



Guidelines for remediation tools producers

Accessible Backlist Ebooks Laboratory
(ABE Lab)



Creative
Europe

TABLE OF CONTENTS

INTRODUCTION	5
PART 1 - THE CONTEXT	6
Accessibility matters.....	6
Be prepared for June 2025: the European Accessibility Act is approaching	6
The ABE Lab project.....	7
The partners	7
Objectives and outcomes of the project	7
Objectives of the guidelines	8
A rapidly evolving landscape	9
The rapid advance of Artificial Intelligence	10
PART 2 - GUIDELINES FOR REMEDIATION TOOLS DEVELOPERS.....	11
What should a remediation tool do?	11
Find, fix, check: the three key aspects of accessibility remediation.....	11
In-tool support.....	11
Improve, don't disrupt	12
What about multimedia resources, scripts, forms, and formulas?	12
Maintaining the look & feel vs resetting the styles	12
Functionalities to fix accessibility issues	13
Semantics.....	17
PART 3 - INFORMING USERS TO HELP THEM MAKING THE RIGHT CHOICE.....	18
EPUB Accessibility, WCAG, PDF/UA, EAA, ADA,: which target?	18
Simple or complex text?	18
Formats of input and output.....	18
Different tools, different targets, different workflows	18
Support and documentation.....	19
A safe and transparent work environment.....	19
Type of licence and plan	20
What else?.....	20
An accessible interface is an added value	20
Don't reinvent the wheel: few open source tools available	20

PART 4 - EVALUATE THE REMEDIATION FUNCTIONALITIES OF YOUR TOOL	22
For EPUBs	24
EPUB Version.....	24
Market validators.....	24
Metadata	24
Graphical resources.....	24
Styling and aspects	24
Textual content.....	24
Publication structure.....	25
Tables.....	25
Navigation	25
Links	25
Rare content (Multimedia and scripts).....	26
For PDFs.....	27
Market validators.....	27
Metadata	27
Graphical resources.....	27
Styling and aspects	27
Textual content.....	27
Document structure	27
APPENDIX.....	28
Acknowledgments	28
Contacts	28
DOCUMENT CONTROL INFORMATION	29
Document Approver(s) and Reviewer(s):.....	29
Document history:	29
Statement of originality:	29
Document Location.....	29

INTRODUCTION

This white paper is a public deliverable of the **Accessible Backlist Ebooks Laboratory (ABE Lab)**, a Creative Europe funded project, and can be freely downloaded from the *Deliverables* section on the official ABE Lab website.

In recent years, the topic of digital accessibility has become increasingly important, also in the light of the imminent entry into force of the Directive 2019/882 on accessibility requirements for products and services, usually referred to as *European Accessibility Act (EAA)*.

Ebooks are among the services to which the Directive applies. Some publishers have already started to produce and distribute *born accessible ebooks* to be in line with the EAA when it comes into force in June 2025. But how to manage ebooks that have been produced and published in previous years, the so-called *backlist*?

This is where all those tools, software, platforms, cloud solutions, APIs and other applications that can be used to transform an inaccessible ebook into an accessible version come into play, correcting accessibility issues to obtain a digital book that meets the requirements defined by European legislation and international accessibility guidelines. Sometimes, these solutions are marketed as and used in the context of conversion services. These technological products can differ greatly in terms of technology used, functions available, target audience and methods of integration within the remediation workflows of publishers and publishing industry players. Some are developed with the aim of outputting a file that is already fully accessible and compliant with international accessibility requirements, while others may be specialised in solving specific accessibility problems. The latter tools can then be integrated into workflows to support a specific remediation activity within the process of converting the ebook into an accessible version.

In this white paper, we use the term *remediation tool* as an umbrella term that includes technological and IT products of different types. Remediation and conversion *services*, however, are outside the scope of the project. With remediation and conversion services we refer to the services offered by commercial companies or realities specialised in producing accessible versions, where the remediation activity is carried out by a third party on behalf of a client.

This document aims to provide high-level guidelines that can be used by producers and developers of remediation tools to design, develop or improve their own tools, highlighting the requirements these tools should include to be effective and to support the human operators by making the process as simple, intuitive and fluid as possible - depending also on the competence of the target users.

PART 1 - THE CONTEXT

Accessibility matters

Accessibility means that tools, technologies, websites, ebooks and digital documents are designed and developed so that everyone, regardless of their disabilities - physical, cognitive, motor, permanent or temporary, situational or age-related - can use them. Accessibility is fundamental to providing equal access to information, study, entertainment, and equal opportunities also to people with disabilities.

In addition to being required by the regulatory frameworks in force in many countries around the world, accessibility is a social value and a basic human right recognized by the United Nations Convention on the Rights of Persons with Disabilities (UN CRPD).

Accessibility is also good for business: an accessible tool, website or ebook is a product of high quality that improves the general user experience and satisfaction¹.

Be prepared for June 2025: the European Accessibility Act is approaching

The European Accessibility Act is an EU Directive (2019/882)² that aims to improve the functioning of the internal market for accessible products and services, by removing barriers created by divergent rules in Member States. It will enter into force in June 2025. The Directive establishes binding accessibility targets that must be met by many different products and services to strengthen the rights of people with disabilities to access goods and services, including ebooks. Since ebooks are considered a service, the concept of a service provider now includes publishers and all the other economic operators involved in their distribution. Every step of the publishing chain must integrate accessibility to create an accessible digital publishing ecosystem.

For publishers, this means that they are required to produce their digital publications in an accessible format, according to the *born accessible paradigm*.

The white paper “E-books for all - Towards an accessible digital ecosystem” by Fondazione LIA³ provides detailed information on the impact of the EAA on the publishing supply chain.

