



Secure and lasting livelihoods for Canadians

# CanWorks Job Site Product Requirements Document

<b>Executive Summary</b>	<b>2</b>
<b>Problem Statement &amp; Context</b>	<b>2</b>
The Problem	2
Target Audience	2
Strategic Fit	3
<b>Goals &amp; Success Metrics</b>	<b>3</b>
<b>Requirements</b>	<b>3</b>
Student	3
Recent Graduates	5
Administrator	5
Superuser	6
Employer	7
Security Officer	7
Regulatory & Compliance Officer	8
<b>User Experience &amp; Design</b>	<b>11</b>
<b>Appendix A: Scraping Compliance Audit</b>	<b>12</b>

# Executive Summary

CanWorks PATHFINDER is a specialized job aggregation platform designed to bridge the "Skills-Experience Gap" for young Canadians. While Canada leads the G7 in educational attainment, post secondary students and recent graduates face a fragmented and confusing job market where "entry-level" and "internships" roles are often mislabeled and scattered across hundreds of individual company career pages and job sites.

PATHFINDER solves this by acting as a centralized intelligence layer. Instead of requiring users to manually check dozens of sites, our platform aggregates and categorizes across the web opportunities relevant to students and early careers.

## Problem Statement & Context

### The Problem

Most job platforms are not built for early-career users. Searching for entry-level, internship, and co-op roles is often time-consuming, frustrating, and demotivating, as opportunities are often mislabeled and scattered across multiple job boards and employer websites. Many young Canadians lack access to reliable career guidance and professional networks. Without family or industry connections, they face significant barriers to understanding current labour-market trends, navigating the job search, and securing early-career opportunities in an increasingly competitive environment.

### Target Audience

students 19-25 in their final years of postsecondary studies and recent graduates within two years of completion, seeking co-ops, internships, and true entry level roles.

### Strategic Fit

The development of this one-stop-shop aggregator aligns with Canworks' core mission of Workforce Development. By scraping and centralizing fragmented opportunities from various company career pages, Canworks reduces the "search friction" for students. It supports the Pathfinder and Launchpad initiatives by providing a clear, filtered pathway to meaningful employment. By acting as an aggregator, Canworks can gather data on where the biggest gaps in junior hiring exist, allowing Canworks to better advocate for policy changes or new training programs.

## Goals & Success Metrics

Category	KPI	Goal
----------	-----	------

Impact	Number of users finding employment using the job hub	1 year after the formal launch: 100+
Growth	Number of registered users	6 months after the formal launch: 300+ users
	Number of employers (posting directly)	
	Number of sources scraped)	
	Number of jobs posting per month - for students	
	Number of jobs posting per month - for students	
Engagement	Job saved per user	Avg. 3 saved jobs per registered user
Effectiveness	Search-to-click rate	20% of searches lead to an outbound click
	User Rating	
Retention	Returning Visitor Rate	30% of users return within 30 days

## Requirements

### Student

#### User Stories

- As a student, I want to filter jobs by 'Co-op,' 'Internship,' or 'New Grad' status, so I do not waste time looking at roles that require prior experience.
- As a student, I want to see if a job qualifies for my school's co-op credits, so I can ensure the role meets my academic requirements.
- As a student, I want to see exactly when a job was scraped from the original site, so I do not waste time applying to roles that have already been filled.
- As a student, I want to search for keywords related to my transferable skills (e.g., 'Communication,' 'Writing,' 'Research') rather than just job titles, so I can find entry-level roles I am qualified for but did not know existed.
- As a busy student, I want to 'Follow' a specific search (e.g., 'Marketing Internships in Vancouver') and get an email alert when the scraper finds a match, so I can be among the first to apply.

- As a student in a mandatory Co-op program, I want a filter that specifically flags jobs mentioning 'Academic Credit' or 'School Enrollment Required, so I can meet my graduation requirements.
- As a busy student, I want to save job postings so that I can return to it later.
- As a student, I want to see a clear 'Apply on Company Site' button with the source name (e.g., 'View on Shopify Careers'), so I know exactly where I am going to finish my application.
- As a student, I want the job card to clearly highlight the 'Work-Integrated Learning' type (Co-op vs. Internship) even if the original site did not label it clearly.
- As a user (student, new graduate, and employer), I want a user friendly, and responsive system.

