

Evaluating DSpace For an Industrial Application

Department of Computer Science
University of Cape Town



by

Gavin Asary (Student Number: ASRGAV001)
Supervised by: Professor Hussein Suleman

Minor Dissertation presented in partial fulfilment of the
requirements for the degree of
Master of Science
in the
Department of Computer Science

February 2024

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

Declaration

I know the meaning of plagiarism and declare that all of the work in the dissertation (or thesis), save for that which is properly acknowledged, is my own.

Signed by candidate

Signature:

Date: 22 February 2024

Contents

Chapter 1: Introduction	5
1.1 Background	5
1.2 Project Objectives	5
1.3 Motivations	6
1.4 Dissertation Organisation	7
Chapter 2: Literature Review	10
2.1 Introduction	10
2.2 Background	10
2.3 Institutional Repositories	11
2.4 Enterprise Content Management (ECM)	12
2.5 Design of Repository Features	14
2.6 Content Management in Business	15
2.7 Summary	16
Chapter 3: Methodology	17
3.1 Introduction	17
3.2 DSpace	17
3.2 Research Question	19
3.3 Methodology	19
3.3.1 Experimental Design and Scope	19
3.3.3 Data Collection and Analysis	24
3.3.4 Research Instrument	25
3.4 Ethical Considerations	28
3.5 Summary	28
Chapter 4: Results	30
4.1 Introduction	30
4.2 Background responses - 'All' respondents irrespective of IT proficiency level	30
4.3 Summary for "All" respondents	34
4.4 Background responses - IT proficiency level of 'Expert or Advanced' vs 'Intermediate & Beginner'	35
4.5 Summary for 'Expert/Advanced' versus 'Intermediate/Beginner' users.	38
4.6 Discussion	39
Chapter 5: Conclusion	41
5.1 Overview of Research	41
5.2 Major Findings and Literature Comparison	41
5.3 Answer to Research Question	41

5.4 Limitations	41
5.5 Future Work.....	42
References	43
Appendix A: Survey Questionnaire (LimeSurvey)	48
Appendix B: Survey Questionnaire Results	53
Appendix C: Ethical Approval (FSREC Ethics Approval 087-2022 Asary)	68
Appendix D: Organisation Consent Approval Letter	69
Appendix E: Online Informed Voluntary Consent to Participate in Research Study	70
Appendix F: Key Business Requirements	71
<i>Table 3-1 - Background Questions</i>	<i>26</i>
<i>Table 3-2 - Functional Questions.....</i>	<i>27</i>
<i>Table 3-3 - Non-functional questions</i>	<i>28</i>
<i>Table 4-1 - Functional Categories.....</i>	<i>32</i>
<i>Table 4-2 - Non-Functional Categories</i>	<i>33</i>
<i>Table 4-3 - Expert and Advanced vs Intermediate and Beginner Users</i>	<i>36</i>

Abstract

The industrial sector is experiencing challenges in managing the escalating complexity and volume of digital content. Traditional Enterprise Content Management (ECM) systems often fall short in addressing these sophisticated needs, leading to inefficient data handling and potential loss of crucial information. This problem is critical as it directly impacts operational efficiency, regulatory compliance, and decision-making processes within industries.

This study proposes the adaptation of DSpace, a tool predominantly used in academic settings, for industrial digital content management in a downstream oil and gas environment in the Western Cape. The goal of the study is to bridge the gap between academic and industry practices in digital content management by examining how DSpace can meet the intricate demands of a business setting.

An online survey assessed the toolkit's functionality, usability, and impact on user experience within the industrial context. Participants, varying in IT proficiency, provided critical insights into the toolkit's adaptability to industrial needs. The results indicate that DSpace is a viable option for businesses seeking efficient digital repository or content management systems. It excels in usability, content management, search capabilities, and metadata handling, suggesting that, with certain adaptations and enhancements, DSpace could meet the requirements of end-users and administrators in an industrial environment.

However, the study also identifies limitations and areas where DSpace differs from traditional ECM systems, indicating potential avenues for adaptation. Specific industrial demands may necessitate further investigation and customisation. This study opens new possibilities for digital content management in industries and advances practices in digital repository management, suggesting that academic solutions like DSpace could be reimaged for commercial use with the right modifications.

Chapter 1: Introduction

1.1 Background

The management of digital assets has become crucial in both commercial and academic enterprises amidst the digital transformation era. As an increasing range of "born digital" content is generated, robust mechanisms to manage, share, and preserve these digital assets are required. This study investigates the feasibility of utilising DSpace, a tool primarily used in academia, in an industrial setting, aiming to bridge the gap between academic and industrial digital content management practices.

The industrial sector is grappling with the complexities of managing digital content. Corporations generate vast amounts of data, including technical reports, project documentation, regulatory compliance materials, and research data, often in unstructured formats. This challenge is compounded by the diverse nature of content, varying file formats, and the need for secure, long-term preservation (O'Callaghan & Smits, 2005). Traditional content management systems often fall short in addressing these sophisticated needs, leading to inefficient data handling and potential loss of critical information. The need for an integrated approach to manage this escalating complexity has never been more urgent.

Enterprise Content Management (ECM) systems, typically used in commercial settings, are designed with strong versioning capabilities, complex access controls, and extensive workflow mechanisms that facilitate document lifecycle management. These systems support dynamic content creation and modification processes. In contrast, DSpace, first co-developed by Hewlett-Packard and MIT (Naik & Shivalingaiah, 2006), is renowned for its capability to manage and preserve digital content primarily through archival ingest, where content remains unchanged once added. Its workflow, metadata, and language are tailored to the submission of final versions of documents, aligning with academic publishing standards.

This research is driven by the curiosity of whether a system performing well in academic settings can be successfully adapted to meet the intricate requirements of an industrial setting. The specific challenges in the industrial sector, such as handling diverse and unstructured content, maintaining stringent security, and ensuring long-term preservation, differ significantly from academic usage. These unique industrial needs necessitate a thorough examination of DSpace's adaptability and efficacy in such environments.

The significance of this study lies in its potential to offer innovative solutions to the growing challenges of digital content management in the industrial sector by examining and investigating the adaptability of DSpace. This research aims to unlock new avenues for digital content management and advance practices in digital repository management.

The goal of this study is to investigate the efficacy and adaptability of the DSpace digital repository software in an environment where commercial Enterprise Content Management (ECM) systems have historically had a dominant position. By conducting a detailed analysis of DSpace's functionality, usability, and impact on user experience within the specified industrial context, this research seeks to ascertain whether the services provided by DSpace meet the requirements of end users and administrators in an industrial environment.

1.2 Project Objectives

The primary goal of this study is to investigate the efficacy and adaptability of DSpace digital repository software within a commercial Enterprise Content Management (ECM) environment,

specifically in the context of the downstream oil and gas industry in the Western Cape. This research gains practical relevance and depth by being a convenience study, as the researcher is conducting the study at his place of work. This unique position provides direct access to both functional and technical resources, including skilled professionals familiar with the current ECM toolset and business processes in this industrial sector. Such access is invaluable in evaluating and comparing the capabilities of DSpace against the existing ECM tool used by organisation.

The increasing complexity and volume of digital material in corporate environments, particularly where unstructured data is prevalent, necessitates this exploration (O'Callaghan & Smits, 2005). The industrial setting of this study, with its specific challenges in managing digital repositories, presents an ideal case study to apply and assess the adaptability of DSpace, a system originally co-developed by Hewlett-Packard and MIT (Naik & Shivalingaiah, 2006) and well-established in academic settings. Its widespread acceptance and use in academic institutions globally and in South Africa (OpenDOAR, 2023) further validate the choice of DSpace for this study.

This convenience study, leveraging the resources and expertise available in the researcher's professional environment, aims to answer the research question: "Are the services provided by the DSpace Institutional Repository toolkit suitable to meet the requirements of end users and administrators in an industrial environment?" The investigation will encompass a detailed analysis of DSpace's functionality, usability, and impact on user experience within this specific industrial context.

1.3 Motivations

The exploration of utilising the DSpace digital repository software toolkit in an industrial environment is driven by several compelling factors that highlight the urgent need for innovative solutions in managing digital content within commercial sectors.

1. Case Study: A Convergence of Context and Opportunity

The selection of the target organisation, situated in the downstream oil and gas industry in the Western Cape, for this case study is deeply intertwined with the researcher's employment there, a factor that influenced the study's methodology and scope. This convenience study, as categorised, leverages the feasibility of conducting research within the researcher's current workplace, thereby aligning with Bryman's (2016) emphasis on the importance of accessibility to necessary resources, staff, and the environment for effective research. Such direct access not only facilitates the collection of rich, detailed data but also provides a thorough understanding of the organisation's unique internal processes and challenges, crucial for the effective implementation and evaluation of the DSpace digital repository software toolkit.

This organisation was selected due to the clear presence of the issue under investigation, making it a relevant setting for examining the practical applications and effectiveness of the DSpace toolkit. The researcher's familiarity and integration within the organisation allow for a tailored approach (Creswell, 2014) in evaluating DSpace's adaptability and efficacy, especially in this specific industrial context. Conducting the study in an environment where the problem is actively present ensures immediate relevance and addresses real-world challenges, in line with Yap and Choy's (2018) assertion that studying issues within their natural environment yields genuine insights and practical solutions. This approach is particularly beneficial in workplace studies, where the researcher's unique insights can enhance data collection and interpretation (Cuya & Palaoag, 2023).

While the research offers insights into DSpace's application and impact in this industrial setting, it is crucial to recognise that the findings are specifically tailored to this organisation and are not

universally applicable to other organisations or industries. This inherent limitation, characteristic of convenience-based studies (Dowling & Mantai, 2017), acknowledges that while some insights may resonate with similar industries, making broad generalisations is not the primary objective of this research. The focus is squarely on addressing the specific needs and challenges of the selected organisation, with an understanding that these findings may offer preliminary guidance for similar contexts, but without claiming universal applicability.

In summary, the rationale behind selecting this organisation for the case study is twofold: the accessibility of the organisation to the researcher and the explicit manifestation of the research problem within it. Although this approach limits the generalisability of the findings, it facilitates a focused, contextually rich exploration of the issue, contributing valuable field insights. The deliberate choice of this organisation and the acknowledgment of the study's specific scope and limitations are intended to pre-empt any queries regarding the research design rationale, underscoring a methodological approach that balances practical feasibility with academic rigour.

2. The Prevalence and Success of DSpace

The choice of DSpace is further motivated by its widespread adoption and success in the academic world. According to OpenDOAR (2023), DSpace is one of the most popular open-source repository systems globally and is extensively used in South African institutions. This popularity suggests a mature, well-supported platform that could potentially be leveraged in new, innovative ways outside of its traditional academic environment.

1.4 Dissertation Organisation

Chapter 2: Related Work

This section reviews prior research on enterprise content management and institutional repositories, examining their features, designs, and applications in business. It also traces the historical development of digital repositories, defines institutional repositories and enterprise content management systems, compares their designs, and discusses their use in supporting business processes and information management.

Chapter 3: Methodology

This section outlines the experimental design, methodologies, and participant selection process used to investigate the applicability of DSpace in industrial settings, including data collection and analysis procedures. It also states the central research question, elaborates on the research instrument and methodology, and addresses ethical considerations to ensure compliance with research ethics protocols.

Chapter 4: Results

This dissertation examines the effectiveness of the DSpace Institutional Repository toolkit, focusing on user IT proficiency levels, with insights drawn from a survey of 30 participants. The findings highlight high satisfaction with DSpace's core functionalities but indicate a need for improvements in advanced features and overall usability to better meet diverse business requirements and user proficiency levels.

Chapter 5: Conclusion

The research evaluated the adaptability and effectiveness of DSpace, traditionally an academic tool, in an industrial setting, addressing whether it could meet the intricate demands of end-users and

administrators in this sector. The findings indicate DSpace's strong capabilities in usability and content management, suggesting its potential for industrial use, though further adaptation and enhancement are needed; limitations and future research areas such as scalability, integration, and comparative analysis with other ECM solutions are also discussed.

Chapter 2: Literature Review

2.1 Introduction

This chapter offers a thorough examination of the digital transformation in information management, highlighting the transition to digital content creation in educational and research institutions and the parallel challenges in corporate environments. It delves into the concept of Enterprise Content Management (ECM) as a strategic response to manage, share, and preserve the burgeoning volume of unstructured content in organisations. The chapter also explores the role of digital repositories, particularly in academic contexts. A detailed analysis is presented on the DSpace digital repository software toolkit, examining its features and applications. The role of institutional repositories in managing research data and facilitating academic communication is discussed, highlighting the dynamic technical environment that necessitates adaptability in these systems. Notably, no other study looked at DSpace usage in industry. The chapter concludes by underscoring ECM's critical role in various industries, illustrating its adaptability and essential utility in managing digital data across professional environments.

2.2 Background

Progressively more research and educational material is created in digital format ('born digital'). Organisations and learning institutions are realising the necessity for a stable platform to manage, share, and preserve content in digital format for long-term access. The Massachusetts Institute of Technology (MIT) is an example of a learning institution who had this need as a considerable amount of material produced by faculty, including, but not limited to, conventional document-based material (articles & reports), experimental results, datasets, rich media data, etc. was saved on individual's hard drives or departmental web servers, which were often lost when faculty and departments changed over time (Koehler, 2001).

Employees and information workers of corporate or industrial environments experience a similar challenge as documents are stored in different locations, systems, in different versions, formats, or languages. Collaboration on documents is difficult and these are shared using email. Content management on an enterprise level is difficult as unstructured data accounts for 80% of content (O'Callaghan & Smits, 2005). Data and information quality of unstructured data is crucial as it contains critical, innovative, and decision-relevant information that has become a critical business resource. Due to the rise of content chaos, enterprise content management (ECM) was created as an integrated information management approach; ECM has garnered a lot of attention in the commercial industry (Smith & McKeen, 2003). Over time, ECM solutions tailored for commercial and industrial use have evolved, becoming more advanced in their capabilities and execution. This progression indicates a marked enhancement in the complexity, functionality, and user accessibility of these systems, providing increasingly effective and efficient content management solutions for organisations (Tyrväinen, Päivärinta, Salminen, & Iivari, 2006).

Digital repositories on the other hand are online archive software systems used to manage, share, and preserve content and make the content available in digital form. The term 'Digital Repository' can also be used to refer to an organisation or organisations responsible for the long-term maintenance or preservation of digital resource content and providing access to these digital resources (Wolski, Simons, & Richardson, 2013).

Digital repository toolkits provide the basis for the creation of digital libraries (Dale, 2001). While initially custom-created for each instance, modern institutional digital libraries are based on toolkits with common functionality and architectural models. These toolkits have evolved in tandem with the evolution of metadata practices and standards and developments in information retrieval. In contrast, document management systems or content management systems are used in industrial organisations to manage documents in mostly commercial settings, and their evolution has tracked developments in databases and Web technology.

The increasing adoption of digital repositories in academic environments is driven by the need to organise and offer access to an extensive range of digital content, encompassing faculty work, student theses, dissertations, e-journals, and datasets (Foster & Gibbons, 2005). These repositories serve a variety of important functions. In the field of cultural heritage distributed object repositories act as essential platform for experts to access, share, and enhance digital content which plays a vital role in the preservation and dissemination of cultural knowledge (Sabharwal, 2021). In the academic domain, specialised repositories such as the HZSK Repository cater to particular data types like spoken language corpora, meeting the specific requirements of linguists and socio-psychological researchers (Jettka & Stein, 2014).

There are various types of digital repository software toolkits available that share similar features and to a certain degree are developed on the same technology. Some of the commonly used digital repository software toolkits include DSpace, EPrints and Fedora (Lumpa & Suleman, 2019).

DSpace is an example of a digital institutional repository toolkit that is used to capture, store, index, preserve, and redistribute a university's research faculty's scholarly intellectual material in digital formats (Naik & Shivalingaiah, 2006). The software toolkit manages and distributes digital content and facilitates the creation, indexing, and searching and retrieval of items based on associated metadata. It was designed and co-developed by the Massachusetts Institute of Technology (MIT) Libraries and Hewlett-Packard (HP) as an open-source application for institutions and organisations to run with moderately few resources. It supports the long-term preservation of the digital material that is stored in the repository and designed to facilitate the easy submission, management, and access of digital content.

2.3 Institutional Repositories

The purpose of hosting institutional repositories has changed over the last decade as institutional repositories arose from the need to archive and store academic material, particularly research publications (Jain, 2011). Lynch (2003) described the scope of institutional repositories as a set of services that extend beyond pure archiving applications to perform administrative roles.

Lougee (2009) identified movements towards improving discovery and data sharing services with added focus on content lifecycle management from data collection to publication and retention. The role of institutional repositories has been described as a richer virtual environment that encourages community engagement (Reznik-Zellen & Adamick, 2012).

Modern institutional repositories are rich ecosystems of data storage, content management functions, discovery, collaboration services and access management. Jain (2011) positions institutional repositories as ongoing services with evolving requirements and technologies that require a clear vision, strategy, and tactics.

Research data is identified by the requirement to link to external services and tools due to the importance of external discoveries, using multiple formats within and between research projects.

The need to manage research data throughout the entire life cycle to meet the needs of a variety of specific subjects has created an institutional repository landscape that spans multiple technologies and services. The growing need for scholarly or academic communication and collaboration and the associated dynamic technical environment requires institutional repositories to be flexible and adaptable to new requirements or trends.

2.4 Enterprise Content Management (ECM)

Enterprise Content Management (ECM) encompasses a range of techniques, methods, and tools which Alalwan (2012) describes as pivotal for capturing, managing, storing, preserving, and delivering materials and documents integral to organisational processes. Furthermore, Alalwan (2012) notes that ECM technologies and strategies are crucial for managing unstructured information within an organisation. ECM was defined by Smith and Mckeen (2003) as the strategies, procedures, methods, systems, and technologies required to capture, create, manage, use, publish, store, preserve, and dispose of material within and between organisations. Alalwan and Weistroffer (2012) describes document management, online content management, records management, document imaging, document-centric communication, and workflow as essential components of ECM.

ECM is all about the creation and dissemination of information, which can include documents and their management, web content management, and records management, among other things (Iverson & Burkart, 2007). These are accomplished through the five major components of ECM: capture, manage, store, preserve, and deliver (Dammak, Dkhil, & Gardoni, 2019). “Capture” entails creating information by converting paper documents to electronic formats, obtaining, and accumulating electronic files into a cohesive structure, and organising information. Students' profiles, grades, contracts, invoices, research reports, and so on are examples of such information. The “Manage” component entails connecting, modifying, and employing information through collaborative software, document management, records management, and web content management. The “Store” component requires temporarily backing up frequently changing information into flexible folder structures to allow respondents to see or change information. The “Preserve” component involves the medium and long-term backup of infrequently changing information, which is often accomplished using records management features. This is typically used to assist firms in complying with government and other requirements. The “Deliver” component entails supplying desired information to end respondents and clients (Munkvold, et al., 2006).

Enterprise Content Management (ECM) offers organisational benefits, primarily through the effective management of both structured and unstructured content across its entire lifecycle. This management includes simplifying heterogeneous data, which encompasses a variety of formats and originates from multiple sources. ECM efficiently processes information, whether it's structured data, organised in a specific format, or unstructured data that lacks a predefined structure (Mendes & Bax, 2018). This enables organisations to create content about their brand and culture that is useful not only to their employees but also to their customers and other stakeholders. As a result, information that formerly required a consumer to visit an organisation's office may now be received online with the click of a button when accessing the company's website and is verified with a valid account (O'Callaghan & Smits, 2005).

O'Callaghan & Smits (2005) defined Enterprise Content Management (ECM) as a synthesis of Document Management (DM) and Content Management (CM) for managing unstructured information in organisations.

The goal of DM systems is to organise and make files accessible and comprises of the following components (O'Callaghan & Smits, 2005)

- File storage: The method used to store and organise data that is in the DM system;
- File categorisation: The user can allocate file types and groups based on their criteria;
- Metadata services: The user can add metadata;
- Collaboration services: The user can collaborate on files while they are checked in or out;
- Workflow services: Content is routed via a series of defined phases that support a business process;
- Versioning services: The system saves all versions of a document and allows users to retrieve them when required; and
- Access services: Tables of contents, indexes, and full-text searches can be used to locate files.

Originally, content management (CM) was associated with Web development, however with the evolution of the web the need for more powerful administration tools became critical and CM systems were born (Hackos, 2002). CM refers to the complete process of gathering, managing, and publishing information to any source. There are several similarities between CM and DM systems with both striving to categorise information, apply metadata, organise development through workflow and collaboration, and provide full access to end users. However, the systems differ in the following areas (Boiko, 2002):

- DM systems deal with files, while CM systems deal with content components. Files are not good containers for content as they can hold only one component per file, so their metadata needs to be clearly defined to find individual components.
- DM systems manage files created by other sources while CM systems create and manage content. While both systems control the creation of metadata, only in CM does the user control the creation of content.
- DM systems provide access to individual files while CM systems provide access to all the components of the content.

The principles and components outlined in ECM, such as document management, web content management, and the overall strategy for handling digital content, lay the foundation for understanding how digital repositories are structured and function. These repositories not only incorporate similar components and management strategies but also extend these concepts to accommodate the specific needs of storing, preserving, and providing access to digital collections.

As we explore the design of repository features, it's important to recognise how ECM strategies influence these designs. The transition from ECM to digital repository design involves a shift in focus from a broad organisational content management perspective to a more specialised, feature-driven approach tailored to support the efficient retrieval and preservation of digital assets. The following section will explore how the features of digital repositories are designed to meet the complex needs of users and organisations, drawing on the foundational principles of ECM to highlight the evolution of content management strategies into the digital realm.

2.5 Design of Repository Features

The intention behind this section is to present a focused review of related work that highlights the broader context of digital libraries, specifically how the design principles and features of these systems, such as DSpace, can be leveraged in various environments, including industrial settings. This sets the stage for investigating the potential use of DSpace in an industrial context.

Papadopoulou (2011), describes the term "digital library" as an organised collection of digital content that is accessible to the public and can include material that has been converted into digital format, like digital versions of books and other physical material originally from archives or libraries. These systems range from digital objects, metadata repositories, archiving, reference linking systems and content management systems to complex systems integrating advanced digital library services (Soergel, 2009). They represent the meeting point of many disciplines and domains, such as data management, information retrieval, document management, library science, information systems, the web, image processing, artificial intelligence (AI), human-computer interaction (HCI), and digital curation (Candela, et al., 2007). The following design principles assist with the design of digital library systems (Vishwanthan, 2001): to expect changes, know the content, involve the appropriate resources, design a usable system, ensure open access, be conscious of data rights, provide as much automation as possible, adopt standards, ensure quality and be concerned with persistence. These principles are particularly relevant when considering the application of digital libraries like DSpace in industrial environments, where adaptability, usability, and data integrity are crucial.

