

**REQUEST FOR INFORMATION (RFI) REGARDING THE
GLOBAL PUBLIC HEALTH INTELLIGENCE NETWORK (GPHIN) RENEWAL
PROCESSING PIPELINE
FOR PUBLIC HEALTH AGENCY OF CANADA (PHAC)**

TABLE OF CONTENTS

A) BACKGROUND AND PURPOSE OF THIS REQUEST FOR INFORMATION (RFI)	3
B) NATURE OF REQUEST FOR INFORMATION	4
C) NATURE AND FORMAT OF RESPONSES REQUESTED	5
D) RESPONSE COSTS	5
E) TREATMENT OF RESPONSES	5
F) CONTENTS OF THE RFI	5
G) SPECIFIC QUESTIONS TO INDUSTRY	6
H) FORMAT OF RESPONSES	7
I) ENQUIRIES	8
J) SUBMISSION OF RESPONSES	8
ANNEX A - STATEMENT OF REQUIREMENTS	9
1.0 Overview and Context	9
2.0 Current Metrics	9
3.0 Terminology	9
4.0 Requirements	10
ANNEX B – CROSS REFERENCING VENDOR SOLUTION TO ANNEX A ARTICLE 4	14

**REQUEST FOR INFORMATION (RFI) REGARDING
THE GLOBAL PUBLIC HEALTH INTELLIGENCE NETWORK (GPHIN) RENEWAL
PROCESSING PIPELINE
FOR PUBLIC HEALTH AGENCY OF CANADA (PHAC)**

A) Background and Purpose of this Request for Information (RFI)

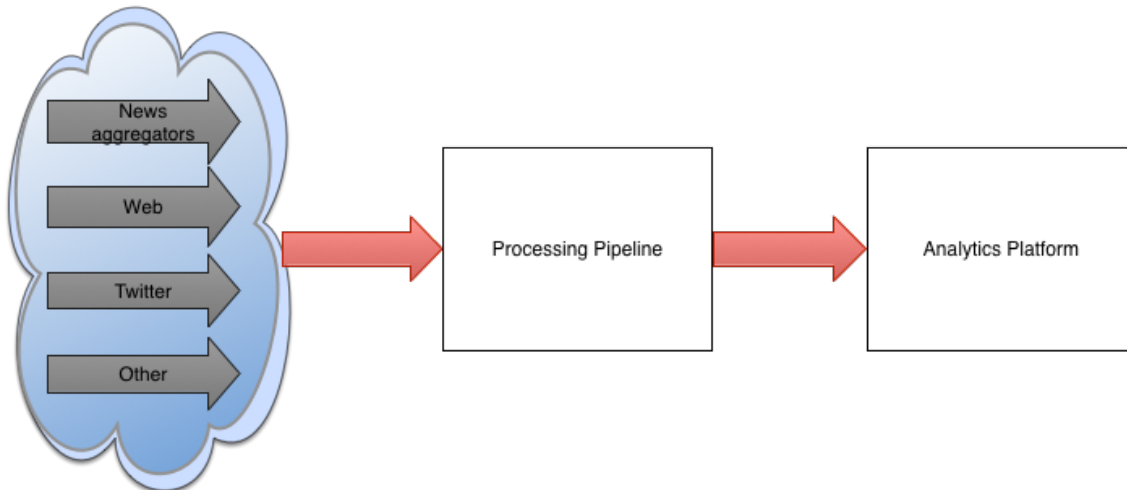
The Global Public Health Intelligence Network (GPHIN) is an Internet-based public health early-warning surveillance program owned by the Public Health Agency of Canada (PHAC) and is part of the World Health Organization's (WHO) Global Outbreak and Alert Response Network (GOARN). GPHIN links together a global network of public health professionals and organizations for situational awareness and early detection of emerging public health threats. GPHIN is relied on to help sift through media content in 9 languages from around the globe. The system is moderated by a team of multicultural and multidisciplinary human analysts, who further filter, interpret and disseminate relevant information.