## Requirements

- Automated Classification: The system must categorize the positions based on the job title and keywords in the job descriptions (e.g., "Intern," "Junior," "0-2 years") to automatically assign a category: Co-op, Internship, or New Grad.
- Academic Credit Detection: The scraper must flag roles that contain phrases like "enrollment required," "academic credit," or "for credit only" to find WIL eligibility.
- Keyword Mapping: The search engine must support "OR" logic, allowing a search for "Writing" to return roles like "Communications Assistant" or "Content Creator".
- External CTA Labeling: The primary action button must dynamically display the source name (e.g., "Apply on Lever" or "Apply on Shopify Careers").
- Data Freshness Timestamp: Every job card must display two timestamps: 1) When the job was originally posted (if available) and 2) When Canworks last synchronized the data.
- Saved Jobs Library: Logged-in users must have a "Heart" or "Save" icon on each job card that adds the role to a "Saved Opportunities" personal tab.
- Automated Job Alerts: Users must be able to save a search filter and opt-in to daily or real-time email notifications for new matches.
- Educational Profile Collection: During onboarding, the system must require users to select their University/College, year in university/college, number of past-internship experience (0 and up), Major/Field of Study, and Province of Residence from standardized dropdown lists.

## Recent Graduates

### User Stories

- As a recent graduate, I want the platform to suggest roles based on my degree and the skills I learned in school, so I can discover industries I had not considered.
- As a graduate living in a rural area, I want to filter specifically for 'Remote within Canada' roles, so I do not get excited about 'Remote' jobs that are restricted to US-based applicants.
- As a graduate, I want to see entry-level roles from 20 different career pages in one feed, so I do not have to manually check every individual company website daily.

## Requirements

- "Remote Canada" Validation: The system must distinguish between "Global Remote" and "Remote within Canada" by checking for Canadian tax or residency mentions in the description.

- Skill-to-Job Suggestions: The user profile should allow students to input their "Major" and "Skills," which the system then uses to highlight "Recommended for You" roles.
- Personalized Feed Logic: The system must use the user's "Major" and "Skills" to rank matching job postings higher in the search results and home feed.
- A student and graduate must be able to update their profile.

## **Administrator**

### **User Stories**

- As an Admin, I want a weekly report of 'Dead Links' (404 errors) from specific sources, so I can find if a company's career page has changed its URL structure and update our scraper.
- As an Admin, I want to see which job categories have the highest 'Click-to-Apply' rates, so I can write blog posts or guides for students on the most in-demand fields.
- As an Admin, I want the system to automatically tag a job as 'Entry Level' based on keywords in the description (like 'New Grad,' 'Junior,' or '0-2 years'), so the feed stays relevant to our niche.
- As an Admin, I want an automated check to flag or remove jobs when the original source link returns a 404 error, so we keep a high-quality user experience.
- As an Admin, I want the scraper to recognize if the same job is posted on both LinkedIn and a company site, so the user does not see the same role twice in their feed.
- As an administrator, I want several ways to feed the system with input of positions: Spreadsheets conformed to a standard template via SFTP and manual upload; GUI accessible by employers granted with access; Screen scraping.
- As an administrator, I want to have a welcome page for new users where they will find their persona, create user id and password, and provide consent for non-PII data to be captured and used for analytics.
- As an administrator, I want to be able to add and remove data sources and users through configuration using my admin access.
- As an administrator, I want to see a daily systems generated log summarizing the status (success or failure) of all the data feeds.
- As an administrator, I want a low-maintenance and responsive system. When a feed fails, it should not affect the processing of the other feeds, and I need to be able to reload the failed feed.
- As an administrator, I must be able to easily keep reference tables used for all drop down boxes using my admin access.
- As an administrator I must be able to update and keep the "Career Insight" column to provide practical labour-market information and guidance that improves career readiness.

### **Requirements**

- Deduplication Logic: The system must use a hashing algorithm (comparing Company Name + Job Title + Location) to ensure duplicate postings from various sources are merged into one.
- Automated Link Validator: The system must run a daily background check on all active URLs. If a link returns a 404 or 500 error, the job must be automatically moved to "Archived".
- Error Reporting Dashboard: Admins must have a view showing which sources (e.g., Workday) are producing the most broken links to troubleshoot scraper scripts.

## **Superuser**

### **User Stories**

- As a superuser, I want to be able to easily query, aggregate, and drill down and generate reports based on the data captured.

### **Requirements**

- Engagement Analytics: The system must track "Outbound Clicks" per job category to find which industries are most popular among Canadian students.
- Employer information is only accessible by the employer and those with admin and super user access.

## **Employer**

### **User Stories**

- As an employer, I want to be able to keep my profile in an easy and secure way.