The European ebook market, however, is not only made up of new titles, which can be placed on the market already accessible, but also of ebooks that have been published in previous years, and which publishers will have to make accessible if they want to keep them on the European market.

¹ Introduction to Web Accessibility, W3C Web Accessibility Initiative (WAI),
<https://www.w3.org/WAI/fundamentals/accessibility-intro/>

² See online at: <https://ec.europa.eu/social/main.jsp?catId=1202>

³ E-books for all. Towards an accessible publishing ecosystem, Fondazione LIA, 2020. Available online at:
<https://www.fondazionelia.org/en/resources/e-books-for-all/>

The ABE Lab project

The partners

The **ABE Lab - Accessible Backlist Ebooks Laboratory** is a European Cooperation project funded by Creative Europe. The project is coordinated by the *European Digital Reading Lab (EDRLab)*; *Fondazione LIA* and the *National Library of the Netherlands* (Koninklijke Bibliotheek - KB) are partners.

EDRLab is a non-profit transnational laboratory that develops open standards and open-source tools, to accelerate the diffusion of digital technologies in the publishing industry. It is an association of more than 70 members from the publishing industry, including publishers, providers of publishing services, distributors, tools developers and accessibility specialists. It is also the developer of the Thorium Reader application, a free and highly accessible reading application for EPUB 3 ebooks⁴.

Fondazione LIA is an Italian non-profit organisation and an internationally recognized expert in the certification of accessibility of EPUB 3 ebooks, having already certified more than 30.000 ebooks available in the *libri italiani accessibili* catalogue⁵. Fondazione LIA carries out research and development activities in the field of digital accessibility, organises awareness-raising events, offers training courses and consulting activities. It works with 76 publishing imprints to create and distribute the new ebooks as *born accessible publications* and it's part of an international network of organisations dealing with accessibility of contents⁶.

KB - The National Library of the Netherlands has a long history in library research and promotes the visibility, usability and longevity of the Dutch Library Collection, defined as the collective holdings of all publicly funded libraries in the Netherlands. It has three main roles: it is the national depot where many Dutch publishers send their ebooks, it operates an ebook lending platform for all Dutch public libraries and it commissions the conversion of non-accessible (e)books to accessible formats for people with a print impairment. With the EAA on the horizon, the KB strives to get all of its public services accessible⁷.

Both EDRLab and LIA are part of the W3C Publishing Business Group, W3C Publishing Working Group, W3C EPUB Community Group, and are active members of the EPUB Community Group Accessibility Task Force and of the EPUB Community Group Fixed Layout Accessibility Task Force.

Objectives and outcomes of the project

ABE Lab has two main objectives:

1. to collect and analyse in detail the composition of the European ebook backlist
1. to provide guidelines to European publishers to boost the remediation of the ebooks from their backlist.

⁴ More information available on EDRLab's official website: <https://www.edrlab.org/>

⁵ <https://catalogo.fondazioneia.org/>

⁶ More information available on Fondazione LIA's official website: <https://www.fondazioneia.org/>

⁷ More information available on KB's official website: <https://www.kb.nl/en>

Without precise data about the composition of the ebook backlist in Europe, it is not possible to evaluate the best workflows and estimate the time and expertise that making these ebooks accessible would require. The first part of the project was therefore focused on the collection and analysis of data to define the size and composition of the European ebook backlist - defined as the collection of all ebooks available on the EU market - and on identifying the most recurrent accessibility issues that can be found in these ebooks. The results are detailed in the public deliverable *Report on Backlist Data and gap analysis*, freely downloadable from the official ABE Lab website both in PDF and EPUB format. An online version is also available⁸.

The core part of the project has been to analyse and test the processes and workflows by which ebooks of different types and formats can be transformed into accessible EPUB files or accessible PDF files. This implies knowing which remediation tools are available on the market and what they can do, for which formats and types of books (fiction, complex layouts, STEM) they are most effective and which are their shortcomings.

It will in fact be essential to have tools that are able to correct the accessibility errors identified in the backlist files, or that can offer the human operators options and features to support them in their ebook remediation journey.

Consequently, testing the remediation tools has represented a fundamental part of ABE Lab. The results of the tests carried out in the context of the project form the basis of these guidelines.

All public future deliverables of the project will be available, after publication, on the *Deliverables* page⁹ of the ABE Lab website.

Objectives of the guidelines

To the best of our knowledge, this white paper represents the first publicly-available work of its kind. The document outlines high-level guidelines that stakeholders, producers and developers of technological tools and solutions aimed at fixing the accessibility issues of digital documents and publications can use to independently evaluate their tools. It can also be a useful resource to improve them through the integration of new features, or to design new ones that offer the functionalities and support necessary to produce fully accessible ebooks starting from inaccessible versions. Remediation tools should provide all the possible support for the human operators responsible for the file correction activity, making the process as simple, intuitive and fluid as possible. This, of course, will depend also on the competences of the target users and on the type of tool used - remediation tools can be desktop applications, cloud platforms, API, SDK and more.

The information these guidelines provide are based on the accessibility requirements for ebooks defined by the European Accessibility Act listed in Annex I sections III and IV Linea f). EPUB (Electronic

⁸ Report on Backlist Data and gap analysis, ABE Lab, October 30, 2023. Available online at <https://www.abelab.eu/outcomes/deliverables/#report-on-backlist-data-and-gap-analysis>

⁹ The page can be found at the following link: <https://www.abelab.eu/outcomes/deliverables/>

PUBlications)¹⁰ is recognized as the format of choice to respond to those requirements, as detailed by the *EPUB Accessibility - EU Accessibility Act Mapping*¹¹. Therefore, the EPUB Accessibility represents the target for the remediation of ebooks in EPUB format, since EPUB files conforming to EPUB Accessibility guidelines are responding to the European Accessibility Act requirements. The EPUB Accessibility guidelines are built on the *Web Content Accessibility Guidelines 2 (WCAG 2)*¹², which they include and expand with specific requirements for digital publications.