GPHIN gathers preliminary reports of public health significance in nine languages (Arabic, English, French, Persian (Farsi), Portuguese, Russian, Simplified Chinese, Traditional Chinese, and Spanish) on a real-time, 24/7 basis. This unique, multilingual system gathers and disseminates relevant information on disease outbreaks and other public health events by monitoring global media sources such as news wires and web sites. The information is filtered for relevancy by an automated process, and then analyzed by PHAC officials. The output is categorized and made accessible to users around the world via the web. Notifications about public health events that may have serious public health consequences are immediately forwarded to subscription users.

In Sept. 2013, Shared services (SSC) made the assessment that the GPHIN platform was not compliant with the newly established Government of Canada Information (IT) policies and standards. This, is why combined with the ageing infrastructure PHAC is seeking input on a new Solution.

The high-level architecture of the new GPHIN system under the GPHIN Renewal Project will consist of two principal components:

1. An automated information **Processing Pipeline** to collect data from open Web sources, parse the data, extract relevant information and prepare the data for analysis. This is the subject of this Request for Information.
2. An **Analytics Platform** for data storage, analysis, rapid risk assessment, rendering and dissemination. This will be developed by the Information Management Services Directorate (IMSD) resources of Health Canada / Public Health Agency of Canada by integrating standard corporate analytic tools including SAS Text Analytics / SAS Visual Analytics.



The GPHIN Renewal Project will address the strategic need for a 24/7 integrated situational awareness network and platform in order to facilitate decision-making for Canada's Health Portfolio in response to domestic emergencies and/or events including those with international dimensions. The GPHIN Renewal Project must also align with the creation of Shared Services Canada and Treasury Board standards such as accessibility, security and usability.

The objectives of this RFI are as follows:

1. To determine if there is a Commercial-off-the-Shelf (COTS) solution that will most closely meet the needs for the Processing Pipeline and what functionality is available in what product.
2. To determine which software providers have a COTS solution that will most closely meet the needs for the Processing Pipeline and what functionality is available in what product.
3. To determine as a result, the Project budget and timeline for the planning of the next phase of the Project.
4. To determine the level of effort required to integrate the COTS product into the overall GPHIN Project End of State.

B) Nature of Request for Information

This is not a bid solicitation. This RFI will not result in the award of any Contract. As a result, potential suppliers of any goods or services described in this RFI should not reserve stock or facilities, nor allocate resources, as a result of any information contained in this RFI. Nor will this RFI result in the creation of any source list. Therefore, whether or not any potential supplier responds to this RFI will not preclude that supplier from participating in any future procurement. Also, the procurement of any of the goods and services described in this RFI will not necessarily follow this RFI. This RFI is simply intended to solicit feedback from industry with respect to the matters described in this RFI.

C) Nature and Format of Responses Requested

Respondents are requested to provide their comments, concerns and, where applicable, alternative recommendations regarding how the requirements or objectives described in this RFI could be satisfied. Respondents are also invited to provide comments regarding the content, format and/or organization of any draft documents included in this RFI. Respondents should explain any assumptions they make in their responses.

D) Response Costs

Canada will not reimburse any respondent for expenses incurred in responding to this RFI.

E) Treatment of Responses

1. **Use of Responses:** Responses will not be evaluated. However, the responses received may be used by Canada to develop or modify procurement strategies or any draft documents contained in this RFI. Canada may review any responses received by the RFI closing date. Canada may, in its discretion, review responses received after the RFI closing date.
2. **Review Team:** A review team composed of representatives of HC/PHAC and possibly PWGSC will review the responses. Canada reserves the right to hire any independent consultant, or use any Government resources that it considers necessary to review any response. Not all members of the review team will necessarily review all responses.
3. **Confidentiality:** Respondents should mark any portions of their response that they consider proprietary or confidential. Canada will handle the responses in accordance with the *Access to Information Act*.
4. **Follow-up Activity:** Canada may meet with each Respondent upon request (one-on-one meeting). Following the closing date, the Contracting Authority may follow up individually with all Respondents who indicate in their responses that they wish to meet with Canada. Canada may request that the Respondent provide an overview of the functionalities of the proposed solution and deliver a demonstration of commercial products so that Canada may obtain a better understanding of the Solution. During any such demonstration, Canada may interact with the Respondent to ask questions in order to gain a better understanding of the capabilities of the proposed solution.
5. **Documentation or any other information** of the proposed Solution, tool suite, or supporting third party applications is welcome.