The purpose of a digital library, as described by Sun & Yuan (2012), is to gather, organise, safeguard, and provide easy access to digital content. It achieves this through a user-friendly interface, supports traditional library tasks, offers improved search and retrieval options, allows for complex searches difficult to do manually, safeguards the rights of content creators, and uses digital and network technologies to preserve special collections.

Content Management - This feature is associated with the convenience with which content material is created, submitted, reviewed, prepared and versioned inside the digital library and includes the searching and browsing functions such as metadata searches, full-text searches, and hierarchical subject browsing (Austerberry, 2007). It further includes the support of content encoding in various formats, including text (ASCII, UNICODE, etc), image (GIF, JPEG, TIFF, etc), presentation (Adobe PDF, etc), structured formats (HTML and XML), audio and video (MP3, AVI, MPEG, etc). These capabilities can be instrumental in managing the vast amounts of data generated in industrial environments.

User interface - Witten and Bainbridge (2002) contend the user interface (UI) requirement includes the flexibility to customise the UI to meet the requirements of different types of digital library implementations, including the support for multilingual access. Multilingual access allows users to specify the language for the digital library's UI and the cataloguing information stored within the digital library. A customisable and user-friendly interface is critical in industrial settings to ensure that the system is accessible and usable by a diverse workforce.

System administration - The system administration requirement focuses on back-end maintenance functions (Cordeiro, 2004), including automated content acquisition, harvesting, automatic metadata generation encompassing named entity recognition and automatic subject indexing and classification. It further includes support of preservation standards and persistent document

identification to facilitate the transfer of digital materials from one platform to another without compromising reference citations and other links. Efficient system administration is essential for maintaining the reliability and performance of digital libraries in industrial environments.

Other requirements – Digital repositories benefit greatly from supporting interoperability protocols, such as OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) and Z39.50 as these protocols facilitate seamless communication between different systems while allowing each to evolve independently (Kaczmarek & Naun, 2005). Adherence to standards for digital library collections and services is crucial; this includes the proper representation of information and images, the management of metadata, multilingual support, and the facilitation of information interchange (Goh, et al., 2006). In addition to the technical requirements, it is important to have formalised support structures in place, which includes a helpdesk for immediate support, easily accessible system documentation, the presence of active online forums, effective mechanisms for tracking defects or bugs, and avenues for suggesting new features (Steehouder, 2002).

As we transition to discussing "Content Management in Business," it's important to understand that the concepts and tools of digital repositories go beyond academic or archival uses and play a significant role in the business sector. Efficient content management is key to business success and strategic planning. We will explore how digital repository designs are applied in business settings, emphasising how they revolutionise the way organisations manage their digital assets, spread knowledge, and make informed decisions.

The core features of digital libraries—like adaptability, awareness of content, user-focused design, and strict data management standards—easily integrate into business practices. These features help businesses streamline operations, increase access to information, and protect their data, effectively connecting the theoretical ideas behind digital repository design with their real-world application in business. This review lays the groundwork for the subsequent investigation into how DSpace, a digital repository system, can be effectively utilised within industrial environments to enhance data management and operational efficiency.

2.6 Content Management in Business

Information is an organisation's most valued asset; this includes company records, internally generated intellectual property, customer, and partner interactions. Most of this information is digitised, which benefits and transforms the entire organisation as digital media is easy to capture, store, share and dispose of when needed. However, most of this information is siloed and exists in isolation as it can only be used within the process or application where it was created, making the data ungoverned, harder to access, and difficult to distribute across the organisation (Opentext, 2023). Organisations struggle to manage and extract value from information as the amount of information grows and business models require new ways of interacting, exchanging, and collaborating to meet increasingly competitive, regulatory, and customer service requirements (Tankard, 2012). Content Management Services (CMS) platforms are designed to go beyond traditional organisational and management capabilities by delving into the enterprise resource planning (ERP) systems responsible for generating information and seamlessly integrate with ERP systems such as SAP, Oracle, PeopleSoft, and others. This integration allows for the immediate analysis of content at the point of creation, thus ensuring a more streamlined and efficient flow of data across the various ERP systems (Palaniswamy & Frank, 2000).

Smith and McKeen (2003) explore content management as a holistic approach to handling every type of information asset within an organisation, ranging from traditional documents and reports to web pages and digital assets. The application of content management in automating document workflows and managing digital assets underlines its importance in optimising business processes and boosting organisational efficiency. By adopting content management solutions, companies can streamline document management tasks, minimise the need for manual handling, and enhance the efficiency of document access and storage. In their study, Gupta, Govindarajan, and Johnson (2001) examine a range of content management techniques, with a particular focus on the distributed model. This model promotes the decentralisation of content creation and governance across different organisational divisions, an approach that is highly beneficial for large entities with department-specific content management requirements. Allowing individual departments to independently manage their content ensures that information remains relevant and specifically tailored to meet the needs of its intended audience, ultimately boosting engagement and relevance. Dobrojevic (2018) explains how web content management software helps make customer experiences better and business operations smoother. In retail, a good content management system can make online shopping better by providing personalized content and easier navigation, which can lead to more sales and stronger customer loyalty.

Overall, the widespread adoption of content management tools in various sectors reflects their importance in enhancing information management, decision-making efficiency, collaboration, and security across professional environments.

2.7 Summary

In conclusion, this chapter addressed the pivotal role of digital transformation in information management within both educational and industrial landscapes. It highlights the increasing shift towards 'born digital' content and the subsequent challenges posed in managing, sharing, and preserving this digital material. Central to this discourse is the exploration of ECM systems, which emerge as crucial solutions for managing the vast and complex array of unstructured data in organisations. The discussion extends to the significant role and continuous evolution of digital repositories, particularly in academic settings, where they are instrumental in organising and providing access to diverse digital content. Overall, the chapter encapsulated the ongoing developments in digital information management, underscoring the necessity and challenges of maintaining dynamic, adaptable systems capable of supporting the growing demands of e-research, e-learning, and organisational information management in the digital era.

Chapter 3: Methodology

3.1 Introduction

In this chapter, we discuss the suitability of DSpace, an established digital content management tool, for industrial applications. The focus is on evaluating its ability to meet the intricate needs of administrators and end users in these settings, through the lens of an online survey. The chapter provides a brief overview of DSpace, which includes an examination of its key functionalities, architectural design, and its applicability across various institutional landscapes. The methodology section details the experimental approach, designed to test the feasibility of DSpace in handling typical industrial tasks such as content creation, editing, and distribution. The section on experimental design and scope explores the objectives of the experiment by evaluating DSpace against the critical business requirements deemed essential for a successful ECM system. Our investigation encompasses both functional and non-functional aspects, providing a comprehensive view of the system's capabilities. Data collection and analysis are discussed, identifying the use of LimeSurvey to gather detailed insights into the functionality and user experience of DSpace. The chapter concludes by addressing the ethical considerations in the research.

The suitability of solutions such as the DSpace Institutional Repository toolkit in fulfilling the complex requirements of administrators and end users in industrial environments is critical in the ever-changing field of digital content management. The purpose of this study, using an online survey is to investigate if the services offered by the DSpace toolkit are appropriate in an industrial setting. A key component of the approach is a user assessment process that is specifically designed to determine whether the solution is feasible in such an environment, especially considering common procedures such as the production, editing, and sharing of digital information. These procedures, which depend on the smooth operation of modules, are essential for preserving control over corporate records and content throughout organisations.

This experiment includes both administrator and end user tasks. The participants' input will be critical in assessing how well the toolkit's features and services meet the unique needs of an industrial environment and, consequently, how adaptable it is. Their opinions will be very important in determining how effectively the toolkit's features and services satisfy the requirements of an industrial setting and, in turn, how flexible it is. The information gathered will be useful in assessing if the current services are necessary, sufficient, or need to be improved to better serve the needs of industrial users.

3.2 DSpace

DSpace is an open-source platform that serves as a digital repository for storing and managing digital research and educational materials produced by organisations or institutions (Tansley, et al., 2003). The software was developed to address the fundamental needs of digital repository services while providing a basis for further enhancements, particularly in the area of long-term digital preservation. The software toolkit allows users to upload, save, store, and share content with other users with the primary focus to preserve stored data. It uses digital storage features such as saving checksums with digital objects to track and verify the integrity of files with the original. It can be used by various communities within an institution such as departments, laboratories, and centres with each having their own distinct locations within the system. Community members can upload content directly through a Web user interface or as a batch item importer for bulk content importing. Each community may designate gatekeepers or administrators who examine and edit submissions before they are added to the main repository. The system indexes the submission's metadata with the digital content to make it available based on the access rights or privileges granted to the user.

DSpace was created 'breadth-first' to provide a functional service in the time available, that is each basic requirement of an institutional digital repository was met in a reasonably simple manner, allowing functionality to expand with the service in production. The underlying architectural model controls the way that users interact with the system, how content is submitted, utilised and how administrators configure the system. DSpace provides APIs for administrators and third-party applications to interact with the system. The system allows for configurable submission and a workflow process that can be configured to align to an individual organisation's practices and policies (Tansley, et al., 2003).

DSpace is a full stack web application that comprises of a three-layer architecture (Bass, et al., 2002), which includes a Services or Application Layer , a Business Layer and a Storage layer, as indicated in Figure 3-1 - DSpace system architecture.

1. **Application Layer:** This layer functions as the system's interface with the outside world, which includes the Web user interface (UI) and the support of metadata harvesting services (Tansley, et al., 2003).
2. **Business Logic Layer:** This layer focuses on management of the functional operations of the archive, which includes the archive's content, its users, workflow management and authorisation processes (Xiang, et al., 2019).
3. **Storage Layer:** This layer is primarily responsible for the physical storage of both metadata and content (Wu, Liu, Li, & Cai, 2011).

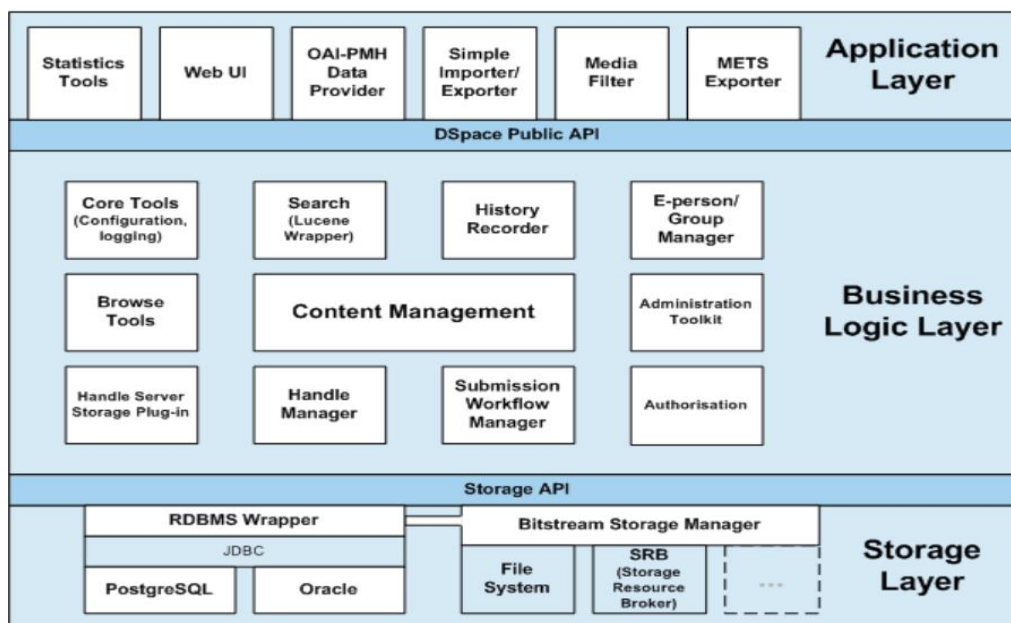


Figure 3-1 - DSpace system architecture (Tansley, et al., 2003)

A critical feature of this architecture is the hierarchical interaction between layers with each layer only being able to interact with the next layer down in the hierarchy. The application layer has no direct contact with the storage layer. Additionally, the storage and business logic layers come with their own public APIs which could be made available as Web services through components in the application layer (Hu, Xue, & Liu, 2011). The modular design of each component within the storage and business logic layers enables components to be modified or replaced without significant system alterations which enhances system flexibility and adaptability. The system's layered trust assigns the responsibility of authentication to the application layer. This design choice introduces the risk of

insecure applications directly accessing the Public API; consequently, securing the application layer is of utmost importance (Arsanjani & Pour, 2001).

DSpace supports a wide range of user roles, ranging from general users engaging with its interface to system administrators overseeing its operation. General users have access to features like search and browse, discoverability, content access, and alerts for new additions. Registered users can submit and manage their work, while community and collection administrators oversee submission workflows, metadata, and collection management. System administrators have comprehensive control, including repository configuration, user management, and system maintenance, alongside accessing usage statistics for insights into repository impact (Bass, et al., 2002).

By integrating foundational digital repository functions with a multi-layered architecture, DSpace emerges as a multifaceted platform capable of meeting the needs of both academic and institutional content management. Its array of features is meticulously designed to cater to the specific needs of various user groups, thereby enriching the repository's overall functionality and the user experience. From facilitating basic browsing and content submission to enabling sophisticated administrative and system management activities, DSpace enables user to effectively engage with and contribute to the repository. This adaptability and potential for customisation establish DSpace as a crucial resource for institutions aiming to efficiently manage and share digital content (Tansley, et al., 2003).

3.2 Research Question

Companies use a range of different ECM software solutions for document management; research institutions and academia also have document management requirements and these institutions use different solutions. This research is investigating if these two ideas can come together by investigating the applicability of using the DSpace software solution from academia in an industrial or commercial setting. This will be done from an end user perspective, in an industrial environment, to determine the suitability of the services and functionality provided by the DSpace software toolkit. This will be done by assessing the software toolkit's features, with common functionality and architectural models to understand the impact on the user environment.

The research seeks to answer the following question: Are the services provided by the DSpace Institutional Repository toolkit suitable to meet the requirements of end users and administrators in an industrial environment?

The research question will be tested in the context of a downstream Oil and Gas industrial environment in the Western Cape only. This is chosen for convenience as the researcher has access to the organisation and understands the needs of the organisation related to digital repositories or ECM toolkits.

3.3 Methodology

This experiment leverages the DSpace digital repository, installed on a virtual server within the UCT Computer Science Department's server room. The goal is to evaluate the system's feasibility in an industrial setting, focusing on typical user and administrator tasks such as content creation, editing, and collaboration.

3.3.1 Experimental Design and Scope

The primary objective of this study was to assess the effectiveness of the DSpace system against specific business requirements identified as crucial for a successful Enterprise Content Management

(ECM) system deployment. The key requirements for the experiment were based on a business requirements document, an internal document of the organisation, that was formulated by a project team tasked with implementing an ECM system. This document contains key requirements identified as vital for the successful deployment of the ECM system within the organisation. These requirements were used to understand the expected functionality and tasks against which DSpace was evaluated.

The list of 12 key requirements includes:

1. Central Document Repository.
2. Ability to extract specialised clauses from documents.
3. Ability to perform document renewals & terminations.
4. Ability to monitor milestones.
5. Ability to create automated workflows with email notification.
6. Ability to store and access generic document templates.
7. Easy Search Capability.
8. Ability to scan and upload documents.
9. Ensure secure document data & information, including appropriate access control.
10. Reporting & Analysis.
11. 3rd Party Tool Integration i.e., SharePoint.
12. Multi Country Support.

These requirements provided a structured framework to evaluate the functionality and performance of DSpace, ensuring it meets the specific needs of our organisation. Refer to Appendix F: Key Business Requirements for a detailed list and description of these requirements.

The scope of the experiment encompassed user testing of various functional and non-functional aspects of the DSpace system. The functional aspects included access management, workspaces and sub-workspace management, item submission, metadata management, workflow, and notification management, browsing and searching, preservation, storing of multiple file formats, lineage, submissions finalisation, and document uploading. The non-functional aspects included ease of use, user interface, navigation, efficiency, performance, and limitations on file size.

The tasks assigned to participants were carefully designed to reflect the daily activities and specific requirements of users within the oil and gas industry. These tasks were structured to evaluate the effectiveness of the DSpace system against the key business requirements essential for the organisation. Participants were required to:

1. **Easily Access and Navigate DSpace:** This task assessed the user interface and ease of navigation, ensuring that users could intuitively find and access the information they needed.
2. **Creation of Workspaces and Sub-Workspaces:** Participants were tasked with creating organisational structures within DSpace, simulating how workspaces and sub-workspaces are used to manage projects and documentation in the oil and gas sector.
3. **Populating and Editing Workspaces:** This involved adding and modifying content within workspaces, testing DSpace's capabilities in content management and collaborative editing.
4. **Advanced Search Capabilities:** Users needed to search for documents using various criteria such as document name, metadata, keywords, or specific clauses, reflecting the need for robust search functionality in managing vast amounts of documentation.

5. **Content Upload, Submission, and Storage:** Participants tested the ability to upload, submit, and save documents in different formats, essential for maintaining a comprehensive digital repository.
6. **Using DSpace as a Central Repository:** This task focused on evaluating DSpace's effectiveness in serving as a central repository for storing and preserving critical documents.
7. **Tracking Changes and Version Control:** Participants were required to track changes and manage document versions, ensuring proper version control and historical accuracy.
8. **Setting Up Automated Controls and Workflows:** This included creating automated workflows, setting up notifications, and configuring controls, crucial for streamlining business processes.
9. **Ensuring Multi-Country Support:** Tasks were designed to test DSpace's ability to support operations across multiple countries, which is a key requirement for global oil and gas operations.

These tasks directly align with the organisation's key business requirements, which include the need for a central document repository, the ability to extract specialised clauses, perform document renewals and terminations, monitor milestones, create automated workflows with email notifications, store and access generic document templates, facilitate easy search capabilities, scan and upload documents, ensure secure document data with appropriate access control, provide robust reporting and analysis, integrate with third-party tools like SharePoint, and support multi-country operations. By evaluating DSpace against these tasks, the study aims to determine its suitability and effectiveness in meeting the specific needs of the oil and gas industry.

Testing Environment

Setup: DSpace was installed on a virtual server located in the UCT Computer Science server room. The installation process followed the standard instructions provided.

User Preparation: Four user accounts were created, with access to DSpace granted for conducting user functionality tests. Workspaces and sub-workspaces were set up for participants to store and manage digital content. Users received three test files in different formats (MS-Word, PDF, and Excel) via email for upload during the experiment.

Test Case Preparation

A comprehensive set of high-level test scenarios was developed. These scenarios encompassed detailed test cases addressing each functionality and requirement outlined in the business requirements document.

Test Scenarios:

- **Access Verification:**
 - Users logged onto the DSpace repository with provided credentials to verify system access.
- **Workspace Navigation:**

- Participants located specific workspaces or sub-workspaces for file addition, utilizing browsing or searching functions.
- **Submission Initiation:**
 - Users selected the option to start a new submission in the identified workspace.
- **Metadata Entry:**
 - Participants entered metadata for the item, including title, author(s), date, keywords, etc.
- **File Upload:**
 - Users uploaded the three test files to examine the system's capability to handle multiple file formats.
- **Submission Review:**
 - Following metadata entry and file upload, users reviewed and submitted their submissions.
- **License Granting:**
 - Participants granted a license for their submitted items.
- **Confirmation:**
 - Users confirmed the successful upload and integration of files within the workspace.

Test Case Execution and Sample of Users

This study utilised a convenience sampling method, leveraging the access to resources and participants within the researcher's workplace. Due to the researcher's access within the organisation, the study could gather in-depth data and understand the organisation's unique processes. The experiment consisted of a group of 20 participants who participated due to their readiness to engage in the research.

Participants were recruited through their job roles, which gave them exposure to the existing ECM system at the organisation. The recruitment process included the following groups:

- **Graduates from the ECM OEM's Skills Development Initiative:** These participants were employed by the ECM OEM and placed at the researcher's organisation to gain on-the-job training.
- **Job Role-Based Recruitment:** Participants whose job roles required them to perform tasks in the ECM systems, such as superusers, first-line support staff, technical administrators, and functional administrators.
- **Departments Involved:** Participants were drawn from business departments, training departments, and the IT department.

All participants were computer literate and used technology as part of their job roles, albeit with varying skill levels. The sample included:

- **Technology Experts:** Some participants were technology experts who, while proficient with general IT systems, might not be experts in ECM technologies.
- **ECM System Administrators:** These participants were responsible for the administration of the ECM system.
- **IT Experts:** Participants who technically administered the ECM system but might not have specialised knowledge of ECM functionalities.

Most participants actively used the current ECM system in their jobs, providing a relevant and practical perspective for the study. This diverse participant pool ensured that the study captured a wide range of insights and experiences with the ECM system, enhancing the robustness and applicability of the research findings.

The participants were divided into three sub-groups to ensure a balanced mix of experience levels to ensure that each subgroup consisted of members with comparable levels of experience, thereby promoting uniformity in the experimental process. Participants were invited to attend an interactive Microsoft Teams meeting where they were briefed the experiment procedure explained and test cases executed on DSpace. Post test case execution, they were directed to complete a survey via LimeSurvey, capturing their experiences and feedback.

3.3.2 Research Philosophy and Approach to Theory

This research followed a positivist approach, focusing on gaining knowledge through clear, measurable observations and analysing data. It used a quantitative method to carefully study and measure the observed events, aiming to provide factual conclusions based on statistical evidence and reproducibility (Avis, 2003). The research method used deductive reasoning, beginning with testing a premise to confirm or deny if the theory applies in certain situations. It relied on organised research methods and mathematics models to keep the results objective and reliable (Babones, 2016).

The positivist philosophy used in this study is consistent with the opinions of academics such as Bryman (2016), who contend that reality is objective and that it is possible to observe and characterise it from an objective vantage point without affecting the phenomena under study. When it comes to quantitative research, positivism works especially well since it focuses on measuring observable phenomena to test theories. This is a deductive technique that follows the guidelines provided by Saunders et al. (2009). It is based on the formulation of theories and research-based hypotheses that are then verified by empirical observation. The purpose of this study is to evaluate the DSpace software tool's functionality, an important topic in digital repository administration. We will empirically investigate the idea that user satisfaction is influenced by functional and non-functional features of the software, as well as IT background variables.

The methodology used is quantitative, and Creswell's (2014) work supports this decision. It involves gathering numerical data using an online poll. This approach is suitable for testing predetermined theories and drawing generalisations about the DSpace userbase as a whole. The survey measures respondents' impressions and experiences with the DSpace software application using a 5-point Likert scale, a method that has been verified in a study by Allen and Seaman (2007).

