Accessibility for data analysis: using digital tools to unlock archives potential

Kathryn Dan, ANU and Ingrid Mason, AARNet
ANU Archives Program

Noel Butlin Archives Centre
The Noel Butlin Archives Centre (NBAC) collects business and labour records from Australian companies, trade unions, ...

University Archives
The University Archives hold a range of material dating from the 1920s relating to the history of The Australian...

Pacific Research Archives
The Pacific Research Archives (PRA) at The Australian National University is the premier Australian centre for research...

National AIDS Archive Collection
The National AIDS Archive Collection documents the history of AIDS education and prevention and related concerns in all...
Outline

• Previous digitisation projects
• The archives: Sydney Stock Exchange Registers
• Project objectives and partners
• Work so far … and where to next
Search for Tooth and Company hotels

This interactive map lets you locate hotels owned or supplied by Tooth and Company between the 1920s and the 1970s. From the map you can see photographs and other information about hotels on 'yellow cards'. These are index cards the company kept about its own hotels and those of its competitors.

Each card covers roughly a decade. While the cards vary slightly, they all provide the same basic information. The front of the card normally has the hotel name, location, owner, licensee, a graph featuring the number of barrels or dozens of bottles of Tooth's beer supplied to the hotel, and information on the total trade with Tooth. The back of the card usually provides a photograph of the hotel and information about the building, land, amenities, nearby hotels and local industries. If this information was not provided the card was not digitised.
## Sydney Stock Exchange

**Wednesday, July 3rd, 1918**

### Official List of Prices Current

<table>
<thead>
<tr>
<th>Stock</th>
<th>Copper</th>
<th>Silver</th>
<th>Gold</th>
<th>Shillings</th>
<th>Pence</th>
<th>全日</th>
<th>1918年</th>
<th>1919年</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Gold

- Victorian
- Birthday Tunnel
- Cock's Pioneer
- Garden Gully Consols
- Golden Age
- Gt. Southern Consols
- Hansel Mandy
- Hobart's Rest
- Long Tunnel
- Monument Hill
- New Long Lagoon
- Olive Branch
- Princess Dagnall
- Sea Amalgamated
- South New Moon
- Victorian Manganese
- Wilson's Hill
- West Australia Associated
- North

### Tin

- Brineis
- Carpathia
- Carpathia South
Project objectives

• Demonstrate data extraction to enhance research potential of digitised archival resources
• Test digitisation workflows and data transfer
• Test applicability of automated handwriting recognition
• Investigate automating description
• Experiment with providing archival collections as ‘data’
• Provide a case study for future work
### Sydney Stock Exchange

**Wednesday, July 4th, 1871**

#### Official List of Prices Current

<table>
<thead>
<tr>
<th>Shares</th>
<th>Palings</th>
<th>Sales</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Tin**
- Briseis
- Carpathia
- Carpathia South
- contributing

**Gold**
- Victorian
  - Victoria
  - Birthday Tunnel
  - Cook’s Pioneer
  - Garden Gully Consols
  - Golden Age
  - G. Southern Consols
  - Hamed Mundy
  - Hustler’s Road
  - Long Tunnel
  - Monument Hill
  - New Langi Logan
  - Olive Branch
  - Princess Dagmar
  - Sea Amalgamated
  - South New Moon
  - Victorian Manganese
- contributing

**West Australia**
- Associated
- North

---

**Prices**
- Briseis: 6½
- Carpathia: 5½
- Carpathia South: 5½
- Victoria: 15½
- Birthday Tunnel: 2½
- Cook’s Pioneer: 2½
- Garden Gully Consols: 2½
- Golden Age: 6½
- G. Southern Consols: 6½
- Hamed Mundy: 6½
- Hustler’s Road: contributing
- Long Tunnel: contributing
- Monument Hill: contributing
- New Langi Logan: contributing
- Olive Branch: contributing
- Princess Dagmar: contributing
- Sea Amalgamated: contributing
- South New Moon: contributing
- Victorian Manganese: contributing
- Wilson’s Hill: contributing
- North: contributing
Deposit N193 - Sydney Stock Exchange stock and share lists

Identity area
- Reference code: N193
- Title: Sydney Stock Exchange stock and share lists
- Date(s): 1901 - 1950
- Level of description: Deposit
- Extent and medium: 199 items

Context area
- Name of creator: Sydney Stock Exchange (1871 - 1987)
  Administrative history:
  The Sydney Stock Exchange was formed to allow brokers and traders to trade stocks and bonds for companies listed in New South Wales. It formed an association with the stock exchanges in Adelaide, Melbourne, Brisbane, Perth and Hobart called the Australian...

Content and structure area
- Scope and content: These are large format bound volumes of the official lists that were posted up for the public to see - 3 times a day, forenoon, noon and afternoon - at the close of the trading session in the call room at the Sydney Stock Exchange. The closing prices of...

Conditions of access and use area
- Conditions governing access: Researchers must sign an access agreement.

