

REPORT

Project Acronym: EODOPEN
Grant Agreement Number: 607666-CREA-1-2019-1-AT-CULT-COOP2
Project Title: EODOPEN | eBooks-On-Demand-Network
Opening Publications for European Netizens
Project Website: <https://eodopen.eu/>

A12 Evaluating delivery formats and solutions

D12a Technical report on the implementation of conversion services

Author(s):

Andreja Hari
Alenka Kavčič – Čolić

Project Co-funded by the Creative Europe Programme of the European Union		
Dissemination Level		
P	Public	X
C	Confidential, only for members of the project partners and the Commission Services	



DOCUMENT INFORMATION

Activity number:	A12
Activity title:	Evaluating delivery formats and solutions
Contractual date of activity:	31. 10. 2023
Actual date of activity:	1. 7. 2022–14. 11. 2023
Peer review:	14. 11. 2023–22. 1. 2024
Author(s):	Andreja Hari, Alenka Kavčič Čolić
Contributor(s):	
<i>Participant(s):</i>	<i>EODOPEN-project members</i>
Working group:	WG4
Working group title:	Delivery formats of digitised material for special needs
Working group leader:	Alenka Kavčič – Čolić
Dissemination Level:	P



HISTORY OF VERSIONS

Version	Date	Status	Author (organisation)	Description/Approval Level
1	3. 10. 2023	Draft	Andreja Hari (NUK) Alenka Kavčič – Čolić (NUK)	First draft
2	23. 1. 2024	Final	Andreja Hari (NUK) Alenka Kavčič – Čolić (NUK)	Final version after the review from partners.

EODOPEN PROJECT SUMMARY

Libraries all over Europe face the difficult challenge of managing tremendous amounts of 20th and 21st century textual material that has not yet been digitised due to the complex copyright situation. These works cannot be accessed by the general public and are hidden deep in library stacks, as they are often out of print or have never even been printed at all, while reprints or facsimiles are out of sight.

The **EODOPEN** project focuses on making 20th and 21st century library collections digitally visible by **directly engaging with communities** in the selection, digitisation and dissemination processes. As a leading partner, the University Library of Innsbruck, joined by 14 European libraries from 11 nations, has set itself the goal of making 15,000 pieces of textual material digitally available, and of reaching more than one million people in Europe by 2024.

Among other goals, such as building a common portal to display the project outcomes, EODOPEN aims to stimulate interest in and improve access to 20th and 21st century textual material, including grey and scientific literature. EODOPEN continuously carries out social media campaigns in order to attract new audiences. Furthermore, the participating libraries establish contact with commemorative institutions all over Europe, as well as with researchers and doctoral study boards, history associations and local publishing houses, in order to obtain suggestions from a broad audience.

In collaboration with local institutions, all of the project partners select hidden library treasures, **deal with rights clearance questions** and put new content online, while dissemination activities display the digital content via international channels.

In addition, EODOPEN aims to provide alternative delivery formats suitable **for blind or visually impaired users**. An international survey gathers data from a broad European public about the use of e-books. By evaluating this data, the project broadens its scope to alternative delivery formats in order to fulfil the needs of **blind or visually impaired users**.

In order to promote best practice in rights clearance among the library community, EODOPEN provides handouts and tools to make 20th and 21st century books available beyond the project's lifetime. In this regard, the project partners cooperate closely to develop an online tool for the documentation of rights clearance, especially suited for out-of-print and orphan works. Interactive workshops investigate needs related to **dealing with rights clearance** questions in order to implement the requirements of the international community in establishing the online tool.

ABSTRACT

The aim of the Report on trial implementations of conversion services (hereinafter The Report) is to help librarians and other organizations in the field of culture to make digitized content available to broader communities either by using conversion services or to inform users about such options. The Report complements the EODOPEN Project Deliverable 11: *Guidelines and recommendations for the provision of alternative and special formats*, which addresses delivery formats and criteria for increasing the quality of digitization results for users of mobile devices and blind and partially sighted users. It also complements the EODOPEN Project Deliverable 12b&c: *Report on trial implementations for mobile devices and print disabled users* which evaluated the delivered digitisation outputs by project partners and some of these outputs were used in further research. The outcomes of the research are covered by The Report, which focuses on trial implementation of conversion services in regard to providing different delivery file formats, depending on different software, file format type and accessibility criteria.

Statement of originality:

This report contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

TABLE OF CONTENTS

DOCUMENT INFORMATION	2
HISTORY OF VERSIONS	3
EODOPEN PROJECT SUMMARY	4
ABSTRACT	5
TABLE OF CONTENTS	6
LIST OF TABLES	7
LIST OF FIGURES	7
1. Introduction	8
1.1 Purpose.....	8
1.2 Description of the Report	8
1.3 Key concepts explanation.....	8
1.4 Context description: User needs for delivery formats	9
2. Evaluation of delivery formats: trial implementation	11
2.1 Background.....	11
2.1.1 Introduction of conversion services used for testing.....	13
2.2 Methodological approach	17
2.2.1 Testing sample	17
2.2.2 Evaluation	18
3. Evaluation results	25
3.1 Results according to conversion service used	26
3.1.1 Calibre	26
3.1.2 RoboBraille	29
3.1.3 Balabolka	33
3.1.4 Mathpix.....	35
3.1.5 WordToEpub.....	36
3.1.6 EasyConverter Express	37
3.2 Results according to established criteria.....	38
3.3 Results according to output format	46
4. Findings, possible solutions and recommendations	50
5. Summary	51
6. References.....	53
7. Vocabulary	54
8. Used acronyms	55
9. Annexes	I
Annex 1: Testing samples	I
Annex 2: Conversion report for Calibre	XVIII
Annex 3: Conversion report for RoboBraille	XIX
Annex 4: Conversion report for Balabolka	XXVI
Annex 5: Conversion report for Mathpix	XXIX

Annex 6: Conversion report for WordToEpubXXX
Annex 7: Conversion report for EasyConverter Express XL

LIST OF TABLES

Table 1: The appearance of the criteria across the 16 scans in the testing sample.....21
 Table 2: Textual outputs made by Calibre26
 Table 3: Textual outputs made by RoboBraille.....29
 Table 4: Audio outputs made by RoboBraille31
 Table 5: Audio outputs made by Balabolka33
 Table 6: Textual outputs made by Mathpix35
 Table 7: Textual outputs made by WordToEpub36
 Table 8: Textual and audio outputs made by EasyConverter Express.....37

LIST OF FIGURES

Figure 1: An example of using PDF Accessibility Checker 2021 to evaluate PDF structured tags and text order.23
 Figure 2: An example of source code for HTML format. Visible are sections: heading 1, paragraphs and mathematical expression for which MathML was used.....24
 Figure 3: Humanware’s Victor Reader Stratus4 M24
 Figure 4: Percentage of fully achieved criteria among all outputs.....45
 Figure 5: Percentage of fully achieved criteria when converting to DOCX based on the input format.46
 Figure 6: Percentage of fully achieved criteria when converting to EPUB based on the input format.47
 Figure 7: Percentage of fully achieved criteria when converting to PDF and PDF tagged over image format based on the input format.47
 Figure 8: Percentage of fully achieved criteria when converting to RTF based on the input format.48
 Figure 9: Percentage of fully achieved criteria when converting to HTML based on the input format.48
 Figure 10: Percentage of fully achieved criteria when converting to MP3 based on the input format.49
 Figure 11: Percentage of fully achieved criteria when converting to DAISY based on the input format.49

1. Introduction

1.1 Purpose

The aim of the Report on the trial implementation of conversion services (hereinafter referred to as “The Report”) is to assist librarians and other organizations in the field of culture in making digitized content accessible to wider communities. The Report serves as a supplement to the EODOPEN Project Deliverable 11: *Guidelines and recommendations for the provision of alternative and special formats based on the survey on special needs, as well as Deliverable 12b&c: Report on trial implementations for mobile devices and print disabled users* (hereafter referred to as Deliverable 12b&c). It is founded upon technical requirements related to delivery formats and criteria aimed at enhancing the quality of digitization outcomes for users of mobile devices, as well as blind and partially sighted users. The Report gathers the findings obtained from testing various conversion services that facilitate the conversion from the input format to a range of output delivery formats.

1.2 Description of the Report

The Report provides the outcomes of a trial implementation of various conversion services. It consists of a concise introduction, followed by a description of the methodology and the test results, and concludes with the findings and recommendations.

The introductory Chapter ([Chapter 1](#)) provides a description of the purpose, scope, and key concepts utilized in the text. Additionally, it defines the needs for users of mobile devices, blind and partially sighted users. [Chapter 2](#) offers the background information, the methodological approach, a description of the conversion services utilized, details of the testing samples, and the definition of evaluation criteria for delivery formats. [Chapter 3](#) presents the testing results of the conversion services. This Chapter is succeeded by a discussion of the findings and a set of recommendations ([Chapter 4](#)). The Report includes a list of literature sources and recommended references, definition of terms used and a list of acronyms. The [Annexes](#) feature the testing samples and the questionnaires used in the testing report.

1.3 Key concepts explanation

In the Report, mobile devices are defined as smartphones, notebooks, tablet computers, as well as e-readers. Within this text, the term “blind and partially sighted users” is used in accordance with the European Blind Union (EBU), instead of the term “blind and visually impaired users.” The term “print disability” encompasses a wide range of disabilities or difficulties when reading text. However, the focus of the Report is specifically on blind and partially sighted individuals, aligning with one of the primary target groups of the project.

Digitisation refers to the process of digitally converting information from analogue carriers. Format conversion entails changing the original delivery format to various output delivery formats (e.g., converting PDF to EPUB, DOCX, MP3 etc.). The target communities encompass individuals who access digitized and digital content in libraries and other cultural organizations.

Typically, the term “eBook” refers to publications that are born digital. However, in this document the term “eBook” is used to encompass digital publications produced as a result of digital conversion, including formats designed for special needs such as audiobooks, which aligns with the objectives of the EODOPEN project. Nonetheless, this term does not exclude digitally born publications, as they may share the same delivery format or serve the same purpose or functions. eBooks can be accessed through e-readers or simply read on personal computers (PCs) or mobile devices such as smartphones, tablets or notebooks.

1.4 Context description: User needs for delivery formats

Mobile devices have become indispensable tools in today’s global information society, and smartphone technology has seamlessly integrated into our daily lives, facilitating constant connectivity with other tools and people through diverse networks and social media. The advancement of mobile devices has also influenced the evolution of their operating systems and tools, which is a crucial consideration for the service sector, including libraries and cultural organisations. It is of utmost importance to plan and publish content in file formats that are currently supported and will continue to be supported by these devices.

eBooks can be read on various types of mobile devices, including e-readers (such as Kindle, Kobo, Midia Inkbook, NOOK, etc.), smartphones, tablets, and portable computers (notebooks). The selection of file format delivery and/or access depends on the device’s characteristics (screen size, visual presentation) and the existing platform (Microsoft, Android, and iOS being the most commonly used). While accessing PDF files on devices with larger screens poses no significant issues, they are not recommended for smaller devices like smartphones or e-readers due to PDF’s lack of responsiveness. The most recommended formats for smaller devices are EPUB 3, AZW/MOBI, HTML, Microsoft Office Word documents (RTF, DOCX, etc.), or audio books (MP3, DAISY). It is worth noting that certain platforms support specific formats exclusively. For example, Kindle e-readers initially did not support the EPUB format, necessitating the conversion of EPUB files to AZW/MOBI before uploading them to the devices. However, recent developments at Amazon have made EPUB format conversion available to users, while the obsolete MOBI format is no longer supported (Amazon, s.a.).

When considering blind and partially sighted users, as well as for other users with print disabilities, it is crucial to take into account the varying degree to which they can utilize their vision, which can fluctuate from day to day or be affected by factors such as light conditions,

tiredness, or stress levels. Therefore, it is important to provide users with the ability to adjust the visual presentation of text to suit their needs. Some common challenges faced by partially sighted individuals include difficulties in focusing on text, reduced contrast sensitivity, reduced field of vision, sensitivity to movement, visual fatigue, and similar issues. For these users, the most beneficial adjustments involve font size, font type, colour themes, margins, and spacing. It is also important to provide the option to access the full text, preferably with verified optical character recognition (OCR) as it enables more fluent use of assistive technologies such as braille display or screen readers. Although blind and partially sighted users primarily access documents through larger screens, they are also active users of smaller mobile devices. The most recommended formats for this user group include Microsoft Office Word documents (RTF, DOCX, etc.), audio books (MP3, DAISY), HTML, EPUB 3 and AZW/MOBI. However, the tagged PDF format is also suitable: “PDF tags are the key to accessing a PDF document’s content with assistive technologies such as screen readers. When a tagged PDF is created, each page element in the document is ‘tagged’. Each tag identifies the type of content and stores some attributes about it. They also arrange the document content into a hierarchical architecture (or a ‘tag tree’). The tag tree forms the logical structure of the document (reading order).” (Accessible document solutions, s.a.)

For a more comprehensive overview of user needs regarding delivery formats, please refer to Deliverable 11: *Guidelines and recommendations for the provision of alternative and special formats*, which is based on a survey on special needs of users and technical requirements. Additionally, Deliverable 12b&c provides a report on trial implementations for mobile devices and print disabled users, including testing conducted among EODOPEN project partners on delivery formats.

2. Evaluation of delivery formats: trial implementation

2.1 Background

The questionnaire survey conducted among the EODOPEN project partners in 2020 revealed that libraries use various devices and software tools in the digitization process. The digital conversion is either automated or carried out in different stages, depending on financial resources and the availability of adequately trained personnel. To achieve the best possible outcomes, a combination of technological and software solutions is needed. As a result, digitized materials are diverse and made accessible to users through digital libraries in various delivery formats. However, these delivery formats also differ in terms of their provided functionalities.

The *Guidelines and recommendations for the provision of alternative and special formats* (Deliverable A11), prepared within the framework of project's working group 4¹, place special emphasis on the possibilities and methods of adapting digitized materials into formats that ensure accessibility for blind and partially sighted users. Additionally, the Deliverable 12b&c presents testing results among EODOPEN project partners regarding delivery formats. These results demonstrated a wide range of delivery file formats, and it was found that additional manual corrections give better results. Among the automatically generated results, the PDF/UA format, based on universal accessibility standard, provided the best outcomes.

According to research and various guidelines provided for accessible digital publications, the proper structuring and personalisation of text and its elements is crucial for reading digital material, as it facilitates navigation within the text. This becomes even more important for blind and partially sighted individuals who rely on assistive technology and specialized software for reading. OCR tools and their software adaptations enable the optical recognition of characters, including letters, numbers, and punctuation marks, as well as text structures. With advancements in machine learning technology, OCR errors have become negligible. However, errors may still occur when reading special characters, such as chemical formulas, mathematical operations and equations. Additionally, errors can arise in identifying headings, sub-headings, and in identifying graphical elements within a text, such as images or graphs. In cases where there are multiple text columns, the correct text flow is often not recognized, resulting in a linear sequence of the text instead. For a blind person using speech synthesis to read, the text becomes unreadable without a proper interpretation of special characters, the specific sequence of the text, and accompanying descriptions of graphic elements.

¹ The Working Group is led by National and university library (Slovenia).

Perception, operability, understanding, and robustness, as defined by The World Wide Web Consortium (W3C) through the Web Accessibility Initiative (WAI) and included in the Web Content Accessibility Guidelines (WCAG), serve as the umbrella criteria for ensuring the accessibility of websites and digital materials for blind and partially sighted users as well as other user groups with disabilities. Within the framework of the aforementioned working group, our objective was to address these accessibility criteria, which also apply to born-digital materials or eBooks.

The objectives of the trial implementations were as follows:

- To assess the impact of input formats by selecting five different formats from Deliverable 12b&c, which would serve as input formats for conversion. We sought to include automatically generated results as well as results with manual corrections that attempted to meet the accessibility criteria.
- To choose the most commonly used, newest or recommended conversion services for the five input formats, with a focus on open-source software and tools.
- To obtain detailed information on the capabilities and features of the chosen conversion services for their users, including librarians and library users.
- To determine the most recommended output formats for the conversion process.
- To compare the same output formats converted with different conversion services, allowing for the identification of the conversion services that produce better results.
- To evaluate all received output formats based on the established accessibility criteria, aiming to identify the most suitable solutions.
- To identify which conversion services or input formats yield the best results with minimal additional work, particularly in terms of accessibility for blind and partially sighted individuals.

The purpose of the testing is to evaluate the most effective solutions in the format conversion process after digitization has been completed and to identify areas for improvement in providing digitized materials that meet accessibility criteria. This enables libraries to review and enhance the range of formats they offer to their users, considering both mobile devices users and individuals who are blind or partially sighted, depending on the available resources.

As mentioned earlier, sighted individuals do not require the same level of precision in text processing to access the content of digitized and digital works. However, responsive technologies also rely on accurate OCR, and it should be noted that sighted individuals may utilize screen readers to have the text read aloud to them. Consequently, our entire testing process was based on criteria that are essential for blind and partially sighted individuals, aligning with the principle of universal design (for everyone). In [Chapter 5](#) of this document, potential solutions and recommendations are provided for both mobile device users and print disabled users, in case libraries choose to focus on one particular user group. Nonetheless, the utilization of solutions for print-disabled users is strongly recommended.

2.1.1 Introduction of conversion services used for testing

Three of the conversion services were previously introduced in Deliverable 11. However, for a comprehensive understanding in the report, it is crucial to introduce all the services that were used for testing. We specifically selected the most commonly used, newest, or recommended conversion services that offer a range of options for delivery format conversion. Although none of the partner institutions in Deliverable 12b&c produced audiobooks, we placed greater emphasis on this section as well. As mentioned in the background, audiobooks are highly utilized by blind and partially sighted individuals, and their popularity is increasing among other print disabled users and users without disabilities.

Calibre version 6.1.0

Available at: <https://calibre-ebook.com>

Cost: free.

About: The open-source software is developed by various contributors from the user community.

Calibre is a “cross-platform open-source suite of e-book software. Calibre supports organizing existing e-books into virtual libraries, displaying, editing, creating and converting e-books, as well as syncing e-books with a variety of e-readers. Editing books is supported for EPUB and AZW3 formats.”^{2, 3} Calibre primarily focuses on text-based documents and does not support conversion into audio formats.

It supports the following operating systems: Windows, macOS, Linux.

Calibre also features a viewer application that allows users to read eBook, and adjust basic features such as font, size, margins. An interesting feature is the ability to adjust the text according to the size of the viewer window, even on a computer.

Input format options: AZW3, DOCX, EPUB, PDF, RTF.

Output format options: EPUB, AZW3, MOBI, DOCX, FB2, HTMLZ, LIT, LRF, PDB, PDF, PMLZ, RB, RTF, SNB, TCR, TXT, TXTZ and ZIP.

First impressions: Conversion process is simple, enabling bulk conversion of multiple books simultaneously, and offering a wide range of output formats. However, it is recommended to modify settings and metadata as needed. It should be noted that Calibre does not support the conversion of a single book into multiple formats simultaneously.

RoboBraille

Available at: <http://robobraille.org>

Cost: free online tool for personal use, subscription fee for institutional use.

² Wikipedia. (2021). Calibre (software): [https://en.wikipedia.org/wiki/Calibre_\(software\)](https://en.wikipedia.org/wiki/Calibre_(software))

³ The software works similarly as Sigil as it enables entrance to each XHTML file, navigation file, styles used in the book, media included in the book etc. Preview is also enabled so that the changes are immediately visible.

About: The web service is developed by RoboBraille, Denmark. “RoboBraille is an e-mail and web-based service, capable of automatically transforming documents into a variety of alternate formats for the visually and reading impaired.”⁴ It offers braille, audio, e-book, and accessibility services. The braille services provide conversion to contracted or uncontracted Braille for 18 Braille code languages. Audio services offer conversion into MP3 files and DAISY talking books, with audio available for 28 languages. EBook services enable document conversion primarily to EPUB and MOBI formats. Accessibility services allow the conversion of inaccessible documents (GIF, TIFF, JPG, PNG, BMP, PCX, DCX, J2K, JP2, JPX, DJV, image-only PDF) into more accessible formats (tagged PDF, DOC, DOCX, Word XML, XLS, XLSX, CSV, TXT, RTF and HTML).

Input format options: DOC, DOCX, PDF, PPT, PPTX, TXT, XML, HTML, HTM, RTF, EPUB, MOBI, TIFF, TIF, GIF, JPG, JPEG, BMP, PNG, PCX, DCX, J2K, JP2, JPX, DJV, TEX, ZIP and ASC. The maximum size of the file is limited to 64 MB, which may cause difficulties when working with larger documents.

Output format options: EPUB, MOBI, tagged PDF, DOC, DOCX, Word XML, XLS, XLSX, CSV, TXT, RTF and HTML, MP3, DAISY full text and audio, DAISY Math full text and audio.

First impressions: Conversion can be cumbersome when a user needs to convert a single input format to multiple output formats. This process requires the user to upload the file each time and receive a separate mail for each output format. Furthermore, the output format is stored on the RoboBraille server for a few days before being deleted, necessitating the user to convert the file again if they require it in the same format. It is important to note that different input formats offer varying options for output formats; not all input formats provide the same conversion choices.

For instance, when uploading an EPUB file, only a limited number of conversion options are available, while DOCX files offer more choices, including the DAISY format. When using the English interface, not all languages for audiobooks are visible. However, changing the interface reveals additional options (e.g., the Slovenian language is missing in the English interface, and a similar case occurs with the Portuguese language). Some languages are still missing altogether (e.g., Estonian). Additionally, the range of available languages varies depending on the output format. For example, the DAISY format offers only 10 language options compared to 42 options available in MP3 format when using the English interface.

When receiving audiobooks from the tool, a message appears stating, “Please note that the e-book is made strictly for personal, non-commercial use. You are not allowed to distribute the e-book or to use it for commercial purposes.” This message emphasizes caution when using the tool for personal purposes. For institutional purposes, the website specifies that “for institutional use (including academic institution, organisation, association or similar), the RoboBraille web interface is available in customised versions to academic institutions and other organisations that wish to include it on their websites as an accommodation. The

⁴ RoboBraille (s.a.) Introduction to RoboBraille. Available on 20. April 2023 at <https://www.robobrainle.org/introduction-robobrainle/>

document conversion capabilities of RoboBraille can also be integrated with library systems and learning management systems in order to support users with special needs. Institutional use is subject to a subscription fee."⁵ Additionally, libraries could consider providing this service to offer a variety of input format options to their users.

Balabolka version 2.15.0.821

Available at: <http://www.cross-plus-a.com/balabolka.htm>

Cost: free.

About: Balabolka is a Text-To-Speech (TTS) software designed for converting text into audio files. It supports synthetic voices installed on a computer system and can convert from various text file formats. The software is compatible with Microsoft Windows XP/Vista/7/8/10/11 operating systems. In addition, the program allows users to adjust a voice's parameters, such as rate and pitch. A special substitution list can also be applied to enhance the quality of the voice's articulation⁶. The output audio can be generated as a single file or split into smaller audio files, facilitating the swift creation of audiobooks.

It is important to note that Balabolka supports non-commercial voices. However, if users choose to utilize commercial voices, they are required to contact the voice creator and purchase the license for commercial use or audio broadcast.⁷ The guidelines for using commercial voices for non-commercial purposes are not entirely clear, posing potential challenges for libraries, as there is no institutional solution similar to RoboBraille's agreement between parties. Balabolka is better suited for individual purposes, and libraries could assist users by providing accessible input format options.

Input format options: ODT, DOC, FB2, EPUB, RTF, PDF, HTML and clipboard content.

Output format options: MP3, WAV, WMA, OGG, MP4.

First impressions: The software to creating text-to-speech based audiobooks is user-friendly and requires basic knowledge. It enables text editing before conversion, allowing for additional changes such as adding elements (e.g., alt-text which is typically a hidden element not visible in the text and needs to be manually inserted) or removing elements (e.g., certain parts that are unnecessary in the audiobook, such as a list of all original pages in the table of contents). It also provides additional settings to optimize the audio output quality.

Mathpix's Snip

Available at: <https://mathpix.com>

Cost: free, free educational subscription or pro subscription for \$4,99 per month.⁸

⁵ RoboBraille (s.a.) Introduction to RoboBraille. Available on 22. January 2024 at <https://www.robobrace.org/introduction-robobrace/>

⁶ Cross+A (s.a.) Balabolka. Available on 20. April 2023 at <http://www.cross-plus-a.com/balabolka.htm>

⁷ Balabolka 2.15.0.844. Available on 20. April 2023 at <https://www.techspot.com/downloads/5814-balabolka.html>

⁸ Mathpix Products Pricing. Available on 20. April 2023 at <https://mathpix.com/pricing>

About: The software, developed by Mathpix, includes Snip, a tool designed for extracting text, particularly mathematical equations, from images and documents. It enables conversion for PDFs using AI-powered document conversion technology.

Input format options: PDF.

Output format options: LaTeX, DOCX, HTML, Overleaf, Markdown and more.

First impressions: In terms of output formats, the recognition of Math as MathML is excellent, even for complex equations. However, there are some issues encountered. The tool utilizes its own OCR recognition, resulting in differences from the original content, and it may struggle with complex structures. Additionally, some pages may be missing in the output formats.

WordToEpub 1.0.10

Available at: <https://daisy.org/activities/software/wordtoepub/>

Cost: free.

About: The WordToEpub software, developed by The DAISY Consortium, is used to convert documents from various word processors (Microsoft Word, Google Docs, LibreOffice Writer or Apple Pages) to EPUB and HTML formats. The supported operating systems include Windows 7, 8, 10, and 11. An additional feature called EpubToWord became available since the version 1.0.10 (February 2023) of the software, allowing for the conversion of EPUB format to Word. This option became available to users who are accustomed to editing documents in word processors rather than utilizing more complex software like Sigil or Calibre, where access to the underlying code requires at least basic or advanced knowledge of HTML and CSS. It is worth noting that according to the developers, the conversion from EPUB to Word is only applicable to DRM-free documents⁹.

Input format options: doc, DOCX.¹⁰

Output format options: EPUB, HTML. Additionally, it is also connected to Calibre for further conversion to other formats.

First impressions: If the input format is well-structured and properly prepared, the resulting EPUB and HTML formats appear excellent and offer good accessibility. The tool functions as a wizard and guides users through a few steps in simple or advanced mode. The software's development is focused on ensuring accessibility for end users and is continually improving.

EasyConverter Express 3.0.1.55

Available at: <https://yourdolphin.com/EasyConverter-Express> or Microsoft store.

Cost: free 30 days trial or one time pay for 71.99 €, available also as add-in or as EasyConverter Express Online for 150 €.

About: The software is developed by Dolphin Computer Access and is used for simple 3-step conversion.

⁹ Documents which are not protected by Digital Rights Management services (DRM).

¹⁰ For EpubToWord the input format is ePub and output DOCX format.

Input format options: doc, DOCX or Office 365 documents.

Output format options: braille, large print, MP3 and text only EPUB.

First impressions: While there were no issues encountered with the EPUB output format, the software encountered difficulties with the MP3 format. Sometimes, it was necessary to attempt the conversion two or three times to obtain the desired result, or in one case the software would stop at 10%.

2.2 Methodological approach

2.2.1 Testing sample

The testing phase of the conversion services took place between 22nd July 2022 and 21st March 2023. From the testing results of Deliverable 12b&c (comprising 23 results), we selected five different delivery formats provided by EODOPEN partners: PDF, PDF/UA, RTF, DOCX, and EPUB. These delivery formats all contained the same content, which was a testing sample consisting of 16 scanned pages (refer to [Annex 1](#)). Among the five delivery formats, two were automatically generated digitized results (PDF and PDF/UA), while the remaining three had manual corrections applied by project partners (RTF, DOCX and EPUB). Among the last three, EPUB format has been most accessible. These chosen delivery formats served as the input formats for the conversion services described in chapter [2.1.1 Introduction of conversion services used for testing](#). Not all results from Deliverable 12b&c were selected, as it was necessary to limit the number of examples to manage the research effectively without overwhelming the analysis with excessive output conversion results.

The samples used in the testing consisted of both textual and non-textual elements, including plain text, chapters and sub-chapters, columns, tables, footnotes, flowcharts, images, and accompanying text for images (captions). The text also contained special examples such as chemical formulas, mathematical equations, and special characters (£, °C, etc.). Two scans contained two pages on a single scan. The first scan having the title and author's name spread on both pages, and the second scan displaying the chemical periodic table across both pages. Most of the scans had complex structures with elements that could disrupt the order of the text (e.g., captions) or pose challenges for elements recognition (e.g., tables). Only three scans (8, 9, 13) had a basic layout with text in a single column and accompanying image, which should not cause significant difficulties in terms of reading order.

The objective of the testing was to compare and identify differences among conversion services, input and output formats, and to determine which conversion services or input formats yield the best results with minimal additional work, particularly in terms of accessibility for blind and partially sighted individuals.

We utilized six different conversion services with the selected input formats and focused on the most commonly used output formats that would enhance usability for both mobile devices users and blind and partially sighted users. The details of the work performed with the conversion services, along with screenshots depicting their usage, are provided in Annexes 2–8.

2.2.2 Evaluation

For evaluating the outputs, we established **25 criteria**¹¹ based on WCAG guidelines and other best practices, with a primary focus on accessibility for blind and partially sighted individuals. These criteria were developed individually for each scan since not all of them were applicable to every scan. Additionally, some criteria were specific to particular scans due to potential variations in the OCR process, such as page rotation and double pagination.

The criteria used to evaluate each output of the digitisation process were:

- 1. ALT-TEXT PICTURE:** Alt-text or alternative text for pictures provides a textual description for non-text content such as pictures, graphics, diagrams. It is essential for blind and partially sighted users to understand the content of the visual material and not miss any information conveyed through graphics. This criterion primarily benefits blind and partially sighted users but can also be useful for sighted users utilizing speech synthesis.
- 2. ALT-TEXT PICTURE (CHEMICAL FORMULA):** This criterion is the same as the “Alt-text picture” criterion but specifically applies to the two special images in the test sample that present molecular reactions (refer to [Scan 2 in Annex 1](#)).
- 3. CAPTION:** Some images and tables in the scans are accompanied by captions. In the document, it should be clearly indicated that the text is a caption associated with a picture, distinguishing it from general paragraph text.¹² This criterion is particularly important for blind and partially sighted users.
- 4. FOOTNOTES:** Footnotes provide additional information related to the main text and should be technically separated from the main text, allowing readers to skip them if desired. When creating or editing footnotes, the result should enable readers to navigate from the main text to the footnote and easily return to the corresponding section in the text.¹³ This criterion is mainly important for blind and partially sighted users.
- 5. HEADING 1:** Headings of the chapters should be clearly marked and structured hierarchically (e.g., Heading 1, 2, 3, etc.) for navigation purposes. Headings can also be utilized to generate a table of contents. This facilitates easier navigation from chapter to chapter for users of assistive technologies, allowing them to navigate within the

¹¹ Compared to D12b&c, we added the criteria Personalization.

¹² Captions can be inserted technically. In tagged PDFs, for example, a specific tag can be added in Adobe Acrobat Pro. When working in Microsoft Word, the “insert caption” option can be used.

¹³ Good results can be achieved, for example, in Microsoft Word, HTML or EPUB by providing two-way hyperlinks.

document without having to read the entire content. This criterion is important for all users.

