul>

4	<p><b>Define the scanning settings and automatic naming and saving of images:</b></p> <ol style="list-style-type: none"> <li>a. Actor defines the following scanning setting in the software: observation method, scan area, magnification, sample detection sensitivity, focus map density, camera settings in accordance with the staining type of tissue sample.</li> <li>b. Actor specifies the structure of the image name using three placeholders: &lt;SlideName&gt;_&lt;Staining&gt;_&lt;Magnification&gt;.vsi; The corresponding values are extracted from the CSV metadata file (or slide properties). Once the structure is defined, software automatically generates image name for all slides that are going to be scanned.</li> <li>c. Actor specifies a saving location on the server as follows: digital images should be stored in a subfolder named with the Specimen ID. This subfolder is automatically created on the server when the corresponding images are generated by the scanner. The actor specifies predefined placeholder &lt;Preparation&gt; for the name of the subfolders. Therefore, multiple subfolders will be created automatically on the server.</li> </ol> <p><i>Remark: The placeholder &lt;Preparation&gt; is used for creating the names of subfolders. Values for the names are extracted from column "Preparation" of the CSV metadata file.</i></p>
5	<p><b>Scan overview images and review batch scan settings</b></p> <ol style="list-style-type: none"> <li>a. After defining the scanning settings, software needs about 90min (for 100 slides) to scan overview images with 2x magnification;</li> <li>b. Once the overview images are acquired, the user needs time (15-20min for 100 slides) to review batch scan settings: <ul style="list-style-type: none"> <li>• check scan area that have been automatically determined (resize if it needs)</li> <li>• review the scanning settings for all slides</li> </ul> </li> </ol>
6	<p><b>Execute the batch scanning mode for multiple images (maximum for 100 glass slides)</b></p> <p><i>Remark: It is automated process that requires approximately 15 hours to scan 100 glass slides in the batch scanning mode.</i></p>
7	<p><b>Estimate the result of scanning (optional step):</b></p> <ol style="list-style-type: none"> <li>a. Once the digital images are generated and stored on the server, actor checks the results on the log file for any error during the scanning process and if found, addresses the mistakes. The log file is automatically generated by scanning software once the batch scanning process is completed; If a barcode is not recognized by Olympus scanner it is noted with "ND" (not detected) in the log file;</li> <li>b. Actor checks the subfolders with accompanying images in the case folder and views the thumbnails of digital images. A specific type of image errors is likely to be discovered during this step, such as blurred images or image distortions. Actor can also recognize low quality images through the file size. A damaged image has a significantly smaller file size than a normal image in the same subfolder (in each subfolder there are 5 images that represent the same tissue block with almost the same size of scan area but stained with different staining types; scanning process revealed if they are scanned correctly, they should have almost the same file size);</li> <li>c. Compare Glass IDs entered from two separated sources: to avoid erroneous linkages between an image and its metadata, Glass ID is registered from two different sources. First, the WSI scanner recognizes a 32-digit barcode on the label of the glass slide and stores it within the appropriate fields of the slide</li> </ol>

		properties. Second, Glass ID is also imported via the metadata entry. If there are no mistakes, the two values—the scanned barcode and Glass ID registered via CSV file—should be the same. Actor checks whether these two fields in slide properties are identical. A couple of minutes are usually sufficient to check this step for 100 images using the administrative software.
	8	Actor identifies the mistakes during the scanning process. Once errors are addressed, actor discusses these issues with the main pathologist. According to the discussion, some of the glass slides with bad image quality might be rescanned and stored on the server.
<b>Post-conditions</b>	Digitized glass slides are stored on the image server	
<b>Alternative path(s)</b>	3a	If the metadata XLS file is not imported automatically, user can enter it manually for each slide (time consuming step)
<b>Exceptional scenario(s)</b>	2a	If the Olympus VS120 scanner is not available for scanning, user can use the Olympus dotSlide scanner in exceptional cases, store digital images on the external hard disk, and transfer them to the image server later.
	6	If scanning is interrupted due to specific errors, digitization process must be restarted (proceed with Step 2c).
<b>Result</b>	Glass slides are digitized and high-resolution digital images with accompanying metadata are stored on the image server.	
<b>Relationship to other scenarios</b>	S3 – Folder structure for images needs to be customized on the server	

Table B.3. Use case specification for the management of user roles and images

<b>Scenario 3 (S3) - Management of user roles and administration of images within the DPS</b>	
<b>Short description</b>	Step-by-step procedures for managing user accounts and customizing imaging data on the server
<b>Responsible stakeholders</b>	Pathologist, research assistant, MOLCI administrator
<b>Actors</b>	<ul style="list-style-type: none"> <li>• Pathologist and research assistant with the project administrator account</li> <li>• MOLCI administrator with the super administrator account</li> </ul>
<b>Goal(s)</b>	<ul style="list-style-type: none"> <li>• To give specific access permissions to the users and researchers of KKNMS;</li> <li>• To manage a database and images on the server.</li> </ul>
<b>Pre-conditions</b>	<ul style="list-style-type: none"> <li>• Super administrator or project administrator account is needed to make changes on the server;</li> <li>• All administrative changes are executed in the Olympus acquisition (VS-ASW) or managing (OlyVIA) software.</li> </ul>
<b>Main path</b>	<p>1     <b>Manage different accounts and user roles:</b></p> <ol style="list-style-type: none"> <li>a. <b>Administrator</b> accounts for the project database: <ul style="list-style-type: none"> <li>▪ The super administrator of the microscopy system registers the actors (pathologist and research assistant) as project administrators and assigns the full access to the corresponding database. Using this account, actor can modify database structure and manage user rights.</li> <li>▪ The project administrator can only make changes in the specific database assigned by super administrator;</li> </ul> </li> <li>b. <b>User</b> accounts within the UMG: <ul style="list-style-type: none"> <li>▪ In the Olympus management software, the administrator creates the “user” accounts for pathologists or research assistants of the UMG, who need access to the image server of the MS-BB biorepository. This account has access to the database, can read, insert and update records in the database, but cannot modify database structure or manage user accounts.</li> <li>▪ The main responsibility of these users is to use their accounts for scanning, saving, and/or viewing images on the server.</li> </ul> </li> <li>c. <b>Guest</b> account for the researchers of KKNMS: <ul style="list-style-type: none"> <li>▪ External users (KKNMS researchers) need to have access to the individual image according to their needs. The main pathologist (project administrator) should define the roles and permissions for individual external researcher and create user accounts accordingly.</li> <li>▪ If there are multiple researchers, who need the same permissions to use the database, the administrator can create the user group, define the permission for all members of this group only once and add multiple user accounts in the group.</li> <li>▪ For the researchers of KKNMS, the administrator can create a “Guest” account and define the specific permissions for this role. For example, for individual images or specific folder(s) with including subfolders (or images), the administrator defines single permission by assigning to the individual user (or user group).</li> <li>▪ While using a guest account, the user only has permission to view the specific digital image with accompanying metadata and add the</li> </ul> </li> </ol>