Ebooks, however, are not exclusively produced in EPUB format. Other formats coexist, most of them derived from this main format or based on the same web technologies (HTML, JavaScript, CSS)¹³.

A notable exception to this is the Portable Document Format (PDF), a format initially proposed for print reliability and digital conservation of documents. Accessibility of PDF is based on WCAG 2 and is made possible thanks to an additional semantic descriptive layer composed of 28 elements expressed as XML language tags. The PDF/UA ISO¹⁴ standard provides definitive terms and requirements for accessibility in PDF documents and applications.

Pre-paginated ebooks (like PDFs and Fixed Layout EPUBs) do not comply with the EAA requirements for the criterion of flexibility and choice in the presentation of the content, a key functionality for persons facing cognitive difficulties (like dyslexia) or with sight impairments. Hence, to make the content of a PDF fully accessible, a change of format might be necessary, from a fixed layout one to a reflowable one.

A rapidly evolving landscape

Technology providers and developers of technical tools for the publishing industry are increasingly integrating functionalities in their products and IT solutions that can be used to convert existing inaccessible digital files into accessible digital versions, or to enhance their level of accessibility. Some tools already exist on the market with this aim, and we are also aware of several new projects in development. It is therefore easy to predict that more dedicated technological solutions will be made available in the near future.

This rapidly evolving landscape is also driven by an increasingly widespread use of new types of automation powered by systems based on Artificial Intelligence and Machine Learning, whose great

¹⁰ EPUB® 3 defines a distribution and interchange format for digital publications and documents. The EPUB format provides a means of representing, packaging, and encoding structured and semantically enhanced web content — including HTML, CSS, SVG, and other resources — for distribution in a single-file container.

The publishing industry, worldwide, agrees that EPUB is the only digital book format which allows, with a few efforts, the creation of fully accessible ebooks. The standard is available online at: <https://www.w3.org/TR/epub-33/>

¹¹ EPUB Accessibility - EU Accessibility Act Mapping, W3C Group Note, 27 December 2023. Available online at: <https://www.w3.org/TR/epub-a11y-eaa-mapping/>

¹² The Web Content Accessibility Guidelines (WCAG) are the international standard for web accessibility, developed and maintained by the W3C. WCAG 2.2 was published on 05 October 2023. An overview of WCAG 2 is available at: <https://www.w3.org/WAI/standards-guidelines/wcag/>

¹³ The backlist data analysis showed that some markets have other formats, like HTML books or apps (18% in France and 25% in Spain). Those formats are based on or derived from the same web technologies of the EPUB format, therefore they should comply with the Web Content Accessibility Guidelines too.

¹⁴ ISO 14289-1:2014, Document management applications - Electronic document file format enhancement for accessibility

potential could play a key role in improving and speeding up processes for correcting accessibility issues - in particular, they represent a possible solution to quickly process large amounts of files.

The rapid advance of Artificial Intelligence

When we defined the objectives and activities of the ABE Lab project, the topic of Artificial Intelligence was not yet as hot a topic as it would become just a few months later. Artificial Intelligence applied to an ever-increasing number of fields began to be talked about very widely during the course of the project. Therefore, the study of the use of Artificial Intelligence in the context of digital accessibility and in particular its correction - i.e., in remediation tools - was not among the aims of the ABE Lab project.

Publishers and other actors in the publishing chain have been using automation, including Natural Language Processing and Artificial Intelligence technologies, for several years already. Now facing the revolutions of Large Language Models (LLM) and Generative AI technologies (GAI), they are observing, testing and defining positions towards this technology and its use in editorial workflows. This has led in some cases to express concerns related to the uncertainty about how their contents and files are used by such systems.

However, it should also be highlighted that, in the current state of the art, Artificial Intelligence and Machine Learning systems cannot replace the competence and evaluation ability of a human being. If it is possible, for example, to automatically generate alternative descriptions for images with AI, it has yet to be proved that these alternative descriptions can provide information equivalent to that conveyed by the images themselves, or that they are comparable to those produced by a subject expert. Therefore, human decision is still key to stating whether such descriptions are appropriate or not for a given context.

Developers and technology specialists are aware of the power of AI, but they are also aware of the risks implied in and of the need for a wise and detailed study of side effects, edge cases, caveats and other surprises coming with intents to bypass human driven operations.

PART 2 - GUIDELINES FOR REMEDIATION TOOLS DEVELOPERS

The following guidelines present indications of different types: some strictly concern the functioning of the remediation tool, while others concern the information that developers should provide to allow an informed and adequate choice by those who have to fix the accessibility of the ebooks in their backlist. The applicability or otherwise of a guideline to a specific tool may depend on the technological characteristics of the tool itself - for example, for tools that work from the command line, guidelines that imply the presence of a graphical interface are not relevant. The functionalities detailed in the following sections might be implemented in a variety of ways: automatic, semi-automatic or manual.

What should a remediation tool do?

Find, fix, check: the three key aspects of accessibility remediation

To solve an accessibility problem, it is first necessary to identify it, and then understand its impact and how to correct it. To support the human operators, the remediation tool should identify the accessibility problem, report it to the users, and guide them in its correction by providing dedicated options to do so.

For the remediation to be effective, the tools should therefore:

- find the accessibility issues contained in the ebook;
- provide a way for correcting the identified accessibility issues, for example, via dedicated options;
- check the conformance of the file to ensure the output will be valid with the file format specification and compliant with the target accessibility guidelines or standards. Accessibility checks carried out before export, or at the request of the operator, are also strategically important to ensure that the ebook does not have accessibility problems before it is exported and checked with external tools.