6. **HEADING 2:** This criterion is the same as the "Heading 1" criterion.
7. **HEADING 3:** This criterion is the same as the "Heading 1" criterion.
8. **INITIAL:** The larger first letter at the beginning of a chapter is often not recognised or not recognised correctly (refer to [Scan 7 in Annex 1](#)). This criterion is important for all users.
9. **LANGUAGE SEGMENTS:** This criterion is related to the Primary Language criterion. Language Segments were assessed on six different occasions in the test sample (Italian + Latin, Italian, French twice and German twice) where text appeared in a language other than the primary language (English). Proper identification of the language is crucial for users of screen reading technologies, as it allows for appropriate audio settings and pronunciation.¹⁴ This criterion is primarily important for blind and partially sighted users but can also benefit sighted users utilizing speech synthesis.
10. **MATH (SIMPLE):** The recognition of mathematical or chemical elements was divided into two criteria based on complexity. The first criterion pertains to simple mathematical elements that appear in a single line, which pose fewer problems for OCR (e.g., $e = e' - AB(t - t')$). This criterion primarily impacts blind and partially sighted users.
11. **MATH (ADVANCED):** The second criterion for mathematical and chemical elements encompasses expressions that span two or more lines. These elements, often featuring subscripts, superscripts, Greek letters, fractions, and complex expressions (e.g., x^2 , $2H_2O$, 10^{-4} , $C_6H_{12}O_6$ or $\Delta p = \rho_v gh$ or $\Delta p = \frac{2T\rho_v}{R(\rho_w - \rho_v)}$), are typically not recognized accurately during OCR. The test sample includes various problematic elements. This criterion is primarily important for the blind and partially sighted.
12. **OCR ERRORS (TEXT IN PICTURE 4):** One image displayed text written on a tombstone (refer to [Scan 7 in Annex 1](#)). Ideally, this type of text should not be recognized, but the purpose was to evaluate the results obtained. The criterion is relevant for all users.
13. **PAGE ROTATION:** This criterion was applied in a single case where a table appeared horizontally on a page. To improve OCR and structural results, rotating the page to correctly orient the table would be beneficial. This criterion is relevant for all users.
14. **PAGINATION:** This criterion was created with blind and partially sighted users in mind. It is common for these users to prefer pagination as the initial information when accessing a page. Even if the pagination physically appears at the bottom of the page, blind and partially sighted users find it helpful for it to be presented first. This criterion is primarily important for blind and partially sighted users but can also be useful for sighted users utilizing speech synthesis or for easier navigation to specific pages within the document.
15. **PAGINATION-DOUBLE:** This criterion was applied in two cases where content spanned across two pages. The first case involved an image of the periodic table of elements, while the second case involved the title and author of the article, both stretching across two

¹⁴ For example, a German text that is read aloud with an English voice sounds strange.

pages. In both cases, better results would be achieved if the pages were not split. This criterion is important for all users.

- 16. PERSONALISATION:** This special criterion evaluates how the format behaves when the text is enlarged on a computer, particularly if reflow is available, allowing the text to adapt to smaller screens. It also assesses the ability to modify the visual appearance of the text, such as changing fonts, colour themes etc. This criterion applies to the entire book rather than individual scans. It is important for all users.
- 17. PICTURE:** A graphical element that should be identified as a separate element and include alt-text for users of assistive technologies. This criterion is primarily important for blind and partially sighted users.
- 18. PICTURE (CHEM. FORMULA):** Same as the Picture criterion. This criterion specifically applies to two images that depict molecular reactions and should also include alt-text. It is primarily important for blind and partially sighted users.
- 19. PRIMARY LANGUAGE:** The Primary Language should be specified for each document. This is important for users of screen reading technologies that provide sound in the correct language. Since the text in the test sample was in English, the Primary Language should be set to English. This criterion is mainly important for blind and partially sighted users, but could also be useful for sighted users utilizing speech synthesis.
- 20. SPECIAL CHARACTER:** This criterion appeared in three different cases involving the characters °C, £ and decimal numbers. The goal was to determine if there would be any issues recognising these characters. In the scan of decimal numbers, the numbers in the original were written with an apostrophe (') instead of a full stop (.) which is the norm in English. The test was conducted to see if any correct results would be obtained. This criterion is important for all users.
- 21. STAMP REMOVAL:** Library stamps in books can affect the recognition of nearby characters. The objective was to determine whether removing the stamp from the scan would result in clearer OCR in that area. In the provided example, the stamp was positioned directly over the text, and it was expected to negatively impact OCR results. This criterion is important for all users.
- 22. TABLE:** Tables are structural elements that should be appropriately marked and not solely represented as images. The table header and rows should also be included following the structure.¹⁵ This criterion is primarily important for blind and partially sighted users, but could also be useful to sighted users using speech synthesis.
- 23. TABLE HEADER:** This refers to an element of a table that typically appears at the top, but can also be located in the first column. It provides key information about the data in the rows that follow, aiding users of assistive technologies in easier navigation and

¹⁵ The structure of the table can be created technically. For example, in tagged PDFs, tags appear for table, table header, table rows and table data, much like in HTML formatting. Microsoft Word, for instance, also has the option to set a table header.

comprehension of the table. This criterion is primarily important for blind and partially sighted users, but it could also be useful for sighted users utilizing speech synthesis.

- 24. TABLE ROWS:** These are structural elements that follow the table header. In the test sample, which did not include the gridlines to mark the lines in the table, it was of interest to observe if the rows had been properly identified and how well the OCR tool recognised the number of rows. This criterion is primarily important for blind and partially sighted users, but could also be useful for sighted users using speech synthesis.
- 25. TEXT ORDER:** This criterion addresses the organization and sequence of the text, particularly when the page structure is more complex, such as with multiple columns and additional graphical elements. When users copy text, convert formats, or use assistive technology, it is crucial that the text is presented in the correct order to avoid confusion (e.g., captions appearing in the middle of paragraphs) or burdening users with the task of editing the content themselves. Some OCR software tools also offer the option to correct the type and order of recognized elements.¹⁶ Additionally, assistive technologies provide users with text linearly from top to bottom, making the correct text order vital for understanding and navigating the content. This criterion is important for all users.

Additionally, when evaluating audiobooks, the criteria were treated slightly differently, particularly in relation to headings¹⁷, the understanding of text flow, and mathematical expressions. Furthermore, three criteria - picture, picture (chem. formula), and personalisation – were excluded as they are not relevant to audiobooks. It should also be noted that in audio format, there are instances where text-to-speech technology may introduce errors, requiring us to listen to the entire audiobook.

Table 1 provides an overview of the occurrence of the 25 mentioned criteria, organized by scan number in the testing samples. As previously mentioned, personalization is an additional criterion that applies to the entire document.

Table 1: The appearance of the criteria across the 16 scans in the testing sample.

Criteria\scan	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	=
alt-text picture	1	1	1	1	3	2	4					1	1	3			18
alt-text picture (chem. formula)		2															2
caption	1	1	1	1	3	2	4			1	2	1	1			1	19
footnotes												1					1
heading 1	1				1	1	1		1				1		1		7

¹⁶ The most frequently used software for OCR – Abbyy FineReader desktop version – has this option during processing the digitised content. For post-processing, an example of software of this kind is Adobe Acrobat Pro.

¹⁷ Headings criterion was accepted if the audio files were either separated per each heading 1 or if DAISY audiobook enabled moving across separate headings.

Criteria\scan	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	=
heading 2	1	1	2			4			1							1	10
heading 3									1								1
initial							1										1
language segments						1	1			1	1	1	1				6
math (simple)	1	1										1					3
math (advanced)	1	1			1							1					4
OCR errors (text in picture 4)							1										1
page rotation										1							1
pagination	1	1		1	1		1	1	1	1	1	1	1			1	12
pagination–double				1			1										2
picture	1	1	1	1	3	2	4					1	1	3			18
picture (chem. formula)		2															2
primary language	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16
special character	1											1	1				3
stamp removal			1														1
table										1	2						4
table header										1	2						4
table rows										1	2						4
text order	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16
personalisation																	
=	11	13	8	7	14	14	20	3	6	9	12	11	9	8	3	8	

Same as in Deliverable 12b&c, three levels were used to evaluate the set criteria:

- *Criterion was fully achieved (A)*: This designation is used when both the technical and content parts of the criterion were met. For example, if the table rows were technically correct (each row contained the right number of rows and the correct content), this level would be assigned.
- *Criterion was partly achieved (B)*: This designation is used when either the technical or content part of the criterion was achieved, but not both, or when there was a very minor mistake in the criterion. For instance, if the alt-text is technically correct, but the content is either the text of the caption or other surrounding text, or if there was a minor mistake in the text order, this level would be assigned.
- *Criterion was not achieved (blank cell)*: This designation is used when neither the technical nor content part of the criterion was met. For example, if pagination was present but was not the first element on the page, this level would be assigned.

The evaluation was conducted using various software, tools and hardware according to different output formats:

- DROID: This software was used to determine the versions of the formats (PDF 1.4 or RTF 1.9).
- Adobe Acrobat Reader Pro: This software was used to check the content of PDFs, as well as the reading order and structure (tags) when the PDF was tagged;
- PDF Accessibility Checker 2021: This software was used to identify errors in the PDF file according to standards, and to check if the language of the document was set. It also checked the reading order and structure (tags) when the PDF was tagged (Figure 1).
- Thorium Reader: This software was used to review the content of EPUB files and determine the available options for visual adjustments and navigation, including the presence of a table of contents within the software.
- Sigil: This software was used to examine the content of EPUB files, as well as the reading order and structure (HTML tags).
- Edge browser: The browser was used to inspect HTML formats, including the design, view, and source code of the HTML (Figure 2).
- Microsoft Office Word: This software was used to review the content of DOCX and RTF files.
- Windows Narrator: This speech synthesis software was used in special cases to assess how the content is presented to the user.
- Windows Media player: This software was used to listen to MP3 audiobooks.
- Humanware’s Victor Reader Stratus4 M: This Daisy and MP3 player¹⁸ hardware was used to evaluate the results in DAISY format (Figure 3).

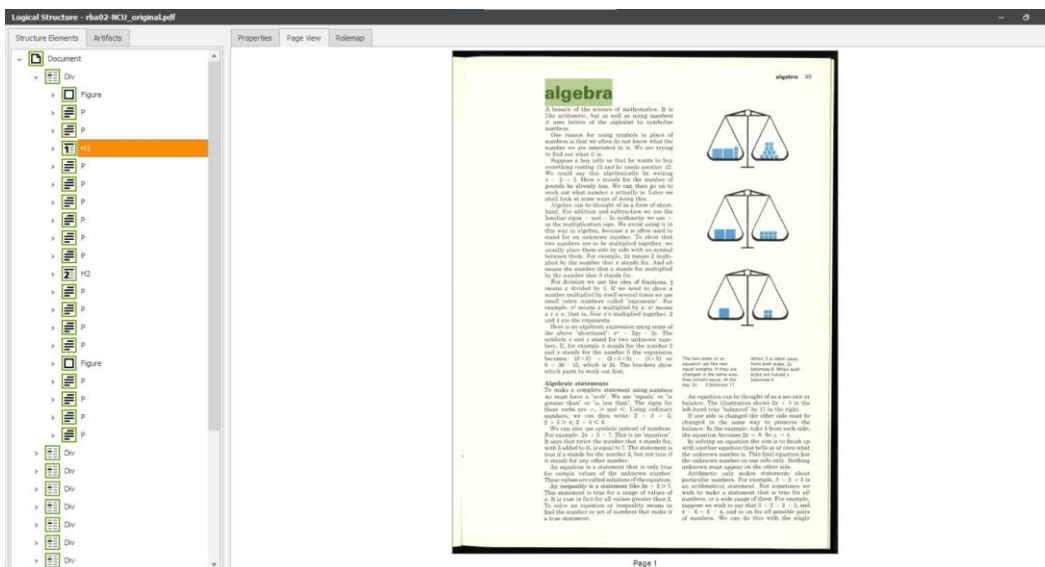


Figure 1: An example of using PDF Accessibility Checker 2021 to evaluate PDF structured tags and text order.

¹⁸ Victor Reader Stratus4 M – Daisy and MP3 player. Available on 20. April 2023 at <https://store.humanware.com/ireland/victor-reader-stratus-4m-daisy-mp3-player.html>


```

170 style="text-align:right;"/>algebra 39</p>
171 <h1
172 id="algebra"><b>algebra</b></h1>
173 <p>A branch of the science of mathematics. It is like arithmetic, but as
174 well as using numbers it uses letters of the alphabet to symbolize
175 numbers.</p>
176 <p>One reason for using symbols in place of numbers is that we often do
177 not know what the number we are interested in is. We are trying to find
178 out what it is.</p>
179 <p>Suppose a boy tells us that he wants to buy something costing
180 <i>£</i>5 and he needs another <i>£</i>2. We could say this
181 algebraically by writing <i>x</i> + 2 = 5. Here <i>x</i> stands for
182 the number of pounds he already has. We can then go on to work out what
183 number <i>x</i> actually is. Later we shall look at some ways of doing
184 this.</p>
185 <p>Algebra can be thought of as a form of shorthand. For addition and
186 subtraction we use the familiar signs + and -. In arithmetic we use × as
187 the multiplication sign. We avoid using it in this way in algebra,
188 because <i>x</i> is often used to stand for an unknown number. To show
189 that two numbers are to be multiplied together, we usually place them
190 side by side with no symbol between them. For example, 2<i>x</i> means
191 2 multiplied by the number that <i>x</i> stands for. And <i>ab</i>
192 means the number that <i>a</i> stands for multiplied by the number
193 that <i>b</i> stands for.</p>
194 <p>For division we use the idea of fractions.
195 <math
196 display="inline" xmlns="http://www.w3.org/1998/Math/MathML"><semantics><math><mi>x</mi></math></semantics></math>
197 encoding="application/x-tex">\frac{x}{3}</annotation></semantics></math>
198 means <i>x</i> divided by 3. If we need to show a number multiplied by
199 itself several times we use small index numbers called 'exponents'. For
200 example, <i>x</i><sup>2</sup> means <i>x</i> multiplied by
201 <i>x</i>. <i>x</i><sup>4</sup> means <i>x</i> <i>x</i>
202 <i>x</i> <i>x</i>; that is, four <i>x</i>'s multiplied together. 2
203 and 4 are the exponents.</p>
204 <p>Here is an algebraic expression using some of the above 'shorthand':
205 <i>x</i>2 + 2<i>xy</i> - 3<i>y</i>. The symbols <i>x</i> and
206 <i>y</i> stand for two unknown numbers. If, for example <i>x</i>
207 stands for the number 3 and <i>y</i> stands for the number 5 the
208 expression becomes: (3×3) + (2×3×5) - (3×5) or 9 + 30 - 15, which is 24.
209 The brackets show which parts to work out first.</p>

```

Figure 2: An example of source code for HTML format. Visible are sections: heading 1, paragraphs and mathematical expression for which MathML was used.



Figure 3: Humanware's Victor Reader Stratus4 M

This is Daisy and MP3 player which was used to evaluate DAISY format audiobook. Source: <https://store.humanware.com/ireland/victor-reader-stratus-4m-daisy-mp3-player.html>

3. Evaluation results

A total of 50 tests were conducted using the conversion services, and 42 results were deemed suitable for evaluation¹⁹. These evaluations only include automatically converted results, without any additional modifications made to the final output formats. The conversion services used for testing were Calibre, RoboBraille, Balabolka, Mathpix, WordToEpub, and EasyConverter Express.

The provided outputs were in the following formats:

- Textual outputs: DOCX (9), EPUB (12), PDF (6), RTF (7), HTML (2).
- Audio outputs: MP3 (11) and DAISY (3).

According to conversion service the formats of the outputs were:

- Calibre: DOCX (4), EPUB (4), PDF (3), RTF (4). We didn't receive 3 of these outputs, all made from PDF/UA.
- RoboBraille: DOCX (2), EPUB (4), PDF (3), RTF (3), MP3 (4) and DAISY (3). We didn't receive 4 of these outputs, 1 from PDF, 2 from PDF/UA and 1 from RTF.
- Balabolka: MP3 (5)
- Mathpix: DOCX (2)
- WordToEpub: EPUB (2), HTML (2) and with EpubToWord: DOCX (1)
- EasyConverter Express: EPUB (2X) and MP3 (2). We didn't receive 1 output from DOCX.

Based on the review of the completed test report questionnaires (see Annexes 2–7), the following summary was derived:

- Each conversion software has different process for converting formats, with some requiring fewer steps and other requiring additional settings.
- Some conversion software required manual intervention prior to conversion (e.g., Balabolka).
- Some conversion tools save output formats on their own servers for a few days, requiring the user to convert the file again if they need it in the same format (e.g., RoboBraille).
- Some conversion tools use their own OCR recognition, resulting in differences from the original content, and it may struggle with complex structures (e.g., Mathpix's Snip).
- Audio outputs formats may present challenges regarding the use of commercially available voices by libraries.
- Most of the outputs were received quickly.

¹⁹ Further evaluation shows also the missing outcomes but all criteria were evaluated as not achieved.

3.1 Results according to conversion service used

The testing results for each conversion service and accessibility criteria for each output format are presented in the tables below. The results are classified as fully achieved, partially achieved, or not achieved, and additional observations are also noted. For easier comprehension, the top results for each criterion are marked in bold text and fully achieved results are further described in Chapter 2.2 Results according to criteria.

3.1.1 Calibre

Table 2: Textual outputs made by Calibre

TEXTUAL OUTPUTS																
Input format	Ref. no.	PDF	PDF/UA*	RTF	EPUB	PDF	PDF/UA*	RTF	DOCX	RTF	DOCX	EPUB	PDF	PDF/UA*	DOCX	EPUB
Output format		DOCX	DOCX ²⁰	DOCX	DOCX	EPUB	EPUB ²¹	EPUB	EPUB	PDF	PDF	PDF	RTF	RTF ²²	RTF	RTF
Version		2007+	2007+	2007+	2007+	2.0	2.0	2.0	2.0	1.4	1.4	1.4	/	/	/	/
Alt-text picture	18			11B	18A			11B	18A	11B						
Alt-text chemical formula	2			2B	2A			2B	2A	2B						
Caption	19	1B		14B	19A	1B		14B		14B		19A	1B			17A 2B
Footnotes	1				1A							1A				
Heading 1	7				7A				6A						6A	7A
Heading 2	10				10A				8A						8A	10A
Heading 3	1				1A											1A
Initial	1	1A		1A	1A	1A		1A	1A	1A	1A	1A	1A		1A	1A
Language segment	6				5A 1B				3B							
Math. (simple)	3	1A		3A	1A	1A		3A	2A	3A	2A	3A	2B		2A	1A
Math. (adv.)	4			4A				4A	2A	4A	2A	4A			2A	
OCR errors	1			1A	1A			1A	1A	1A	1A	1A			1A	1A
Page rotation	1			1A	1A			1A	1A	1A	1A	1A			1A	1A
Pagination	12	8A		10A 1B	12A	8A		10A 1B	9A	10A 1B	9A	12A	6A 1B		9A	12A
Pagination double	2	1A		2A	2A	1A		2A	2A	2A	2A	2A			2A	2A
Picture	18				18A				18A		18A	18A			18A	18A
Picture chem. formula	2				2A				2A		2A	2A			2A	2A
Primary language	1	1A		1A	1A			1A	1A				1A		1A	1A
Special character	3	1A		1A	2A 1B	1A		1A	2A	1A	2A	2A 1B	2A		2A	1A 1B
Stamp removal	1			1A	1A			1A	1A	1A	1A	1A			1A	1A
Table	4			4A	4A			4A	4A	4B	4B	4B				
Table header	4				4A							4B				
Table rows	4			4A	4A			4A	4A	4B	4B	4B				
Text order	16	4A		8A 8B	15A 1B	4A		8A 8B	13A 2B	8A 8B	12A 3B	15A 1B	4A		11A 3B	15A 1B
Personalisation	1			1A	1A			1A	1A						1A	1A

Used codes: A = fully achieved criterion; B = partly achieved criterion; empty cell = criterion was not achieved; * = tagged PDF

²⁰ We could not evaluate this format because it contained only images of the scans. All criteria are not achieved.

²¹ Same as footnote 14.

²² Same as footnote 14.

With Calibre we were unable to produce audio output formats, so this aspect is not evaluated by this conversion service.

Additional observations for textual outputs (Table 2):

1. DOCX as output format:

- **EPUB to DOCX:** The first page is empty, and the table of contents (TOC) at the beginning appears twice. The first TOC is derived from list of files in the EPUB, and second is the actual TOC from the input file, containing working hyperlinks to original pages. Alt-text is present. Mathematical content appears in both linear writing and additionally in LaTeX (EPUB had MathML), which can be confusing for screen readers. In scans 2, 8, 10, 11, 12 and 16, there are extra heading 1 elements. In scan 6, a language segment appears in Latin but not in Italian. In scan 12, footnotes are transformed into endnotes and lack hyperlinks to work in both directions. In scan 15, hyperlinks in the TOC function correctly.
- **PDF to DOCX:** Each word is on a new line, and there are headings 1 and 2, but they are incorrect.
- **RTF to DOCX:** The TOC at beginning appears twice, which is working but not useful to have it duplicated. Table headers are present, but there are also column headers. In scan 12, footnotes have been converted into endnotes and lack hyperlinks to work in both directions. In scan 16, the table is not entirely correct.

2. EPUB as output format:

- **DOCX to EPUB:** The first element appearing is the image from scan 3. The sidebar TOC is present and layered. Mathematical content is challenging to read as some signs are missing. MathML was not converted, resulting in the loss of all MathML content.
- **PDF to EPUB:** Each word appears on a new line, and there is no TOC.
- **RTF to EPUB:** The TOC is present at the beginning, but not all hyperlinks function correctly. The links that do work are also found in the sidebar TOC. The HTML code is messy. Footnotes have been transformed into endnotes, and mathematical content is not rendered correctly.

3. PDF as output format:

- **DOCX to PDF:** The resulting PDF is not tagged. The first element to appear is the image from scan 3. Text exhibits varying alignments. In scans 10, 11, and 16, table rows seem fine visually, but without tags, it's challenging to verify their correctness. Mathematical content is missing because MathML was not converted. Footnotes do not work. The TOC is generated as bookmarks, but their structure cannot be confirmed without tags. Adobe presented issues when using a screen reader, making it difficult to detect some elements.
- **EPUB to PDF:** The PDF is not tagged. The first page is empty. At the beginning, there is a TOC with functioning hyperlinks, and a TOC is present as sidebar bookmarks where chapter levels appear correct. However, heading levels cannot be determined as there are no tags. Checking alt-text is problematic due to Adobe-related screen reader issues. Captions are marked as figures, which is acceptable. Tables appear to have table headers

visually, but there are no tags to confirm this. Footnotes work. In scan 15, the original fake hyperlinks for the TOC remain, and they function correctly.

RTF to PDF: The PDF is not tagged. The TOC is generated as bookmarks but is not entirely correct as some levels are missing. Not all hyperlinks in the TOC at the beginning are functional. The document has a clean appearance with strong contrasts. In scans 10 and 11, tables are present, but their structure cannot be discerned due to the absence of tags. Mathematical content is partially accepted. Footnotes lack hyperlinks in both directions and appear as endnotes rather than footnotes, which could be improved by placing them on the same page where they are referenced. Scan 15 is not recognized as a table.

4. RTF as output format

- **DOCX to RTF:** The pagination is on the right side. Not all advanced math content is correct. In scan 12, math is missing, and footnotes are not functioning.
- **EPUB to RTF:** The first page is empty. The TOC at the beginning includes a list of all pages and landmarks with non-functional hyperlinks, which is problematic and should be removed. The sidebar TOC shows additional empty headings after each heading. All MathML content is missing and was not converted. There is no alt-text from the original. In scans 2, 8, 10, 12, and 16, there are extra heading 1 elements. In scan 5, a picture is too large, and not all of it appears on the screen. In scan 10, the table is not present, making it confusing to read with a screen reader. In scan 15, there are no hyperlinks in the TOC that existed in the input format. In scan 16, the table is not present.
- **PDF to RTF:** Each word appears on a separate line, which is problematic. The image of the scan appears first, followed by the text (OCR). Heading levels are present but incorrectly set. In scans 1, 2, 3, 4, 5, and 6, text is wrongly recognized through rows instead of columns. In scan 11, text in tables is missing. In scans 15 and 16, there are only two pages that do not have one word per row, and in scan 16, no text from the table is recognized, resulting in missing content.

3.1.2 RoboBraille

Table 3: Textual outputs made by RoboBraille

TEXTUAL OUTPUTS													
Input format	Ref. no.	PDF	PDF/UA*	PDF	PDF/UA*	RTF	DOCX	PDF	PDF/UA*	DOCX	PDF	PDF/UA*	EPUB
Output format		DOCX	DOCX	EPUB	EPUB ²³	EPUB	EPUB	PDF tagged image over text*	PDF tagged image over text*	PDF*	RTF	RTF	RTF
Version		2007 +	2007+	2.0	2.0	2.0	2.0	1.5	1.5	1.5	1.5-1.6	1.5-1.6	/
Alt-text picture	18			13B		11B	18A	11B		18A			
Alt-text chemical formula	2			2B		2B	2A	2B		2A			
Caption	19	1B		1B		14B		2A 6B	6A		1B		15A
Footnotes	1					1A							
Heading 1	7						6A	2A	2A	6A			6A
Heading 2	10			4A			8A	1A	4A	8A			8A
Heading 3	1												1A
Initial	1					1A	1A			1A			1A
Language segment	6	2A	1A 3B			2B	3B		1B	3B	2A	1A 3B	
Math. (simple)	3	1A 2B	3A	1A 1B		3A	3A	1A	2A	3A	1A 2B	2A 1B	1A
Math. (adv.)	4	1B	2B			4A	2A			3A 1B	1B	1A	
OCR errors	1					1A	1A			1A			1A
Page rotation	1		1A			1A	1A		1A	1A		1A	1A
Pagination	12		1A	6A 2B		10A 1B	9A	7A	2A 1B	9A		1A	9A
Pagination double	2	1A	2A			2A	2A		2A	2A	1A	2A	2A
Picture	18	10A 1B	11A	8A 4B			18A	11A	10A	18A	11A 1B	11A	17A
Picture chem. formula	2	2B		1A 1B			2A	2A		2A	2B		2A
Primary language	1	1A	1A				1A	1A	1A		1A	1A	1A
Special character	3	2A	2A	2A		2A	2A	2A	2A	2A	2A	2A	1A 1B
Stamp removal	1	1B				1A	1A			1A	1B		1A
Table	4	1A	4A			4A	4A	1A	4A	4A	1A	4A	
Table header	4							1A	4A				
Table rows	4		1B			4A	4A	1B	1A 3B	4A		1A 3B	
Text order	16	4A 1B	4A 3B	6A		8A 8B	13A 2B	7A	4A 3B	13A 2B	4A 1B	3A 5B	12A 1B
Personalisation	1					1A	1A						1A

Used codes: A = fully achieved criterion; B = partly achieved criterion; empty cell = criterion was not achieved; * = tagged PDF

Additional observations for textual outputs (Table 3):

1. DOCX as output format:

- PDF to DOCX:** The visual appearance remains consistent, with all elements in the same position. However, on a page that is only 22x28 cm in size, it appears crowded, and some text is not visible. In certain areas, text is densely layered, making it challenging to read. Text-boxes present a problem, as they are not suitable for screen readers. Mathematical content is not entirely correct, and there are issues with the use of super- and subscripts. In scan 3, a stamp has been removed, and almost all text is correct. In scan 10, a table is treated as an image. Similar problems exist with PDF to RTF conversions.

²³ We could not evaluate this format because it contained only images of the scans. All criteria are not achieved.

- **PDF/UA to DOCX:** Text-boxes are present, which can be problematic for screen readers. Mathematical content is only partially accurate. In scan 6, text order follows lines rather than columns. In scan 8, not all text is displayed. In scan 10, a table has one extra column.
- 2. EPUB as output format:**
- **DOCX to EPUB:** This format contains an in-depth and adjustable TOC. MathML is missing and has not been converted.
 - **PDF to EPUB:** For each scan, the image of the scan appears first, followed by the text, which is not acceptable from an accessibility standpoint. There is no sidebar TOC. Alt-text is present but is labeled as "image 3," which is not acceptable. The HTML is decent, but there are too many HTML files (89), making navigation complex. The lines of the text are divided excessively, resulting in significant text breaks. Text in two columns is recognized in straight lines, leading to a messier presentation than the original input format. In scan 10, a table is presented as an image. In scan 11, the rows in the table are not formatted correctly and do not appear as a table.
 - **RTF to EPUB:** The TOC at the beginning is present, and hyperlinks are working. There is also a detailed sidebar TOC. The document is divided into 16 HTML files, one for each scan, which is an ideal setup and allows for easy potential editing. However, the text can be messy in terms of language, with a mix of Danish and some paragraphs in English or German. Footnotes function in both directions, but mathematical content is not read correctly.
- 3. PDF tagged image over text and PDF as output format:**
- **PDF to PDF text over image:** It appears as though OCR has been performed again, resulting in the recognition of many additional small dots or specks. Furthermore, the order of elements is different.
 - **PDF/UA to PDF tagged image over text:** In scan 1 and 3, captions disrupt the order of the text. In scan 4, an attempt has been made to present chemical elements as a table, but it is not well-structured. Scan 8 shows the header as an image. In scan 10, rows of a table are not recognized correctly, and an extra column is added at the end. In scan 11, the page number is marked as a caption, and the rows of the table are not correctly recognized. In scan 13, the page appears twice, with the second instance rotated 90° to the right. In scan 14, the second column is tagged as a caption for the image of the entire page. There are no TOC sidebar bookmarks for navigation. Each page has a separate tree tag structure, which is a positive aspect.
- 4. RTF as output format:**
- **PDF/UA to RTF:** The presence of background images can be disturbing. There are occasional instances of Slovak language appearing. Text colors used for contrast can be problematic, as text in some places appears faintly grey. The appearance of text-boxes poses issues for screen readers. Captions are missing. RTF doesn't support inserting equations in Word, so mathematical content isn't fully correct. In scan 3, a stamp was removed, but the text is also missing. Scan 4 presents chemical elements as a table. In scan 6, the text does not follow the correct order. In scan 8, the page header is marked as

an image. In scan 10, there is an unnecessary additional column in the table. Scan 11 contains mistakes in table rows. Scan 12 includes pagination in the header of the document. Scans 13 and 15 have missing text. In scan 14, it is difficult to determine the reading order.

- **EPUB to RTF:** The first page is empty. The TOC at the beginning remains in XML/HTML code, which is not user-friendly. The sidebar TOC reveals additional empty headings that appear after each heading. MathML is missing and was not converted. In scan 4, the image is too large. In scan 10, there is no table, which is confusing for screen readers. In scan 11, the page appears in XML/HTML code, which can be disturbing. Hyperlinks that were present in the input file are missing. In scan 16, the page also appears in XML/HTML code, which is disruptive.