		<p>annotations on it. However, a guest cannot add, modify or delete imaging data on the image server.</p>
	2	<p><b>Handle digital images on the server:</b></p> <p>a. Customize the database structure:</p> <ul style="list-style-type: none"> <li>▪ The project Database stores images in a hierarchical tree structure. There are multiple subfolders (case folders) in the main folder dedicated to the MS-BB project. One case folder corresponds to one individual brain (named with appropriate Brain ID). Within the case folder, there are numerous subfolders (Specimen Folder). Each specimen folder corresponds to one specimen of the current brain (named with appropriate Specimen ID). Under each specimen folder, corresponding images are stored (named as &lt;Slide_Name&gt;_&lt;staining&gt;_&lt;magnification&gt;). At any time, the user can insert case folders, specimen folders or any other kind of folders in this tree structure.</li> </ul> <p>b. Insert, update, and delete images on the server:</p> <ul style="list-style-type: none"> <li>▪ Using administrator or user accounts, digital images can be inserted onto the server either manually from external hard disk or automatically during the scanning process.</li> </ul> <p>c. Manage the image property fields:</p> <ul style="list-style-type: none"> <li>▪ The administrator can add new fields to the list of image properties and manage them accordingly</li> <li>▪ If the pathologist does not want specific fields to appear in the web viewer such as Specimen IDs, he/she can customize the viewing interface and hide unneeded fields in the web-viewer. It is also possible to assign these changes to specific (or all) users, who will be able to view images with new configuration.</li> </ul>
<b>Post-conditions</b>		<p>Project administrator, user, and guest accounts are created for the actors, UMG users, and KKNMS researchers respectively. Grouping properties, individual roles and permissions are assigned to the users. Images are well-structured on the server and viewing fields are adjusted according to the user's needs.</p>
<b>Result</b>		<p>Different types of user credentials are created and proper user permissions are registered; digital images are administrated and effectively handled on the image server.</p>
<b>Relationship to other scenarios</b>		<p>S2 – store images on the server during the digitization of glass slide</p>

Table B.4. Use case specification for the viewing of images

<b>Scenario 4 (S4) - Viewing of digital images by researchers of the KKNMS or by a pathologist of the MS-BB</b>		
<b>Short description</b>	Step-by-step procedures for viewing digital images using web viewer	
<b>Responsible stakeholders</b>	Pathologist, research assistant, IT technician	
<b>Actors</b>	Researchers of the KKNMS (Path A), pathologist/research assistant (Path A and Path B).	
<b>Goal(s)</b>	To view digital representation of samples using web viewer	
<b>Pre-conditions</b>	Researchers need to have own user credentials (Guest user) and image URL(s) provided by the main pathologist. Web interface of the MS-BB database is required to be published (responsible person is an IT technician)	
<b>Main path</b>	<b>Path A</b>	<b>Viewing digital images in the web viewer of the DPS by all users (researchers of KKNMS or a pathologist/research assistant of the MS-BB)</b>
		1   Once the pathologist supplied the researcher with the user credentials and image URL(s) of the requested specimen, the researcher can open the link and enter the username and password into the authorization form.
		2   Once the user is signed in, the system is automatically redirected to the appropriate image URL.
		3   To navigate over the image smoothly, the following navigation tools are likely to be used: zoom in, zoom out, rotation bar, image overview, full screen viewing.
		4   To change the color settings of digital images, researchers can use color settings bar and modify the following parameters: brightness, contrast, hue, and saturation.
		5   By using the advanced annotation tools, users can add different kind of annotations (rectangle, circle, point, flag, etc.) on the region of interest. Adding annotation requires selecting the region on the image and then adding text in a proper field.
		6   Users can take snapshots of the region of interests and save it as a JPG file.
		7   If the user has access to the multiple images on the server, he/she can compare images side by side. It is very useful for the researchers to compare the images of specific specimen stained with different staining types.
		8   If the user has access to some folders on the database, he/she can view the folder structure in the web-viewer and use the advanced search modules as required.
	<b>Path B</b>	<b>Viewing digital images using the interface between the DPS and the LIMS (only by the main pathologist/research assistant of the MS-BB)</b> <i>Note: Actor (main pathologist or research assistant) can login in the DPS as a separate and independent system, as well as access it directly from Starlims.</i>
1	When the actor is viewing the specimen information in Starlims, he/she can view the thumbnails of corresponding digital images in the "Remote images" tab with the links to the DPS system.	

			Furthermore, additional imaging data, such as barcode and staining type, appears in this tab.
		2	The user can click on any image thumbnail to open a new window and load the image web viewer.
		3	In Starlims, the pathologist can also copy the URL of a desired image to share it with other researchers.
<b>Post-conditions</b>	Digital images are viewed and annotated in the image web viewer by various types of users		
<b>Exceptional scenario(s)</b>	A-1	If the DPS web viewer is not available from outside of UMG, the responsible person should contact the IT department.	
	A-1	If the user credentials are not working or digital image is not loaded in the web viewer, researchers should contact the administrator (pathologist or research assistant).	
<b>Result</b>	A user can access digital images of MS-BB biorepository, use navigation and image analysis tools and add various type of annotations on the region of interest. Researchers of the KKNMS have access only to the DPS (Path A), whereas main pathologist/research assistant has access to the both systems (also Path B) —the DPS and the LIMS— providing additional functionality for them.		
<b>Relationship to other scenarios</b>	S5 - interfacing solution between the DPS and the LIMS		

Table B.5. Use case specification for data sharing between the LIMS and DPS

<b>Scenario 5 (S5) - Data sharing between the LIMS and the DPS using middleware interface</b>		
<b>Short description</b>	Manual and automated processes for effective linkage between the DPS and the LIMS	
<b>Responsible stakeholder</b>	Biobank interface technician	
<b>Actors</b>	<ul style="list-style-type: none"> <li>Automated process is handled by triggers;</li> <li>Manual tools are managed by a pathologist/research assistant.</li> </ul>	
<b>Goal(s)</b>	To link digital images to the corresponding specimen through an SOAP-based web service interface	
<b>Pre-conditions</b>	Digital images are stored on the server along with barcodes which are necessary for the linkage between images and specimens. The LIMS and the DPS should support SOAP-based web services; both systems are running in the DMZ (demilitarized zone) within the local environment;	
<b>Main path</b>	<b>Path A</b>	<b>Run web-service interface automatically (handled by triggers)</b>
		1   Middleware application integrated into Starlims runs automatically once in every 24 hours.
		2   The interface sends the query to the image server in order to select metadata (barcodes, thumbnails, Specimen IDs, URLs) of multiple images that are modified or inserted (scanned) on the image server during the previous 24 hours.
		3   Retrieved data is written in an XML format and processed with the Starlims Scripting Language. As a result, image metadata is distributed for each specimen into separated array elements.
		4   Using the standard database commands, image barcode and other imaging metadata are inserted into the Starlims DB and linked to the corresponding specimen using the barcodes of glass slides.
	<b>Path B</b>	<b>Run web-service interface manually (performed by pathologist/research assistant)</b>
		1   The pathologist/research assistant selects a desired specimen within Starlims and activates a tab called "remote images".
		2   If there is no corresponding image in the "remote images" tab (or if the imaging data requires an update), the actor can run middleware application by clicking on the button "NIS" located on the bottom right side in the "remote images" tab.
		3   After refreshing the specimen (select any other specimen, and reselect the desired specimen), corresponding imaging data will be updated in the "Remote Images" tab.
	<b>Post-conditions</b>	Digital images are linked to the corresponding specimens.
<b>Exceptional scenario(s)</b>	If Path A or B is not working properly, the actor should contact the interface technician or use the manual uploading tools implemented in the "remote images" tab.	
<b>Result</b>	Middleware interface integrated in Starlims links multiple digital images stored on the DPS to the corresponding specimen in Starlims using a SOAP-based web service technology.	

