

# Artificial Intelligence: Implications for archival methods and case studies

Archives Society of Australia

2nd August 2024

**Gaia Resources**

+61 8 9227 7309

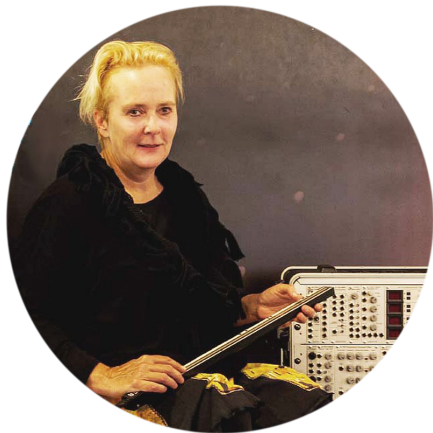
Level 6, FLUX 191 St Georges Terrace, Perth

[enquiries@gaiaresources.com.au](mailto:enquiries@gaiaresources.com.au)





**FEAR  
IS FULL OF  
OPPORTUNITY**



Like people doing  
research on the Internet,  
AI is limited by its inputs,  
but it can be trained to  
do useful work!



AI's fine... but:

It should only be used to  
do grunt work or do  
things faster...  
so that we have more  
time to do the **more  
important stuff.**

Artificial Intelligence: Implications for archival methods and case studies

# What is Artificial Intelligence?

Archivists



The  
Tsunami of  
digital

Archivists

The  
Tsunami of  
digital

AI



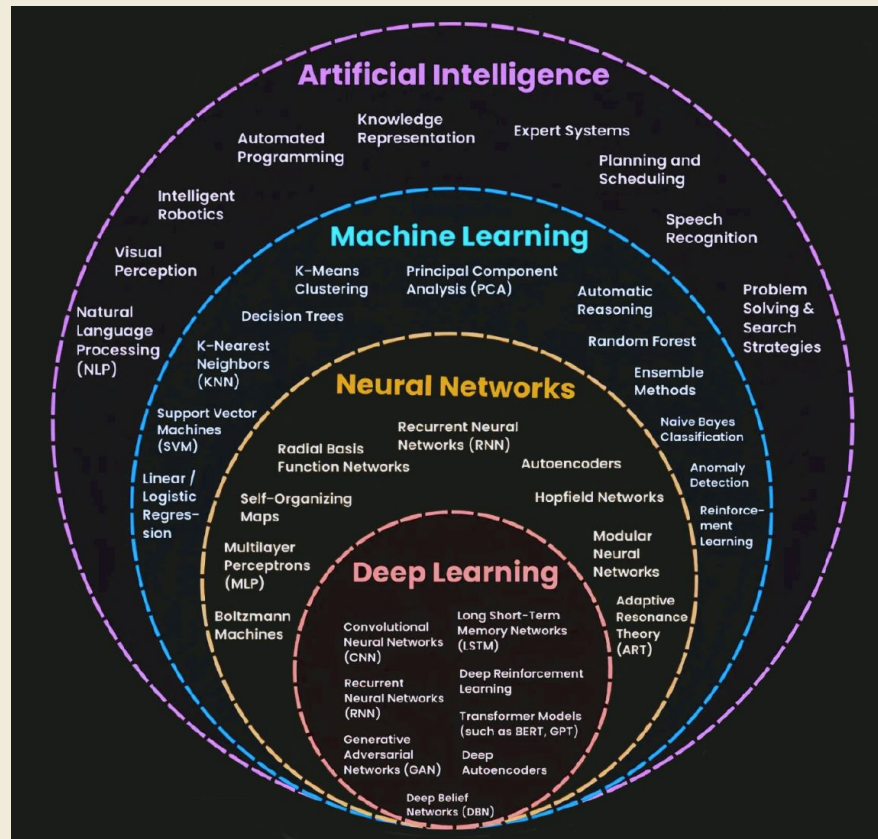


# What is Artificial Intelligence?

*The ability of machines to exhibit intelligent behavior.*

You've probably been using it already:

- Spam filtering
- Music and video recommendations
- Google/Alexa/Siri
- Facial recognition
- Language translation
- Self-driving vehicles
- Medical imaging analysis
- Online games



# Why is it useful?

Massively increases speed of processing of digital information:

- Both information in your collection, and about your collection
- Spot popular topics and prioritise access strategies