These activities should not be interpreted as steps that must be performed in a predefined sequence in the remediation process, but rather as three fundamental aspects that should be implemented by remediation tools to provide comprehensive support and ensure a quality output.

In-tool support

There should be attention to providing as much support as possible to guide the human operators in fixing the accessibility errors, with the aim of reducing the number of iterations needed to the minimum.

Even if some accessibility problems can be effectively solved automatically, a report of all automatic operations run should always be provided and manual evaluation and intervention should always be possible.

For actions where no automatic intervention is done, there should be guidance to allow the human operator to manually fix the accessibility issues of the files, for example providing dedicated menus, popups, options, in-tool guidance, ad-hoc functionalities.

Providing a code editor (such as an HTML editor for EPUB remediation tools) can be an option, but this solution requires the human operator to have a technical knowledge and although it can be really useful, this alone doesn't provide it itself much guidance and support.

Improve, don't disrupt

The remediation tool must not cause a loss of information, must not add accessibility errors and must not compromise the technical integrity of the file. This may seem obvious, yet we have encountered tools whose solutions to correct accessibility errors generated others, making the file invalid; in other cases the output files, although passing the tool's internal validation, were reported by other software and applications as "corrupt" and thus not readable. In these cases it is not a question of accessibility, but rather of formal validity of the output and compliance with the standards that define the file format itself.

What about multimedia resources, scripts, forms, and formulas?

Multimedia elements (audio and video) are quite rare in digital books. The same goes for specific contents like form fields, scripts, and buttons, which could be used to create interactive components. Users of the tool need to know in advance whether audio, video, interactive forms (and therefore JavaScript) or formulas tagged with MathML are correctly imported into the tool and can also be retained in the output ebook when this content is present in the publication to be remediated. Support for these elements must be made explicit in the documentation or in the official channels of the tool, to avoid a loss of content and functionality of the ebook.

These elements must meet specific accessibility requirements. Tools that provide support for them should at least notify the human operators that such special content exists in the ebook, to allow them to find them and check them to verify if specific activities with dedicated tools are needed. In fact, making these components accessible may not be one of the purposes and functions of digital publishing dedicated tools. For example, audio and video content require a textual alternative (caption or transcription), which can be done with solutions for creating accessible multimedia content.

Whether they can be fixed in-tool or with external additional technologies, for an ebook to actually comply with accessibility requirements, audio and video, as well the other special content, must also be taken into consideration.

Maintaining the look & feel vs resetting the styles

When it comes to converting a PDF into an accessible PDF, we expect the original layout and design to be maintained: fonts, colours and graphic elements are not affected during the remediation phase, and the accessible file respects the graphics of the source file. There may be cases in which a

reflowable EPUB presents graphic characteristics that would also be desired to be maintained in the accessible reflowable EPUB. Maintaining the *look & feel* of the inaccessible input file may not always be the desired solution for publishers, for example if the goal is to turn the PDF into an accessible reflowable EPUB.

It is therefore important to specify whether the tool removes by default all the graphic properties of the input file, thus returning black and white text and assigning predefined styles to the text components (headings, bold, italics, links, ...), whether these are maintained, or whether it is possible to choose which solution to apply. In some cases, specific fonts are necessary to represent special characters. If styles are automatically removed and replaced, these characters may no longer be represented correctly, resulting in a loss of content.

Maintaining the graphics may also imply maintaining accessibility errors related to colour contrasts or the use of colour alone to convey information. To solve this problem, a possible solution is to allow manual editing of styles or provide image or graphics editors inside the tool itself or as external linked resources.

No solution is intrinsically better than the other, but the user must be informed.

Functionalities to fix accessibility issues

An ebook can contain many different types of accessibility errors. By analysing a sample of backlist files, we found that some accessibility problems recur with particular frequency¹⁵.

Based on these results, and by analysing the characteristics of the remediation tools tested in the context of the project, we have identified a series of features that we deem important to remediate backlist ebooks.

Adding accessibility metadata

The lack of accessibility metadata, required by both EPUB Accessibility 1.1 and the EAA, is an error that was found in 100% of the backlist EPUB analysed. Accessibility metadata can be created by publishers or by a third-party certifier who checks the file and certifies it if accessible. Just as it is usually possible to set the general metadata of the document, such as title and author, it should also be possible to specify the accessibility metadata of the ebook as required by the EPUB Accessibility, therefore using the metadata defined in the Schema.org standard. In this regard, a remediation tool should:

- *detect missing accessibility metadata.* The lack of accessibility metadata can be easily detected with an automatic check;
- *provide a way to add accessibility metadata.* This could be done, for example, adding an “Accessibility information” section in the general metadata panel, or including a specific menu item that the human operator can activate to open a popup to add them;

¹⁵ Report on Backlist Data and gap analysis, ABELab, October 30, 2023. Available online at https://www.abelab.eu/outcomes/gap_analysis/

- *allow to review accessibility metadata* - this is satisfied simply by making the accessibility metadata available for the human operators in the popup to add them.

There are also other useful functionalities regarding accessibility metadata:

- *automatically create accessibility metadata* based on the analysis of the EPUB features;
- *automatically check for correct accessibility metadata*. This would require the tool to compare the declared values of the accessibility metadata with the actual content of the file and warn the human operators if something seems incorrect.

While accessibility metadata is a fundamental requirement for the accessibility of EPUB files, there is currently no defined standard for its support by the PDF format.

Alternative text for images

After missing accessibility metadata, this is the most common issue found in backlist files. The topic of image description is a delicate one and represents one of the greatest difficulties when it comes to accessible digital content: how can alternative image descriptions be validated? Who should write them?