Table 4: Audio outputs made by RoboBraille

AUDIO OUTPUTS								
Input format	Ref. no.	PDF	PDF/UA*	RTF	DOCX	RTF	DOCX	DOCX
Output format		MP3 ²⁴	MP3 ²⁵	MP3	MP3	DAISY full text and audio ²⁶	DAISY full text and audio	DAISY math full text and audio
Version		/	/	/	/	/	2.02	2.02
Alt-text picture	18			10A 1B			18A	18A
Alt-text chemical formula	2			2A			2A	2A
Caption	19			14A	1B		1B	1B
Footnotes	1							
Heading 1	7						6B	6B
Heading 2	10						10B	10B
Heading 3	1							
Initial	1			1A	1A		1A	1A
Language segment	6						1B	3B
Math. (simple)	3			1B	1B		2A	2A
Math. (adv.)	4			3B	3B		1A 3B	1A 3B
OCR errors	1			1A	1A		1A	1A
Page rotation	1			1A	1A		1A	1A
Pagination	12			10A 1B	9A		9A	9A
Pagination double	2			2A	2A		2A	2A
Picture	18							
Picture chem. formula	2							
Primary language	1			1A	1A		1A	1A
Special character	3			1A 1B	1A 1B		2A	2A
Stamp removal	1			1A	1A		1A	1A
Table	4							
Table header	4							
Table rows	4							
Text order	16			9A 7B	13A 2B		13A 2B	12A 1B
Personalisation	1							

Used codes: A = fully achieved criterion; B = partly achieved criterion; empty cell = criterion was not achieved; * = tagged PDF

²⁴ We could not evaluate this format because no results came back from RoboBraille. All criteria are not achieved.

²⁵ Same.

²⁶ Same.

Additional observations for audio outputs (Table 4):

1. MP3 as output format:

- **RTF to MP3:** The entire TOC and pages were read. The marker "orig. page" sounds unclear, and it is suggested to use the whole name "original page" or only "page" for clarity. In mathematical content, the minus sign is not read, and "<" and ">" signs were not consistently read. Greek letters were also not read aloud. In scan 4, the table is confusing to listen to, with descriptions like "5 B 10 81 6 C 12 01," etc. In scan 12, footnotes are not recorded, and the degree symbol "°C" wasn't pronounced.

2. DAISY as output format:

- **DOCX to DAISY full text and audio and DAISY math full text and audio:** It's possible to navigate between marked chapters. Simple math signs (=, +, -, x, <) are read, with the exception of the ">" sign, which is pronounced as "mgt." Where MathML was used in the input format, the math is read as "mathematical expression x over 3 end of mathematical expression" – which is the best way to listen to math. There are no issues with Greek letters (e.g., $\Delta\rho = p_v g h$ is pronounced: "increment rho equals p sub v g h"²⁷). **Advanced math content is better compared to other cases.** Alt-text is marked with a comma before the text and is also pronounced. The audiobook has two landmarks: the beginning and end of the book. There are small pauses around each new chapter. The title of the book is "NUK," which may not be suitable. The only significant difference between the two output formats is that DAISY Math Full Text and Audio also **contain the best language segments in French and German, where the voice shifts to the mentioned languages.** However, in scans 1 and 3, part of the text is repeated.

²⁷ Another example: $h = \frac{\Delta\rho}{p_v g} = \frac{2T}{R(p_w - p_v)g}$ -> »h equals fraction increment rho over p sub v g end fraction equals fraction 2 cap t over cap r open p sub w minus p sub v close g end fraction«

3.1.3 Balabolka

Table 5: Audio outputs made by Balabolka

AUDIO OUTPUTS						
Input format	Ref. no.	PDF	PDF/UA	RTF	DOCX	EPUB
Output format		MP3	MP3	MP3	MP3	MP3
Version		/	/	/	/	/
Alt-text picture	18			10A 1B		
Alt-text chemical formula	2			2A		
Caption	19	1B	1B	11A	1B	19A
Footnotes	1					
Heading 1	7	7B	7B	7B	7B	7B
Heading 2	10					
Heading 3	1					
Initial	1	1A		1A	1A	1A
Language segment	6					
Math. (simple)	3	1A	1A	1A	1A	1A
Math. (adv.)	4	3B	2B	4B	3B	4B
OCR errors	1			1A	1A	1A
Page rotation	1		1A	1A	1A	1A
Pagination	12	9A	8A	10A 1B	9A	12A
Pagination double	2	1A	2A	2A	2A	2A
Picture	18					
Picture chem. formula	2					
Primary language	1	1A	1A	1A	1A	1A
Special character	3	2A 1B	2A	2A	2A	2A 1B
Stamp removal	1			1A	1A	1A
Table	4					
Table header	4					
Table rows	4					
Text order	16	7A 1B	9A 3B	9A 7B	13A 2B	15A 1B
Personalisation	1					

Used codes: A = fully achieved criterion; B = partly achieved criterion; empty cell = criterion was not achieved; * = tagged PDF

Additional observations for audio outputs (Table 5):

1. MP3 as output format:

- **PDF to MP3:** The audiobook is divided into 16 separate files, each corresponding to one of the defined scans. However, the reading is paused after each line of text, resulting in constant interruptions, which can be challenging for the listener. In mathematical content, the minus sign is either missing or pronounced as "dash," and the symbols "<" and ">" also presented pronunciation issues. The "-O-" was pronounced as "tired face," possibly related to emojis. Text from tables is missing.
- **PDF/UA to MP3:** Similar to the PDF to MP3 conversion, the audiobook is split into 16 separate files, following the defined scan settings. However, the reading is paused slightly after each line of text, which can be difficult to listen to, especially in cases like scan 7, where text lines are short (2-3 words). The minus sign is pronounced as "dash" or "to," and the symbols "<" and ">" were not consistently pronounced. The "-O-" was pronounced as "tired face," possibly related to emojis. The mathematical expression

$C_6H_{12}O_6$ was not pronounced at all. There is also an issue with distinguishing between empty table cells and cells with data.

- **EPUB to MP3:** The audiobook is divided into 17 separate files, each corresponding to one of the defined scans, with an extra file for the TOC. Page numbers in the TOC are not pronounced, which can affect navigation. The first file lists all landmarks and pages, which is lengthy and possibly unnecessary. An alternative solution may be considered. Alt-text from the original input file is missing and requires manual insertion. Mathematical content in MathML is read twice: once as written in linear form and secondly in LaTeX format. For example, it says "frac," which is an abbreviation for "\frac" for fractions.²⁸ Superscript is pronounced as "caret", which is derived from LaTeX, as it uses the "^" symbol for superscripts. (e.g., x^2 is written as x^2 or $x\{2\}$). Similarly, "backslash" is pronounced because it is used in LaTeX. For subscript, it is pronounced as "underscore" because LaTeX uses the "_" symbol. For example, $C_6H_{12}O_6$ is pronounced as "C underscore 6 H underscore 12 O underscore 6," while in LaTeX, it would be "C_6H_12O_6." The "-O-" is pronounced as "tired face," which might be related to emojis. There is a good solution for empty table cells, where the slash is pronounced. Greek letters like delta and rho are pronounced correctly. Footnotes do not work.
- **RTF to MP3:** The audiobook is divided into 17 separate files, with an extra file for the TOC. The TOC does not cause problems, but it may be unclear which level of the heading is being read. The "orig. page" marker sounds unclear, and it is suggested to use the whole name "original page" or only "page". The text flows fluidly without breaks, which is different from PDFs. In mathematical content, the minus sign is either missing or pronounced as "dash" or "to," and the signs "<" and ">" also present pronunciation problems. Footnotes are missing. There are no problems with Greek letters delta and rho.

²⁸ Example of math: $h = \frac{\mathrm{\Delta}p}{\rho_{\{v\}}g} = \frac{2T}{R(\rho_{\{w\}} - \rho_{\{v\}})g}$

3.1.4 Mathpix

Table 6: Textual outputs made by Mathpix

TEXTUAL OUTPUTS			
Input format	Ref. no	PDF	PDF/UA
Output format		DOCX²⁹	DOCX³⁰
Version		2007+	2007+
Alt-text picture	18		
Alt-text chemical formula	2		
Caption	19	1B	
Footnotes	1		
Heading 1	7		
Heading 2	10		
Heading 3	1		
Initial	1		
Language segment	6		
Math. (simple)	3	3A	2A
Math. (adv.)	4	3A 1B	3A
OCR errors	1		
Page rotation	1		1A
Pagination	12		
Pagination double	2		
Picture	18	11A	7A
Picture chem. formula	2	2A	
Primary language	1	1A	1A
Special character	3	2A 1B	3A
Stamp removal	1	1A	
Table	4		
Table header	4		
Table rows	4		
Text order	16	4A 2B	2A 4B
Personalisation	1	1A	1A

Used codes: A = fully achieved criterion; B = partly achieved criterion; empty cell = criterion was not achieved; * = tagged PDF

Additional observations for textual outputs (Table 6):

1. **DOCX as output format:**

- **PDF to DOCX:** The entire document maintains a consistent and visually clean presentation. However, there are occasional instances where the text order becomes messy. All mathematical content is in MathML, although on scan 2, some math is missing. Tables are also missing, and in scan 7, the second page is missing. Scan 10 is completely missing.
- **PDF/UA to DOCX:** The document contains all mathematical content in MathML and is overall very clean. However, there are issues with missing pages in scans 2, 3, and 7. In scan 6, not all text is recognized or displayed as a table, and columns are not recognized.

²⁹ Some pages are missing in the output format. Those pages were evaluated as “not achieved” criteria.

³⁰ Same.

Scans 5, 9, 12, and 15 also have instances of missing text. Scan 10, 11, and 16 present tables as pictures, which is not acceptable.

3.1.5 WordToEpub

Table 7: Textual outputs made by WordToEpub

TEXTUAL OUTPUTS						
Input format	Ref. no.	RTF	DOCX	RTF	DOCX	EPUB
Output format		EPUB	EPUB	HTML	HTML	DOCX
Version		3.0	3.0	5	1.0, 5	2007+
Alt-text picture	18	11B	18A	11B	18A	18A
Alt-text chemical formula	2	2B	2A	2B	2A	2A
Caption	19	14B		14B		19A
Footnotes	1	1A		1A		
Heading 1	7		6A		6A	7A
Heading 2	10		8A		8A	10A
Heading 3	1					1A
Initial	1	1A	1A	1A	1A	1A
Language segment	6	1B	3B	1B	3B	4A 2B
Math. (simple)	3	3A	3A	3A	3A	3A
Math. (adv.)	4	4A	3A 1B	4A	3A 1B	4A
OCR errors	1	1A	1A	1A	1A	1A
Page rotation	1	1A	1A	1A	1A	1A
Pagination	12	10A 1B	9A	10A 1B	9A	12A
Pagination double	2	2A	2A	2A	2A	2A
Picture	18		18A		18A	18A
Picture chem. formula	2		2A		2A	2A
Primary language	1	1A	1A	1A	1A	1A
Special character	3	2A	2A	2A	2A	2A 1B
Stamp removal	1	1A	1A	1A	1A	1A
Table	4	4A	4A	4A	4A	4A
Table header	4		1A			4A
Table rows	4	4A	4A	4A	4A	4A
Text order	16	8A 8B	13A 2B	8A 8B	13A 2B	16A
Personalisation	1	1A	1A	1B	1B	1A

Used codes: A = fully achieved criterion; B = partly achieved criterion; empty cell = criterion was not achieved; * = tagged PDF

Additional observations for textual outputs (Table 7):

1. EPUB as output format

- **DOCX to EPUB:** An additional TOC is present at the start due to an error during conversion. The sidebar TOC is enabled, containing two landmarks and enabling Word document page navigation since original page navigation was not possible. However, there is a problem with the HTML files being broken into too many separate files due to incorrectly marked headings.
- **RTF (DOCX) to EPUB:** The TOC is kept from the original document, hyperlinks work, but navigation through reading software is not possible as there is no sidebar TOC. In scan 4, a table is very large, but everything can be accessed.

2. HTML as output format

- **DOCX to HTML:** The document includes an in-depth TOC, and hyperlinks are functioning correctly. Some text, particularly pagination, is right-aligned, while other text and all pictures are center-aligned. Mathematical content looks great, with MathML in the background. However, footnotes are not working.

3. DOCX as output format

- **EPUB to DOCX:** The TOC at the beginning is not complete, and hyperlinks for the original page and landmarks are not working. Headings are marked, allowing navigation through the sidebar. In scan 6, Swiss German is marked instead of the Latin language segment. MathML is present.

3.1.6 EasyConverter Express

Table 8: Textual and audio outputs made by EasyConverter Express

		TEXTUAL OUTPUTS		AUDIO OUTPUTS	
Input format	Ref. no.	RTF	DOCX	RTF	DOCX
Output format		EPUB	EPUB	MP3	MP3 ³¹
Version		3.0	3.0	/	/
Alt-text picture	18	11B	18A	10A 1B	
Alt-text chemical formula	2	2B	2A	2A	
Caption	19	11B		11A	
Footnotes	1	1A			
Heading 1	7		6A		
Heading 2	10		8A		
Heading 3	1				
Initial	1	1A	1A	1A	
Language segment	6	2B	3B		
Math. (simple)	3	3A	3A	1B	
Math. (adv.)	4	4A	3A 1B	4B	
OCR errors	1	1A	1A	1A	
Page rotation	1	1A	1A	1A	
Pagination	12	10A 1B	9A	10A 1B	
Pagination double	2	2A	2A	2A	
Picture	18		18A		
Picture chem. formula	2		2A		
Primary language	1			1A	
Special character	3	2A	2A	2A	
Stamp removal	1	1A	1A	1A	
Table	4	4A	4A		
Table header	4				
Table rows	4	4A	4A		
Text order	16	9A 7B	13A 2B	9A 7B	
Personalisation	1	1A	1A		

Used codes: A = fully achieved criterion; B = partly achieved criterion; empty cell = criterion was not achieved; * = tagged PDF

³¹ We could not evaluate this format because the software stopped at 10%. All criteria are not achieved.

Additional observations for audio outputs (Table 8):

1. MP3 as output format

- **RTF to MP3:** The conversion service split the audio into two separate files, with the split occurring at scan 4. There is no TOC in the audio, and footnotes are located at the end of the MP3 file. It is not clear which cells in the table are empty.

General observation for all audio outputs:

MP3 Format:

- MP3 audio has difficulties with empty cells in tables, making it unclear which cells are empty.
- The minus sign is often overlooked and pronounced as "dash" or "to."
- The symbols "<" and ">" are often not pronounced, particularly as standalone characters, although sometimes they are pronounced within expressions like " $2x+3>7$."
- Sub- and superscripts are mostly read as, for example, "H two" or "x four," without specifying whether they are subscripts or superscripts. Only DAISY formats provide an option to clarify when something is a subscript. Similarly, only DAISY clearly indicates where fractions, brackets, etc., begin and end. DAISY also handles Greek letters and other symbols without issues.
- In comparison, MP3 is more suitable for literature and less suitable for scientific content. For scientific materials, creating DAISY files with Robobrainle may be a more appropriate choice.

3.2 Results according to established criteria

Tables 2-8 present the best results based on the established criteria for each of the tested conversion services. To enhance clarity, the fully achieved results for each criterion are further described ([Figure 4](#)).

- **ALT-TEXT PICTURE (18)**

10 outputs fully met this criterion:

- Calibre: DOCX converted from EPUB and EPUB from DOCX
- RoboBraille: EPUB converted from DOCX, PDF from DOCX, DAISY full text and audio from DOCX, and DAISY math full text and audio from DOCX
- WordToEpub: EPUB converted from DOCX, HTML from DOCX, and DOCX from EPUB
- EasyConverter Express: EPUB converted from DOCX

In all cases, the input formats included additional manual corrections.

- **ALT-TEXT PICTURE (CHEM. FORMULA) (2)**

13 outputs fully met this criterion:

- Calibre: DOCX converted from EPUB and EPUB from DOCX

- RoboBraille: EPUB converted from DOCX, PDF from DOCX, MP3 from RTF, DAISY full text and audio from DOCX, and DAISY math full text and audio from DOCX
- Balabolka: MP3 converted from RTF
- WordToEpub: EPUB converted from DOCX, HTML from DOCX, and DOCX from EPUB
- EasyConverter Express: EPUB converted from DOCX and MP3 from RTF

In all cases, the input formats included additional manual corrections.

- **CAPTION (19)**

4 outputs fully met this criterion:

- Calibre: DOCX converted from EPUB and PDF from EPUB
- Balabolka: MP3 converted from EPUB
- WordToEpub: DOCX from EPUB

In all cases, the input format was EPUB, which included additional manual corrections.

- **FOOTNOTES (1)**

6 outputs fully met this criterion:

- Calibre: DOCX converted from EPUB and PDF from EPUB
- RoboBraille: EPUB converted from RTF
- WordToEpub: EPUB converted from RTF, HTML from RTF
- EasyConverter Express: EPUB converted from RTF

In all cases, the input formats included additional manual corrections.

- **HEADING 1 (7)**

3 outputs fully met this criterion:

- Calibre: DOCX converted from EPUB and RTF from EPUB
- WordToEpub: DOCX converted from EPUB

In all cases, the input format was EPUB, which included additional manual corrections.

- **HEADING 2 (10)**

3 outputs fully met this criterion:

- Calibre: DOCX converted from EPUB and RTF from EPUB
- WordToEpub: DOCX converted from EPUB

In all cases, the input format was EPUB, which included additional manual corrections.

- **HEADING 3 (1)**

4 outputs fully met this criterion:

- Calibre: DOCX converted from EPUB and RTF from EPUB
- RoboBraille: RTF converted from EPUB
- WordToEpub: DOCX converted from EPUB

In all cases, the input format was EPUB, which included additional manual corrections.

- **INITIAL (1)**

32 outputs fully met this criterion:

- Calibre: DOCX converted from PDF, RTF, and EPUB; EPUB converted from PDF, RTF, and DOCX; PDF converted from RTF, DOCX, and EPUB; RTF converted from PDF, DOCX, and EPUB
- RoboBraille: EPUB converted from RTF and DOCX; PDF converted from DOCX; RTF converted from EPUB; MP3 converted from RTF and DOCX; DAISY full text and audio from DOCX; DAISY math full text and audio from DOCX
- Balabolka: MP3 converted from PDF, RTF, DOCX, and EPUB
- WordToEpub: EPUB converted from RTF and DOCX; HTML converted from RTF and DOCX; DOCX converted from EPUB
- EasyConverter Express: EPUB converted from RTF and DOCX; MP3 converted from RTF

Most input formats included additional manual corrections, although some results were fully achieved with non-tagged PDF as the input format.

- **LANGUAGE SEGMENTS (6)**

0 outputs fully met this criterion.

- **MATH (SIMPLE) (3)**

16 outputs fully met this criterion:

- Calibre: DOCX converted from RTF; EPUB converted from RTF; PDF converted from RTF and from EPUB
- RoboBraille: DOCX converted from PDF/UA; EPUB converted from RTF and DOCX; PDF converted from DOCX
- Mathpix: DOCX converted from PDF
- WordToEpub: EPUB converted from RTF and DOCX; HTML converted from RTF and DOCX; DOCX converted from EPUB
- EasyConverter Express: EPUB converted from RTF and DOCX

Most input formats included additional manual corrections, although some results were fully achieved with tagged PDF/UA as the input format.

- **MATH (ADVANCED) (4)**

9 outputs fully met this criterion:

- Calibre: DOCX converted from RTF; EPUB converted from RTF; PDF converted from RTF and EPUB
- RoboBraille: EPUB converted from RTF
- WordToEpub: EPUB converted from RTF; HTML converted from RTF; DOCX converted from EPUB
- EasyConverter Express: EPUB converted from RTF

All input formats included additional manual corrections.

- **OCR ERRORS (1)**

28 outputs fully met this criterion:

- Calibre: DOCX converted from RTF and EPUB; EPUB converted from RTF and DOCX; PDF converted from RTF, DOCX and EPUB; RTF converted from DOCX and EPUB
- RoboBraille: EPUB converted from RTF and DOCX; PDF converted from DOCX; RTF converted from EPUB; MP3 converted from RTF and DOCX; DAISY full text and audio converted from DOCX; DAISY math full text and audio converted from DOCX
- Balabolka: MP3 converted from RTF, DOCX and EPUB
- WordToEpub: EPUB converted from RTF and DOCX; HTML converted from RTF and DOCX; DOCX converted from EPUB
- EasyConverter Express: EPUB converted from RTF and DOCX; MP3 converted from RTF

All input formats included additional manual corrections.

- **PAGE ROTATION (1)**

33 outputs fully met this criterion:

- Calibre: DOCX converted from RTF and EPUB; EPUB converted from RTF and DOCX; PDF converted from RTF, DOCX and EPUB; RTF converted from DOCX and EPUB
- RoboBraille: DOCX converted from PDF/UA; EPUB converted from RTF and DOCX; PDF tagged image over text converted from PDF/UA; PDF converted from DOCX; RTF converted from PDF/UA and EPUB; MP3 converted from RTF and DOCX; DAISY full text and audio converted from DOCX; DAISY math full text and audio converted from DOCX
- Balabolka: MP3 converted from PDF/UA, RTF, DOCX and EPUB
- Mathpix: DOCX converted from PDF/UA
- WordToEpub: EPUB converted from RTF and DOCX; HTML converted from RTF and DOCX; DOCX converted from EPUB
- EasyConverter Express: EPUB converted from RTF and DOCX; MP3 converted from RTF

Most input formats included additional manual corrections; however, some results fully met the criterion when using tagged PDF/UA as the input format.

- **PAGINATION (12)**

5 outputs fully met this criterion:

- Calibre: DOCX converted from EPUB; PDF converted from EPUB; RTF converted from EPUB
- Balabolka: MP3 converted from EPUB
- WordToEpub: DOCX converted from EPUB

In all cases, the input format was EPUB, which contained additional manual corrections.

- **PAGINATION-DOUBLE (2)**

32 outputs fully met this criterion:

- Calibre: DOCX converted from RTF and EPUB; EPUB converted from RTF and DOCX; PDF converted from RTF, DOCX, and EPUB; RTF converted from DOCX and EPUB

- RoboBraille: DOCX converted from PDF/UA; EPUB converted from RTF and DOCX; PDF tagged image over text from PDF/UA; PDF converted from DOCX; RTF converted from PDF/UA and EPUB; MP3 converted from RTF and DOCX; DAISY full text and audio converted from DOCX; DAISY math full text and audio converted from DOCX
 - Balabolka: MP3 converted from PDF/UA, RTF, DOCX and EPUB
 - WordToEpub: EPUB converted from RTF and DOCX; HTML converted from RTF and DOCX; DOCX converted from EPUB
 - EasyConverter Express: EPUB converted from RTF and DOCX; MP3 converted from RTF
- Most input formats included additional manual corrections; however, some results fully met the criterion when using tagged PDF/UA as the input format.

- **PERSONALISATION (1)**

16 outputs fully met this criterion:

- Calibre: DOCX converted from RTF and EPUB; EPUB converted from RTF and DOCX; RTF converted from DOCX and EPUB
- RoboBraille: EPUB converted from RTF and DOCX; RTF converted from EPUB
- Mathpix: DOCX converted from PDF and PDF/UA
- WordToEpub: EPUB converted from RTF and DOCX; DOCX converted from EPUB
- EasyConverter Express: EPUB converted from RTF and DOCX

Most input formats included additional manual corrections; however, some results fully met the criterion when using untagged PDF and tagged PDF/UA as the input format.

- **PICTURE (18)**

12 outputs fully met this criterion:

- Calibre: DOCX converted from EPUB; EPUB converted from DOCX; PDF converted from DOCX and EPUB; RTF converted from DOCX and EPUB
- RoboBraille: EPUB converted from DOCX; PDF converted from DOCX
- WordToEpub: EPUB converted from DOCX; HTML converted from DOCX; DOCX converted from EPUB
- EasyConverter Express: EPUB converted from DOCX

All input formats were the ones that contained additional manual corrections.

- **PICTURE (CHEM. FORMULA) (2)**

15 outputs fully met this criterion:

- Calibre: DOCX converted from EPUB; EPUB converted from DOCX; PDF converted from DOCX and EPUB; RTF converted from DOCX and EPUB
- RoboBraille: EPUB converted from DOCX; PDF tagged image over text converted from PDF; PDF converted from DOCX; RTF converted from EPUB
- Mathpix: DOCX converted from PDF
- WordToEpub: EPUB converted from DOCX; HTML converted from DOCX; DOCX converted from EPUB

- EasyConverter Express: EPUB converted from DOCX

Most input formats included additional manual corrections; however, some results fully met the criterion using untagged PDF as the input format.

- **PRIMARY LANGUAGE (1)**

33 outputs fully met this criterion:

- Calibre: DOCX converted from PDF, RTF and EPUB; EPUB converted from RTF and DOCX; RTF converted from PDF, DOCX and EPUB
- RoboBraille: DOCX converted from PDF and PDF/UA; EPUB converted from DOCX; PDF tagged image over text converted from PDF and PDF/UA; RTF converted from PDF, PDF/UA and EPUB; MP3 converted from RTF and DOCX; DAISY full text and audio converted from DOCX; DAISY math full text and audio converted from DOCX
- Balabolka: MP3 converted from PDF, PDF/UA, RTF, DOCX and EPUB
- Mathpix: DOCX converted from PDF and PDF/UA
- WordToEpub: EPUB converted from RTF and DOCX; HTML converted from RTF and DOCX; DOCX converted from EPUB
- EasyConverter Express: MP3 converted from RTF

Most input formats included additional manual corrections; however, some results fully met the criterion using untagged PDF and tagged PDF/UA as input formats.

- **SPECIAL CHARACTER (3)**

1 output fully achieved this criterion:

- Mathpix: DOCX converted from PDF/UA

Input format PDF/UA was automatically generated.

- **STAMP REMOVAL (1)**

29 outputs fully achieved this criterion:

- Calibre: DOCX converted from RTF and EPUB; EPUB converted from RTF and DOCX; PDF converted from RTF, DOCX and EPUB; RTF converted from DOCX and EPUB
- RoboBraille: EPUB converted from RTF and DOCX; PDF converted from DOCX; RTF converted from EPUB; MP3 converted from RTF and DOCX; DAISY full text and audio converted from DOCX; DAISY math full text and audio converted from DOCX
- Balabolka: MP3 converted from RTF, DOCX and EPUB
- Mathpix: DOCX converted from PDF
- WordToEpub: EPUB converted from RTF and DOCX; HTML converted from RTF and DOCX; DOCX converted from EPUB
- EasyConverter Express: EPUB converted from RTF and DOCX; MP3 converted from RTF

Most input formats included additional manual corrections; however, some results fully met the criterion using untagged PDF as the input format.

- **TABLE (4)**

15 outputs fully met this criterion:

- Calibre: DOCX converted from RTF and EPUB; EPUB converted from RTF and DOCX
- RoboBraille: DOCX converted from PDF/UA; EPUB converted from RTF and DOCX; PDF tagged image over text converted from PDF/UA; PDF converted from DOCX; RTF converted from PDF/UA
- WordToEpub: EPUB converted from RTF and DOCX; HTML converted from RTF and DOCX; DOCX converted from EPUB
- EasyConverter Express: EPUB converted from RTF and DOCX

Most input formats included additional manual corrections; however, some results fully met the criterion using tagged PDF/UA as the input format.

- **TABLE HEADER (4)**

3 outputs fully met this criterion:

- Calibre: DOCX converted from EPUB
- RoboBraille: PDF tagged image over text converted from PDF/UA
- WordToEpub: DOCX converted from EPUB

Most input formats included additional manual corrections; however, some results fully met the criterion using tagged PDF/UA as the input format.

- **TABLE ROWS (4)**

14 outputs fully met this criterion:

- Calibre: DOCX converted from RTF and EPUB; EPUB converted from RTF and DOCX
- RoboBraille: EPUB converted from RTF and DOCX; PDF converted from DOCX
- WordToEpub: EPUB converted from RTF and DOCX; HTML converted from RTF and DOCX; DOCX converted from EPUB
- EasyConverter Express: EPUB converted from RTF and DOCX

All input formats included additional manual corrections.

- **TEXT ORDER (16)**

1 output fully met this criterion:

- WordToEpub: DOCX converted from EPUB

All input formats included additional manual corrections.

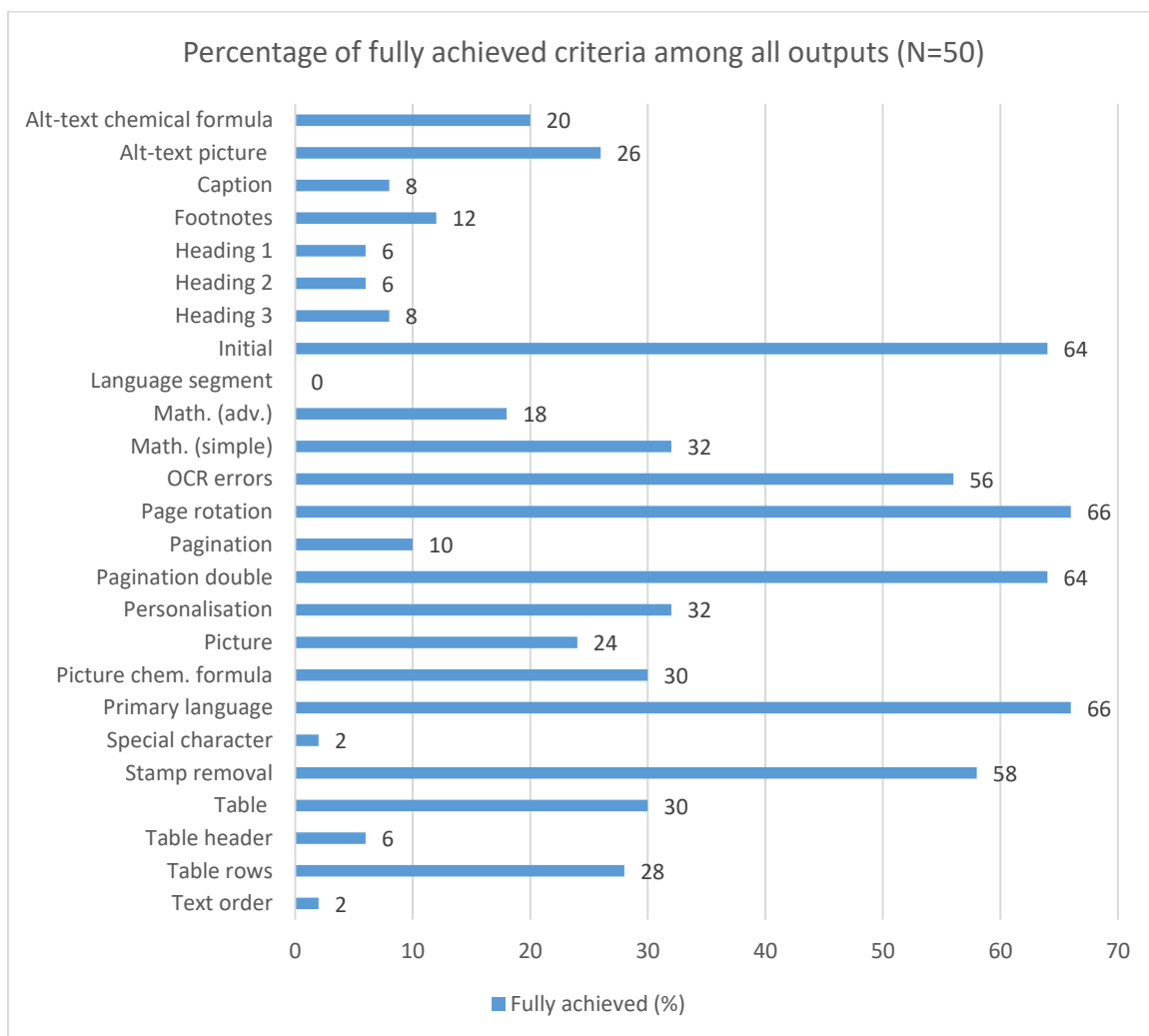


Figure 4: Percentage of fully achieved criteria among all outputs.