Can also identify linkages from small details

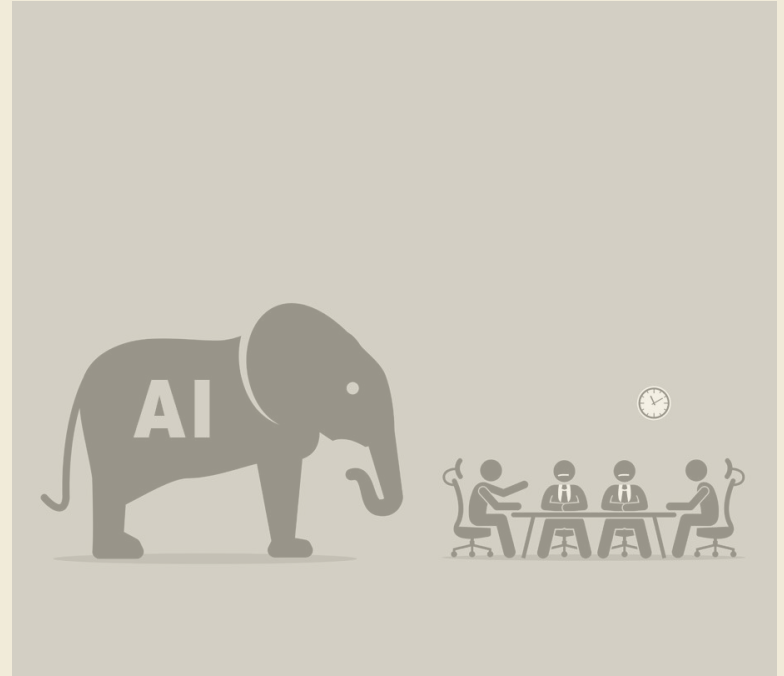


*UN International Criminal Tribunal AV archive*



# What are the barriers?

- Little use in the Archives (and GLAM) sectors as yet
- Lack of **skills** in those areas
- Need to create **trust** and shared ethics - between record creators, records managers, archivists, and researchers
- **Governance** is coming across all sectors



Artificial Intelligence: Implications for archival methods and case studies

How does this impact  
archival practices?

# Why should Archives engage with AI?

For government archives, you will be dealing with AI content and AI systems from agencies:

- ["Automated Decision Making"](#) (Commonwealth Ombudsman)
- ["The new machinery of government"](#) (Ombudsman NSW)
- ["Interim guidance on government use of public generative AI tools"](#) (Digital Transformation Agency)



*The West Australian*

# Why should Archives engage with AI?

For non-government archives, you will be dealing with AI content and AI systems from agencies:

- Many people are not aware they are making use of AI systems already
- Businesses are taking up AI at a faster rate than government

