

Project Report: The Ian Williams Motocross Collection: A Digital Transcription

Introduction	1
Documentation of the methodological work	2
Division of labour	7
The project's relevance in the area of cultural heritage digitisation	8
A critical analysis of our choices and methods we have used	9
Conclusion	10
References	10
Appendix	11

Introduction:

For this project, we have chosen to digitise a selection of pages and photographs from a scrapbook which belonged to Jacqueline Williams' late grandfather, Ian Williams. The artefact documents his time as a motorcycle racer in 1950s Dublin through an array of newspaper clippings and photographs. The digitisation process serves to preserve family history as the artefact itself has not been well maintained. Apart from its relevance as a piece of family memorabilia, it also contributes to the cultural heritage sector as it documents a part of the sporting history of Ireland. Little is known about source of many of the articles, due to the way the document has been constructed, the journalists or publications had not been included. Therefore, one of the main obstacles of this project has been to navigate copyright law in Ireland. This report details the processes and purpose behind this digitisation project and seeks to justify its relevance to cultural heritage digitisation.

Documentation of the methodological work:

To document our methodological work, we have used the digitisation chain from the lecture slides in the course as a reference, however, for the scale of this project, we have omitted certain sections where they did not seem pertinent.

Planning and budget:

Since the project team members live in different countries (Germany and Norway, respectively), and due to other engagements, the planning of this project began with how to work around the distance between us - we had also chosen to digitise a physical object, so we had to rely on photos and video through different digital communication channels in the beginning.

The timeline and end date for the project was largely set by the course plan, though due to unforeseen circumstances, we had to delay the delivery of the project.

As the artefact is a family heirloom and we used software which we already owned (a licence to Adobe Photoshop and Oxygen XML editor) or had access to through public and university libraries (scanning), there were no budget restraints to this project.

Selection:

The aim of this digitisation project was to preserve, transcribe/encode and publish a scrapbook, both for the Williams family, and those interested in motorsports or Irish history. The scrapbook consists of 29 pages and several, loose photographs. We decided to select ten pages from the scrapbook and ten additional photographs. We wanted to digitise a varied selection of the general content. With the loose photographs, as these were not encoded, the selection was based mainly on variation, quality and photographs supporting the narrative of the scrapbook.

Document analysis and preparation:

The current owner of the document requested that it not be altered in any way. This imposed some minor restrictions in relation to our approach when analysing and preparing the document for digitisation. One limitation being that any metadata and information that may have been present on the back of the photographs affixed to the scrapbook (like those on a select few of the loose photographs) would be inaccessible. An analysis of the source document was carried out to ascertain which pages and photographs would be selected for the project. No physical restoration was implemented, any alterations or 'cleaning' was done digitally.

Text and image capture:

The first scanning strategy for the materials was twofold, the digital facsimiles of the scrapbook pages were collected using a book scanner and the photographs using a flatbed scanner.

Several issues were encountered while using the book scanner, the most crucial being the lack of output and formatting options, which only allowed the user to render scans in PDF format. Upon review of the files, it was found that the resolution was not adequate for our purposes. We decided to use the flatbed scanner for all the pages and photographs. As the page contents were mixed media (photographs and text), which could not be separated from the scrapbook page itself, it was beneficial to scan all pages with a higher resolution than would be necessary for the text portions alone.

After scanning the material, we used Adobe Photoshop to edit our selected scrapbook pages and photographs. The goal was to straighten the scanned images, crop out excess whitespace from the scanning process and to adjust the images slightly to be as true to the original artefact as possible. To achieve this, each master file (in TIFF format) was opened in Photoshop, straightened and cropped (some of the whitespace was kept in the edited files, to ensure that the whole page was contained in the finished image file), then the base layer was copied as an edit layer. This edit layer was then retouched, if needed, and a layer was added for contrast and to adjust black/white balance in the images. In the scrapbook images, the retouching was kept to a minimum to ensure the correct documentation of the state of the artefact. Some of the loose photographs were damaged, so a choice was made to remove spills, cracks and other distracting blemishes, to be presented in an album on our web page.

With both the scrapbook pages and the loose photographs, all files were saved as high resolution JPG files for publication, as editable TIFF files (in addition to the TIFF master files) and PNG thumbnail images, both in colour and in black and white.

OCR:

Optical character recognition (OCR) was utilised for part of this project. As the pages we chose to include were not overtly damaged, the quality of the output was adequate. As expected, there was some distortion to text located in areas of damage. However, as the text itself was a recognised font, the OCR software was easily able to assume the correct character and produce matching words and phrases, even when certain letters were slightly unclear, for

example, when two characters were too close together, due to the printing techniques of the time. The remainder of text was transcribed manually, as this was the preferred method of one team member. All text was proofread following its input from each method.

Text encoding:

The raw text was encoded in alignment with the current TEI guidelines. Before encoding began, decisions were made regarding what types of information should be described, how it should be described, utilising the TEI namespace, and how to arrange the text to reflect the format of the original document. We decided that, keeping in mind the potential primary audience for this work being family members and those interested in the history of motorsport, the most relevant items to encode would be names, motorcycle models, dates, publication details (where applicable) and races and their results.

As the source material is a newspaper, it possesses a unique format. We felt this was important to represent within the text encoding. Columns are marked and divided into sections, which are labelled using journalistic terminology, for example ‘lead’ for paragraphs that function as an introduction to a piece. Elements such as `<lb>` were important to include as it provided information regarding the structure of the text in its original form.

Structurally, we began with a format based on the TEI Encoding Guidelines for the University of Virginia Library

(<https://dcs.library.virginia.edu/digital-stewardship-services/tei-encoding-guidelines/#newspapers>) for newspapers. This was later revised to include a more comprehensive representation of the scrapbook itself rather than just its content. The original format relied heavily on `<div>` elements to label and differentiate the content, this was updated to utilise the `<facsimile>` tags to represent each page and the `<surface>` tag for each attachment on the page, allowing for a more standardised structure to be used across the various text and photographic forms.

Testing and validating:

GitHub has been used throughout the coding phases of the project (both TEI and HTML), and have therefore been able to test the code continuously through the process. Even when working on different parts of the project, the team members have been able to communicate and edit each other’s work.

Metadata and the TEI Header:

The TEI header of our encoded transcription of the scrapbook pages consists of compulsory elements of a TEI header and additional sections which best describe our project.

The <fileDesc> contains the bibliographic description of the file, in our case, the title of the project, publication information and a description of the source. There is a description of the artefact itself (<physDesc>), information about copyright and the history of the scrapbook.

In <encodingDesc>, an element meant to describe how the electronic document relates to the source, we found <projectDesc> And <editorialDecl> to be sufficient. Here, we were able to comment on what the project entails, in short, and to add that we have not edited the original source in our encoding. The element <encodingDesc> can contain several other elements, e.g., <unitDecl> or <samplingDecl>, but we did not feel the need to use these with the material we worked with due to the scale of the project.

The <profileDesc> section of the TEI header can contain a lot of information about the source material. In our project, the focus was on the person Ian Williams and his experiences, and the metadata in this section reflects that. If the project was to be expanded, metadata related to a wider range of racers and individuals related to the subject would be added to the <particDesc>. In <creation>, we chose an approximation of the <origDate>, since most of the clippings and photographs did not contain this information. Since the scrapbook in question contained several different types of texts, <textClass> and <textDesc> made it possible to distinguish between them. We also added a list of keywords, which according to the TEI Guidelines (<https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-keywords.html>, June 2022), can contain a set of keywords that appears in the content that is encoded.

XSL Transformation:

Firstly, we created a template based on the instruction from the residential week and a combination of provided video lectures and templates. As our transcription and our entire TEI is contained within the same HTML file, we created just one XSL transformation to HTML from the XML TEI file. This transformation was based on the ‘transcription’ page. We had issues with rendering a complete transformation, therefore much time was spent entering data manually. We chose to omit the line breaks in the HTML file as we felt that if we had followed the exact layout in the HTML render, the readability of the text would be compromised.