3.3 Results according to output format

When converting to the DOCX format (as shown in Figure 5), the two best results were achieved when using the EPUB format as the input format. Specifically, Calibre achieved 80% success rate, while WordToEpub reached an impressive 88% success rate. On the other hand, when the input format was RTF and Calibre was used, the success rate reached 44%. Unfortunately, neither the PDF nor the PDF/UA input formats yielded good results with any of the conversion services used.

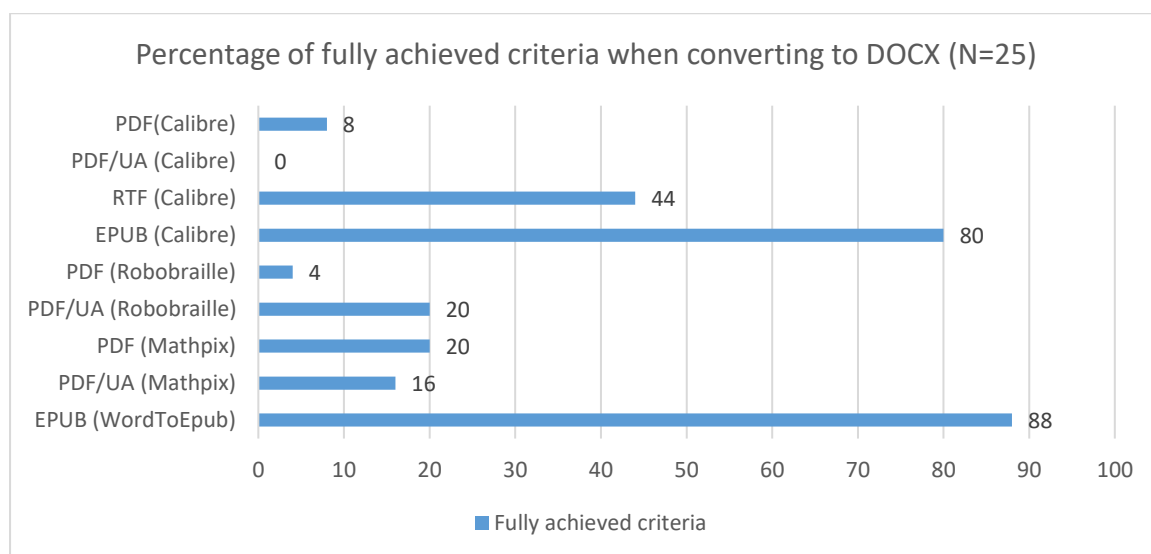


Figure 5: Percentage of fully achieved criteria when converting to DOCX based on the input format.

When converting to the EPUB format (as shown in Figure 6), the best results were achieved when using the RTF and DOCX format as the input formats. Both formats behaved similarly in all four used conversion services. DOCX behaved slightly better than RTF. RTF achieved mostly 44% or 48% success rate and DOCX achieved mostly above 50% success rate. Unfortunately, neither the PDF nor the PDF/UA input formats yielded results with any of the conversion services used.

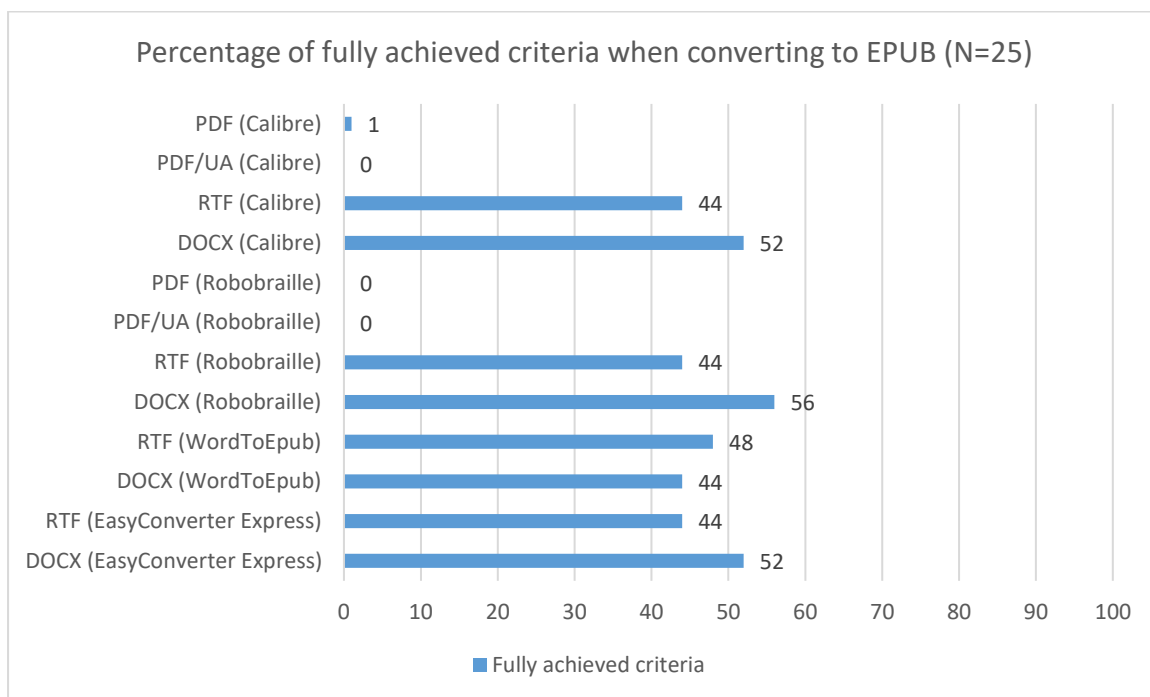


Figure 6: Percentage of fully achieved criteria when converting to EPUB based on the input format.

When converting to the PDF and PDF tagged over image format (as shown in Figure 7), the two best results were achieved when using the EPUB and DOCX format as the input format. EPUB when Calibre was used achieved 48% success rate, while DOCX when RoboBraille was used achieved the same 48% success rate.

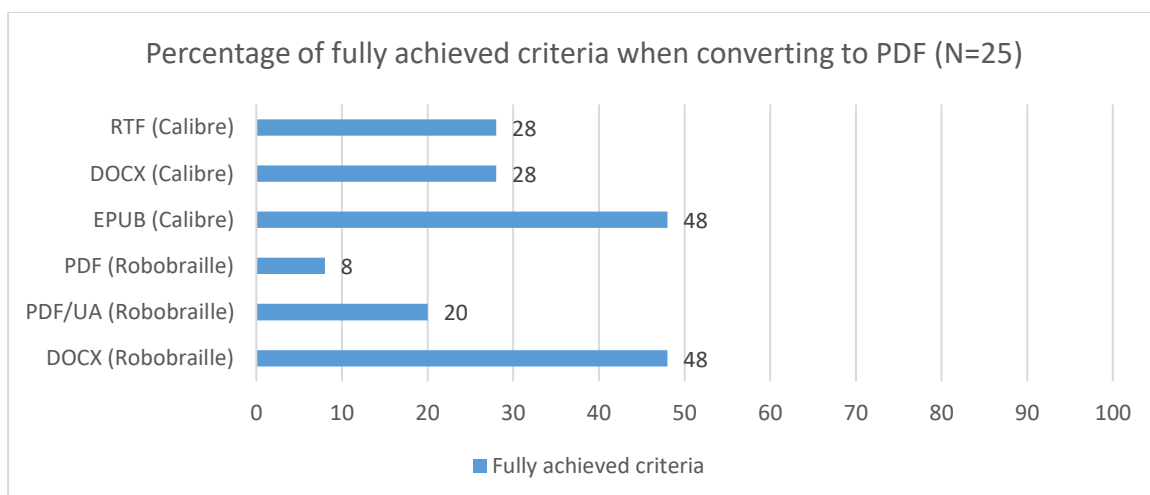


Figure 7: Percentage of fully achieved criteria when converting to PDF and PDF tagged over image format based on the input format.

When converting to the RTF format (as shown in Figure 8), the best result (48%) was achieved when EPUB format was used in Calibre as the input format. Following by EPUB using RoboBraille which achieved 48 % success rate and DOCX format using Calibre which achieved the same result. Unfortunately, neither the PDF nor the PDF/UA input formats yielded good results with any of the conversion services used.

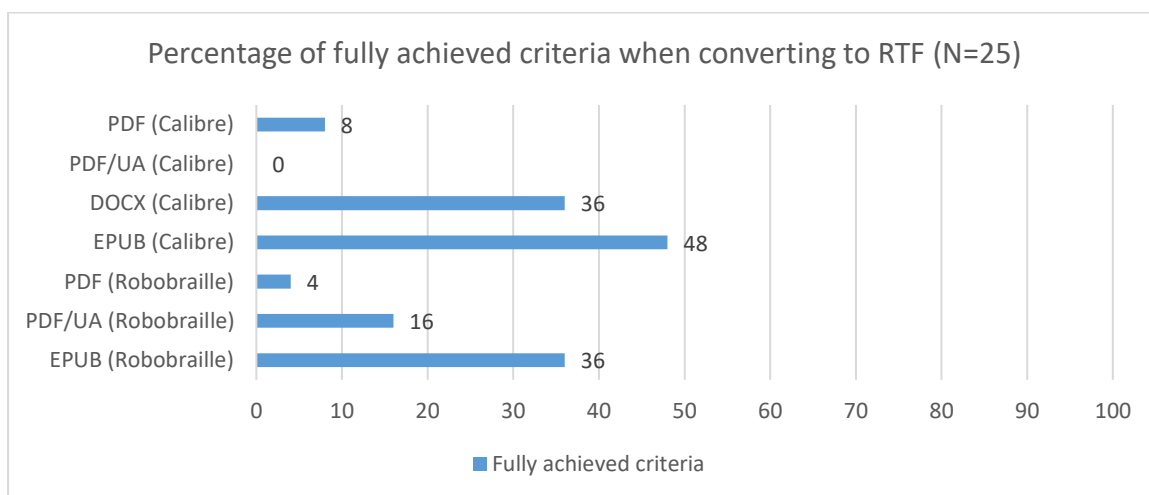


Figure 8: Percentage of fully achieved criteria when converting to RTF based on the input format.

When converting to the HTML format (as shown in Figure 9), only one conversion software was used: WordToEpub. Between the two input formats, DOCX achieved 52% success rate while RTF achieved 44% success rate.

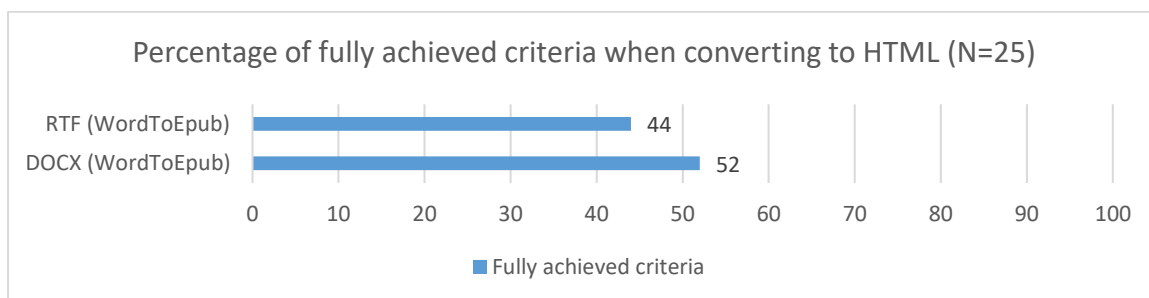


Figure 9: Percentage of fully achieved criteria when converting to HTML based on the input format.

When converting to the MP3 format (as shown in Figure 10) only 22 out of 25 criteria were used for evaluation. The two best results were achieved when using the EPUB and RTF as the input formats. Specifically, when EPUB was converted with Balabolka it achieved 36,4% success rate while RTF in all conversion services achieved the same 31,8% success rate.

Unfortunately, neither the PDF nor the PDF/UA input formats yielded good results with any of the conversion services used.

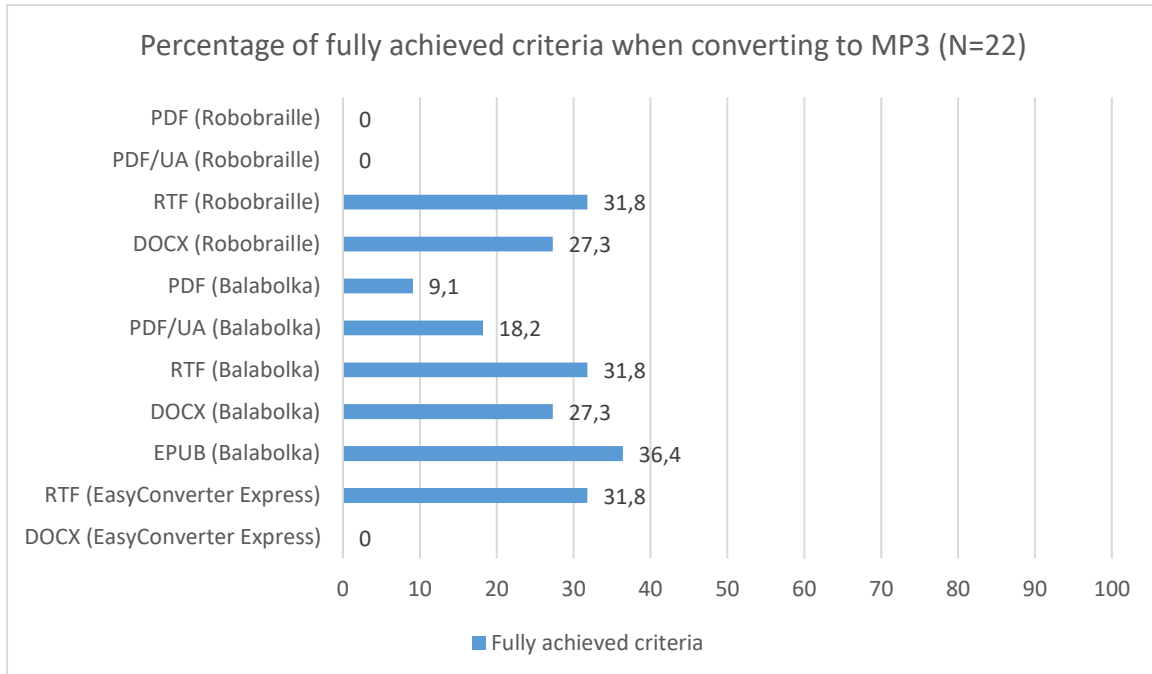


Figure 10: Percentage of fully achieved criteria when converting to MP3 based on the input format.

When converting to the DAISY format (as shown in Figure 11), only 22 out of 25 criteria were used for evaluation and only one conversion service was used: RoboBraille. Between the three outputs, two reached the same (36,4%) success rate³².

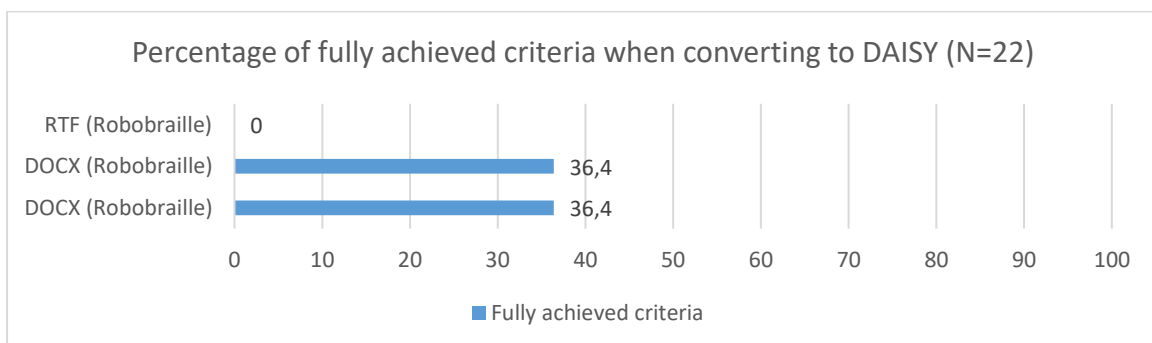


Figure 11: Percentage of fully achieved criteria when converting to DAISY based on the input format.

³² The two outputs were DAISY full text and audio and DAISY math full text and audio.

4. Findings, possible solutions and recommendations

According to the results, the most favourable conversion results were obtained when manual interventions were applied to the input. Therefore, it is advisable that before utilizing conversion services, ensure that the input format aligns with the WCAG criteria for accessibility to achieve optimal output results. Inaccessible input will yield inaccessible output. The conversion outputs excelled when dealing with DOCX, RTF and EPUB as input formats. DOCX format demonstrates excellent compatibility with most conversion services, offering a wide range of options and consistently delivering good results. RTF gave similar outcomes but content was often lost. While EPUB format demonstrated good results when the input format was done well. It is advisable to use DOCX, RTF or EPUB as input formats. All three outputs excelled in usability for mobile devices and blind and partially sighted as well.

Mathematical expressions were in most cases best handled when MathML was used in the input formats. During conversion it was handled well in most cases, in few instances the mathematical expressions were lost.

When evaluating audio formats generated through conversion services, it is crucial to be mindful of potential limitations in speech synthesis, particularly how it manages the conversion from text to audio. Unfortunately, the results from the Balabolka conversion service do not include an institutional license for speech voices, requiring organizations to reach out to the right holder(s) for licensing, especially for commercial voices. In this scenario, the conversion can be utilized for personal use, which may not be ideal for libraries.

On the other hand, RoboBraille, the second tool tested for creating audiobooks, offers an institutional license for a submission fee. When utilizing conversion to audio, emphasis should be placed on text order, language, alt-text, and, before conversion, manual adaptation—especially concerning mathematical expressions or special characters. Additionally, with regard to mathematical expressions and language segments, DAISY audiobooks created with RoboBraille handled them more effectively than the MP3 format.

To provide users with audiobooks, we suggest offering them accessible DOCX format files and informing them about the use of conversion services like RoboBraille and Balabolka for generating MP3 and DAISY formats. Regarding the choice of conversion services that are best suited for formats adaptable to different screen sizes, we recommend considering Calibre, RoboBraille, or WordToEpub. However, as mentioned previously, careful consideration of the input format is essential.

For users with special needs, we can only recommend WordToEpub, as it addressed most accessibility problems effectively. Once again, the accessibility of the input format should be considered.

5. Summary

The purpose of the *Report on the trial implementation of conversion services* (project Deliverable D12a) is to assist librarians and other organizations in the field of culture in making digitized content accessible to wider communities. The Report serves as a supplement to EODOPEN Project Deliverable 11: *Guidelines and recommendations for the provision of alternative and special formats based on the survey on special needs*, as well as Deliverable 12b&c: *Report on trial implementations for mobile devices and print-disabled users*. In this report (D12b&C), it was revealed that EODOPEN project partners use a wide range of delivery file formats, and it was found that additional manual corrections yield better results. Among the automatically generated results, the PDF/UA format, based on universal accessibility standards, provided the best outcomes.

The Report provides the outcomes of a trial implementation of various conversion services. The purpose of the testing is to evaluate the most effective solutions in the file format conversion process after digitization has been completed and to identify areas for improvement in providing digitized materials that meet accessibility criteria. This enables libraries to review and enhance the range of formats they offer to their users, considering both mobile device users and individuals who are blind or partially sighted, depending on the available resources.

The objectives of the trial implementations were as follows:

- To assess the impact of input formats by selecting five different formats from Deliverable 12b&c, which would serve as input formats for conversion.
- To select the most commonly used, newest, or recommended conversion services for the five input formats, with a focus on open-source software and tools.
- To gather detailed information on the capabilities and features of the chosen conversion services for their users, including librarians and library users.
- To determine the most recommended output formats for the conversion process.
- To compare the same output formats converted with different conversion services, allowing for the identification of the conversion services that produce better results.
- To evaluate all received output formats based on established accessibility criteria, aiming to identify the most suitable solutions.
- To identify which conversion services or input formats yield the best results with minimal additional work, particularly in terms of accessibility for blind and partially sighted individuals.

The trial implementation took place from 22nd July 2022 to 21st March 2023, utilizing the same testing results as outlined in Deliverable 12b&c. Among the five delivery formats, two were automatically generated digitized results (PDF and PDF/UA), while project partners applied manual corrections to the remaining three (RTF, DOCX, and EPUB). Among these three formats, EPUB was identified as the most accessible. These selected delivery formats served as input for the most commonly used conversion tools/services: Calibre version 6.1.0, RoboBraille, Balabolka version 2.15.0.821, Mathpix's Snip, WordToEpub version 1.0.10 and EasyConverter Express version 3.0.1.55.

A total of 50 tests were conducted using these conversion tools/services, with 42 results considered suitable for evaluation. These evaluations only encompass automatically converted results, without any further modifications to the final output formats.

The evaluation focused on the type of conversion tool/service, 25 established criteria according to Web Content Accessibility Guidelines (WCAG), and the output format.

The testing findings revealed that the most favorable conversion results were achieved when manual interventions were applied to the input. Exceptional conversion outputs were observed when working with DOCX, RTF, and EPUB as input formats. The DOCX format demonstrated excellent compatibility with most conversion services, providing a wide range of options and consistently delivering good results. When considering audio formats created with conversion services, it's essential to be aware of potential limitations in speech synthesis. However, among two audio formats, DAISY format gave better results than MP3.

As a result of the evaluation, the following recommendations were proposed:

- Consider input format before the conversion: It is advisable to use DOCX, RTF, or EPUB as input formats.
- Consider accessibility of input format before the conversion: Ensure that the input format aligns with the WCAG criteria for accessibility to achieve optimal output results. Inaccessible input will yield inaccessible output.
- Select conversion service based on desired output format: Conversion services offer a wide range of input and output formats and the quality of output format may vary when using the same input format with different conversion services. Additionally, not all conversion services provide options of the desired output format. Consider using open, and not proprietary file formats. To provide users with audiobooks, offer them accessible DOCX format files and inform them about using the RoboBraille and Balabolka conversion tools/services for generating MP3 and/or DAISY formats. Also, when preparing text for users to create audiobooks, avoid special characters and mathematical expressions by converting them into words (e.g. “one plus three equals four”; “ α – alpha, β – beta,” etc.). Alt-text can’t be directly extracted from images by the conversion services; it must be written in line with other text.

6. References

Accessible document solutions. (s.a.). *An Introduction to PDF Tags: The key ingredients in an accessible tagged PDF*. Available at: <https://accessible-docs.com/tagging-accessible-pdf/>

Guidelines and recommendations for the provision of alternative and special formats based on the survey on special needs of users and technical requirements. (2022). EODOPEN Project Deliverable D11. Available at: <https://eodopen.eu/outputs>.

Learn about sending documents to your Kindle library. (s.a.) Amazon. Available at: https://www.amazon.com/gp/help/customer/display.html?ref=hp_left_v4_sib&nodeId=G5WYD9SAF7PGXRNA

Report on Trial Implementations for Mobile Devices and Print-Disabled Users. (2023). EODOPEN Project Deliverable D12 b&c. Available at: <https://eodopen.eu/outputs>

7. Vocabulary

ALTERNATIVE TEXT (ALT-TEXT) – Alternative text provides a textual description for non-text content (pictures, graphics, diagrams ...).

ASSISTIVE TECHNOLOGIES – “... any item, piece of equipment, software program, or product system that is used to increase, maintain, or improve the functional capabilities of persons with disabilities.” (Source: ATIA, <https://www.atia.org/home/at-resources/what-is-at/>)

CONVERSION SERVICES – software that enables file format conversion from input to output format (ex. conversion of DOCX format into EPUB).

DELIVERY FILE FORMAT – the final file formats accessed by the users.

DIGITAL CONVERSION – digitisation.

DIGITISATION WORKFLOW – all the processes implemented during the digitisation process from image capturing, image processing, OCR production ..., to the conversion of scanning file format to archival and access file formats.

EBOOK – Usually, the term eBook refers to digitally born publications. However, we use the term of eBook especially referring to digital publications produced as a result of digital conversion, including formats for special needs (audiobooks), which is also the aim of EODOPEN project.

IMAGE CAPTURING – scanning.

IMAGE PROCESSING – “Image processing is a method to perform some operations on an image, in order to get an enhanced image or to extract some useful information from it.” (Source: Digital Image Processing, University of Tartu, <https://sisu.ut.ee/imageprocessing/book/1>).

MOBILE DEVICES – were mobile or smartphones, laptops, and tablet computers.

PARTIALLY SIGHTED – “People who are partially sighted are not completely blind but are able to see very little.” (Source Cambridge Dictionaire, <https://dictionary.cambridge.org/dictionary/english/partially-sighted>). Use for visually impaired.

PRINT DISABLED – “The term “print disabled” was coined by George Kerscher, Ph.D. around 1989 to describe persons who could not access print. He used it to refer to: A person who cannot effectively read print because of a visual, physical, perceptual, developmental, cognitive, or learning disability.” (Source: <https://myblindspot.org/mbs-accessibility-defined/>).

PROPRIETARY FILE FORMATS – formats that rely on specific software for using and the content of the file can’t be read without that software, ex. MOBI, AZW.
RESPONSIVE FILE FORMAT – is a format that enables the text to adjust to any screen size.

SCREEN READER – “Screen readers perform a text to speech role, but also allow audio-only access to the menus and other features of the delivery platform.” (McNaught and Alexander, 2014)

TAGGED PDF – PDF which contains tags for each page element and enables easier access to document’s content with assistive technologies.

TEXT TO SPEECH – “Text to speech is a mature technology that allows text on screen to be voiced by software. (McNaught and Alexander, 2014)

VISUALLY IMPAIRED – see partially sighted.

8. Used acronyms

DRM – Digital Rights Management services

E2W – EpubToWord

EBU – European Blind Union.

EODOPEN – eBooks-On-Demand-network Opening Publications for European Netizens – European project cofinanced under Creative Europe program from 2019-2023.

EODOPEN PARTNERS ACRONYMS

BNP -National Library of Portugal

CVTI SR - Slovak Centre of Scientific and Technical Information

MZK - Moravian Library

NCU - Nicolaus Copernicus University in Torun

NLE - National Library of Estonia

NLS - National Library of Sweden

NUK - National and University Library

OSZK - National Széchenyi Library

UG, University of Greifswald

UIBK - University of Innsbruck

UREG - University of Regensburg

UT - University of Tartu

VKOL - Research Library Olomouc

OCR – Optical Character Recognition

TOC – Table of contents

WCAG – Web Content Accessibility Guidelines

W2E – WordToEpub

Acronyms for file formats:

AZW – Amazon Word

DOCX - Microsoft Word Open XML Format

EPUB - electronic publication

HTML - Hyper Text Markup Language

MOBI - MOBI file format (Mobipocket eBook format)

PDF – Portable Document Format

RTF – Rich Text Format

9. Annexes

Annex 1: Testing samples

Scan 1

the
beer
the
ears
and
This
iter.
ring

ting
is is
tivist
reds
ical
the
h is
oper
nol.
nol,

been
e by
lled
they
isms
uce
thyl
the
l in
ave
tely

this
ake
own
mall
as a

any
per-
they
ices,
Gly-
king

from
con-
l, or
ilar.
o) the
con-
s as
ases.
rics,
pro-

h as
with
vap-
rbon
d in
zing
s an

etyl
vaxy
.

algebra

A branch of the science of mathematics. It is like arithmetic, but as well as using numbers it uses letters of the alphabet to symbolize numbers.

One reason for using symbols in place of numbers is that we often do not know what the number we are interested in is. We are trying to find out what it is.

Suppose a boy tells us that he wants to buy something costing £5 and he needs another £2. We could say this algebraically by writing $x + 2 = 5$. Here x stands for the number of pounds he already has. We can then go on to work out what number x actually is. Later we shall look at some ways of doing this.

Algebra can be thought of as a form of shorthand. For addition and subtraction we use the familiar signs $+$ and $-$. In arithmetic we use \times as the multiplication sign. We avoid using it in this way in algebra, because x is often used to stand for an unknown number. To show that two numbers are to be multiplied together, we usually place them side by side with no symbol between them. For example, $2x$ means 2 multiplied by the number that x stands for. And ab means the number that a stands for multiplied by the number that b stands for.

For division we use the idea of fractions. $\frac{x}{3}$ means x divided by 3. If we need to show a number multiplied by itself several times we use small index numbers called 'exponents'. For example, x^2 means x multiplied by x . x^4 means $x \times x \times x \times x$; that is, four x 's multiplied together. 2 and 4 are the exponents.

Here is an algebraic expression using some of the above 'shorthand': $x^2 + 2xy - 3y$. The symbols x and y stand for two unknown numbers. If, for example x stands for the number 3 and y stands for the number 5 the expression becomes: $(3 \times 3) + (2 \times 3 \times 5) - (3 \times 5)$ or $9 + 30 - 15$, which is 24. The brackets show which parts to work out first.

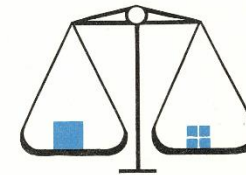
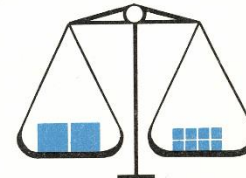
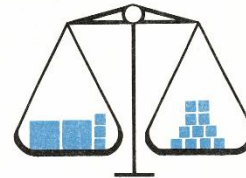
Algebraic statements

To make a complete statement using numbers we must have a 'verb'. We use 'equals' or 'is greater than' or 'is less than'. The signs for these verbs are $=$, $>$ and $<$. Using ordinary numbers, we can then write: $2 + 3 = 5$; $2 + 3 > 4$; $2 + 3 < 6$.

We can also use symbols instead of numbers. For example: $2x + 3 = 7$. This is an 'equation'. It says that twice the number that x stands for, with 3 added to it, is equal to 7. The statement is true if x stands for the number 2, but not true if it stands for any other number.

An equation is a statement that is only true for certain values of the unknown number. These values are called *solutions* of the equation.

An *inequality* is a statement like $2x + 3 > 7$. This statement is true for a range of values of x . It is true in fact for all values greater than 2. To solve an equation or inequality means to find the number or set of numbers that make it a true statement.



The two sides of an equation are like two equal weights. If they are changed in the same way, they remain equal. At the top, $2x + 3$ balances 11.

When 3 is taken away from both sides, $2x$ balances 8. When both sides are halved x balances 4.

An equation can be thought of as a see-saw or balance. The illustration shows $2x + 3$ in the left-hand tray 'balanced' by 11 in the right.

If one side is changed the other side must be changed in the same way to preserve the balance. In the example; take 3 from each side; the equation becomes $2x = 8$. So $x = 4$.

In solving an equation the aim is to finish up with another equation that tells us at once what the unknown number is. This final equation has the unknown number on one side only. Nothing unknown must appear on the other side.

Arithmetic only makes statements about particular numbers. For example, $3 + 2 = 5$ is an arithmetical statement. But sometimes we wish to make a statement that is true for all numbers, or a wide range of them. For example, suppose we wish to say that $3 + 2 = 2 + 3$, and $4 + 6 = 6 + 4$, and so on for *all* possible pairs of numbers. We can do this with the single

experiments could show which substances were elements, and there might be many more than the four 'elements' of the ancients.

