



Public Health
Agency of Canada

Agence de santé
publique du Canada



Global Public Health Intelligence Network (GPHIN)

Machine Translation Summit
September 2005

Michael Blench

GPHIN Technical Advisor and Project
Coordinator



Global Public Health Intelligence Network (GPHIN)

This presentation will discuss:

- What is GPHIN?
- How does it work?
- SARS
- Where are we going?
- GPHIN's Machine Translation



What is GPHIN?

- A 'solution' comprised of a team of analysts directing a very innovative and unique technology.
- The 'Team':
 - Multi-ethnic
 - Multi-cultural
 - Multi-lingual
 - Multi-disciplined
 - Both genders and all age groups





Goal

- Provide to decision makers, early warning intelligence of potential public health threats to aid them in the avoidance, mitigation and management of an 'event'.



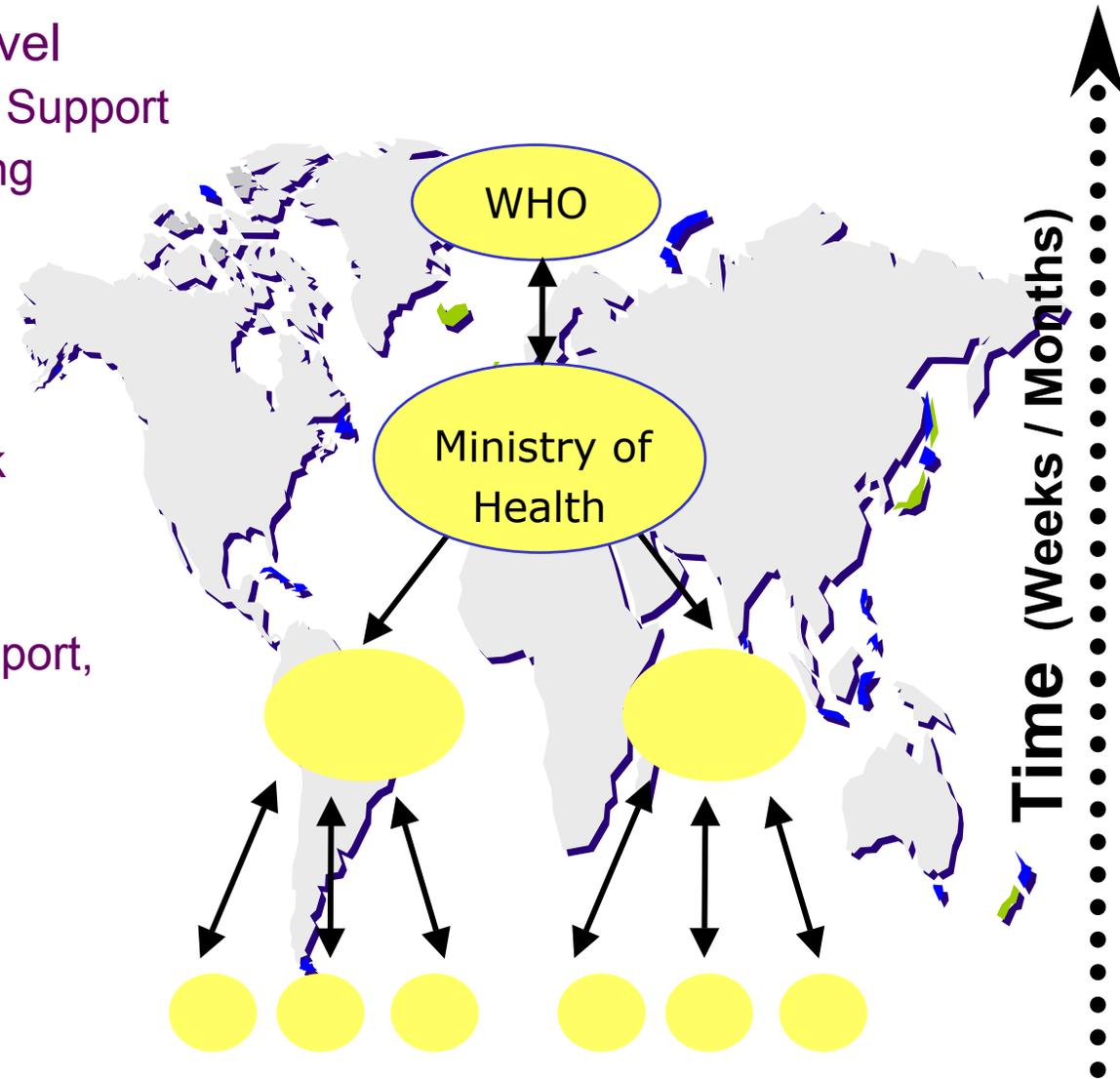
Public Health Organizations at the International Level

- WHO and GOARN
- PHAC – Public Health Agency of Canada
- CDC – Center for Disease Control
- Others