As recommended by Trochim and Donnelly (2006), respondents are chosen based on their IT background, guaranteeing a diversity of experiences and competence levels. This methodology enables a thorough comprehension of the ways in which various user groups engage with and view

the DSpace application. Functional and non-functional survey questions cover both the software's specialised functions and more general operational elements, as defined by Nielsen's research (1993).

This study aims to improve the management of digital repositories and the evaluation of software by carefully and fairly reviewing the DSpace software toolkit in a specific organisation. Acknowledging that the findings, as noted by Lumpa, Munyaradzi, and Suleman (2011), will be unique to this organisation and not widely applicable, we design this research to help refine the DSpace toolkit and similar software systems. By employing a positivist and deductive framework alongside quantitative methods, we aim to generate insights that, while specific to the researcher's organisation, offer valuable lessons for the targeted improvement of digital repository management.

3.3.3 Data Collection and Analysis

The purpose of the online questionnaire is to seek to answer whether the services provided by the DSpace Institutional Repository toolkit was suitable to meet the requirements of end users and administrators in an industrial environment. The online survey consisted of three sections with questions first related to the respondent's work background. The second section consisted of functional questions based on the DSpace software toolkit & user business requirements and the third section related to non-functional requirements.

Online surveys are an efficient and cost-effective way to collect data from varied geographical regions, increasing the sample's representativeness (Sue & Ritter, 2012). Furthermore, they enable the collection of standardised data, which is required for objective analysis and comparison (Creswell & Creswell, 2017). Online surveys' intrinsic scalability and replicability further add to the robustness of quantitative research, making it a preferred choice in studies intending to establish patterns, test hypotheses, or investigate connections (Smith J. , 2013).

When assessing a software application via an online survey, it is important to include background questions about respondents' years of Information Technology (IT) work experience, IT competence level, ECM experience, and perceived value of employing ECM solutions. These background questions offer a more in-depth insight of the respondents' operating environment, allowing for a more nuanced interpretation of the survey results. Years of IT work experience, for example, can provide insight into the respondent's knowledge with technological trends and adaptability to new technologies (Johnson & Turner, 2003). Similarly, analysing IT competence level aids in determining the ease of use and learning curve associated with software from the perspective of users of varied skill levels (Nielsen, 1993). Experience with ECM in particular is important since it directly correlates to how users perceive and interact with the tested tool, influencing their comments on usability and functionality (Wilson, 2018). Furthermore, understanding the importance respondents place on using ECM tools shows their intentions and expectations, which are critical in interpreting how they evaluate the program (Baker & Sinkula, 1999). Such demographic and experiential data augment the survey findings, providing a holistic view that extends beyond simple functionality evaluation to include user experience and acceptance variables (Venkatesh & Bala, 2008).

Non-functional questions are critical in evaluating a software application, especially when incorporated into an online survey. These questions are intended to elicit information about the software's practical aspects, such as usability, efficiency, reliability, and compatibility (Nielsen, 1993). Researchers can acquire significant data on how effectively the tool functions in real-world circumstances by asking participants about their direct experiences with the software's performance (Rubin & Chisnell, 2008). Functional questions aid in the identification of potential areas for improvement, ensuring that the software meets the expectations and requirements of the users (Lewis, 2014). They also provide insights into user satisfaction and overall user experience, both of

which are critical measures for the success and acceptance of any software tool (Tullis & Albert, 2013). This method is consistent with user-centred design concepts, emphasising the significance of understanding user wants and experiences in software development (Norman, 2013).

3.3.4 Research Instrument

The process of data collection was conducted using the LimeSurvey online survey tool. A new account was registered on LimeSurvey, and a new online survey was created. The survey was structured into three parts; background questions, functional, and non-functional questions. The functional questions came out of a design process that was based on a business requirements document that was previously created by a project team that implemented an ECM system. A review of previous research and case studies on the implementation and evaluation of similar systems informed the inclusion of questions that have been proven to yield valuable insights into non-functional requirements. The document contained key business requirements identified as vital for the successful deployment of the ECM system. The survey questions were based on the key business requirements and evaluated against functionality and user experience of the DSpace system. Please refer to Appendix A: Survey Questionnaire (LimeSurvey).

The background questions were developed to understand the respondent's years of Information Technology (IT) work experience, IT proficiency level, ECM experience and the value they see in using ECM tools.

Each question had different options as per *Table 3-1 - Background Questions*. The questions were asked in the order as specified in the table.

Seq #	Question	Option Type	Response Options
1	Number of years using Information Technology in a work environment?	Dropdown list	1 2 to.. 20+
2	How would you rate your overall proficiency in using Information Technology in a work environment.	Dropdown list	Beginner Intermediate Advanced Expert
3	How often do you use Enterprise Content Management (ECM) technologies?	Dropdown list	Never Yearly Monthly Weekly Daily
4	I find ECM technologies useful as a tool for my work (1=Strongly Disagree and 5=Strongly Agree)	5-Point choice	1=Strongly Disagree 2=Disagree 3=Neutral

			4=Agree 5=Strongly Agree
--	--	--	-----------------------------

Table 3-1 - Background Questions

The second section of the questionnaire relates to functional questions based on the DSpace software toolkit & user business requirements. The user business requirements were based on commonly used tasks as identified by business users. Each question had 5 options with the following ratings (1=Strongly Disagree, 2=Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree). The questions were asked in the order as specified in *Table 3-2 - Functional Questions*.

Sequence #	Question	Option Type
5	I can easily access DSpace.	5-Point choice
6	I can easily create Workspaces and Sub-Workspaces (Communities and Collections).	5-Point choice
7	I can easily edit a workspace or sub-workspace (communities and collections).	5-Point choice
8	I can easily populate a workspace or sub-workspace with a document, image, or other digital artifact.	5-Point choice
9	I can easily update metadata fields.	5-Point choice
10	I can easily browse through workspaces and sub-workspaces (communities and collections).	5-Point choice
11	I can easily search using keywords.	5-Point choice
12	I can easily submit and save documents.	5-Point choice
13	Uploading files to DSpace is straightforward and user-friendly.	5-Point choice
14	I can quickly locate the resources I need in DSpace.	5-Point choice
15	The search functionality in DSpace helps me find specific documents effectively.	5-Point choice
16	DSpace serves as a central repository to store documentation.	5-Point choice
17	DSpace supports my needs in terms of long-term preservation of digital content.	5-Point choice
18	DSpace allows easy searches of the content via metadata.	5-Point choice
19	DSpace allows me to setup a centralised work area to store standardised reusable templates.	5-Point choice
20	I can easily store digital content that has been scanned or captured.	5-Point choice
21	I am satisfied with the variety of file formats supported by DSpace.	5-Point choice

22	I can easily track changes and versions of documents within DSpace.	5-Point choice
23	DSpace has multi-country support and is designed to support multiple languages, character sets, and time zones, which enables it to be used across different countries and regions.	5-Point choice
24	Dates can easily be managed in DSpace.	5-Point choice
25	Customisable workflows can easily be setup in DSpace.	5-Point choice
26	Automated alerts and notifications can easily be setup in DSpace.	5-Point choice
27	Workflows for approving items in DSpace are efficient.	5-Point choice
28	Managing and controlling user access in DSpace is easy.	5-Point choice
29	DSpace meets all my business requirements.	5-Point choice
30	DSpace meets some of my business requirements.	5-Point choice
31	DSpace provides functionality that ECM systems do not provide.	5-Point choice
32	ECM system provide functionality that DSpace does not provide.	5-Point choice
33	DSpace and ECM systems have a high degree of overlap in functionality.	5-Point choice
34	Are there any features that you believe are missing and would improve DSpace?	Text Input

Table 3-2 - Functional Questions

The third section of the questionnaire relates to non-functional questions, which relate to DSpace operational elements rather than its specific behaviours or functionalities. The non-functional requirements are concerned with how the system accomplishes specific processes, as opposed to what the software does. Each question had 5 options with the following ratings (1=Strongly Disagree, 2=Disagree, 3=Neutral, 4=Agree, 5=Strongly Agree). The questions were asked in the order as specified in *Table 3-3 - Non-functional questions*.

Sequence #	Question	Option Type
36	The navigation menu within DSpace is easy to understand.	5-Point choice
37	The user interface of DSpace is cluttered or confusing.	5-Point choice
38	I find it easy to learn and become proficient with DSpace, even as a new user.	5-Point choice
39	I can easily navigate, search, and interact with DSpace.	5-Point choice
40	I can efficiently perform common tasks such as uploading, searching, and retrieving content.	5-Point choice

41	The system easily handles large volumes of content.	5-Point choice
42	The overall speed and performance of DSpace meet my expectations.	5-Point choice
43	The response times for searches or content retrieval operations is acceptable.	5-Point choice
44	There are limitations on the number of files I can upload.	5-Point choice
45	I would recommend DSpace to colleagues or peers.	5-Point choice
46	General comments and feedback.	Text Input

Table 3-3 - Non-functional questions

3.4 Ethical Considerations

This study's ethical framework was strictly followed, assuring compliance with accepted norms of research integrity and participant protection. The UCT Science Faculty Ethics Committee granted formal clearance to the research procedure, which included both the data utilised and the process used, under the approval code FSREC 087 - 2022. This certification represents adherence to ethical norms as well as the protection of participant welfare and data privacy. Refer to Appendix C: Ethical Approval (FSREC Ethics Approval 087-2022 Asary).

Furthermore, as demonstrated by the consent letter from the company, organisational consent was received. Refer to Appendix D: Organisation Consent Approval Letter. This phase was critical in ensuring that the study activities were in accordance with organisational regulations and ethical issues at work.

The study's participation was fully voluntary, emphasising the notion of autonomy. Prior to the survey, participants were thoroughly informed about the research through a detailed background provided. Participants needed to acknowledge and agree to an Online Informed Voluntary Consent to Participate in Research Study. A necessary confirmation phase was included in this procedure, in which participants had to actively declare their understanding and willingness to engage by ticking a box. This process achieved informed consent, which is a key component of ethical research. Refer to Appendix E: Online Informed Voluntary Consent to Participate in Research Study. Course participants made no objections to the usage of their data.

Data was properly processed to maintain confidentiality and preserve participant identities. It was gathered, analysed, and reported in an aggregated manner. This method made it easier to extract significant categories, themes, and trends while keeping individual responses anonymous. The research complies with the principle of beneficence by anonymising the data, protecting participants from potential harm or discomfort that could result from personal data publication.

Overall, these indicators demonstrate a complete and ethical approach to research, harmonising with both academic and professional ethical standards in scientific investigation.

3.5 Summary

In conclusion, this chapter presented an experiment to examine the suitability of DSpace as a digital content management tool for industrial use. Using an online survey, we evaluated its ability to meet the complex requirements of both administrators and end users in industrial environments. The

assessment covered a comprehensive overview of DSpace, highlighting its functionalities, architectural design, and adaptability across diverse institutional landscapes. The methodology adopted for this study has been instrumental in assessing the practicality of DSpace for standard industrial tasks like content creation, editing, and distribution. The experimental design and scope provide insights into how DSpace aligns with the critical business requirements necessary for a successful ECM system in an industrial context. Furthermore, the data collection and analysis, facilitated by LimeSurvey, offer a deeper understanding of DSpace's functionality and user experience. This approach enables us to draw meaningful conclusions about its efficacy and user satisfaction in an industrial setting. Finally, the research has been conducted with a strong commitment to ethical considerations, ensuring the integrity and reliability of the results.

Chapter 4: Results

4.1 Introduction

The data in this section serves as the foundation for the analysis and conclusions drawn from the research. The main goal of gathering this data was to learn more about how the DSpace Institutional Repository toolkit's efficacy is evaluated, with an emphasis on how this differs depending on the IT ability level of the user (Smith J. , 2020). For this investigation twenty participants finished the survey in its entirety, supplying all the data needed for analysis while five surveys were returned incomplete, and five people did not answer at all. Although the survey included text input fields for respondents to provide additional comments, none of the respondents entered any data in these fields. As a result, there is no presentation of qualitative data from text inputs in this section.

The data analysed and presented is separated into two main categories. The first category includes 'All' respondents irrespective of their IT proficiency level. The second category compares respondents where their IT proficiency levels are 'Expert or Advanced' against respondents where their IT proficiency levels are 'Intermediate & Beginner'. This distinction allows for a nuanced understanding of the responses based on the technical acumen of the respondents. The overall results encompass responses from all 20 respondents who completed the survey, providing a holistic view of the data.

4.2 Background responses - 'All' respondents irrespective of IT proficiency level.

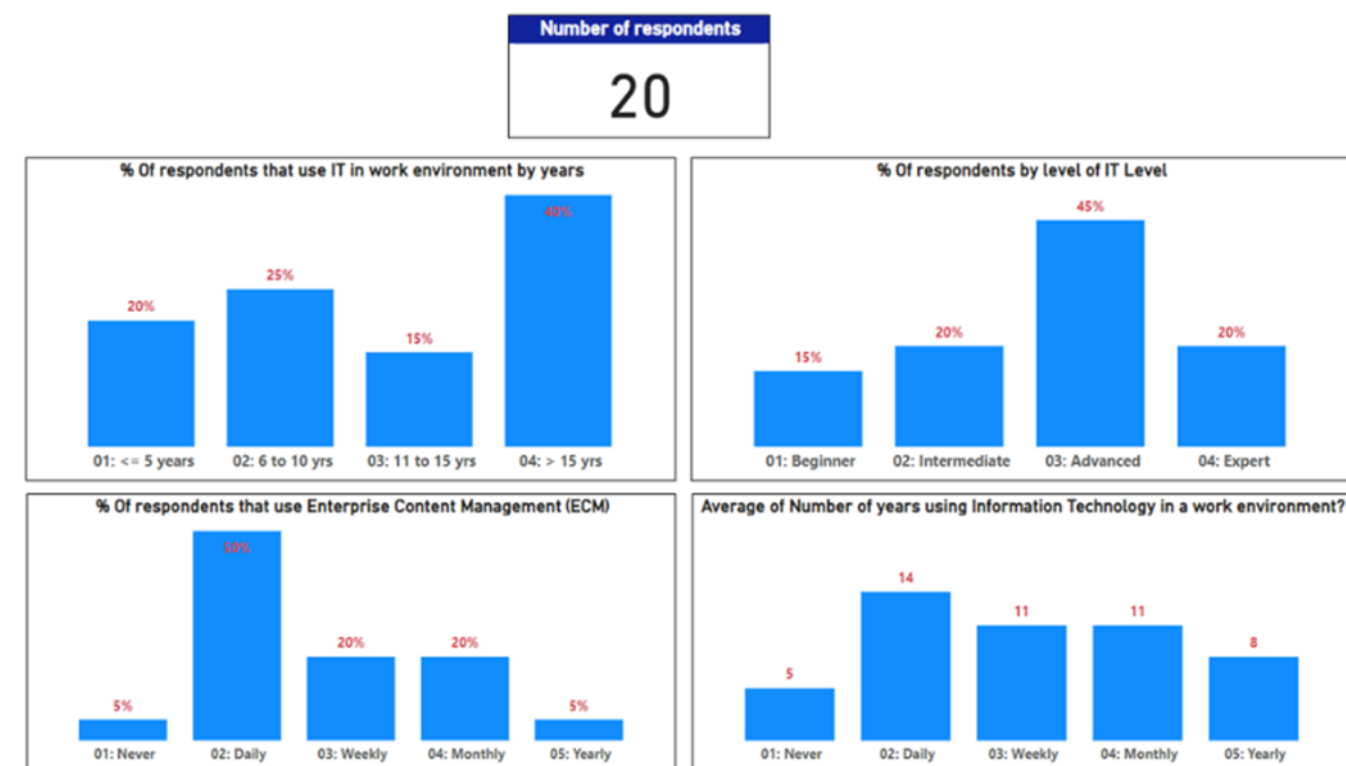


Figure 4-1 - Background responses - 'All' respondents irrespective of IT proficiency level

From Figure 4-1 - Background responses - 'All' respondents irrespective of IT proficiency level:

1. Years of IT Experience:

- The histogram of IT experience years shows a spread across respondents, with a notable number of individuals indicating '15+' years of experience.
- The specific percentages for each category indicate that 40% of the respondents have 15 or more years of experience, followed by 15% with 11 to 15 years of experience. There's also a reasonable spread across other experience levels.

2. Proficiency in IT:

- 45% of respondents rate themselves as 'Advanced' in IT proficiency, 20% as 'Expert', another 20% as 'Intermediate', and 15% as 'Beginner'. This distribution suggests that most respondents feel quite competent in their IT skills, which could influence their perception of DSpace.

3. Frequency of ECM Usage:

- 50% of respondents use ECM technologies daily, indicating a high level of engagement with such systems. Monthly and weekly usage are next, each with 20%, while 'Never' and 'Yearly' both have 5%. This high frequency of use suggests that respondents are familiar with ECM technologies, which could mean they have well-formed opinions on the suitability of DSpace in their work environment.

4. Usefulness of ECM Technologies:

- The analysis of the survey responses revealed a high average rating of 4.74 out of 5, indicating a strong positive consensus among the participants about the usefulness of ECM technologies. This endorsement is underscored by a relatively low standard deviation of 0.44, which suggests that the responses were not only positive but also consistent. This tight clustering of data points around the high mean value points to a general agreement among the respondents, rather than a divided opinion. The high average rating, coupled with the low standard deviation, provides evidence that the effectiveness and benefits of ECM technologies are widely recognised and appreciated among the respondents. This underscores the critical role that ECM technologies play in the surveyed organisation.

Given the high proficiency in IT among respondents, the frequent use of ECM technologies, and the high usefulness rating, these factors could suggest a readiness and a positive disposition toward adopting DSpace in an industrial setting. Respondents who are experienced and proficient with technology, and who frequently use ECM systems, likely recognise the potential benefits of integrating a system like DSpace into their workflow.

The analysis of the responses to the functional questions from the questionnaire was divided into four distinct categories to offer perspectives on the user experience within DSpace. The categories are “Access and Workspace Management”, “Content Management and Search Functionality”, “System and Content Management”, and “Advanced Features and Overall Satisfaction” as detailed in Table 4-1 - Functional Categories.

Category	Process	Mean	Std Deviation
Access and Workspace Management	Access to DSpace	4.55	0.60
	Creating Workspaces and Sub-Workspaces	4.50	0.61
	Editing Workspaces	4.50	0.61
	Populating Workspaces with Content	4.55	0.69
Content Management and Search Functionality	Updating Metadata Fields	4.53	0.61
	Browsing Workspaces	4.55	0.69
	Keyword Search	4.74	0.45
	Submitting and Saving Documents	4.80	0.41
	Uploading Files	4.80	0.41
	Locating Resources	4.60	0.50
	Search Functionality Effectiveness	4.68	0.48
System and Content Management	Central Repository Function	4.60	0.60
	Long-Term Preservation Support	4.50	0.51
	Metadata-Based Content Search	4.68	0.48
	Storage of Scanned Content	4.60	0.50
	Variety of File Formats Supported	4.55	0.69
Advanced Features and Overall Satisfaction	Tracking Document Changes	4.35	0.75
	Multi-Country Support	4.53	0.61
	Customisable Workflows	4.37	0.60
	Automated Alerts	4.21	0.79
	Efficient Approval Workflows	4.37	0.68
	User Access Management	4.50	0.61
	Meeting Business Requirements	3.95	0.78

Table 4-1 - Functional Categories

In the 'Access and Workspace Management' category, the survey data reveals a consensus among respondents about the use of DSpace in terms of access and workspace management. This is reflected in the high mean ratings, approximately 4.5 out of 5, demonstrating a positive sentiment. Although there is some variability, as indicated by the standard deviations ranging from 0.60 to 0.69, it does not impact the overall positive user experience in this aspect.

Moving to the 'Content Management and Search Functionality' category, the results demonstrate a highly favourable perception of these functions among respondents. Critical tasks such as document submission, file uploading, and keyword searching receive high ratings, with means exceeding 4.7, indicative of strong satisfaction. The low standard deviations in this category suggest that this positive experience is consistent and uniform across respondents.

Regarding the 'System and Content Management' category, the survey presents a generally favourable view. Respondents express significant satisfaction with features like the central repository function, long-term preservation, and metadata-based search, as shown by high mean scores ranging between 4.50 and 4.68. The moderate average deviations further emphasise a shared agreement among respondents regarding these features.

However, the survey paints a more complex picture in the 'Advanced Features and Overall Satisfaction' category. Here, a diverse range of satisfaction levels with DSpace's advanced features

and overall user satisfaction is evident. The mean score of 3.95 for meeting business requirements, in particular, points to an area ripe for improvement. Additionally, higher standard deviations in aspects like document change tracking and automated alerts suggest variability in respondent experiences and perceptions within these advanced functionalities.

In summary, the survey findings indicate that DSpace is highly regarded for its core functionalities, especially in content management and ease of use, as evidenced by the high average ratings. Nonetheless, the variability in respondent experiences and perceptions regarding the advanced features, along with the need for enhancement in meeting comprehensive business requirements, highlight areas for potential development. While DSpace excels in fundamental functionalities, this analysis underscores the opportunity for refining its advanced features to better cater to holistic business needs.

The analysis of responses to the non-functional questions from the questionnaire was categorised into six distinct sections to provide insights into the user experience with DSpace. These categories, as outlined in Table 4-2 - Non-Functional Categories, are “Functionality Comparison with ECM Systems”, “User Interface and Navigation”, “Learning and Proficiency”, “Performance and Efficiency”, and “Limitations and Recommendations”.

Category	Process	Mean	Std Deviation
Access and Workspace Management	Access to DSpace	4.55	0.60
	Creating Workspaces and Sub-Workspaces	4.50	0.61
	Editing Workspaces	4.50	0.61
	Populating Workspaces with Content	4.55	0.69
Content Management and Search Functionality	Updating Metadata Fields	4.53	0.61
	Browsing Workspaces	4.55	0.69
	Keyword Search	4.74	0.45
	Submitting and Saving Documents	4.80	0.41
	Uploading Files	4.80	0.41
	Locating Resources	4.60	0.50
	Search Functionality Effectiveness	4.68	0.48
System and Content Management	Central Repository Function	4.60	0.60
	Long-Term Preservation Support	4.50	0.51
	Metadata-Based Content Search	4.68	0.48
	Storage of Scanned Content	4.60	0.50
	Variety of File Formats Supported	4.55	0.69
Advanced Features and Overall Satisfaction	Tracking Document Changes	4.35	0.75
	Multi-Country Support	4.53	0.61
	Customisable Workflows	4.37	0.60
	Automated Alerts	4.21	0.79
	Efficient Approval Workflows	4.37	0.68

Table 4-2 - Non-Functional Categories

In the "Functionality Comparison with ECM Systems" category, a mean of 3.34 indicates a moderate satisfaction level among respondents about DSpace's unique functionalities compared to other ECM systems. This moderate satisfaction, coupled with a relatively high standard deviation of 1.09, points to a substantial variability in opinions, possibly stemming from diverse respondent needs, different

content management requirements, or varying familiarity levels with both DSpace and other ECM systems.

