**Hospital Scraper Pattern Intelligence Reference**

**Version:** 1.0  
**Date:** October 17, 2025  
**Based on:** 38 successfully configured hospitals

**Quick Reference Guide**

This document provides a comprehensive lookup table for all 13 scraping patterns, including real code samples from successfully configured hospitals.

**Pattern Success Rankings**

**Tier 1: Highly Reliable (Try First)**

1. **div\_classes** - 8 hospitals (21% of total) - HIGHEST SUCCESS
2. **combined\_h2** - 4 hospitals (11% of total) - VERY HIGH SUCCESS
3. **h2\_name\_h3\_title** - 3 hospitals (8% of total) - HIGH SUCCESS
4. **h2\_name\_p\_title** - 3 hospitals (8% of total) - HIGH SUCCESS
5. **list\_items** - 3 hospitals (8% of total) - HIGH SUCCESS

**Tier 2: Specialized Patterns**

1. **table\_rows** - 2 hospitals (5% of total) - HIGH SUCCESS
2. **boardcard\_gallery** - 1 hospital (3% of total) - MEDIUM SUCCESS
3. **field\_content\_sequential** - 1 hospital (3% of total) - MEDIUM SUCCESS
4. **nested\_list\_with\_ids** - 1 hospital (3% of total) - MEDIUM SUCCESS
5. **custom\_table\_nested** - 1 hospital (3% of total) - COMPLEX CASES

**Tier 3: Last Resort**

1. **manual\_entry\_required** - 4 hospitals (11% of total) - BLOCKED/IRREGULAR SITES

**Pattern 1: h2\_name\_h3\_title (Sequential Different Elements)**

**Success Rate:** HIGH - 3 hospitals  
**When to Use:** When you see h2 elements for names followed by h3 elements for titles

**Successful Hospitals**

* **FAC-707**: Trillium Mississauga - h2 for names, h3 for titles
* **FAC-624**: Joseph Brant - h2.entry-title for names, h3 for titles
* **FAC-596**: Collingwood - h2 sequential with h3

**Key Indicators**

* ✓ h2 and h3 elements present
* ✓ Sequential ordering (h2 followed by h3)
* ✓ Different element types for name and title
* ✓ Name in h2, title in h3

**Code Samples**

<h2>John Smith</h2>

<h3>Chief Executive Officer</h3>

<h2 class="entry-title">Jane Doe</h2>

<h3>Chief Financial Officer</h3>

**YAML Configuration**

pattern: "h2\_name\_h3\_title"

html\_structure:

name\_element: "h2"

title\_element: "h3"

notes: "Sequential h2→h3 pattern"

**Pattern 2: combined\_h2 (Combined in Single Element)**

**Success Rate:** VERY HIGH - 4 hospitals  
**When to Use:** When name and title are in the same element with a visible separator

**Successful Hospitals**

* **FAC-941**: Chatham-Kent - h2 text: "Name - Title", separator: " - "
* **FAC-952**: Hamilton St Joseph's - h2 text: "Name, Title", separator: ", "
* **FAC-619**: Brockville - p text: "Name, Title", separator: ","
* **FAC-970**: Woodstock - h2 with " – " separator

**Key Indicators**

* ✓ Name and title in same element
* ✓ Visible separator: dash (-), comma (,), pipe (|)
* ✓ Format: 'Name [separator] Title'
* ✓ Usually h2 or p elements

**Code Samples**

<h2>John Smith - Chief Executive Officer</h2>

<h2>Jane Doe, Chief Financial Officer</h2>

<p>Bob Johnson | President</p>

<h2>Alice Brown – Vice President</h2>

**YAML Configuration**

pattern: "combined\_h2"

html\_structure:

combined\_element: "h2" # or "p"

separator: " - " # Common: " - ", ", ", " | ", " – "

notes: "Name and title combined with separator"

**Pattern 3: table\_rows (Table Structure)**

**Success Rate:** HIGH - 2 hospitals  
**When to Use:** When executives are listed in a table with distinct columns

**Successful Hospitals**

* **FAC-661**: Markham Stouffville - table with td columns, name in col 1, title in col 2
* **FAC-781**: North York General - table structure with distinct columns

**Key Indicators**

* ✓ <table> tags present
* ✓ Names in one column, titles in another
* ✓ Usually td or th cells
* ✓ Structured tabular layout

**Code Samples**

<table>

<tr>

<td>John Smith</td>

<td>Chief Executive Officer</td>

</tr>

<tr>

<td>Jane Doe</td>

<td>Chief Financial Officer</td>

</tr>

</table>

**YAML Configuration**

pattern: "table\_rows"

html\_structure:

structure\_type: "table"

name\_location: "td\_column\_1" # or th\_column\_1 for header cells

title\_location: "td\_column\_2"

notes: "Table with name/title columns"