F) Contents of the RFI

1. This RFI contains a draft Statement of Requirements (see Annex A). This document remains a work in progress, requirements may be added or

modified or deleted. ***Comments regarding any aspect of the draft document are welcome.***

2. This RFI also contains specific questions addressed to the industry.

G) Specific Questions to Industry

This section asks the Vendor Community to respond to the questions below:

- 1) The Processing Pipeline is composed of independent steps such as duplicate removal, translation, scoring, and information extraction. Based on the requirements identified in Annex A what products in the Vendor's Solution provide what functionality?
- 2) What steps in the Processing Pipeline are not currently covered in the Vendor's Solution? What would be involved in incorporating the rest of the steps in the Solution?
- 3) What other capabilities are available in the Vendor's Solution that are not covered in Annex A?
- 4) How would the Solution be deployed and are there any options? In particular, is the Vendor proposing to provide the Solution as a service or as a product that the Government of Canada will host?
- 5) If the Vendor is providing a product that the Government of Canada will host, will the Solution be able to integrate with or work with the full suite of SAS analytics tools including SAS VA, Enterprise Miner and Text Miner?
- 6) What is the ideal technical environment for deployment of the Processing Pipeline Solution and would it be compliant with the environment identified in Annex A?
- 7) What are the various metrics on processing functions and output to allow for monitoring and evaluation?
- 8) Describe your approach and costing in terms of reliability, availability, scalability and operational monitoring of the pipeline.
- 9) How do you typically license your COTS product (i.e. type of users, by servers, by CPU, concurrent users, enterprise-wide)?
- 10) What is the price estimated for each of the components?
- 11) Describe your Solution's licensing, maintenance and support model, including pricing, volume discounts and any optional add-on modules.
- 12) How do you provide support for your Solution, including the application of security patches and fixes, the implementation of known vulnerability remediation actions and troubleshooting and resolution of incidents or problems?

13) What steps could be undertaken to minimize the risk to all Parties involved?

H) Format of Responses

1. **Cover Page:** If the response includes multiple volumes, Respondents are requested to indicate on the front cover page of each volume the title of the response, the solicitation number, the volume number and the full legal name of the Respondent.
2. **Title Page:** The first page of each volume of the response, after the cover page, should be the title page, which should contain:
 - a) the title of the Respondent's response and the volume number;
 - b) the name and address of the Respondent;
 - c) the name, address, email and telephone number of the Respondent's contact;
 - d) the date; and,
 - e) the RFI number.
3. **Corporate Profile:** Each Respondent should provide a brief description of the Company background (Expertise; Related products; Current (in production) clients' reference and URLs; Partners; Product roadmap; and history of development of the product).
4. **Solution Description:** Each Vendor should provide the following information:
 - a) A Solution identifier such as a model number, version number and a description of all components required for the Solution cross referenced if possible against the requirements identified in Article 4.0 of Annex A.;
 - b) Costing model;
 - c) Licensing model; and
 - d) Support and Maintenance Service levels.
5. **Cross Referenced Grid:** Respondents are requested to complete a grid included in Annex B indicating which of the functional requirements identified in Annex A, Article 4 they meet; as well as any additional information that would be relevant for the specific criteria.
6. **Numbering System:** Respondents are requested to prepare their response using a numbering system corresponding to the one in this RFI. All references to descriptive material, technical manuals and brochures included as part of the response should be referenced accordingly.
7. **Response to be sent electronically:** Canada requests that respondents submit their responses electronically by e-mail to the Contracting Authority define below.

