AI powered archive indexing
Improving accessibility across heritage collections

The Story at Mount Oswald is a new cultural venue on the outskirts of Durham City. It brings together archive, museum and library collections from across the county, including from institutions such as the Durham County Record Office and Durham Light Infantry Collection.

The historical content in these collections, including artworks, artefacts and texts, was difficult to bring together as they were unstructured and difficult to organise into a database. What's more, metadata (such as author, artist or date of creation), which could help categorise and link pieces, appeared inconsistently across collections.

Durham County Council wanted to make all of this content easily available and searchable online for visitors to its website. To do this, the solution needed to enrich, index and interlink the historical content so users could access all collections from a single point of search, whether that was online or through consoles at The Story.

What they needed   What we did   How it went

What we did

We worked with Durham County Council to create the required single point of search for their multiple collections using cloud-based AI and digital technologies from Microsoft.
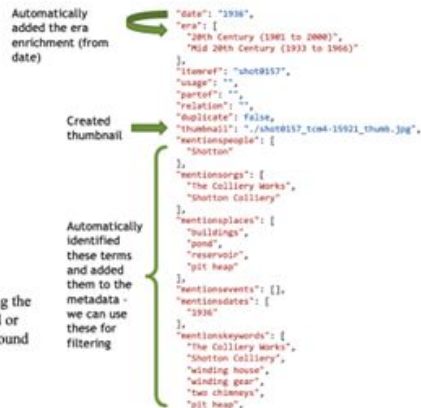
Here's how it works:

- We used AI techniques, like Natural Language Processing (NLP), to assign tags to items and convert audio/video recordings into text. NLP also identified additional metadata to build new links between items across collections.

- We used Cognitive Search (from Microsoft) to index and provide querying capabilities that supported rich user search experiences. These included autocomplete and auto suggest capabilities, as well as being able to identify similar items through metadata and semantic links.

- We also included Azure Data Factory and Azure Data Lake Service to manage the flow of information across collections and storage. With this, we created an automated, end-to-end solution that scales with the collection. Any new content added will flow into a central storage space (data lake) and become easily accessible via search.

The Colliery Works – Shotton

Photograph of the buildings of Shotton Colliery showing the winding house, the winding gear, two chimneys, a pond or reservoir in the foreground and a pit heap in the background

Date: 1936

Automatically added the era enrichment (from date)

Created thumbnail

Automatically identified these terms and added them to the metadata - we can use these for filtering

"date": "1936",
"era": [
    "20th Century (1901 to 2000)",
    "Mid 20th Century (1933 to 1966)"
],
"itemref": "shot0157",
"usage": "",
"partof": "",
"relation": "",
"duplicate": false,
"thumbnail": "./shot0157_tcm4-15521_thumb.jpg",
"mentionspeople": [
    "Shotton"
],
"mentionsorgs": [
    "The Colliery Works",
    "Shotton Colliery"
],
"mentionsplaces": [
    "buildings",
    "pond",
    "reservoir",
    "pit heap"
],
"mentionsevents": [],
"mentionsdates": [
    "1936"
],
"mentionskeywords": [
    "The Colliery Works",
    "Shotton Colliery",
    "winding house",
    "winding gear",
    "two chimneys",
    "pit heap",

How it went

By building an AI Search Index in the cloud we connecting the collections and making them more accessible, Durham County Council can now open its archives like never before. The council has already indexed multiple collections and in Autumn 2023, the database will become accessible online via APIs.

The database will democratise heritage as visitors will no longer need to rely on expert advice or travel across institutions to access the breadth of content in Durham's collections.

Digitisation also opens the door to crowdsourcing. Durham plans to get local volunteers, amateur archivists and historians involved in tagging, donating digital assets or commenting and sharing online exhibitions.

Most importantly, our work provides a framework, which Durham can use to develop, extend, optimise and improve their collections going forward.

Durham
County Council

"We have always encouraged the use of our collections as an inspiration for the creative arts – commissions, writing and biographies. By putting all this information at people's fingertips, we hope to support the cultural heritage economy better than ever."

Gill Parkes
Principal Archivist

"That flexibility, in viewing and using the information, provides some really exciting future opportunities. We're already talking about its use in VR exhibitions, sharing it out to the community through our Area Action Partnerships and even establishing a cross–council database for cultural heritage in the North East."

Brian Dougherty
Senior IT Project Leader at Durham County Council