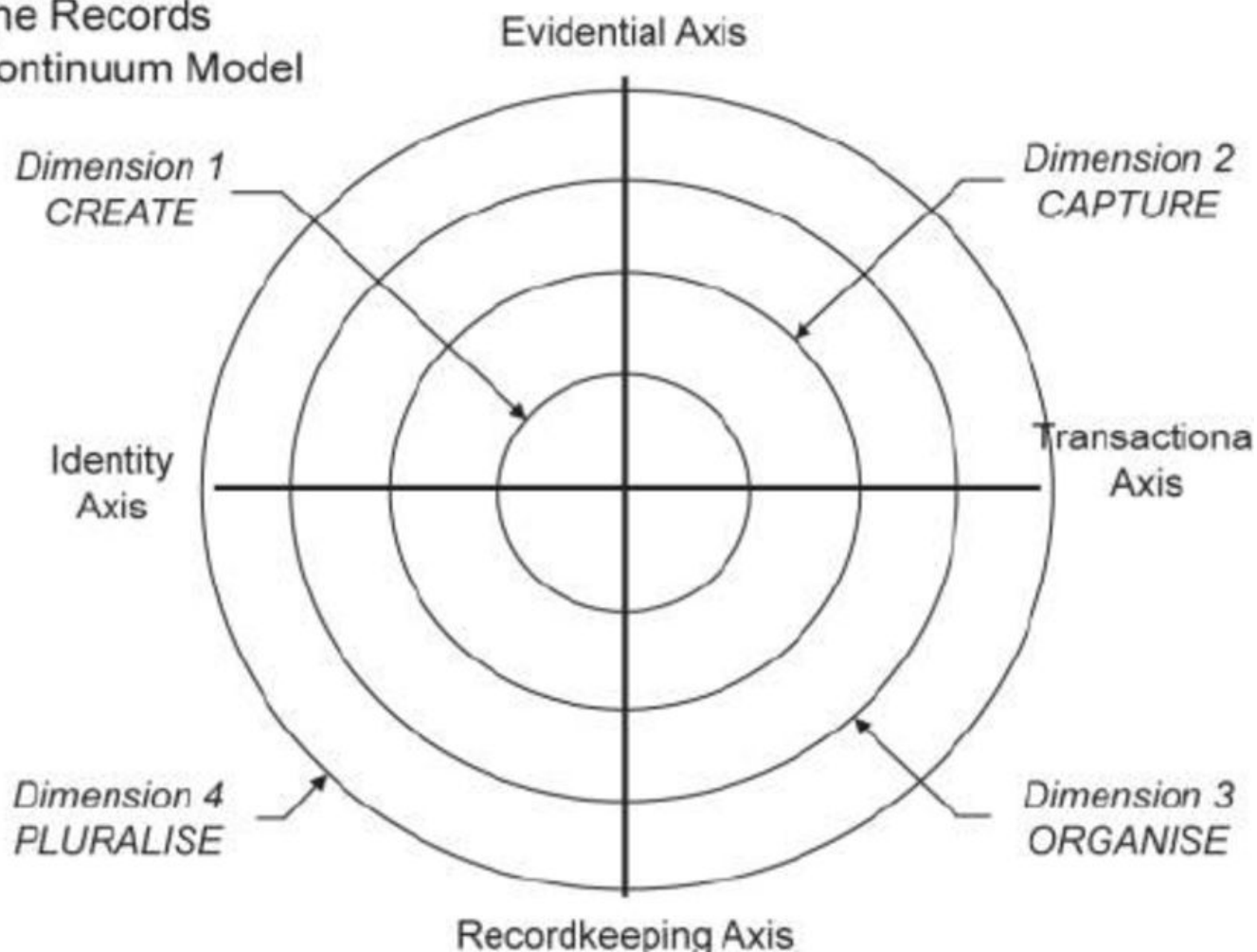
## TOP 10 MACHINES FOR SMALL BUSINESS



## START WITH LOW INVESTMENT

*New Business Ideas*

# The Records Continuum Model



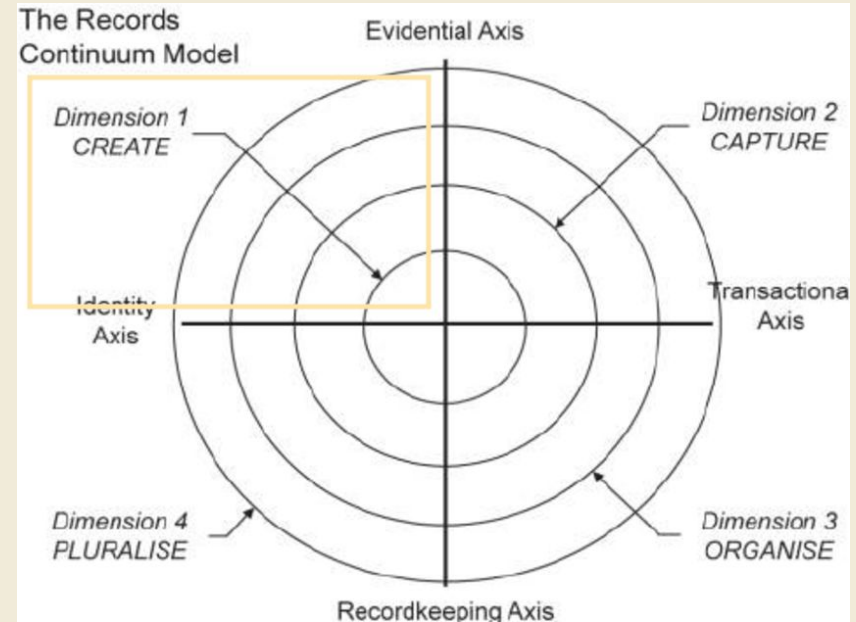


# How does this impact Archival practices?

## *Dimension 1 - CREATE*

*A document is created in an information system.*

- Using AI to create content



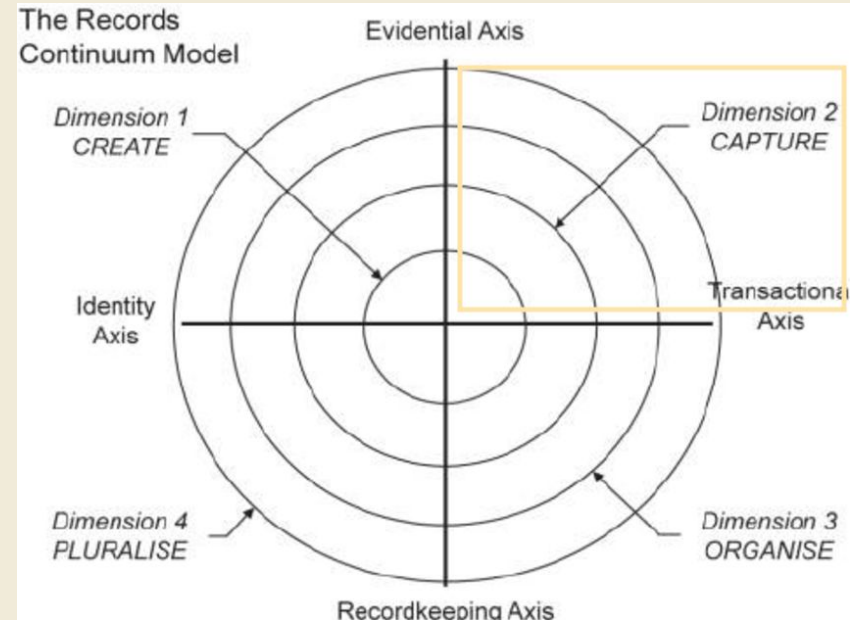
*Records Continuum Model*

# How does this impact Archival practices?

## *Dimension 2 - CAPTURE*

*Records are captured into a Recordkeeping System*

- Ephemeral or private records identified prior to capture
- PII identified prior to capture
- Security content identified prior to capture



