

# Putting It All Together

# End-to-end Example

Input: PDF Document



Condensed Consolidated Financial Statements and Notes  
Quarter 2 2015

## CONSOLIDATED STATEMENT OF COMPREHENSIVE INCOME (LOSS)

| Unaudited<br>(Canadian dollars in millions)             | Three months ended<br>June 30 |               | Six months ended<br>June 30 |                 |
|---|-------------------------------|---------------|-----------------------------|-----------------|
|   | 2015                          | 2014          | 2015                        | 2014            |
| <b>Comprehensive income (loss)</b>                      |                               |               |                             |                 |
| Net income (loss) for the period                        | \$ 296                        | \$ 223        | \$ (13)                     | \$ (118)        |
| Other comprehensive income (loss), net of taxes of nil: |                               |               |                             |                 |
| Items that will not be reclassified to net income       |                               |               |                             |                 |
| Remeasurements on employee benefit liabilities          | 542                           | (90)          | 790                         | (293)           |
| Items that will be reclassified to net income           |                               |               |                             |                 |
| Loss on derivatives designated as cash flow hedges      | (1)                           | -             | (1)                         | -               |
| <b>Total comprehensive income (loss)</b>                | <b>\$ 837</b>                 | <b>\$ 133</b> | <b>\$ 776</b>               | <b>\$ (411)</b> |
| <b>Comprehensive income (loss) attributable to:</b>     |                               |               |                             |                 |
| Shareholders of Air Canada                              | \$ 836                        | \$ 132        | \$ 774                      | \$ (413)        |
| Non-controlling interests                               | 1                             | 1             | 2                           | 2               |
|   | <b>\$ 837</b>                 | <b>\$ 133</b> | <b>\$ 776</b>               | <b>\$ (411)</b> |

## CONSOLIDATED STATEMENT OF CHANGES IN EQUITY

| Unaudited<br>(Canadian dollars in millions)    | Share capital | Contributed<br>surplus | Deficit      | Total<br>shareholders'<br>equity | Non-controlling<br>interests | Total equity |
|--|---------------|------------------------|--------------|----------------------------------|------------------------------|--------------|
| January 1, 2014                                | \$ 827        | \$ 80                  | \$ (2,367)   | \$ (1,460)                       | \$ 63                        | \$ (1,397)   |
| Net income (loss)                              | -             | -                      | (120)        | (120)                            | 2                            | (118)        |
| Remeasurements on employee benefit liabilities | -             | -                      | (293)        | (293)                            | -                            | (293)        |
| <b>Total comprehensive income (loss)</b>       | -             | -                      | <b>(413)</b> | <b>(413)</b>                     | <b>2</b>                     | <b>(411)</b> |
| Share-based compensation                       | -             | (4)                    | (13)         | (17)                             | -                            | (17)         |
| Shares issued                                  | 1             | -                      | -            | 1                                | -                            | 1            |

# End-to-end Example

## Table Extraction

### A. Identify table borders



Condensed Consolidated Financial Statements and Notes  
Quarter 2 2015

#### CONSOLIDATED STATEMENT OF COMPREHENSIVE INCOME (LOSS)

| Unaudited<br>(Canadian dollars in millions)             | Three months ended<br>June 30 |               | Six months ended<br>June 30 |                 |
|---|-------------------------------|---------------|-----------------------------|-----------------|
|   | 2015                          | 2014          | 2015                        | 2014            |
| <b>Comprehensive income (loss)</b>                      |                               |               |                             |                 |
| Net income (loss) for the period                        | \$ 296                        | \$ 223        | \$ (13)                     | \$ (118)        |
| Other comprehensive income (loss), net of taxes of nil: |                               |               |                             |                 |
| Items that will not be reclassified to net income       |                               |               |                             |                 |
| Remeasurements on employee benefit liabilities          | 542                           | (90)          | 790                         | (293)           |
| Items that will be reclassified to net income           |                               |               |                             |                 |
| Loss on derivatives designated as cash flow hedges      | (1)                           | -             | (1)                         | -               |
| <b>Total comprehensive income (loss)</b>                | <b>\$ 837</b>                 | <b>\$ 133</b> | <b>\$ 776</b>               | <b>\$ (411)</b> |
| <b>Comprehensive income (loss) attributable to:</b>     |                               |               |                             |                 |
| Shareholders of Air Canada                              | \$ 836                        | \$ 132        | \$ 774                      | \$ (413)        |
| Non-controlling interests                               | 1                             | 1             | 2                           | 2               |
|   | <b>\$ 837</b>                 | <b>\$ 133</b> | <b>\$ 776</b>               | <b>\$ (411)</b> |

