# Digitization

Simon Shezi



# Introduction

- What is digitization?
- Also referred to as image capture,
- Process of creating a digital representation or image of an original through scanning or digital photography.
- In summary, digitization converts materials that can be read by people to a format that can be read only by machines (digital). Flatbed scanning, digital cameras, planetary cameras, and a number of other devices can be used



# Why digitize

- The main reason to digitize are to enhance access and improve preservation.
- By digitizing collections, institutions can make information accessible that was previously available to a select group of researchers.
- Digitization can also help preserve precious materials.
- Making high-quality digital images available electronically can reduce wear and tear on fragile items.
- This does not mean, however, that digital copies should be seen as a replacement for the original piece



### Selecting a scanner

1. Select your scanner in proportion to its intended use

2. Understand the product specifications

3. Consider the nature of the material to be scanned



# Scanner setting

🐇 EPSON Scan - EPSON GT-1	5000
EPSON Sca	Mode: Professional Mode
_ Settings	
Name:	Current Setting
	Save Delete
Original	
Document Type:	Reflective
Document Source:	Document Table
Auto Exposure Type:	Document
Destination	
🛨 Image Type:	8-bit Grayscale 🔽
Resolution:	600 🔽 dpi
🛨 Target Size:	Briginal 🕑 🔝
Adjustments	
	Reset
<u>Preview</u>	<u>S</u> can
Help Conf	figuration Close

Contents Vol. 1 No. 1 Marc	1978	Ahour Staffrider
Soweto Speaking to Mirian Thuk	2.6	
Mr X, unlicensed fresh produce and poultry dealer	2 Like Herr	er is, let's here it, a wheth of aorts, nes or Mercury - the messenger
Mrs Icath Koae, dieseruker	4 atomotication	ts in classical mythology — he is rtainly as light-fingered as he is
Mis Schenzile Lekoto, market researcher	6 of mexity	ed. A skill uler Acrainer, a bringer En a useful person but slightly
Magawulana, a stirry K.I.S. Nuni, Plessislart, Natal	7-10 Min, hut t	ble. Our cersors may not like they should consider putting up
Two Poems Nkatharo kuMuyayisa, Mpumalanga. Vatal	10 mg at the	A whole new merature is specie- e door, and if our society is to
Rockville, Soweto	11-12 mesages i	it can get the bad acwell as the
Three Poems Mandia Ndlazi	II guan.	au or flot, he is part of the present
Soweth Hijuck, a sterr Miriam Tali	12 phase of 0	our corginon history, nding suit t suid dangerous trains of our late
Four Poems Oups Thando Mchimkuha	21 vene. He a	He is part of the idiom of this superior on Tage VI of our first
Johannesburg	25-27 Legoale of	128 to arrest William 'Checks' fifte Creative Youth Association.
Three Paems Mike Nicol	25 Diephluot The rus	usation which benrs his name has
Van, a straty Peter Wilhelm	26 Attempt to	blished by RAYAN Press in an o respond, us publishers, to the
Two Postna Reg Rummer	27 boon one	se of creative activity which has of the more hopeful signs of
Staffeider. a poeti	28 feacht tun	tts. w writing hus altered the scope
Creative Youth Association, Diepkloof, Soweto	29-35 in whys we	tion of literature in South Attuck chave shift of descriver. The sum of
Three Poems Masilo Isaac Rabothata	29 hut to put	orne is not to impose standarda
Sketch William: 'Cheeks' Legnale	30 which will	I have to shape the future of our
Introducing CYA Masilo Isaac Rabothata	32 A ficerature.	ce of much of the new writing is
From Poems Matsumela Cam Manuka	32 the writer	EVA. This is a two-way have the
Poems I, Maketha, Hanyane Nelson Shikwumbane	3.3 munityse	attempting to voice the com-
l'wo Puems Lebethe R. Lebethe	34 ('An I n	dists authence is the community (ehc?) Community drama, 'say'
Tribute to Victor Nelazilwane Matsemeta Cuin Manuka	34 inspired b	by music, this is the heart of the
Puerus Manishile I., Manaka, J.C. Kgaba, Jcabuth Maubane	35 forms are	re energing in a new mould.
Сарс Тоwn	3(+38 ing in ST	AFFRIDER fues the flag of its
Hero and Bad Mother in Epic, a poem I.M. Coccee	36 groups of	by. We know that there are many writers in Southern Africa whom
Two Poems Keith Gottschalk	37 would we	elevane their contributions. We
Ice-ulation, extracts from a play Muthobi Muthoatse, Pimville	41 magazine	will be selected and edited as far
Durban	Ad-51 Manual 12-44	is prepared for publication by
The Maniage Pontrait, a power	44 editorial h	heard in the usual Senio.
I wo Poems Shabbi Banuohhai	46 strength to	our protey: to encourage and gyo to a new literature based on com-
Widows' Talk, a nory fientude Strauss	47 of contract	unication between Uses writers,
Journeyman, a poers Douglas Livingstone	49 At the san	no menes, and the general puote.
Mrs Ogle's Despair, a poem and a case history Ruth Keech	50 attrobed	individuals, yet find the STAFF.
Old Homes, a powm D. Mphusn, Ca-Rankuwa	S] NILTER	HUMANICA MIANIMATIAN
Five Poems Vapc à Motaua, Mabopune Lust Pretoria	54	
Ngikusongele, a paem L.B.Z. Buthelevi, Dundee, Natal	56	
Th: Day A Leader Died, a story. Auhrey Ngcoho, Edendale, Natal	57	
Photographs Cmar Bedsha	22.39.58	
AL ANIMAL MALE AND ALL AND A	VANCE IN THE REAL	the second secon