## Appendix C. Result of the market analysis in digital image handling systems and viewers

Table C.1. List of the image management systems and viewers for microscopic images

N	Name	Developed by	Supported file formats	Navigation tools	Based on	Advantages	Disadvantages	Commercial/open source
1	Net Image Server (v2.8)	Olympus	.vsi, TIFF, JPEG2000 and all standard file formats	STANDARD <sup>11</sup> + Annotation, Joystick, take snapshots, send region of interest via email, side-by-side viewing tools, color settings	HTML5	<ul style="list-style-type: none"> <li>Support for vsi and all standard image formats</li> <li>Viewing the whole metadata</li> <li>Support for SOAP-based web services</li> <li>Access by web-viewer, Olyvia software, or mobile applications</li> <li>Rapid access of virtual slide image</li> <li>Security with password and user login</li> <li>Support for adding advanced annotations</li> <li>Automated data storage on the server</li> </ul>	<ul style="list-style-type: none"> <li>Connection with only Olympus scanners</li> <li>Old scanner (dotSlide) is unlikely to be connected directly to the updated NIS</li> </ul>	Commercial
2	PathXL	PathXL	Olympus, Zeiss, Aperio, Hamamatsu, leica, Philips, Ariol, Microscan, 3Dhistech	STANDARD + Annotation, upload media, take snapshots	HTML5 + JavaScript	<ul style="list-style-type: none"> <li>Support for modules for biobank, research, education, and clinical usage</li> <li>User management system</li> <li>Image analysis framework</li> <li>Support for annotations, multiple windows, TMA toolbox, and case modules</li> <li>Support for API for integration into the LIMS</li> </ul>	<ul style="list-style-type: none"> <li>High cost</li> <li>Complexity of the consolidation methods for the multiple scanners provided by different vendors</li> </ul>	Commercial
3	Open Seadragon	Silverlight	.dzi, .xml (.jpg)	STANDARD	HTML, AJAX	<ul style="list-style-type: none"> <li>Switching images</li> <li>Adding custom controls</li> </ul>	Converted by deep zoom composer	Open source
4	Digital Image Hub	Slidepath, Leica Biosystems	Standard file format, Aperio, Hamamatsu, Leica	STANDARD + Joystick, Free rotate, brightness panel	Flash player	<ul style="list-style-type: none"> <li>Support for annotations</li> <li>Support for multiple windows</li> <li>Well organized panel</li> </ul>	<ul style="list-style-type: none"> <li>No support of other proprietary file formats</li> <li>Lack of the user management system</li> </ul>	Commercial
5	Web scope /Image scope	Aperio	.svs, .TIFF (BigTiff file) Aperio - svx	STANDARD	Flash player	<ul style="list-style-type: none"> <li>Add and view annotations</li> <li>Adjust brightness and contrast</li> </ul>	<ul style="list-style-type: none"> <li>No support of other proprietary file formats</li> <li>Lack of the user management system</li> </ul>	Commercial

<sup>11</sup> STANDARD (navigation tools) include the following tools: drag&drop, zoom in, zoom out, pan, reset view, and full screen mode.

N	Name	Developed by	Supported file formats	Navigation tools	Based on	Advantages	Disadvantages	Commercial/open source
6	ZEN browser	Zeiss (ZEISS Axio Scanner Z1)	Zeiss (Mirax), and all standard file formats	STANDARD+ annotation, side-by-side viewing, color settings	Flash player	<ul style="list-style-type: none"> <li>• Saving metadata and supporting documents</li> <li>• User management system</li> <li>• Remote access for mobile devices</li> </ul>	<ul style="list-style-type: none"> <li>• Only connected to the Zeiss scanners</li> </ul>	Commercial
7	Zoomify HTML5 Enterprise	Zoomify	.TIFF, .PFF (Pyramidal File Format), JPEG	STANDARD + built-in help, side-by-side comparison, measuring, rotation	HTML5 Javascript	<ul style="list-style-type: none"> <li>• Web-based annotation viewing and editing</li> <li>• Image Filters including brightness, contrast, sharpness</li> <li>• Javascript source code</li> <li>• Multiple window (up to 30 images synchronized in one display)</li> <li>• Save region of interest as a JPEG or PNG image</li> </ul>	<ul style="list-style-type: none"> <li>• No support of microscopic image formats (needs to convert to PFF or TIFF)</li> <li>• Support for folder-based storage or own PFF storage solution</li> </ul>	Commercial (Up to 500USD)
8	OMERO	University of Dundee & Open Microscopy Environment	Almost all microscopic images including proprietary file formats + OMETIFF	STANDARD + multiple dimensions, multiple channel, rendering settings	Python for web viewer, Java for desktop app	<ul style="list-style-type: none"> <li>• Support for desktop software, web applications and server platform</li> <li>• Hierarchical folders for better viewing and analysis of images; image management system</li> <li>• Support for histogram display for each channel of image</li> <li>• expended metadata support using OME-XML standard</li> <li>• Image analysis and measuring tools</li> </ul>	<ul style="list-style-type: none"> <li>• Needs image uploading on the OMERO server (time consuming task)</li> <li>• Difficulties while uploading large size of images (more than 3GB)</li> </ul>	Open source
9	JCB Data Viewer (OMERO Web)	The Journal of Cell Biology and Glencoe Inc.	.jp2,.tiff, .ome.tiff, standard file formats	STANDARD + color image rendering, annotations	HTML, Javascript (Based on OMERO system)	<ul style="list-style-type: none"> <li>• Support for sharing of image data associated with articles published in the Journal of Cell Biology</li> <li>• Metadata represented with OME-TIFF format</li> <li>• Downloading image as OME-TIFF file</li> <li>• Viewing the images merged in a stack or separated with different color channels</li> </ul>	<ul style="list-style-type: none"> <li>• Works only for microscopic images published in the journal</li> </ul>	Open source
10	mScope	Aurora Ltd.	Standard file formats, proprietary digital slide formats (not mentioned in detail), DICOM	STANDARD+ rotation, modifying transparency, color filters, viewing history, annotations	Java	<ul style="list-style-type: none"> <li>• scanner independent web-based viewer for desktop, tablet, and smart-phone platforms</li> <li>• Image sharing and archiving</li> <li>• Support for encryption streaming of data</li> <li>• Support for collaborative session using the real-time chats for multiple users</li> <li>• Test management and grading system for education purposes</li> </ul>	<ul style="list-style-type: none"> <li>• Requires server infrastructure</li> </ul>	Commercial