#### CONSOLIDATED STATEMENT OF CHANGES IN EQUITY

| Unaudited<br>(Canadian dollars in millions)    | Share capital | Contributed<br>surplus | Deficit      | Total<br>shareholders'<br>equity | Non-controlling<br>interests | Total equity |
|--|---------------|------------------------|--------------|----------------------------------|------------------------------|--------------|
| January 1, 2014                                | \$ 827        | \$ 80                  | \$ (2,367)   | \$ (1,460)                       | \$ 63                        | \$ (1,397)   |
| Net income (loss)                              | -             | -                      | (120)        | (120)                            | 2                            | (118)        |
| Remeasurements on employee benefit liabilities | -             | -                      | (293)        | (293)                            | -                            | (293)        |
| <b>Total comprehensive income (loss)</b>       | -             | -                      | <b>(413)</b> | <b>(413)</b>                     | <b>2</b>                     | <b>(411)</b> |
| Share-based compensation                       | -             | (4)                    | (13)         | (17)                             | -                            | (17)         |
| Shares issued                                  | 1             | -                      | -            | 1                                | -                            | 1            |

# End-to-end Example

## Table Extraction

### B. Identify table structure

| Unaudited   |        | Three months ended June 30 |               | Six months ended June 30 |                 |
|---|--------|----------------------------|---------------|--------------------------|-----------------|
| (Canadian dollars in millions)                          |        | 2015                       | 2014          | 2015                     | 2014            |
| <b>Comprehensive income (loss)</b>                      |        |                            |               |                          |                 |
| Net income (loss) for the period                        |        | \$ 296                     | \$ 223        | \$ (13)                  | \$ (118)        |
| Other comprehensive income (loss), net of taxes of nil: |        |                            |               |                          |                 |
| Items that will not be reclassified to net income       |        |                            |               |                          |                 |
| Remeasurements on employee benefit liabilities          | Note 5 | 542                        | (90)          | 790                      | (293)           |
| Items that will be reclassified to net income           |        |                            |               |                          |                 |
| Loss on derivatives designated as cash flow hedges      | Note 9 | (1)                        | -             | (1)                      | -               |
| <b>Total comprehensive income (loss)</b>                |        | <b>\$ 837</b>              | <b>\$ 133</b> | <b>\$ 776</b>            | <b>\$ (411)</b> |
| <b>Comprehensive income (loss) attributable to:</b>     |        |                            |               |                          |                 |
| Shareholders of Air Canada                              |        | \$ 836                     | \$ 132        | \$ 774                   | \$ (413)        |
| Non-controlling interests                               |        | 1                          | 1             | 2                        | 2               |
|   |        | <b>\$ 837</b>              | <b>\$ 133</b> | <b>\$ 776</b>            | <b>\$ (411)</b> |