Moving to the "User Interface and Navigation" category, the analysis reveals that respondents generally find the navigation in DSpace user-friendly, as indicated by a consolidated mean of 4.6 and a standard deviation of 0.6. This suggests a consistent and positive experience across respondents. However, when it comes to the user interface's clarity, a consolidated mean of 2.11 coupled with a standard deviation of 1.41 indicates a variability in user experiences, emphasising the need for accommodating various user preferences in interface design.

Regarding "Learning and Proficiency", respondents express a high level of satisfaction, with a consolidated mean score of 4.52, which is above the midpoint of typical scales. This score reflects that respondents find DSpace intuitive, easy to navigate and search, and efficient for common tasks, with a standard deviation of 0.63 pointing to a relatively consistent experience among respondents.

In the 'Limitations and Recommendations' category, the consolidated mean of 3.87 suggests a generally positive view towards the "File Upload Limitations" and "Recommendation to Peers" aspects. The balance between moderate concerns about file upload limitations and a high propensity to recommend DSpace is reflected here. However, the standard deviation of 1.01 indicates noticeable variability in user experiences, possibly influenced by different usage patterns, types of content managed, or individual expectations.

Finally, the "Performance and Efficiency" category, with a consolidated mean of 4.27, indicates a high level of user satisfaction regarding DSpace's performance and efficiency. This score suggests that respondents find DSpace effective in managing large content volumes and efficient in search and retrieval operations, despite some variation in experiences as suggested by a standard deviation of 0.70. This variation might be attributed to differing user expectations, the nature of their work, or the diversity in content and data volumes managed within DSpace.

This comprehensive analysis of the non-functional aspects of DSpace provides valuable insights into areas where the system excels and where there are opportunities for improvement, guiding future enhancements to better meet respondent needs and expectations.

The results imply that DSpace is a capable and approachable solution for digital content management in a learning environment or business. It is a good option for businesses wishing to put in place an efficient digital repository or content management system because of its strong points in terms of usability, content management, search and metadata capabilities, and document handling. Although specific industrial demands and circumstances may require more examination, the excellent ratings in these functional categories suggest that DSpace could potentially meet the requirements of end users and administrators in an industrial environment.

4.3 Summary for "All" respondents

The survey findings from the questionnaire, categorised into functional and non-functional aspects, provide an in-depth understanding of user experiences with DSpace. In the functional categories, which include "Access and Workspace Management", "Content Management and Search Functionality", "System and Content Management", and "Advanced Features and Overall Satisfaction", DSpace demonstrates strong performance, particularly in core functionalities. High consolidated mean ratings, especially in access, workspace management, and content management, indicate a high level of user satisfaction and a perception of DSpace as a user-friendly platform. However, the variability observed in the responses, especially in the advanced features category,

suggests that while DSpace excels in basic functionalities, there are opportunities for further development, particularly in meeting comprehensive business requirements.

The non-functional categories, namely “Functionality Comparison with ECM Systems”, “User Interface and Navigation”, “Learning and Proficiency”, “Performance and Efficiency”, and “Limitations and Recommendations”, further reinforce this narrative. While DSpace largely meets business-related needs as per the respondents' feedback, variability in user experiences and perceptions indicates the need for DSpace to address specific respondent concerns to enhance functionality and applicability. The moderate satisfaction levels in comparison with other ECM systems and the variability in respondent interface experiences highlight areas where DSpace can improve to better align with diverse user needs and preferences.

The high satisfaction levels in "Learning and Proficiency" and "Performance and Efficiency" categories are promising, as they suggest that DSpace is not only intuitive and easy to navigate but also efficient in handling content and performance. Yet, the variability observed in the “Limitations and Recommendations” category, especially concerning file upload limitations, points to the need for DSpace to continuously evolve to cater to the varying demands and expectations of its users.

In summary, DSpace emerges as a competent and user-friendly solution for digital content management, with strong capabilities in usability, content management, search functionality, and document handling. While it shows potential to meet the needs of end-users and administrators in industrial environments, the insights from the survey indicate that there is room for improvement. Addressing the specific industrial demands and enhancing the advanced features are critical steps to ensure that DSpace not only maintains its strengths but also evolves to meet the changing and diverse needs of its user base. This analysis provides valuable guidance for future enhancements, aiming to optimise DSpace to better meet user needs and expectations in varied environments.

4.4 Background responses - IT proficiency level of ‘Expert or Advanced’ vs ‘Intermediate & Beginner’.

The analysis provides insights into the background questions concerning 'Years of IT Experience', 'Proficiency in IT', 'Frequency of ECM Usage', and 'Usefulness of ECM Technologies':

The survey data highlights differences in the frequency of ECM technology usage across various experience groups, with a significant trend towards daily use among both "Expert/Advanced" and "Intermediate/Beginner" users, especially notable in those with more than 15 years of IT experience. This observation underscores the prevalence of consistent ECM technology interaction among experienced IT professionals, regardless of their self-rated proficiency levels. A closer look at the data shows that within the dataset, most respondents (10 out of 20) have over 15 years of IT experience, indicating a significant representation of highly experienced users. On the other hand, the second largest group, comprising individuals with 6 to 10 years of experience (5 out of 20 respondents), and the least represented groups—with 5 years or less and 11 to 15 years of IT experience (3 and 2 respondents respectively); demonstrate the varied engagement levels with ECM technologies. For those with 6 to 10 years of experience, the usage frequency ranges widely from "Never" to "Yearly" among "Intermediate/Beginner" users, suggesting a sporadic engagement with ECM technologies. In contrast, "Expert/Advanced" users within the 6 to 10 years and 11 to 15 years' experience brackets report more consistent usage patterns of "Monthly" and a mix of "Daily" and "Monthly," respectively, highlighting a regular and sustained use of ECM technologies among users with greater expertise and experience. This synthesis of survey findings paints a comprehensive picture of ECM technology use across different IT experience levels, revealing a general trend towards frequent use among more experienced professionals while also indicating varied

engagement levels among those with mid-range experience. Users at both proficiency levels demonstrate regular engagement with ECM technologies, with many incorporating them into their daily professional activities, highlighting the crucial role of ECM in their work routines irrespective of their skill level. The fact that "Intermediate/Beginner" users also frequently use ECM technologies daily points to the system's broad accessibility, suggesting that DSpace's design and support strategies should prioritise this inclusiveness. However, the lack of monthly engagement among "Intermediate/Beginner" users, combined with the existence of a user who never engages with ECM, may signal potential obstacles to sustained use or challenges in leveraging the technology effectively for those at lower proficiency levels. This situation underscores the need for enhanced user training, support, and possibly adjustments in system usability to promote more regular use.

The analysis indicates a common recognition of the utility of ECM technologies among users with varying IT experience and proficiency levels. Nonetheless, the observed differences in the degree of consensus, as reflected by the standard deviations, point to a more varied perspective among those with greater IT experience and higher proficiency. This variation implies that users who are more experienced and skilled in ECM technologies tend to have more complex and detailed perceptions regarding the technology's efficacy and applications.

The analysis of responses to functional and non-functional questions for "Advanced & Expert vs Intermediate & Beginner respondents" from the questionnaire was categorised into four sections for functional questions and six sections for non-functional questions to provide insights into the user experience with DSpace. The categories, as outlined in Table 4-3 - Expert and Advanced vs Intermediate and Beginner Users, for functional questions are "Access and Workspace Management", "Content Management and Search Functionality", "System and Content Management", "Advanced Features and Overall Satisfaction" and "Business Requirements" and for non-functional questions are, "Functionality Comparison with ECM Systems", "User Interface and Navigation", "Learning and Proficiency", "Performance and Efficiency", and "Limitations and Recommendations".

Category	Process	Expert and Advanced		Intermediate and Beginner	
		Mean	Standard Deviation	Mean	Standard Deviation
Access and Workspace Management	Functional	4.54	0.58	4.50	0.71
Content Management and Search Functionality		4.66	0.55	4.73	0.46
System and Content Management		4.65	0.54	4.59	0.53
Advanced Features and Overall Satisfaction		4.67	0.49	4.33	0.69
Business Requirements		3.88	0.94	4.08	1.14
Functionality Comparison with ECM Systems	Non-Functional	3.35	1.15	3.30	0.89
User Interface and Navigation		4.23	0.81	4.06	1.08
Learning and Proficiency		3.67	0.96	3.75	0.69
Performance and Efficiency		4.29	0.62	4.43	0.78
Limitations and Recommendations		3.94	0.78	4.14	1.10

Table 4-3 - Expert and Advanced vs Intermediate and Beginner Users

In the 'Access and Workspace Management' category a comparative analysis highlighted a distinction between 'Expert/Advanced' and 'Intermediate/Beginner' users. 'Expert/Advanced' users demonstrated high satisfaction, indicated by elevated mean scores and lower standard deviations, showcasing their proficiency in utilising DSpace's workspace management features. This suggests that DSpace's interface and tools are well-suited to those with advanced technical skills, allowing them to fully leverage the system's capabilities. In contrast, 'Intermediate/Beginner' users showed lower mean scores, pointing to difficulties in accessing and managing workspaces. Their varied experiences, inferred from higher standard deviations, likely arise from differing familiarity levels

with such systems. This disparity underlines a critical focus area for DSpace's development: enhancing user experience for less experienced users through more intuitive navigation, streamlined workspace management, and targeted educational resources.

The 'Content Management and Search Functionality' category revealed differences between user groups. 'Expert/Advanced' users rated DSpace more favourably, suggesting that their advanced skills enable them to fully utilise its content management and search features. However, 'Intermediate/Beginner' users, facing a steeper learning curve, reported lower mean scores and higher standard deviations, reflecting diverse experiences and possibly difficulties in leveraging complex functionalities. These findings emphasise the importance of designing DSpace with varying user proficiencies in mind. Enhancing guidance, simplifying interfaces, and ensuring intuitive search and content management tools could improve satisfaction and efficiency across all user levels.

In the 'System and Content Management' category, a disparity emerged. 'Expert/Advanced' users showed high proficiency, indicated by higher mean scores and lower standard deviations, suggesting DSpace's effectiveness in catering to users with advanced technical knowledge. Conversely, 'Intermediate/Beginner' users displayed lower mean scores and higher standard deviations, highlighting challenges in system navigation and content management. This variance suggests the need for DSpace to improve its interface and provide more accessible features for less experienced users. Tailoring training and support systems to these users can enhance inclusivity and user-friendliness.

In the "Advanced Features and Overall Satisfaction" category, there's a notable difference in how 'Expert/Advanced' and 'Intermediate/Beginner' users perceive the system. Expert users, showing high mean scores and low standard deviations, are generally satisfied with DSpace's advanced features, thanks to their proficiency and experience with ECM technology. They effectively utilise these sophisticated functionalities, appreciating the system's ability to meet their complex needs. In contrast, beginner users demonstrate varied experiences, indicated by lower mean scores and higher standard deviations, suggesting challenges in accessing and fully benefiting from these advanced features. This highlights a key area for improvement in DSpace: enhancing the intuitiveness and accessibility of advanced features for less experienced users, without reducing their complexity and utility for experts.

In assessing the Business Requirements' category, a contrast in perceptions emerged between 'Expert/Advanced' and 'Intermediate/Beginner' users. 'Expert/Advanced' users, while generally positive about DSpace's business capabilities, showed higher standard deviations in their mean scores. This pattern indicates that, despite recognising DSpace's proficiency, their extensive familiarity, and more complex needs result in a wider range of expectations and opinions. Their heightened scrutiny likely stems from a nuanced understanding of how DSpace functions in varied business contexts. Conversely, 'Intermediate/Beginner' users displayed more uniform perceptions, with lower standard deviations suggesting a basic but consistent satisfaction with DSpace's features. This contrast underscores the importance of evolving DSpace's functionalities to cater to the diverse needs of users across different proficiency levels, ensuring it remains effective.

The analysis of responses to the non-functional categories provides insights into the user experience with DSpace. These categories, as outlined in Table 4-3 - Expert and Advanced vs Intermediate and Beginner Users, are "Functionality Comparison with ECM Systems", "User Interface and Navigation", "Learning and Proficiency", "Performance and Efficiency", and "Limitations and Recommendations".

In the 'Functionality Comparison with ECM Systems' category, clear distinctions were noted between the two user groups. 'Expert/Advanced' users, demonstrating an in-depth understanding, provided moderate mean scores but higher standard deviations. This reflects their appreciation of DSpace's unique features, tempered by an awareness of its limitations when compared to other ECM systems. In contrast, 'Intermediate/Beginner' users, likely less experienced with ECM technologies, showed more uniform opinions, focusing primarily on basic functionalities. These findings emphasise the need for DSpace to innovate continually, addressing functional gaps identified by experienced users while maintaining simplicity for newcomers.

Regarding 'User Interface and Navigation', differences between 'Expert/Advanced' and 'Intermediate/Beginner' users were pronounced. 'Expert/Advanced' users generally viewed the interface favourably, yet the varied standard deviations suggest a range of opinions about its design and usability. This diversity indicates their awareness of potential enhancements to support complex tasks. On the other hand, 'Intermediate/Beginner' users demonstrated consensus on certain challenges within the interface, indicated by more uniform responses. This highlights the need for an adaptable interface that caters to all proficiency levels, balancing ease of use with advanced functionality.

In the 'Learning and Proficiency' category, 'Expert/Advanced' users showed high mean scores and low standard deviations, indicating a consensus on DSpace's user-friendliness. This suggests their previous ECM technology experience translates to a quick adaptation to DSpace. However, 'Intermediate/Beginner' users displayed a wider range of experiences, implying challenges in navigating and mastering DSpace. This disparity underscores the importance of tailored educational resources to support the learning curve of less experienced users.

Analysis of 'Performance and Efficiency' revealed that 'Expert/Advanced' users generally reported high satisfaction, yet the varied standard deviations indicate differing experiences, likely due to diverse usage patterns. In contrast, 'Intermediate/Beginner' users, while positive, showed slightly lower mean scores and more uniform responses, reflecting their limited scope of interaction with DSpace. These insights suggest the need for ongoing optimisation of DSpace to meet the performance expectations of advanced users without compromising accessibility for beginners.

Finally, in the 'Limitations and Recommendations' category, 'Expert/Advanced' users were more critical of DSpace's limitations, such as file upload constraints, evidenced by varied responses. However, their likelihood of recommending DSpace remained high. 'Intermediate/Beginner' users, with more uniform experiences, showed slightly less inclination to recommend DSpace, possibly due to a limited understanding of its full capabilities. This emphasises the need to address specific limitations to enhance user experience across proficiency levels, fostering broader acceptance within the user community.

4.5 Summary for 'Expert/Advanced' versus 'Intermediate/Beginner' users.

The comprehensive analysis of responses from 'Expert/Advanced' versus 'Intermediate/Beginner' users in DSpace, covering both functional and non-functional aspects, provides critical insights into user experience and system efficacy.

In functional categories like 'Access and Workspace Management', 'Content Management and Search Functionality', 'System and Content Management', and 'Advanced Features and Overall Satisfaction', clear differences were observed. 'Expert/Advanced' users showed high satisfaction and proficiency, leveraging DSpace's capabilities effectively, especially in managing advanced features. This group's responses indicate that DSpace aligns well with the needs of technically adept users. In

contrast, 'Intermediate/Beginner' users encountered challenges, particularly in utilising complex functionalities and managing content, as evidenced by lower mean scores and higher standard deviations in their responses. This suggests a need for more intuitive interfaces and tailored educational resources to enhance their experience.

In non-functional categories, including 'Functionality Comparison with ECM Systems', 'User Interface and Navigation', 'Learning and Proficiency', 'Performance and Efficiency', and 'Limitations and Recommendations', the responses again underscored the divergence in user proficiency. While 'Expert/Advanced' users appreciated DSpace's capabilities, even critically evaluating its limitations, 'Intermediate/Beginner' users generally expressed basic satisfaction but highlighted areas needing improvement. Particularly, they faced difficulties in navigation, learning, and leveraging DSpace's full potential due to less experience with ECM technologies.

The analysis underscores DSpace's effectiveness in meeting advanced users' needs but also highlights crucial areas for development to make the system more accessible and user-friendly for beginners. By enhancing the intuitiveness of advanced features, simplifying the user interface, providing comprehensive learning resources, and addressing specific limitations identified by users are key to broadening DSpace's appeal and utility across diverse user groups.

By focusing on these areas, DSpace can ensure it not only maintains its fundamental competencies but also evolves to meet the nuanced requirements of its varied user base, fostering a more inclusive and efficient experience for all users.

4.6 Discussion

The survey, conducted in an industrial environment, yielded responses from 20 participants, providing a diverse range of insights into the efficacy of the DSpace Institutional Repository toolkit. The survey findings from DSpace users, encompassing both 'All respondents' and specific 'Expert/Advanced' vs. 'Intermediate/Beginner' groups, offer a multi-faceted understanding of the system's utility and areas for growth.

Based on the background responses analysis, the representation of respondents with over 15 years of IT experience and their self-rated advanced IT proficiency levels highlights a user base that is well-versed in technology, likely influencing their perceptions of DSpace. This experienced group's daily interaction with ECM technologies underscores a familiarity that shapes their expectations and evaluations of DSpace's capabilities.

For all respondents, DSpace showcases strong performance across functional categories like "Access and Workspace Management", "Content Management and Search Functionality", "System and Content Management", and "Advanced Features and Overall Satisfaction". The overall high satisfaction levels, particularly in core functionalities like access, workspace management, and content management, reflect DSpace as a user-friendly and efficient platform. However, the presence of variability, especially in the advanced features category, highlights potential avenues for further refinement to meet comprehensive business requirements.

Moving to non-functional aspects, such as "Functionality Comparison with ECM Systems", and others, DSpace largely fulfils business-related needs as per general user feedback. Nonetheless, the noted variability in user experiences and perceptions underscores the importance of addressing specific concerns to enhance functionality and applicability. The moderate satisfaction levels in comparison with other ECM systems and the diverse experiences with the user interface signal opportunities for improvement to align better with varied user preferences.

The high satisfaction in "Learning and Proficiency" and "Performance and Efficiency" categories is encouraging, indicating DSpace's intuitive nature and effective performance handling. Yet, the "Limitations and Recommendations" category, particularly concerning file upload limitations, suggests a need for ongoing evolution to match the diverse demands and expectations of users.

Transitioning to the 'Expert/Advanced' versus 'Intermediate/Beginner' respondents, the analysis reveals distinct user experiences. 'Expert/Advanced' users display high satisfaction and adeptness in utilising DSpace, particularly in managing advanced features, signifying the system's alignment with the needs of technically skilled users. Conversely, 'Intermediate/Beginner' users face hurdles in leveraging complex functionalities, indicated by lower mean scores and higher standard deviations, pointing towards a requirement for more intuitive interfaces and customised learning resources.

In non-functional areas, while 'Expert/Advanced' users demonstrate a deeper appreciation of DSpace's capabilities, critically assessing its limitations, 'Intermediate/Beginner' users express basic satisfaction but also identify areas for improvement. Challenges in navigation, learning, and fully utilising DSpace's potential due to lesser ECM technology experience are prominent among these users.

This analysis aims to assess the adaptability and suitability of DSpace in an industrial context, focusing on user experience for both novice and experienced users. The survey results highlight the need for DSpace to cater to diverse user proficiencies by offering an intuitive design for beginners and advanced features for experienced users.

DSpace stands out as a robust, efficient, and user-friendly digital content management solution, effectively meeting the needs of users across varying levels of IT proficiency. Advanced and expert users particularly commend DSpace for its ease of navigation, swift performance, and efficiency in handling common tasks, underscoring its suitability for more technically skilled users. The system's proficiency in speed, performance, and overall user satisfaction is evident, with a high likelihood of being recommended within professional circles.

However, the survey insights also point to critical areas for improvement, essential for DSpace's continuous evolution to cater to its diverse user base. Notably, aspects such as user interface clarity, content handling capacity, and file upload limitations have emerged as key areas needing enhancement. Intermediate and beginner users, while finding DSpace accessible and efficient in workspace management, search functionality, and document handling, have highlighted the need for improvements in multi-country support and other specific features. This indicates potential areas for further development or investigation to enhance the overall user experience.

To maintain and amplify DSpace's appeal, a dual approach is necessary. On one hand, the system should become more approachable for beginners through an intuitive design and comprehensive educational resources. On the other, it should continue to meet the sophisticated demands of advanced users. Addressing the perceived limitations in file upload capacity and enhancing user interface clarity, especially for handling large volumes of content, are crucial steps in this direction.

In essence, these findings underscore the importance of a balanced development strategy for DSpace. By aligning its features and functionalities with the needs and expectations of both beginner and advanced users, DSpace can ensure a more inclusive, efficient, and user-friendly experience, catering effectively to a broad spectrum of users with varying IT proficiencies. This balanced approach is key to reinforcing DSpace's position as a leading solution in digital content management, capable of adapting to and satisfying the evolving needs of its diverse user community.

Chapter 5: Conclusion

5.1 Overview of Research

This research investigates the adaptability and effectiveness of DSpace, a tool primarily used in academia, within a downstream oil and gas industrial environment in the Western Cape. The core research question explores whether DSpace's services can meet the intricate demands of end users and administrators in this specific industrial setting.

5.2 Major Findings and Literature Comparison

The research findings indicate that DSpace, an institutional repository toolset, has a great deal of potential for use in industrial settings. Research suggests that DSpace facilitates data transmission between academic and industrial domains as well as in other settings (Ma, Wang, Zhu, & Tang, 2009). It is appropriate for industrial application because its capabilities in content generation, management, community and collection definition, and the creation of cutting-edge institutional repositories, as demonstrated by the installation (Roa, 2020). The system's minimum investment requirements and low demand for application support knowledge, together with its easy-to-use self-archiving, logical document organisation, and retrieval features (Das & Chatterjee, 2015), further increase its appropriateness for industrial environments.

This study, while providing valuable insights into DSpace's functionality and usability, encounters inherent limitations that must be acknowledged and addressed for a comprehensive understanding. Primarily, the transferability of the findings is constrained due to the testing being conducted in a single, controlled environment. Inferring these results to other settings without considering contextual differences may lead to overgeneralisation. Hence, it is crucial to approach the study's conclusions with an understanding that they are primarily indicative of DSpace's performance within the specific environment tested and may not fully represent its capabilities or shortcomings in varied contexts.