N	Name	Developed by	Supported file formats	Navigation tools	Based on	Advantages	Disadvantages	Commercial/open source
11	Simagis Live	Smart Imaging Technologies Co	Aperio, Zeiss, Hamamatsu, Leica-.scn, Ventana (.tif, jp2), BigTiff, standard image formats	STANDARD + advanced annotations, color settings, histograms, measuring tools.	HTML, Javascript	<ul style="list-style-type: none"> <li>• Support for interface for LIS, image analysis and other applications</li> <li>• Support for secure private cloud hosting or in-house server</li> <li>• Automatic image analysis tools are integrated</li> <li>• Sharing images with unauthorized users</li> <li>• Measure particles and download distribution statistics in XLS file</li> <li>• Generating reports containing information about images and samples</li> <li>• Support for image management server</li> </ul>	<ul style="list-style-type: none"> <li>• Needs to install software for uploading images on the server</li> <li>• Do not support all proprietary file formats</li> <li>• Average file size is 100MB. Difficulties with very large files (such as 2-3 GB)</li> </ul>	Free trial version, different packages of commercial product
13	Digital SlideBox	Slidepath	Aperio, Hamamatsu, Zeiss, and standard image formats	STANDARD + Joystick, different types of brightness, annotations, multiple windows	Flash player	<ul style="list-style-type: none"> <li>• Support for learning management system using web</li> <li>• Helps to standardize course materials in histology</li> <li>• well organized panel and user friendly working environment</li> <li>• Adjust brightness, contrast and color settings</li> <li>• Saving snapshots as JPEG files</li> <li>• Performing measurements on the digital image</li> </ul>	<ul style="list-style-type: none"> <li>• Created and used mostly for educational purposes</li> </ul>	Commercial
14	VM Slide Explorer	VMscope GmbH	VMscope, 3DHitech, Hamamatsu, Aperio ScanScope, Olympus webImage, Leica, Zoomify, JPEG2000, .jp2	STANDARD + touch screen via Apple iPad, annotations	Most probably on HTML and Javascript	<ul style="list-style-type: none"> <li>• Ability to integrate into the LIMS (using simple tools)</li> <li>• Automatic quantification analysis tools for immune stained slides</li> </ul>	<ul style="list-style-type: none"> <li>• Do not support all proprietary file formats</li> <li>• Difficulties during operation with different vendors using automated tools.</li> <li>• LIMS integration without webservice</li> </ul>	Commercial
15	IIPMoo Viewer	IIPImage	JPEG2000 (decoded via Kakadu library) TIFF	STANDARD + annotations	HTML5, Javascript	<ul style="list-style-type: none"> <li>• Compatible with mobile devices</li> <li>• Working with the IIPImage server or static tiling protocols (e.g., Zoomify, Deepzoom)</li> <li>• Dedicated for most imaging applications, including telescope images or paintings</li> </ul>	<ul style="list-style-type: none"> <li>• Kakadu decoding system for JPEG2000 files</li> </ul>	Open source, licensing cost
16	The Brain Maps API	The BrainMaps Team	TIFF	STANDARD	HTML, Javascript	<ul style="list-style-type: none"> <li>• Including viewer in any applications with iFrames</li> </ul>	<ul style="list-style-type: none"> <li>• Works only with the tiled images</li> <li>• Does not support advanced tools, mostly used only for viewing images</li> </ul>	Free

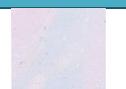
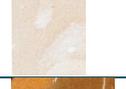
N	Name	Developed by	Supported file formats	Navigation tools	Based on	Advantages	Disadvantages	Commercial/open source
17	DMetrix	Dmetrix Inc	.dmi (DMetrix), standard image formats (Tiff, JPEG, BMP, PNG)	Zoom in and zoom out, full screen, annotations	HTML, javascript	<ul style="list-style-type: none"> <li>• Access images via web or desktop application (digitalEyepiece)</li> </ul>	<ul style="list-style-type: none"> <li>• No advanced tools for image viewing and analysis</li> <li>• Mostly operates with the DMetrix DX-40 scanner</li> </ul>	Commercial
18	Collibio	Pixcelldata	Standard image formats	STANDARD + annotations,		<ul style="list-style-type: none"> <li>• Interactive web or desktop application</li> <li>• Sharing images with colleagues</li> <li>• Able to connect to existing server infrastructure</li> <li>• Able to connect the LIMS/LIS</li> <li>• Support for case management system</li> <li>• Monitor changes to the data</li> <li>• Manage image libraries for students</li> </ul>	<ul style="list-style-type: none"> <li>• Difficulties with proprietary file formats such as .vsi</li> <li>• Difficulties with large image data</li> </ul>	Commercial

## Appendix D. Results of test scanning process

Table D.1. Scanning result 1 (Olympus dotSlide scanner)

N	Image	Staining	Type	Scanned area	Objective Lens	Scanning time	Saving time	Size (.vsi)	Convert to Tiff	Convert to JPEG2000	Convert to Tiff Image Series	Convert to web image series (JPG)
1		H&E	Chronic active	10mmx24mm	x40	1h 25min	12min	2.79 GB	Crashed the system after 25 min			
2		H&E	Chronic active	10mmx24mm	x20	27min	5min	0.99GB	35min canceled by us due to the large size	After 1h 52min canceled by software	8.82GB (20 files)	3.16GB
3		Bielschowsky	Chronic active	10mmx24mm	x40	1h 23min	24min	3.12 GB				6,53GB
4		H&E	Shadow plaque	25mmx28mm	x40	2h 54min	26min	6.71GB				
5		LFB-PAS	Cortical lesion	23mmx23mm	X40	2h 51min	14min	3.92GB				

Table D.2. Scanning result 2 (using both scanners - Olympus dotSlide and Olympus VS120)

N	Image	Staining	Type	Scanned area	Objective Lens	Olympus dotSlide		Olympus VS120	
						Scanning time	Size (.vsi)	Scanning time	Size (.vsi)
1		H&E	Chronic active	13mmx10mm	x40	20min	907MB	4min	1.05GB
2		Bielschowsky	Chronic active	13mmx10mm	X40	24min	1.65GB	5min	2.15GB
3		LFB-PAS	Chronic active	13mmx10mm	x40	21min	899MB	4min	1.2 GB
4		KiM1P	Chronic active	13mmx10mm	x40	21min	526MB	3min	814MB
5		PLP	Chronic active	13mmx10mm	X40	26min	856MB	4min	1.04GB
6		LFB-PAS	Shadow plaque	5mmx7mm	X40	7min	197MB	2min	280MB
7		PLP	Shadow plaque	4mmx7mm	X40	3min	162MB	1min	207MB
8		Bielschowsky	Shadow plaque	5mmx6mm	X40	5min	477MB	1min	526MB
9		H&E	Shadow plaque	4mmx6mm	X40	7min	188MB	1min	251MB

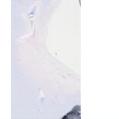
N	Image	Staining	Type	Scanned area	Objective Lens	Olympus dotSlide		Olympus VS120	
						Scanning time	Size (.vsi)	Scanning time	Size (.vsi)
10		PLP	Cortical lesion	7mmx8mm	X40	9min	347MB	3min	550MB
11		Bielschowsky	Inactive lesion	7mmx8mm	X40	9min	806 MB	3min	1.14GB
12		LFB-PAS	Inactive lesion	6mmx9mm	X40	10min	242MB	2min	288MB
13		KiM1P	Inactive lesion	9mmx10mm	X40	17min	296MB	3min	505MB
14		PLP	Inactive lesion	10mmx6mm	X40	12min	351MB	3min	592MB
15		H&E	Inactive lesion	8mmx5mm	X40	8min	272MB	2min	390MB

Table D.3. Conversion tools of VS120. One digital image (that was stained with H&E and scanned with 20x magnification) was converted into different open file formats. Original file format is .vsi.