I) Enquiries

Because this is not a bid solicitation, Canada will not necessarily respond to enquiries in writing or by circulating answers to all potential suppliers. However, respondents with questions regarding this RFI may direct their enquiries to:

Contracting Authority:	Jason Solomon
E-mail Address:	Jason.Solomon@hc-sc.gc.ca
Telephone:	613-668-6410

J) Submission of Responses

5. **Time and Place for Submission of Responses:** Suppliers interested in providing a response should email their response to Jason Solomon at Jason.solomon@hc-sc.gc.ca.
6. **Responsibility for Timely Delivery:** Each respondent is solely responsible for ensuring its response is delivered on time to the Contracting Authority.
7. **Identification of Response:** Each respondent should ensure that its name and return address is provided and that the solicitation number and the closing date appear legibly in the subject line of the e-mail.

ANNEX A - STATEMENT OF REQUIREMENTS

This Statement of Requirements provides an overview of the Processing Pipeline, some of the terminology, interface requirements, function and non-functional capabilities of the required solution:

1.0 Overview and Context

The new Processing Pipeline will comprise of several automated stages to acquire the data from open Web sources, to filter and to reduce the volume of data, to extract specific public-health information relevant to GPHIN and prepare the data for analysis and dissemination by the GPHIN analysts using the Analytics Platform. The Processing Pipeline is responsible for cleaning the data source of unwanted noise, elimination of duplicates, translation, scoring and linguistic processing.

2.0 Current Metrics

GPHIN volumes (per day)

- Articles from a data feed or news aggregator: 2,300
- Duplicates articles: 720 (31%)
- Articles presented to GPHIN analysts: 1,580

3.0 Terminology

Within the Statement of Requirements the following terms will have the meaning as ascribed in the table below:

Term	Meaning
Article	Text from a given source describing an event. Running Example: A news article from a traditional media source on the Web describing an update in the number of cholera cases in Haiti
Disease	IHR 2005: "disease" means an illness or medical condition, irrespective of origin or source, that presents or could present significant harm to humans; Running Example: cholera in Haiti
Event	An event is an activity, occurrence or incident that may have public health implications or become an emerging public health threat, including, but not limited to those that are of infectious, zoonotic, food safety, chemical, radiological or nuclear origin or source. Running Example: The event is the cholera outbreak occurring in Haiti.
Event-based surveillance	IHR 2005: Event-based surveillance is the organized and rapid capture of information about events that are a potential risk to public health. This information can be rumours and other ad-hoc reports transmitted through

Term	Meaning
	formal channels (i.e. established routine reporting systems) and informal channels (i.e. media, health workers and NGO reports).
Story	A story connects one or more topics based on a common thread and evolves over time. A topic may fall within multiple stories. Running Example: The cholera article is related to the overarching Haiti earthquake story.
Taxonomy	A set of keywords such as disease names and their variants, symptoms and other terms in multiple languages arranged according to a hierarchy as defined by the Public Health Agency of Canada used to filter and classify text.
Topic	A topic connects one or more events based on a common thread and evolves over time. An event may fall within multiple topics. Running Example: The cholera article is related to the overarching Cholera Outbreak in Haiti topic.

4.0 Requirements

In terms of requirements, it is expected that the proposed Solution must meet the following requirements:

#	Criteria	Detail
1.	Technical Environment	If the Vendor is providing a product that the Government of Canada will host, the Solution must be compatible with: <ul style="list-style-type: none"> a. Java 2 Enterprise stack or Microsoft .Net stack; b. Oracle database servers; and c. Not require any physical access to servers.
2.	Interface Requirements	The Solution must have an operational system that will process data continuously as it is collected and at a minimum meets all of the following specific characteristics: <ul style="list-style-type: none"> a) Real-time; b) Variety of sources; and c) Multiple languages.
3.	Interface Requirements	The Solution must provide automated server-server processing without manual intervention.
4.	Input	The Solution must process text files from various sources as identified by the Public Health Agency of Canada (PHAC) which at a minimum include XML, HTML and textual content from sources.