Publication:

Ongoing efforts are being made to secure copyright permission for the publication of the digitised material. As most of the newspaper clippings do not bear any information regarding the publisher, author, photographer, or date, it has been difficult to establish the source of the articles. According to the Copyright and Related Rights Act, 2000, “The author of a work shall be the first owner of the copy-right unless... a work... is made by an author in the course of employment by the proprietor of a newspaper or periodical” (23). The two known authors in the collection were ‘Basil Brindley’ (deceased 2019) and Barry Mason (unknown), we decided not to contact the family of Mr. Brindley as the permission should be granted by the publisher and no information could be found on the latter mentioned journalist.

We began by attempting to contact the Irish Independent and Irish Times newspapers, two of the only known publishers throughout the selected documents. The Irish Independent is known to have produced one article in our collection. The Sunday Review, which has been out of print since 1963, was published by the Irish Times (Irish Times, 2018) and occurs once in our collection. As both the Times and the Independent are large national newspapers, contact via phone and email has not been successful. We investigated whether the works could fall under Fair Dealing or be declared as orphan works and contacted the Irish Copyright Licensing Agency (<https://www.icla.ie/>). The ICLA referred to Section 5 of the Orphan Works Regulations 2014 (SI 490 of 2014), which refers to diligent search which shall be carried out by a relevant body, the stipulations for declaring the works in their entirety as orphan works would not be feasible as some of the articles have identifiable rights holders. Furthermore, the ICLA stated that Orphan Works Legislation “would normally apply to the works in the collection of a library, archive or cultural heritage institution.” (S. Holman, personal communication, June 6th, 2022), that is, a relevant body. The ICLA also communicated that the project is unlikely to fall under a Fair Dealing exception (S. Holman, personal communication, June 6th, 2022). Additionally, we contacted the National Library of Ireland (<https://www.nli.ie/>) via phone, and a staff member recommended that a disclaimer may be used if the works were to be published. (J. Furlong, personal communication, May 2nd, 2022). Considering the myriad of unknown sources and the above restrictions, we have decided that we will not, for the foreseeable future, publish our project.

Documentation and delivery media and distribution forms:

Documentation for this project has been kept in the form of a log, detailing the progress of each team member with accompanying dates and information related to the work carried out. We used Github to document the progress of the markup pages (XML, XSL, HTML, CSS). Throughout the project, we have engaged in multiple meetings via Zoom and in person to discuss and make decisions on various aspects of the digitisation process. The summation of the above has been detailed in this report and was written by both members.

Division of labour:

Phase	Jacqueline Williams	Elisabeth Nonstad	Combined hours
Studying relevant course literature and researching other literature related to the project	49	57	106
Contacting various bodies related to copyright	15	-	15
Team meetings - strategy, workflow, and content related decisions	20	20	40
Reviewing, reading and selection of materials	17	33	50
Scanning and image capture	13	-	13
Image processing and organisation	16	37	53
Transcription and OCR	15	22	37
Meta-data research	16	15	31
Text transcription markup	25	30	55
TEI header	21	20	41
Preparing HTML and CSS files from templates	15	-	15
Edits and Adjustments to the HTML and CSS	6	-	6

XSL Transformation and edits	16	12	28
Writing and editing the report	26	25	51
Total hours	270	271	541

The project's relevance in the area of cultural heritage digitisation:

The scrapbook and the accompanying photographs constitute a small-scale documentation of the motor cross community in Dublin and surrounding areas in the 1950s, from the perspective of the creator of the artefact, Ian Williams (grandfather of Jacqueline Williams). Lindley (2012) puts forward an argument for the relationship between an individual's own memories that construct a sense of personal history and how memory is produced through the processing of archives and the creation of new, in this case digital, artefacts, in the latter case, positioning memory as a resource. The images, text, and in a broader sense, the artefact itself captures Williams' own memories, which function as the resources from which the digital facsimile is made. The uniqueness of the narrative enhances the scrapbook's relevance for cultural heritage and its digitised forms. It can be argued that the narratives of individuals "form an important part of our identity and communication, providing an opportunity to explore self, culture and others" (Hausknecht et al., 2018, p. 2714). While researching and collecting metadata on motocross in Dublin during the 1950s, there was little to be found online and even less related to Ian Williams.

Hausknecht et al., (2018) insists on the importance of 'creating legacy', both for family but also for society in general, this is especially pertinent to our project, where we have perceived a general lack of available documentation. Although this can be viewed as a personal project, it is also an important type of contemporary documentation which Berkaak (2002) mentions in his article "Samtidsdokumentasjon: En spøkelseshistorie". The language used by the journalists in the newspaper clippings, the techniques of the photographers and the events in this scrapbook, are all part of the artefact we have chosen to digitise. In terms of selection and relevance it can be argued that "there is nothing created by the hands of mankind that is of not potential research value for humanities scholars, even the humblest scrap of data" (Schreibman et al., 2004, p. 588), with the current overabundance of data, this assertion may not be as appropriate nearly two decades on. However, the relevance of ordinary citizens'

narratives and experiences, especially against the backdrop of a decade in Ireland marred by a severe macroeconomic crisis and the highest rates of emigration in the 20th century (Bielenberg & Ryan, 2013) gives this project adequate grounds for preservation.

In the article “Fotoalbum - digital praksis og prøvelse” (Photoalbum - digital practices and tribulation) (Bratland, 2018), Nina Bratland discusses photo albums as a historical artefact in museums, and the digitisation of such albums. Scrapbooks and photo albums are personal, they tell a story and give a snapshot of history. The act of curating the photographs in an album is also a part of the photo album as a historical artefact. In the case of our scrapbook, the newspaper clippings together with photographs tell a story of the actual happenings during the motor races (results lists, injuries, which motorcycle each rider drove, etc.), but the language used, the details the journalist focused on and the black and white photographs, is a direct snapshot of the 1950's Dublin and the motor cross community. Art historian Anna Dahlgren (Bratland, 2018, p. 62) compares photo albums to other image-based forms of disseminating cultural and historical content, for instance magazines, newspapers, books, exhibitions and so forth, and emphasises on photo albums' “intermedial character” (p. 62). In addition, the person who made the scrapbook, subtly gives the viewer his or her viewpoint through the organising of it. As such, it is both a personal project, but it is also very much a part of the cultural heritage of Dublin.

A critical analysis of the choices and methods used:

Several pages of the original scrapbook, the artefact that is the subject of this project, and many of the loose photographs accompanying it, have been damaged through wear and tear, improper storage, and accidental spills. One of the preservation considerations that is offered as part of the Digital Imaging Tutorial (Cornell University Library, 2003) relates to whether the digital surrogate would reduce further handling of the original, this was a strong motivating factor due to both the age and the condition of the materials. For this project, we have chosen image capture (scanning) of the pages and loose photographs themselves, we have used OCR technology to transcribe the text from the pages and encoded the texts using TEI to describe the contents of the texts. By digitising these artefacts in this manner, they will be preserved in a way that is not possible with the physical originals in their current state.

Regarding selection, we decided on a set of ten photographs and ten scrapbook pages which contained varying amounts of content. It is the aim of Jacqueline Williams that the entirety of the collection will be digitised in the future, however, for the purpose of this project, we had to take a methodological approach to selection based on several parameters. It was of importance to be aware how the chosen documents would affect how this new digitised form would be perceived, in relation to the original. A study by Lindley (2012), although primarily focused on preservation of artefacts within the family, posits many interesting points related to the tension between personal and family memory. Family archives are not usually “a neatly packaged set. They need to be identified and stored... and this entailed making decisions about what to add, maintain, and discard” (Lindley, 2012, p.21). A decision was made to include the less damaged pages while still displaying a narrative that encapsulated a varied depiction of the subject's time as a motocross racer, detailing both the wins and losses, to somewhat align personal and family memory.