# End-to-end Example

## Table Understanding

### A. Identify table regions (column/row headers)

| Unaudited   |        | Three months ended June 30 |               | Six months ended June 30 |                 |
|---|--------|----------------------------|---------------|--------------------------|-----------------|
| (Canadian dollars in millions)                          |        | 2015                       | 2014          | 2015                     | 2014            |
| <b>Comprehensive income (loss)</b>                      |        |                            |               |                          |                 |
| Net income (loss) for the period                        |        | \$ 296                     | \$ 223        | \$ (13)                  | \$ (118)        |
| Other comprehensive income (loss), net of taxes of nil: |        |                            |               |                          |                 |
| Items that will not be reclassified to net income       |        |                            |               |                          |                 |
| Remeasurements on employee benefit liabilities          | Note 5 | 542                        | (90)          | 790                      | (293)           |
| Items that will be reclassified to net income           |        |                            |               |                          |                 |
| Loss on derivatives designated as cash flow hedges      | Note 9 | (1)                        | -             | (1)                      | -               |
| <b>Total comprehensive income (loss)</b>                |        | <b>\$ 837</b>              | <b>\$ 133</b> | <b>\$ 776</b>            | <b>\$ (411)</b> |
| <b>Comprehensive income (loss) attributable to:</b>     |        |                            |               |                          |                 |
| Shareholders of Air Canada                              |        | \$ 836                     | \$ 132        | \$ 774                   | \$ (413)        |
| Non-controlling interests                               |        | 1                          | 1             | 2                        | 2               |
|   |        | <b>\$ 837</b>              | <b>\$ 133</b> | <b>\$ 776</b>            | <b>\$ (411)</b> |

What does this cell represent?

# End-to-end Example

## Table Understanding

### B. Identify additional context within table

| Unaudited   |        | Three months ended June 30 |               | Six months ended June 30 |                 |
|---|--------|----------------------------|---------------|--------------------------|-----------------|
| (Canadian dollars in millions)                          |        | 2015                       | 2014          | 2015                     | 2014            |
| <b>Comprehensive income (loss)</b>                      |        |                            |               |                          |                 |
| Net income (loss) for the period                        |        | \$ 296                     | \$ 223        | \$ (13)                  | \$ (118)        |
| Other comprehensive income (loss), net of taxes of nil: |        |                            |               |                          |                 |
| Items that will not be reclassified to net income       |        |                            |               |                          |                 |
| Remeasurements on employee benefit liabilities          | Note 5 | 542                        | (90)          | 790                      | (293)           |
| Items that will be reclassified to net income           |        |                            |               |                          |                 |
| Loss on derivatives designated as cash flow hedges      | Note 9 | (1)                        | -             | (1)                      | -               |
| <b>Total comprehensive income (loss)</b>                |        | <b>\$ 837</b>              | <b>\$ 133</b> | <b>\$ 776</b>            | <b>\$ (411)</b> |
| <b>Comprehensive income (loss) attributable to:</b>     |        |                            |               |                          |                 |
| Shareholders of Air Canada                              |        | \$ 836                     | \$ 132        | \$ 774                   | \$ (413)        |
| Non-controlling interests                               |        | 1                          | 1             | 2                        | 2               |
|   |        | <b>\$ 837</b>              | <b>\$ 133</b> | <b>\$ 776</b>            | <b>\$ (411)</b> |

# End-to-end Example

|   |   |
|---|---|
| <b>AIR CANADA</b>  | Condensed Consolidated Financial Statements and Notes<br>Quarter 2 2015 |
| <b>CONSOLIDATED STATEMENT OF COMPREHENSIVE INCOME (LOSS)</b>  |   |

## Table Understanding

### C. Identify context within document

| Unaudited   |        | Three months ended June 30 |               | Six months ended June 30 |                 |
|---|--------|----------------------------|---------------|--------------------------|-----------------|
| (Canadian dollars in millions)                          |        | 2015                       | 2014          | 2015                     | 2014            |
| <b>Comprehensive income (loss)</b>                      |        |                            |               |                          |                 |
| Net income (loss) for the period                        |        | \$ 296                     | \$ 223        | \$ (13)                  | \$ (118)        |
| Other comprehensive income (loss), net of taxes of nil: |        |                            |               |                          |                 |
| Items that will not be reclassified to net income       |        |                            |               |                          |                 |
| Remeasurements on employee benefit liabilities          | Note 5 | 542                        | (90)          | 790                      | (293)           |
| Items that will be reclassified to net income           |        |                            |               |                          |                 |
| Loss on derivatives designated as cash flow hedges      | Note 9 | (1)                        | -             | (1)                      | -               |
| <b>Total comprehensive income (loss)</b>                |        | <b>\$ 837</b>              | <b>\$ 133</b> | <b>\$ 776</b>            | <b>\$ (411)</b> |
| <b>Comprehensive income (loss) attributable to:</b>     |        |                            |               |                          |                 |
| Shareholders of Air Canada                              |        | \$ 836                     | \$ 132        | \$ 774                   | \$ (413)        |
| Non-controlling interests                               |        | 1                          | 1             | 2                        | 2               |
|   |        | <b>\$ 837</b>              | <b>\$ 133</b> | <b>\$ 776</b>            | <b>\$ (411)</b> |