#	Criteria	Detail
5.	Output	The Solution must allow XML data files to be submitted using a Web service or pushed to a HC/PHAC file server using FTP.
6.	Duplicate detection	At the article level, the Solution must at a minimum: <ul style="list-style-type: none"> a) Identify identical articles from the same source; and b) Identify identical articles from a different source.
7.	Similarity detection	At the article level, the Solution must at a minimum identify similarity and differences between articles.
8.	Language detection	At the article level, the Solution must at a minimum be capable of detecting the language of any input.
9.	Translation	At the article level, the Solution must ensure that Source articles are output in their original language along with a translated version.
10.	Translation	At the article level, the Solution must ensure that Source language article are translated to English and French using own or third-party translation engines specified by the Public Health Agency of Canada. At a minimum the current source languages to be translated include: <ul style="list-style-type: none"> a) English, b) Arabic, c) French, d) Persian, e) Portuguese, f) Russian, g) Spanish, h) Simplified Chinese; and i) Traditional Chinese.
11.	Translation	At the article level, the Solution must ensure that the Solution can augment the translation mechanisms as new input languages are added. An example is Indonesian.
12.	Article Summarization	At the article level, the Solution must summarize article content to capture salient points in a brief paragraph.
13.	Internal representation of input	At the article level, the Solution must create an internal representation of input to meet retention restrictions of input and potential future needs for re-processing and re-mining of the article content.
14.	Taxonomy	The Solution must use a taxonomy to filter and classify article content for further processing and annotation.
15.	Taxonomy Management	The Solution must provide a mechanism for the Public Health Agency of Canada to define and update a taxonomy used for processing article content.

#	Criteria	Detail
16.	Relevance Score	The Solution must provide the functionality to generate a relevancy score that conveys the article's relevance to its assigned event, topic and story (i.e. one relevancy score for each).
17.	Signal Score	The Solution must provide the functionality to generate a score that identifies the event's potential significance from a public health perspective and used to determine whether or not an article needs further immediate analysis and attention.
18.	Meta-data	The Solution must extract meta-data from the source article. For purposes of this criterion, the Meta-data includes information about the source of the original article like news provider, date published and date received by the system, language, and reporting location.
19.	Processing of content from manually identified URLs	The Solution must be able to process content from manually identified URLs i.e. allows the Solution to process input based on a given URL.
20.	Re-mining	The Solution must be able to re-process internal representation of input using new taxonomy or algorithm changes.
21.	Tuning	The Solution must provide the ability to iteratively enhance relevance and signal scoring based on user input / feedback.
22.	Supplement processing by integrating with Academia and Research partners	The Solution must provide the ability to collaborate with other organizations to facilitate research and development of specific enhancements identified by PHAC.
23.	Information Extraction	The Solution must provide the ability to identify and extract important / relevant information for subsequent analysis, contextualization and linkage based on the type of input. At minimum, the following information must be extracted when available: <ul style="list-style-type: none"> a) Location: both source of the information and the location of the event to be extracted at the most granular level possible; b) Disease; c) Symptoms; and d) Cases, deaths, hospitalizations, recoveries.
24.	Topic Building	The Solution must provide the ability to build topics from the processed articles and their associated events.
25.	Story Building	The Solution must provide the ability to build stories from the identified topics.
26.	Topic Summarization	The Solution must provide an iterative process to summarize topics as they unfold to capture salient points in multiple brief paragraphs.

#	Criteria	Detail
27.	Story Summarization	The Solution must provide an iterative process to summarize stories as they unfold to capture salient points in multiple brief paragraphs.