According to Manžuch (2017), a digitised version of such artefacts like the one in this project cannot “reflect” all features of the original. It may be more apt to align with the argument posed by Björk (2015) whereby the digitised version is viewed more as a copy or ‘instance’ of the original. In relation to materiality, we were aware that, ultimately, elements of the original would be lost and creating a visceral and accurate reproduction was impossible. It can be argued that “no true representation of three-dimensional space can be achieved within the two-dimensional confines of a computer screen” (*Schreibman et. al.*, 2004, p. 493). However, taking these points into account, we were still mindful to work towards a more accurate representation including a full pictorial representation of the scrapbook by including the page itself instead of solely the photographs and clippings which are affixed to it. This was especially pertinent for the TEI structure, which was revised during the project. One “aim of digital conversion might be to capture the content of a source without necessarily capturing its form. So, an edition of the work of a literary author might be rekeyed and re-edited in electronic form without particular reference to the visual characteristics of an existing print or manuscript version.” (*Schreibman et. al.*, 2004, p. 493). We considered what information was important to include and decided against this approach, instead creating a structure that is more inclusive of the photographic content and the physicality of the scrapbook itself. We viewed this as important information to include as its value lies in its connection to its creator to which this project also serves as an homage.

In the case of this scrapbook, however, parts of it were quite damaged, and without proper storage and restoration, it would continue to deteriorate. By scanning the artefact and enhancing the files to reflect the original, the features of the original are reflected in the digitised version. In addition, by being able to transcribe the texts (both handwritten notes and the texts from the newspaper clippings), while information about the happenings described in them still was available to us, we were able to give context to the content.

In Perry Willetts essay, (ch. 18, A Companion to Digital Humanities, Schreibman, et al., 2004) he describes advantages to digital representations, such as searchability, ease of manipulation and editing especially when used by people with disabilities, for example, visual impairments. This is supported by Renear (2004) who mentions using descriptive markup language as a method to distinguish between different things with the same name, which is in turn helpful for all audiences. In our case, by actively using tags in the text encoding of the texts found in the scrapbook, it is easier for future readers, with little to no previous knowledge of motorsport, to understand the contents of the newspaper clippings. In our project, descriptive markup lets readers of the document easily understand what the different abbreviations and proper names are.

Conclusion:

This project has raised many interesting questions regarding the nature of digitisation, what is lost, what can be retained, especially within the confines of limited time, budget, and expertise. The idea of memory and how the history of an individual is communicated through the digitised document is also of central importance when considering how to present a reproduced artefact. In regard to technical aspects, we hope that by keeping the digital alterations minimal, the integrity of the document was preserved. Questions of selection strategies, what should be encoded, and even how the text encoding should be structured were integral to forming how the theoretical concerns would be resolved. It is evident that the process of digitisation itself possesses many facets, the chain of digitisation has many links and to digitise in a way that adheres to best practices and considers the conceptual aspect of the process requires extensive resources. This feeds into the trade-offs that exist in the field of mass digitisation. Due to the small scale of this collection, it is the intention to digitise the entire scrapbook and accompanying pictures and to publish the site, pending copyright approval. In terms of scalability and the expansion of this work, the aim would be to provide more detailed metadata that could benefit not only the family of the subject but also the Irish

motocross community and in a wider sense make a small contribution to the sporting history and cultural heritage sector in Ireland.

References:

- Bielenberg, & Ryan, R. (2013). *An economic history of Ireland since independence*. Routledge. <https://doi.org/10.4324/9780203094952>
- Björk, L. (2015). *How reproductive is a reproduction? Digital transmission of text-based documents*. Borås: University of Borås.
- Bratland, N., 2018, Fotoalbum, Digital praksis og prøvelse. *Mediehistorisk tidsskrift*. No. 2 2018 (no. 30). pp. 61-73) <https://dms-cf-06.dimu.org/file/013AksaFq4bH> (In Norwegian)
- Breen, J. and O'Brein, M., (2018, November 5). *The Sunday Papers: From investigative brilliance to scurrilous gossip*. The Irish Times. Retrieved from <https://www.irishtimes.com/culture/tv-radio-web/the-sunday-papers-from-investigative-brilliance-to-scurrilous-gossip-1.3681465>
- Copyright and Related Rights Act , Irish Statute Book, 28 s. (2000). Retrieved from <https://www.irishstatutebook.ie/eli/2000/act/28/enacted/en/pdf>
- Cornell University Library (2003). *Moving Theory into Practice: Digital Imaging Tutorial*. Retrieved from <http://preservationtutorial.library.cornell.edu/>
- Hausknecht, S., Vanchu-Orosco, M., & Kaufman, D. (2019). Digitising the wisdom of our elders: connectedness through digital storytelling. *Ageing and Society*, 39(12), 2714–2734. <https://doi.org/10.1017/S0144686X18000739>
- Johansen, A., Losnedahl, K. G., Ågotnes, H-J. (red). 2002. Tingenes Tale: Innspill til museologi. Bergen Museums skrifter nr. 12. Ch. Samtidsdokumentasjon: En spøkelseshistorie. Berkaak, O. A. (pp. 183-191) (In Norwegian)

Lindley, S. E. (2012). *Before I Forget*: From Personal Memory to Family History. *Human-Computer Interaction*, 27(1-2), 13–36.
<https://doi.org/10.1080/07370024.2012.656065>

Manžuch, Z. (2017). Ethical Issues in Digitization of Cultural Heritage. *Journal of Contemporary Archival Studies*, 4 (2); article 4
[<http://elischolar.library.yale.edu/jcas/vol4/iss2/4>]

Schreibman, S., Siemens, R., & Unsworth, J., (Eds.) (2004). *A Companion to digital humanities*. Oxford: Blackwell.

- Ch. 17: Allen Renear, *Text encoding*
- Ch. 18: Perry Willett, *Electronic Texts: Audiences and Purposes*
- Ch. 32: Marilyn Deegan and Simon Tanner, *Conversion of Primary Sources*
- Ch. 37: Abby Smith. *Preservation*

TEI: Text Encoding Initiative. (Accessed May and June, 2022). Retrieved from
<https://tei-c.org>

University of Virginia Library. (n.d.) TEI Encoding Guidelines. Retrieved from
<https://dcs.library.virginia.edu/digital-stewardship-services/tei-encoding-guidelines/#newspapers>

Appendix:

1.1 Physical Description:

Object dimensions: scrapbook: photographs: (mixed)

Object type: Scrapbook, mixed content (newspaper articles, photographs, handwritten and typed lettering on paper)

Binding: string bound

Total number of items: 29 pages and 15 photographs

Additional: Total number of items digitised (at the time of this report): 10 pages and 10 photographs.

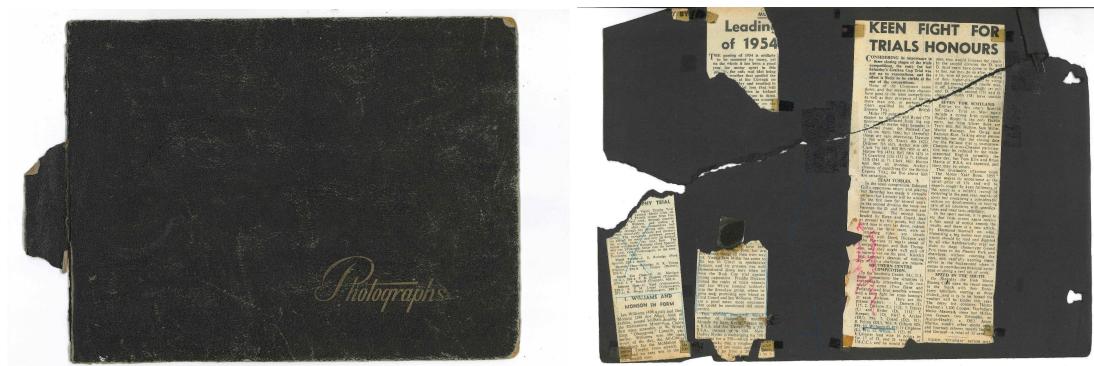


Fig 1.1 A scanned image of the scrapbook's front cover (left) and a severely damaged page from the collection that was not chosen for digitisation (right).

1.2 Technical parameters:

Scanner model: RICOH MP C3504ex Colour Laser Multifunction Printer

- Colour profile: full colour
- DPI: 600

(Earlier scanning attempts: BookEye 2 Plus, PDF format)

Photo-editing software: Adobe Photoshop 2022

OCR: ABBYY FineReader PDF

Text Encoding: Oxygen XML Editor and Sublime Text

HTML, CSS: Sublime Text and Visual Studio Code

Communication and project management: Zoom, WhatsApp, GitHub



Fig 1.2 BookEye 2 Plus book scanner (left) and resulting scan (top right) compared to RICOH scan (lower right)

1.3 Selection criteria:

Below we have a list of the selection of the pages from the scrapbook for digitisation, with short notes that formed the basis for the larger discussion of the selection criteria used for each of the different pages.