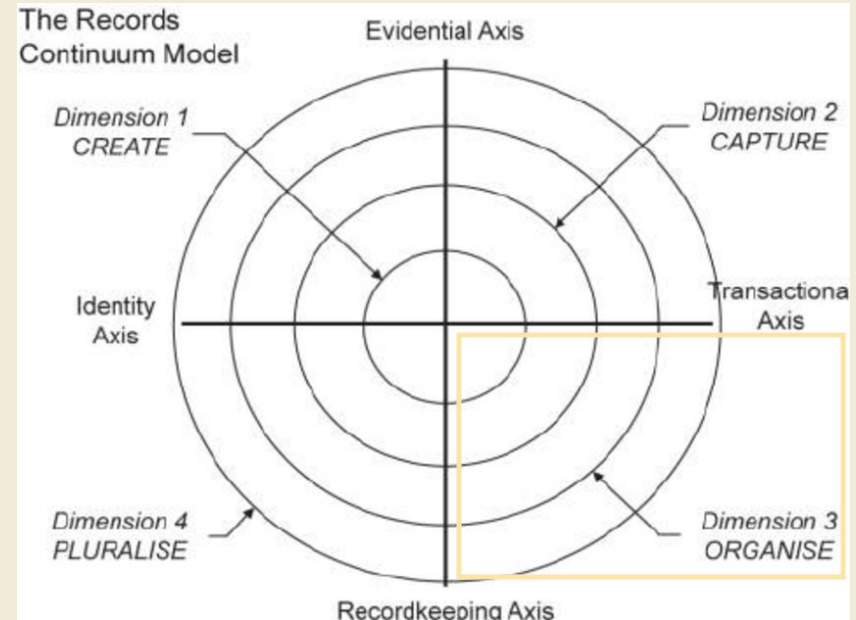
*Records Continuum Model*

# How does this impact Archival practices?

## *Dimension 3 - ORGANISE*

*Records are organised in an Archive*

- Descriptions generated
- Linkages between Agents, Locations, Series and Items identified
- Matching Agents and Locations
- De-duplication



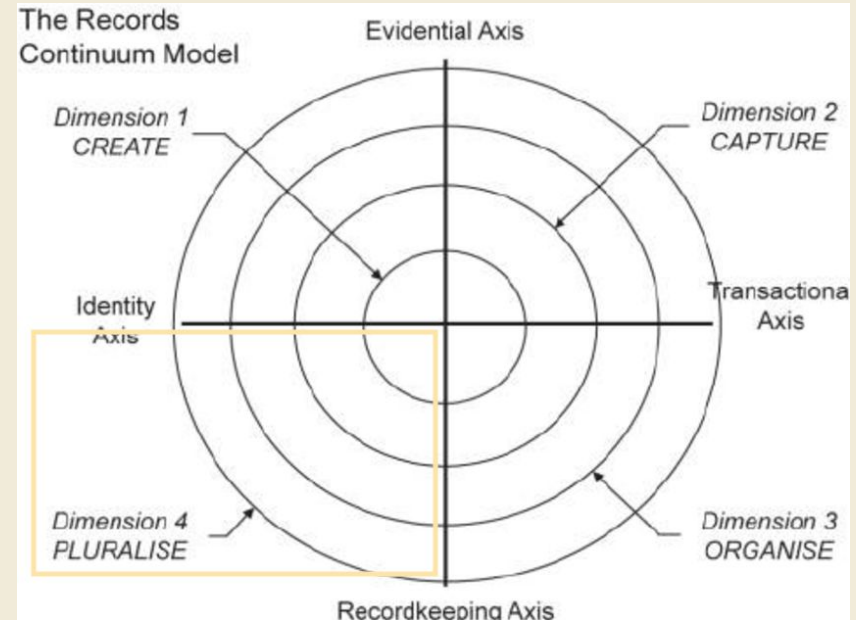
*Records Continuum Model*

# How does this impact Archival practices?

## *Dimension 4 - PLURALISE*

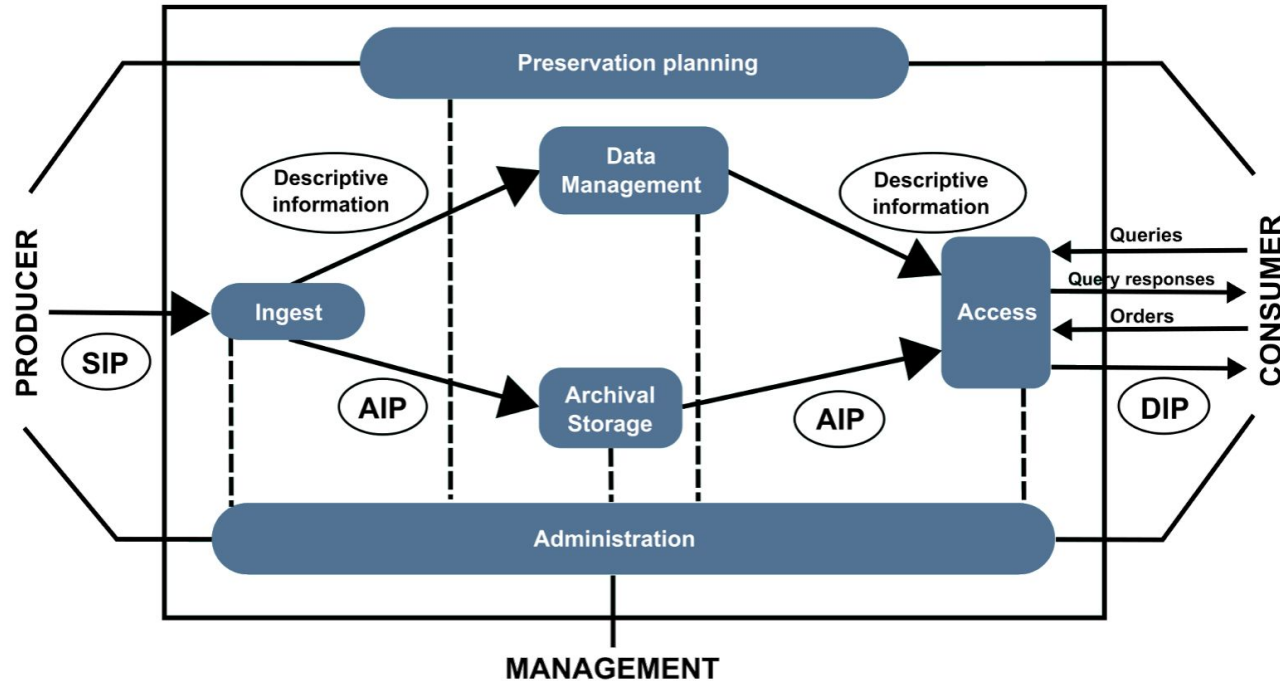
Records are embedded in society by adding metadata and context, and making them accessible