My cautious approach to generalisation stems from the specific environment in which the study was conducted. The controlled setting of the research limits the direct application of the findings to other contexts without further validation. This cautious approach was taken to avoid overgeneralisation and to encourage future studies to explore DSpace's performance in varied environments.

Hence, it is crucial to approach the study's conclusions with an understanding that they are primarily indicative of DSpace's performance within the specific environment tested and may not fully represent its capabilities or shortcomings in varied contexts.

5.3 Answer to Research Question

The findings suggest that while DSpace can meet most business requirements in an industrial setting, its full potential is yet to be completely realised. Continuous adaptation and enhancement are necessary to harness its capabilities fully. Thus, while DSpace demonstrates viability as an efficient digital repository or content management system for industrial applications, specific customisation and further investigation are required to meet unique industrial demands comprehensively.

5.4 Limitations

It is crucial to note that the findings of this study are specific to the selected organisation and should not be universally generalised. While the insights may resonate with similar industries, the primary

focus has been on addressing the distinct challenges and needs of this organisation. The choice of this organisation, driven by its relevance to the research problem and the accessibility it offered, has enabled an in-depth, context-specific exploration of the issue at hand.

While this study provides critical insights and advances our understanding of DSpace's utility, it also highlights the need for further research and development. Exploring DSpace's application in a variety of settings, understanding its interaction with different industrial demands, and continuously adapting it to meet these challenges are essential steps in fully harnessing its capabilities and redefining its role in digital content management beyond academic realms.

5.5 Future Work

There are several areas where future research should focus; some of the areas are discussed below:

- **Enhanced Customisation for Industrial Needs:** Investigate how DSpace can further be customised to meet the specific requirements of various industrial sectors, considering the diverse nature of digital artifacts in these settings.
- **Scalability and Performance in Large-scale Industrial Environments:** Examine the scalability of DSpace in handling large volumes of industrial data and assess its performance under different operational loads.
- **Integration with Industrial Systems and Protocols:** Explore the potential for DSpace to integrate seamlessly with existing industrial information systems and protocols for smoother data interchange and management.
- **User Experience and Accessibility:** Conduct detailed studies on the user experience of DSpace in an industrial context, focusing on ease of use, accessibility, and satisfaction among a diverse range of industrial users with varying levels of IT expertise.
- **Long-term Preservation and Security:** Assess the long-term digital preservation capabilities of DSpace in the context of industrial data, which often requires stringent security measures and reliable long-term storage.
- **Economic and Organisational Impact:** Study the economic implications of implementing DSpace in an industrial setting and its impact on organisational workflows and processes.
- **Comparative Analysis with Other ECM Solutions:** Compare DSpace with other ECM solutions specifically designed for industrial use to identify areas where DSpace excels or needs improvement.

Through these research directions, the applicability and effectiveness of DSpace in industrial environments can be thoroughly evaluated and enhanced, contributing to better digital content management in various industrial sectors.

References

- Alalwan, J., & Weistroffer, H. (2012). Enterprise content management research: a comprehensive review. *Journal of Enterprise Information Management*, 25(5), 441–461.
- Alalwan, J., Thomas, M., & Weistroffer, H. (2014). Decision support capabilities of enterprise content management systems: An empirical investigation. *Decis. Support Syst.* (68), 39-48.
- Allen, M., & Seaman, C. A. (2007). Likert Scales and Data Analyses. *Quality Progress*, 40(7), 64-65.
- Arsanjani, A., & Pour, G. (2001). Best Practices in Business Rule Design and Implementation. *Technology of Object-Oriented Languages, International Conference* (p. 407). Santa Barbara: IEEE Computer Society.
- Austerberry, D. (2007). Chapter 2 - Document and Content Management. In D. Austerberry, *Digital Asset Management (Second Edition)* (pp. 21-34). Focal Press.
- Avis, M. (2003). Do we Need Methodological Theory to Do Qualitative Research? *Qualitative Health Research* (13), 995-1004.
- Babones, S. (2016). Interpretive Quantitative Methods for the Social Sciences. *Sociology* (50), 453-469.
- Baker, W. E., & Sinkula, J. M. (1999). The Synergistic Effect of Market Orientation and Learning Orientation on Organizational Performance. *Journal of the Academy of Marketing Science*, 27(4), 411-427.
- Bass, M., Stuve, D., Tansley, R., Branschofsky, M., Breton, P., Carmichael, P., . . . Ng, J. (2002). *DSpace Internal Reference Specification – Technology & Architecture*. Cambridge MA.
- Boiko, B. (2002). *Content Management Bible*. New York: Hungry Minds.
- Brocke, J., Simons, A., & Cleven, A. (2011). Towards a business process-oriented approach to enterprise content management: the ECM-blueprinting framework. *Information Systems and e-Business Management* (9), 475-496.
- Bryman, A. (2016). *Social Research Methods*. Oxford University Press.
- Candela, L., Castelli, D., Ferro, N., Ioannidis, Y., Koutrika, G., Meghini, C., . . . Schuldt, H. (2007). *The DELOS Digital Library Reference Model - Foundations for Digital Libraries*. Retrieved from DELOS:
http://delosw.isti.cnr.it/files/pdf/ReferenceModel/DELOS_DLReferenceModel_0.98.pdf
- Chiu, D., & Hung, P. (2005). Privacy and Access Control Issues in Financial Enterprise Content Management. *Proceedings of the 38th Annual Hawaii International Conference on System Sciences* (pp. 95c-95c). Big Island, HI, USA: IEEE.
- Cordeiro, M. (2004). From rescue to long-term maintenance: Preservation as a core function in the management of digital assets. *VINE Vol. 34 No. 1*, 6-16.
- Creswell, J. W. (2014). *Research Design: Qualitative, Quantitative and Mixed Methods Approaches*. Sage Publications.
- Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage Publications.

- Cuya, K., & Palaoag, T. (2023). Assessment of a Learning Management System as a Blended Workplace Using the UTAUT2 model. *IEEE 7th International Conference on Information Technology and Digital Applications (ICITDA)* (pp. 1-8). Yogyakarta: IEEE.
- Dale, R. (2001, August). *OCLC.org*. Retrieved from Research Libraries Group Inc (RLG) and Online Computer Library Center (OCLC), Attributes of a Trusted Digital Repository: Meeting the Needs of Research Resources (Draft for Public Comment), August 2001: <https://www.oclc.org/content/dam/research/activities/trustedrep/attributes01.pdf>
- Dammak, H., Dkhil, A., & Gardoni, M. (2019). Graph-Based Tools for ECM Search Result Analysis to Support the Ideation Step. *16th IFIP WG 5.1 International Conference, PLM 2019* (pp. 170-180). Moscow: Springer.
- Das, D., & Chatterjee, P. (2015). Institutional Repository at Central Library IIT Kharagpur: An Overview. *International Journal of Engineering Development and Research* (3), 321-328.
- Dawson, C. (2019). *Practical Research Methods*. Uptown Books.
- Dhouib, S., & Halima, R. (2013). Surveying Collaborative and Content Management Platforms for Enterprise. *Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises*, 299-304.
- Dobrojevic, M. (2018). Acceleration of Web Content and Page Structure Management Using Shortcode. *International Scientific Conference on Information Technology and Data Related Research* (pp. 61-67). Belgrade: Sinteza.
- Dowling, R., & Mantai, L. (2017). Placing researcher identifications: labs, offices and homes in the PhD. *Area, Vol 40, Issue 2*, 200-207.
- Foster, N., & Gibbons, S. (2005). Understanding Faculty to Improve Content Recruitment for Institutional Repositories. *D-Lib Magazine Vol. 11 No. 1*, 1-2.
- Ganti, A. (2022, March 06). *Central Limit Theorem (CLT)*. Retrieved from Investopedia: https://www.investopedia.com/terms/c/central_limit_theorem.asp
- Goh, D., Chua, A., Khoo, D., Khoo, E., Mak, E., & Ng, M. (2006). A checklist for evaluating Open Source Digital Library software. *Online Information Review* 30(4), 360-379.
- Gupta, V., Govindarajan, S., & Johnson, T. (2001). Overview of content management approaches and strategies. *Electronic Markets* (11), 281-287.
- Hackos, J. T. (2002). *Content Management for Dynamic Web Delivery*. New York: John Wiley & Sons Inc.
- Hu, X., Xue, J., & Liu, H. (2011). Research of architecture pattern based on .NET distributed system. *International Conference on Mechatronic Science, Electric Engineering and Computer (MEC)*, (pp. 1270-1273). Jilin: IEEE.
- Iverson, J., & Burkart, P. (2007). Managing electronic documents and work flows: Enterprise content management at work in nonprofit organizations. *Nonprofit Management and Leadership* (17), 403-419.
- Jain, P. (2011). New trends and future applications/directions of institutional repositories in academic institutions. *Library Review*, 60 (2), 125-141.

- Jettka, D., & Stein, D. (2014, September 28). *The HZSK Repository: Implementation, Features, and Use Cases of a Repository for Spoken Language Corpora*. Retrieved from D-Lib Magazine: <https://www.dlib.org/dlib/september14/jettka/09jettka.html>
- Johnson, B., & Turner, L. (2003). *Data Collection Strategies in Mixed Methods Research*. Thousand Oaks: Sage Publications.
- Johnson, T., & Turner, C. (2003). Survey Response Styles and Data Quality. *Journal of Survey Statistics and Methodology*, 2(1), 18-36.
- Jones, P. (2017). Challenges in Online Survey Research. *Internet Research Methods* 4(1), 45-60.
- Kaczmarek, J., & Naun, C. (2005). A statewide metasearch service using OAI-PMH and Z39.50. *Library Hi Tech* (23), 576-586.
- Khan, S., Usman, A., Irfan, R., & Hayat, A. (2011). DSpace@SEECs: SEECs institutional repository system. *7th International Conference on Emerging Technologies* (pp. 1-6). Islamabad, Pakistan: IEEE XPlore.
- Koehler, W. (2001). Web Page Change and Persistence -A Four-Year Longitudinal Study. *Journal of the American Society for Information Science and Technology* 53, 2, 162-180.
- Lewis, J. (2014). Usability: Lessons Learned and Yet to Be Learned. *International Journal of Human-Computer Interaction*, 30(9), 663-684.
- Lougee, W. (2009). The diffuse library revisited: aligning the library as strategic asset. *Library Hi Tech* 27(4), 610-623.
- Lumpa, M., & Suleman, H. (2019). Investigating the feasibility of digital repositories in private clouds. *Proceedings of 21st International Conference on Asia-Pacific Digital Libraries (ICADL)* (pp. 6-7). Kuala Lumpur: Springer.
- Lumpa, M., Munyaradzi, N., & Suleman, H. (2011). Interconnecting DSpace and LOCKSS. *International Conference on Theory and Practice of Digital Libraries* (pp. 493-496). Berlin: Springer.
- Lynch, C. (2003). Institutional repositories: Essential infrastructure for scholarship in the digital age. *Portal: Libraries and the Academy*, 3(2), 327-336.
- Ma, J., Wang, Y., Zhu, Z., & Tang, R. (2009). An attempt of data exchange between the institutional repository and the information environment for the management of scientific research - ARP. *Library Collections, Acquisitions, and Technical Services* (33), 1-7.
- Masor, M., & Kassim, N. (2020). Measuring Success Factors and Performance of Institutional Repositories. *Journal of Administrative and Business Studies* 6(3), 78-83.
- Mendes, M., & Bax, M. (2018). BPM and ECM: Similarities, differences, conceptual, and technological limits. *Transinformacao* (30), 95-105.
- Munkvold, B., Päiväranta, T., Hodne, A., & Stangeland, E. (2006). Contemporary issues of enterprise content management. *Scandinavian Journal of Information Systems*, 18(2), 4.
- Naik, U., & Shivalingaiah, D. (2006). Digital Library Open Source Software : A Comparative Study. *4th International Convention CALIBER-2006* (pp. 32-33). Gulbarga: INFLIBNET Centre.

- Nicholas, D., Rowlands, I., Watkinson, A., Brown, D., & Jamali, H. (2012). Digital repositories ten years on: what do scientific researchers think of them and how do they use them? *Learned Publishing* (25), 195-206.
- Nicholas, D., Rowlands, I., Watkinson, A., Brown, D., & Jamali, H. R. (2012). Digital repositories ten years on: what do scientific researchers think of them and how do they use them? *Learned Publishing Vol (25) Issue (3)*, 195-206.
- Nielsen, J. (1993). *Usability Engineering*. Boston, MA: Academic Press.
- Norman, D. A. (2013). *The Design of Everyday Things: Revised and Expanded Edition*. Basic Books.
- Oak, M., & Patil, S. K. (2015). Planning and implementation of institutional repository: A case study of IMCC. *4th International Symposium on Emerging Trends and Technologies in Libraries and Information Services* (pp. 253-256). Noida, India: IEEE Xplore.
- O'Callaghan, R., & Smits, M. (2005). A Strategy Development Process for Enterprise Content Management. *European Conference on Information Systems (ECIS)*.
- O'Callaghan, R., & Smits, M. (2005). A Strategy Development Process for Enterprise Content Management. *13th European Conference on Information Systems* (pp. 1271-1282). Regensburg: ResearchGate.
- OpenDOAR. (2023, 11 24). *OpenDOAR*. Retrieved from Directory of Open Access Repositories: https://v2.sherpa.ac.uk/view/repository_visualisations/1.html
- Opentext. (2023, November 12). *Four ways Content Services drive a more Intelligent and Connected Enterprise*. Retrieved from Opentext: <https://www.opentext.com/>
- Palaniswamy, R., & Frank, T. (2000). Enhancing Manufacturing Performance with ERP Systems. *Information Systems Management* (17), 43-55.
- Papadopoulou, M. (2011). The Issue of 'Orphan' Works in Digital Libraries. *E-Publishing and Digital Libraries: Legal and Organizational Issues.*, 198-231.
- Prieto, A. (2009). From conceptual to perceptual reality: trust in digital repositories. *Library Review* (58), 593-606.
- Reznik-Zellen, R., & Adamick, J. (2012). Supporting virtual communities through disciplinary repository development. *Library Hi Tech*, 30(2), 275-290.
- Roa, M. (2020). Use of Institutional Repository for Information Dissemination and Knowledge Management. *Digital Libraries and Institutional Repositories: Breakthroughs in Research and Practice*, 409-422.
- Robson, C., & McCartan, K. (2016). *Real World Research*. Wiley.
- Rubin, J., & Chisnell, D. (2008). *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*. Wiley.
- Sabharwal, A. (2021). Functional Frameworks for Socialized Digital Curation: Curatorial Interventions and Curation Spaces in Archives and Libraries. *Library Trends* (69), 672-695.
- Saunders, M., Lewis, P., & Thornhill, A. (2009). *Research Methods for Business Students* (6th Edition ed.). Retrieved from www.pearson.com/uk

- Smith, H. A., & McKeen, J. D. (2003). Developments in practice VIII: Enterprise content management. *The Communications of the Association for Information Systems*, 11(1), 41.
- Smith, J. (2013). *Quantitative versus qualitative research: The key differences*. University Press.
- Smith, J. (2015). Detecting Dishonest Responses in Online Surveys. *Survey Methodology Insights*, 7(2), 24-31.
- Smith, J. (2018). Online Surveys in Educational Research. *Journal of Educational Research*, 11(2), 213-227.
- Smith, J. (2020). Evaluating Digital Repository Software in Industrial Environments. *Journal of Digital Management*, 15(3), 45-59.
- Soergel, D. (2009). Digital Libraries and Knowledge Organization. *Semantic Digital Libraries*, 9-39.
- Steehouder, M. (2002). Beyond technical documentation: users helping each other. *Proceedings. IEEE International Professional Communication Conference* (pp. 489-499). Portland, OR: IEEE.
- Sue, V. M., & Ritter, L. A. (2012). *Conducting online surveys*. Sage Publications, Inc.
- Sun, J., & Yuan, B.-Z. (2012). Development and Characteristic of Digital Library as a Library Branch. *IERI Procedia* 2, 12–17.
- Tankard, C. (2012). Big Data Security. *Network Security Volume 2012, Issue 7*, 5-8.
- Tansley, M., Bass, M., Stuve, D., Branschovsky, M., Chudnov, D., McClellan, G., & Smith, M. (2003). The DSpace Institutional Digital Repository System: Current Functionality. *Joint Conference on Digital Libraries (JCDL'03)*. Houston Texas: Computer Society.
- Taylor, H. (2016). The Impact of Non-Functional Questions on Online Survey Results. *Online Research Methods*, 8(3), 155-167.
- Trochim, W., & Donnelly, J. P. (2006). *The Research Methods Knowledge Base*. Atomic Dog Publishing.
- Tullis, T., & Albert, B. (2013). *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics*. Elsevier.
- Tyrväinen, P., Päivärinta, T., Salminen, A., & Iivari, J. (2006). Characterizing the evolving research on enterprise content management. *European Journal of Information Systems* (15), 627-634.
- Venkatesh, V., & Bala, H. (2008). Technology Acceptance Model 3 and a Research Agenda on Interventions. *Decision Sciences*, 39(2), 273-315.
- Vishwanthan, K. (2001). Digital Libraries: development and principles. *Library Review* 50 (1), 10-16.
- Williams, M., & Brown, T. (2014). Using Non-functional Questions to Improve Survey Engagement. *International Journal of Market Research*, 56(4), 461-483.
- Wilson, P. (2018). Evaluating User Experience in Information Systems. *Journal of Information Technology*, 33(4), 324-338.
- Witten, I., & Bainbridge, D. (2002). *How to Build a Digital Library*. San Francisco, CA: Morgan Kaufmann.

- Wolski, M., Simons, N., & Richardson, J. (2013). ECMs and Institutional Repositories: The Case for a Unified Enterprise Approach to Content Management. *THETA: The Higher Education Technology Agenda*.
- Wu, Q., Liu, M., Li, X., & Cai, W. (2011). Environmental Datasets-oriented Spatial On-Line Analytical Processing System. *Journal of Geo-information Science* (15), 90.
- Xiang, L., Qixu, W., Xiao, L., Xingshu, C., Ning, Z., & Daijiang, C. (2019). *Enhancing Cloud-Based IoT Security Through Trustworthy Cloud Service: An Integration of Security and Reputation Approach* (Vol. 7). IEEE Open Access Journals and Conferences. Retrieved from https://uct.primo.exlibrisgroup.com/discovery/fulldisplay?docid=cdi_crossref_primary_10_1109_ACCESS_2018_2890432&context=PC&vid=27UCT_INST:27UCT&lang=en&search_scope=MyInst_and_CI&adaptor=Primo%20Central&tab=Everything&query=any,contains,IoT%20and%20Cloud
- Yap, K., & Choy, S. (2018). Learning and praxis for workplace safety. *Journal of Workplace Learning, Vol. 30, No. 4*, 230-244.
- Yin, R. K. (2003). *Case study research: Design and methods* (3rd ed., Vol. 5). Thousand Oaks, CA: Sage.

Appendix A: Survey Questionnaire (LimeSurvey)



Load ui

Purpose of research

Commercial companies use a range of different ECM software solutions for document management; research and academia also have document management requirements and these academic institutions use different solutions for academia. This research is investigating if these two ideas can come together by investigating the applicability of using the DSpace software solution from academia in an industrial or commercial setting. This will be done from an end user perspective, in an industrial environment, to determine the suitability of the services and functionality provided by the DSpace software toolkit. This will be done by assessing the software toolkit's features, with common functionality and architectural models to understand the impact on the user environment.

Before proceeding, it is important to note that DSpace is primarily intended to be used as an institutional repository for capturing, storing, indexing, preserving, and redistributing a community's intellectual output. However, when considering its application in an industrial setting or environment, you may be assessing its suitability as a tool for managing various types of digital artifacts relevant to industrial operations.

This research seeks to answer the following question:

Are the services provided by the DSPACE Institutional Repository toolkit suitable to meet the requirements of end users and administrators in an industrial environment?

There are 10 questions in this survey.

Partial or incomplete responses may be given either by partially completing the survey and clicking the 'Next' button or by clicking the 'Back' button. A partially completed survey will be considered incomplete and not form part of the final analysis.

Feedback: Feedback of the results of the survey will be available on this link once the survey has closed and analysis has been completed.

Risk: There are no risks to participation in this study.

Confidentiality: Participation is anonymous as you will not be required to provide any personal information to identify you and all data collected from the survey will be anonymised. Pseudonyms will be used to remove your data if you withdraw. No company information or documents to be shared as part of this survey.

Disclaimer: Participation is voluntary, you may withdraw from the study at any stage without providing a reason and your data will be removed

What you agreeing to the survey means: By selecting "I agree to participate in this research" and submitting the consent form you are agreeing to participate in this study, and you have read and understood the benefits, procedure, feedback, risk, confidentiality, and disclaimer. You may withdraw at any stage.

Questions: Should you have any questions regarding this research, please feel free to contact Gavin Asary on 0837072885 or email: asrgav001@myuct.ac.za

☐ I have read the Online Informed Voluntary Consent to Participate in Research Study and agree to participate in this research

Next

Background Questions

Number of years using Information Technology in a work environment?

Choose one of the following answers

Please choose... ▾

How would you rate your overall proficiency in using Information Technology in a work environment.

Choose one of the following answers

Please choose... ▾

How often do you use Enterprise Content Management (ECM) technologies?