Conversion tools of VS120		
File format	Size	Duration
.vsi	390MB	-
TIFF	4.6GB	2min
BigTiff	4.16GB	3min
JPEG2000	166MB	4min
PNG	1.52GB	16min
JPG	216MB	4min
Web images with html file	1.12GB	2min
export to image series (multiple tiff) max size 50MB for each image	164 GB (3364 images)	32min

Appendix E. Metadata for multiple glass slides

	A	B
1	Brain ID (without "Kit_BB")	1234567890
2	From	1
3	To (including)	10
4	Date of autopsy	01/13

Figure E.1. Metadata generator template. A user enters the following four parameters to generate the metadata table for scanning of multiple glass slides: Brain ID, starting and ending specimen index, and date of autopsy.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Tray No	Slide No	SlideName	SlideInfo	Specimen	Species	Tissue	Preparation	Staining	Sample ID	Date of Autopsy	Institution	Project Name
2	1	1	Kit_BB_1234567890_TIS_01_IMG_001	1234567890_TIS_01_HE	Gyrus frontalis medialis L			1234567890_TIS_01	HE	Kit_BB_1234567890_TIS_01	01/13	UMG Neuropathologie	MS-BB
3	1	2	Kit_BB_1234567890_TIS_01_IMG_002	1234567890_TIS_01_LFB-PAS	Gyrus frontalis medialis L			1234567890_TIS_01	LFB-PAS	Kit_BB_1234567890_TIS_01	01/13	UMG Neuropathologie	MS-BB
4	1	3	Kit_BB_1234567890_TIS_01_IMG_003	1234567890_TIS_01_Biel	Gyrus frontalis medialis L			1234567890_TIS_01	Biel	Kit_BB_1234567890_TIS_01	01/13	UMG Neuropathologie	MS-BB
5	1	4	Kit_BB_1234567890_TIS_01_IMG_004	1234567890_TIS_01_KIM1P	Gyrus frontalis medialis L			1234567890_TIS_01	KIM1P	Kit_BB_1234567890_TIS_01	01/13	UMG Neuropathologie	MS-BB
6	1	5	Kit_BB_1234567890_TIS_01_IMG_005	1234567890_TIS_01_PLP	Gyrus frontalis medialis L			1234567890_TIS_01	PLP	Kit_BB_1234567890_TIS_01	01/13	UMG Neuropathologie	MS-BB
7	1	6	Kit_BB_1234567890_TIS_02_IMG_001	1234567890_TIS_02_HE	Gyrus frontalis medialis R			1234567890_TIS_02	HE	Kit_BB_1234567890_TIS_02	01/13	UMG Neuropathologie	MS-BB
8	1	7	Kit_BB_1234567890_TIS_02_IMG_002	1234567890_TIS_02_LFB-PAS	Gyrus frontalis medialis R			1234567890_TIS_02	LFB-PAS	Kit_BB_1234567890_TIS_02	01/13	UMG Neuropathologie	MS-BB
9	1	8	Kit_BB_1234567890_TIS_02_IMG_003	1234567890_TIS_02_Biel	Gyrus frontalis medialis R			1234567890_TIS_02	Biel	Kit_BB_1234567890_TIS_02	01/13	UMG Neuropathologie	MS-BB
10	1	9	Kit_BB_1234567890_TIS_02_IMG_004	1234567890_TIS_02_KIM1P	Gyrus frontalis medialis R			1234567890_TIS_02	KIM1P	Kit_BB_1234567890_TIS_02	01/13	UMG Neuropathologie	MS-BB
11	1	10	Kit_BB_1234567890_TIS_02_IMG_005	1234567890_TIS_02_PLP	Gyrus frontalis medialis R			1234567890_TIS_02	PLP	Kit_BB_1234567890_TIS_02	01/13	UMG Neuropathologie	MS-BB
12	1	11	Kit_BB_1234567890_TIS_03_IMG_001	1234567890_TIS_03_HE	Gyrus cinguli L			1234567890_TIS_03	HE	Kit_BB_1234567890_TIS_03	01/13	UMG Neuropathologie	MS-BB
13	1	12	Kit_BB_1234567890_TIS_03_IMG_002	1234567890_TIS_03_LFB-PAS	Gyrus cinguli L			1234567890_TIS_03	LFB-PAS	Kit_BB_1234567890_TIS_03	01/13	UMG Neuropathologie	MS-BB
14	1	13	Kit_BB_1234567890_TIS_03_IMG_003	1234567890_TIS_03_Biel	Gyrus cinguli L			1234567890_TIS_03	Biel	Kit_BB_1234567890_TIS_03	01/13	UMG Neuropathologie	MS-BB
15	1	14	Kit_BB_1234567890_TIS_03_IMG_004	1234567890_TIS_03_KIM1P	Gyrus cinguli L			1234567890_TIS_03	KIM1P	Kit_BB_1234567890_TIS_03	01/13	UMG Neuropathologie	MS-BB
16	1	15	Kit_BB_1234567890_TIS_03_IMG_005	1234567890_TIS_03_PLP	Gyrus cinguli L			1234567890_TIS_03	PLP	Kit_BB_1234567890_TIS_03	01/13	UMG Neuropathologie	MS-BB
17	1	16	Kit_BB_1234567890_TIS_04_IMG_001	1234567890_TIS_04_HE	Gyrus cinguli R			1234567890_TIS_04	HE	Kit_BB_1234567890_TIS_04	01/13	UMG Neuropathologie	MS-BB
18	1	17	Kit_BB_1234567890_TIS_04_IMG_002	1234567890_TIS_04_LFB-PAS	Gyrus cinguli R			1234567890_TIS_04	LFB-PAS	Kit_BB_1234567890_TIS_04	01/13	UMG Neuropathologie	MS-BB
19	1	18	Kit_BB_1234567890_TIS_04_IMG_003	1234567890_TIS_04_Biel	Gyrus cinguli R			1234567890_TIS_04	Biel	Kit_BB_1234567890_TIS_04	01/13	UMG Neuropathologie	MS-BB
20	1	19	Kit_BB_1234567890_TIS_04_IMG_004	1234567890_TIS_04_KIM1P	Gyrus cinguli R			1234567890_TIS_04	KIM1P	Kit_BB_1234567890_TIS_04	01/13	UMG Neuropathologie	MS-BB
21	1	20	Kit_BB_1234567890_TIS_04_IMG_005	1234567890_TIS_04_PLP	Gyrus cinguli R			1234567890_TIS_04	PLP	Kit_BB_1234567890_TIS_04	01/13	UMG Neuropathologie	MS-BB
22	1	21	Kit_BB_1234567890_TIS_05_IMG_001	1234567890_TIS_05_HE	Gyros parietalis inferior L			1234567890_TIS_05	HE	Kit_BB_1234567890_TIS_05	01/13	UMG Neuropathologie	MS-BB
23	1	22	Kit_BB_1234567890_TIS_05_IMG_002	1234567890_TIS_05_LFB-PAS	Gyros parietalis inferior L			1234567890_TIS_05	LFB-PAS	Kit_BB_1234567890_TIS_05	01/13	UMG Neuropathologie	MS-BB
24	1	23	Kit_BB_1234567890_TIS_05_IMG_003	1234567890_TIS_05_Biel	Gyros parietalis inferior L			1234567890_TIS_05	Biel	Kit_BB_1234567890_TIS_05	01/13	UMG Neuropathologie	MS-BB
25	1	24	Kit_BB_1234567890_TIS_05_IMG_004	1234567890_TIS_05_KIM1P	Gyros parietalis inferior L			1234567890_TIS_05	KIM1P	Kit_BB_1234567890_TIS_05	01/13	UMG Neuropathologie	MS-BB
26	1	25	Kit_BB_1234567890_TIS_05_IMG_005	1234567890_TIS_05_PLP	Gyros parietalis inferior L			1234567890_TIS_05	PLP	Kit_BB_1234567890_TIS_05	01/13	UMG Neuropathologie	MS-BB
27	1	26	Kit_BB_1234567890_TIS_06_IMG_001	1234567890_TIS_06_HE	Gyros parietalis inferior R			1234567890_TIS_06	HE	Kit_BB_1234567890_TIS_06	01/13	UMG Neuropathologie	MS-BB
28	1	27	Kit_BB_1234567890_TIS_06_IMG_002	1234567890_TIS_06_LFB-PAS	Gyros parietalis inferior R			1234567890_TIS_06	LFB-PAS	Kit_BB_1234567890_TIS_06	01/13	UMG Neuropathologie	MS-BB
29	1	28	Kit_BB_1234567890_TIS_06_IMG_003	1234567890_TIS_06_Biel	Gyros parietalis inferior R			1234567890_TIS_06	Biel	Kit_BB_1234567890_TIS_06	01/13	UMG Neuropathologie	MS-BB
30	1	29	Kit_BB_1234567890_TIS_06_IMG_004	1234567890_TIS_06_KIM1P	Gyros parietalis inferior R			1234567890_TIS_06	KIM1P	Kit_BB_1234567890_TIS_06	01/13	UMG Neuropathologie	MS-BB
31	1	30	Kit_BB_1234567890_TIS_06_IMG_005	1234567890_TIS_06_PLP	Gyros parietalis inferior R			1234567890_TIS_06	PLP	Kit_BB_1234567890_TIS_06	01/13	UMG Neuropathologie	MS-BB
32	1	31	Kit_BB_1234567890_TIS_07_IMG_001	1234567890_TIS_07_HE	Area striata L			1234567890_TIS_07	HE	Kit_BB_1234567890_TIS_07	01/13	UMG Neuropathologie	MS-BB
33	1	32	Kit_BB_1234567890_TIS_07_IMG_002	1234567890_TIS_07_LFB-PAS	Area striata L			1234567890_TIS_07	LFB-PAS	Kit_BB_1234567890_TIS_07	01/13	UMG Neuropathologie	MS-BB
34	1	33	Kit_BB_1234567890_TIS_07_IMG_003	1234567890_TIS_07_Biel	Area striata L			1234567890_TIS_07	Biel	Kit_BB_1234567890_TIS_07	01/13	UMG Neuropathologie	MS-BB
35	1	34	Kit_BB_1234567890_TIS_07_IMG_004	1234567890_TIS_07_KIM1P	Area striata L			1234567890_TIS_07	KIM1P	Kit_BB_1234567890_TIS_07	01/13	UMG Neuropathologie	MS-BB
36	1	35	Kit_BB_1234567890_TIS_07_IMG_005	1234567890_TIS_07_PLP	Area striata L			1234567890_TIS_07	PLP	Kit_BB_1234567890_TIS_07	01/13	UMG Neuropathologie	MS-BB
37	1	36	Kit_BB_1234567890_TIS_08_IMG_001	1234567890_TIS_08_HE	Area striata R			1234567890_TIS_08	HE	Kit_BB_1234567890_TIS_08	01/13	UMG Neuropathologie	MS-BB