**Pattern 4: h2\_name\_p\_title (h2 Name → p Title)**

**Success Rate:** HIGH - 3 hospitals  
**When to Use:** When names are in h2 elements immediately followed by p elements for titles

**Successful Hospitals**

* **FAC-953**: Hanover - h2 for name, immediately followed by p for title
* **FAC-632**: Kingston General - h2 name, p title sequential
* **FAC-932**: Elizabeth Bruyère - h2/p pairing with accented names

**Key Indicators**

* ✓ h2 elements for names
* ✓ p elements immediately following for titles
* ✓ Sequential h2→p pairs
* ✓ No separator, separate elements

**Code Samples**

<h2>John Smith</h2>

<p>Chief Executive Officer</p>

<h2>Mélanie Dubé</h2>

<p>Présidente-directrice générale</p>

**YAML Configuration**

pattern: "h2\_name\_p\_title"

html\_structure:

name\_element: "h2"

title\_element: "p"

notes: "h2 followed by p pattern"

**Pattern 5: div\_classes (CSS Class-Based) ⭐ MOST COMMON**

**Success Rate:** VERY HIGH - 8 hospitals (21% of all configured hospitals)  
**When to Use:** When you see semantic CSS class names like .name, .title, .staff-name, etc.

**Successful Hospitals**

* **FAC-606**: Grand River - div.name + div.title
* **FAC-695**: Niagara Health - div.staff-name + div.staff-title
* **FAC-905**: William Osler - div with semantic class names
* **FAC-979**: Renfrew Victoria - span.name + span.position
* **FAC-837**: Lakeridge Oshawa - div.field-name + div.field-title
* **FAC-976**: Belleville - div classes for structure
* **FAC-736**: Southlake - div.tm-name + div.tm-primary-title
* **FAC-745**: Orillia - div.name + div.position

**Key Indicators**

* ✓ CSS classes with semantic names
* ✓ Classes like: .name, .title, .staff-name, .position, .field-*, .tm-*
* ✓ Separate divs or spans for name and title
* ✓ May have container class
* ✓ Look for: field-, staff-, tm-, or semantic class names

**Code Samples**

<div class="name">John Smith</div>

<div class="title">Chief Executive Officer</div>

<div class="staff-member">

<div class="staff-name">Jane Doe</div>

<div class="staff-title">Chief Financial Officer</div>

</div>

<span class="name">Bob Johnson</span>

<span class="position">President</span>

<div class="tm-name">Alice Brown</div>

<div class="tm-primary-title">Vice President</div>

**YAML Configuration**

pattern: "div\_classes"

html\_structure:

name\_class: "name" # Replace with actual class name

title\_class: "title" # Replace with actual class name

# container\_class: "member" # Optional - only add if mis-pairing occurs

notes: "CSS class-based structure"

**Pro Tip:** Start WITHOUT container\_class. Only add it if you get mis-paired results.

**Pattern 6: list\_items (List Items with Separator)**

**Success Rate:** HIGH - 3 hospitals  
**When to Use:** When executives are listed in <ul> or <ol> with combined name/title text

**Successful Hospitals**

* **FAC-957**: Haliburton Highlands - li elements with "Name | Title" format
* **FAC-790**: Queensway Carleton - ul/li structure with separators
* **FAC-850**: Bluewater Sarnia - list items with combined text

**Key Indicators**

* ✓ <ul> or <ol> present
* ✓ li elements contain combined text
* ✓ Separator visible: |, -, comma
* ✓ List-based layout

**Code Samples**

<ul>

<li>John Smith | Chief Executive Officer</li>

<li>Jane Doe - Chief Financial Officer</li>

<li>Bob Johnson, President</li>

</ul>

**YAML Configuration**

pattern: "list\_items"

html\_structure:

list\_type: "ul" # or "ol"

format: "combined"

separator: " | " # Common: " | ", " - ", ", "

notes: "List with separators, handles non-breaking spaces"

**Pattern 7: boardcard\_gallery (Gallery/Card Layouts)**

**Success Rate:** MEDIUM - 1 hospital  
**When to Use:** When executives are displayed in a visual grid/gallery with card layouts

**Successful Hospitals**

* **FAC-935**: Scarborough Health - div.boardcard gallery layout, name,title format

**Key Indicators**

* ✓ Gallery or card-based layout
* ✓ Container classes like: .card, .boardcard, .profile
* ✓ Visual grid layout
* ✓ Combined name,title in cards

**Code Samples**

<div class="boardcard">

John Smith, Chief Executive Officer

</div>

<div class="boardcard">

Jane Doe, Chief Financial Officer

</div>

**YAML Configuration**

pattern: "boardcard\_gallery"

html\_structure:

container\_class: "boardcard"

text\_format: "name\_comma\_title"

separator: ","

notes: "Gallery card layout"