- Linkages with external systems
- Handwriting recognition
- Transcriptions of audio and video created
- Interpreting old phraseology



*Records Continuum Model*

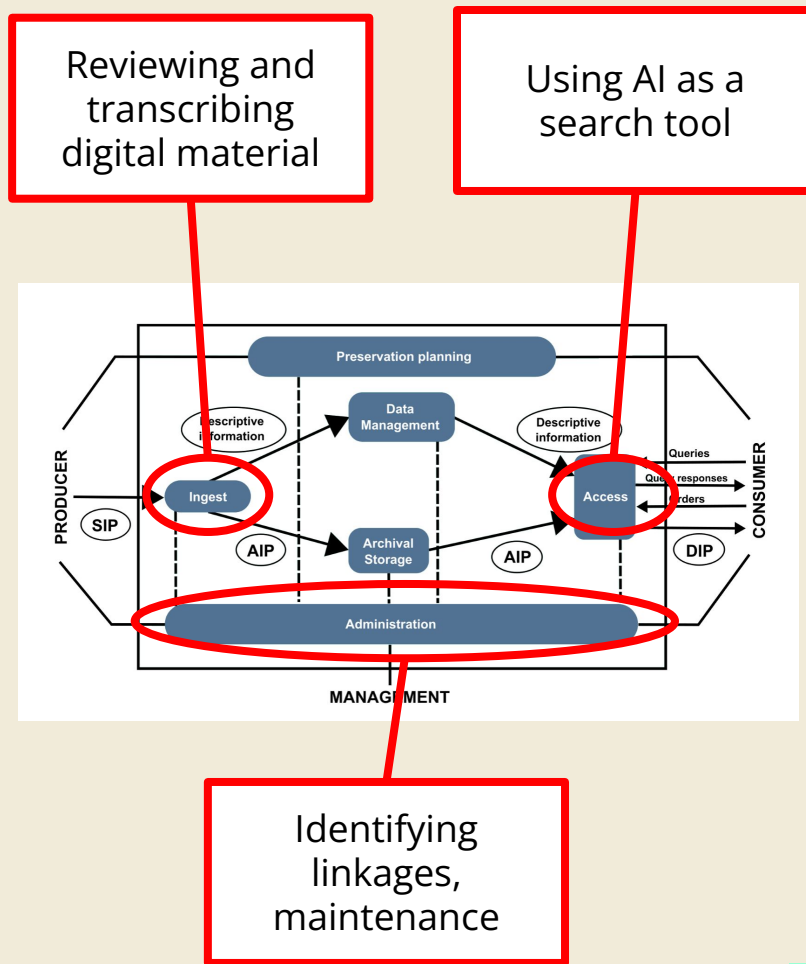
# Open Archival Information Systems Model





# OAIS

- Whilst AI works on digital materials, it's functionality isn't represented in the OAIS model
- Transcription, additional metadata, and linkages are outside OAIS - in archival description.
- Likely places to see AI in the future in OAIS:
  - **Ingest**
  - **Access**
  - **Administration**



# It's all about Trust and Governance



# Building Trust in the Public Record

Artificial Intelligence can help Archives to undertake the “grunt” work of collection management, and free Archivists up to perform governance and be the trusted interface to their collections.

*Processes to check and confirm the accuracy of records / data generated by or through the use of AI technologies must be in place and overseen by an appropriately skilled human being with relevant authority and documented line of responsibility*

[Artificial Intelligence Technologies and Recordkeeping](#)  
(PROV)

# Building Trusted Availability in the Public Record

When information is managed well-

- enables the Australian Government to be transparent, responsive and accountable to the community it serves
- is foundational to the Australian Government's digital transformation agenda to improve government services to the community
- is available as a resource for re-use by government and community to drive innovation, increase economic productivity and enhance social and cultural outcomes