### **Requirements**

- Data Type 4) Data Source: Employer/Source (a maintainable static table), Employer Contact info, Preferred communication method, Available for events, CanWorks Hires, Special notes, and Sponsor status.
- Special Notes from the Employer: Optional - and must be part of our template for data sources from manual upload.

## **Security Officer**

### **User Stories**

- As an administrator [acting as Security Lead], I want to ensure the site is compliant with Ontario and Canadian compliance and regulatory rules - more specifically - I do not want logging of PII data.

### **Requirements**

- Security: The system must enforce multi factor authentication when anyone logs on, register and perform password reset.
- Access Monitoring: A daily system generated report viewable by the administrator logging access (read, update, delete) to employer and user (students and new graduates) profiles.
- Data Integrity: Certain records coming from certain data sources to be flagged in the database as "Update Prohibited" and allow only read/delete but NO Update to the content of the records.

## **Regulatory & Compliance Officer**

## User Stories

- As a user, I want transparency about my data and the ability to withdraw consent.

## Requirements

- Regulatory Compliance: The system must not capture any PII data other than for user account purposes.
- Mandatory Digital Consent: During sign-up, a mandatory modal must present the "Canworks Data Consent Form." The system must block account creation until the user explicitly clicks "I Accept".
- Consent Management: The user profile settings must include a "Data & Privacy" tab where users can view their signed consent.
- Right to Purge: When a user decides to withdraw consent for data being captured and used for analytics, the system must allow the user to update their preference, prompt the user to confirm, followed by a message to advise them their data will be purged in 24 hours.

## Technical Requirements

**Scrape Frequency:** How often do we update?

### Source List:

- [Civicjobs](#)
- [CareerEdge](#)
- [WorkInNonProfit.ca](#)
- [EcoCanada](#)
- [MaRS Discovery District Job Board](#)

*\*See Appendix A for the scraping compliance audit on all sites listed in the source list.*

### Data Fields to Capture and display:

#### Data Type 1) Position

- Type (Early Career, Internship, Co-Op)
- Term (e.g., Summer, Winter. Not applicable for early career positions)
- Anticipated graduation date, not mandatory and not applicable for early career positions)
- Application Deadline (not mandatory)
- mandatory academic requirements (in the backend we will map them to a fixed list when displaying to the users. The fixed list needs to be easily maintainable by the Administrator. Examples include minimum GPA, majors,)
- mandatory skill requirement eg. experience of specific softwares beyond MS 365 suite, programming languages, spoken languages (choose from a maintainable static table)
- Job Title
- With Pay (Yes/No)
- Company Name
- With pay (yes/no)
- Company size (values based on a drop-down box)
- Job Location (Remote/Hybrid/City, Province - values based on drop down boxes)

- Date Posted
- Job Description Highlights (to be displayed on the "Basic Profile" page) -We can use AI to generate this but must not be applied to jobs flagged as "Update Prohibited" - there needs to be a disclaimer that is for information purpose only.
- Original Application URL (The "Source" link)
- Subscribe to Event Announcements (Yes/No)
- Special Notes from the Employer (Optional - and must be part of our template for data sources from manual upload. The Administrator is allowed to manually update (single record or in bulk) this field for positions coming from any data source other than those marked as "Change Not Permitted)

A student or a new graduate must be able to easily filter by any combination of these fields real time or as an alert sent to their email in their preferred frequency (i.e., daily, weekly, monthly)

### **Data Type 2) Student and New Graduate profile**

- Username (only viewable by the user and Administrator)
- Password (only viewable by the user and Administrator)
- Currently in university /college (Yes/No)
- University/College graduated from
- University/College currently attending
- Year and month of graduation (expected or completed)
- Number of past-internship experience (0 and up)
- Major/Field of Study (we will have a fixed list from an easily maintainable reference table and user can select from a drop-down box)
- Province of Residence from standardized dropdown lists.
- CanWorks Success (Yes and No) and CanWorks internal job no

A student and graduate must be able to update their profile

### **Data Type 4) Data Source**

- Employer/Source (a maintainable static table with employer names or other sources e.g., sites we scrapped or job sites like LinkedIn)
- Employer Contact First Name, last name, email, Phone number
- Preferred communication method (a maintainable static table - email, i.e., text, email, scheduled call)
- Available for events (yes or no)
- CanWorks Hires (0 and up - maintainable based on students and new grades' update)
- Special notes (free form)
- Sponsor (Yes or No)
- Employer login id
- Employer password
- Number of employee (a drop down)

- Update prohibited (viewable & maintainable only by the admin)
- Feed Method (a drop down - viewable & maintainable only by the Administrator)
- An employer must be able to maintain their profile in an easy and secured way
- Employer information is only accessible by the employer and those with admin and super user access

### **Data Analytics**

- The admin and superusers must be able to easily query, aggregate, and drill down and generate reports based on the data captured.