Figure E.2. Metadata file for scanning of multiple glass slides automatically generated in response to the parameters indicated in the metadata generator template. The complete metadata file is represented as a digital format on the accompanied DVD.

## Appendix F. Accompanied DVD

The enclosed DVD contains the following data:

1. Metadata file for scanning of glass slides
2. Sample of the source code for the interface solution that provides data exchange between the LIMS and the DPS
3. Video representation of image viewing process within the DPS and the LIMS
4. Digital version of the dissertation (PDF format)

## List of Figures

Figure 1.1. Difficulties in DP .	1
Figure 1.2. Thesis workflow	8
Figure 2.1. Non digital and digital images	14
Figure 2.2. Components of medical imaging informatics.	15
Figure 2.3. Publication in whole slide imaging in PubMed between 2000–2016.	21
Figure 2.4. Traditional and digital pathology workflows	22
Figure 2.5. Applications of digital pathology from basic research to diagnostic purposes	24
Figure 2.6. Research infrastructure at UMG	35
Figure 2.7. The initial state of digital pathology infrastructure at UMG through 2013	36
Figure 3.1. Requirements engineering framework	42
Figure 4.1. Elements of the DPS	56
Figure 4.2. Myelin damage in MS	57
Figure 4.3. MS lesion stained with different colors	58
Figure 4.4. Hierarchy of information entities in pathology	60
Figure 4.5. Identification of pathology information in STARLIMS	62
Figure 4.6. Dimensions of scanning area and label area of glass slides.	62
Figure 4.7. Interface for identification of specimen-derived assets	64
Figure 4.8. Design of the barcode label for glass slides	65
Figure 4.9. Attaching the labels on the glass slides	67
Figure 4.10. Uniform labeling workflow for two types of glass slides	68
Figure 4.11. UML activity diagram for the preparation of glass slides	69
Figure 4.12. Olympus VS120 workstation at UMG	71
Figure 4.13. Scanning settings for the glass slides of the MS-BB repository	73
Figure 4.14. Types of the damage image revealed after the scanning process	77
Figure 4.15. Communication scheme between the image server and the other components of the DPS	79
Figure 4.16. Folder hierarchy on the database	80
Figure 4.17. Estimation of Storage needs for the digital microscopic images of the MS Brain Bank project	81
Figure 4.18. Different types of user accounts and corresponding roles	82
Figure 4.19. Example of user groups	83
Figure 4.20. Web interface for digital microscopic images	85
Figure 5.1. Three levels of interoperability	88
Figure 5.2. Web service as a layer between the application code and application client	89
Figure 5.3. Architecture of web services. It contains three main components	90
Figure 5.4. Web service architecture for the DPS and the LIMS	92
Figure 5.5. Sequence diagram for the data exchange from the DPS to the LIMS	94
Figure 5.6. Mockup of “Remote images” tab in STARLIMS	97
Figure 5.7. Digital pathology workflow	99
Figure 5.8. W-model for the evaluation of the DPS	100
Figure E.1. Metadata generator template	143
Figure E.2. Metadata file for scanning of multiple glass slides	144