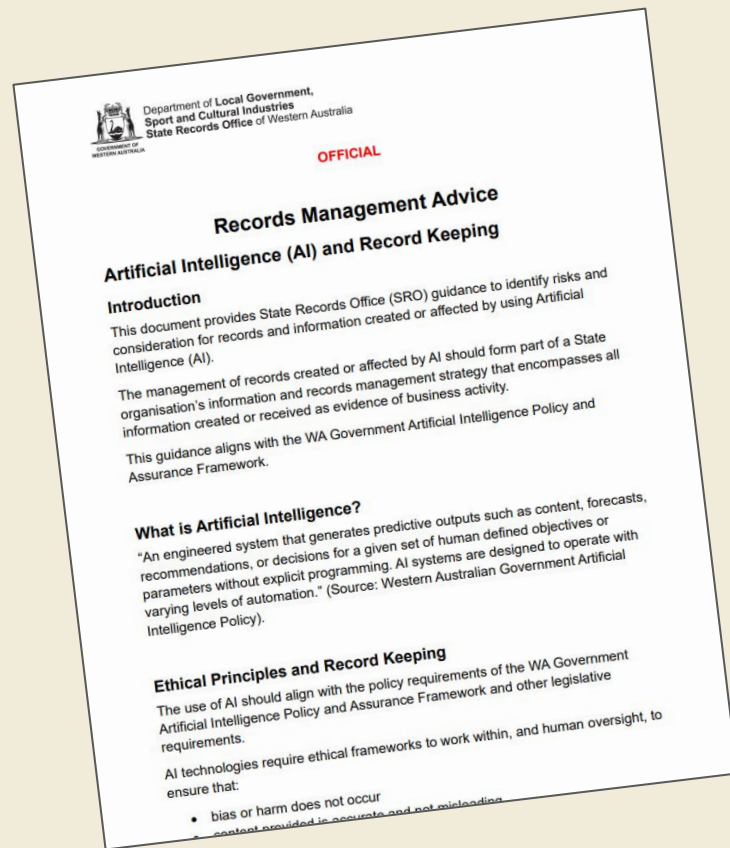
[Digital Archiving - Context is Everything.](#)  
The National Archives UK

*"We have amazing opportunities to rethink how we establish intellectual control using modern technology when we can accession records that are so much richer, with information about how they have been created and used embedded in their very fabric."*

# Governance

“As per section 3(2) of the State Records Act 2000, record keeping **should document the use of AI in business activities of the organisation**. This should include records related to design, testing, implementation, system evaluation and decommissioning. The managing of the digital information should comply with the State Records Commission Standards.”

“While AI can optimise or automate record keeping processes such as document classification and retrieval, these should be processes **overseen by an appropriately skilled human being** with relevant skills, knowledge, authority and documented line of responsibility to check and confirm the accuracy of records. In particular, the **disposal of records process must have human oversight** and be in line with the relevant retention and disposal authority(ies).”



[Records Management Advice](#) (SROWA)



Artificial Intelligence: Implications for archival methods and case studies

# How is AI being used in Archives?

# How is AI being used in Archives?



## It's not widely used

- AI is not a packaged solution, it's a collection of tools
- There is no 'one size fits all'
- Knowing your collection and its users helps you know what kind of tools could be useful



Google Cloud  
Speech API



Tesseract OCR

spaCy



amazon Transcribe



Amazon Comprehend



Google  
Natural Language  
API



OpenAI

# Audio Visual transcription

- Tools: AWS Transcribe vs OpenAI Whisper vs Google Speech-to-Text
- All cloud-based - potential security issues, all different pricing models
- Different languages, accuracy, speed
- Then onto entity recognition and sentiment analysis - AWS Comprehend vs SpaCy vs Google Cloud Natural Language

## How to assess AI transcription?

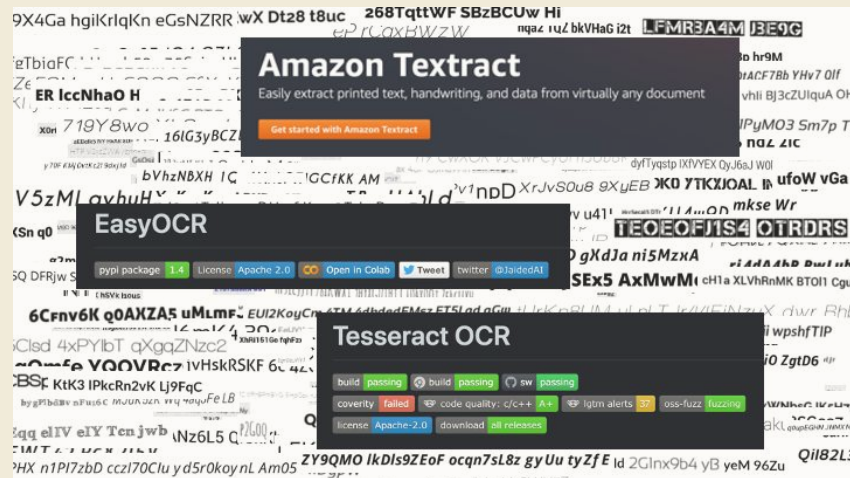
$$\text{WER} = \frac{\begin{array}{c} \text{Nº of words} \\ \text{substitutions} \end{array} + \begin{array}{c} \text{Nº of word} \\ \text{deletions} \end{array} + \begin{array}{c} \text{Nº of word} \\ \text{insertions} \end{array}}{\begin{array}{c} \text{N} \\ \text{Nº of words in} \\ \text{the reference text} \end{array}}$$

**Accuracy** is paramount when it comes to speech recognition. The same goes for **speed**.

The most common metric used to assess transcription quality in this industry is **Word Error Rate (WER)**.

# Other Transcription

- Tools: AWS Textract vs Google Tesseract vs OCR with Google AI
- All cloud-based - potential security issues, all different pricing models
- Different languages, accuracy, speed
- Different support for handwriting, formatted docs.