Public Health Surveillance System

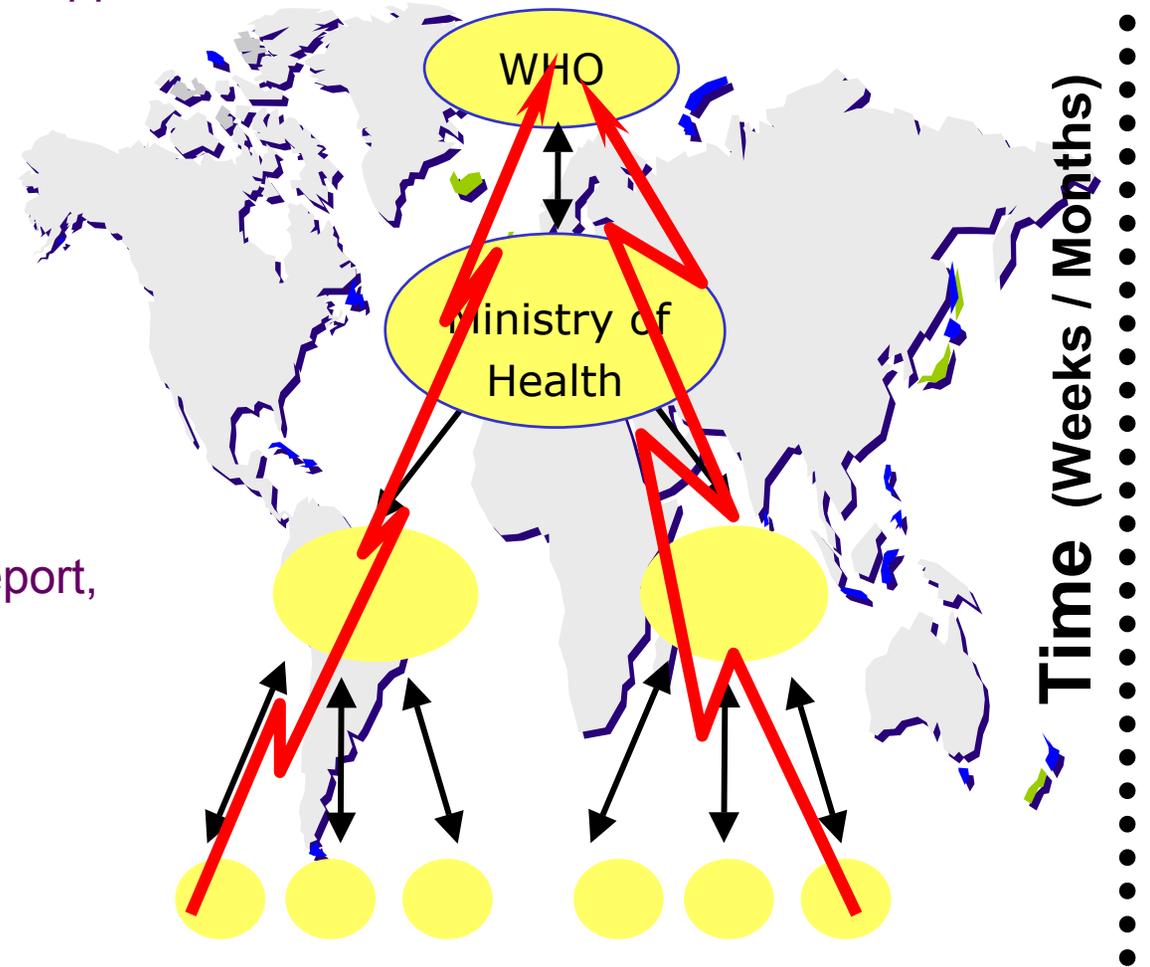
- Regional / International level
 - Analysis and Feedback. Support
 - Policy and targets funding
- Central level
 - Analyze, Investigate, Confirm, Plan, Fund, Respond, and Feedback
- Intermediate level
 - Analyze, Investigate, Report, Respond and Feedback
- Peripheral level
 - Detect, Treat, Report





Public Health Surveillance System with GPHIN

- Regional / International level
 - Analysis and Feedback. Support
 - Policy, target funding
- Central level
 - Analyze, Investigate, Confirm, Plan, Fund, Respond, Feedback
- Intermediate level
 - Analyze, Investigate, Report, Respond, Feedback
- Peripheral level
 - Detect, Treat, Report