Scrapbook	ID	Reason
Order	ID	Reason
1	sb_03	Varied content, lacking dates
2	sb_04	good condition, featuring another rider, mentions brother

3	sb_06	includes date and newspaper photo
4	sb_08	good content of overall achievements
5	sb_09	good quality and content, little damage, record of wins
6	sb_10	same as above
7	sb_11	same as above (11 mentions Tye - from previous photo)
8	sb_15	details of damaged vehicle and loss
9	sb_17	details of fall
10	sb_20	good picture and content, metadata related to pics

Fig 1.3 A table featuring rough notes used in the selection process

1.4 Image Processing:

As discussed above, editing of the original scans were kept to a minimum, however some editing was done, and the below images shows the process.

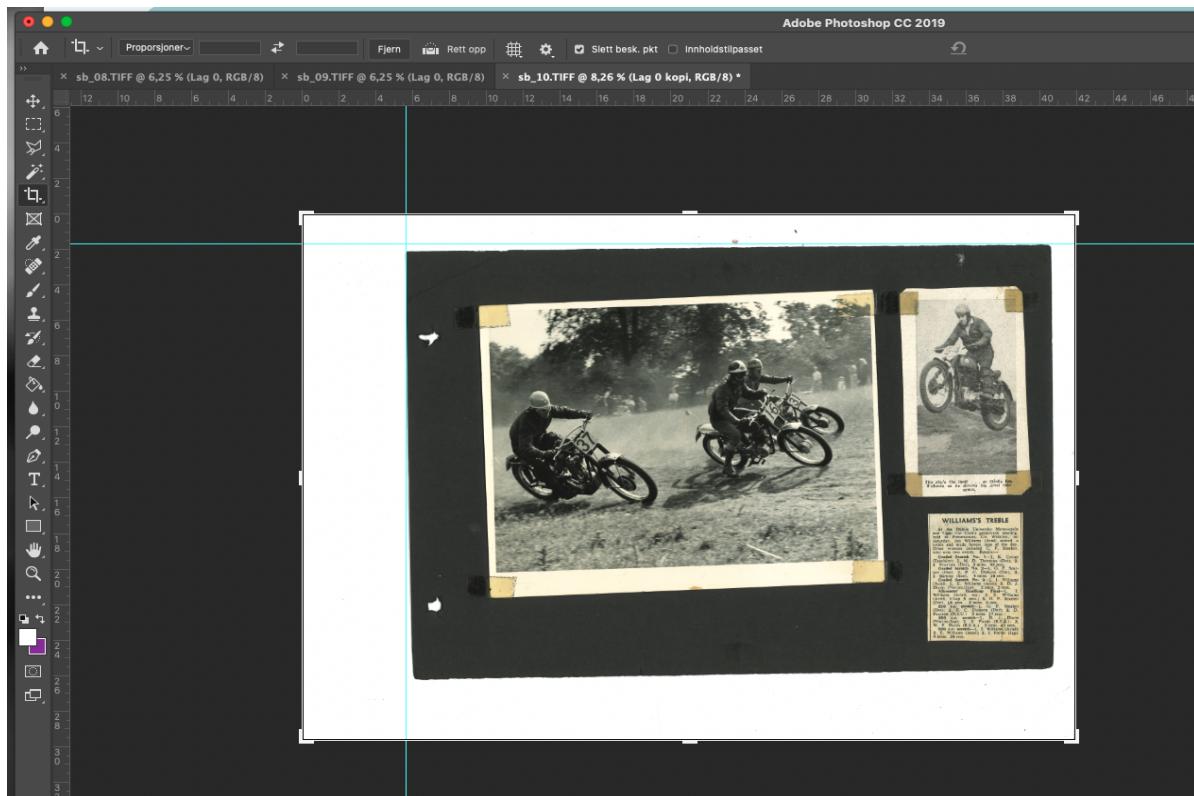
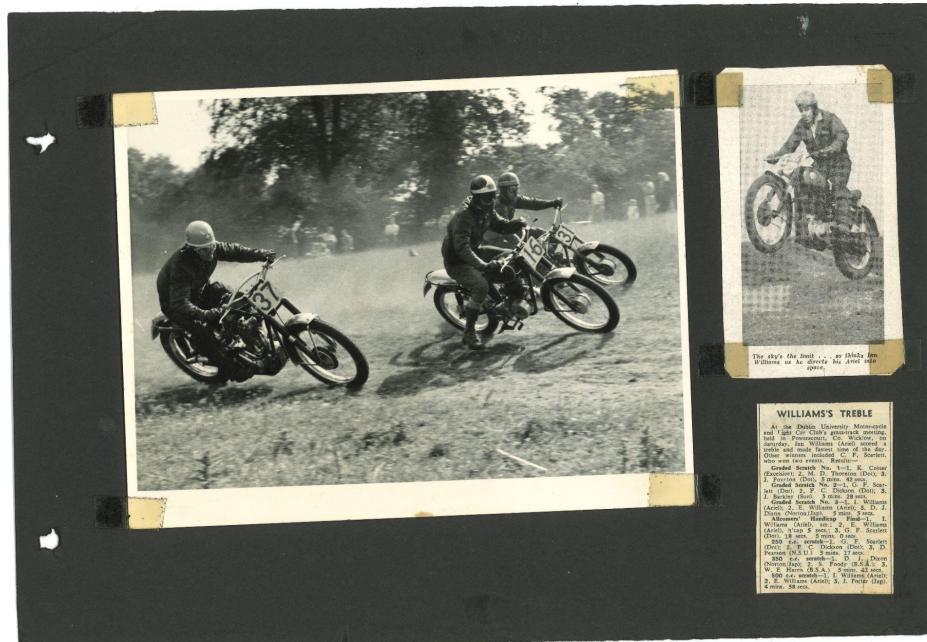


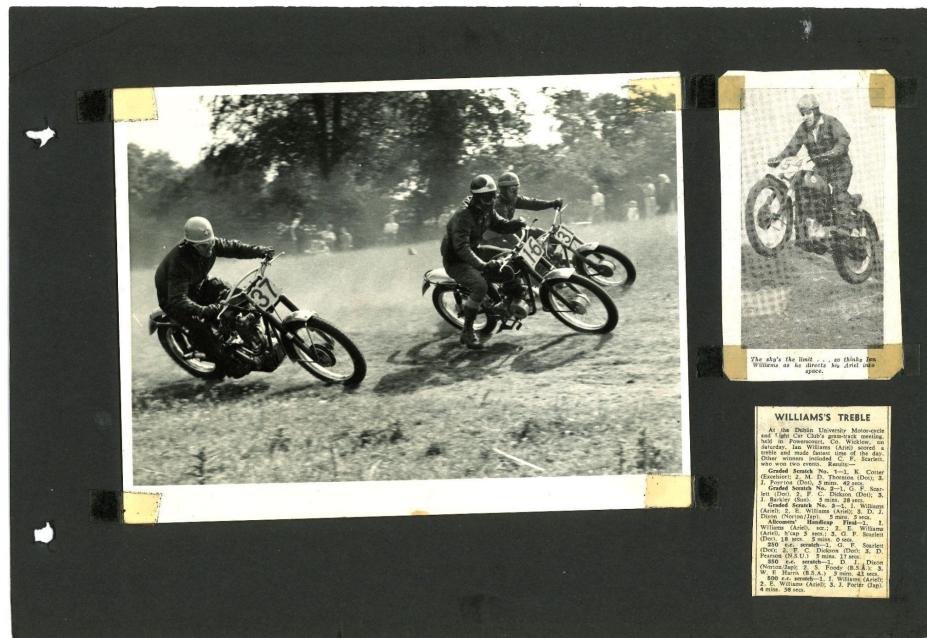
Fig. 1.4.1.1 Screenshot from Adobe Photoshop: Levelling the scrapbook pages before cropping the image.