Choose one of the following answers

Please choose... ▾

I find ECM technologies useful as a tool for my work (1=Strongly Disagree and 5=Strongly Agree)

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☒ No answer

Previous

Next

Functional Questions

Please answer the questions below with 1=Strongly Disagree and 5=Strongly Agree

	1	2	3	4	5	No answer
I can easily access DSpace	☐	☐	☐	☐	☐	☒
I can easily create Workspaces and Sub-Workspaces (Communities and Collections)	☐	☐	☐	☐	☐	☒
I can easily edit a workspace or sub-workspace (communities and collections)	☐	☐	☐	☐	☐	☒
I can easily populate a workspace or sub-workspace with a document, image, or other digital artifact.	☐	☐	☐	☐	☐	☒
I can easily update metadata fields.	☐	☐	☐	☐	☐	☒
I can easily browse through workspaces and sub-workspaces (communities and collections).	☐	☐	☐	☐	☐	☒
I can easily search using keywords.	☐	☐	☐	☐	☐	☒
I can easily submit and save documents.	☐	☐	☐	☐	☐	☒
Uploading files to DSpace is straightforward and user-friendly.	☐	☐	☐	☐	☐	☒
I can quickly locate the resources I need in DSpace	☐	☐	☐	☐	☐	☒
The search functionality in DSpace helps me find specific documents effectively.	☐	☐	☐	☐	☐	☒
DSpace serves as a central repository to store documentation.	☐	☐	☐	☐	☐	☒
DSpace supports my needs in terms of long-term preservation of digital content.	☐	☐	☐	☐	☐	☒
DSpace allows easy searches of the content via metadata?	☐	☐	☐	☐	☐	☒
DSpace allows me to setup a centralised work area to store standardised reusable templates.	☐	☐	☐	☐	☐	☒
I can easily store digital content that has been scanned or captured.	☐	☐	☐	☐	☐	☒
I am satisfied with the variety of file formats supported by DSpace.	☐	☐	☐	☐	☐	☒
I can easily track changes and versions of documents within DSpace.	☐	☐	☐	☐	☐	☒
DSpace has multi-country support is designed to support multiple languages, character sets, and time zones, which enables it to be used across different countries and regions.	☐	☐	☐	☐	☐	☒

Functional Questions

Please answer the questions below with 1=Strongly Disagree and 5=Strongly Agree

	1	2	3	4	5	No answer
Dates can easily be managed in DSpace.	☐	☐	☐	☐	☐	☒
Customisable workflows can easily be setup in DSpace.	☐	☐	☐	☐	☐	☒
Automated alerts and notifications can easily be setup in DSpace.	☐	☐	☐	☐	☐	☒
Workflows for approving items in DSpace are efficient.	☐	☐	☐	☐	☐	☒
Managing and controlling user access in DSpace is easy.	☐	☐	☐	☐	☐	☒

How does DSpace meet your requirements (1=Strongly Disagree and 5=Strongly Agree)

	1	2	3	4	5	No answer
DSpace meets all my business requirements	☐	☐	☐	☐	☐	☒
DSpace meets some of my business requirements	☐	☐	☐	☐	☐	☒
DSpace provides functionality that ECM systems do not provide	☐	☐	☐	☐	☐	☒
ECM system provide functionality that DSpace does not provide	☐	☐	☐	☐	☐	☒
DSpace and ECM systems have a high degree of overlap in functionality	☐	☐	☐	☐	☐	☒

Are there any features that you believe are missing and would improve DSpace?

Previous

Next

Non-Functional Questions

Please answer the questions below with 1=Strongly Disagree and 5=Strongly Agree

	1	2	3	4	5	No answer
The navigation menu within DSpace is easy to understand.	☐	☐	☐	☐	☐	☒
The user interface of DSpace is cluttered or confusing.	☐	☐	☐	☐	☐	☒
I find it easy to learn and become proficient with DSpace, even as a new user.	☐	☐	☐	☐	☐	☒
I can easily navigate, search, and interact with DSpace	☐	☐	☐	☐	☐	☒
I can efficiently perform common tasks such as uploading, searching, and retrieving content?	☐	☐	☐	☐	☐	☒
The system easily handles large volumes of content?	☐	☐	☐	☐	☐	☒
The overall speed and performance of DSpace meet my expectations.	☐	☐	☐	☐	☐	☒
The response times for searches or content retrieval operations is acceptable	☐	☐	☐	☐	☐	☒
There are limitations on the number of files I can upload.	☐	☐	☐	☐	☐	☒
I would recommend DSpace to colleagues or peers.	☐	☐	☐	☐	☐	☒

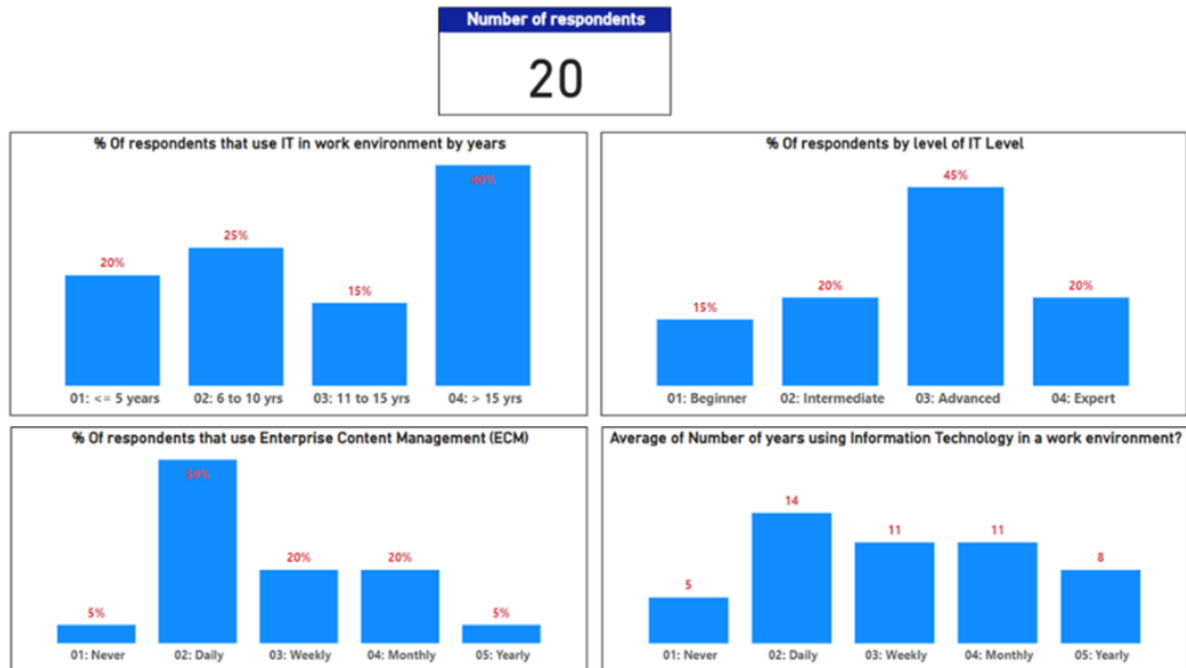
General comments and feedback

Previous

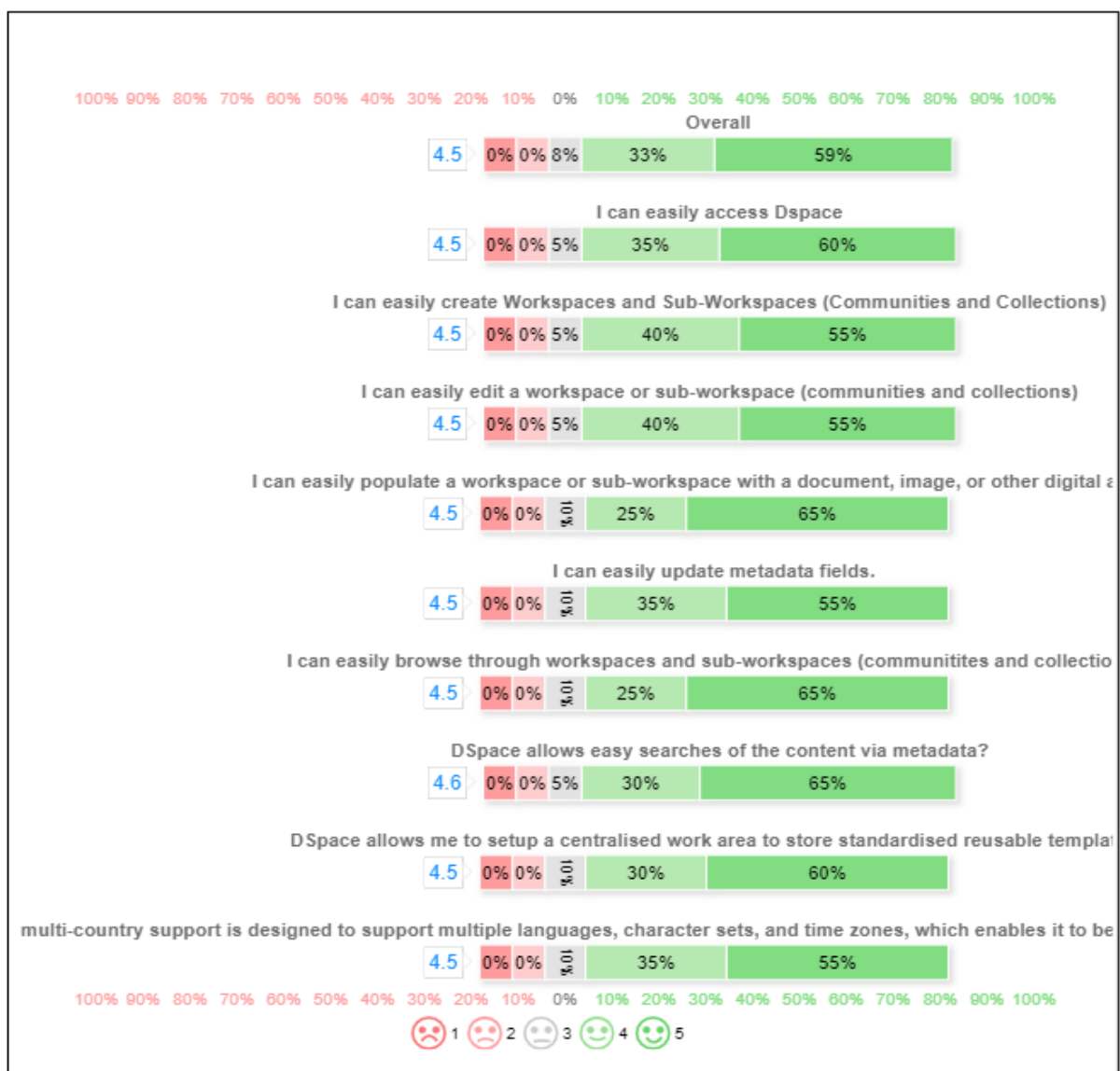
Submit

Appendix B: Survey Questionnaire Results

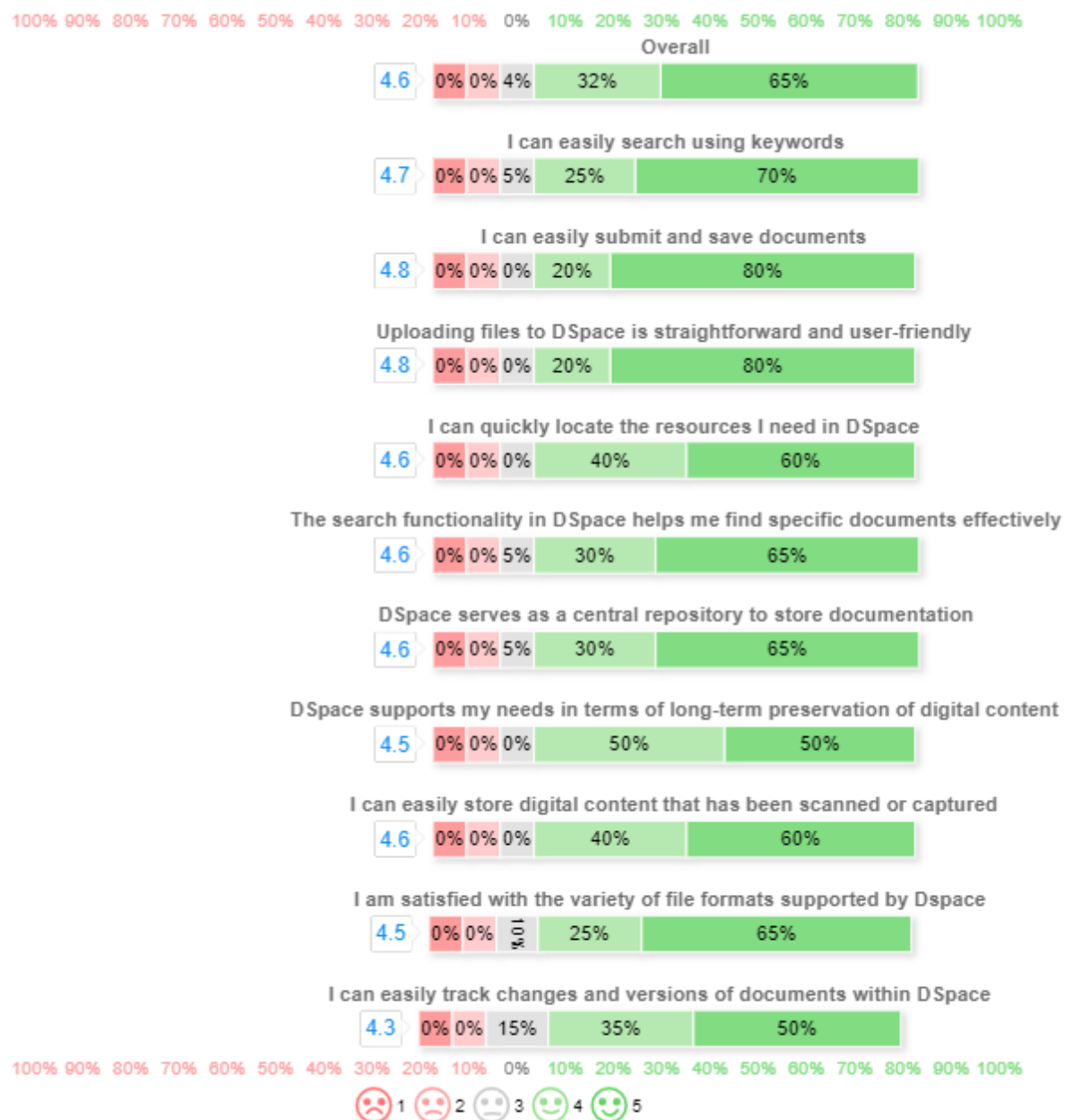
Background responses - 'All' respondents irrespective of IT proficiency level.



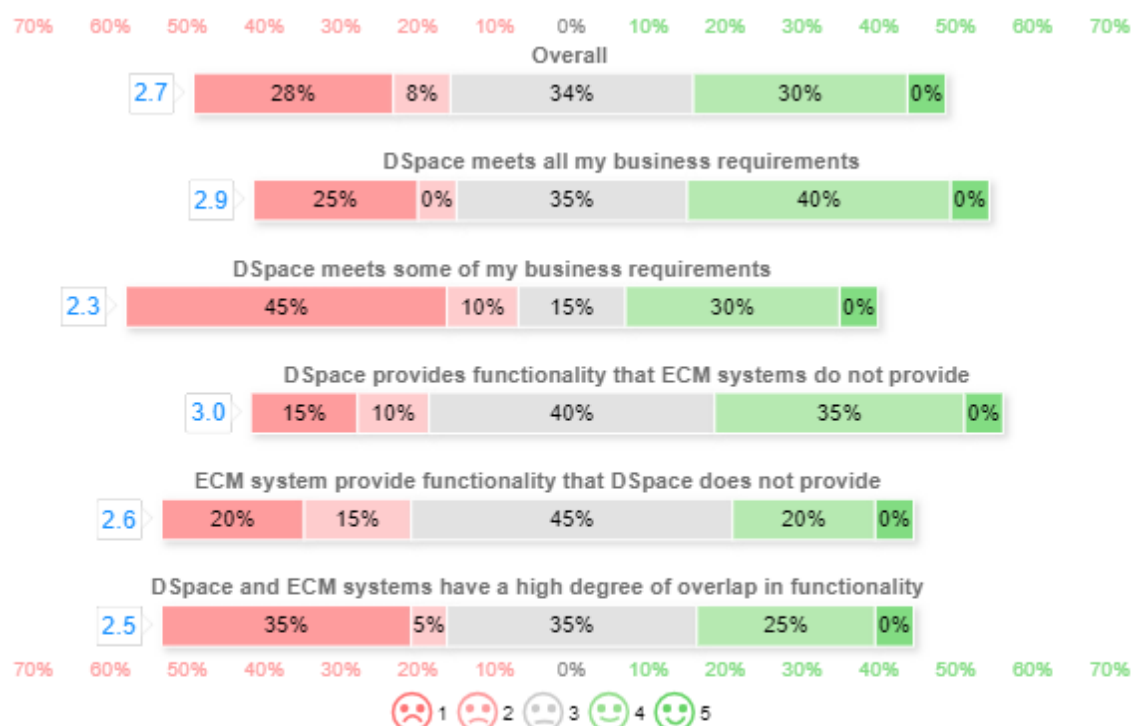
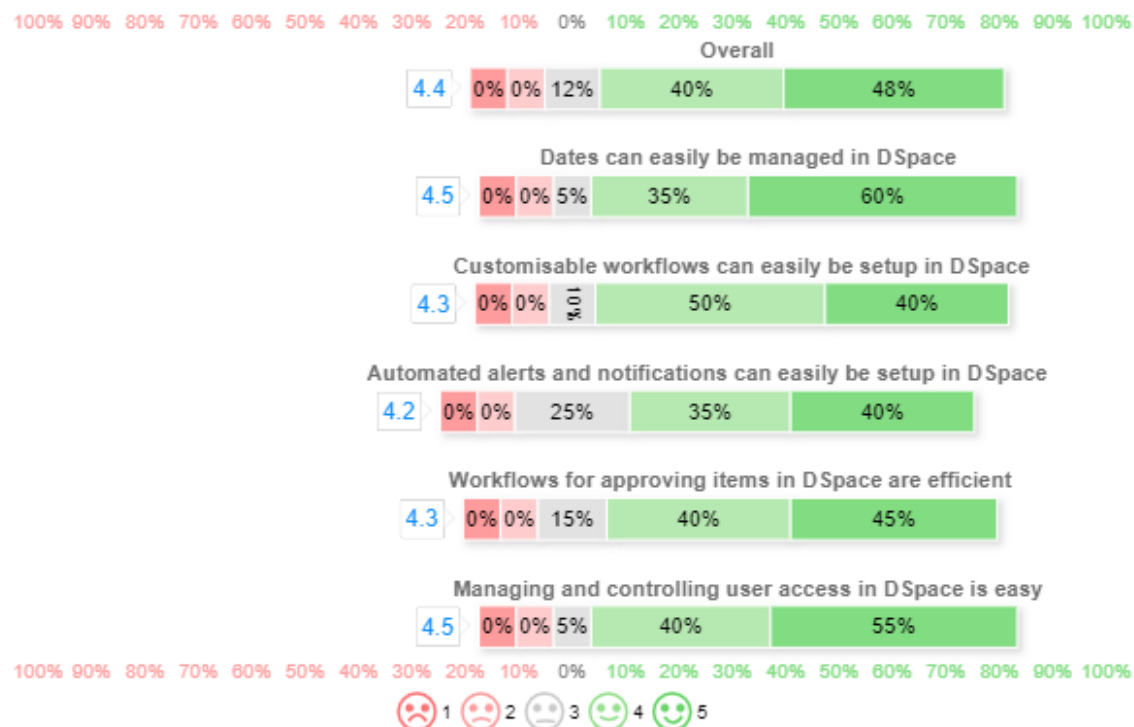
Functional question responses – “All” respondents.



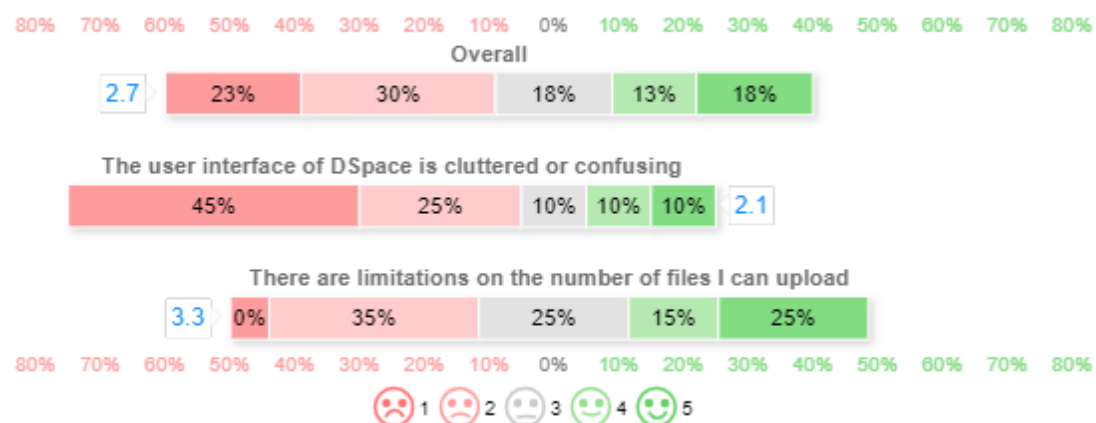
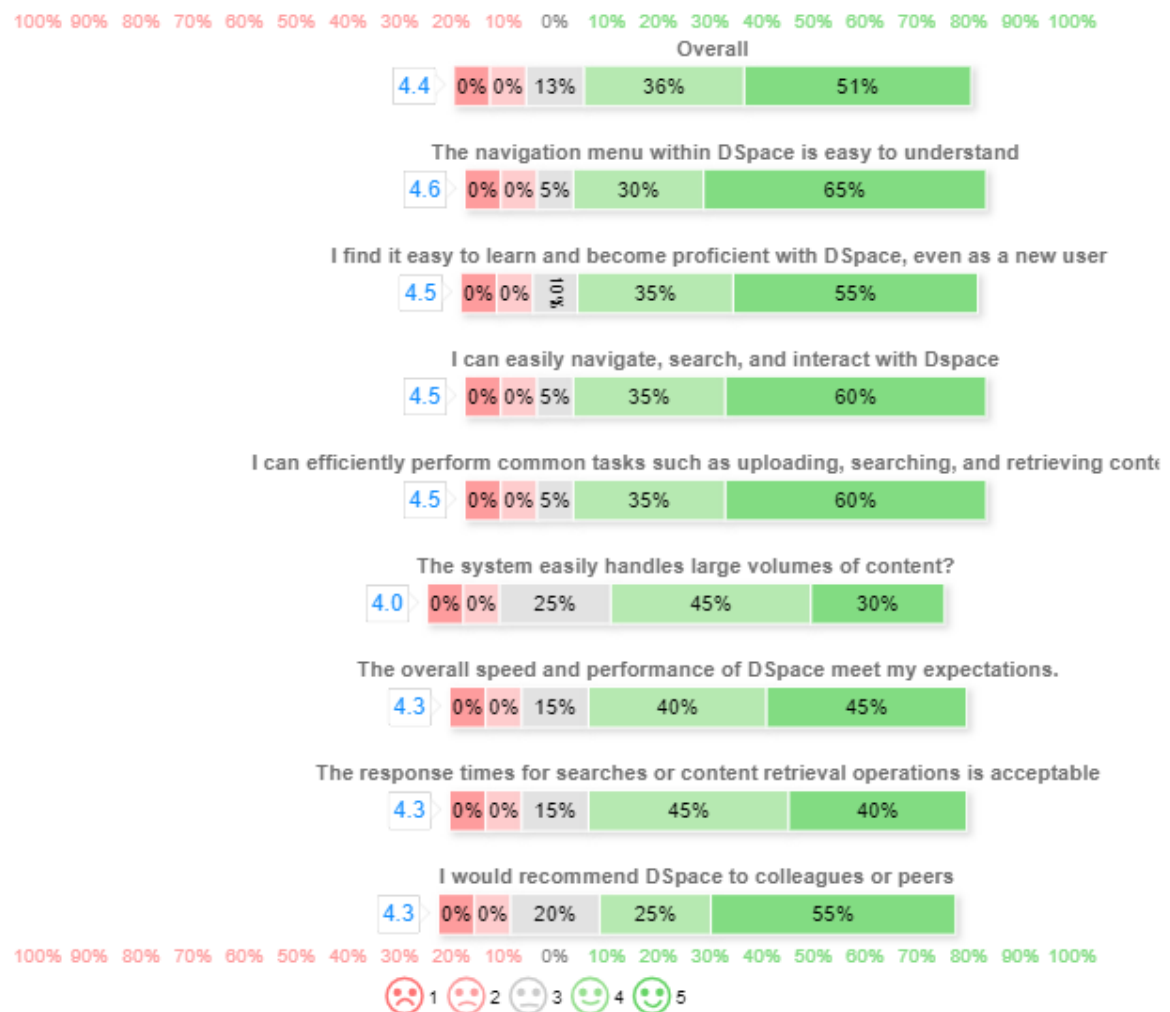
Functional question responses – “All” respondents.