GPHIN – Early Warning System

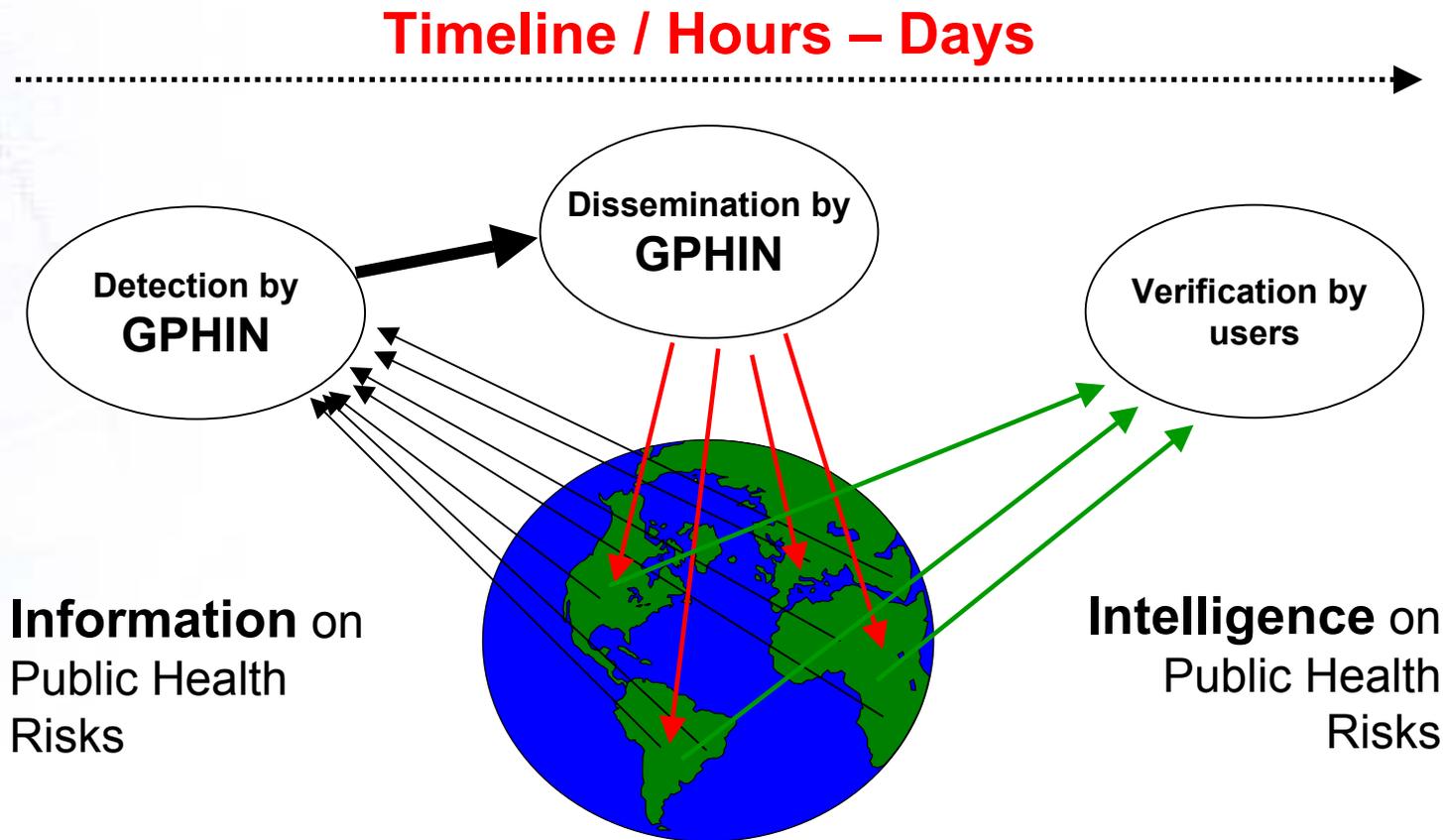


Background

- GPHIN's Role and management
 - GPHIN was developed by the Public Health Agency of Canada
 - Managed by the Agency's Centre for Emergency Preparedness and Response
- GPHIN's history
 - 1997 – Idea of two Health Canada physicians
 - 1998/99 – Prototype developed by Health Canada in partnership with WHO
 - 2000 – Proof-of-concept
 - 2001 – Prototype deployed into production
 - 2002 – SARS
 - 2003 – Collaborative Research Agreement with Nstein Technologies Inc.
 - 2004 – Official launch of GPHIN II at United Nations



GPHIN's Modus Operandi



- **Over 10,000 sources of information**
- **Arabic, English, French, Spanish, Russian, Chinese (Traditional & Simplified), Farsi**

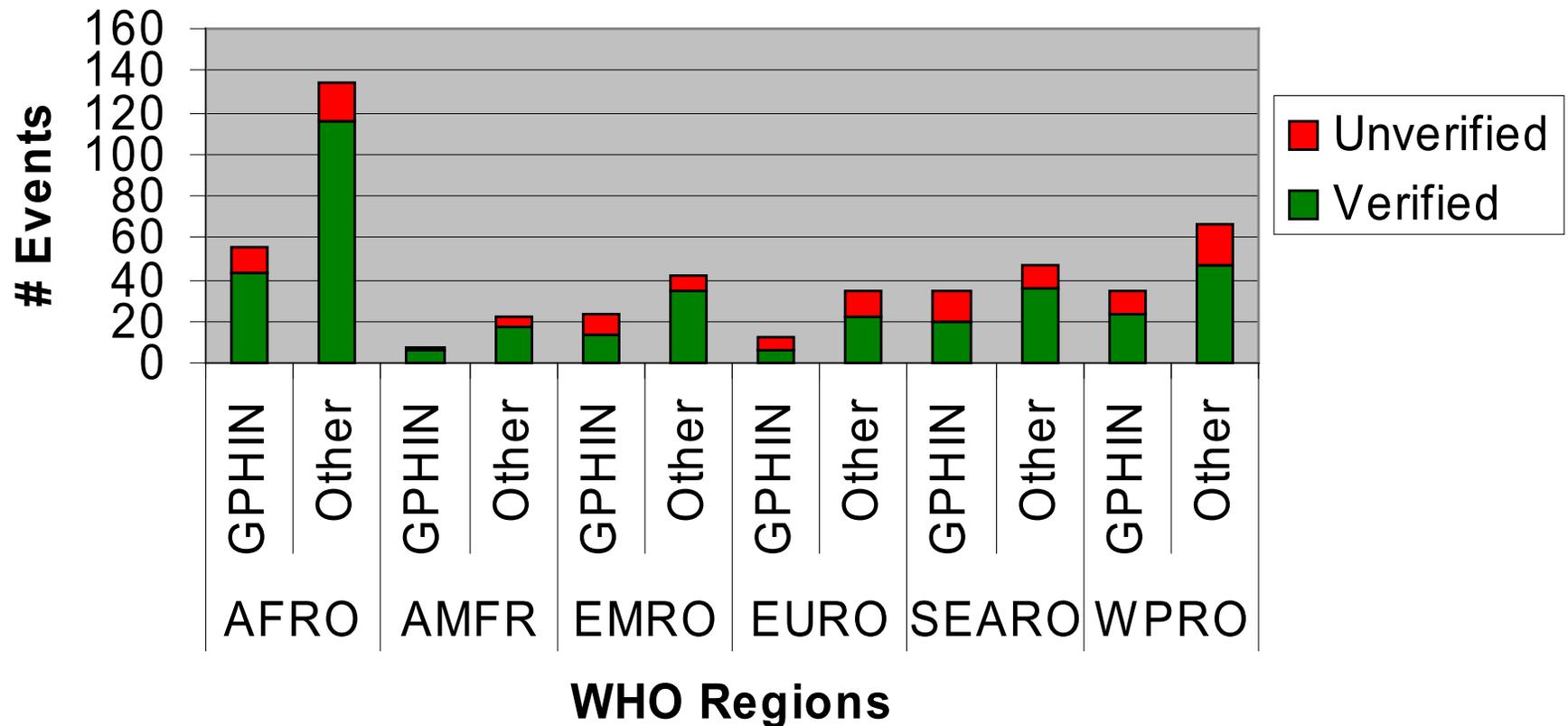


Source: WHO/CDS/CSR/Office of Alert and Response Operations (ARO)
Preliminary Data

WHO Intelligence & Verification Activities

01 Jan 2004 - 30 Jun 2005

Total events - N=513; Verified - n=384





GPHIN

Early Warning Multilingual System

- New Features:
 - Viewing in 8 languages.
 - Meta data querying
 - Immediate email alerts.
 - Robust, secure system