**Pattern 8: custom\_table\_nested (Complex Nested Tables)**

**Success Rate:** MEDIUM - 1 hospital  
**When to Use:** When tables have complex nested structures with divs/p inside cells

**Successful Hospitals**

* **FAC-777**: Queensway Carleton (mixed) - Multiple table formats, complex nested structure

**Key Indicators**

* ✓ Nested table structures
* ✓ Mixed formats in same page
* ✓ Complex td/div nesting
* ✓ Requires custom parsing

**Code Samples**

<table>

<tr>

<td>

<p style="text-align: left">John Smith</p>

<div style="text-align: left">Chief Executive Officer</div>

</td>

</tr>

</table>

**YAML Configuration**

pattern: "custom\_table\_nested"

html\_structure:

structure\_type: "table\_with\_nested\_elements"

name\_selector: "td p[style\*='text-align: left']"

title\_selector: "td div[style\*='text-align: left']"

container: "td"

notes: "Complex nested table structure"

**Pattern 9: field\_content\_sequential (Sequential Same-Class Elements)**

**Success Rate:** MEDIUM - 1 hospital  
**When to Use:** When all data uses the same CSS class in a repeating pattern

**Successful Hospitals**

* **FAC-939**: Holland Bloorview - All in .field-content, pattern: Name(i), Title(i+1), Empty(i+2)

**Key Indicators**

* ✓ All data in same CSS class
* ✓ Repeating pattern: Name, Title, [empty/skip], repeat
* ✓ Sequential index-based extraction
* ✓ No structural difference between name/title

**Code Samples**

<div class="field-content">John Smith</div>

<div class="field-content">Chief Executive Officer</div>

<div class="field-content"></div>

<div class="field-content">Jane Doe</div>

<div class="field-content">Chief Financial Officer</div>

<div class="field-content"></div>

**YAML Configuration**

pattern: "field\_content\_sequential"

html\_structure:

element\_selector: ".field-content"

pattern\_type: "sequential\_every\_3"

start\_index: 3 # Skip first N elements

notes: "Sequential pattern: Name, Title, Empty, repeat"

**Pattern 10: nested\_list\_with\_ids (ID-Based or Complex Selectors)**

**Success Rate:** MEDIUM - 1 hospital  
**When to Use:** When elements use ID patterns or require complex CSS selectors

**Successful Hospitals**

* **FAC-827**: Peterborough - div[id^='t-'] for names, span[id^='d-'] for titles

**Key Indicators**

* ✓ IDs with patterns: id='t-1', id='d-1'
* ✓ Complex CSS selectors needed
* ✓ ID-based pairing
* ✓ No semantic class names

**Code Samples**

<div id="t-1">John Smith</div>

<span id="d-1">Chief Executive Officer</span>

<div id="t-2">Jane Doe</div>

<span id="d-2">Chief Financial Officer</span>

**YAML Configuration**

pattern: "nested\_list\_with\_ids"

html\_structure:

name\_selector: "div[id^='t-']" # CSS attribute selector

title\_selector: "span[id^='d-']" # CSS attribute selector

container: "li.column" # Optional parent container

notes: "ID-based pairing with attribute selectors"

**Pattern 11: qch\_mixed\_tables (Mixed Table Formats)**

**Success Rate:** MEDIUM - 1 hospital  
**When to Use:** When a page has multiple different table formats

**Successful Hospitals**

* **FAC-777**: Queensway Carleton (QCH) - Listed under custom\_table\_nested

**Key Indicators**

* ✓ Multiple table structures on same page
* ✓ Different formats for different sections
* ✓ Requires handling multiple patterns

**Pattern 12: p\_with\_bold\_and\_br (Bold Names with BR Separator)**

**Success Rate:** MEDIUM  
**When to Use:** When names are in <strong> or <b> tags within <p> elements

**Key Indicators**

* ✓ Names in bold/strong tags
* ✓ <br> tags separating name from title
* ✓ Paragraph-based layout

**Code Samples**

<p><strong>John Smith</strong><br>Chief Executive Officer</p>

<p><b>Jane Doe</b><br>Chief Financial Officer</p>

**YAML Configuration**

pattern: "p\_with\_bold\_and\_br"

html\_structure:

container\_selector: "p" # Optional, use if needed to narrow scope

notes: "Names in bold/strong, separated by BR tags"

**Pattern 13: manual\_entry\_required (Last Resort)**

**Success Rate:** N/A - 4 hospitals requiring manual entry  
**When to Use:** Site blocks scraping, titles in prose, or extremely irregular structure