# End-to-end Example

|   |   |
|---|---|
| <b>AIR CANADA</b>  | Condensed Consolidated Financial Statements and Notes<br>Quarter 2 2015 |
| <b>CONSOLIDATED STATEMENT OF COMPREHENSIVE INCOME (LOSS)</b>  |   |

## Table Understanding

### D. Identify context outside document

| Unaudited   |        | Three months ended June 30 |               | Six months ended June 30 |                 |
|---|--------|----------------------------|---------------|--------------------------|-----------------|
| (Canadian dollars in millions)                          |        | 2015                       | 2014          | 2015                     | 2014            |
| <b>Comprehensive income (loss)</b>                      |        |                            |               |                          |                 |
| <b>Net income (loss) for the period</b>                 |        | \$ 296                     | \$ 223        | \$ (13)                  | \$ (118)        |
| Other comprehensive income (loss), net of taxes of nil: |        |                            |               |                          |                 |
| Items that will not be reclassified to net income       |        |                            |               |                          |                 |
| Remeasurements on employee benefit liabilities          | Note 5 | 542                        | (90)          | 790                      | (293)           |
| Items that will be reclassified to net income           |        |                            |               |                          |                 |
| Loss on derivatives designated as cash flow hedges      | Note 9 | (1)                        | -             | (1)                      | -               |
| <b>Total comprehensive income (loss)</b>                |        | <b>\$ 837</b>              | <b>\$ 133</b> | <b>\$ 776</b>            | <b>\$ (411)</b> |
| <b>Comprehensive income (loss) attributable to:</b>     |        |                            |               |                          |                 |
| Shareholders of Air Canada                              |        | \$ 836                     | \$ 132        | \$ 774                   | \$ (413)        |
| Non-controlling interests                               |        | 1                          | 1             | 2                        | 2               |
|   |        | <b>\$ 837</b>              | <b>\$ 133</b> | <b>\$ 776</b>            | <b>\$ (411)</b> |

|  |  |
|--|--|
| DBpedia  Browse using  Formats  |  |
| <b>About: net income (\$)</b>  |  |
| An Entity of Type : DatatypeProperty, from Named Graph : http://dbpedia.org  |  |
| <b>Property</b>  | <b>Value</b>   |
| rd:type  | <ul style="list-style-type: none"> <li>owl:FunctionalProperty</li> <li>rdf:Property</li> <li>owl:DatatypeProperty</li> </ul> |
| rd:domain  | db:Company   |
| rd:isDefinedBy   | http://dbpedia.org/ontology  |
| rd:label   | net income (\$) (en)   |
| rd:range   | xsd:double   |
| wd:describedby   | db:data/definitions.ttl  |
| prov:wasDerivedFrom  | http://mappings.dbpedia.org  |
| is: http://www.w3.org/2002/07/owl#DatatypeProperty   | http://dbpedia.org/ontology  |
| is: http://open.vocab.org/terms/describes of   | db:data/definitions.ttl  |



# End-to-end Example

## Output: Semantic cell annotations

| Value | Row Headers  | Column Headers                 | Currency         | Scale    | Audited | Table Title   | Document Title  | DBPedia Entities |
|-------|--|--------------------------------|------------------|----------|---------|---|---|------------------|
| -13   | Comprehensive income (loss), Net income (loss) for this period | Six months ended June 30, 2015 | Canadian dollars | Millions | No      | Consolidated Statement of Comprehensive Income (Loss) | Air Canada Second Quarter 2015 – Interim Unaudited Financial Statements and Notes | Net Income       |
| ...   | ...  | ...                            | ...              | ...      | ...     | ...   | ...   | ...              |