## List of Tables

<i>Table 2.1. Statements of the guideline for the validating whole slide imaging for diagnostic purposes.....</i>	<i>30</i>
<i>Table 3.1. Requirements elicitation questions .....</i>	<i>50</i>
<i>Table 3.2. List of the requirements for the DPS .....</i>	<i>51</i>
<i>Table 3.3. List of DPS scenarios. ....</i>	<i>52</i>
<i>Table 3.4. List of the image management systems and viewers in digital pathology .....</i>	<i>55</i>
<i>Table 4.1. Requirements for uniform labeling of glass slides divided into two subgroups .....</i>	<i>63</i>
<i>Table 4.2. Extract from an XLS exported file.. ....</i>	<i>65</i>
<i>Table 4.3. Examples of metadata for multiple glass slides. ....</i>	<i>74</i>
<i>Table 4.4. Scanning results.....</i>	<i>76</i>
<i>Table 5.1. Fields of the table for the microscopic images in the LIMS database.....</i>	<i>95</i>
<i>Table B.1. Use case specification for the preparation of glass slides.....</i>	<i>125</i>
<i>Table B.2. Use case specification for the digitization of glass slides.....</i>	<i>127</i>
<i>Table B.3. Use case specification for the management of user roles and images .....</i>	<i>130</i>
<i>Table B.4. Use case specification for the viewing of images.....</i>	<i>132</i>
<i>Table B.5. Use case specification for data sharing between the LIMS and DPS.....</i>	<i>134</i>
<i>Table C.1. List of the image management systems and viewers for microscopic images .....</i>	<i>135</i>
<i>Table D.1. Scanning result 1 (Olympus dotSlide scanner) .....</i>	<i>139</i>
<i>Table D.2. Scanning result 2 (using both scanners - Olympus dotSlide and Olympus VS120).....</i>	<i>140</i>

## Abbreviations

3D	Three-dimensional
ATA	American Telemedicine Association
CAD	Computer Assisted (Aided) Diagnosis
CAP	College of American Pathologists
CE	European Conformity (Conformite Europeenne)
CFR	Code of Federal Regulations
CSV	Comma-separated Value
DICOM	Digital Imaging and Communications in Medicine
DMZ	Demilitarized Zone
DP	Digital Pathology
DPA	Digital Pathology Association
DPS	Digital Pathology System
EURO-TELEPATH	Telepathology Network in Europe
FDA	Food and Drug Administration
GIF	Graphics Interchange Format
H&E	Hematoxylin and Eosin
HL7	Health Level-7
HTTPS	Hypertext Transfer Protocol Security
ID	Identifier
IHE	Integrating the Healthcare Enterprise
IIS	Internet Information Service
IT	Information Technology
JPEG	Joint Photographic Experts Group
KKNMS	Competence Network Multiple Sclerosis
LAN	Local Area Network
LIMS	Laboratory Information Management System
MI	(Department of) Medical Informatics
MOLCI	Molecular and Optical Live Cell Imaging
MRI	Magnetic Resonance Imaging
MS	Multiple Sclerosis
MS-BB	Multiple Sclerosis Brain Bank
Nifti	Neuroimaging Informatics Technology Initiative
NIS	Net Image Server
NIS-SQL	Net Image Server SQL
OME-TIFF	The Open Microscopy Environment
PACS	Picture Archive and Communication Systems
PNG	Portable Networks Graphics
RBAC	Role-based Access Control
REST	Representational State Transfer
SDLC	System Development Life Cycle

SOAP	Simple Object Access Protocol
TIFF	Tagged Image File Format
TMF	Technology, Methods and Infrastructure
UMG	University Medical Center Göttingen
URL	Uniform Resource Locator
WADO	Web Access to DICOM Persistent Objects
WSDL	Web Services Description Language
WSI	Whole Slide Image (Imaging)
XLS	Microsoft eXcel Spreadsheet
XML	Extensible Markup Language
QA	Quality Assurance
QC	Quality Control

# Curriculum Vitae

## Personal data

---

Name Lika Svanadze  
Date of birth 07.06.1987  
Nationality Georgian  
Email likasvanadze@gmail.com

## Education

---

2013 – 2019 Doctoral student in Computer Science, Department of Medical Informatics  
Georg-August-University Göttingen, Germany  
PhD thesis: “The Integration of the Brain Bank Imaging Workflow  
into the Infrastructure of the Multiple Sclerosis Research Network”

2008 – 2013, Bachelor’s program at the Faculty of Medicine; Tbilisi State Medical University,  
2017 – 2019 Georgia

2008 – 2010 Master’s degree in Computer Science  
Ivane Javakhishvili Tbilisi State University, Georgia  
Master Thesis: Automation of a complex application with script language

2004 – 2008 Bachelor’s degree in Computer Science  
Ivane Javakhishvili Tbilisi State University, Georgia

## Research experience

---

2013 - 2017 Research associate at the Department of Medical Informatics, University Medical  
Center Göttingen

## Current Job

---

2018 - Today Ivane Javakhishvili Tbilisi State University, Georgia  
Position: Invited Lecturer  
Subject: Research Skills for Medical Students (Georgian and English courses)

2018 - Today Business and Technology University, Georgia  
Position: Lecturer  
Subject: Programming Language – Python

## Fellowships and Awards

---

- ALRAKIS scholarship to study at the Georg-August-University Göttingen from 2013 to 2015
- Scholarship from the International Education Center in Georgia for studying at the Georg-August-University Göttingen from 2015-2017
- Student's Scientific Conference, Ivane Javakhishvili Tbilisi State University, 2007; Second place; Thesis: "Elliptic Curve in Cryptography"

## Conference Publication

---

Svanadze L, Franke T, Buckow K, Bahn E, Rienhoff O. Requirements for integration of microscope images into the laboratory information management system of a research network. 59. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie; 2014.

## Conference contributions

---

Svanadze L, Franke T, Buckow K, Romanus G, Bahn E, Rienhoff O. An Integrated Pathology Workflow Solution for Biobank Research Consortia. Selected for podium presentation; Pathology Informatics Summit 2016, Pittsburgh, PA; 2016.

Svanadze L, Franke T, Buckow K, Romanus G, Bahn E, Rienhoff O. Integration of the Microscopic Images into a Research Network Using Web Services. Poster presentation. Digital Pathology Congress, London, UK, 2015.

Svanadze L, Franke T, Buckow K, Bahn E, Rienhoff O. IT- Tools to Overcome Integration Problems in Digital Microscopy within Collaborative Research Networks. Podium presentation. Pathology Informatics Summit 2015, Pittsburgh, PA.