Fig. 1.4.1.2 Screenshot from Adobe Photoshop with included layers (original/edit layer/light & contrast layer). Not shown: Additional black & white layer, which was added in case we wanted to use black and white versions of the images.

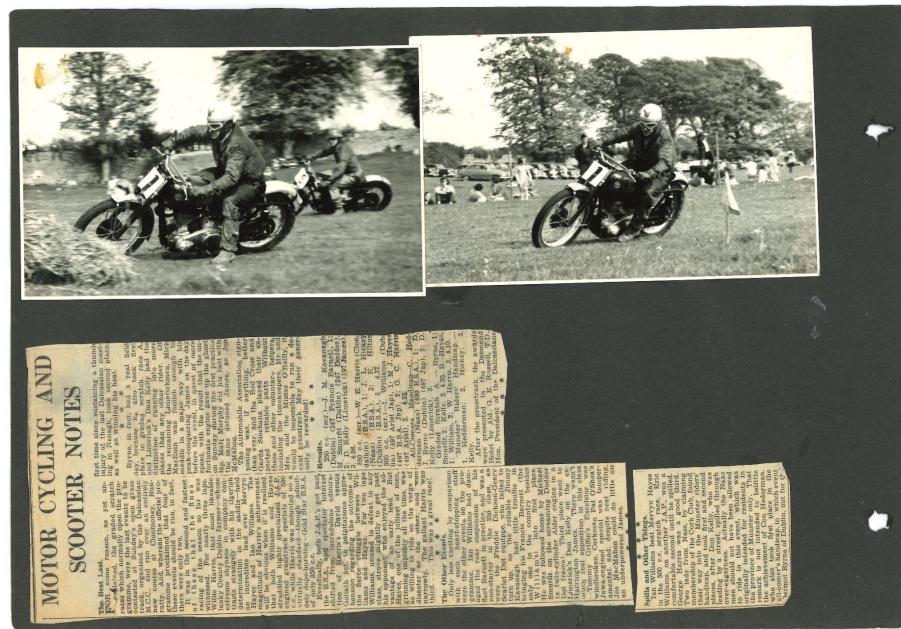


Original



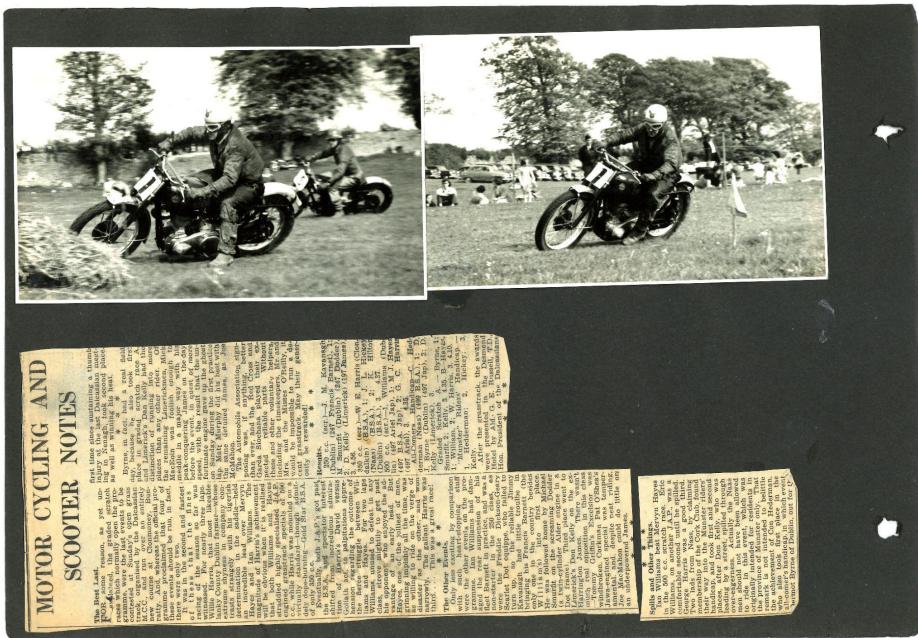
Edited

Fig 1.4.2.1 Examples of original and edited versions of the scrapbook pages.



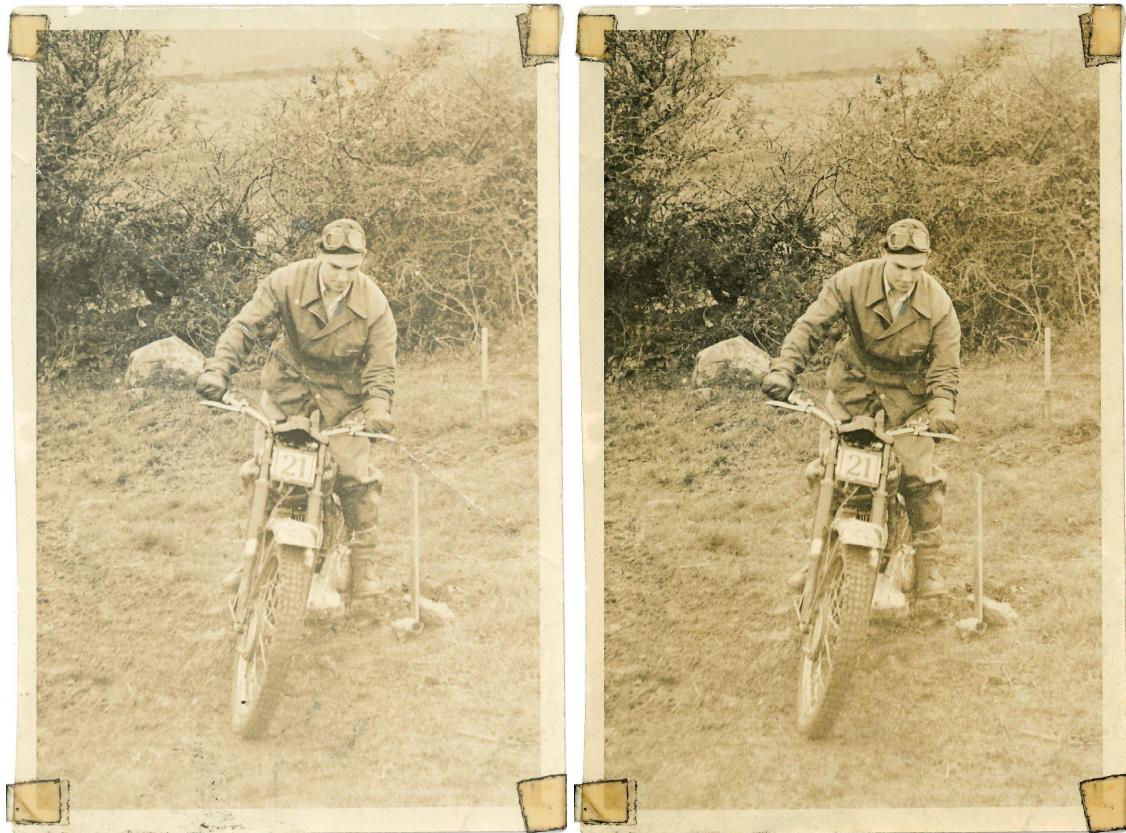
MOTOR CYCLING AND SCOOTER NOTES

Original



Edited

Fig. 1.4.2.2 Examples of original and edited versions of the scrapbook pages.