Downstream  
Applications

# Interaction between Table Extraction & Understanding

## 1. Extraction should maintain formatting info helpful for understanding

| Unaudited   |        | Three months ended June 30 |               | Six months ended June 30 |                 |
|---|--------|----------------------------|---------------|--------------------------|-----------------|
| (Canadian dollars in millions)                          |        | 2015                       | 2014          | 2015                     | 2014            |
| <b>Comprehensive income (loss)</b>                      |        |                            |               |                          |                 |
| Net income (loss) for the period                        |        | \$ 296                     | \$ 223        | \$ (13)                  | \$ (118)        |
| Other comprehensive income (loss), net of taxes of nil: |        |                            |               |                          |                 |
| Items that will not be reclassified to net income       |        |                            |               |                          |                 |
| Remeasurements on employee benefit liabilities          | Note 5 | 542                        | (90)          | 790                      | (293)           |
| Items that will be reclassified to net income           |        |                            |               |                          |                 |
| Loss on derivatives designated as cash flow hedges      | Note 9 | (1)                        | -             | (1)                      | -               |
| <b>Total comprehensive income (loss)</b>                |        | <b>\$ 837</b>              | <b>\$ 133</b> | <b>\$ 776</b>            | <b>\$ (411)</b> |
| <b>Comprehensive income (loss) attributable to:</b>     |        |                            |               |                          |                 |
| Shareholders of Air Canada                              |        | \$ 836                     | \$ 132        | \$ 774                   | \$ (413)        |
| Non-controlling interests                               |        | 1                          | 1             | 2                        | 2               |
|   |        | <b>\$ 837</b>              | <b>\$ 133</b> | <b>\$ 776</b>            | <b>\$ (411)</b> |



Indentation may help identify row header hierarchies



Text and background styling can help distinguish headers from data cells

# Table Extraction & Understanding Interaction

## 2. Errors in extraction can affect the quality of understanding

Original Table

| Species                           | Total for<br>Beach Drive | South of Broad<br>Branch Road | Sections Closed to<br>Traffic on Weekends<br>and Holidays | Sections Open to<br>Traffic North of<br>Broad Branch |
|-----------------------------------|--------------------------|-------------------------------|---|--|
| Raccoon                           | 8                        | 5                             | 3   | 0  |
| Squirrel                          | 7                        | 3                             | 3   | 1  |
| Deer                              | 4                        | 1                             | 0   | 3  |
| Water snake                       | 2                        | 0                             | 1   | 1  |
| Unidentified bird                 | 2                        | 2                             | 0   | 0  |
| Red-eyed vireo                    | 1                        | 0                             | 0   | 1  |
| Snapping turtle                   | 1                        | 0                             | 1   |  |
| Box turtle                        | 1                        | 0                             | 0   |  |
| Opossum                           | 1                        | 1                             | 0   |  |
| Domestic cat                      | 1                        | 0                             | 1   |  |
| Total                             | 28 (100%)                | 12 (43%)                      | 9 (32%)   | 7 (25%)  |
| Percent of road length            | 100%                     | 13%                           | 46%   | 41%  |
| Average annual road-kill per mile | 28/5.8 = 4.8             | 12/0.70 = 17.1                | 9/2.7 = 3.3   | 7/2.4 = 2.9  |

Extracted & Understood Table

| Total for                                    |              | South of Broad | Section Closed to<br>Traffic on Weekends | Sections Open to<br>Traffic North of |
|--|--------------|----------------|--|--------------------------------------|
| Species                                      | Beach Drive  | Branch Road    | and Holidays                             | Broad Branch                         |
| Raccoon                                      | 8            | 5              | 3  | 0                                    |
| Squirrel Deer                                | 7 4          | 3 1            | 3 0                                      | 1 3                                  |
| Water snake                                  | 2            | 0              | 1  | 1                                    |
| Unidentified Bird                            | 2            | 2              | 0  | 0                                    |
| Red-eyed vireo Snapping<br>turtle Box turtle | 1 1 1        | 0 0 0          | 0 1 0                                    | 1 0 1                                |
| Opossum Domestic cat                         | 1 1          | 1 0            | 0 1                                      | 0 0                                  |
| Total  | 28 (100%)    | 12 (43%)       | 9 (32%)                                  | 7 (25%)                              |
| Percent of road length                       | 100%         | 13%            | 46%                                      | 41%                                  |
| Average annual road-kill per mile            | 28/5.8 = 4.8 | 12/0.70 = 17.1 | 9/2.7 = 3.3                              | 7/2.4 = 2.9                          |