We identified three functionalities that the remediation tools should offer:

- detect missing alternative text. Images and graphics that are used for decorative purposes only don't need an alternative text, but sometimes in the EPUB files what happens is that the HTML tags for the images (, <svg>) are not just declared as decorative, but instead are missing the attributes required to specify the alternative text or to declare that the image is decorative, and so it is not possible to decide whether they were meant to be informative or not. In the case of the tag, the alt attribute used to set the alternative text is always required and the HTML code is not well-formed if it's missing. Warning the user about this error helps ensuring that all images either have an alternative text, or have been marked as non-informative;
- provide a way to add an alternative text to images and graphical elements, and allowing to mark as decorative (or "artefact" in the PDF world) all the images and graphical elements that do not convey information and can be ignored by Assistive Technologies;
- review all images. Providing a way to review all the images and the associated alternative text is extremely useful: it allows to check if all the images have been described and if all the decorative images have been marked as such. This is also helpful to check if the alternative descriptions are appropriate for the images and the context in which each image occurs.

Remediation tools should not provide placeholder image descriptions just to pass the accessibility validation. We have found out that some tools provide non-meaningful alternative texts like generic single words (example: "image") or the image file name (example: "final26ok72.jpg"). Unfortunately, those practices have been in place since a long time ago. Implementing some search patterns may help easily identify them and point them as errors.

A particular type of textual alternative for graphic content is represented by the so-called *extended descriptions*. If the alternative text represents a summary description of the image, the extended descriptions can be longer and more structured, including also tables, lists and links. Extended descriptions should be used for complex images that can be found in certain categories of books, like STEM. These images must be described in detail to ensure that their content is correctly and thoroughly conveyed to users with visual impairments using screen readers and other Assistive Technologies. At the state of the art, there is no standard way to insert extended descriptions in either the EPUB and HTML-based formats, or the PDF format. The *Daisy's Best practices for authoring Extended Descriptions in EPUB*¹⁶ provide some useful guidance on the topic.

While providing an alternative text for non-decorative graphic elements is mandatory to reach compliance with the EAA and the WCAG, whether or not the image should be described with an extended description depends on the type of image and the context in which it is placed. The possibility of inserting them in the remediation phase represents an added value for the tools.

Missing or bad language tag

The WCAG requires that the language of the content - therefore of each HTML documents that make up the EPUB file, and of the PDF document -, as well as the language of words, sentences and passages of text in a different language are correctly identified, to allow user agents and Assistive Technologies to present them properly (with the right spelling and pronunciation by speech synthesis) to the end user.

This implies that:

- the tool should *allow to set the main language of the content*;
- the tool should *allow to set the language of words, phrases or passages of text*;
- it would be ideal if it could recognize *a mismatch between the declared language and the actual one*, and report it to the human operators.

Open source language detection libraries and open dictionaries in different languages can be used to help identify the language of the content - both for the whole document and for single passages of text, even automatically - and report cases where a language assignment may potentially not match the actual language of the content.

If the recognition of the languages used in the content and their marking can be done automatically, then it must be specified which languages are supported and can be correctly identified.

What the tool *should not do* is to set a default value for the language of the document regardless of the actual language of the contents. If this happens, it implies the tool can only be used to remediate content in that specific default language.

¹⁶ Available at <https://daisy.github.io/transitiontoepub/best-practices/extended-desc/ExtendedDescriptionsBestPractices.html>

Headings hierarchy

The content of the ebooks, just like that of paper books, can be structured in sections, subsections, chapters or articles. The headings, generally visually stylized to stand out from the actual text, allow to recognize these parts and understand how a book is organised, what the main themes are and how they are connected. The same information must be made available also for visually impaired readers who can't access the visual presentation of the text. Headings, just like other text components, must be tagged with the semantic tags available for the format used. Whether it is EPUB or PDF, it is necessary to use the heading tags in an appropriate way, with the appropriate level, implementing correct nesting.

Non-tagged PDF documents are commonly found, but we also identified EPUBs with missing or incorrect heading structures. The headings hierarchy must reflect the hierarchy of the publication or document, so an H2 should always be used for secondary level headings, H3 for tertiary, and so on. Headings of the same level may have different styles attached, but gaps in numbering must be avoided whenever possible, meaning that a H1 should always be followed by H2, and not H3 or a lower level.

Both for PDFs and EPUBs, the remediation tools should:

- allow to identify (tag) a text as a heading of the appropriate level. No restriction should be applied to the available levels (i.e. from level 1 to level 6) since it's not possible to know beforehand what is the structure of the ebook;
- detect incorrect heading nesting. It is not required to use all the levels, but levels cannot be skipped and the hierarchy must be respected;
- provide a way to review the headings hierarchy. This allows you to check whether the heading hierarchy respects the structure of the ebook and whether the headings are correctly nested.

Publication formats like EPUB files usually include more than one HTML document (the EPUB Content Document). In these cases, the readers' expectation is that the main headings of each document reflects their position in the overall hierarchy of the publication. Which means that the first heading of an EPUB Content Document might not always be a first-level heading. Hence, for EPUB files, remediation tools must not force each EPUB Content Document to start with a first-level heading, as this might not actually reflect the right structure of the ebook.

This is different for single document formats such as PDF, where the order of headings must reflect the structure of the document.

Colour contrast

The WCAG defines minimum colour contrast ratio requirements between text and background, and between graphic elements necessary to understand the content and their backgrounds. Even if the remediation tool doesn't allow to change the graphic appearance of the ebook - as it may happen for PDF to PDF/UA conversion tools -, it should *check if the colour contrasts between the text and*

its background is sufficient and complies with the minimum requirements set by the WCAG. If feasible, a way should be provided to change the colours to meet the requirements.

This also applies to graphic elements necessary to understand the content.