The final break away of chemistry from alchemy came when the theory of phlogiston was proved wrong. This theory was put forward by the German Georg Stahl in the 18th century. He said that phlogiston was a substance given out when things burn. Things that contained phlogiston would burn; those that did not would not burn. The more phlogiston present, the more inflammable the substance.

Many people studied this theory, like the well-known British chemist Joseph Priestley. It became generally accepted that phlogiston explained the burning of substances. It was the French scientist Lavoisier who finally proved this wrong.

In 1774, he repeated one of Priestley's experiments and weighed everything carefully. He heated mercury in a closed vessel. A dark crust was formed on the surface of the mercury. He found that the mercury had increased in weight while it was burning, and that the amount of air in the vessel had decreased. Thus he concluded that the mercury had not given off phlogiston. Instead it had taken in something from the air. This something he called oxygen. He found it was possible to get back this oxygen by heating the crust of mercury vigorously. In this way, the oxygen was released again.

Many new chemical substances were discovered after this time. They were given names in a logical way by Lavoisier and his followers. Chemists now began to try to make chemicals in their laboratories, starting with simple substances and trying to make them react and combine. Later on, with the help of electricity, they were able to break down natural mineral products and build them up again. But it did not seem possible to do this with products from animals and plants. These chemicals were called 'organic' because they came from living things (see: *carbon*). Organic and inorganic chemistry eventually became two distinct branches of science.

For many years, it was believed that there was something missing, a 'vital principle' that was lost - rather like phlogiston - when an organic substance was broken down into its component parts. The chemist Wohler, in 1828, accidentally changed the inorganic chemical ammonium cyanate (which he got from strictly mineral sources) into the organic chemical urea (which occurs in urine). This was the start of a new approach to chemistry.

Atoms, molecules and equations

The chemist deals in very exact quantities: only precise numbers of atoms or molecules react with each other. And he needs to know exactly what quantities of chemicals to mix together. For this reason he states his chemical facts in a mathematical form.

First of all he needs a set of symbols to represent all the different elements. Lavoisier was the first to give the elements logical names like oxygen ('acid-maker') and hydrogen

('water-maker'). The symbols for these elements are now O and H. Other symbols are equally obvious: S for sulphur, P for phosphorus, C for carbon, Ca for calcium ('chalk metal'), and so on. Other elements were given symbols derived from their Latin names: Na for sodium, from natrium; K for potassium, from kalium.

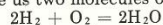
In 1808 John Dalton published his theory of the atomic nature of matter. His suggestions have had to be slightly modified as we have found out more about atoms, but they are basically unchanged. In brief: every element is made up of atoms, which do not break up during chemical reactions. All atoms of any one element are the same in weight and other properties. But atoms of different elements differ from one another. Chemical compounds are formed by atoms of different kinds joining together.

The symbol of an element, therefore, is used to represent one atom of an element. Compounds are represented by combinations of the atomic symbols, showing how many of each kind of atom there are in one molecule. Oxygen, for instance, exists normally as a gas with molecules that contain two atoms. We write this O₂. Hydrogen similarly exists as the gas H₂. If we mix hydrogen and oxygen together and explode them with an electric spark, water is formed.

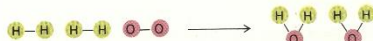
But to make water, we need two molecules of hydrogen to every one of oxygen:



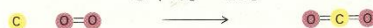
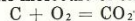
This will give us two molecules of water:



This is a chemical equation, and it tells us very simply about the chemicals that have reacted together, the quantities of each, their molecular structure and what the reaction has produced.



In the same way, one molecule of oxygen, containing two atoms, and one atom of carbon will produce one molecule of carbon dioxide:



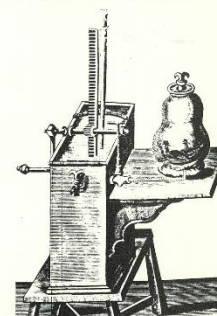
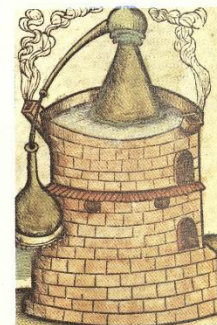
The chemical equation is just like an equation in algebra; the numbers of molecules that are involved are reduced to the smallest whole number. But in reality, of course, many millions of molecules are present.

Different elements combine together in different proportions. It is as if the atoms had little hooks with which they could attach themselves to other atoms. (In fact, this is due to the number of electrons present in the outer orbits of the atom.) Hydrogen has only one hook, H-. Oxygen however has two hooks, -O-. This is why two atoms of hydrogen and one of oxygen make water, H-O-H.

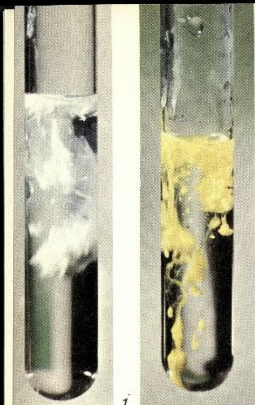
The number of hooks of this kind that an atom possesses is called its valence. Hydrogen has a valence of only 1; oxygen has a valence of 2. Calcium also has a valence of 2, aluminium has a valence of 3, and carbon has a valence of 4. Some elements have more than one valence.



Iron mixed with sulphur (top) responds to magnetism (middle). Heat creates a new substance (above) that is non-magnetic



Scan 3



Scientists have now made artificial elements of high atomic weights. In honour of Mendeleev, one of these was named mendeleevium in 1955 (see: *physics*).

What chemists do

As we have seen, all chemistry is divided into inorganic and organic chemistry. This is a rather artificial division, but it recognises the difficulty that the early chemists had in dealing with the products of living things.

Roughly speaking, inorganic chemistry is concerned with mineral substances. The possible number of inorganic chemicals is limited. Most naturally occurring minerals are either simple compounds of a metal with chlorine, oxygen or sulphur, or with sulphuric, phosphoric or carbonic acid. Or they are rather more complex structures held together by the element silicon. Most of the inorganic chemist's work concerns finding better ways of preparing pure elements, methods of separating pure compounds, and ways to make different mixtures of metals (alloys) that have special properties.

Organic chemistry is concerned with substances that are compounds of carbon with one or more other elements. Carbon has very special properties. It can make many thousands of compounds and these include the most complex substances known.

Silicon can also form many complex compounds. It is similar in its properties to carbon.



The chemist at work:

- 1) In this test a white cloud of silver chloride shows a chloride is there.
- 2) This yellow cloud shows that lead is present.
- 3) The brownish cloud is ferric hydroxide.
- 4) Bright Prussian blue

- indicates iron.
- 5) These two 'Soxhlet extractors' are used to make solutions of substances that are slow to dissolve.
- 6) A spectrometer detects the presence of lithium by its pink flame, which also shows the quantity

Some important organic substances – silicones – have been made, in which silicon partly replaces carbon. It has been suggested that life on some other planets could be based on compounds of silicon rather than carbon compounds.

Organic chemists concern themselves with analysis. When they discover a new compound they try to find out what elements it is made of and how they are put together in the molecule. Then they try to make the compound in the laboratory. This is called synthesis. If they succeed in making the compound, they often try to make other compounds like it, but which may be even more useful.

Chemistry in industry

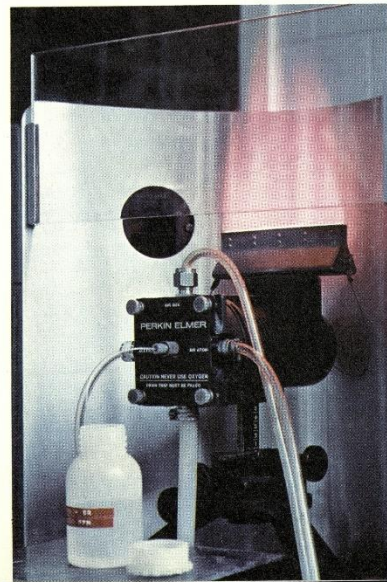
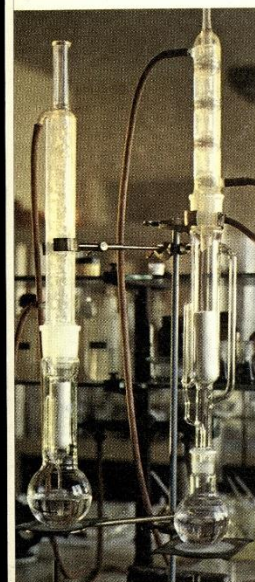
Some of the world's earliest industries were chemical ones: drugs, brewing, mining and metallurgy. The chemists of the 18th century laid the foundations of today's 'heavy chemical' industry. They found that by heating coal in closed vessels coal gas was formed. This could be used for lighting and heating. A number of liquids and a residue of tar were also formed. The liquids could be separated from each other by distillation. Most of them were oily, inflammable compounds of carbon and hydrogen. The organic chemists analysed these substances and gave them names: the hydrocarbons benzene, toluene, naphthalene and anthracene; the phenols carbolic acid, cresol and so on.

Organic chemists everywhere now turned their attentions to improving on natural products. They took the separate compounds obtained from coal tar and experimented with them. They tried to find ways to change them into other compounds by adding or taking away groups of atoms. They made drugs like aspirin, lots of different dyes, explosives, fruit flavours and perfumes. They also started to make substitutes like margarine, and celluloid (see: *cellulose*), rayon and saccharine. Later they made detergents and artificial rubber from the by-products of the oil industry.

Chemists have also devised fertilizers that are now indispensable to modern farming, and additives, colorants and preservatives to alter the properties of food and cosmetics. Useful as they are in providing us with convenient products, some of these chemicals can be a threat to health. Many people now prefer to buy food that is free from artificial substances or that has not been grown with synthetic fertilizers.

The foundations of the plastics industry were laid when Otto Baekeland heated two simple chemicals – phenol and formaldehyde – together and made bakelite. This artificial resin is formed by many molecules joining together and giving off water. The very large molecule that is formed in this way is called a polymer.

Many substances occurring in nature are polymers of this kind. Silk and rubber are examples. In 1930 the American chemist Carothers began to study polymerization, and in 1935 he produced a polymer that was very similar to the protein in silk. Nylon was developed from this discovery. Carothers's research also led to the synthesis of rubber. Nylon was the first synthetic textile fibre developed. See: *acids and bases, atom, physics*.

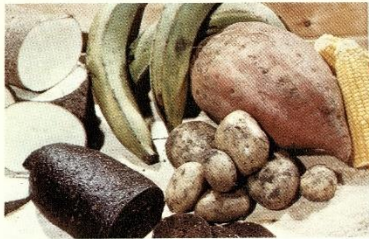


Source: *The Lexicon Encyclopedia of Science* (p. 5 volumes). (1984). Lexicon Publications, Inc., vol. 1, p. [156].

Scan 5

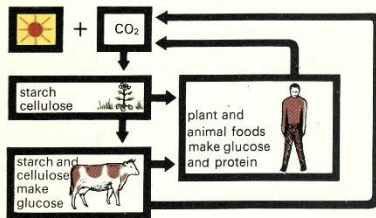
132 carbohydrate

carbohydrate



△ Basic foods, like yams, potatoes, bread, bananas, corn, and sweet potatoes, are rich in carbohydrates

▽ Carbohydrates are made by plants from light, water and CO₂. Man gets energy by eating plants and animals



What do a sugar lump, a loaf of bread and a potato have in common? They all contain carbohydrates. The sugar lump is made entirely of one of the simpler types of carbohydrate, which are all called sugars. The bread is made from flour, which contains a more complicated carbohydrate, starch. The potato also contains starch, food stored by the plant to feed its new shoots in the spring.

There are many types of sugar. Plants make sugars by the process of photosynthesis (see: *photosynthesis*). Plant sap is full of sugar. Both plants and animals change these sugars into one type, glucose, which they use for energy. Plants store glucose by combining many molecules of it to form into a large starch molecule. Starch does not dissolve in water and is too large to enter or leave a cell. When the plant needs the stored energy, it breaks down the large molecule into small glucose molecules and uses them. Glucose is a small molecule that dissolves in water, and passes easily into and out of the cell. Carbohydrates are the main source of energy in our diet. When we eat stored plant starch in flour or potatoes, our bodies too break it down into glucose.

The name carbohydrate refers to the chemical composition of these compounds. The formula shows that for each carbon atom (C) there are two hydrogen atoms (H) and one oxygen atom (O), as in water (H₂O). For instance, a sugar that contains six carbon atoms usually has enough hydrogen and oxygen atoms to make six molecules of water: C₆H₁₂O₆.

It is more accurate and useful to draw a diagram of the molecule. There are many different sugars with the formula C₆H₁₂O₆. Each of these sugars has a different taste and takes part in different chemical reactions, according to the arrangement of its atoms.

Sugars are the simplest carbohydrates. The simplest sugars consist of just one type of sugar molecule. Such a molecule is called a monosaccharide, that is, 'one sugar'. Usually this molecule contains from three to seven carbon atoms. Glucose, which is found in all plants, is the main sugar found in fruit. It has six carbon atoms. Ribose, a sugar that is found in the nucleus of the cell, has five carbon atoms. Two other five-carbon sugars are found in the sticky gums with which plants heal wounds in their bark. These are arabinose and xylose.

Two monosaccharide molecules can be joined together to make a double sugar, a disaccharide. We are most familiar with the disaccharide sucrose. This is the sugar we eat in our food. Sucrose is found in the sap of many plants. Maple syrup is mostly sucrose. The commercial sugar industry makes its brown or white crystals of sucrose from sugar cane and sugar beets. In warm climates these plants are fairly easy to grow in large quantities and yield plenty of sucrose. Honey also contains sucrose, mixed with glucose and another sugar, fructose.

Each molecule of sucrose is made up of a molecule of glucose and one of fructose joined together. These double molecules are too large to be absorbed by our cells. When we eat

Much of the sugar we eat is made from the sweet sap of sugar cane. This sap tastes good when it is fresh. It is boiled down to concentrate its sugar content into syrup or crystals. Sugar cane grows only in hot countries. Here it is being harvested in Barbados

▽
tr
i
pou
b

sucr
and
carb
gluc
Al
plan
leavi
gree
uses
For
carb
oxyg
from
gluc
used
hydr
Wl
large
polys
port
gluc
Gluc
sacci
glyce
goes
glyce
level
Pls
polys
tree
comp
stren
cean
Sir
be a
most
cellu
passe
absor
anim
have
or ye

THE COMMON CAUSE

Fortnightly of the Polish Section of the "Sword of the Spirit"

EDITORIAL OFFICE: 27 GROSVENOR STREET, LONDON, W.1

Telephone: MAY 2928

Vol. 4 Nos. 4 and 5

SUNDAY, MARCH 24th, 1945

Price One Shilling

ST. STROŃSKI

THE RESURRECTION

HAPPY INDEED ARE THE nations for which the idea of resurrection exists only in the meaning of a creed—the belief of a man in eternal life. Since the time of the partitions—from the end of the eighteenth century—the idea of resurrection has been for Poland a fundamental idea of national existence. It has been the hope of generations, and the object of longing for those millions who, for countless days, saw the sun rise immutably over Polish soil when it was trodden by conquerors and oppressors. Belief in the earthly resurrection of the nation became one of Poland's guiding stars along the highway of her history.

The country's unhappiness filled to overflowing the minds of her people. Was it deserved or not? Was it the result of our own guilt, or of the crimes of other nations? This question has long been at the core of the enigma of Polish existence.

It is a hundred years now since the great Polish poets, led by Mickiewicz, merged their loftiest inspiration with belief in God's justice; they gave to the first generation "born in slavery and chained when in the cradle," this slogan: God has selected Poland to regenerate all mankind by her martyrdom.

So said the Polish messiahism. The consciousness and sober steadiness of the Poles, added to the general trend of history, have produced the conviction which is held both in Polish and in world science, that the virtues and defects of the Polish nation are neither greater nor less than those of other nations. But no other country in the world has such mighty, rapacious neighbours as Germany and Russia. Any nation would find, just as Poland found, that it is hard to stand up to such colossal and usually simultaneously-applied double pressure.

So say simple geopolitics. There is something which many still fail to realise. In the recent debates in the Houses of Parliament, were uttered on March 1st, 1945—among numerous noble and prudent declarations—such words as these: The Poles possess many virtues, but they are deficient in political wisdom and statesmanship. That, of course, depends on what one considers political wisdom and statesmanship to be. Some people understand them to mean that if someone stronger takes half your country and removes from the other half its independence, you must accept. For Poles this breed of wisdom is unacceptable.

The sombre reality of Poland's ominous neighbours weighed on her destiny from the end of the eighteenth century till the beginning of the twentieth, and again from the outbreak of the present war. The Russo-German understanding of August 1939 was followed in September of the same year by the double invasion—Germany from the west, Russia from the east—and the dividing of Poland between the invaders. From 1941-44 the whole of Poland was subjugated by Germany. By the end of 1945 the whole of Poland may be subjugated by Russia. A continuous and bloody Golgotha.

The world looks on with sincere emotion and general compassion, but the world powers pass sentence in the Crimean resolutions.

Before the resurrection of Christ there was the buying of Him for silver, the washing of hands, the mocking and the martyrdom, but it came and it was victorious.

What is there left for Poland? Christian, unbroken and active faith in resurrection. And that is not little. It is very much. It is more than the surrender to evil. It is the only way into the light.

Mgr. ZYGMUNT KACZYŃSKI

NOT DICTATORSHIP, BUT UNDERSTANDING

ONE OF THE BRITISH PUBLICISTS after the debate in Parliament on the Yalta resolutions said that the House had indeed accepted them but that the shadow of the Polish problem remained. In my opinion not only the shadow remained but the very essence of the matter remained; even according to the Crimean declaration, many points are still unsettled and afford a field for many interpretations. Above all the conscience of the civilized world remained—a conscience uneasy and dissatisfied with the Crimean resolutions on the Polish problem.

The highest guardians in the world of moral values criticize the attitude of the three powers towards the Polish problem. Catholic and Orthodox bishops, eminent dignitaries of the Protestant Churches, rabbis and followers of the Mosaic Law, stand forth in defence of the just rights of Poland. If to these voices we add the actions of eminent statesmen, politicians and publicists, we are convinced that it is not possible at this moment nor will it be possible in the future to ignore the Polish problem.

During this war, which now seems to be approaching its end, we have experienced very much. During these terrible, but yet not very numerous years, many, many events and changes have passed before our eyes. During this period of time, the final seal has not been placed

on these or other declarations, decisions and resolutions of the powerful ones of this world; and on the edifice of the new order erected by them the inscription of Dante, "Lasciate ogni speranza . . ." does not appear.

It really does not seem long since, on October 7th, 1939, Reichsmarshall Adolf Hitler announced in Danzig, today besieged by the Russians, that the Polish state had once and for all ceased to exist; and Benito Mussolini, at almost the same time, pronounced the memorable words, "Polonia e liquidata." Intoxicated with the victory over France, the Governor General Frank proudly asserted in 1940 in Cracow that "Hitler is the leader of the world, unlimited in his power," and that "the Germans are proud that they rule the world." *Quantum mutatur ab illis!*

Yes, not much time has elapsed, and terrible punishment is already overtaking the criminals.

I do not want to belittle or make light of the significance of the three-power conference in Yalta. It does not seem to me, however, that the conference pronounced the last word in this war, that its decisions cannot undergo a revision or changes at the future peace conference, or even through negotiations and mutual understanding. The final solutions, also, need not be come to through armed conflict between the

"Allies," in which I personally do not believe, because all states are tired and exhausted by this war and desire a speedy peace. For that matter, armed conflict is not the only means of exerting pressure and supporting just demands. Economic arguments and world opinion have sometimes a greater importance than the eloquence of guns.

If the future peace is not based on the principles of justice and respect for the sovereign rights of every country particularly in central Europe and in the Balkans, we shall have to look forward to a state of general ferment and discontent, of increased hunger and general poverty. By force and terror alone it will not be possible to master the situation and maintain governments over nations thirsting for freedom and desirous of deciding their own constitutions and fates.

The relations of Russia to Poland have become the chief problem in this war. Without any doubt a long peace in Europe after the defeat of Germany depends on a favourable solution of the Polish-Russian conflict and on the establishment of good-neighbourly relations between Russia and Poland. If, therefore, Russia is really anxious to cooperate with Poland, and to maintain peace and exchange economic products, she should avoid leaving in the Polish soul the feeling that a great wrong has been done. Such a feeling must remain if Poland is deprived of almost half the territory that has hitherto belonged to her. The more so, that Polish land has no vital significance for Russia.

Under the present conditions, therefore, in the interest of European peace, it is essential to work hard and strive for a satisfactory solution of the Polish-Russian problem on the basis of a sensible compromise, and of the principles of justice and honour for the two neighbouring Slav States. At the Yalta conference Poland was spoken of and decisions taken without her voice being heard and without her participation. It is, therefore, high time to enter upon direct conversations with the Polish nation and to make agreements with her that will be binding upon her in the future.



Wooden Church in Carpathian Mountains

EDMUND SPENSER

EASTER DAY

Most glorious Lord of life, that on this day,
Didst make Thy triumph over death and sin;
And having borrow'd hell, didst bring away
Captivity thence captive us to win:
This joyous day, dear Lord, with joy begin,
And grant that we for whom Thou didst die
Being with Thy dear blood clean wash'd from sin,
May live for ever in felicity,
And that Thy love we weighing worthily,
May likewise love Thee for the same again;
And for Thy sake that all like dear didst buy,
With love may one another entertain.
So let us love, dear love, like as we ought,
Love is the lesson which the Lord us taught,
(From Amoretti: Sonnet LXXVIII)

G. K. CHESTERTON

THE BALLAD OF GOD-MAKERS

A bird flew out at the break of day
From the nest where it had carried,
And ere the eve the bird had set,
Fear on the kings of the world.
The first tree it lit upon
Was green with leaves unshed;
The second tree it lit upon
Was red with apples red.
The third tree it lit upon
Was barren and was brown,
Save for a dead man nailed thereon
On a hill above a town.
That night the kings of the earth were gay
And filled the cup and can;
Last night the kings of the earth were chill
For dread of a naked man.
"If he speak two more words," they said,
"The slave is more than the free;
If he speak three more words," they said,
"The stars are under the sea."
Said the King of the East to the King of the West,
I wot his frown was set,
"Lo, let us slay him and make him as dung,
It is well that the world forget."
Said the King of the West to the King of the East,
I wot his smile was dread,
"Nay, let us slay him and make him a god,
It is well that our god be dead."
They set the young man on a hill,
They nailed him to a nail;
And there in darkness and in blood
They made themselves a god.
And the mightiest word was left unsaid,
And the world had never a mark,
And the strongest man of the sons of men
Went dumb into the dark.
Then lyones and harps of praise they brought,
Incense and gold and myrrh,
And they thronged above the seraphim,
The poor dead carpenter.
"Thou art the prince of all," they sang,
"Ocean and earth and air,"
Then the bird flew on to the cruel cross,
And hid in the dead man's hair.
"Thou art the sun of the world," they cried,
"Speak if our prayers be heard,"
And the brown bird stirred in the dead man's hair,
And hid in the dead man's hair.
Then a shriek went up like the world's last cry
From all nations under heaven,
And a master fell before a slave
And begged to be forgiven.
They covered, for dread in his weakened eyes
The ancient wrath to see,
And a bird flew out of the dead Christ's hair,
And lit on a lemon-tree.



The Pilgrims before Czestochowa

Scan 7

100TH ANNIVERSARY OF THE LWOW POLYTECHNIC INSTITUTE

by JERZY W. MEIER



Kazimierz Bartel, former Prime Minister and Professor of the Lwów Polytechnic, was assassinated by the Germans in Lwów for refusing to cooperate with them.

LONG before the war, plans to celebrate the one-hundredth anniversary of the founding of the Lwów Polytechnic Institute were made by the directors of the Institute, now in exile. They undertook the construction of several buildings for the Mechanical and Electrotechnical Departments,

historical uprising of 1848 that strove to break the despotism of the Holy Alliance. The Galician massacre in the Cracow region, the Austrian occupation of Lwów, the uprising in Poznań, the temporary freedom of Lwów and the revengeful bombardment of the city by General Hammerstein which destroyed the old University of Lwów followed. All this along with later repression under Austrian rule developed the character of the Poles and the technical progress of their school for more than the program of education dictated by Vienna. Once again progress was disrupted by the Insurrection of 1863, in Russian-held Poland, when youths from Lwów as well as from every part of Poland crossed the border and flocked to join the volunteers. This time repression was not so severe, as the artificial dual monarchy was already shaken by its disastrous defeat at the hands of the Germans at Sedan in 1871. To rescue their decaying empire, the Hapsburgs adopted a constitution and granted limited self-government to their subject peoples. Education was reorganized by the Provincial Sejm in

Lwów. In 1872, the Lwów Technical Academy was changed to the Polytechnical School. A rector and senate were placed in charge of it with the right to hold state examinations and grant state diplomas in engineering. This placed the Institute on a par with the Polytechnic School at Vienna. At the same time, classes were transferred to a new building. At last the school was assured of steady growth and progress. This was the era of the pioneer engineer Stanisław Szczepanowski and of positivist tendencies in Russian-held Poland. Technology grew in popularity, and the school gained an ever-increasing number of students. The failure of the insurrections and the proverbial poverty of bureaucratically governed Galicia further stimulated the desire to create on Polish soil technological schools equal to any found in Western Europe. The Lwów Institute thus became not only an institute of technology but also a school of Polish patriotism. Many of its students were among the first to join the ranks of Piłsudski's Legions in 1914, and fought through the World War at its side.



Lwów Polytechnic Institute.

along with new dormitories for aviation students and testing laboratories for the civil engineering students. All these workshops of the school were used during the development of the Central Industrial Area of Poland.

A special anniversary book was to have been published which would have given a brief review of past achievements and a detailed report of Polish technical work and the contribution of Polish thought to the development of technical study. A series of monographs was to have been published on the work of alumni of the Lwów Polytechnic Institute and of the general Congress of Polish Engineers. This work was disrupted by Germany's aggression and the ensuing occupation. Polish professors—those who still live—are in German concentration camps, young Polish engineers and students are either in the Allied armies, have joined the Home Army, or are prisoners of war.

The beginnings of this Institute reach back to 1825, when Stanisław Stasiak became head of the Polytechnical Council called by the Polish Government to organize a Polytechnic under the direction of Kajetan Garbiński. Stasiak opened this school two weeks before his death. The school progressed steadily and its program was enlarged and improved until the Uprising of 1831. Not until 1844 was another attempt made to establish another such school under the name of "Technical Academy."

Normally, in the life of a nation, a hundred years is no great length of time, but the last century before the rebirth of Poland was a most stormy period during which political and economic events affected the development of the Lwów Technical Academy. Just as the school began to function again, its progress was interrupted by the so-called Spring of the Nations, the revo-



Monument to the students who fell in the defense of Lwów in 1918-1919, in the square of the Lwów Polytechnic. The garden had been used as a first aid station during the war.

Following the first war, after regaining the right to rule itself, the enrollment of the Lwów Polytechnic Institute rose to more than 3,000 students, from the pre-war total of 300. Buildings, class rooms, and dormitories were inadequate. The Institute, however, was fortunate in obtaining the buildings of the Magdalena Convent.

In 1917, construction of new buildings for the Department of Mechanical Engineering began on land donated by the city of Lwów. Seven buildings were to have been erected for the new departments of theoretic and practical engineering, metallurgical technology and the use of scrap materials, industrial and metallurgical laboratories, etc. The first two of these projected

buildings were in process of completion and their equipment was ready for installation at the outbreak of the war.

The organization and program of the Polytechnic Institute underwent many changes after the rebirth of Poland. In 1919, two other independent educational institutions, the Agricultural Academy in Dublany and the Forestry Academy in Lwów were absorbed and made into the Department of Agricultural and Forestry. To the older



Memorial tablet to "Students of the Lwów Polytechnic who died in the defense of Lwów and their Country in the years 1918-1921."

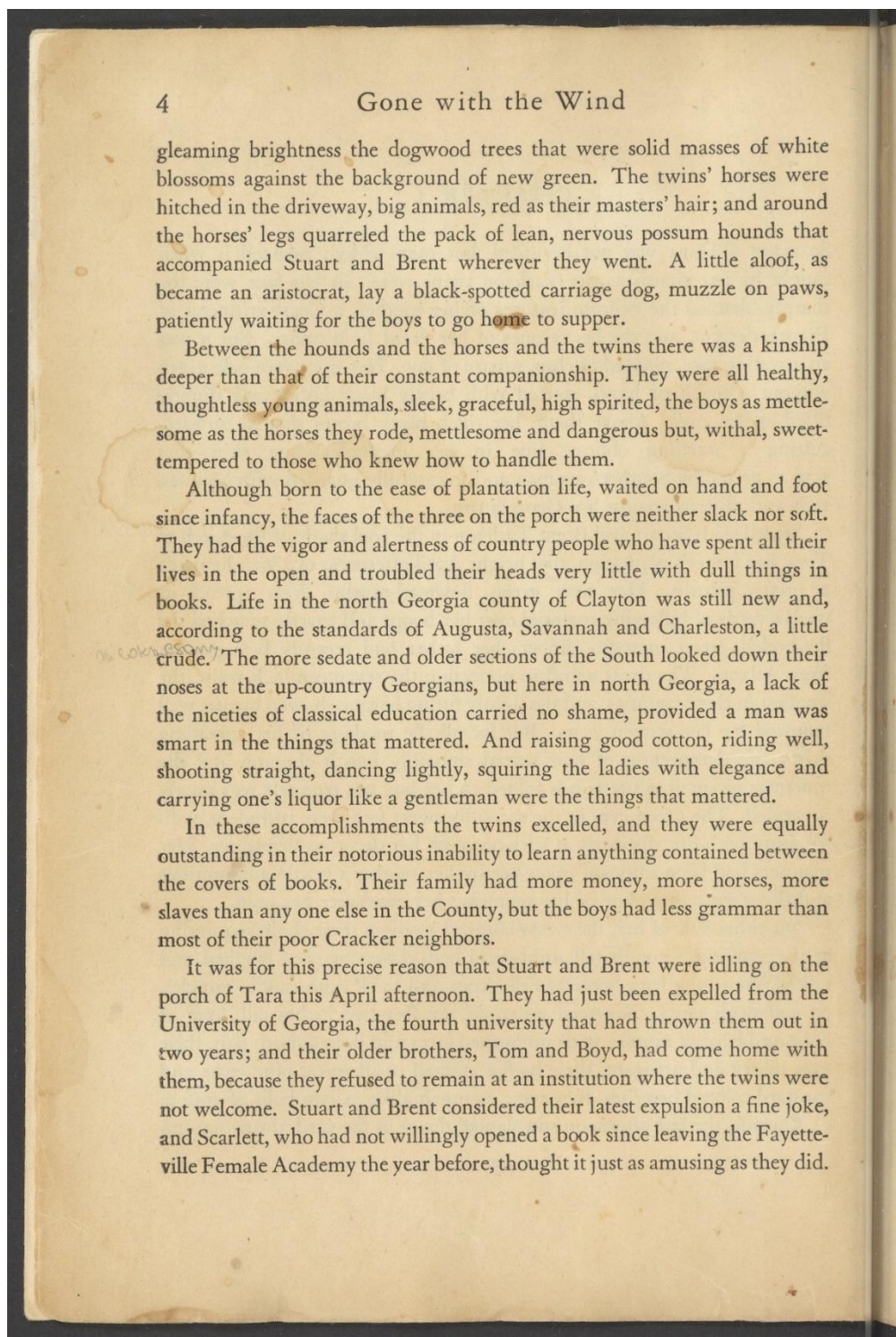
departments of architecture, civil and marine engineering, mechanical engineering, and chemistry was also added the department of military and general engineering. The entire program was brought up to date and enlarged.