# Table Extraction & Understanding Interaction

## 2. Errors in extraction can affect the quality of understanding

| Class of Error  | Error or Complication               | Tables | Frequency | Affected Downstream Changes                |
|---|-------------------------------------|--------|-----------|--|
| <b>Detecting Tables in PDF Documents</b>                |                                     |        |           |  |
| Detection Errors  | Table Not Detected                  | 620    | 52.95%    | Table unavailable for data reuse           |
|   | Detected Non-Table                  | 200    | 20.96%    | Result not appropriate for data reuse      |
| <b>Extracting Tables from PDF Documents</b>             |                                     |        |           |  |
| Missing Table Data                                      | Table Spans Multiple Pages          | 279    | 29.25%    | Extracted data is incomplete               |
|   | Table Only Partially Extracted      | 116    | 12.16%    | Extracted data is incomplete               |
| Excessive Table Data                                    | Includes Non-Table Content          | 179    | 18.76%    | Incoherent extracted data values           |
|   | Merged Multiple Tables              | 13     | 1.36%     | Incoherent extracted data values           |
| Incorrect Table Structure                               | Split Cell into Multiple            | 230    | 24.11%    | Incoherent cell contents                   |
|   | Merged Adjacent Cells               | 220    | 23.06%    | Improper cell feature extraction           |
|   | Missing Whitespace Cells            | 32     | 3.35%     | Structural mismatch between cells          |
|   | Cell Misalignment                   | 5      | 0.52%     | Uninterpretable table rows/columns         |
| Incorrect Cell Content                                  | Incorrect Symbol                    | 67     | 7.02%     | Does not accurately reflect original table |
|   | Incorrect Number                    | 22     | 2.31%     | Misleading data values                     |
|   | Incorrect Contents                  | 22     | 2.31%     | Misleading cell contents                   |
| <b>Analyzing Table Type and Structure</b>               |                                     |        |           |  |
| Type Prediction   | Table Misclassified as Data Table   | 52     | 5.45%     | Table not appropriate for data reuse       |
|   | Table Misclassified as Layout Table | 22     | 2.31%     | Data values not recognized as usable       |
| Structural Identification                               | Incorrect Orientation               | 110    | 15.28%    | Data entities improperly extracted         |
|   | Incorrect Hierarchies               | 154    | 21.39%    | Structural relationships not recognized    |
|   | Incorrect Headers                   | 440    | 61.11%    | Data entities improperly labeled           |
|   | Incorrect Keys                      | 521    | 72.36%    | Data entities improperly identified        |
| <b>Parsing the Table Data for Downstream Data Reuse</b> |                                     |        |           |  |
| Parsing and Data Reuse Challenges                       | Repeated Table Structure            | 12     | 1.26%     | Data entities incorrectly merged           |
|   | Cells Contain Units/Type            | 79     | 8.28%     | Cells incorrectly typed                    |
|   | Cells Annotate Missing Data         | 55     | 5.77%     | Cells incorrectly typed                    |
|   | Cells Contain Uncertainty           | 45     | 4.72%     | Uncertainty influences data reuse          |
|   | Cells Contain Excessive Information | 36     | 3.77%     | Non-numeric values impact reuse            |
|   | Cells Contain Footnote              | 21     | 2.20%     | Cells incorrectly typed                    |

Common errors at different stages of table processing



J. Hoffswell et al. “Interactive Repair of Tables Extracted from PDF Documents on Mobile Devices”. CHI ‘19