Document title

It must be possible to set the title of a PDF, which is usually found in the document properties (or metadata) panels of PDF remediation tools and reading applications.

For EPUBs, however, the title of each individual EPUB Content Document (represented by the <title> tag) must be meaningful. Using the book title for each Content Document is neither correct nor meaningful. If this tag is automatically populated by the remediation tool, it is important to ensure that a more meaningful text is provided (for example, the first heading, or the top-level heading of the document, eventually preceded by the book's title).

To support this requirement, the remediation tool should provide the following features:

- allowing to set the value for the <title> tag of each EPUB Content Document;
- detecting empty <title> tags in EPUB Content Documents and additionally detecting wrong titles such as titles identical to the main title of the publication;
- provide a way to review <title> tags. This means providing a way to ensure that each EPUB Content Document has a meaningful title.

The content of the <title> tag is often used by EPUB reading applications to inform the reader about their current position in the publication (functionality “Where am I?”), so it’s important it provides meaningful information.

Semantics

Semantics is fundamental for accessibility. HTML5 and the PDF standard provide a rather rich list of semantic tags, which must be used to convey the correct information to user agents, including Assistive Technologies such as screen readers and the people who use them. This is why a remediation tool should *allow semantic tags to be assigned to content* quickly and easily. Available semantic tags should not be limited to headings, paragraphs, and lists. The rich semantics the standards offer are much broader and should be used correctly.

PART 3 - INFORMING USERS TO HELP THEM MAKING THE RIGHT CHOICE

EPUB Accessibility, WCAG, PDF/UA, EAA, ADA,: which target?

Juggling standard acronyms and regulatory references can be complicated. It should be stated in the documentation of the remediation tool or on the official website what accessibility standard the remediated files can achieve compliance with. This naturally depends on the functions of the tool and on how many and which accessibility problems it allows to solve. It must be possible to output a file that actually complies with the requirements of the indicated standard or regulation.

Simple or complex text?

Books of different types (fiction, biographies, law, children's books, art books, ...) often have very different complexity of structure and layout. Fixing a text-only book doesn't require as much effort as a book with a complex layout. Remediating a STEM book requires specific features, such as for example an editor to add mathematical or chemical formulas in an accessible format (like MathML for EPUB, or LaTeX for PDF). Providing clear indications allows publishers and operators in the publishing chain to choose the remediation tool best suited to their needs.

Formats of input and output

Usually remediation tools do not support all ebook formats, but only some. This information is generally included in the presentation and description of the tool, but it is worth mentioning this aspect also in this context: it must be clear which input formats are accepted and which remediation processes are supported, i.e. which formats are obtained in output. For example, we can have remediation solutions that only accept PDF as input, and can output only PDF or Fixed Layout EPUB, but not accessible reflowable EPUB. This is a key aspect that must be declared.

Not only that: for ebooks in EPUB format it is necessary to specify how the tool behaves with older files, and whether these are correctly accepted as input and can be worked on successfully. EPUBs produced and distributed in past years may not be valid if checked with the latest version of EPUBCheck¹⁷. For this reason it is necessary to specify the minimum version of EPUBCheck according to which the input file must be valid.

Different tools, different targets, different workflows

Since remediation tools can take many different forms and have different levels of complexity, it is important to clearly identify which kind of audience of operators the tool is aimed at.

In fact, it must be considered that the remediation of ebooks could be done by actors in the publishing chain with skills that can be very variable: the remediation can be done by technicians or

¹⁷ EPUBCheck is an open source W3C project, maintained by the DAISY Consortium, that evaluates EPUB publications against the official EPUB specifications. More information available at: <https://www.w3.org/publishing/epubcheck/>

by operators with a more editorial training. Knowing whether the tool requires technical skills or whether it can also be used by non-expert operators is important information for the user.

The level of expertise required is closely linked to the type of tool and the way in which it can be integrated into the publishing house's ebook remediation workflows. Solutions provided as APIs or SDKs require specific technical and IT skills; instead, tools that have a user-friendly graphical interface could be easily used even by those who do not possess these technical skills.

It is therefore important that this information is specified.

A versatile remediation tool should also allow the creation of user profiles of different types, with functionalities adapted based on the type of profile.

Support and documentation

Providing clear, detailed documentation of how the remediation tool works and what it offers makes the work easier and faster. If the tool does not have any documentation it can be difficult to understand how the interface is structured, what the function of its components is, or what to do and which feature to use to obtain a certain result or correct a specific accessibility error. This information can be provided in various ways: a guide available online, downloadable or shared with licence subscribers, or one or more video tutorials. Providing clear directions on how to receive support or report problems is also helpful.

A safe and transparent work environment

Ensuring the security of the working environment and the protection of user data, as well as of the information relating to ebooks processed within the tools and platforms, is an essential requirement. This aspect takes on even greater relevance when it comes to Artificial Intelligence and Machine Learning: there are more and more remediation tools that use these technologies and this often requires a training phase on a collection of ebooks with similar characteristics to those you want to remediate. This improves performance but also means that AI or ML models base their training on information extracted from the provided files. It is necessary to provide precise indications on how the files, their contents and the information obtained from their analysis will be used, and it must be declared whether they will be used to train Artificial Intelligence or Machine Learning models, including the models used by the tool.

None of the remediation tools analysed in the context of the ABE Lab project provided any information in this regard. Artificial Intelligence and Machine Learning were not mentioned in the "Terms and conditions of use," the "Privacy Policy," or the description of the licence to use the product.

While we recognize the need for developers to not disclose the technological details of their infrastructures and to maintain trade secrecy, those who share their content with these tools have the right to know what use will be made of it.

In addition, it must also be declared how long the files will be kept within the platform (for online platforms) and whether they will be permanently deleted once the user decides to delete them.