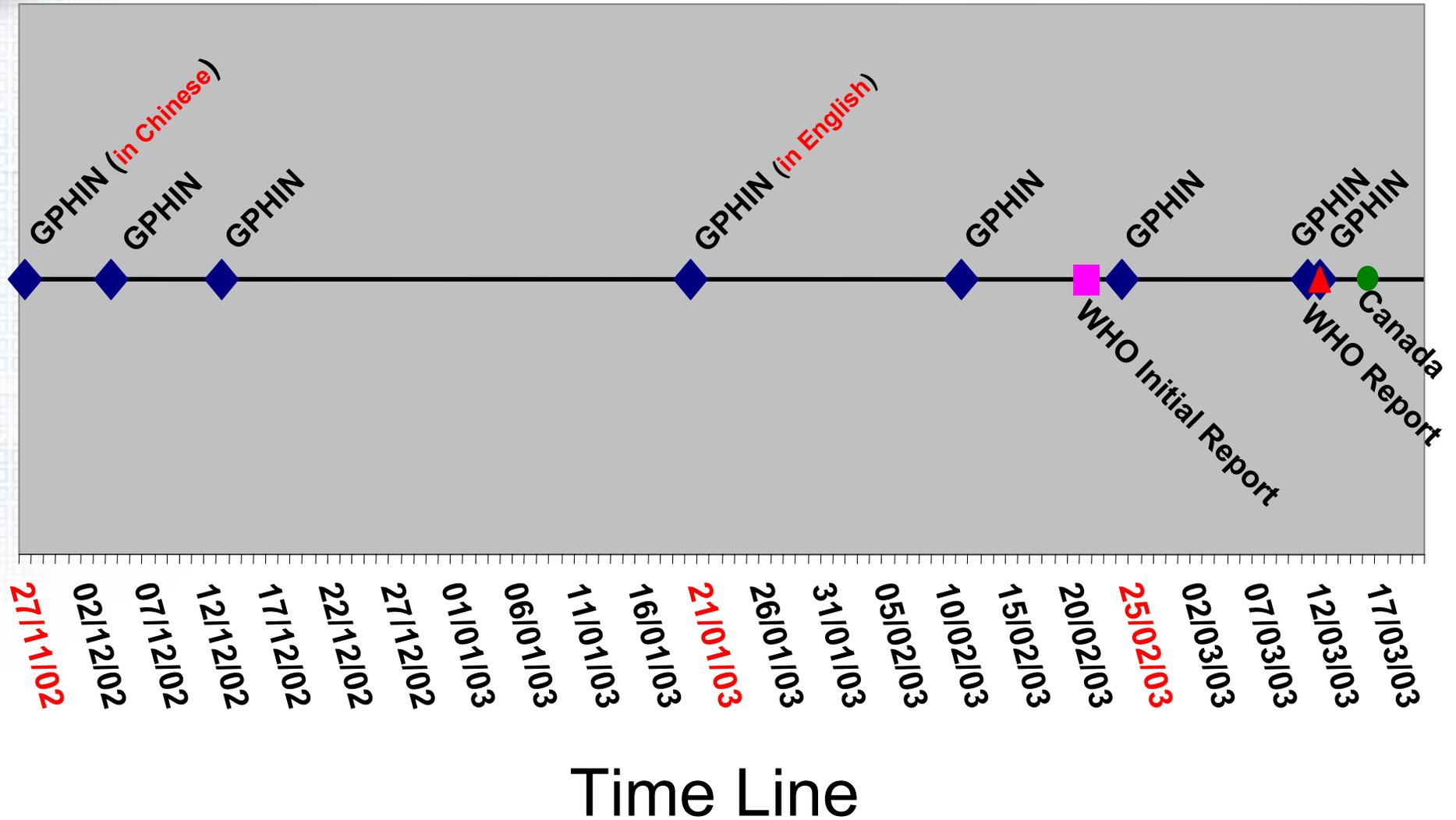


GPHIN Taxonomy Topics

- Human diseases – e.g. ebola
- Animal diseases – e.g. streptococcus suis
- Plant diseases – e.g. Asian soybean rust
- Other Biologics – e.g. anthrax
- Chemical incidents – e.g. chlorine plume
- Radioactive exposures – e.g. Chernobyl disaster
- Unsafe products – e.g. Prozac, NSAID
- Natural Disasters – e.g. tsunami, 26 Dec 04



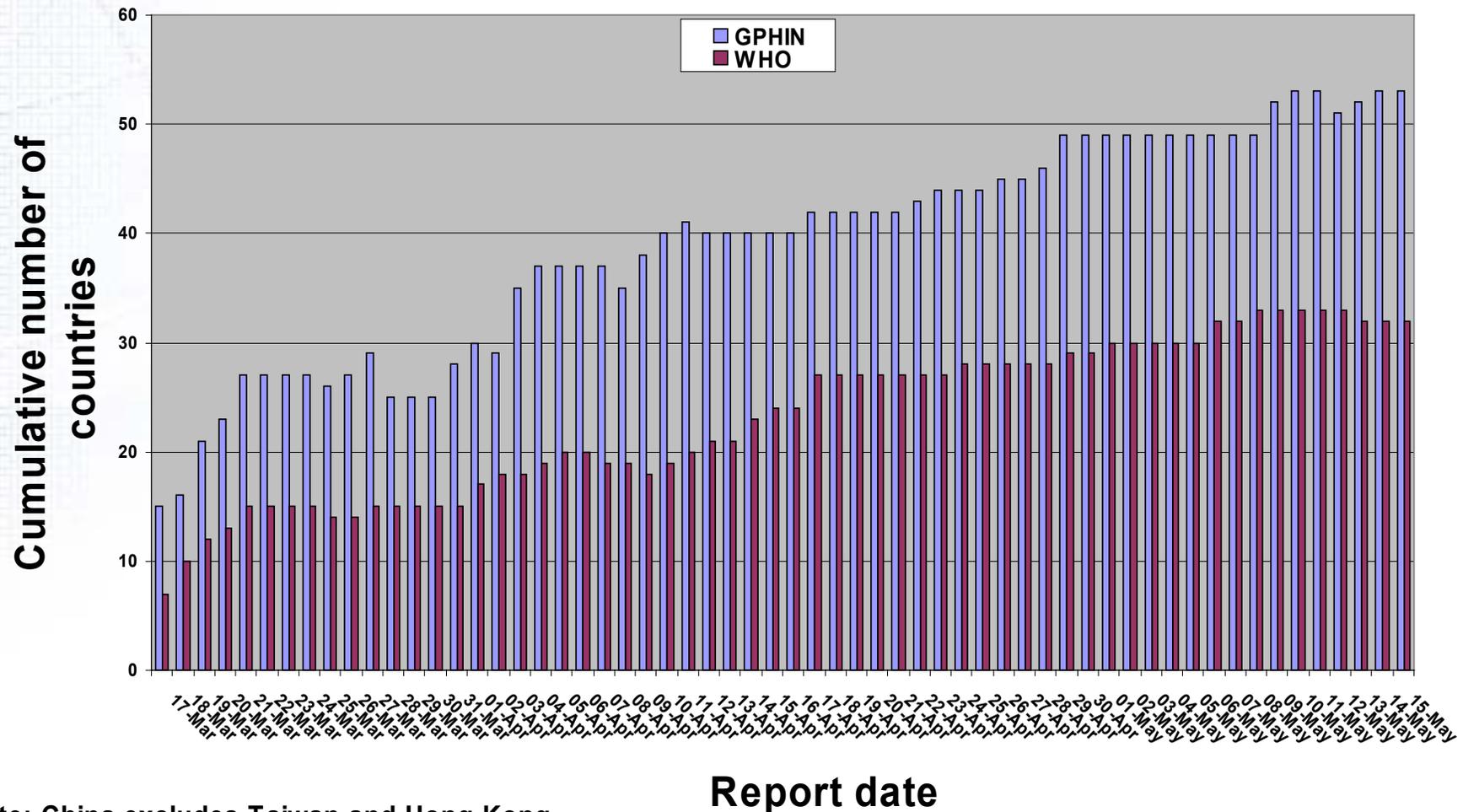
GPHIN's Reporting suggestive of an atypical event (SARS?)