Access points
- Name access points: Sydney Stock Exchange (Creator)

Description control area
- Dates of creation revision deletion: Entered from deposit description on 25 February 2013.
Sydney Stock Exchange: [199]

The Sydney Stock Exchange was formed to allow brokers and traders to trade stocks and bonds for companies listed in New South Wales. It formed an association with the stock exchanges in Adelaide, Melbourne, Brisbane, Perth and Hobart called the Australian Associated Stock Exchanges but remained an independent body. These six stock exchanges amalgamated on 1 April 1987 to form the Australian Stock Exchange Limited (ASX).

Collection's Items (Sorted by Submit Date in Descending order): 1 to 20 of 199

Sydney Stock Exchange Stock Official List of Prices Current (199) - March 1950 - July 1950

Author(s): Sydney Stock Exchange (1871 - 1987)
Type: Text
Date Published: 1950
Date Created: March 1950 - July 1950
<table>
<thead>
<tr>
<th>Stock Symbol</th>
<th>Price</th>
<th>Quantity</th>
<th>Buyer</th>
<th>Seller</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABC</td>
<td>5.00</td>
<td>10</td>
<td>John</td>
<td>Mary</td>
<td>17</td>
</tr>
<tr>
<td>DEF</td>
<td>10.50</td>
<td>5</td>
<td>Paul</td>
<td>Steve</td>
<td>18</td>
</tr>
<tr>
<td>GHI</td>
<td>20.00</td>
<td>20</td>
<td>Mark</td>
<td>Lily</td>
<td>19</td>
</tr>
</tbody>
</table>

SYDNEY STOCK EXCHANGE
OFFICIAL LIST OF PRICES CURRENT
Yesterday Friday 8th July 1992
Steps

1. Digitise and extract
2. Convert ledger to data – columns, rows
3. Recognise text – transcription and automation
4. Automate recognition
5. Develop and test further extraction

Using CloudStor

Commonly asked questions and troubleshooting

- Preferred Client Tools
- Accessing public links over WebDAV
- Accessing your CloudStor Directory using Finder in macOS
- Can I view 3D Models on CloudStor?
- Audio in CloudStor
- Where can I download the CloudStor sync client?
- Add your CloudStor directory as a network drive in Windows 10
- How can I upload datasets from a scientific instrument to CloudStor?
- How do I restore deleted files and folders?
- Is there any restriction on the FileSender file types I can send? My organisation/institute blocks .zip and .exe
- Can I just send files without logging in to CloudStor?
- I have reached my quota - how do I get more storage?
- Why do I sometimes get conflict files/messages when syncing?
- Why does my video take so long to play on CloudStor?
Find straight lines in the edges

  # Note that the minLineLength and maxLineGap values can have a dramatic effect on the number of lines detected.
  # Note too that what looks to the human eye as a single straight line,
  # can actually be a series of short line with tiny gaps between them,
  # so while increasing the minLineLength reduces noise from text, it can also filter out columns.
• directory
• name
• path
• referenceCode
• startDate
• endDate
• year
• width – width of the image
• height – height of the image
• columns – number of columns
• column_positions – comma-separated string of x values of columns in pixels
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>41076</td>
</tr>
<tr>
<td>4</td>
<td>26917</td>
</tr>
<tr>
<td>2</td>
<td>4825</td>
</tr>
<tr>
<td>1</td>
<td>19</td>
</tr>
<tr>
<td>0</td>
<td>6</td>
</tr>
</tbody>
</table>

Name: columns, dtype: int64
### PUBLIC INCOME AND EXPENDITURE

- **Revenue and Other Receipts**
  - [Table of revenue and other receipts]

### TRAFFIC RECEIPTS

- **Domestic, Overseas and Colonial**
  - [Table of traffic receipts]

### TRAINWAYS AND GROUNDWORKS

- [Table of trainways and groundwork receipts]

### MISCELLANEOUS

- [Table of miscellaneous receipts]

### FOREIGN RAILWAYS

- [Table of foreign railways receipts]

### ARGENTINE RECEIPTS

- [Table of Argentine receipts]

### AFRICAN EXCHANGE RATES

- [Table of African exchange rates]

---

*Figures for the past week.