**Hospitals Requiring Manual Entry**

* **FAC-947**: North Bay - Site blocks scraping
* **FAC-927**: Peterborough St Joseph - Site blocks scraping
* **FAC-763**: Pembroke - Titles in prose, not structured
* **FAC-975**: Custom nested - Too complex for patterns

**Key Indicators**

* ✓ Site blocks automated scraping
* ✓ Captcha or bot detection
* ✓ Titles in prose/paragraph form (not structured data)
* ✓ Extremely irregular structure
* ✓ No consistent HTML pattern

**YAML Configuration**

pattern: "manual\_entry\_required"

html\_structure:

notes: "Site blocks scraping - Cloudflare protection"

# OR

notes: "Titles in prose paragraphs, not extractable"

missing\_people:

- name: "John Smith"

title: "Chief Executive Officer"

- name: "Jane Doe"

title: "Chief Financial Officer"

**Decision Tree: Which Pattern to Use?**

**Step 1: Check for Obvious Blockers**

* **Site blocked/captcha?** → Pattern 13 (manual\_entry\_required)
* **Titles in prose paragraphs?** → Pattern 13 (manual\_entry\_required)

**Step 2: Identify Structure Type**

* **Table present?** → Check Pattern 3 (table\_rows) or Pattern 8 (custom\_table\_nested)
* **List (ul/ol) present?** → Check Pattern 6 (list\_items)
* **Gallery/card layout?** → Check Pattern 7 (boardcard\_gallery)

**Step 3: Check for CSS Classes**

* **Semantic class names present?** (.name, .title, .staff-, etc.) → Pattern 5 (div\_classes) ⭐ MOST COMMON
* **Same class for everything?** → Pattern 9 (field\_content\_sequential)
* **ID patterns?** → Pattern 10 (nested\_list\_with\_ids)

**Step 4: Check Element Types**

* **h2 + h3 pairs?** → Pattern 1 (h2\_name\_h3\_title)
* **h2 + p pairs?** → Pattern 4 (h2\_name\_p\_title)
* **Name + separator + title in same element?** → Pattern 2 (combined\_h2)
* **Bold/strong in p with br?** → Pattern 12 (p\_with\_bold\_and\_br)

**Step 5: Complex Cases**

* **Multiple patterns on same page?** → Pattern 11 (qch\_mixed\_tables) or Pattern 8
* **Nested complex structure?** → Pattern 8 (custom\_table\_nested)

**Common Troubleshooting**

**Problem: Names Being Rejected**

**Cause:** Name patterns in enhanced\_hospitals.yaml don't cover the name style  
**Solution:** Add appropriate pattern to global name\_patterns section

**Problem: No Executives Found**

**Cause:** Wrong pattern, wrong elements, or wrong separator  
**Solution:**

1. Run helper$analyze\_hospital\_structure(FAC, "Name", "URL")
2. Check actual HTML structure
3. Update pattern and html\_structure
4. Test again

**Problem: Found More/Fewer Than Expected**

**Cause:** expected\_executives count is wrong  
**Solution:** Update expected\_executives to match reality

**Problem: Mis-Paired Names and Titles**

**Cause:** Pattern incorrectly pairing names with titles  
**Solution:**

* For div\_classes: Add container\_class to limit pairing scope
* For sequential patterns: Check start\_index and pattern\_type
* For combined patterns: Verify separator is correct

**Statistics Summary**

**Total Configured Hospitals:** 38  
**Pattern Distribution:**

* div\_classes: 8 hospitals (21%)
* combined\_h2: 4 hospitals (11%)
* manual\_entry\_required: 4 hospitals (11%)
* h2\_name\_h3\_title: 3 hospitals (8%)
* h2\_name\_p\_title: 3 hospitals (8%)
* list\_items: 3 hospitals (8%)
* table\_rows: 2 hospitals (5%)
* Others: 11 hospitals (29%)

**Most Reliable Patterns:** div\_classes, combined\_h2, h2\_name\_h3\_title

**Quick Reference: Pattern Selection Priority**

**For new hospitals, try in this order:**

1. ✅ **div\_classes** (if you see .name, .title, or similar CSS classes)
2. ✅ **combined\_h2** (if you see "Name - Title" or "Name, Title" format)
3. ✅ **h2\_name\_h3\_title** (if you see h2/h3 pairs)
4. ✅ **h2\_name\_p\_title** (if you see h2/p pairs)
5. ✅ **table\_rows** (if you see a table)
6. ✅ **list\_items** (if you see ul/ol lists)
7. ⚠️ Consider specialized patterns for complex cases
8. 🛑 **manual\_entry\_required** (last resort only)

**Document End**  
For interactive pattern prediction, use the published web tool.  
For R integration, see SESSION\_LOG.md future work section.