# CLIO - Image

- Online system for the upload and transcription of any document using AI

## Build recipe:

- Take an AWS Textract model
- Wrap it in an interface for human review:
  - Display the raw image alongside the text
  - Display the confidence level with the text
- Allow data to be exported and used

049 ITM17699 No.07 1862  
'booklet' style and Split Files



Page 3 (and 4)

I need hardly recording that in not say that I deeply lament very distant times those the unhappy fate of these Europeans, nor can I venture to blame those who in the existing state of the Queens -landly back Country at once avenged this terrible outrage by such means as were at hand, yet in reading the account of these occurrences I cannot dismiss from my mind the impression produced by the Report on the Native Police force which you have recently forwarded to me recording

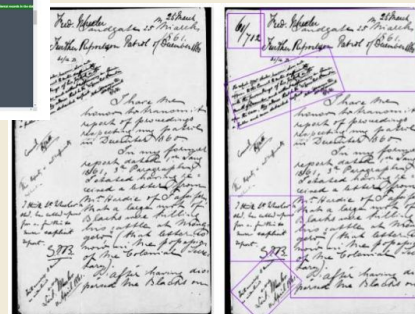
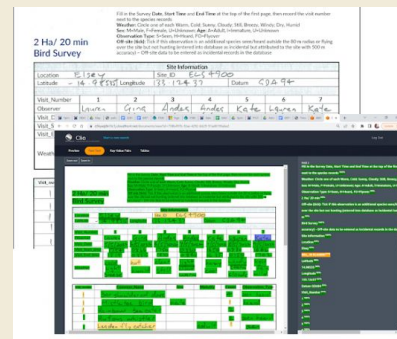
Page 3

I need hardly say that I deeply lament the unhappy fate of these Europeans, nor can I venture to blame those who in the existing state of the Queens -landly back Country at once avenged this terrible outrage by such means as were at hand, yet in reading the account of these occurrences I cannot dismiss from my mind the impression produced by the Report on the Native Police force which you have recently forwarded to me recording

Page 4

recording that in not very distant times these Natives were in some parts destroyed indiscriminately by strychnine, and that even now common destruction of men and women, is by no means unknown. While such things or anything like them occur, it is impossible to hope that any efforts for elevating the character of the natives will be successful, on to doubt that the mutual animosity of whites and

Good



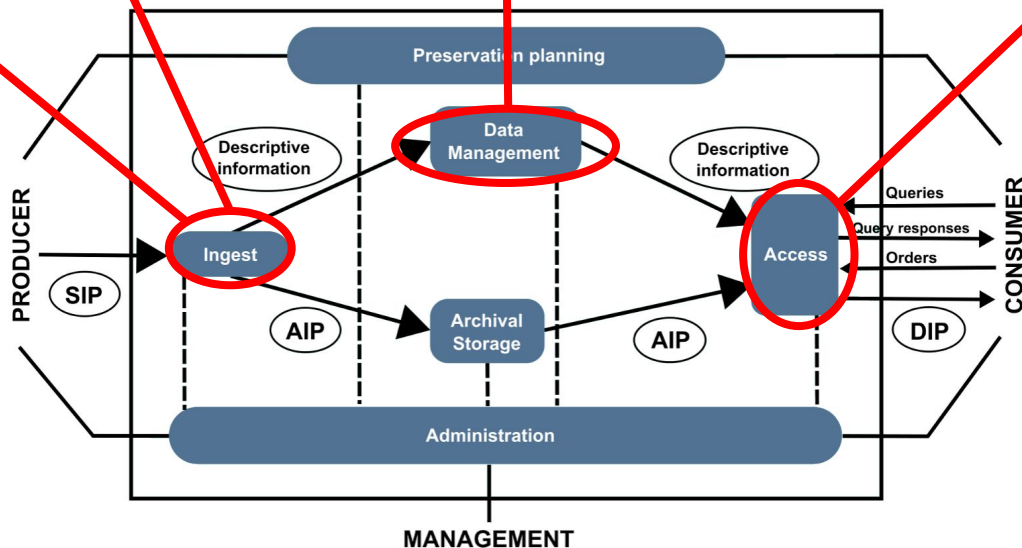
# CLIO

Archivist reviews the transcriptions

Store transcriptions along with the object

Allow users to search the catalogue **and the transcription**

Transcribe images of documents on ingest



# Other Australian Archives

- Archives and Manuscripts (2019) paper by Rolan et al
- **Public Records Office Victoria** - email appraisal POC (28 petabytes of content)
- **NSW State Archives** - classifying unstructured data against retention and disposal strategy (EDRMS extract, 42,653 files)
- **National Archives of Australia** - disposal authorisations for objects



# Other International Archives

- **US National Archives** - extracting data from forms, AI redaction of information, personalised visitor AI
- **UK National Archives** - undertaking a sensitivity review of historical records using AI
- **Britain/France** - using AI to identify sensitive images prior to access
- **Japan** - AI used to transcribe archival film footage - to create infographics for an art installation



# Apply it in the right areas

*Massively increases speed of processing of digital information:*

- *Both information in your collection, and about your collection*
- *Spot popular topics and prioritise access strategies*