#### Image scanned at 24-bit colour





# Image specifications

- Books and journals
  - Front covers with colour 600 dpi (dots per inch) 24-bit RGB (Red, green, blue) colour at 100% of pixel width, resampled to 8-bit paletted
  - Front covers with no colour 600dpi 8-bit grayscale at 100% of pixel width Individual pages: 600dpi 8-bit grayscale at 100% of pixel width



# Image specification cont.

#### Posters

- 300 dpi RGB (24bit), saved as an uncompressed .tif image
- For scanning of posters from slides (unknown original size), scan at 300dpi RGB (24bit), with the short side set at 5000 pixels or if preferred the long side at 7000 pixels. Save as an uncompressed .tif image.



# Image specification cont.

- Photographs
  - Archival colour photographs: 600dpi RGB (24bit) at 100% of pixel width.
  - Photographs on book and journal pages: 200dpi greyscale (8bit) with pixel width reduced to 500 pixels.
  - Archival black and white photographs: 600dpi grayscale (8bit) at 100% of pixel width.



# Image specification cont.

- Archival material (diaries, letters, manuscripts, memorandums etc)
  - All archival material except photographs: 600dpi 8-bit grayscale at 100% of pixel width



#### Enhancements

- Images must be straightened horizontally.
- Archival images retain original size of physical document
- Any marks integral to the original document, annotations, watermarks, etc. should not be removed.



# **Post Processing**

#### PDF (Portable Document Format )creation

- Open Adobe Acrobat or equivalent PDF writer Click Create PDF Button Choose "From Multiple Files ,,."
- Browse to web-ready folder Choose only the 800dpi GIFs Make sure all check boxes are un-ticked Save PDF to the document folder, using folder name, eg: rep19631012.032.009.743.pdf
- Open PDF and check page sequence.
- Use PDF to QC images.



# **OCR** -Optical Character Recognition

- Open ScanSoft Omnipage or equivalent OCR software
- Open tiffs in Omnipage and perform OCR on selected tiffs.
- Save text twice :One for each image file and one for all pages.



#### **OCR within Adobe Acrobat PDF**



# poor recognition

🚦 NoteTab Ligh	t - Nol	Name0	1.sgm										_ 7 🗙
File Edit Search	View M	1odify	Document	- Favorites	Tools Help								
) 🖻 🖪 🖳	🗊 🖆	<b>6</b> 6	<b>3</b> X I	è 🖹 🖻	a 🖻 🖉	) 🗠 🏘 🐁	🖙 😭 ₩	' 🐴 A .	a" 🛍 🧿 😪	🊄 📭			
	142.003.0	008.Dec	1964.17.x	ml 🖪 NoNa	ame01.sgm								
HTML-ar			✓	'	".","	."."h.".o.'o	c,.""."	·····	, _0	1 ~_ """	"','		
About this library <pre> <a <a="" <a<="" td=""><td></td><td></td><td>• Out</td><td>"""a I""""""" </td><td> 10 " "',,,, = ', les ,,, ,, ,, , , ,</td><td></td><td> </td><td>,_"', _"'."'_" . no.l&lt;"" " O&lt; """"" </td><td></td><td> </td><td></td><td>.*-""",.''" 01 fD'~.".</td><td><u>-</u>, 0,</td></a></pre>			• Out	"""a I""""""" 	 10 " "',,,, = ', les ,,, ,, ,, , , ,		 	,_"', _"'."'_" . no.l<"" " O< """"" 		 		.*-""",.''" 01 fD'~.".	<u>-</u> , 0,
🛃 start 🔰	10 6	🕑 👌	è 🛯 🐑 N	ovell G	📄 6 Wind.	🔹 🔞 Corel	рн 🔯	Microsof	Docume	🔁 lav21n2	🚹 NoteTab	EN 🔇 🚅 👽	02:22 PM

# OCR in abbyy fine reader



### Better recognition – editing possible



### Example of OCR Text in abbyy fine reader

🕂 NoteTab Light - NoName02.s	gm
File Edit Search View Modify Docu	iment Favorites Tools Help
D 🛩 🔒 🖳 🕼 🖬 🛍 🍯 .	% 🖻 🖻 🖉 🛍 🖗 🗠 🚧 🔔 🖓 😭 🐨 🙀 🚣 🖍 💉 🛍 🍳 🏶 🏑 👫
B PVDec64.0032.9142.003.008.Dec1964	1.17.xml BNoName01.sgm BNoName02.sgm
HTML-ar	Heata and satory Issues lave tang been part of Eabour's ag "ida, and with goad reason: rtungredj m [1,:",• [-] (11:1[4][1][1]). rd., i] Imany alluroc, sutler ths fongning dined? c=f rdethilaving disease, Apartheid planning led td the "itikg of working class residential areas next to industry These coniiTiurfibes are (ha first to (Tperleoce I:); !?", y" envi:

### References

- Maxine, K. 2000. Handbook for digital projects: a management tool for preservation and access. Massachusetts: Conservation Center.
- Trevor, Jones 2001. An introduction to digital projects for Libraries, Museums and Archives. Accessed 03 July 2009 from <u>http://illinoisheritage.prairienet.org/</u>