Monitoring During SARS Outbreak

Cumulative number of countries reporting SARS as of May 15, 2003 by GPHIN and WHO



Note: China excludes Taiwan and Hong Kong
Source: GPHIN, 2003; WHO, 2003



GPHIN's Versatility

– Information for Action

- ***Detect*** potential public health threats
- ***Estimate*** magnitude of the public health threat
- ***Determine*** geographic distribution of public health threat
- ***Identify*** control and preventative measures considered and implemented
- ***Monitor*** concerns of the general public
- ***Observe*** situational politics



Next Steps: Medium Range

- English <-> Portuguese translation software
- Enhanced translation engine
- Summarizer
- Enhanced Search
- User profiles
- Text-to-Speech
- Basic GIS (drill down mapping)



Next Steps: Long Range

- Enhanced GIS
- Other languages (e.g. Asian)
- Predictive modelling
- Economic intelligence for predicting public health events
- Advanced search – Boolean, meta data & phonetics
- Speech-to-Text (Speech recognition)
- Speech-to-Speech

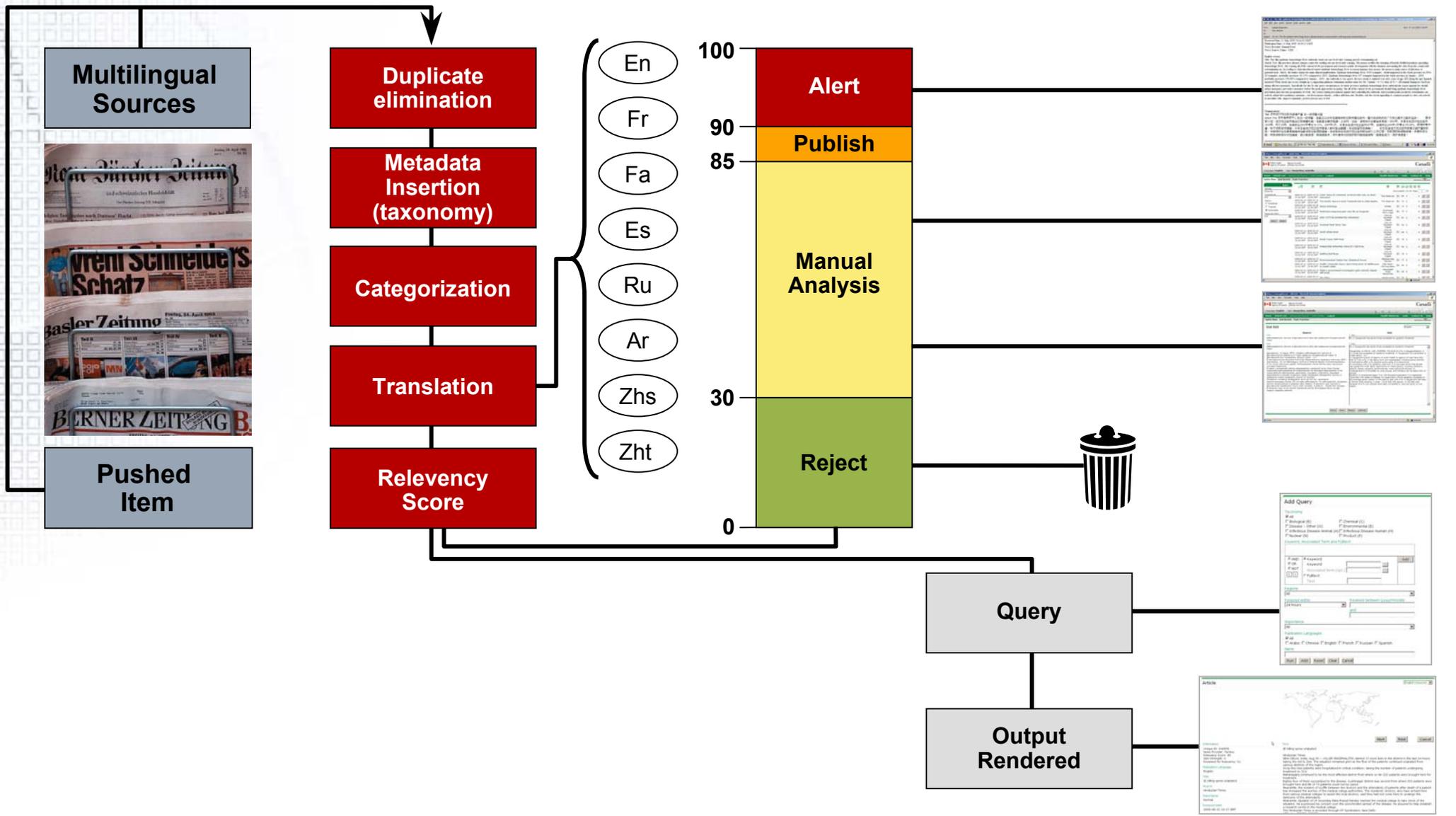


GPHIN Workflow

Data Collection

Advanced Analytics

Multilingual Real Time Intelligence Rendering





Machine Translation

- Gatekeeper / Analysts with topical expertise plus linguistic skills provide human ‘quick’ translation of ‘GISTS’ to improve ‘comprehensibility’
- “Best-of-Breed” approach
- Dictionaries / lexicons
 - Corpora
 - Taxonomy
 - Root term
 - Synonyms
 - Spelling variants of root terms and synonyms
- Meta Data – trans-lingual text mining