*Can also identify linkages from small details*



Title repeat

# Summary

# Summary

## Trust

Ensure that you have a review on anything created by AI to build the trust



## Governance

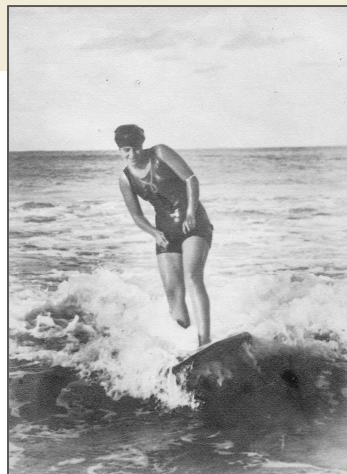
Ensure that you have the right governance around AI

Use it in the right applications



## Skills

Build the skills that you need





# Questions?

**Bring out your ideas, thoughts,  
problems!**

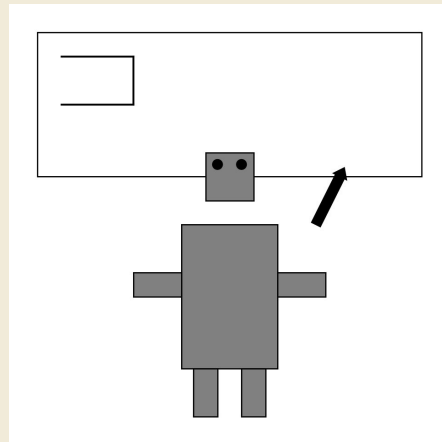
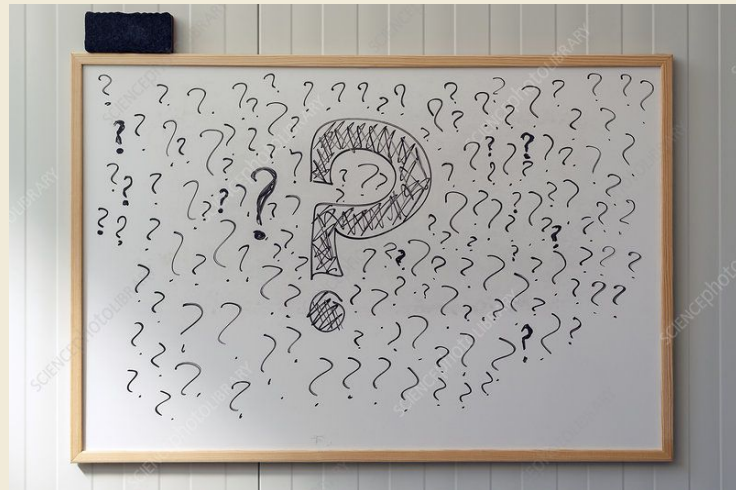
(or contact us later)

Meg Travers

[meg.travers@gaiaresources.com.au](mailto:meg.travers@gaiaresources.com.au)

Piers Higgs

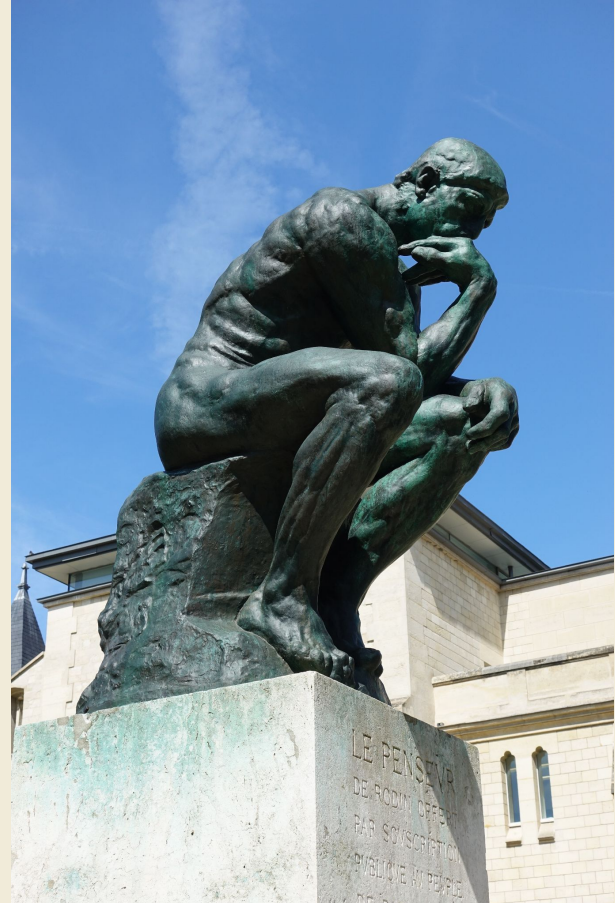
[piers.higgs@gaiaresources.com.au](mailto:piers.higgs@gaiaresources.com.au)



# Future thinking

Upcoming Archives events are looking at a range of AI topics:

- AI at the National Library of NZ webinar, 20 August (online)
- Fantastic Futures Conference - 15-18 October, Canberra
- Archives Society of Australia Conference - 22-25 October, Christchurch NZ



# Thank you.

**Gaia Resources**

+61 8 9227 7309

Level 6, FLUX 191 St Georges Terrace, Perth

[enquiries@gaiaresources.com.au](mailto:enquiries@gaiaresources.com.au)