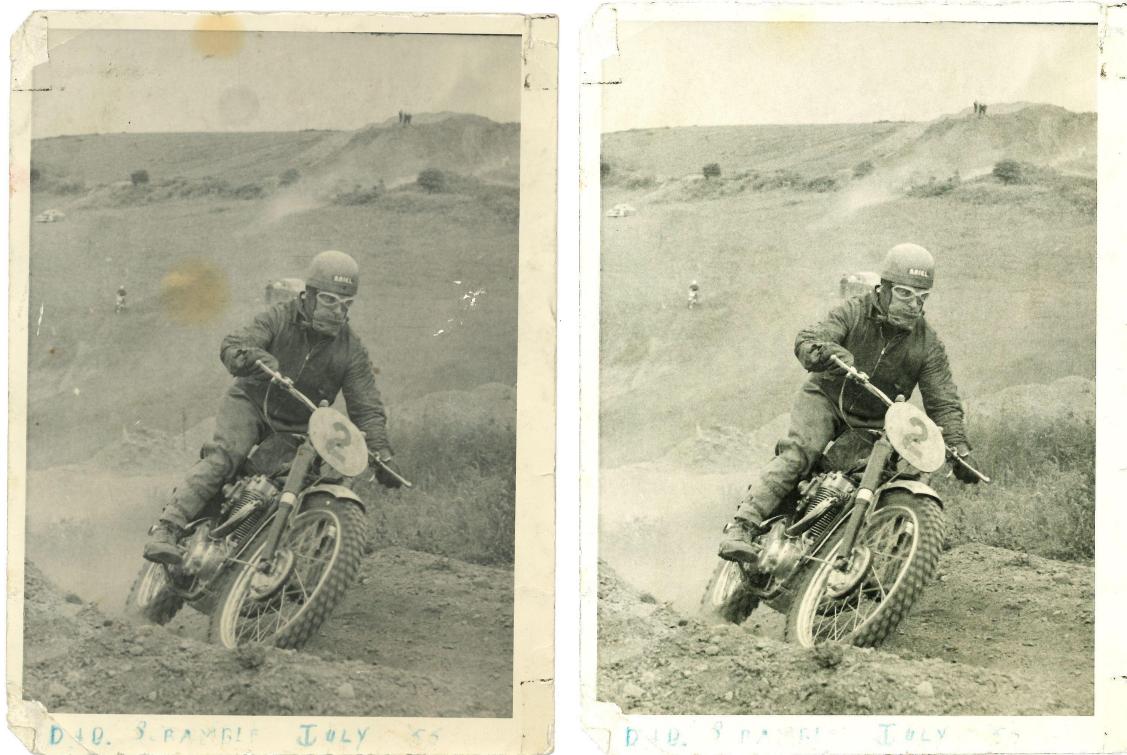


Fig. 1.4.3 Examples of edited loose photographs:

Original to the left and the edited image to the right. The aim was not to edit the images to be without damage or traces of time, but to remove distracting elements without disturbing the integrity of the original.

1.5 Optical Character Recognition (OCR):

Below is a representation of an OCR result and subsequent edited text. ABBYY FineReader PDF was used to create plain text files formatted to UTF-8(Unicode Standard). The following pages utilised the OCR output (sb_03, sb_04, sb_06, sb_11, sb_20) the remaining pages were transcribed manually. It can be observed that many errors are a result of minor ink residue (due to the printing methods of the time) and damage to the source material (tears, stains and worn or faded areas)

OCR Result sb_04_03	Original sb_04_03	Edited OCR Text sb_04_03
<p>Brother</p> <p>I-lood</p> <p>IAN WILLIAMS returned to grass track racing in sparkling form and at the Dublin University grass track races at Powerscourt yesterday his victories included ~ an exciting finish in the handicap final where he beat his brother Eric by half a wheel.</p> <p>Graded Scratch Race A — 1. K. Cotter (248 Excelsior); 2, M. D. Thornton (197 Dot); 3, J. D. Poynton (197 Dot). 5 mins. 42 secs.</p> <p>Graded Scratch Race B — 1, G. F. Scarlett (197 Dot); 2, F. C. Dickson (197 Dot); 3, J. Barkley (197 Dot). 5 mins. 28 secs. Graded Scratch Race C—1, R. Williams (497 Aerial); 2, E. Williams (497 Aerial); 3, D. J. Dixon (344 Norton Jap). 5 mins. 5 secs.</p> <p>Open Handicap—Heat 1 — 1 W. Hogan (197 Dot); 2, I. Williams (497 Aerial); 3 L. Armstrong (500 Jap). 5 mins. 47 secs. Heat 2—1, F. C. Dixon (197 Dot); 2, S. Foody (348 B.S.A.); 3, J. D. Roche (348 B.S.A.). 5 mins. 22 secs. Heat 3—1, E. Williams (497 Aerial); 2, G. F. Scarlett (197 Dot); 3, K. Cotter (248 Excelsior). 5 mins. 7 secs. Handicap Final — 1, I. Williams (497 Aerial); 2, E. Williams (497 Aerial); 3, G. F. Scarlett (197 Dot). 5 mins. dead.</p> <p>250 c.c. Scratch Race—Heat 1—1, D. Pearson (248 N.S.U.). Heat 2—G. F. Scarlett (197 Dot). Final —1, G. F. Scarlett (197 Dot); 2, F. C. Dickson (197 Dot); 3, D. Pearson (248 N.S.U.). 5 mins. 17 secs.</p> <p>350 c.c. scratch race—1, D. Dixon (344 Norton Jap); 2, S. Foody (348 B.S.A.); 3, W. E. Harris (348 B.S.A.). 5 mins. 11s. 500 c.c. scratch race—1, I. Williams (497 Aerial); 2, E. Williams (497 Aerial); 3, J. Porter (500 Jap). 4 mins. 58 secs.</p> <p>Results of Autocross</p> <p>THE M.G. Car Club held an autocross at Lynch Park, Sagart, Co. Dublin yesterday. The course was ½ mile long and each race was over 4 laps. Results:—</p> <p>Saloon Class—Scratch — J. C. Millarad (Fiat 600), 4 mins. 17 secs.</p> <p>Handicap—Miss E. Fenston (V.W.), 4 mins. 22 secs.</p> <p>Open Class—Scratch — M. Hayes (Triumph TR2), 4 mins. 46 secs.</p> <p>Handicap—M. Kennedy (Triumph TR2), 4 mins. 45 secs.</p> <p>Specials—Scratch — B. K. Hood (Ford Special), 4 mins. 9 secs.</p> <p>Handicap—S. V. Baker (Ford Special), 4 mins. 17 secs.; R. Redmond (Ford Special), 4 mins. 18 secs.</p>	 <p>Ian Williams Beats Brother</p> <p>IAN WILLIAMS returned to grass track racing in sparkling form and at the Dublin University grass track races at Powerscourt yesterday his victories included ~ an exciting finish in the handicap final where he beat his brother Eric by half a wheel.</p> <p>Graded Scratch Race A — 1. K. Cotter (248 Excelsior); 2, M. D. Thornton (197 Dot); 3, J. D. Poynton (197 Dot). 5 mins. 42 secs.</p> <p>Graded Scratch Race B — 1, G. F. Scarlett (197 Dot); 2, F. C. Dickson (197 Dot); 3, J. Barkley (197 Dot). 5 mins. 28 secs. Graded Scratch Race C—1, R. Williams (497 Aerial); 2, E. Williams (497 Aerial); 3, D. J. Dixon (344 Norton Jap). 5 mins. 5 secs.</p> <p>Open Handicap—Heat 1 — 1 W. Hogan (197 Dot); 2, I. Williams (497 Aerial); 3 L. Armstrong (500 Jap). 5 mins. 47 secs. Heat 2—1, F. C. Dixon (197 Dot); 2, S. Foody (348 B.S.A.); 3, J. D. Roche (348 B.S.A.). 5 mins. 22 secs. Heat 3—1, E. Williams (497 Aerial); 2, G. F. Scarlett (197 Dot); 3, K. Cotter (248 Excelsior). 5 mins. 7 secs. Handicap Final — 1, I. Williams (497 Aerial); 2, E. Williams (497 Aerial); 3, G. F. Scarlett (197 Dot). 5 mins. dead.</p> <p>250 c.c. Scratch Race—Heat 1—1, D. Pearson (248 N.S.U.). Heat 2—G. F. Scarlett (197 Dot). Final —1, G. F. Scarlett (197 Dot); 2, F. C. Dickson (197 Dot); 3, D. Pearson (248 N.S.U.). 5 mins. 17 secs.</p> <p>350 c.c. scratch race—1, D. Dixon (344 Norton Jap); 2, S. Foody (348 B.S.A.); 3, W. E. Harris (348 B.S.A.). 5 mins. 11s. 500 c.c. scratch race—1, I. Williams (497 Aerial); 2, E. Williams (497 Aerial); 3, J. Porter (500 Jap). 4 mins. 58 secs.</p> <p>Results of Autocross</p> <p>THE M.G. Car Club held an autocross at Lynch Park, Sagart, Co. Dublin yesterday. The course was ½ mile long and each race was over 4 laps. Results:—</p> <p>Saloon Class—Scratch — J. C. Millarad (Fiat 600), 4 mins. 17 secs.</p> <p>Handicap—Miss E. Fenston (V.W.), 4 mins. 22 secs.</p> <p>Open Class—Scratch — M. Hayes (Triumph TR2), 4 mins. 46 secs.</p> <p>Handicap—M. Kennedy (Triumph TR2), 4 mins. 45 secs.</p> <p>Specials—Scratch — B. K. Hood (Ford Special), 4 mins. 9 secs.</p> <p>Handicap—S. V. Baker (Ford Special), 4 mins. 17 secs.; R. Redmond (Ford Special), 4 mins. 18 secs.</p>	<p>Ian Williams</p> <p>Beats Brother</p> <p>IAN WILLIAMS returned to grass track racing in sparkling form and at the Dublin University grass track races at Powerscourt yesterday his victories included an exciting finish in the handicap final where he beat his brother Eric by half a wheel.</p> <p>Graded Scratch Race A — 1, G. F. Scarlett (197 Dot); 2, F. C. Dickson (197 Dot); 3, J. Barkley (197 Dot). 5 mins. 28 secs. Graded Scratch Race B — 1, G. F. Scarlett (197 Dot); 2, F. C. Dickson (197 Dot); 3, J. Barkley (197 Dot). 5 mins. 28 secs. Graded Scratch Race C—1, R. Williams (497 Aerial); 2, E. Williams (497 Aerial); 3, D. J. Dixon (344 Norton Jap). 5 mins. 5 secs.</p> <p>Open Handicap—Heat 1 — 1 W. Hogan (197 Dot); 2, I. Williams (497 Aerial); 3 L. Armstrong (500 Jap). 5 mins. 47 secs. Heat 2—1, F. C. Dixon (197 Dot); 2, S. Foody (348 B.S.A.); 3, J. D. Roche (348 B.S.A.). 5 mins. 22 secs. Heat 3—1, E. Williams (497 Aerial); 2, G. F. Scarlett (197 Dot); 3, K. Cotter (248 Excelsior). 5 mins. 7 secs. Handicap Final — 1, I. Williams (497 Aerial); 2, E. Williams (497 Aerial); 3, G. F. Scarlett (197 Dot). 5 mins. dead.</p> <p>250 c.c. Scratch Race—Heat 1—1, D. Pearson (248 N.S.U.). Heat 2—G. F. Scarlett (197 Dot). Final —1, G. F. Scarlett (197 Dot); 2, F. C. Dickson (197 Dot); 3, D. Pearson (248 N.S.U.). 5 mins. 17 secs.</p> <p>350 c.c. scratch race—1, D. Dixon (344 Norton Jap); 2, S. Foody (348 B.S.A.); 3, W. E. Harris (348 B.S.A.). 5 mins. 11s. 500 c.c. scratch race—1, I. Williams (497 Aerial); 2, E. Williams (497 Aerial); 3, J. Porter (500 Jap). 4 mins. 58 secs.</p> <p>Results of Autocross</p> <p>THE M.G. Car Club held an autocross at Lynch Park, Sagart, Co. Dublin yesterday. The course was ½ mile long and each race was over 4 laps. Results:—</p> <p>Saloon Class—Scratch — J. C. Millarad (Fiat 600), 4 mins. 17 secs.</p> <p>Handicap—Miss E. Fenston (V.W.), 4 mins. 22 secs.</p> <p>Open Class—Scratch — M. Hayes (Triumph TR2), 4 mins. 46 secs.</p> <p>Handicap—M. Kennedy (Triumph TR2), 4 mins. 45 secs.</p> <p>Specials—Scratch — B. K. Hood (Ford Special), 4 mins. 9 secs.</p> <p>Handicap—S. V. Baker (Ford Special), 4 mins. 17 secs.; R. Redmond (Ford Special), 4 mins. 18 secs.</p>