At the beginning of the 20th century, the entire department of mechanical engineering was in the hands of three professors, Franek, Jakubowski, and Dykowski. Later they were replaced by Professors Fiedler

(Turn to p. 14)

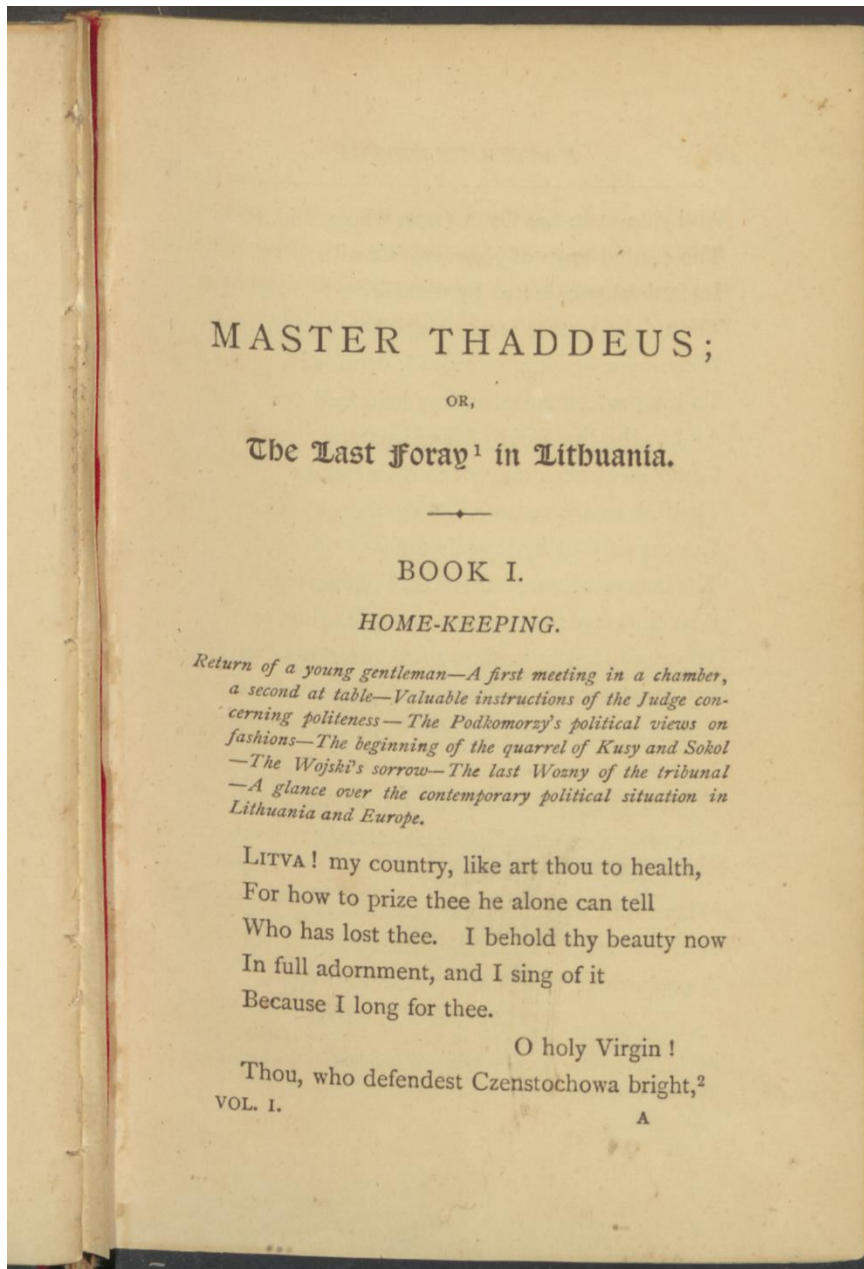
Source: *The Polish Review*. (1944). Vol. 4 no. 28, p. 4-5. Available at: <https://kpbk.umk.pl/dlibra/publication/235389/edition/233504/content>

Scan 8



Source: Mitchell, Margaret. (1936). *Gone with the wind*, p. 4. Available at: <https://kpbk.umk.pl/dlibra/publication/255991/edition/255385/content>

Scan 9



Source: Mickiewicz, Adam. (1885). *Master Thaddeus; or, the Last foray in Lithuania : an historical epic poem in twelve books*. Vol. 1., p. 1. Available at:
<https://kpbc.umk.pl/dlibra/publication/215648/edition/233290/content>

Scan 10

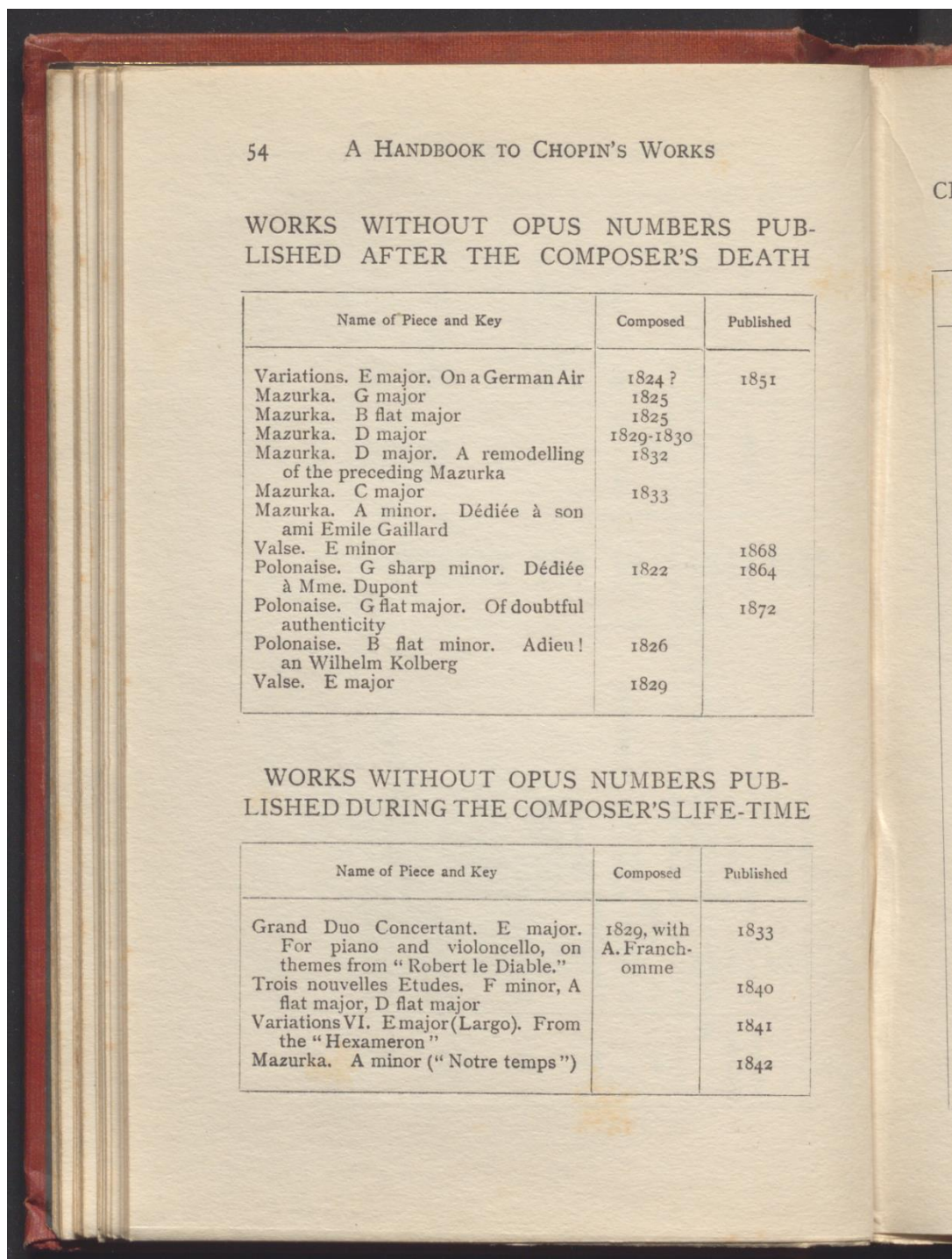
A HANDBOOK TO CHOPIN'S WORKS 45

TABLE OF THE WORKS OF FREDERIC CHOPIN

Opus. No.	Name of Piece and Key	Composed	Published	Dedication.
1	First Rondo in C minor	1828	1825	Mme. de Linde
2	Variations on "Là ci darem," with orchestral accompaniment. B flat major	1829	March 1830	Titus Woyciechowski
3	Introduction et Polonaise brillante, for piano and violoncello. C major	1828	1833	Joseph Merk
4	Sonata in C minor	1828	Posth. 1851	Joseph Elsner
5	Rondo à la Mazurka. F major		1827	Mlle. la Comtesse Alexandrine de Mouriolles
6	Four Mazurkas		1832	Mlle. la Comtesse Pauline Plater
(1st Set)	No. 1. F sharp minor			
	No. 2. C sharp minor			
	No. 3. E major			
	No. 4. E flat minor			
7	Five Mazurkas		Dec. 1832	Mr. Johns
(2nd Set)	No. 1. B flat major (5)			
	No. 2. A minor (6)			
	No. 3. F minor (7)			
	No. 4. A flat major (8)			
	No. 5. C major (9)			
8	Trio for Piano, Violin, and Violoncello. G minor	1828	1833	M. le Prince Antoine Radziwill
9	Three Nocturnes	Probably 1832	Jan. 1833	Mme. Camille Pleyel
	No. 1. B flat minor (Larghetto)			
	No. 2. E flat major (Andante)			
	No. 3. B major (Allegretto)			

Source: Jonson, G. C. Ashton. (1908). *A handbook to Chopin's works : giving a detailed account of all the compositions of Chopin, short analyses for the piano student, and critical quotations from the writings of well-known musical authors : the whole forming a complete guide for concert-goers, pianists and pianola-players, also a short biography, critical bibliography and a chronological list of works, etc.*, p. 45. Available at: <https://kpbc.umk.pl/dlibra/publication/225034/edition/234756/content>

Scan 11



WORKS WITHOUT OPUS NUMBERS PUBLISHED AFTER THE COMPOSER'S DEATH

Name of Piece and Key	Composed	Published
Variations. E major. On a German Air	1824 ?	1851
Mazurka. G major	1825	
Mazurka. B flat major	1825	
Mazurka. D major	1829-1830	
Mazurka. D major. A remodelling of the preceding Mazurka	1832	
Mazurka. C major	1833	
Mazurka. A minor. Dédicée à son ami Emile Gaillard		
Valse. E minor		1868
Polonaise. G sharp minor. Dédicée à Mme. Dupont	1822	1864
Polonaise. G flat major. Of doubtful authenticity		1872
Polonaise. B flat minor. Adieu! an Wilhelm Kolberg	1826	
Valse. E major	1829	

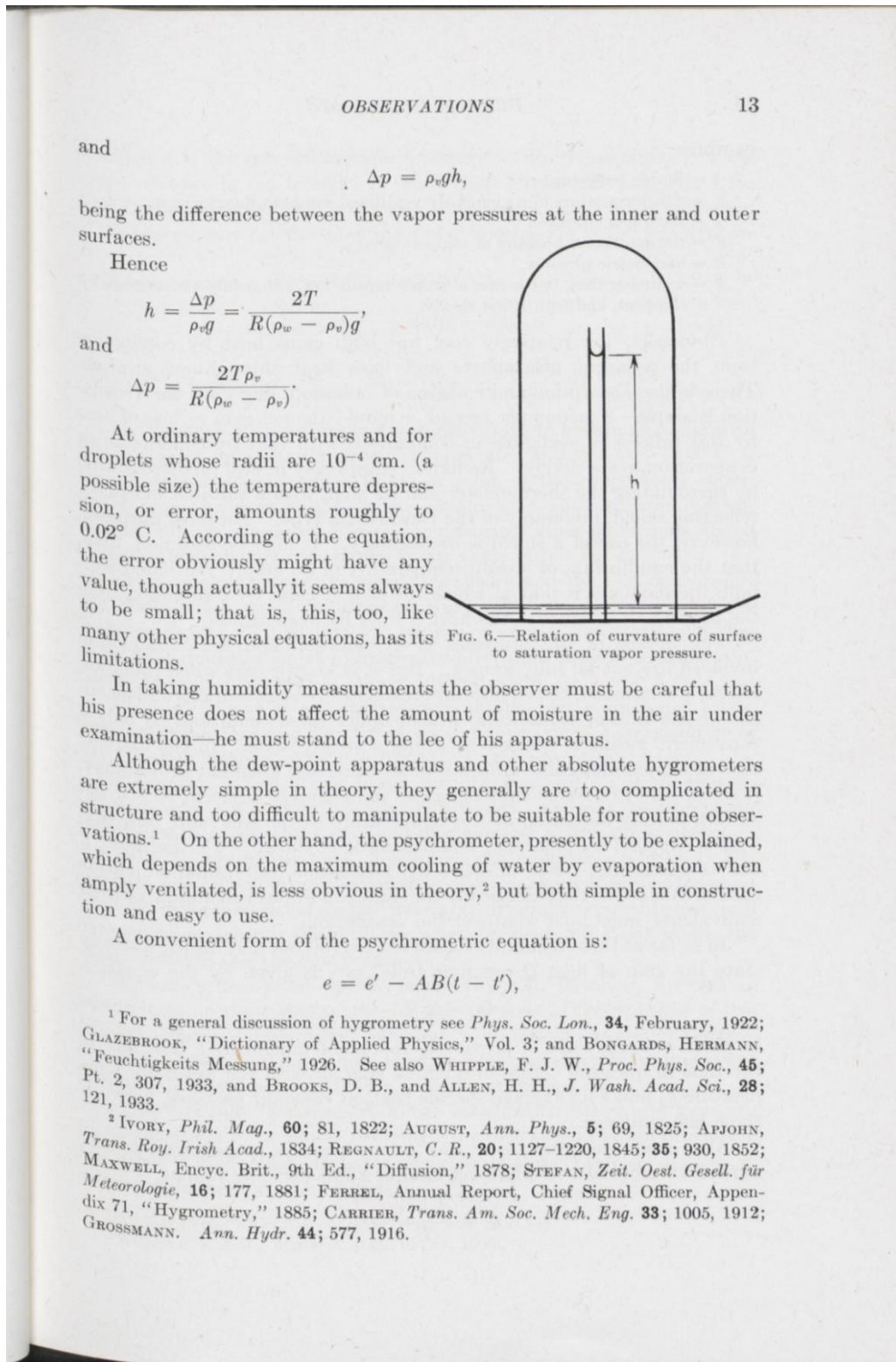
WORKS WITHOUT OPUS NUMBERS PUBLISHED DURING THE COMPOSER'S LIFE-TIME

Name of Piece and Key	Composed	Published
Grand Duo Concertant. E major. For piano and violoncello, on themes from "Robert le Diable."	1829, with A. Franchomme	1833
Trois nouvelles Etudes. F minor, A flat major, D flat major		1840
Variations VI. E major (Largo). From the "Hexameron"		1841
Mazurka. A minor ("Notre temps")		1842

Source: Jonson, G. C. Ashton. (1908). *A handbook to Chopin's works : giving a detailed account of all the compositions of Chopin, short analyses for the piano student, and critical quotations from the writings of well-known musical authors : the whole forming a complete guide for concert-goers, pianists and pianola-players, also a short biography, critical bibliography and a chronological list of works, etc.*, p. 54. Available at:

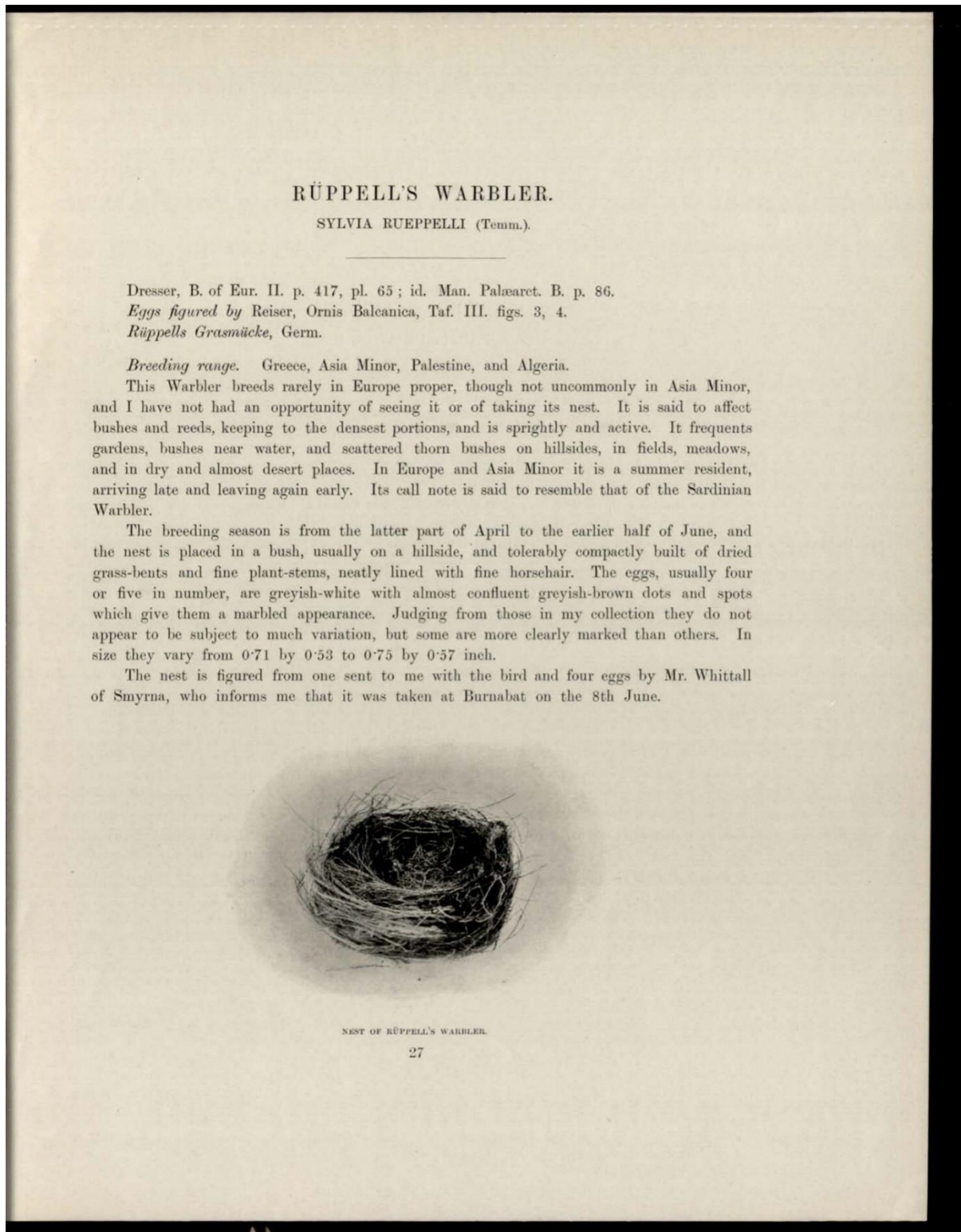
<https://kpbc.umk.pl/dlibra/publication/225034/edition/234756/content>

Scan 12



Source: Humphreys, William Jackson. (1940). *Physics of the air*, p. 13. Available at: <https://kpbc.umk.pl/dlibra/publication/194654/edition/209487/content>

Scan 13



RÜPPELL'S WARBLER.

SYLVIA RUEPPELLI (Temm.).

Dresser, B. of Eur. II. p. 417, pl. 65; id. Man. Palæarct. B. p. 86.

Eggs figured by Reiser, Ornis Balcanica, Taf. III. figs. 3, 4.

Rüppells Grasmücke, Germ.

Breeding range. Greece, Asia Minor, Palestine, and Algeria.

This Warbler breeds rarely in Europe proper, though not uncommonly in Asia Minor, and I have not had an opportunity of seeing it or of taking its nest. It is said to affect bushes and reeds, keeping to the densest portions, and is sprightly and active. It frequents gardens, bushes near water, and scattered thorn bushes on hillsides, in fields, meadows, and in dry and almost desert places. In Europe and Asia Minor it is a summer resident, arriving late and leaving again early. Its call note is said to resemble that of the Sardinian Warbler.

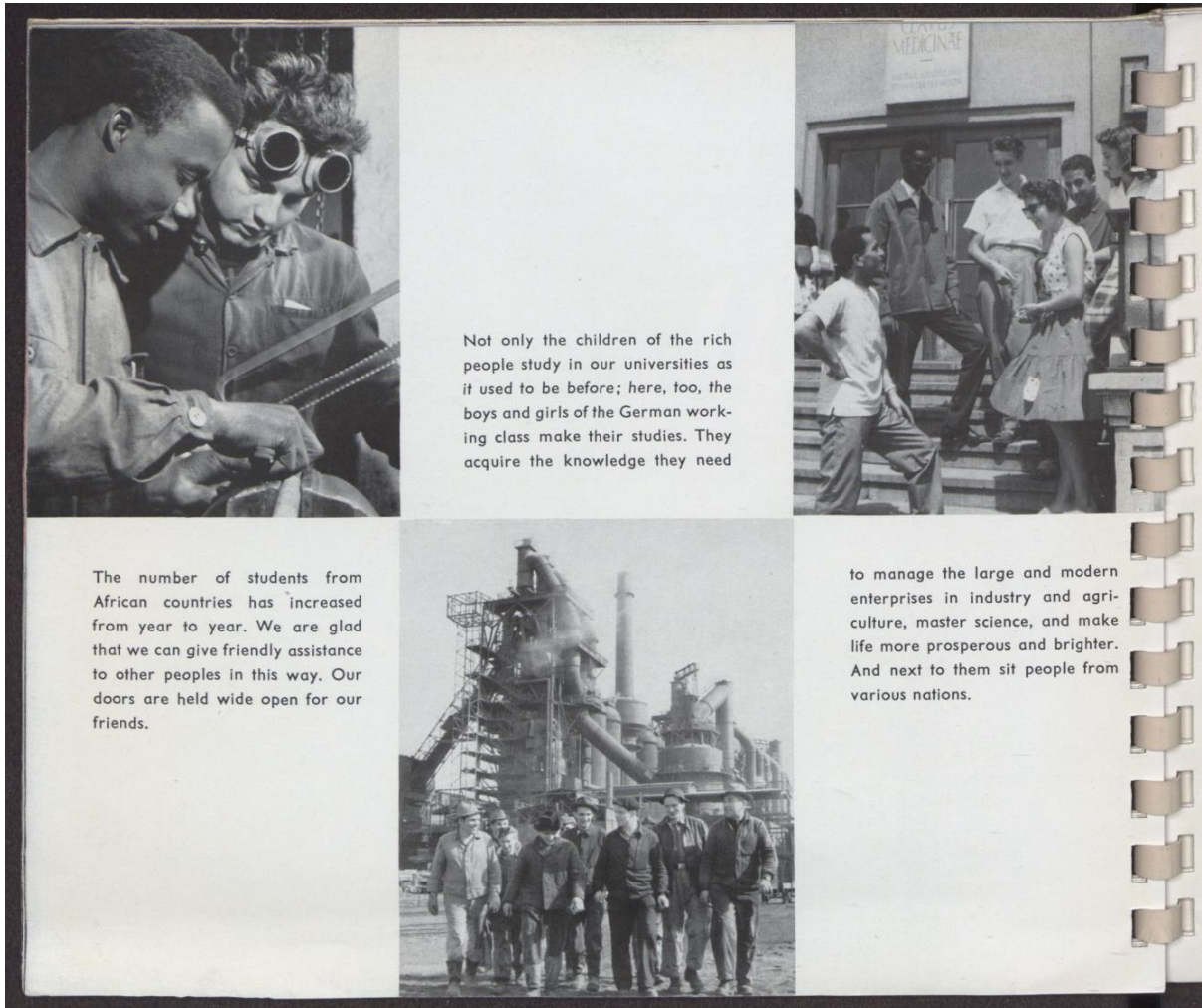
The breeding season is from the latter part of April to the earlier half of June, and the nest is placed in a bush, usually on a hillside, and tolerably compactly built of dried grass-stems and fine plant-stems, neatly lined with fine horsehair. The eggs, usually four or five in number, are greyish-white with almost confluent greyish-brown dots and spots which give them a marbled appearance. Judging from those in my collection they do not appear to be subject to much variation, but some are more clearly marked than others. In size they vary from 0.71 by 0.53 to 0.75 by 0.57 inch.

The nest is figured from one sent to me with the bird and four eggs by Mr. Whittall of Smyrna, who informs me that it was taken at Burnabat on the 8th June.

NEST OF RÜPPELL'S WARBLER.

Source: Dresser, Henry E. (1910). *Eggs of the birds of Europe, Vol. 1*, p. 27. Available at: <https://nbn-resolving.org/urn:nbn:de:bvb:355-ubr11969-2>

Scan 14



Not only the children of the rich people study in our universities as it used to be before; here, too, the boys and girls of the German working class make their studies. They acquire the knowledge they need

The number of students from African countries has increased from year to year. We are glad that we can give friendly assistance to other peoples in this way. Our doors are held wide open for our friends.

to manage the large and modern enterprises in industry and agriculture, master science, and make life more prosperous and brighter. And next to them sit people from various nations.

Source: *German Democratic Republic : International Agricultural Exhibition in Cairo.* (1961). p. [22]. Available at: <https://nbn-resolving.org/urn:nbn:de:bvb:355-ubr10353-8>

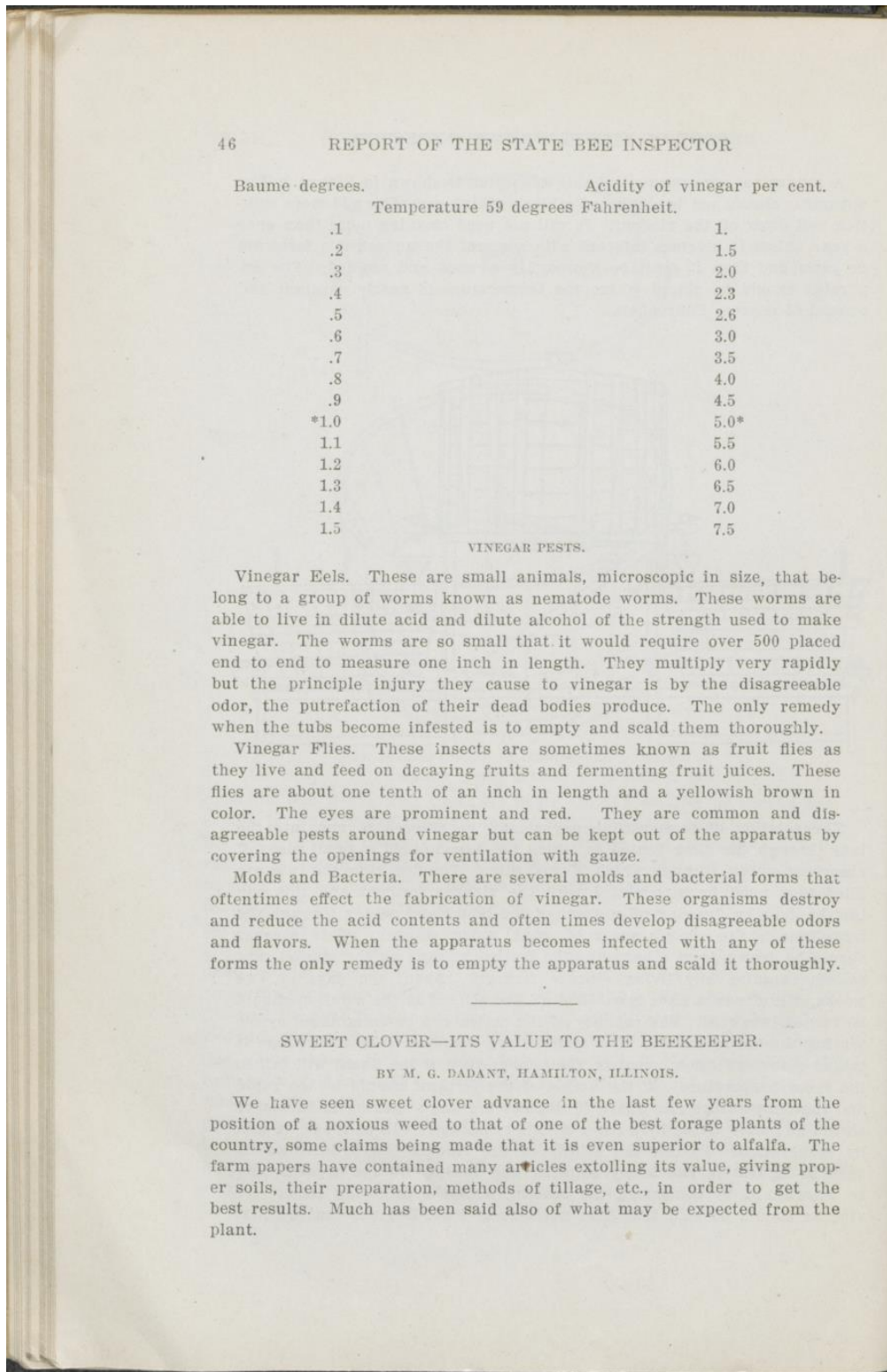
Scan 15

CONTENTS.