# Requirements for integration of microscope images into the laboratory information management system of a research network

L Svanadze<sup>1</sup>, T Franke<sup>1</sup>, K Buckow<sup>1</sup>, E Bahn<sup>2</sup>, O Rienhoff<sup>1</sup>

<sup>1</sup>University Medical Center Göttingen, Department of Medical Informatics, Göttingen

<sup>2</sup>University Medical Center Göttingen, Department of Neuropathology, Göttingen

## Introduction

Digital pathology provides novel ways for pathologists to view, store, transmit, manage, process, and analyze tissue samples. As microscopic images are more and more being used in a digital format, their well-organized integration into laboratory information management systems (LIMS) becomes significantly more important for primary diagnosis, research, and education. Currently, there are no standard guidelines regarding the handling of whole slide images unlike radiology images, for which DICOM (Digital Image Communication in Medicine) became the de- facto standard. The very large virtual microscope images are hitting some limitations of DICOM, such as the maximum size of uncompressed image pixel data (4GB) caused by the underlying 32-bit architecture [1].

As a subproject of the Competence Network Multiple Sclerosis (KKNMS) [2], the Multiple Sclerosis Brain Bank (MS- BB) serves as a use case for the application of Digital Pathology system (DPS) in clinical research. Within the scope of the project, high-resolution scans of histopathologic samples are to be provided to researchers before shipping specific brain tissue to facilitate a digital view into the sample. The objective of the work is to find out significant requirements for the DPS in the environment of the research network, in order to develop an interoperable microscope image handling system, efficiently integrated into the existing IT infrastructure of the KKNMS.

## Materials and methods

In order to determine the requirements, the following five steps were performed. At first, the workflow of digital pathology was examined, where the characteristics of the MS lesion, staining procedures, clinical and biomaterial data were assessed. Second, the scanning process was analyzed and several histopathologic samples supplied by the Department of Neuropathology were scanned. During this, technical details like file size, format, and structure, scanning time, and batch processing functionality were evaluated. Third, the documentation process for biomaterial with the local LIMS was assessed, where tissue samples of patients are registered. Forth, an investigation of image viewers and handling systems for digital pathology was carried out on the market. In the final step, the requirements were drawn from the materials collected in the previous four steps according to the procedure of theoretical assumptions of the requirements engineering [3].

## Results

During the analysis of the workflows, three significant stages were identified: preparation, digitalization, and provision. The preparation process comprises the standard operating procedures for the autopsy of the MS patients, tissue collection, the histological staining procedures and the characterization of the MS lesions with regard to the localization and lesion activity.

During the assessment of the digitalization stage, we used the Olympus dotSlide microscope and significant parameters such as file size and scanning time were regarded. Digital pathology

works with huge data volumes and it requires support of big data storage solutions. Average file size of a 15mm x 20mm tissue specimen varies between 2.8GB for 40x and 1GB for 20x magnification. Additionally, the scanning duration plays an important role, especially when using the higher magnification. Corresponding scanning durations are 1h 25min for 40x and 25min for 20x scan. Historically, scanning speed has been improved with every new generation of scanners, but data volume, especially if fluorescence tissues or z-stack images are scanned, has also gone up. The file format of our images is .vsi, but contemporary digital scanners support conversion of own format to standard file formats such as TIFF or JPEG2000 with loss of accompanying metadata.

To provide an overview of tissue samples for requested research questions, the clinical parameters of patients in conjunction with the digital images are important for researchers of KKNMS. The conclusion of this is that DPS needs to be integrated into a comprehensive IT infrastructure including phenotype and biomaterial data. Tissue samples, registered in the LIMS, are assigned to the specific brain area and have a unique sample ID generated by the LIMS. DPS should be separated from the LIMS and includes an image server, database, and web viewer for microscopic images. Digital slides in conjunction with metadata are manually uploaded to the DPS after the completion of the scanning process. A URL and thumbnail of each digital slide are registered for tissue sample into the LIMS in order to enable launching of a DPS viewer directly from within the LIMS [4]. Most of the LIMS can work with web-services that are the most common mechanism for interfacing with external systems (e.g. SOAP, REST). Finally, the current web viewers for microscopic images were investigated that identified supplementary features for the DPS. 15 different systems were examined, basing on different platform and supporting different file formats. Based on depicted investigations, the following features have been determined as required for a DPS: tools for uploading images; storing of accompanying metadata in conjunction with the images; web viewer with navigation tools for image viewing; support of high resolution images and different file formats; storage infrastructure for large digital files. In addition, desirable functionalities of system are: tools for adding annotations; support of multiple images viewing; questionnaire of digital images for researchers; take snapshot of region of interest. Last but not least, security of data and user account is one of the important features of system.

## **Discussion**

The work presented here sums up the functional, technical, system, and environmental requirements for integration of digital pathology into the already existing IT infrastructure of a matured research network. Based on described features catalog, next stage of research is to determine, implement, and integrate a viable solution of MS-BB IT infrastructure for archiving, retrieving, and distribution of the microscopic images into the LIMS. The limitations of this project need to be considered. First, currently the study presented here has not evaluated the usage of digital scanners of different vendors, but it is planned to develop vendor neutron DPS within the scope of KKNMS. Second, underlying workflows resemble the MS-BB project; Different projects may have distinct workflows and thus different needs, but the goal was to implement a working prototype that can later be extended to other projects. It is expected that web-based DPS will probably become useful product to facilitate collaboration among the researchers of KKNMS.

## Literature

- [1] Digital Imaging and Communication in Medicine (DICOM) Supplement 145: Whole Slide Microscopic Image IOD and SOP Classes; c2010 - [cited 2014 Mar 20]; Available from: [ftp://medical.nema.org/medical/dicom/final/sup145\\_ft.pdf](ftp://medical.nema.org/medical/dicom/final/sup145_ft.pdf)
- [2] KKNMS [internet], TU München: Competence Network Multiple Sclerosis; [cited 2014 Mar 20]. Available from: <http://www.kompetenznetz-multiplesklerose.de/>
- [3] K. Pohl. Requirements Engineering: Fundamentals, Principles, and Techniques. Springer; 2010.
- [4] Validation of Digital Pathology in a Healthcare Environment [internet]. Madison: Digital Pathology Association; 2011 [cited 2014 Mar 20]. Available from: [https://digitalpathologyassociation.org/\\_data/files/DPA-Healthcare-White-Paper--FINAL\\_v1.0.pdf](https://digitalpathologyassociation.org/_data/files/DPA-Healthcare-White-Paper--FINAL_v1.0.pdf)

## Acknowledgements

This work was supported by the Competence Network Multiples Sclerosis (01GI1304B), funded by the German Federal Ministry of Education and Research.

## Informationen zum eingereichten Beitrag:

- Beitrags Id: 245
- Kategorie: Big Data und Forschungsinfrastruktur - Perspektiven für die Medizin
- Fachbereich: Medizinische Informatik
- Themenbereich: Bilddatenmanagement für Forschung und Versorgung
- Beitragstyp: Vortrag