Aerial); 2 E. Williams (497 Aerial); 3, D. J. Dixon (344 Norton Jap). 5 mins. 6 secs. <missing> Hogan (197 Dot); 2, I. Williams 497 Aerial); 3 L, Armstrong (500 Jap). 5 mins. 47 secs. Heat 1 . F. C. Dixon (197 Dot); 2. S. Foody (348 B.S.A.); 3. J- D. Roche (348 B.S.A.). 5 mins. 22 se <?s. Heat 0— I E Williams (497 Aerial); 2, G. B. S h arlett (197 Dot); 3. K. Cotter (244 Excelsior). 5 mins. (secs. Handi- can Final — 1. I Williams (497 Aerial) • 2 E. . Williams (497 Aerial); 3, G. F. Scarlett (197 Dot). 5 mins, dead. <missing, yellow highlight> Results of Autocross	Race C—1, R. Williams (197 Aerial); 2 E. Williams (497 Aerial); 3, D. J. Dixon (344 Norton Jap). 5 mins. 5 secs. Open Handicap- Heat 1 - 1 W. Hogan (197 Dot); 2, I. Williams 497 Aerial); 3 L, Armstrong (500 Jap). 5 mins. 47 secs. Heat 2—1. F. C. Dixon (197 Dot); 2. S. Foody (348 B.S.A.); 3. J- D. Roche (348 B.S.A.). 5 mins. 22 se <?s. Heat 0— I E Williams (497 Aerial); 2, G. B. Scarlett (197 Dot); 3. K. Cotter (244 Excelsior). 5 mins. 7 secs. Handi- cap Final — 1. I Williams (497 Aerial); 2 E. Williams (497 Aerial); 3, G. F. Scarlett (197 Dot). 5 mins. dead. 250 c.c. Scratch Race—Heat
---	---

<p>fp<HE M.G. Car Club held an IX autocross At Lynch Dark. Sag-1 cart, Co. Dublin yesterday. The course was i ncke long and each race was over 4 laps. Results Saloon Class — Scratch — J. C. Miilarad (Fiat 600). 4 mins. 17 secs. Handicap—Miss E. Penston (V.W.), 4 mins. 22 secs. , Open Class—Scratch—M. Hayes (Triumph TR2). 4 mins, 46 secs. Handicap—M. Kennedy (Triumph TR2). 4 mins. 45 secs. - Specials—Scratch—B. » *1 (Ford Special), 4 mins. ----- Handicap—S. V. Baker _ Jord Special). 4 mins. 17 sees; K Red- mond (Ford Special), 4 mms. 18 gees. — <missing>Scratch Race—Heat 1—</p>		<p>1— Pearson (248 N.S.U.). Heat 1s. D. Pearson (248 N.S.U.). Heat 2- G. F. Scarlett (197 Dot.). Final --1, G. F. Scarlett (197 Dot.); 2. F C. Dickson (197 Dot.); 3. D. Pear- son (248 N.S.U.). 5 mins. 17s. 350 c.c. scratch race—1, D. Dixon (344 Norton Jap.); 2, S. Foody (348 B.S.A.); 3, W. E. Harris (348 B.S.A.). 5 mins. 11s. 500 c.c. scratch race—1. I. Williams (497 Aerial); 2, E. Williams (497 Ariel); 3, J. Porter (500 Jap.). 4 mins 58 secs. Results of Autocross THE M.G. Car Club held an autocross At Lynch Dark. Sag- gart, Co. Dublin yesterday. The course was 1/2 mile long and each race was over 4 laps. Results: - Saloon Class — Scratch — J.</p>
---	--	--