Type of licence and plan

A tool can be open source, freely available, or available with a paid plan or licence. Different formulas can be used: it can be a subscription plan, a "pay per project", a "project bundle" solution, or a one-time purchase.

Providing information on the options available and detailing what they include allows publishers and other operators in the publishing supply chain to choose the option that best suits their needs.

What else?

An accessible interface is an added value

Although the objective of this document is not to provide guidance on the design, usability, and accessibility of remediation tools' graphical interfaces, it would be great if these were compliant with the requirements defined by the WCAG 2, the User Agent Accessibility Guidelines (UAAG)¹⁸, and the technical standard EN 301 549. An accessible interface would represent an added value would make the remediation tool usable by many operators.

Don't reinvent the wheel: few open source tools available

Regarding the availability of open source tools that can be securely incorporated into remediation workflows, we discovered a limited number of available solutions beyond the widely recognized conformance checkers such as EPUBCheck and ACE by DAISY.

These tools have become the industry standard and are already integrated into many publishers' production workflows. While incorporating these tools into workflows is a positive step, it's crucial to take note of the developers' disclaimer: *It is important to keep in mind that only a limited portion of accessibility checks can be automated, and therefore, Ace is not a complete accessibility conformance evaluation tool; instead, it is an aid for a broader, human-driven evaluation process.*

The lack of recognized tools is somewhat worrying, as it suggests that most accessibility issues are addressed using in-house developed solutions. This results in redundant efforts when it comes to selecting and evaluating the quality of remediation operations.

The availability of open source solutions does not imply that remediation tools necessarily have to use them, but knowing what the open source community has already produced can be a good starting point when designing and developing, especially if these products are actively maintained and updated.

All primary areas of software development extensively utilise existing components provided by the open source ecosystem. It is vital to the software development landscape, offering scalability,

¹⁸ User Agent Accessibility Guidelines (UAAG) 2.0, W3C Working Group, 15 December 2015, <https://www.w3.org/WAI/standards-guidelines/uaag/>

flexibility, and robustness. It helps everyone in the ecosystem save time and money, enhance quality and security, foster innovation and collaboration, and gain a competitive edge in the market.

PART 4 - EVALUATE THE REMEDIATION FUNCTIONALITIES OF YOUR TOOL

Evaluating a tool is a complex activity and involves multifaceted aspects. One of these is represented by the expectations of those who evaluate the instrument, which in turn largely depends on their specific needs. A tool could be evaluated on the basis of the results it allows to obtain even when used by non-expert users, while in other cases, what is sought is a tool capable of integrating effectively and fluidly into existing workflows and which therefore potentially requires advanced coding and programming skills. The ideal solution may also be platforms or programs that allow making batch modifications to large quantities of files. Combining such broad and varied expectations would make a complex checklist. Hence, the checklists we prepared focus on the features related to the most recurring accessibility issues highlighted by the gap analysis performed in the context of the ABE Lab project.

To facilitate the selection of tools by publishers, stakeholders or service providers, we encourage producers of tools to detail as much as possible the approach they adopt and the targets the tools are aimed at. This includes almost all aspects detailed in part 3 of this document, which refers to the information that should be provided to the customers and the final users. Since these points represent additional information and are on a different level compared to the functionalities necessary to improve the accessibility of ebooks according to the accessibility requirements, they are not included in the evaluation checklists reported in part 4 of this white paper.

These checklists can be used as methodological support to obtain a clearer view of the current state of a remediation tool. They want to help stakeholders, remediation tools producers and developers to check the state of the art of their tools and to plan future developments.

However, they do not claim to be a complete and definitive list to assess the quality or completeness of a remediation tool, and meeting all the points in the list does not guarantee that the remediated files will be fully accessible and compliant with the requirements of the EAA and international accessibility guidelines.

It's also important to highlight that they might not be maintained in the future. Therefore, **they can not be used to claim a level of conformity and *must not be used as a marketing or commercial argument.***

The points of the checklists, as well as the content of the guidelines described in part 2 of this document, are high-level indications that can be relevant for remediation tools regardless of the input file format.

Given that most ebooks are produced in EPUB and PDF and those two formats have important differences, we have decided to provide a separate checklist for EPUB and PDF, even if this means that many checkpoints appear in both checklists.

For other formats like Apps or HTML-based books that can be found in some markets like France (where they represent 18% of the ebooks) and Spain (25%), remediation to mainstream EPUB format should be considered the best solution, and EPUB Accessibility 1.1 should be considered the reference

point for accessibility. If it is not possible, WCAG will apply, along with the accessibility requirements set by the EAA.

The gap analysis, and especially the recurrent accessibility issues detected, provide a list of the most commonly found issues that should be addressed by any remediation tool.

For EPUBs

EPUB Version

- Does the tool check the EPUB version of the file?
- Does the tool automatically update an EPUB 2 to an EPUB 3?
- Does the tool accept files with EPUBcheck errors?
- Does the tool warn of the presence of past proprietary technologies?

Market validators

- Does the tool use the last version of EPUBcheck?
- Does the tool use the last version of ACE by DAISY?

Metadata

- Does the tool detect missing accessibility metadata?
- Does the tool check if the accessibility metadata are correct?
- Is it possible to review accessibility metadata?
- Does the tool automatically generate accessibility metadata?
- Is it possible to add accessibility metadata?

Graphical resources

- Is it possible to add alternative text for images and informative graphical content?
- Is it possible to detect non-decorative images without alternative text?
- Is it possible to mark an image as decorative?
- Does the tool provide a way to review the images and their associated alternative text?
- Does the tool provide a way to properly associate an image with its caption?
- Does the tool provide a way to add long and structured descriptions linked to an image?