ANNEX B – CROSS REFERENCING VENDOR SOLUTION TO ANNEX A ARTICLE 4

#	Criteria	Detail	Included in Solution (Yes = Y; No = N)	Additional Information
1.	Technical Environment	<p>If the Vendor is providing a product that the Government of Canada will host, the Solution must be compatible with:</p> <ul style="list-style-type: none"> a) Java 2 Enterprise stack or Microsoft .Net stack; b) Oracle database servers; and c) Not require any physical access to servers. 		
2.	Interface Requirements	<p>The Solution must have an operational system that will process data continuously as it is collected and at a minimum meets all of the following specific characteristics:</p> <ul style="list-style-type: none"> a) Real-time; b) Variety of sources; and c) Multiple languages. 		
3.	Interface Requirements	<p>The Solution must provide automated server-server processing without manual intervention.</p>		
4.	Input	<p>The Solution must process text files from various sources as identified by the Public Health Agency of Canada (PHAC) which at a minimum include XML, HTML and textual content from sources.</p>		
5.	Output	<p>The Solution must allow XML data files to be submitted using a Web service or pushed to a HC/PHAC file server using FTP.</p>		
6.	Duplicate detection	<p>At the article level, the Solution must at a minimum:</p> <ul style="list-style-type: none"> a) Identify identical articles from the same source; and b) Identify identical articles from a different source. 		
7.	Similarity detection	<p>At the article level, the Solution must at a minimum identify similarity and differences between articles.</p>		

#	Criteria	Detail	Included in Solution (Yes = Y; No = N)	Additional Information
8.	Language detection	At the article level, the Solution must at a minimum be capable of detecting the language of any input.		
9.	Translation	At the article level, the Solution must ensure that Source articles are output in their original language along with a translated version.		
10.	Translation	<p>At the article level, the Solution must ensure that Source language article are translated to English and French using own or third-party translation engines specified by the Public Health Agency of Canada. At a minimum the current source languages to be translated include:</p> <ul style="list-style-type: none"> a) English, b) Arabic, c) French, d) Persian, e) Portuguese, f) Russian, g) Spanish, h) Simplified Chinese; and i) Traditional Chinese. 		
11.	Translation	At the article level, the Solution must ensure that the Solution can augment the translation mechanisms as new input languages are added. An example is Indonesian.		
12.	Article Summarization	At the article level, the Solution must summarize article content to capture salient points in a brief paragraph.		
13.	Internal representation of input	At the article level, the Solution must create an internal representation of input to meet retention restrictions of input and potential future needs for re-processing and re-mining of the article content.		
14.	Taxonomy	The Solution must use a taxonomy to filter and classify article content for further processing and annotation.		

#	Criteria	Detail	Included in Solution (Yes = Y; No = N)	Additional Information
15.	Taxonomy Management	The Solution must provide a mechanism for the Public Health Agency of Canada to define and update a taxonomy used for processing article content.		
16.	Relevance Score	The Solution must provide the functionality to generate a relevancy score that conveys the article's relevance to its assigned event, topic and story (i.e. one relevancy score for each).		
17.	Signal Score	The Solution must provide the functionality to generate a score that identifies the event's potential significance from a public health perspective and used to determine whether or not an article needs further immediate analysis and attention.		
18.	Meta-data	The Solution must extract meta-data from the source article. For purposes of this criterion, the Meta-data includes information about the source of the original article like news provider, date published and date received by the system, language, and reporting location.		
19.	Processing of content from manually identified URLs	The Solution must be able to process content from manually identified URLs i.e. allows the Solution to process input based on a given URL.		
20.	Re-mining	The Solution must be able to re-process internal representation of input using new taxonomy or algorithm changes.		
21.	Tuning	The Solution must provide the ability to iteratively enhance relevance and signal scoring based on user input / feedback.		
22.	Supplement processing by integrating with Academia and Research partners	The Solution must provide the ability to collaborate with other organizations to facilitate research and development of specific enhancements identified by PHAC.		

#	Criteria	Detail	Included in Solution (Yes = Y; No = N)	Additional Information
23.	Information Extraction	<p>The Solution must provide the ability to identify and extract important / relevant information for subsequent analysis, contextualization and linkage based on the type of input. At minimum, the following information must be extracted when available:</p> <ul style="list-style-type: none"> a) Location: both source of the information and the location of the event to be extracted at the most granular level possible; b) Disease; c) Symptoms; and d) Cases, deaths, hospitalizations, recoveries. 		
24.	Topic Building	The Solution must provide the ability to build topics from the processed articles and their associated events.		
25.	Story Building	The Solution must provide the ability to build stories from the identified topics.		
26.	Topic Summarization	The Solution must provide an iterative process to summarize topics as they unfold to capture salient points in multiple brief paragraphs.		