Pearson (248 N.S.U.). Heat <missing>. Scarlett (197 Dot.). Final „F. Scarlett. (197 Dot.); 2. P C. Dickson (197 Dot.); 3. D. Pear- gon (248 N.S.U.). 5 mins. 17s. 350 c.c. scratch race—1, D. Pixon (344 Norton Jap.); 2, S. . Foody (348 B.S.A. > 3, W. E. Harns (348 B.S.A.). 5 mins. Us. 500 c.c. scratch raCe—1. j. Williams (497 Aerial); *. E Williams (497 Ariel); 3, J. Porter (500 Jap.). 4 mins. 58 secs.	C. Millarad (Fiat 600). 4 mins. 17 secs. Handicap—Miss E. Penston (V.W.), 4 mins. 22 secs. , Open Class—Scratch—M. Hayes (Triumph TR2). 4 mins, 46 secs. Handicap—M. Kennedy (Triumph TR2). 4 mins. 45 secs. Specials—Scratch—B. K. Hood (Ford Special), 4 mins. 9 secs. Handicap—S. V. Baker (Ford Special). 4 mins. 17 secs; R. Red- mond (Ford Special), 4 mms. 18 secs.
--	---

Fig 1.4 An OCR comparison

1.6 Text Encoding Amendments:

Below is a representation of the XML before and after the structure was changed.

```

<!--page 1 (sb_03) -->
<facsimile xml:id="page_01" fcs="sb_03_TN_C.FULL.png">
  <surface xml:id="surface_1_1">
    <label>Surface 1.1 of page 1 (sb_03)</label>
    <figDesc>A newspaper article cutting featuring an-in-article photograph </figDesc>
    <graphic xml:id="sb_03_01.jpg" url="assets/img/documents/JPG/sb_03_01.jpg"/>
    <graphic xml:id="sb_03_01_thumb" url="assets/img/documents/Thumbnails/sb_03_TN_C_01.png"/>
  </surface>
  <surface xml:id="surface_1_2">
    <label>Surface 1.2 of page 1 (sb_03)</label>
    <figDesc>A piece of pink paper with text that has been cut. The paper is damaged and missing a section on the right hand side </figDesc>
    <graphic xml:id="sb_03_02.jpg" url="assets/img/documents/JPG/sb_03_02.jpg"/>
    <graphic xml:id="sb_03_02_thumb" url="assets/img/documents/Thumbnails/sb_03_TN_C_02.png"/>
  </surface>
  <surface xml:id="surface_1_3">
    <label>Surface 1.3 of page 1 (sb_03)</label>
    <figDesc>A newspaper article cutting. </figDesc>
    <graphic xml:id="sb_03_03.jpg" url="assets/img/documents/JPG/sb_03_03.jpg"/>
    <graphic xml:id="sb_03_03_thumb" url="assets/img/documents/Thumbnails/sb_03_TN_C_03.png"/>
  </surface>
  <surface xml:id="surface_1_4">
    <label>Surface 1.4 of page 1 (sb_03)</label>
    <figDesc>A black and white photograph of two men race motorcycles in a grass field, in the lead is <persName ref="#IW>Ian Williams</persName>, riding a motorcycle marked number '34' the man behind (unidentified) rides a motorbike marked number '45'. In the background Two onlookers stand next to a parked car. </figDesc>
    <graphic xml:id="photo_1_4.JPG" url="assets/img/documents/JPG/sb_03_04.jpg"/>
    <graphic xml:id="photo_1_4_thumb" url="assets/img/documents/Thumbnails/sb_03_TN_C_04.png"/>
  </surface>

```

Fig 1.5 Original XML structure

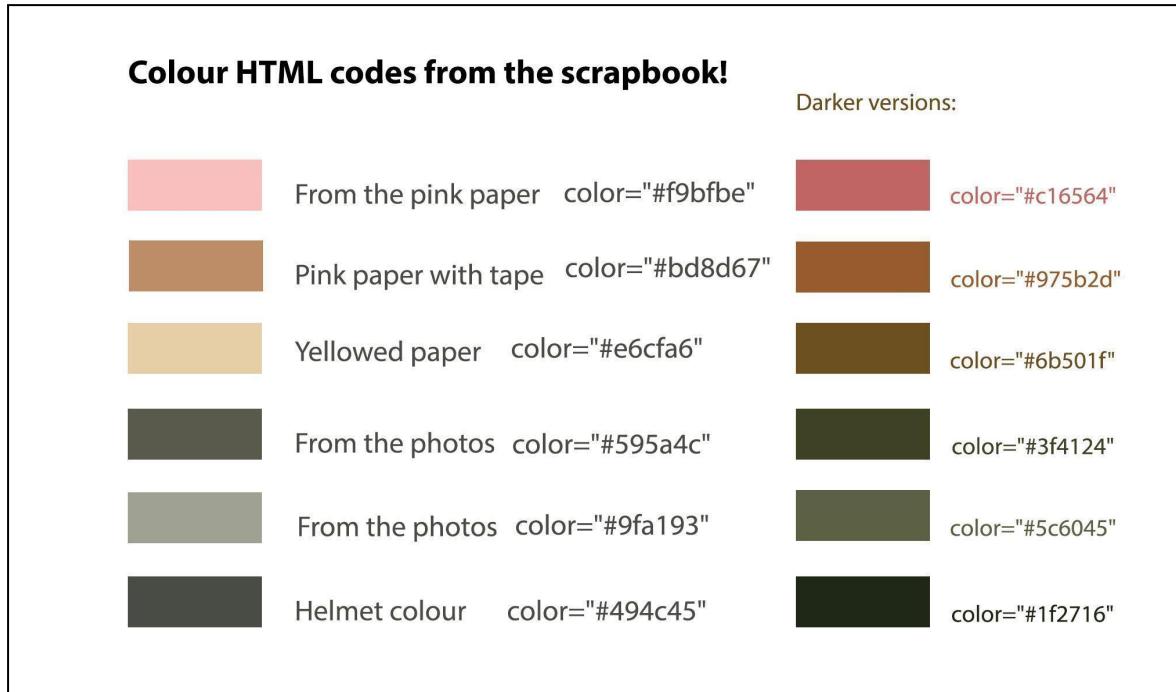
```

<!--page 1 (sb_03) -->
<facsimile xml:id="page_01" fcs="sb_03_TN_C.FULL.png">
  <surface xml:id="surface_1_1">
    <label>Surface 1.1 of page 1 (sb_03)</label>
    <figDesc>A newspaper article cutting featuring an-in-article photograph </figDesc>
    <graphic xml:id="sb_03_01.jpg" url="assets/img/documents/JPG/sb_03_01.jpg"/>
    <graphic xml:id="sb_03_01_thumb" url="assets/img/documents/Thumbnails/sb_03_TN_C_01.png"/>
  </surface>
  <surface xml:id="surface_1_2">
    <label>Surface 1.2 of page 1 (sb_03)</label>
    <figDesc>A piece of pink paper with text that has been cut. The paper is damaged and missing a section on the right hand side </figDesc>
    <graphic xml:id="sb_03_02.jpg" url="assets/img/documents/JPG/sb_03_02.jpg"/>
    <graphic xml:id="sb_03_02_thumb" url="assets/img/documents/Thumbnails/sb_03_TN_C_02.png"/>
  </surface>
  <surface xml:id="surface_1_3">
    <label>Surface 1.3 of page 1 (sb_03)</label>
    <figDesc>A newspaper article cutting. </figDesc>
    <graphic xml:id="sb_03_03.jpg" url="assets/img/documents/JPG/sb_03_03.jpg"/>
    <graphic xml:id="sb_03_03_thumb" url="assets/img/documents/Thumbnails/sb_03_TN_C_03.png"/>
  </surface>
  <surface xml:id="surface_1_4">
    <label>Surface 1.4 of page 1 (sb_03)</label>
    <figDesc>A black and white photograph of two men race motorcycles in a grass field, in the lead is <persName ref="#IW>Ian Williams</persName>, riding a motorcycle marked number '34' the man behind (unidentified) rides a motorbike marked number '45'. In the background Two onlookers stand next to a parked car. </figDesc>
    <graphic xml:id="photo_1_4.JPG" url="assets/img/documents/JPG/sb_03_04.jpg"/>
    <graphic xml:id="photo_1_4_thumb" url="assets/img/documents/Thumbnails/sb_03_TN_C_04.png"/>
  </surface>

```

*Fig 1.6 Updated XML structure*1.7 Design and Aesthetic Choices:

A colour chart was created, elements of which were used in the design of the website.

*Fig 1.5 colour chart*