	Page.
INTRODUCTION	3—74
CHAPTER I.	
Probabilities of a successful mission from the United States to Japan.—Such a mission proposed to the Government by Commodore Perry.—Expedition resolved on.—Vessels selected for the squadron.—Vexatious delay in their equipment.—Commodore Perry sent in the Mississippi to the Gulf of St. Lawrence.—Applications of scientific men to join the expedition refused.—Causes of such refusal.—Dr. Von Siebold.—Visit of the President and Secretary of the Navy to Annapolis to take leave of the expedition.—Failure of the machinery of the steamship Princeton on the passage down Chesapeake Bay.—Substitution of steamer Powhatan in place of the Princeton.—Final departure of Commodore Perry on the mission with the Mississippi alone..	75—80
CHAPTER II.	
Voyage from the capes of the Chesapeake to Madeira.—View of the island.—Funchal.—Hospitality of the inhabitants.—Salubrity of the climate.—Exports of the island.—Novel mode of conveyance.—Departure from Madeira and arrival at the Canaries.—Early failure of northeast trades.—Extraordinary swell from the northwest.—General order as to private journals and communications to public prints.—General order as to scientific investigations by officers.—The “Harmattan,” consideration of hypotheses as to its origin.—Southeast trades.—Ship steered for St. Helena.—Observations on the currents.—Chaplain’s observations on the zodiacal lights.—Arrival at St. Helena.—Description of the island.—Jamestown.—Longwood.—Tomb of Napoleon.—The calculating hospitality of the inhabitants of the island.—Adventure of Lieutenant —.—Fortifications of the island.—Their sufficiency against sailing vessels.—Probable insufficiency against an approach from the west by steam.—Departure from St. Helena	81—96
CHAPTER III.	
Passage to the cape.—Fuel for steamers.—Table rock and Cape Town.—Description of Cape Town.—Climate.—Annoyance from dust.—Violence of winds and difficulty of holding to anchorage.—Supplies at the cape.—Caffre war and its effects.—Mode of transporting produce on land.—Vineyards of Constantia.—Effects of emancipation of slaves on agricultural labor.—Mode of cultivating the vine.—Population of Cape Colony.—Bushmen.—The Caffres.—Physical characteristics.—Fingoes.—Military organization of Caffres.—Condition of the emancipated slaves.—Departure from Table Bay.—Passage to, and arrival at, Mauritius.—Harbor of Port Louis.—Dangers of the harbor.—Skill of port officers in mooring vessels.....	97—106
CHAPTER IV.	
Mauritius, its discovery.—Geological formation and physical aspect.—Production of sugar.—Effect on agriculture of the abolition of slavery.—Coolies.—Population of the island.—State of feeling between English and French residents.—Hospitable treatment of the expedition.—Description of Port Louis.—Grand Port.—Paul and Virginia.—Facts on which St. Pierre founded his story.—Tombs of Paul and Virginia.—Built by an eccentric Frenchman.—Cyclones.—Their probable causes.—Interest felt in them at Mauritius.—Departure of the Mississippi from Port Louis.—Her course thence to Point de Galle, island of Ceylon.—Reasons for taking it.—Point de Galle, description of.—Great rendezvous of steamers.—Difficulty of procuring fuel there.—American consul.—Thoughts on consular system.—Early knowledge of Ceylon.—Its several European possessors.—Climate.—Salubrity.—Causes of its diminished prosperity.—Productions.—Value of cocoonut palm.—Pearl fishery.—Immense numbers of elephants.—Great slaughter of them.—Boa constrictor.—Population of Ceylon.—Physique of Cingalese, Malabars, and Mahomedaus in the island.—Religious condition.—Buddhism.—Pilgrimage to the temples.—Intercourse with a Siamese naval officer at Ceylon.—Commodore’s letter to the second king of Siam.—Departure from Ceylon.—Passage through the Straits of Malacca.—Arrival at Singapore.....	107—124

Source: Hawks, Francis L. & Perry, Matthew Calbraith. (1967). *Narrative of the expedition of an American squadron to the China Seas and Japan : Performed in the years 1852, 1853, and 1854, under the command of Commodore M. C. Perry, United States Navy, by order of the Government of the United States*, p. [VII]. Available at: <https://nbn-resolving.org/urn:nbn:de:bvb:355-ubr07179-4>

Scan 16



Source: *Annual report of the State Bee Inspector /4. 1915. (1916)*, p. 46. Available at: <https://nbn-resolving.org/urn:nbn:de:bvb:355-ubr15095-5>

Annex 2: Conversion report for Calibre

1. Were there any changes to the input file before the conversions?

No, except BNP's separate DOCX files were joined in one single file as it is easier to do conversion with a single file.

2. How long did it take to receive output results?

Few minutes.

3. Is it possible to convert multiple files simultaneously?

Yes, but has separate processes and you need to apply the settings each time.

4. Explain the workflow.

We inserted the input formats into the program. Next, we selected "convert books" from the top main menu. When the settings window appeared, we didn't make any changes to the settings; instead, we only modified the output format on the top right side during each conversion process (refer to Figure 1).

As there were no alternative options for input formats, we proceeded with the following conversions:

- For PDF and PDF/UA we converted to: DOCX, EPUB, RTF
- For DOCX we converted to: EPUB, PDF, RTF
- For RTF we converted to: DOCX, EPUB, PDF
- For EPUB we converted to: DOCX, PDF, RTF

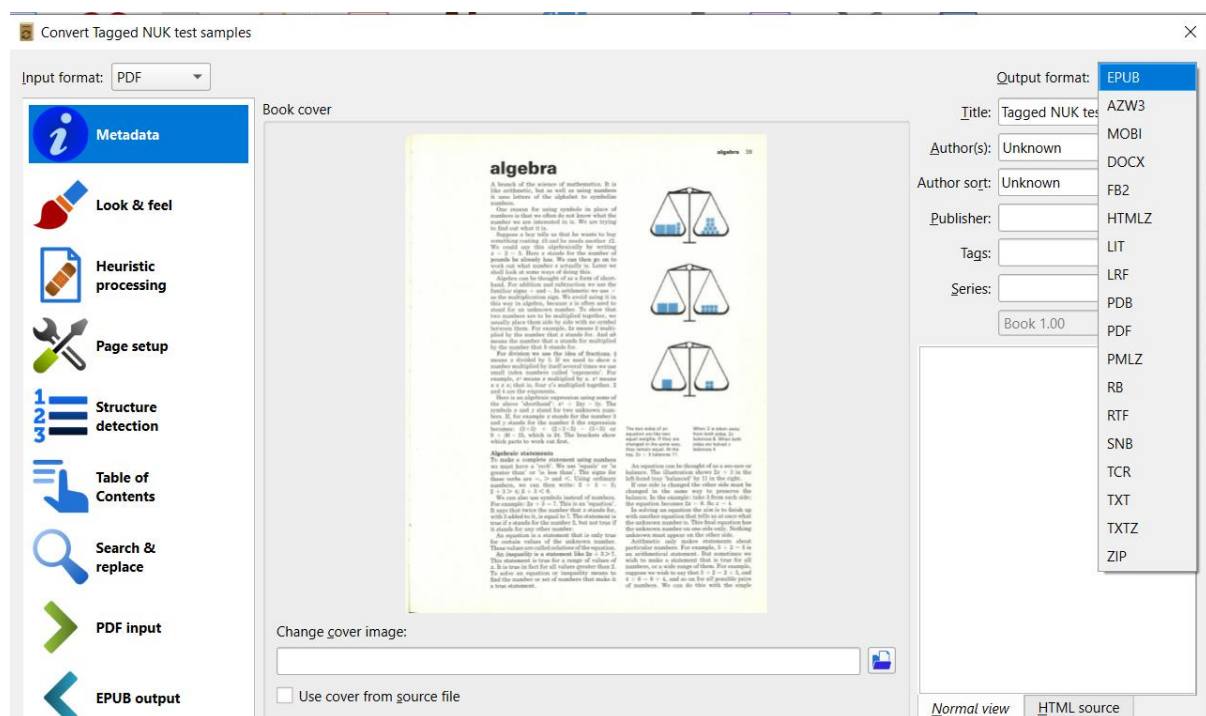


Figure 12: Settings window in Calibre

Annex 3: Conversion report for RoboBraille

1. Were there any changes to the input file before the conversions?

No, except that BNP's separate DOCX files were combined into a single file for easier conversion.

How long did it take to receive the output results?

Most of the results were returned within a day. Some results were even quicker, while others took at least an hour. Some results didn't arrive at all. The entire process is time-consuming because each conversion required a separate upload of the input file.

2. Is it possible to convert multiple files simultaneously?

No.

3. Explain the workflow.

There were different options for input formats, so we utilized various workflows as described below. The input formats were converted as follows:

- a) For PDF and PDF/UA, we converted to: DOCX, EPUB, PDF tagged image over text, RTF, and audio: MP3.
- b) For DOCX, we converted to: EPUB, PDF, RTF, and audio: MP3, DAISY full text and audio, DAISY math full text and audio.
- c) For RTF, we converted to: EPUB and audio: MP3, DAISY full text and audio.
- d) For EPUB we converted to: RTF

Additional notice: When we received audio formats, there was special note: "Please note that the e-book is made strictly for personal, non-commercial use. You are not allowed to distribute the e-book or to use it for commercial purposes." This could potentially pose issues if a library were to use this service.

A. Workflow for PDF and PDF/UA

The first step was to upload the file (refer to Figure 2). In the second step, we selected the target format, which was either "MP3 audio," "braille," "e-book," or "accessibility conversion" (refer to Figure 3). At step three, we chose additional options, either the target output format or audio settings (refer to Figure 4-6).

Alternate media made easy

Follow the four easy steps below to have your document converted into an alternative, accessible format. To upload a file, enter a URL to a file or simply type in the text you wish to have converted. The first step is to upload your document.

Source

File

URL

Text

Step 1 - Upload your document

Select your file and upload it to the server (max 64 MB). Multiple files of the same type may be uploaded. Supported file formats: .HTML, .HTM, .RTF, .EPUB, .MOBI, .TIFF, .TIF, .GIF, .JPG, .JPEG, .BMP, .PNG, .PCX, .DCX, .J2K, .

File name: NUK_original.pdf

Figure 13: First step: uploading input PDF format in RoboBraille

Step 2 - Select target format

Specify the target format of your document. For more information about formats, please visit [target formats](#). For the uploaded document type, the following formats are available:

Target format

MP3 audio

Braille

E-book

Accessibility conversion

Figure 14: Second step: selecting target conversion format in RoboBraille when converting PDF

Step 2 - Select target format

Specify the target format of your document. For more information about formats, please visit [target formats](#). For the uploaded document type, the following formats are available:

Target format

MP3 audio

Braille

E-book

Accessibility conversion

Step 3 - Specify audio options

Specify the natural language of your document and how fast you want the speech. For more information about MP3 audio, please visit [MP3 audio features](#).

Options

Language:

Speed:

Figure 15: Third step: specifying additional options for “MP3 audio” conversion in RoboBraille when converting PDF

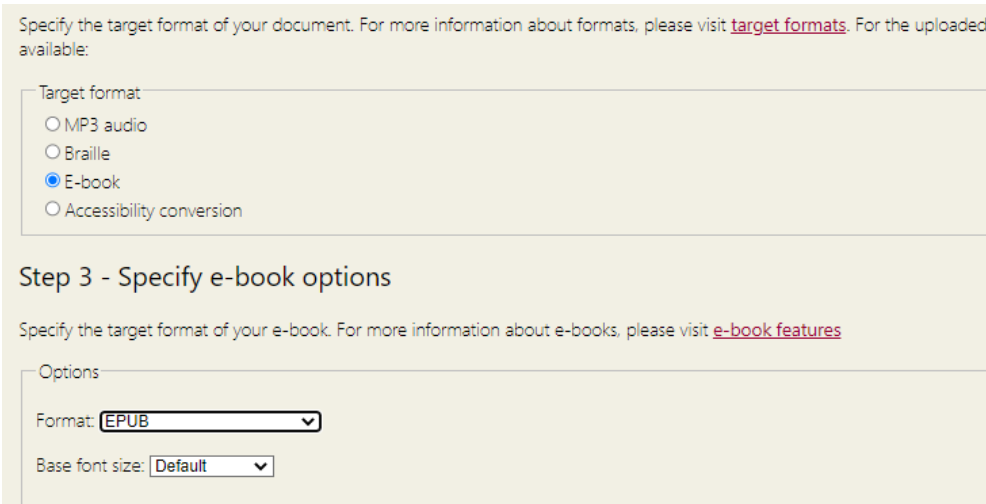


Figure 16: Third step: specifying additional options for “E-book” conversion in RoboBraille when converting PDF

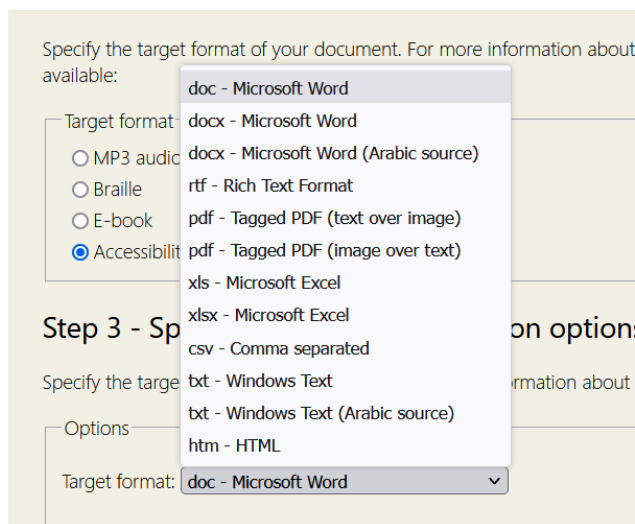


Figure 17: Third step: specifying additional options for “Accessibility conversion” conversion in RoboBraille when converting PDF

B. WORKFLOW IMAGES FOR DOCX

The process was identical to the workflow for PDF, with the only difference being the availability of additional conversion options and a broader range of output formats (including DAISY book options, which were not present for PDF). In this case, during the second step, we selected the target format from the following options: “e-book”, “accessibility conversion”, “MP3 audio”, “DAISY full text and audio” and “DAISY math full text and audio” (refer to Figure 7). In the third step, we chose additional options, including the target output format or audio settings, following the same procedure as with PDF (refer to Figure 8-12).

Target format

- MP3 audio
- DAISY full text and audio
- DAISY Math full text and audio
- Braille
- E-book
- Accessibility conversion

Figure 18: Target format options in RoboBraille when converting DOCX format.

Step 2 - Select target format

Specify the target format of your document. For more information about formats, please visit [target formats](#). For the uploaded document type, the following formats are available:

Target format

- MP3 audio
- DAISY full text and audio
- DAISY Math full text and audio
- Braille
- E-book
- Accessibility conversion

Options

- American English
- Arabic
- Arabic English Bilingual
- British English
- Bulgarian
- Chinese (Cantonese)
- Chinese (Mandarin)
- Chinese (Taiwanese)
- Czech (female, Iva)
- Czech (female, Zuzana)
- Danish (female, Sara)
- Danish (male, Carsten)
- Dutch (female)
- Dutch (male)
- Finnish
- French
- German
- Greek
- Greenlandic
- Hungarian (female)

Language:

Speed:

Step 3 - Specify the natural language of the user reply. For more information about MP3 audio, please visit [MP3 audio features](#)

Options

Language:

Figure 19: Specifying additional options for “MP3 audio” conversion options in RoboBraille when converting DOCX

Target format

- MP3 audio
- DAISY full text and audio
- DAISY Math full text and audio
- Braille
- E-book
- Accessibility conversion

Step 3 - Specify DAISY full text and audio options

Specify the natural language of the user reply. For more information about DAISY, please visit [DAISY features](#)

Options

Language:

Step 4 - Specify email address and submit request

Email address:

Figure 20: Specifying additional options for “DAISY full text and audio” conversion options in RoboBraille when converting DOCX

Target format:

- MP3 audio
- DAISY full text and audio
- DAISY Math full text and audio
- Braille
- E-book
- Accessibility conversion

Step 3 - Specify DAISY Math full text and audio options

Specify the natural language of the user reply. For more information about DAISY, please visit [DAISY features](#)

Options

Language: English Danish English German Polish

Step 4 - email address and submit request

Figure 21: Specifying additional options for “DAISY Math full text and audio” conversion options in RoboBraille when converting DOCX

Target format:

- MP3 audio
- DAISY full text and audio
- DAISY Math full text and audio
- Braille
- E-book
- Accessibility conversion

Step 3 - Specify e-book options

Specify the target format of your e-book. For more information about e-books, please visit [e-book features](#)

Options

Format: EPUB EPUB EPUB3 with media overlay EPUB3 MOBI (Kindle)

Base for

Figure 22: Specifying additional options for “E-book” conversion options in RoboBraille when converting DOCX

- E-book
- Accessibility conversion

Step 3 - Specify accessibility conversion options

Specify the target format of the conversion. For more information about accessibility conversion, please visit [accessibility conversion features](#)

Options

Target format: txt - Windows Text txt - Windows Text pdf - Tagged PDF

Figure 23: Specifying additional options for “Accessibility conversion” conversion options in RoboBraille when converting DOCX

C. WORKFLOW IMAGES FOR RTF

The process mirrored the workflow for PDF, with the sole distinction being the inclusion of extra conversion options and a wider array of output formats (notably, DAISY book options that were absent in the case of PDF). In this instance, during the second step, we selected the target format, which could be either “e-book”, “accessibility conversion”, “MP3 audio” or “DAISY full text and audio” (refer to Figure 13). In the third step, we opted for additional choices, such as the target output format or audio settings, following the same approach as we did with PDF (e.g., as shown in Figure 14).

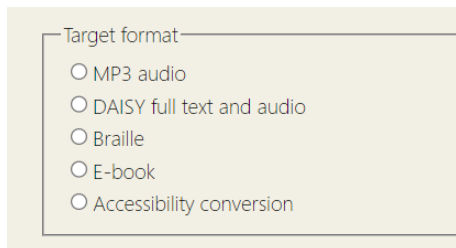


Figure 24: Target format options in RoboBraille when converting RTF format

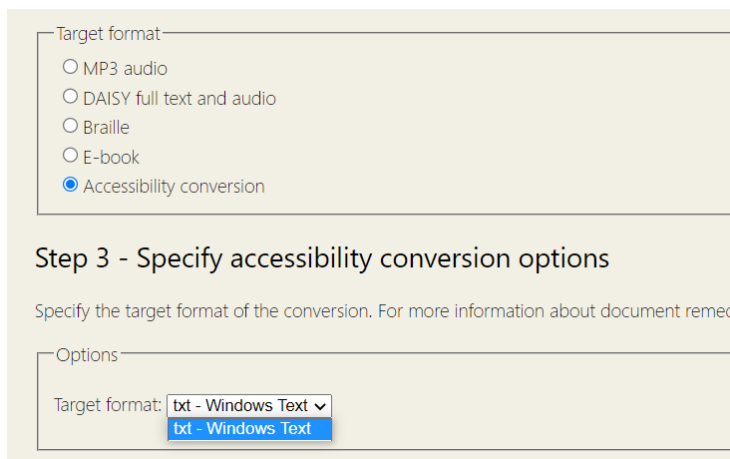


Figure 25: Specifying additional options for “Accessibility conversion” conversion options in RoboBraille when converting RTF format

D. WORKFLOW FOR EPUB

The process closely resembled the workflow for PDF, with the main distinction being that the system provided fewer options. During the second step, we selected the target format which was limited to either “e-book” or “accessibility conversion”, as no additional options were available (refer to Figure 15). Furthermore, these two options provided limited choices for the output format (refer to Figure 16). Only MOBI, RTF and TXT formats were enabled, and we selected the RTF format. In the third step, we made additional selections pertaining to the target output format.

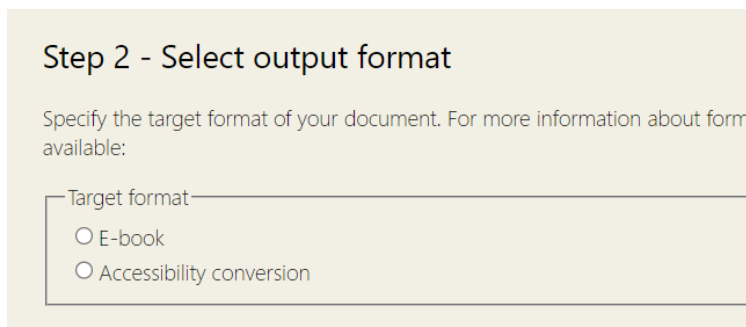


Figure 26: Target format options in RoboBraille when converting EPUB format

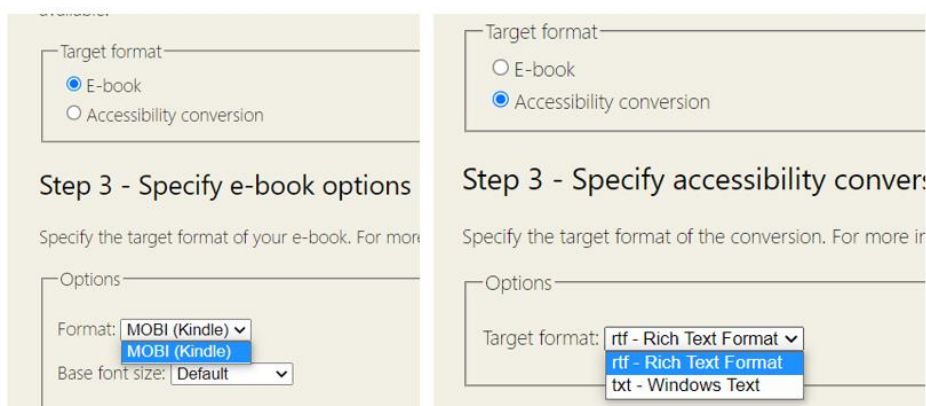


Figure 27: Specifying additional options for “E-book” and “Accessibility conversion” conversion options in RoboBraille when converting EPUB format

Annex 4: Conversion report for Balabolka

1. Were there any changes to the input file before the conversions?

The separate DOCX files from BNP were merged into a single file, simplifying the conversion process. Additionally, minor adjustments were made within the software to facilitate splitting of audio files according to each chapter, including the addition of two empty lines in succession.

2. How long did it take to receive the output results?

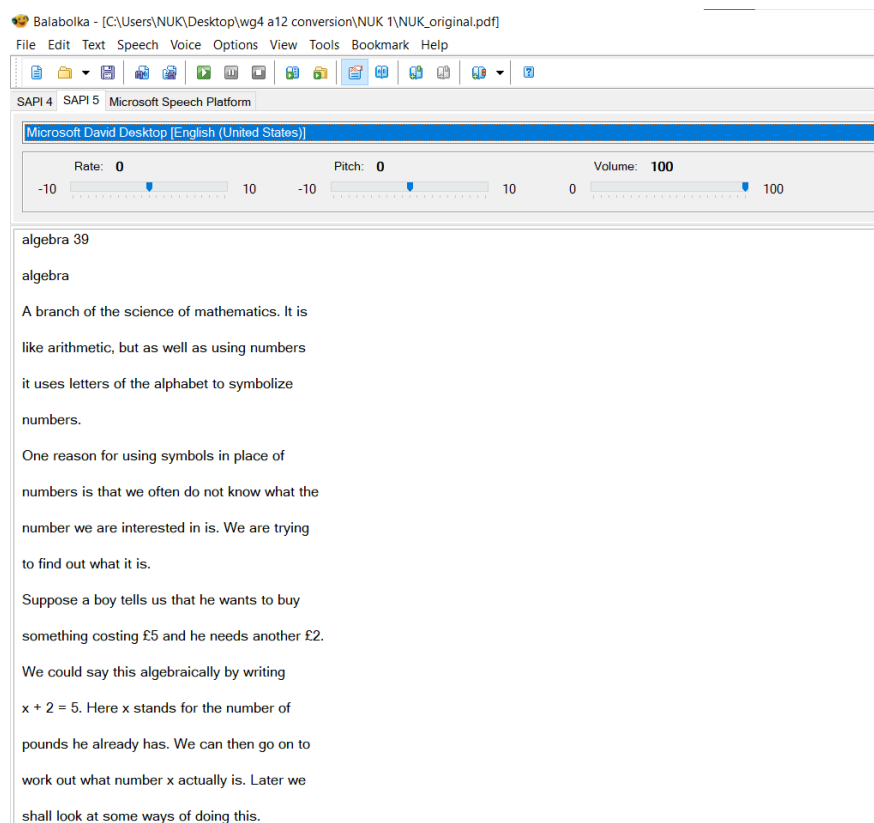
It took approximately one minute after implementing the mentioned changes.

3. Is it possible to convert multiple files simultaneously?

No.

4. Explain the workflow.

The format was inputted into the software, and to ensure proper chapter separation, two empty lines were added before each chapter. The English voice was selected, and no other settings were modified (refer to Figure 17). The process of splitting the text and converting it to audio was initiated, with MP3 chosen as the output format (refer to Figures 18-19). After clicking “split and convert to audio files,” a preview of how the text would be split was provided, offering a final opportunity to correct any errors (refer to Figure 20). Finally, the files were delivered to the designated folder destination.



Balabolka - [C:\Users\NUK\Desktop\wg4 a12 conversion\NUK 1\NUK_original.pdf]

File Edit Text Speech Voice Options View Tools Bookmark Help

SAPI 4 SAPI 5 Microsoft Speech Platform

Microsoft David Desktop [English (United States)]

Rate: 0 Pitch: 0 Volume: 100

algebra 39

algebra

A branch of the science of mathematics. It is like arithmetic, but as well as using numbers it uses letters of the alphabet to symbolize numbers.

One reason for using symbols in place of numbers is that we often do not know what the number we are interested in is. We are trying to find out what it is.

Suppose a boy tells us that he wants to buy something costing £5 and he needs another £2.

We could say this algebraically by writing $x + 2 = 5$. Here x stands for the number of pounds he already has. We can then go on to work out what number x actually is. Later we shall look at some ways of doing this.

Figure 28: Interface of Balabolka after the text has been inserted and audio language was chosen.

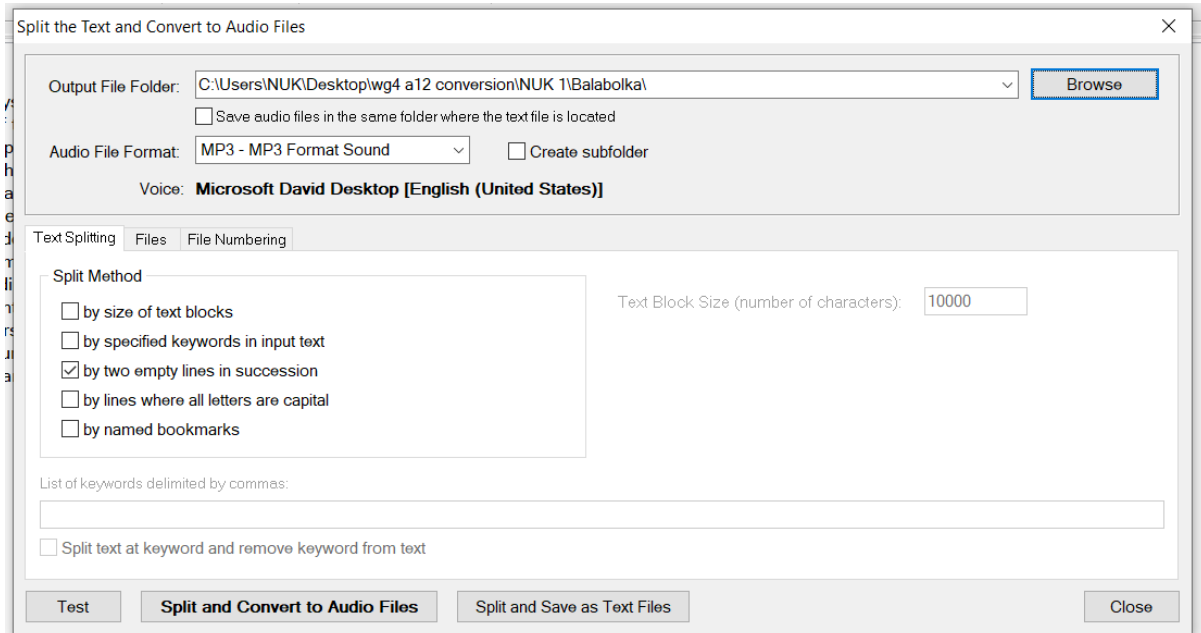


Figure 29: First step of splitting the audio files in Balabolka, according to two empty lines in succession and chosen output format MP3 on top.

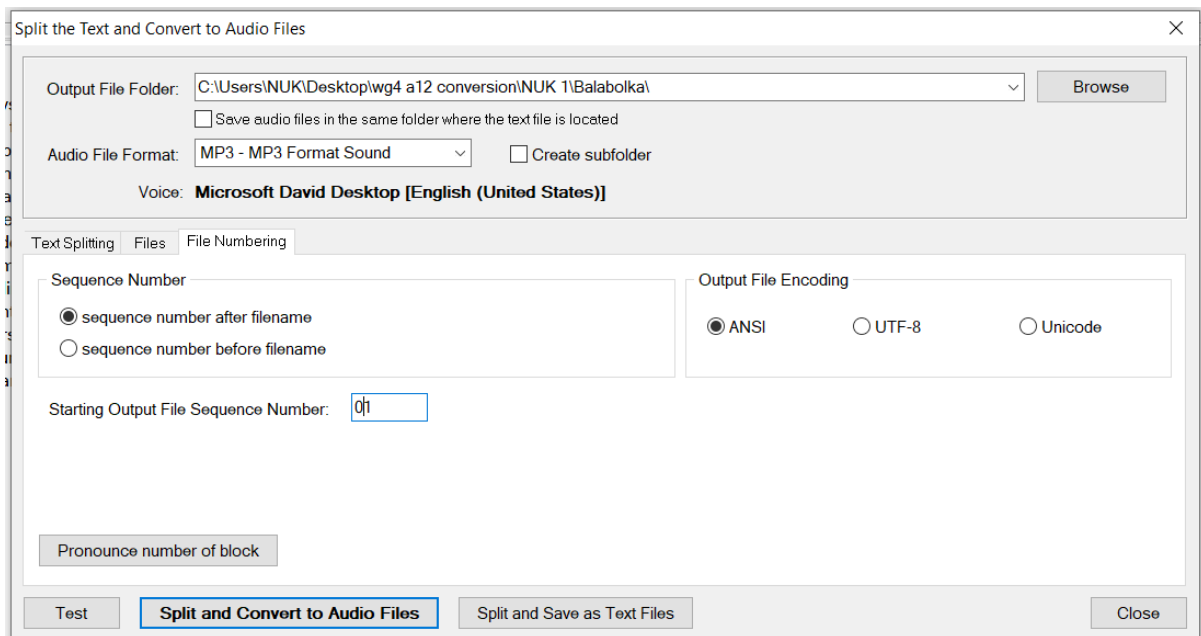


Figure 30: Additional step of splitting the audio files in Balabolka, setting the starting sequence number.

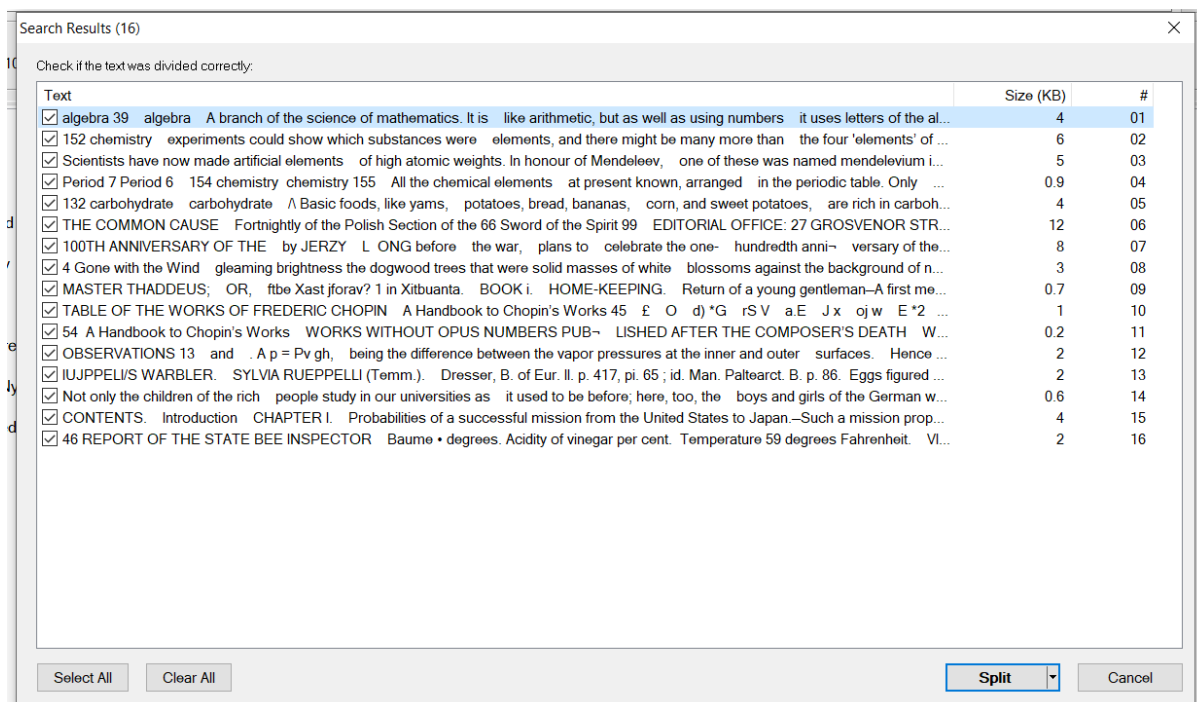


Figure 31: The preview of the audio files that will be created in Balabolka with option “split and convert audio files”

Annex 5: Conversion report for Mathpix

1. Were there any changes to the input file before the conversions?

No. We could only use it for the PDF formats.