Gist Edit

Chinese

Source

Title

Aggressive visitors carry alarming buzz; The bees are known for their potential to swarm, then attack

Text

Aggressive visitors carry alarming buzz; The bees are known for their potential to swarm, then attack

MANATEE

They can deliver 1,000 stings in one minute. People who have died after being stung by them routinely have 40 stings per square inch on every bit of their exposed skin. And now the African honeybee - the killer bee made only slightly more monstrous by Hollywood - has been discovered in the Tampa area and could be in Manatee County, said a University of Florida entomologist who has studied them in Costa Rica and in their native Africa.

"It's hard to tell if they are in Manatee County and I wouldn't venture a guess," associate professor Glenn Hall said by phone Monday. "It will probably be a year or two if they aren't here already. They are getting established."

The killer bee is actually slightly smaller than the European bee, which is used by beekeepers in Manatee County for honey and crop pollination. The African bee also has 10 percent less venom, Hall said.

But what it loses in size, it makes up in numbers. "It is the large number of stings that cause the problem," Hall said. "We usually say 1,000 stings is a lethal dose and they can deliver that in one minute."

The African bee is exceptionally defensive. If its colony is approached, even by a passerby who means no harm, the bees have been known to pursue a fleeing subject for a quarter of a mile, Hall said.

Don't panic, Hall advises, even though his stories of the bee are alarming. "A European bee will pursue 50 to 100 feet and quit," Hall said.

Hall has been able to create genetic markers for the African bee. He can extract the DNA from dead bees to analyze if they are the variety with attitude.

Hall's advice to Manatee County: "Right now I wouldn't worry about it," he said. "We've had some stinging incidents (in Florida), but most have been with animals."

When African bees leave the ports in Florida, they have no choice but to mate

Gist

* Title

侵略性访客带着使人惊恐的吱吱声;蜂被为他们的潜力知道爬,那时攻击

* Text

侵略性访客带着使人惊恐的吱吱声;蜂被为他们的潜力知道爬,那时攻击

海牛

在一分钟中他们能送交1,000刺。已经随而过世,在被他们例行公事地刺痛之后的在的人们有每平方英寸的40刺在全部他们的暴露的皮上。

和现在非洲人蜜蜂经过好莱坞使-the非洲蜜蜂变得仅仅略微更可怕的已经被在坦帕区域中发现和能是在海牛县中,对说一位佛罗里达大学昆虫学家,其已经研究在哥斯达黎加和在他们的故乡非洲的他们。

星期一副教授Glenn霍尔通过电话对说"它是难以说如果他们是在海牛县中和我将冒失去一个猜想"的危险。它将大概是一年或者二,如果他们不已经在这里。他们正变得确立。

非洲蜜蜂事实上是与欧洲蜂,被在海牛县中的养蜂人为蜂蜜和作物传粉使用相比略微更小。霍尔对说非洲人蜂也让更少恶毒10分之一。

但是,它在数量上编造它在大小中失去的。

霍尔对说"正是许多的刺造成的问题"的。"我们通常对说1,000刺是一致死量和在一分钟"中他们能送交那个。

非洲人蜂特别是防御性。霍尔对说如果它的殖民地接近,甚至经过一个有害人的意思没有的过路人,蜂已经被知道一从消失科目是一英里的四分之一例追赶。

不确定惊恐不安,霍尔提供意见,尽管他的蜂的故事是使人惊恐的。

霍尔对说"一只欧洲蜂将追赶50向100脚和离开"。

霍尔已经能建立为非洲人蜂基因标志。他能从死了蜂抽取脱氧核糖核酸分析,如果他们标新立异本来打算种类的。

霍尔的向海牛县的忠告:他对说"现在我不将担心它"。"我们已经中的一些(在佛罗里达)的刺痛的事件但是绝大部分已经是和动物"。

霍尔对说当非洲人的时候,蜂把港口遗留在佛罗里达,他们不是精选让但是对有欧洲蜂的是一好事情的配偶。

霍尔对说中最好的控制蜂的希望在于本地养蜂人,其能生产这样多的雄性欧洲雄蜂使更平静后代和他们生产的非洲人蜂交配。

霍尔对说我们确实有关关注那蜂人口可以达到一非洲人蜂在那里发现彼此和彼此繁殖多产的临界质量

Save View Reset Cancel



GPHIN News Feeds: unvalidated, publicly available

- Poetic licence or polysemic terms
- Deliberate / accidental mis-use of words
- Mis-spellings
- Local vocabulary
- Style of prose
- Presence / absence of diacritics
- Abbreviations
- Spacing of letters of a word
- Capitalization
- Chinese grammar rules
- Quirks & twists



Poetic licence or polysemic terms

- **Example #1**

- “No beef in cattle bailout”
- “Aucun boeuf dans le saut du bétail en parachute”

- **Example #2**

- “Injured bats in rabies alert”
- “Bates heridos en alert de rabia” vs “Murcielago heridos en alerta de rabia”



Poetic licence or polysemic terms cont'd

- **Example #3**

- ... after drinking contaminated water in southwestern Turkey
- ... après l'eau contaminée de boisson dans la Dinde ...

- **Example #4**

- Officer takes a spill as Bush takes birthday spin
- Prises de l'officier qu'un déversement comme l'anniversaire des prises du Buisson tourne



Deliberate / accidental mis-use of words

- **Example #1**

- “China pesa demasiado” = “China weighs too much”

- **Example #2**

- “Aggressive visitors carry alarming buzz ...”