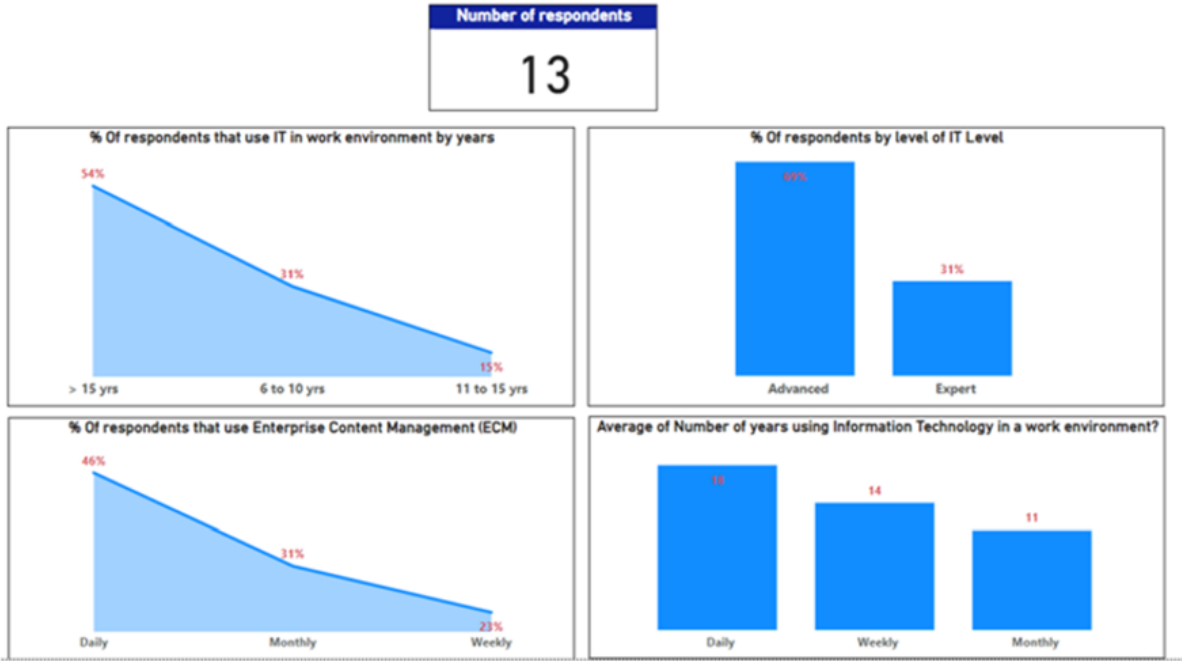
Functional question responses – “All” respondents.



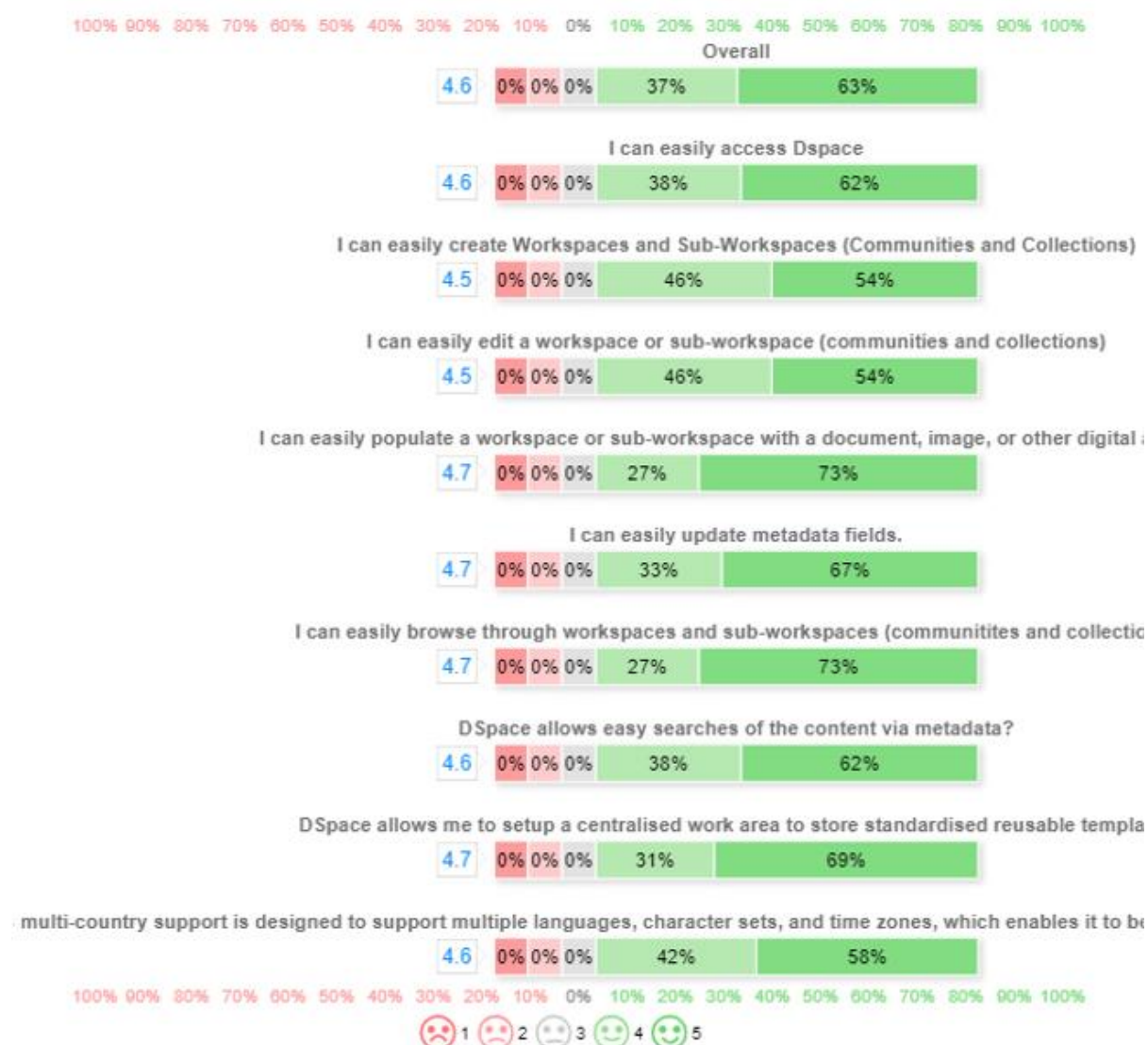
Non-Functional question responses – “All” respondents.



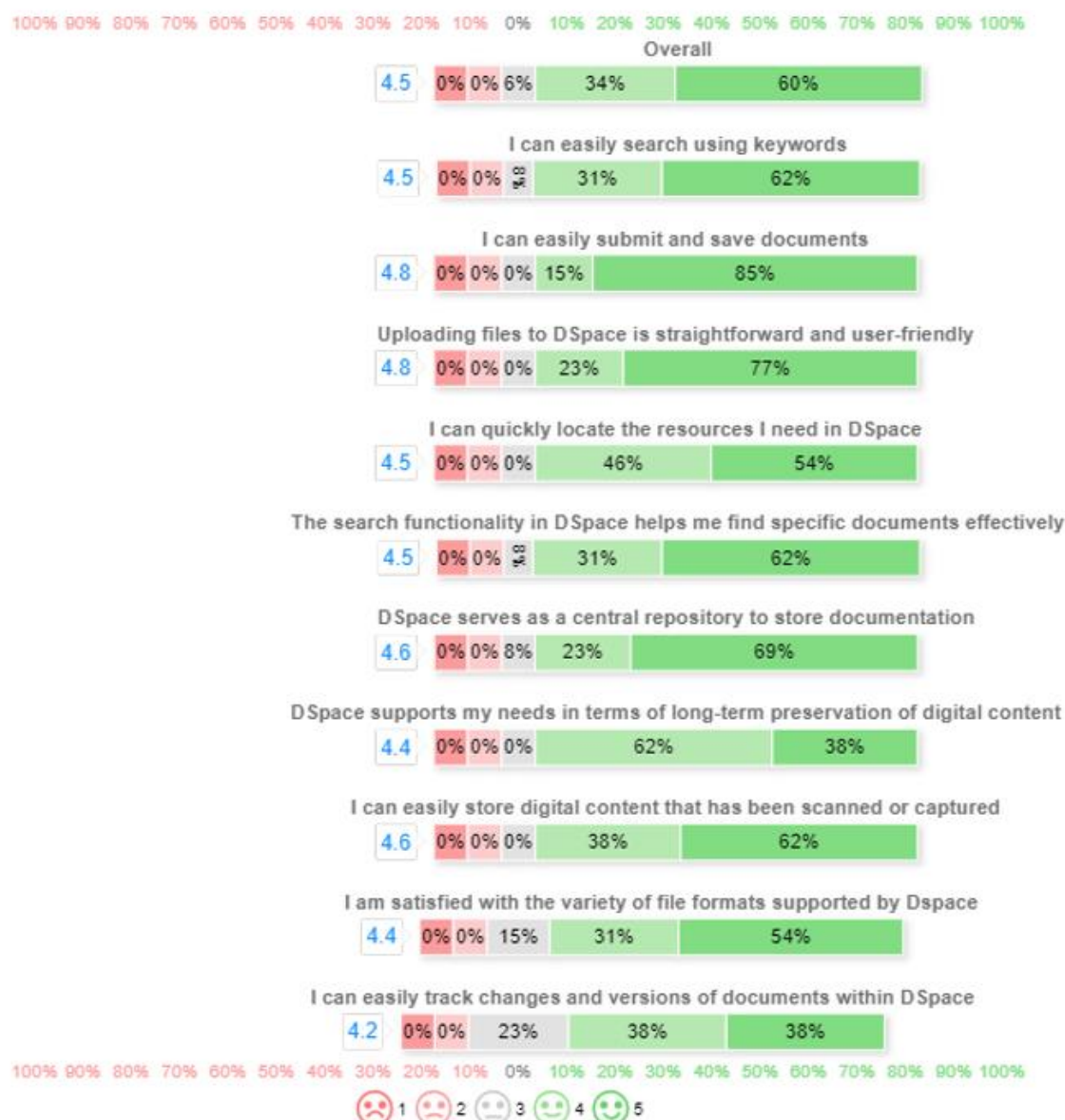
Background Questions – Advanced and Expert



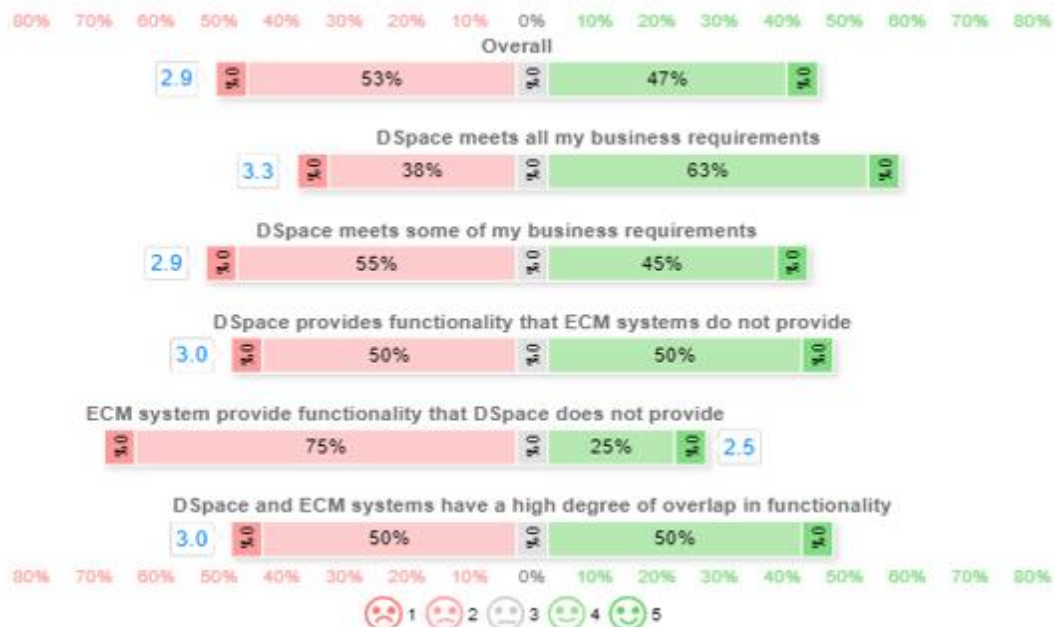
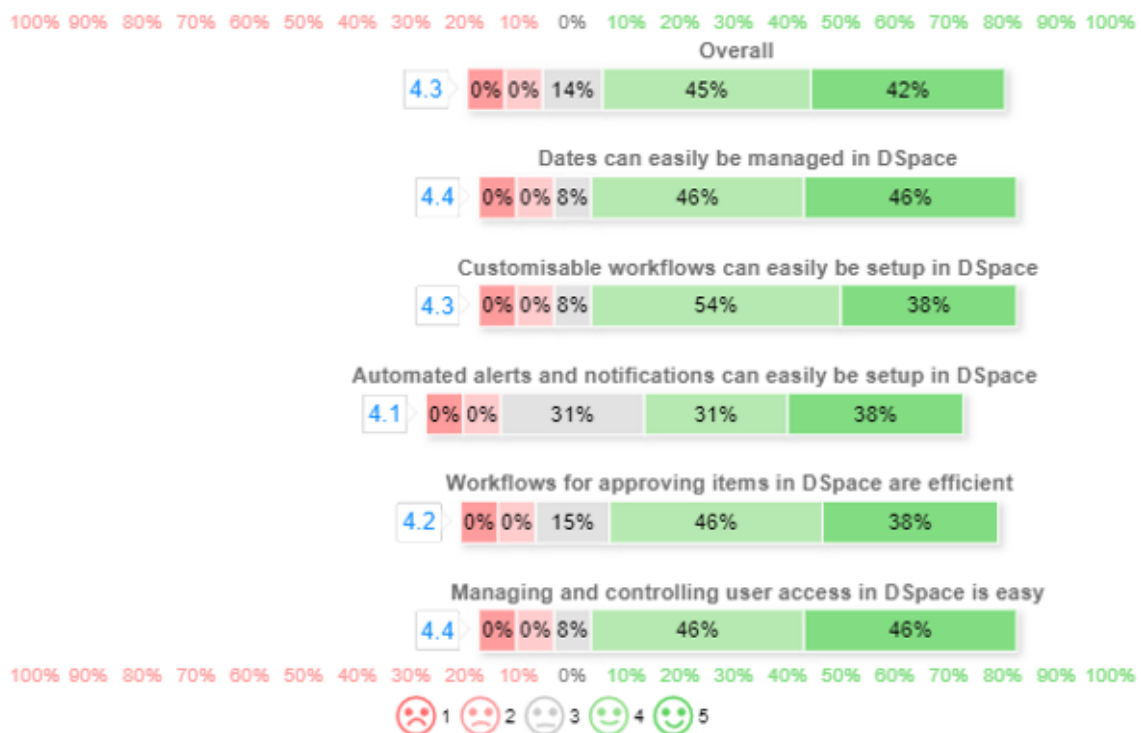
Functional question responses - Advances and Expert respondents



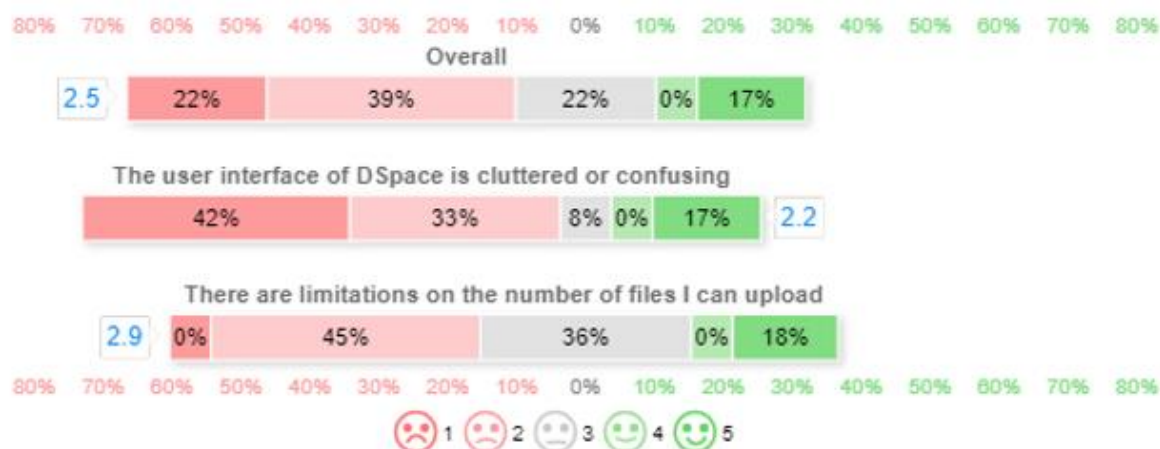
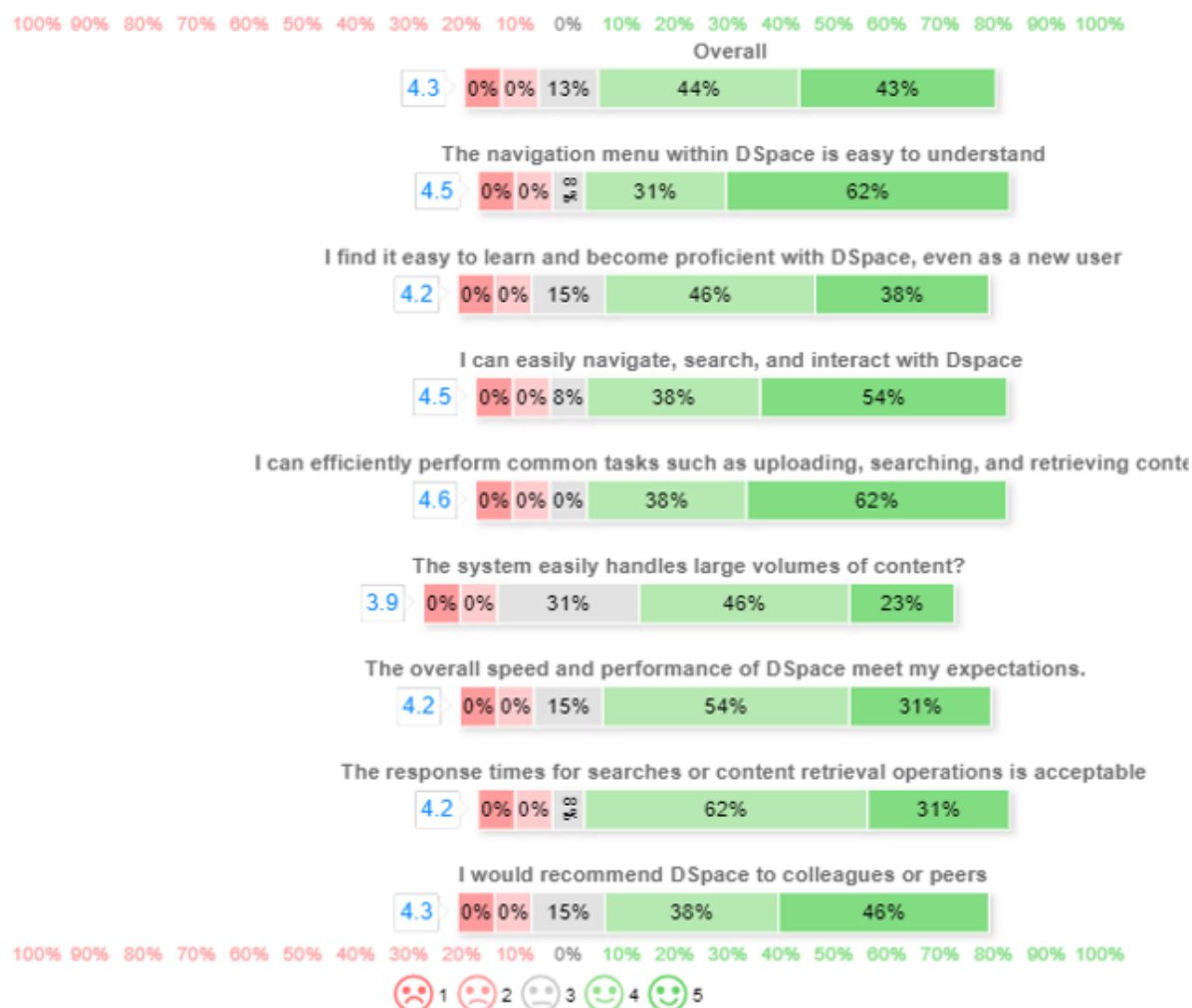
Functional question responses - Advances and Expert respondents