2. How long did it take to receive the output results?

It took just a few minutes.

3. Is it possible to convert multiple files simultaneously?

No.

4. Explain the workflow.

For the testing we used Mathpix’s “snip web” version with an educational subscription, allowing for up to do 100 image snips and up to 35 PDF pages. Currently, this service is offered free of charge. We imported the PDF files and then exported them in DOCX format, using MathML as the language for mathematical elements (refer to Figures 21-22). The DOCX format is automatically downloaded from the web browser.

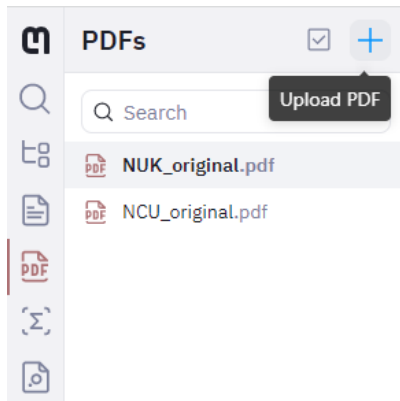


Figure 32: Interface of Mathpix snip web after the two PDFs were imported

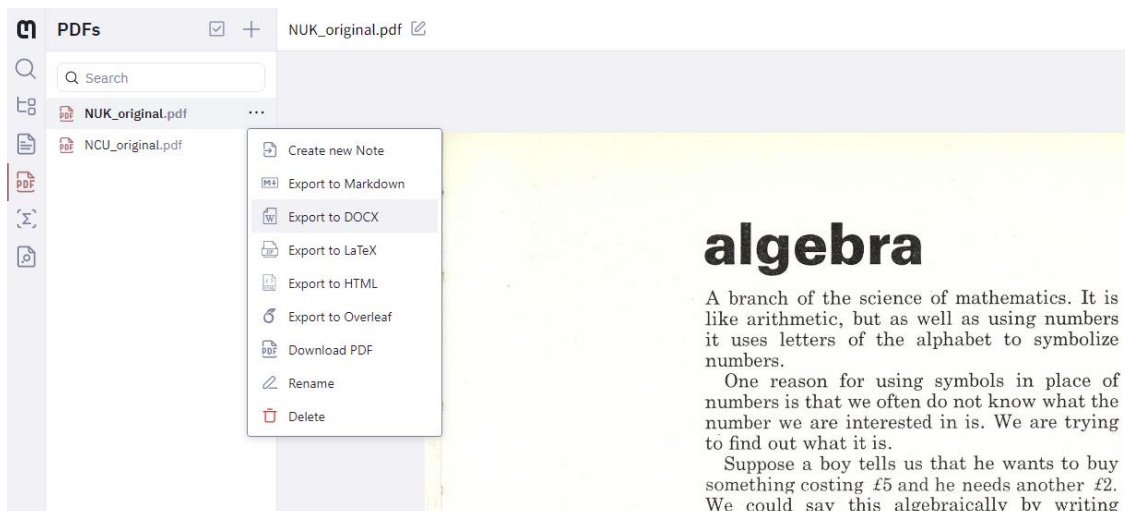


Figure 33: Interface of Mathpix snip web when chosen the option to export as DOCX format

Annex 6: Conversion report for WordToEpub

1. Were there any changes to the input file before the conversions?

The separate DOCX files from BNP were merged into a single file, simplifying the conversion process. Additionally, UIBK's RTF format was opened in Microsoft Office Word and saved as DOCX format to provide an additional sample to test.

2. How long did it take to receive the output results?

We received the output within a minute or two after configuring all the basic settings, which took approximately five minutes.

3. Is it possible to convert multiple files simultaneously?

Yes, batch conversions are supported.

4. Explain the workflow.

After inserting the file, we did not add any metadata (refer to Figure 23) and primarily used general settings (refer to Figures 24–32). For enhanced accessibility, adding metadata and adjusting settings based on the input file would be beneficial. In both cases, we did not include a cover image. Here are the settings we configured and our observations:

- For the BNP's DOCX, we were unable to set the original page numbers, so we opted to use page number from Microsoft Word to provide some navigation (refer to Figure 26). The heading structure appeared to be acceptable without warnings (refer to Figure 30). In cases involving math, MathML was detected in the document, and we selected the option to continue using MathML for math presentation (refer to Figure 29).
- Regarding UIBK's RTF converted to DOCX, it contained "orig. page #" which we selected as a marker for the original pages (refer to Figure 27). There seemed to be an issue with the headings as none were marked except for one heading 3, but we proceeded with the conversion regardless (refer to Figure 31)

In both cases, during the conversion process, we received a warning about merged table cells that may not export well to EPUB.

We followed the same process to export in HTML format, and the only warning we received was also related to merged table cells (refer to Figures 33-34).

For utilizing a special feature of the tool called EPUBToWord, we followed these steps:

- In the file explorer, we right-clicked on the file we wanted to convert and selected "show more options." Once new options appeared, we chose "Convert with EPUBToWord" (refer to Figure 35).
- The interface window opened with basic options for converting EPUB to DOCX file, and we clicked "Create Word document" (refer to Figure 36).
- The conversion process would have been completed at this stage, but we encountered an error message regarding codes for footnotes. We clicked OK, and the conversion proceeded regardless (refer to Figure 37). Although footnote problem could have been manually corrected, it was not done for testing purposes.

- The resulting file was delivered to the designated folder.

A. Workflow for creating EPUB format

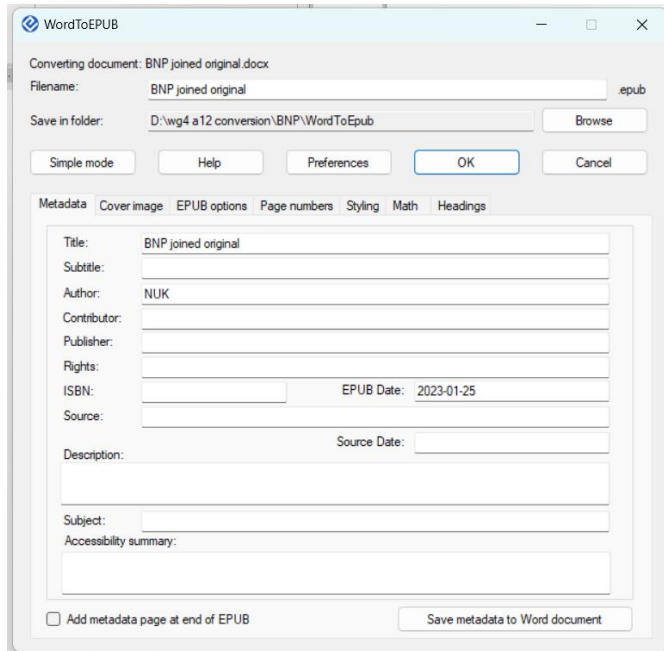


Figure 34: Interface of WordToEpub tool after the DOCX format was opened

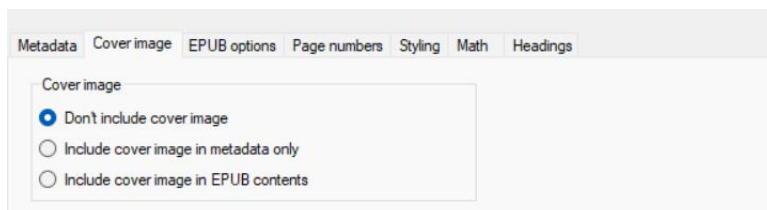


Figure 35: Interface of WordToEpub tool about settings for the cover image

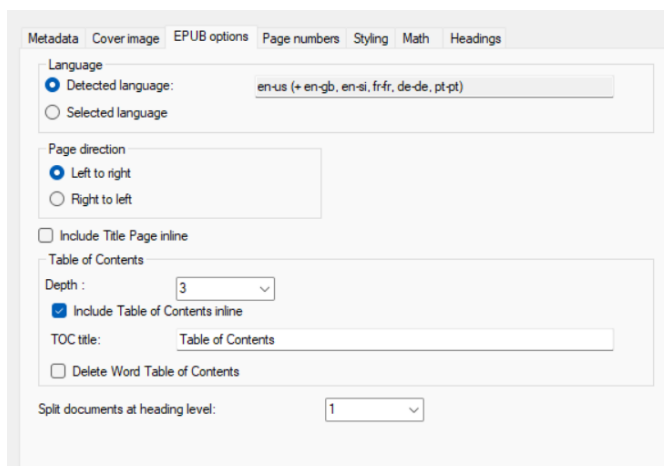


Figure 36: Interface of WordToEpub tool about settings for EPUB

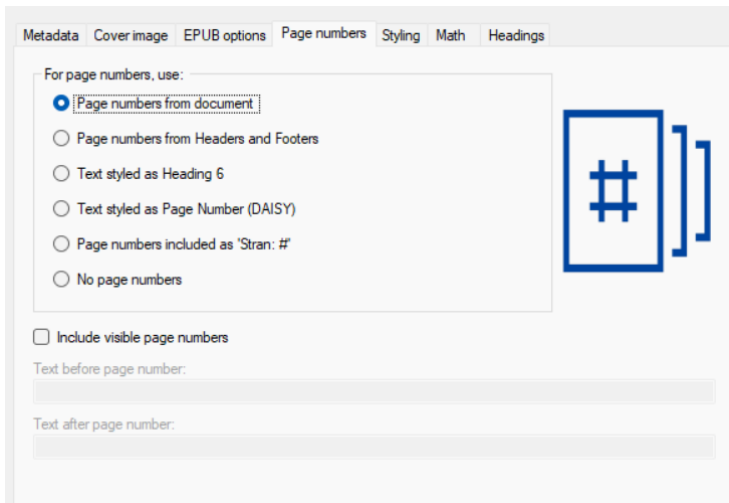


Figure 37: Interface of WordToEpub tool about settings for page numbers in BNP's DOCX format

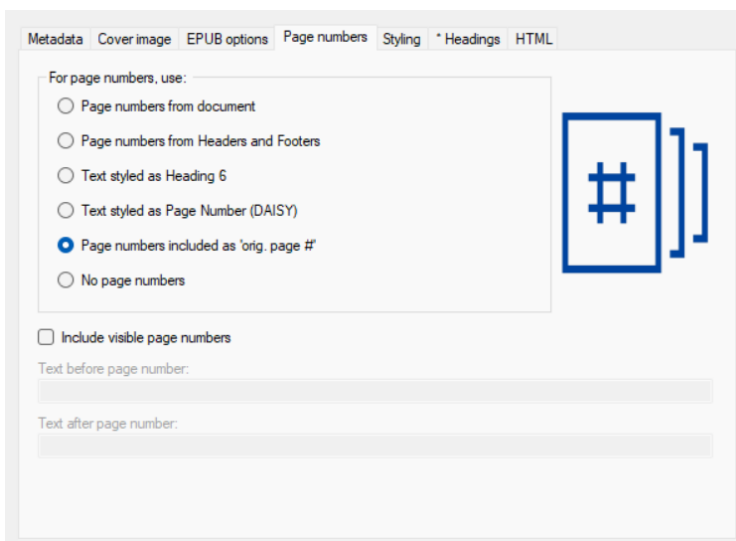


Figure 38: Interface of WordToEpub tool about settings for page numbers in UIBK's RTF turned to DOCX format

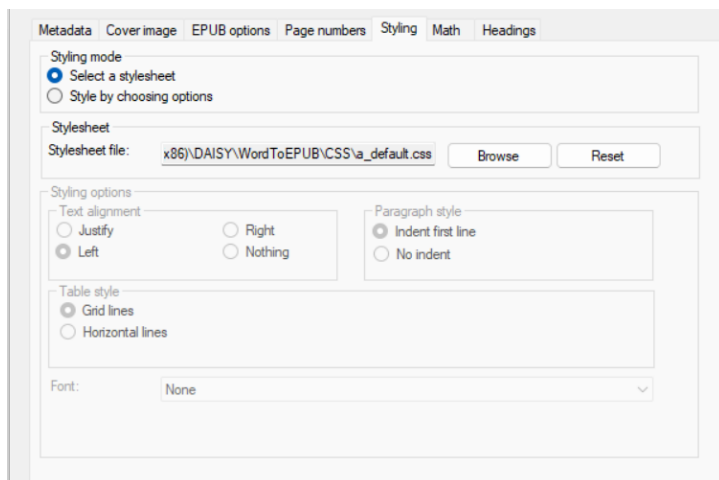


Figure 39: Interface of WordToEpub tool about settings for styling

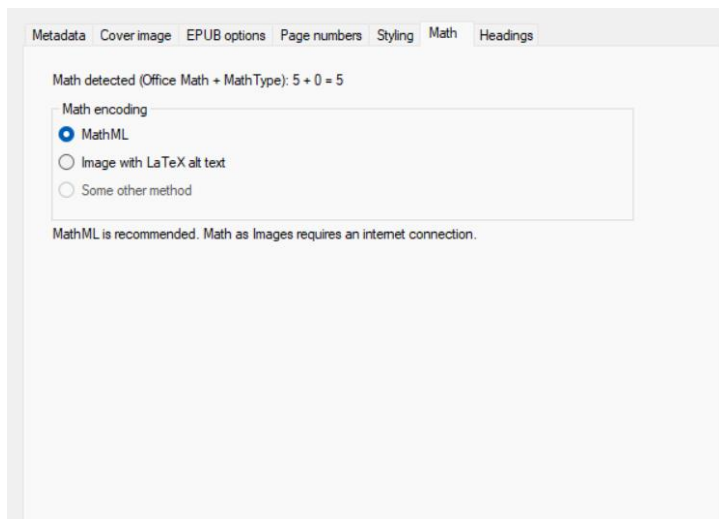


Figure 40: Interface of WordToEpub tool about settings for math where MathML option is chosen

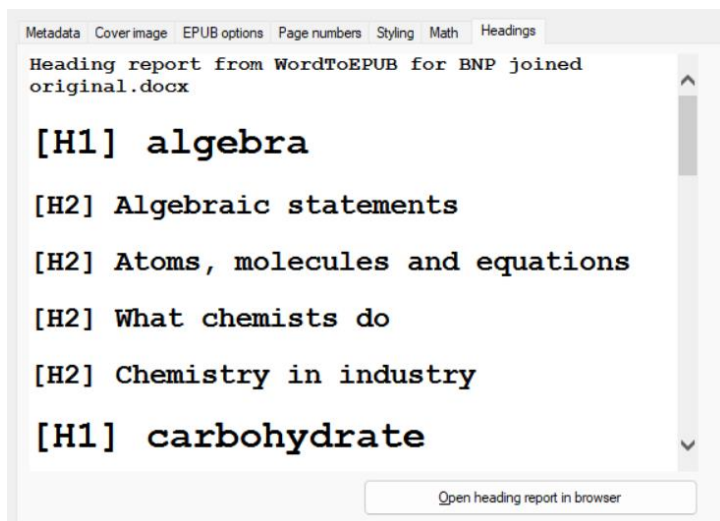


Figure 41: Interface of WordToEpub tool with preview of headings structure without any warnings

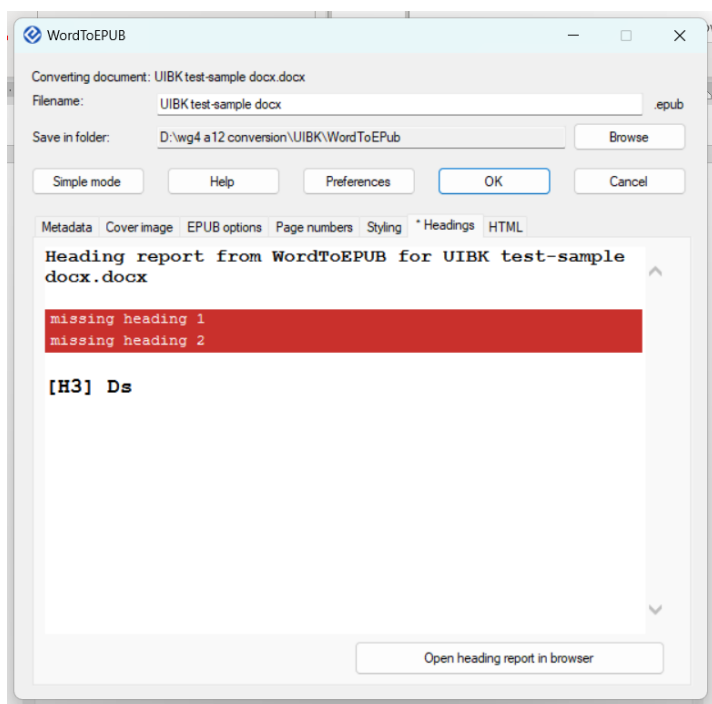


Figure 42: Interface of WordToEpub tool with preview of headings structure with warnings

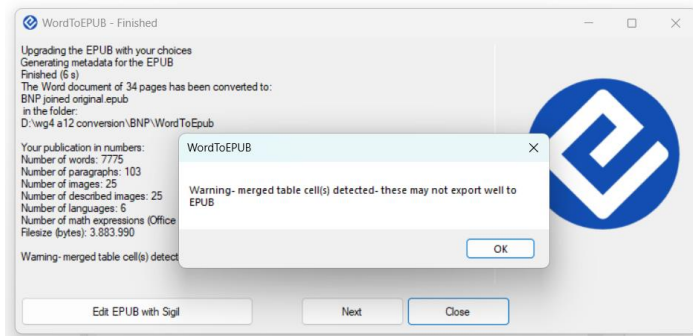


Figure 43: Interface of WordToEpub tool after the conversion into EPUB was finished and warning appeared about merged table cells

The following is the full report on conversion into EPUB, copied from inside the WordToEpub interface.

Now converting: BNP joined original.docx
Starting...
Making a safe copy of your document
File copy success
Getting the names of heading styles
Reading metadata
Analyzing headings
Analyzing languages.....
Detected Office Math: 5
Starting a fresh copy of Word
Opening a copy of the document
Repaginating the document
Inserting navigation for page 34/34
Saving the prepared document
Simplify Word markup
Numbered headings.....
Language set to: en-us
Added language markup into content: en-gb, en-si, fr-fr, de-de, pt-pt
Described images: 25, decorative images: 0, missing alt text: 0
Generating metadata for the EPUB
Building your EPUB with pandoc
Upgrading the EPUB with your choices
Generating metadata for the EPUB
Finished (6 s)
The Word document of 34 pages has been converted to:
BNP joined original.epub

in the folder:

D:\wg4 a12 conversion\BNP\WordToEpub

Your publication in numbers:

Number of words: 7775

Number of paragraphs: 103

Number of images: 25

Number of described images: 25

Number of languages: 6

Number of math expressions (Office Math): 5

Filesize (bytes): 3.883.990

Warning- merged table cell(s) detected- these may not export well to EPUB

B. Workflow for creating HTML format

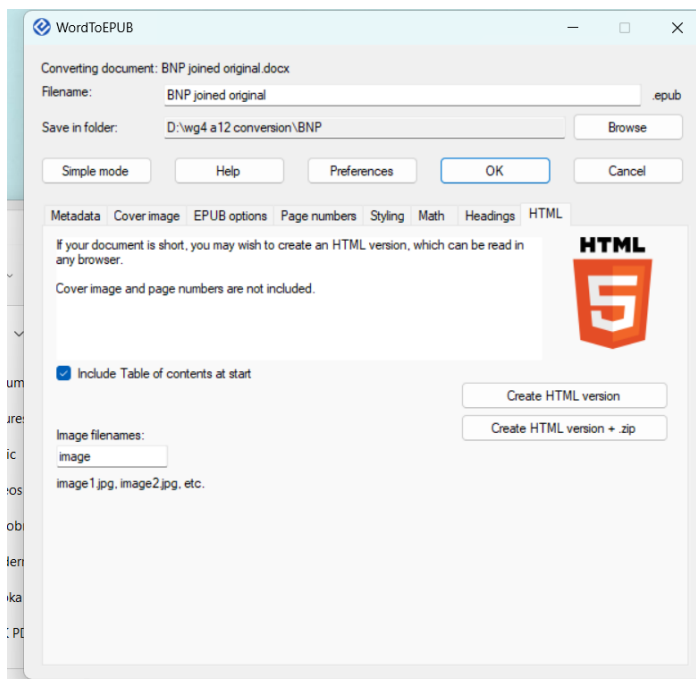


Figure 44: Interface of WordToEpub tool in regard to settings for HTML format

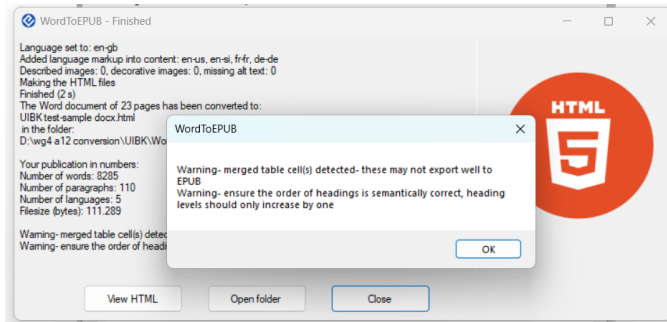


Figure 45: Interface of WordToEpub tool after the conversion into HTML was finished and warning appeared about merged table cells

The following is the full report on conversion into HTML, copied from inside the WordToEpub interface.

Now converting: UIBK test-sample docx.
Starting...
Making a safe copy of your document
File copy success
Getting the names of heading styles
Reading metadata
Analyzing headings
Analyzing languages.....
Simplify Word markup
Numbered headings
Marking the pages
Number of pages marked: 0.....
Language set to: en-gb
Added language markup into content: en-us, en-si, fr-fr, de-de
Described images: 0, decorative images: 0, missing alt text: 0
Making the HTML files
Finished (2 s)
The Word document of 23 pages has been converted to:
UIBK test-sample docx.html
in the folder:
D:\wg4 a12 conversion\UIBK\WordToEPub\UIBK test-sample docx

Your publication in numbers:
Number of words: 8285
Number of paragraphs: 110
Number of languages: 5
Filesize (bytes): 111.289

Warning- merged table cell(s) detected- these may not export well to EPUB

Warning- ensure the order of headings is semantically correct, heading levels should only increase by one

C. Workflow for creating DOCX format with EpubToWord

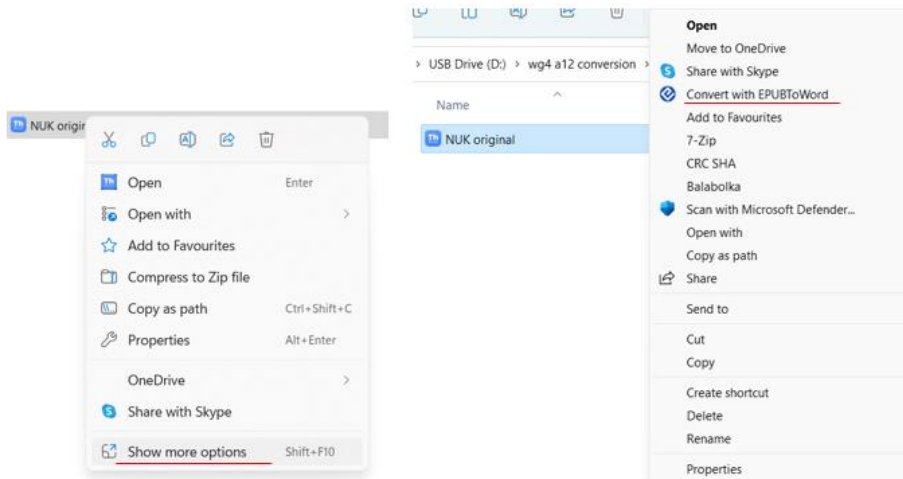


Figure 46: Opening WordToEpub tool's feature EPUBToWord in Windows explorer

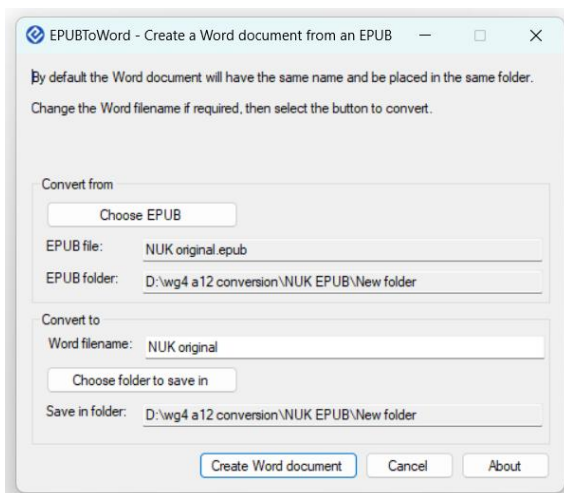


Figure 47: Interface of WordToEpub tool's feature EPUBToWord for converting EPUB to Word document

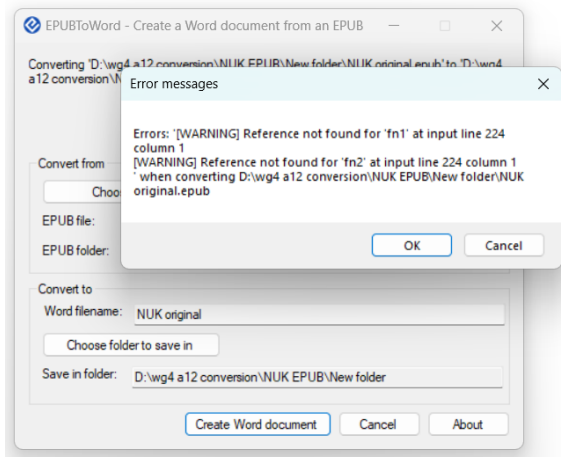


Figure 48: Interface of WordToEpub tool's feature EPUBToWord for converting EPUB to Word document with error for the footnotes

Annex 7: Conversion report for EasyConverter Express

1. Were there any changes to the input file before the conversions?

We combined BNP’s individual DOCX files into a single file to facilitate the conversion process. Additionally, UIBK’s RTF format was opened in Microsoft Office Word and saved as DOCX format to provide an additional sample to test.

How long did it take to receive the output results?

For the EPUB format, it took only a few minutes, and the conversion proceeded without any issues. However, when creating the MP3 files, we encountered some challenges. In one case, we needed to attempt the conversion multiple times to achieve a 100% completion, as it stopped prematurely. In another instance, we were unable to obtain a 100% completion, and the process halted at 10%.

2. Is it possible to convert multiple files simultaneously?

No, batch conversions are not supported.

3. Explain the workflow.

The workflow is straightforward. We opened the input files in the software and selected the desired output format (refer to Figure 38). In the subsequent step, whether for EPUB or MP3, we didn’t make any adjustment to the settings and designated the location for saving the files (refer to Figures 39-40).

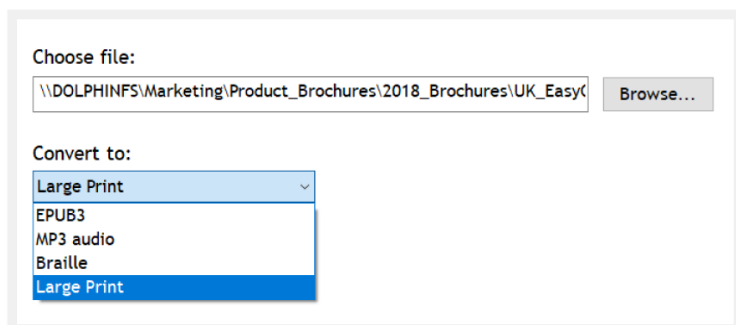


Figure 49: Interface of EasyConverter Express.

Source: Dolphin Computer Access, accessed 27. 6. 2023.

<https://yourdolphin.com/EasyConverter-Express/Features>

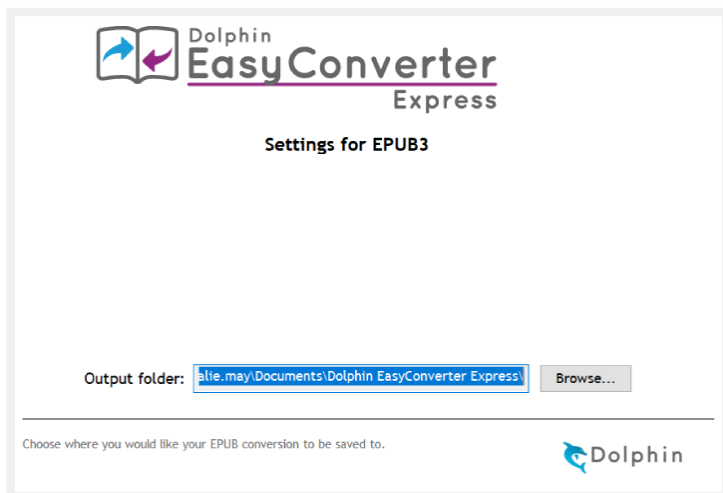


Figure 50: Interface of EasyConverter Express, window of settings for EPUB3.

Source: Dolphin Computer Access, accessed 27. 6. 2023.

<https://yourdolphin.com/EasyConverter-Express/Features>

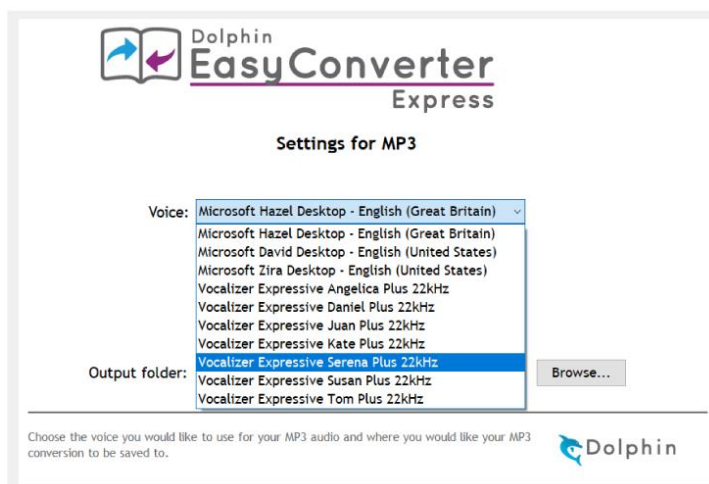


Figure 51: Interface of EasyConverter Express, window of settings for MP3.

Source: Dolphin Computer Access, accessed 27. 6. 2023.

<https://yourdolphin.com/EasyConverter-Express/Features>