---

*Figures for the past week.*
[{'task': 'T4',
     'task_label': 'Can you see a handwritten date?',
     'value': 'Yes'},
{'task': 'T0',
     'task_label': 'Can you identify the session? Look for one of the following handwritten headings.',
     'value': 'No session heading'},
{'task': 'T5',
     'task_label': 'What is the day of the week?',
     'value': 'Saturday'},
{'task': 'T1',
     'value': '29',
     'task_label': 'What is the day of the month? This should be a number between 1 and 31.'},
{'task': 'T2', 'task_label': 'What is the month?',
     'value': 'June'},
{'task': 'T3', 'value': '1929', 'task_label': 'What is the year?'}]
<table>
<thead>
<tr>
<th>date</th>
<th>holiday</th>
<th>year</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/01/1900</td>
<td>New Years day</td>
<td>1900</td>
</tr>
<tr>
<td>26/01/1900</td>
<td>Anniversary Day</td>
<td>1900</td>
</tr>
<tr>
<td>13/04/1900</td>
<td>Good Friday</td>
<td>1900</td>
</tr>
<tr>
<td>14/04/1900</td>
<td>Easter Saturday</td>
<td>1900</td>
</tr>
<tr>
<td>16/04/1900</td>
<td>Easter Monday</td>
<td>1900</td>
</tr>
<tr>
<td>25/05/1900</td>
<td>Queen's Birthday</td>
<td>1900</td>
</tr>
<tr>
<td>6/08/1900</td>
<td>Bank Holiday</td>
<td>1900</td>
</tr>
<tr>
<td>1/10/1900</td>
<td>Labour Day</td>
<td>1900</td>
</tr>
<tr>
<td>12/11/1900</td>
<td>Prince of Wales's Birthday</td>
<td>1900</td>
</tr>
<tr>
<td>25/12/1900</td>
<td>Christmas Day</td>
<td>1900</td>
</tr>
<tr>
<td>26/12/1900</td>
<td>Boxing Day</td>
<td>1900</td>
</tr>
<tr>
<td>1/01/1901</td>
<td>New Years day</td>
<td>1901</td>
</tr>
<tr>
<td>2/02/1901</td>
<td>Day of Mourning for Queen</td>
<td>1901</td>
</tr>
<tr>
<td>5/04/1901</td>
<td>Good Friday</td>
<td>1901</td>
</tr>
<tr>
<td>6/04/1901</td>
<td>Easter Saturday</td>
<td>1901</td>
</tr>
<tr>
<td>8/04/1901</td>
<td>Easter Monday</td>
<td>1901</td>
</tr>
<tr>
<td>3/06/1901</td>
<td>Prince of Wales's Birthday</td>
<td>1901</td>
</tr>
<tr>
<td>5/08/1901</td>
<td>Bank Holiday</td>
<td>1901</td>
</tr>
<tr>
<td>7/10/1901</td>
<td>Labour Day</td>
<td>1901</td>
</tr>
<tr>
<td>11/11/1901</td>
<td>King's Birthday</td>
<td>1901</td>
</tr>
<tr>
<td>25/12/1901</td>
<td>Christmas Day</td>
<td>1901</td>
</tr>
<tr>
<td>26/12/1901</td>
<td>Boxing Day</td>
<td>1901</td>
</tr>
<tr>
<td></td>
<td>actual</td>
<td>predicted</td>
</tr>
<tr>
<td>---</td>
<td>--------</td>
<td>-----------</td>
</tr>
<tr>
<td>0</td>
<td>303</td>
<td>322</td>
</tr>
<tr>
<td>1</td>
<td>312</td>
<td>334</td>
</tr>
<tr>
<td>2</td>
<td>345</td>
<td>351</td>
</tr>
<tr>
<td>3</td>
<td>312</td>
<td>319</td>
</tr>
<tr>
<td>4</td>
<td>305</td>
<td>312</td>
</tr>
<tr>
<td>5</td>
<td>334</td>
<td>341</td>
</tr>
<tr>
<td>6</td>
<td>349</td>
<td>351</td>
</tr>
<tr>
<td>7</td>
<td>318</td>
<td>319</td>
</tr>
<tr>
<td>8</td>
<td>327</td>
<td>329</td>
</tr>
<tr>
<td>9</td>
<td>327</td>
<td>334</td>
</tr>
<tr>
<td>10</td>
<td>350</td>
<td>351</td>
</tr>
<tr>
<td>11</td>
<td>310</td>
<td>319</td>
</tr>
<tr>
<td>12</td>
<td>330</td>
<td>334</td>
</tr>
<tr>
<td>13</td>
<td>329</td>
<td>329</td>
</tr>
<tr>
<td>14</td>
<td>349</td>
<td>351</td>
</tr>
<tr>
<td>15</td>
<td>313</td>
<td>314</td>
</tr>
</tbody>
</table>
Example 1901

'1901-01-07': 3,
'1901-01-18': 4,
'1901-01-23': 0, # Death of the Queen business abandoned https://trove.nla.gov.au/newspaper/article/14371864/1343690
'1901-02-25': 4,
'1901-03-18': 0,
'1901-03-29': 0, # missing
'1901-04-04': 3, # No afternoon, day before Easter
'1901-04-09': 0, # Extra Easter Tuesday
'1901-04-10': 0, # Extra Easter Wednesday
'1901-05-27': 0, # Holiday Duke of Cornwall visiting
'1901-05-28': 0, # Holiday Duke of Cornwall visiting
'1901-07-03': 0, # Holiday for polling day
'1901-09-16': 4, # No morning
Contacts and resources

• ANU Archives http://archives.anu.edu.au/
• Digitised registers on Open Research https://openresearch-repository.anu.edu.au/handle/1885/146659