- **Example #3**

- **تالاح رومظ يلودلا ةئبوالا بتكم عم دصرت ةعارزلا** = “**The agriculture** with the international ...” (the ‘Ministry of Agriculture’ is implied)



Mis-spellings

- **Example #1**

- “Global News Bites” = “Las Noticias Globales Muerden” vs
“Global News Bits” = “Las Noticias Globales Pulsaciones ”

- **Example #2**

- “Penguin Sex Deaths”



Local Vocabulary (slang)

- **Example #1**

- “anthrax”, proper Russian term = сибирская язва
- Local term = сибирк*

- **Example #2**

- “Time for cha” vs “Time for tea”

- **Example #3**

- Proper term: Tuberculosis-----
- Local term: Tuberculosis-----
- Proper term: Chickenpox-----
- Local term: Chickenpox-----



Presence / absence of diacritics

- **Example #1**

- tâche = task
- tache = stain, speck

- **Example #2**

- When an apostrophe replaces a letter and the next letter has a diacritic:
- “.. à l'énergie atomique ..” = “.. at l 'atomic energy ..”



Abbreviations

- **Example #1**

- РОССЕЛЬХОЗНАДЗОР ГАРАНТИРУЕТ БЕЗОПАСНОСТЬ ..

- translates to: ÐÎÑÑÅËÛÕÎÇÍÀÄÇÎÐ

- GUARANTEES SAFETY ..

- РОССЕЛЬХОЗНАДЗОР is the Russian Agriculture Control Authority

- **Example #2**

- .. почти до 2 **тыс (thousand)** гектаров .. = .. up to 2 **òûñ** hectares ..



Spaces within a word

- **Example #1**

- “С Р О Ч Н О Министры стран ...” where
- “С Р О Ч Н О” (with spaces between letters should be “U R G E N T” but translates to:
- “WITH R ABOUT CH N ABOUT Ministers of the countries ...”

- **Example #2**

- □□□□□□□□□□□□ (13 May) =
“ .. Shijiazhuang, **May 1** XinHua ..”



Capitalization (in place names)

- **Example #1**

- АРМЕНИЯ (Armenia) = Ἀρμενία

- **Example #2**

- ЕРЕВАН (Yerevan, capital of Armenia) = Ἐρεβάν



Chinese grammar rules

Public Health Agency of Canada / Agence de santé publique du Canada

Canada

Language: English User: Blench, Michael

Global Public Health Intelligence Network

Home | Article List | Advanced Search | Edit Profile | Logout

Health Ministries | Links | Contact Us | Help

Gphin Main | Gist Bucket | Push Function

EMPOWERED BY Stein

Gist Edit

Chinese

Source	Gist
<p>Title</p> <p>Aggressive visitors carry alarming buzz; The bees are known for their potential to swarm, then attack</p>	<p>* Title</p> <p>侵略性访客带着使人惊恐的吱吱声;蜂被为他们的潜力知道爬,那时攻击</p>
<p>Text</p> <p>Aggressive visitors carry alarming buzz; The bees have their potential to swarm, then attack</p> <p>MANATEE</p> <p>They can deliver 1,000 stings in one minute. People who have died after being stung by them routinely have 40 stings per square inch on every bit of their exposed skin. And now the African honeybee - the killer bee made only slightly more monstrous by Hollywood - has been discovered in the Tampa area and could be in Manatee County, said a University of Florida entomologist who has studied them in Costa Rica and in their native Africa. "It's hard to tell if they are in Manatee County and I wouldn't venture a guess," associate professor Glenn Hall said by phone Monday. "It will probably be a year or two if they aren't here already. They are getting established." The killer bee is actually slightly smaller than the European bee, which is used by beekeepers in Manatee County for honey and crop pollination. The African bee also has 10 percent less venom, Hall said. But what it loses in size, it makes up in numbers. "It is the large number of stings that is the problem," Hall said. "We usually say 1,000 stings is a lethal dose and that is true in one minute." The African bee is exceptionally defensive. When a colony is approached, even by a passerby who means no harm, the bees are known to pursue a fleeing subject for a quarter of a mile, Hall said. Don't panic, Hall advises, even though his stories of the bee are alarming. "A European bee will pursue 50 to 100 feet and quit," Hall said. Hall has been able to create genetic markers for the African bee. He can extract the DNA from dead bees to analyze if they are the variety with attitude. Hall's advice to Manatee County: "Right now I wouldn't worry about it," he said. "We've had some stinging incidents (in Florida), but most have been with animals." When African bees leave the ports in Florida, they have no choice but to mate</p>	<p>* Text</p> <p>侵略性访客带着使人惊恐的吱吱声;蜂被为他们的潜力知道爬,那时攻击</p> <p>海牛</p> <p>在一分钟中他们能送交1,000刺。已经随而过世,在被他们例行公事地刺痛之后的在的人们有每平方英寸的40刺在全部他们的暴露的皮上。</p> <p>和现在非洲人蜜蜂经过好莱坞使-the非洲蜜蜂变得仅仅略微更可怕的已经被在坦帕区域中发现和能是在海牛县中,对说一位佛罗里达大学昆虫学家,其已经研究在哥斯达黎加和在他们的故乡非洲的他们。</p> <p>星期一副教授Glenn霍尔通过电话对说"它是难以说如果他们是在海牛县中和我不将冒失去一个猜想"的危险,它将大概是一年或者二,如果他们不已经在这里,他们正变得确立。非洲蜜蜂事实上是与欧洲蜂,被在海牛县中的养蜂人为蜂蜜和作物传粉使用相比略微更小。霍尔对说非洲人蜂也让更少恶毒10百分之一。</p> <p>但是,它在数量上编造它在大小中失去的。霍尔对说"正是许多的刺造成的问题"的。"我们通常对说1,000刺是一致死量和在一分钟"中他们能送交那个。</p> <p>非洲人蜂特别是防御性。霍尔对说如果它的殖民地接近,甚至经过一个有害人的意思没有的过路人,蜂已经被知道一从消失科目是一英里的四分之一例追赶。不确实惊恐不安,霍尔提供意见,尽管他的蜂的故事是使人惊恐的。霍尔对说"一只欧洲蜂将追赶50到100脚和离开"。</p> <p>霍尔已经能建立为非洲人蜂基因标志。他能从死了蜂摘取脱氧核糖核酸分析,如果他们标新立异本来打算种类的。</p> <p>霍尔的向海牛县的忠告:他对说"现在我不将担心它"。"我们已经中的一些(在佛罗里达)的刺痛的事件但是绝大部分已经是和动物"。</p> <p>霍尔对说当非洲人的时候,蜂把港口遗留在佛罗里达,他们不是精选但是对欧洲蜂的是一好事情的配偶。</p> <p>霍尔对说中最好的控制蜂的希望在于本地养蜂人,其能生产这样多的雄性欧洲雄蜂使更平静后代和他们生产的非洲人蜂交配。</p> <p>霍尔对说我们确实有关关注那蜂人口可以达到一非洲人蜂在那里发现彼此和彼此繁殖多达的临界质量</p>