Functional question responses - Advances and Expert respondents



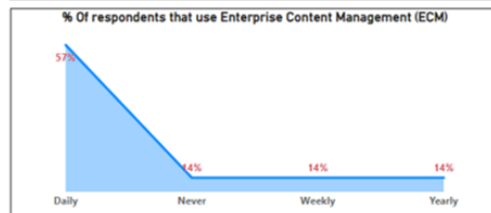
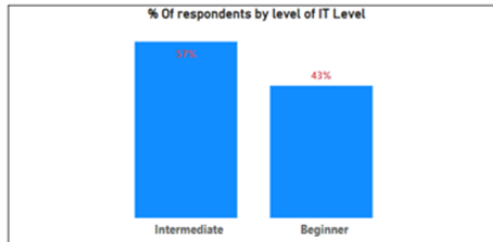
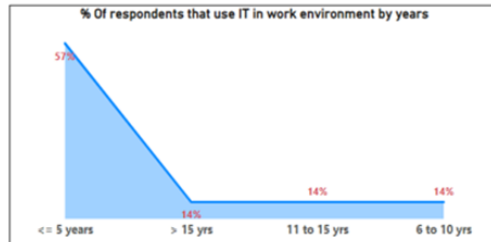
Non-Functional question responses - Advances and Expert respondents



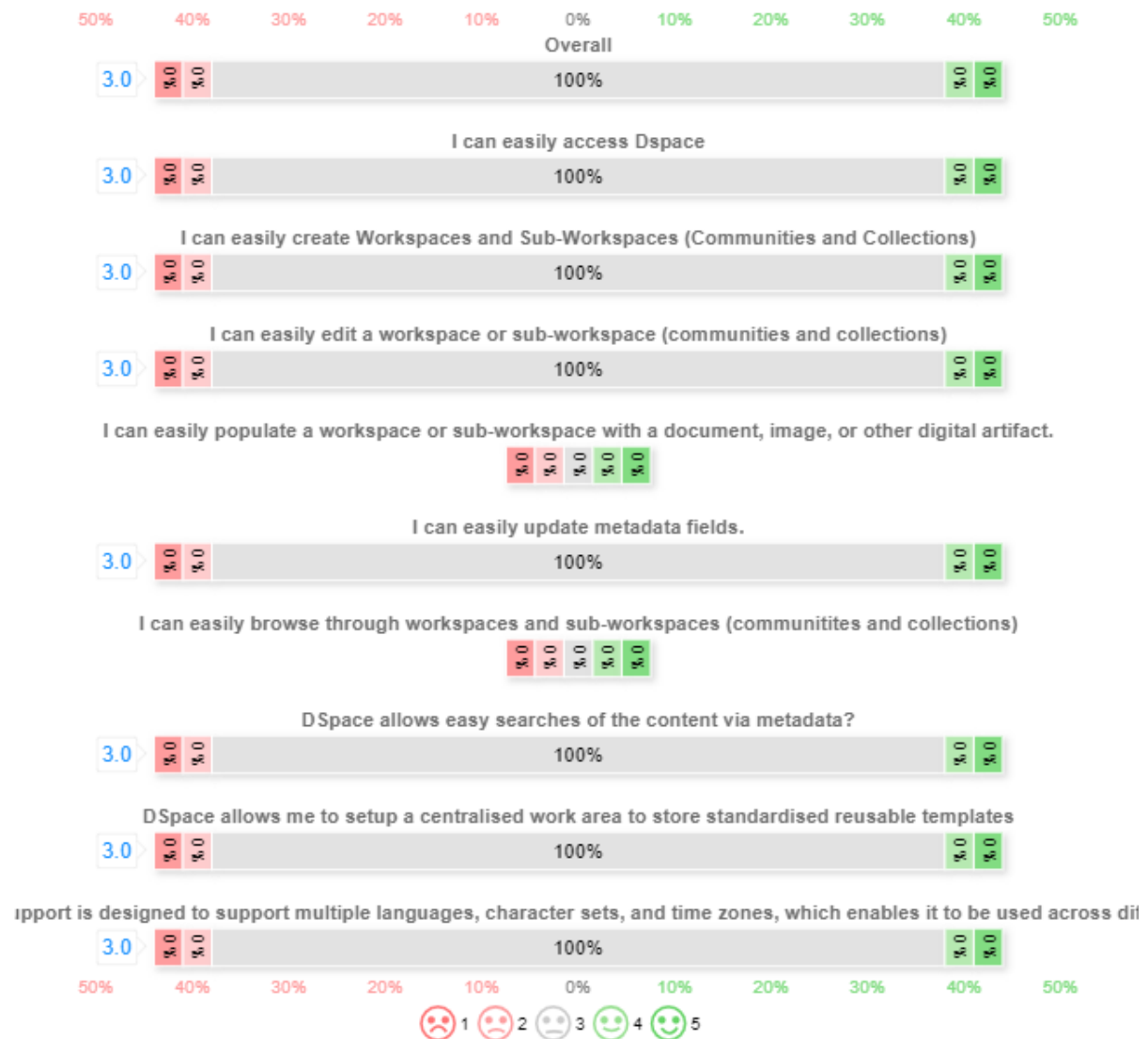
Background Questions – Intermediate and Beginner

Number of respondents

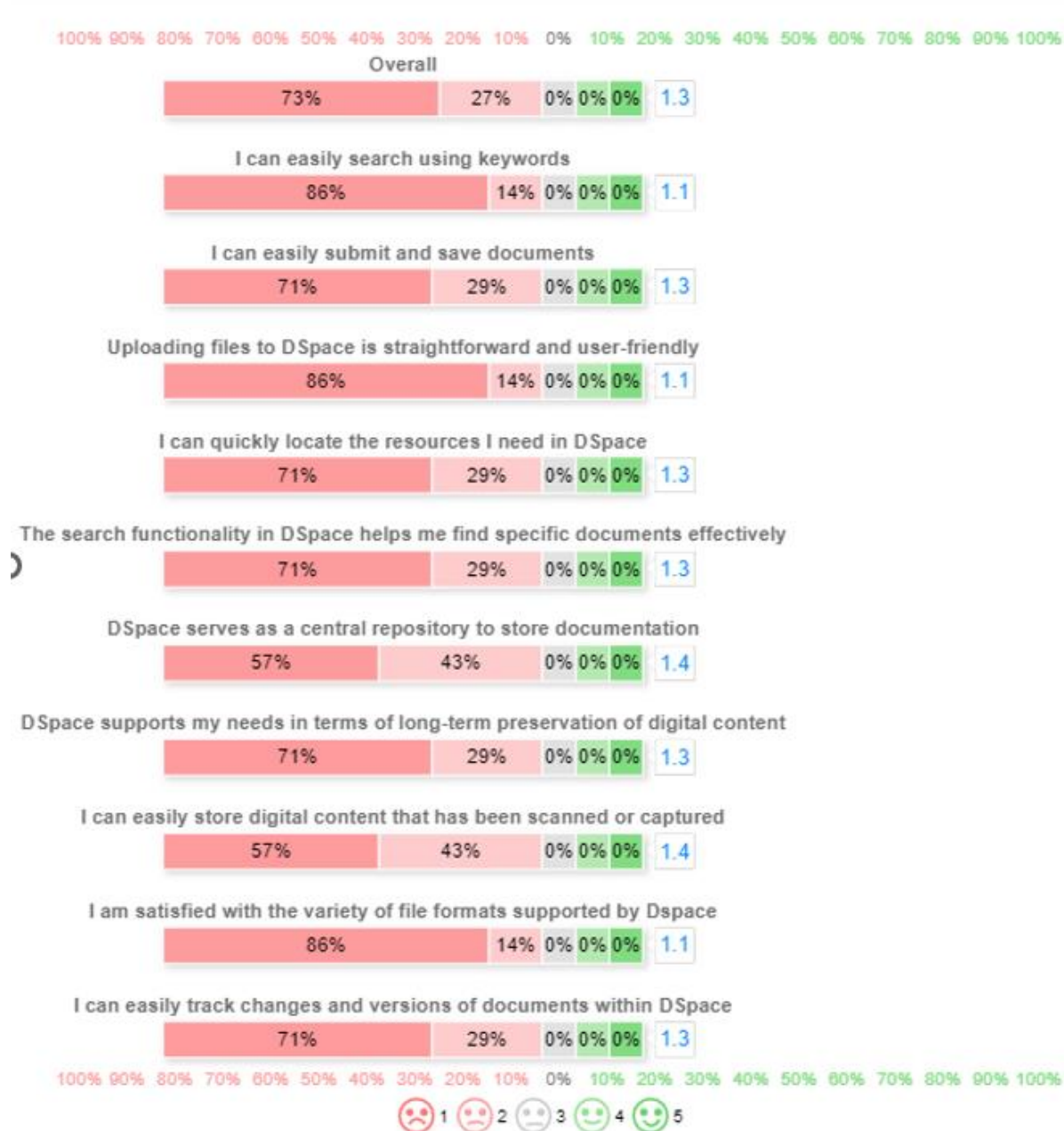
7



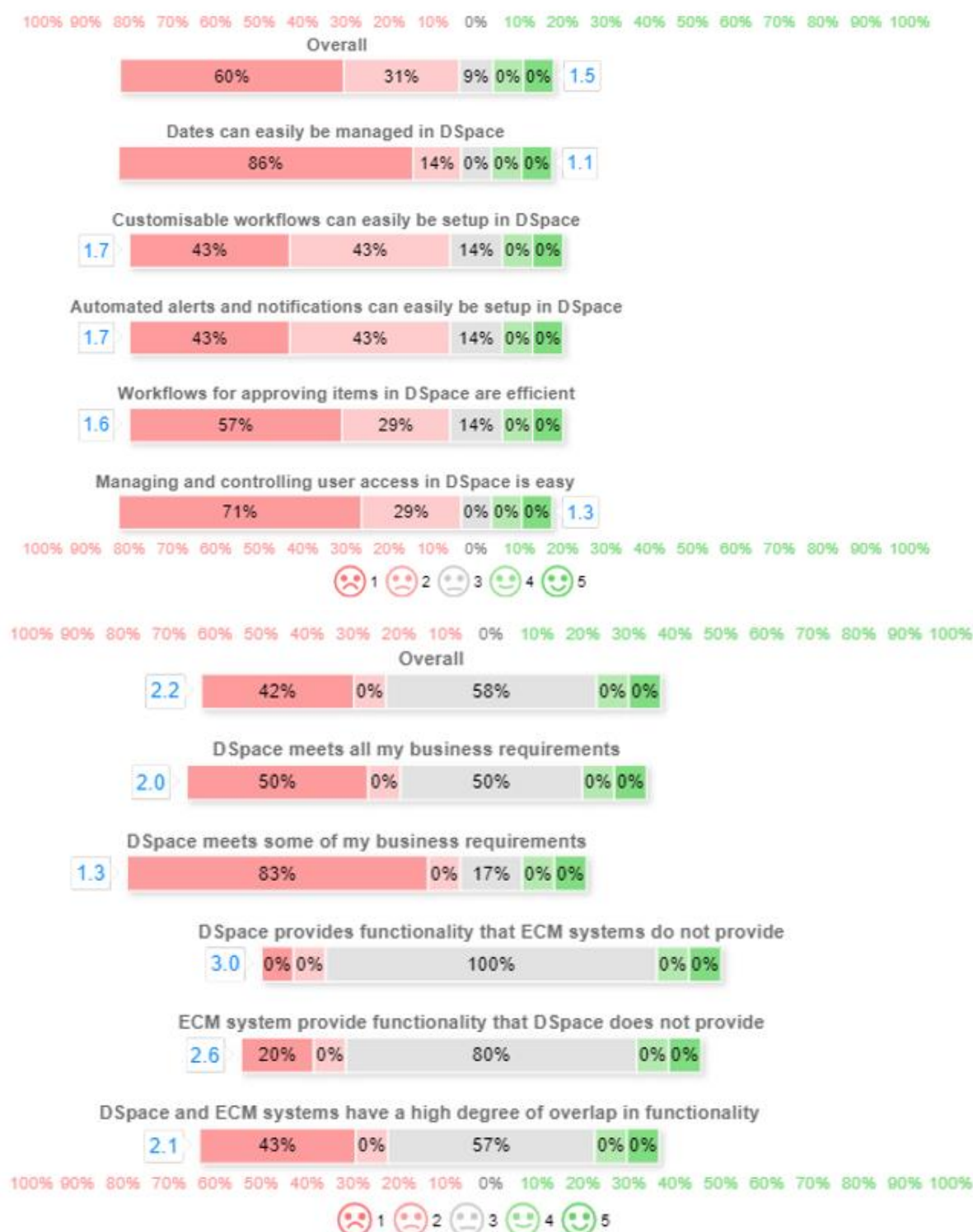
Functional question responses – Intermediate and Beginner respondents



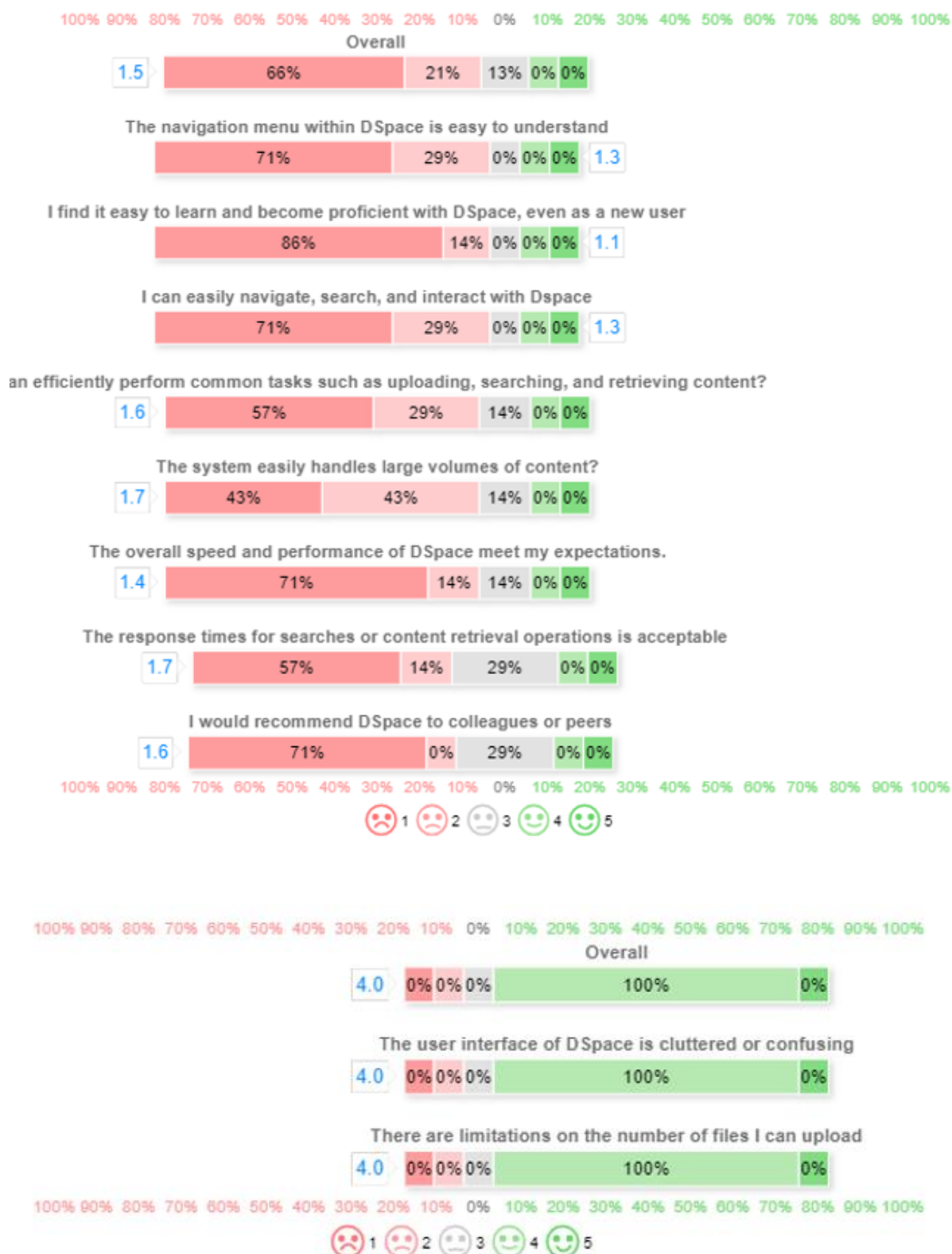
Functional question responses – Intermediate and Beginner respondents



Functional question responses – Intermediate and Beginner respondents



Non-Functional question responses – Intermediate and Beginner respondents



Appendix C: Ethical Approval (FSREC Ethics Approval 087-2022 Asary)



UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

Faculty of Science
University of Cape Town
Rondebosch
South Africa 7701

E-mail: melissa.densmore@uct.ac.za
Tel: 021 650-9111

28 November 2022

Gavin Asary
Department of Computer Science

Investigating the feasibility of utilising the DSPACE digital repository software toolkit in an industrial Environment

Dear Gavin Asary

I am pleased to inform you that the Faculty of Science Research Ethics Committee has approved the above-named application for research ethics clearance, subject to the conditions listed below.

- Applicant must get permission from the company prior to recruiting participants.
- Restrictions on involving human participants in research must be adhered to, given current concerns about the spread of Covid-19. Please ensure that you are aware of and comply with UCT policy on this, as communicated by management.
- Implement the measures described in your application to ensure that the process of your research is ethically sound; and
- Uphold ethical principles throughout all stages of the research, responding appropriately to unanticipated issues: please contact me if you need advice on ethical issues that arise.

Your approval code is: FSREC 087 – 2022

I wish you success in your research.

Yours sincerely

Signed by candidate

A/Prof Melissa Densmore
Chair: Faculty of Science Research Ethics Committee

Appendix D: Organisation Consent Approval Letter



Department of Computer Science

Private Bag X3 Rondebosch 7701

Tel: +27 (0) 21 650 2663

Internet: <https://www.cs.uct.ac.za/>

Request to conduct research and interview participation consent form

August 2023

Dear Sir/Madam,

In terms of the requirements for completing a Masters degree in Information Technology (MIT) at the University of Cape Town, a research study is required.

The researcher, Gavin Asary, has chosen to conduct a research study entitled "Investigating the feasibility of utilising the DSpace digital repository software toolkit in an industrial environment". The researcher would like to request permission to conduct this research study at your organisation.

The objective of the research is to perform an investigation whether the DSPACE Institutional Repository toolkit (DSpace) is suitable to meet the requirements of end users and administrators in an industrial environment. This will be done by requesting participants perform a user evaluation of DSpace to assess the feasibility of using the solution within an industrial environment given the typical practices users would follow. These practices would typically include, however not be limited to, the creation, editing, and collaboration of digital content allowing the user control over business documents and content across the enterprise. The outcome of this study is an initial attempt to bridge the gap between the institutional and industrial communities that develop digital repository toolkits.

Participation in this research is voluntary and participants can choose to withdraw from the research at any time, in accordance with ethical research requirements. All information will be treated with utmost confidentiality and used exclusively for the purpose of this study. No individual or organisational identifiable information will be recorded or published, ensuring anonymity of responses.

The data will be collected using an online survey which should take approximately 30 minutes to complete. If you are willing to participate in this study, please kindly sign the attached form and return to me at your earliest convenience.

Should you have any questions regarding this research, please feel free to contact me on 081 171 2100's or email: asgav001@myuct.ac.za

Your organisation's voluntary participation in this study would be greatly appreciated.

Sincerely

Signed by candidate

Signed by candidate

Gavin Asary

Researcher \ MIT Student, (UCT)
Department of Computer Science
University of Cape Town
Email: asgav001@myuct.ac.za

Supervisor

Prof Hussein Saleman
Head of Department and Professor
Department of Computer Science
University of Cape Town
Email: hussain@cs.uct.ac.za

Management Consent

I, Aveena Mothilall, as Chief Information Officer (CIO) of Engen Petroleum (Engen) give Gavin Asary the researcher of this study consent to conduct their study at Engen. I am aware that participation is voluntary and that participants may choose to withdraw from this study at any time, should they choose to do so.

Signed by candidate

21 August 2023

Signature

Date

Appendix E: Online Informed Voluntary Consent to Participate in Research Study

Informed Consent

Load unfinished survey

tion from academia in an industrial or commercial setting. This will done from an end user perspective, in an industrial environment, to determine the suitability of the services and functionality provided by the DSpace software toolkit. This will be done by assessing the software toolkit's features, with common functionality and architectural models to understand the impact on the user environment.

Before proceeding, it is important to note that DSpace is primarily intended to be used as an institutional repository for capturing, storing, indexing, preserving, and redistributing a community's intellectual output. However, when considering its application in an industrial setting or environment, you may be assessing its suitability as a tool for managing various types of digital artifacts relevant to industrial operations.

This research seeks to answer the following question:

Are the services provided by the DSPACE Institutional Repository toolkit suitable to meet the requirements of end users and administrators in an industrial environment?

There are 10 questions in this survey.
If you do not answer all the questions, your responses will be considered incomplete and not form part of the final analysis.

Feedback: Feedback of the results of the survey will be available on this link once the survey has closed and analysis has been completed.

Risk: There are no risks to participation in this study.

Confidentiality: Participation is anonymous as you will not be required to provide any personal information to identify you and all data collected from the survey will be anonymised. Pseudonyms will be used to remove your data if you withdraw. No company information or documents to be shared as part of this survey.

Disclaimer: Participation is voluntary, you may withdraw from the study at any stage without providing a reason and your data will be removed

What you agreeing to the survey means: By selecting "I agree to participate in this research" and submitting the consent form you are agreeing to participate in this study, and you have read and understood the benefits, procedure, feedback, risk, confidentiality, and disclaimer. You may withdraw at any stage.

Questions: Should you have any questions regarding this research, please feel free to contact Gavin Asary on 0837072885 or email: asrgav001@myuct.ac.za

We are sorry but you can't proceed without first agreeing to our survey privacy policy.

☐ I have read the Online Informed Voluntary Consent to Participate in Research Study and agree to participate in this research

Next

Appendix F: Key Business Requirements

#	Key Business Requirement	Description
1	Central Document Repository	The ability to centralise all documents in a single repository.
2	Ability to Extract Specialised Clauses from Documents	Functionality to extract specific clauses from documents for easy access and management.
3	Ability to Perform Document Renewals & Terminations	Features that support the renewal and termination of documents as needed.
4	Ability to Monitor Milestones	Tools to monitor key milestones and deadlines within document workflows.
5	Ability to Create Automated Workflows with Email Notification	Capability to create workflows that include automated email notifications.
6	Ability to Store and Access Generic Document Templates	Support for storing and accessing generic templates for document creation.
7	Easy Search Capability	A robust search function to quickly locate documents.
8	Ability to Scan and Upload Documents	Features that allow for scanning and uploading physical documents into the system.
9	Ensure Secure Document Data & Information, Including Appropriate Access Control	Security features to protect document data and enforce access control.
10	Reporting & Analysis	Tools to generate reports and analyse document-related data.
11	3rd Party Tool Integration (e.g., SharePoint)	Ability to integrate with third-party tools such as SharePoint.
12	Multi-Country Support	Support for managing documents across multiple countries.