Styling and aspects

- Does the tool provide a colour contrast checker to check the contrast of the text and its backgrounds?
- Does the tool detect styles and properties that prevent textual contents to correctly reflow?
- Does the tool provide an editor to define the style of the content?

Textual content

- Is it possible to set the main language of the content?
- Is it possible to set the language for words, phrases and passages of text?

- Does the tool automatically recognize the language of the content and assign the correct value to the xml:lang and lang attributes?

Publication structure

- Is it possible to identify a text as a heading of the appropriate level?
- Does the tool detect incorrect heading nesting?
- Does the tool provide a way to review the headings hierarchy?
- Is it possible to set the value for the <title> tag for EPUB Content Documents?
- Does the tool detect empty <title> tags for EPUB Content Documents?
- Is it possible to assign semantic tags (e.g. order and unordered lists) to the content?
- Does the tool detect missing or incorrect (DPUB) ARIA roles?
- Is it possible to edit (DPUB) ARIA roles?
- Does the tool propose a mapping for epub:types to ARIA Roles?

Tables

- Does the tool detect missing or empty table headers?
- Is it possible to identify table cells as row or column headers?
- Does the tool provide a way to properly associate a table with its caption?

Navigation

- Does the tool detect a broken pagelist?
- Is it possible to edit page numbers and the page list?
- Does the tool detect non unique landmarks?
- Is it possible to edit landmarks?
- Does the tool detect incorrect TOC (table of content) order?
- Is it possible to edit the TOC?

Links

- Does the tool detect a broken link?
- Is it possible to edit links?
- Does the tool detect potentially non meaningful text for the links or empty links?
- Is it possible to edit the textual label of a link (the name with which a link is displayed)?

Rare content (Multimedia and scripts)

- Does the tool warn the user if the file contains multimedia elements (e.g audio, video)?
- Does the tool warn the user if the file contains scripts?

For PDFs

Market validators

- Does the tool use a PDF/UA validator?

Metadata

- Is it possible to add the *title* metadata?

Graphical resources

- Is it possible to add alternative text for images and informative graphical content?
- Is it possible to detect non-decorative images without alternative text?
- Is it possible to mark an image as decorative?
- Does the tool provide a way to review the images and their associated alternative text?
- Does the tool provide a way to properly associate an image with its caption?

Styling and aspects

- Does the tool provide a colour contrast checker to check the contrast of the text and its backgrounds?

Textual content

- Is it possible to set the main language of the content?
- Is it possible to set the language for phrases and passages of text?
- Does the tool automatically recognize the language of the content and assign the correct language attribute?

Document structure

- Is it possible to identify a text as a heading of the appropriate level?
- Does the tool detect incorrect heading nesting?
- Does the tool provide a way to review the headings hierarchy?
- Is it possible to assign semantic tags (e.g. order and unordered lists) to the content?

APPENDIX

Acknowledgments

The ABE Lab team thanks all the developers and managers of the remediation tools who supported the project.

The following list **does not imply any endorsement** by either the ABE Lab project or by any of its partners or team members.

The list follows the alphabetical order of the names of remediation solutions:

- Access Aide plugin for Sigil: <https://github.com/kevinhendricks/Access-Aide>
- Bookalope, <https://bookalope.net/>
- InclusiveDocs, <https://inclusivedocs.com/>
- Movleen ISICrunch, <https://www.isicrunch.com/studio/>
- Nightingale epub-converter. The tool is still under development and no official resources exist. The official website of the Nightingale platform, dedicated to data management for book publishers, can be found at <https://www.nightingale.io/>
- PDFix Desktop, <https://pdfix.net/products/pdfix-desktop-lite/>
- StreetLib Write, <https://writeapp.io/home>

The following tools were also tested in the context of the project:

- ACE plugin for Sigil, <https://www.mobileread.com/forums/showthread.php?t=294678>
- Adobe Acrobat Pro, <https://www.adobe.com/it/acrobat/acrobat-pro.html>
- Adobe PDF Accessibility Auto-Tag API, <https://developer.adobe.com/document-services/docs/overview/pdf-accessibility-auto-tag-api/>
- Sigil - EPUB Editor, <https://sigil-ebook.com/>

All links were active on 18/04/2024.

Contacts

ABE Lab official website: <https://www.abelab.eu/>

If you want to contact the project team, share your thoughts or feedback on these guidelines, please drop an email to contact@abelab.eu

DOCUMENT CONTROL INFORMATION

Settings	Value
Document Title:	Guidelines for remediation tools producers
Project Title:	ABE Lab
Document Author:	Fondazione LIA
Project Manager (PM):	EDRLab
Doc. Version:	1.1
Sensitivity:	Public
Date:	4/03/24 - 23/04/2024

Document Approver(s) and Reviewer(s):

NOTE: All Approvers are required. Records of each approver must be maintained. All Reviewers in the list are considered required unless explicitly listed as Optional.

Name	Role	Action	Date
Fondazione LIA	Author	Approved	1/03/24 - 22/04/2024
EDRLab	PM	Approved	1/03/24 - 22/04/2024
KB	Project partner	Approved	1/03/24 - 22/04/2024

Document history:

The Document Author is authorised to make the following types of changes to the document without requiring that the document be re-approved:

- Editorial, formatting, and spelling clarification

To request a change to this document, contact the Document Author. Changes to this document are summarised in the following table in reverse chronological order (latest version first).

Revision	Date	Created by	Short Description of Changes
1	1/03/24	EDRLab, Fondazione LIA, KB	Initial publication
2	22/04/2024	EDRLab, Fondazione LIA, KB	Editorial revision, new checklist items added

Statement of originality:

This deliverable 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.

Document Location

The latest version of this controlled document is stored in:

<https://github.com/ABELaboratory/publications/deliverables/guidelines-remediation-tools-producers>