Save View Reset Cancel



Quirks and Twists

- Example #1
 - AIDS: ----□□□ ----- “love growing disease”
- Example #2
 - هديده Hodeida: A city in Yemen
 - هديده Hadida: Iron in Arabic
 - The reader has to understand the sense of the phrase to decide how to pronounce the word
- Example #3
 - نيب اصملا فاعسا يف قمزاو قلوه جمل اى محل اب ةافو **اتلاح**
translates to:
Death **cases** by the unknown fever .. When it should translate to **TWO cases**



Integration Issues

- Awkward API's
- Poor documentation
- Lack of standards for dictionaries / lexicons
- Bugs
- Clashes between different manufacturers



Operational Issues

- Inherently unstable code
- Bugs
- Crashes
- Unpredictable performance



nTranslator™

- Overcomes some of the integration and operational issues:
 - Normalizes API's
 - Detects 'engine' crashes and re-boots
 - Overcomes in-compatibilities
 - Solves display / rendering problems
 - Converts file formats



nTranslator™

nTranslator™

English ↔
Farsi

English ↔
French

English ↔
Spanish

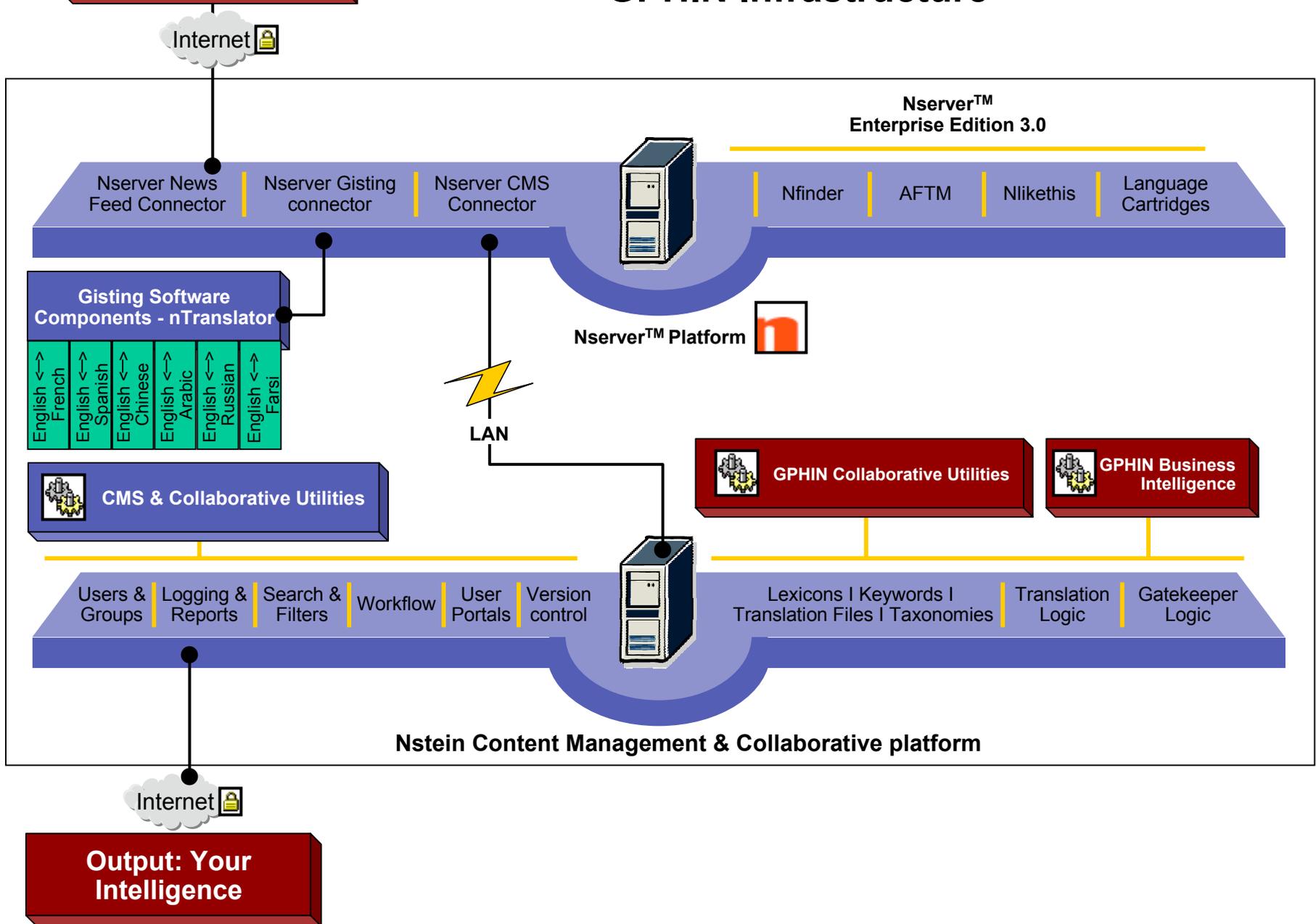
English ↔
Chinese

English ↔
Arabic

English ↔
Russian

GPHIN Data Sources

GPHIN Infrastructure





Recommendations

- Translation memory
- Transliteration
- Tighter code to improve speed and reduce cycles
- Standardization



Public Health
Agency of Canada

Agence de santé
publique du Canada



Questions ??



Contact Information

- Ms. Abla Mawudeku, Manager, GPHIN
(613) 957-2715
Abla_Mawudeku@phac-aspc.gc.ca
- Michael Blench, GPHIN Technical Advisor
(613) 957-3567
Michael_Blench@phac-aspc.gc.ca
- Ms Isabelle Desjardins, GPHIN Account Manager
Nstein Technologies Inc.
(514) 908-5406
isabelle.desjardins@nstein.com