### **Regulatory Compliance**

- The system must not capture any PII data other than for user account purposes
- The system will not screen-scrap sources not allowing automation of interaction with the sites.
- When a user decides to withdraw consent for data being captured and used for analytics, the system must allow the user to update their preference, prompt the user to confirm, followed by a message to advise them their data will be purged in 24 hours. The daily count of such cases needs to be in a report.
- Refer to details about special handling for “Update Not Prohibited” positions.

### **Security**

- The system must enforce multi factor authentication when anyone logs on, register and perform password reset.
- A daily system generated report viewable by the administrator logging access (read, update, delete) to employer and user (students and new graduates) profiles
- Certain records coming from certain data sources to be flagged in the database as “Update Prohibited” and allow only read/delete but NO Update to the content of the records. There needs to be an easily maintainable table to flag those sources.
- 

## **User Experience & Design**

# Appendix A: Scraping Compliance Audit

## 1.1 Civic Jobs

---

```
# robots.txt for https://civicjobs.ca/
```

```
#User-agent: Googlebot  
#Disallow:
```

```
#User-agent: Googlebot-image  
#Disallow:
```

```
User-agent: *  
Crawl-delay: 5  
Disallow: /mobile/
```

```
#User-agent: bingbot  
#Disallow: /
```

```
User-agent: AhrefsBot  
Disallow: /
```

```
User-agent: magpie-crawler  
Disallow: /
```

```
User-agent: M112bot  
Disallow: /
```

```
User-agent: trovitBot  
Disallow: /
```

```
User-agent: BLEXBot  
Disallow: /
```

```
User-agent: Barkrowler  
Disallow: /
```

```
User-agent: SemrushBot  
Disallow: /
```

```
User-agent: SemrushBot-SA  
Disallow: /
```

```
User-agent: SemrushBot-BA  
Disallow: /
```

```
User-agent: SemrushBot-SI  
Disallow: /
```

```
User-agent: SemrushBot-SWA  
Disallow: /
```

```
User-agent: SemrushBot-CT  
Disallow: /
```

```
User-agent: SemrushBot-BM  
Disallow: /
```

```
User-agent: SemrushBot-SEOAB  
Disallow: /
```

## 1.2 [WorkInNonProfit.ca](https://workinnonprofit.ca)

Sitemap: [https://workinnonprofits.ca/static/winp\\_ca\\_sitemap.xml.gz](https://workinnonprofits.ca/static/winp_ca_sitemap.xml.gz)

User-agent: \*  
Disallow: /email/  
Disallow: /account/  
Disallow: /share\_ping/

### 1.3 [Eco.ca](#)

---

User-agent: \*  
Crawl-delay: 15

### 1.4 MARSD

---

User-agent: \*  
Allow: /  
Sitemap: <https://techjobs.marsdd.com/sitemap.xml>  
Crawl-delay: 1

### 1.5 Career Edge

---

```
# START YOAST BLOCK
# -----
User-agent: *
Disallow:

Sitemap: https://www.careeredge.ca/sitemap\_index.xml
# -----
# END YOAST BLOCK
```

### 1.6 Scraping Compliance and source table

Source Name	Robots.txt Status	ToS/Privacy Policy	Compliance Proof
Eco Canada	Strict 15-second crawl-delay needed.	No mention of scraping, data mining, ir crawler on ToS or Privacy Policy.	eco.ca/robots.txt
CivicJobs.ca	5-second crawl-delay required. Disallows /mobile/ path.	No mention of scraping, data mining, ir crawler on ToS or Privacy Policy.	civicjobs.ca/robots.txt
Career Edge	No Disallow directives found. Public sitemap provided.	No mention of scraping, data mining, ir crawler on ToS or Privacy Policy.	careeredge.ca/robots.txt
MaRS Discovery District	Allow: / explicitly stated. 1-second crawl-delay requested.	No mention of scraping, data mining, ir crawler on ToS or Privacy Policy.	techjobs.marsdd.com/robots.txt
WorkInNonProfit.ca	Sitemap provided. Disallows only sensitive /account/ and /email/ areas.	No mention of scraping, data mining, ir crawler on ToS or Privacy Policy.	workinnonprofit.ca/robots.txt