PUBLIC HEALTH AGENCY *of* CANADA
AGENCE DE SANTÉ PUBLIQUE *du* CANADA

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

Canada

Global Public Health Intelligence Network (GPHIN)

AMTA
August 2006

Michael Blench
GPHIN Technical Advisor and Project Coordinator

Global Public Health Intelligence Network (GPHIN)

This presentation will discuss:

1. What is GPHIN?
2. How does it work?
3. Effectiveness
4. Workflow
5. Machine Translation

What is GPHIN?

- A 'solution' comprised of a team of analysts directing a very innovative and unique technology.

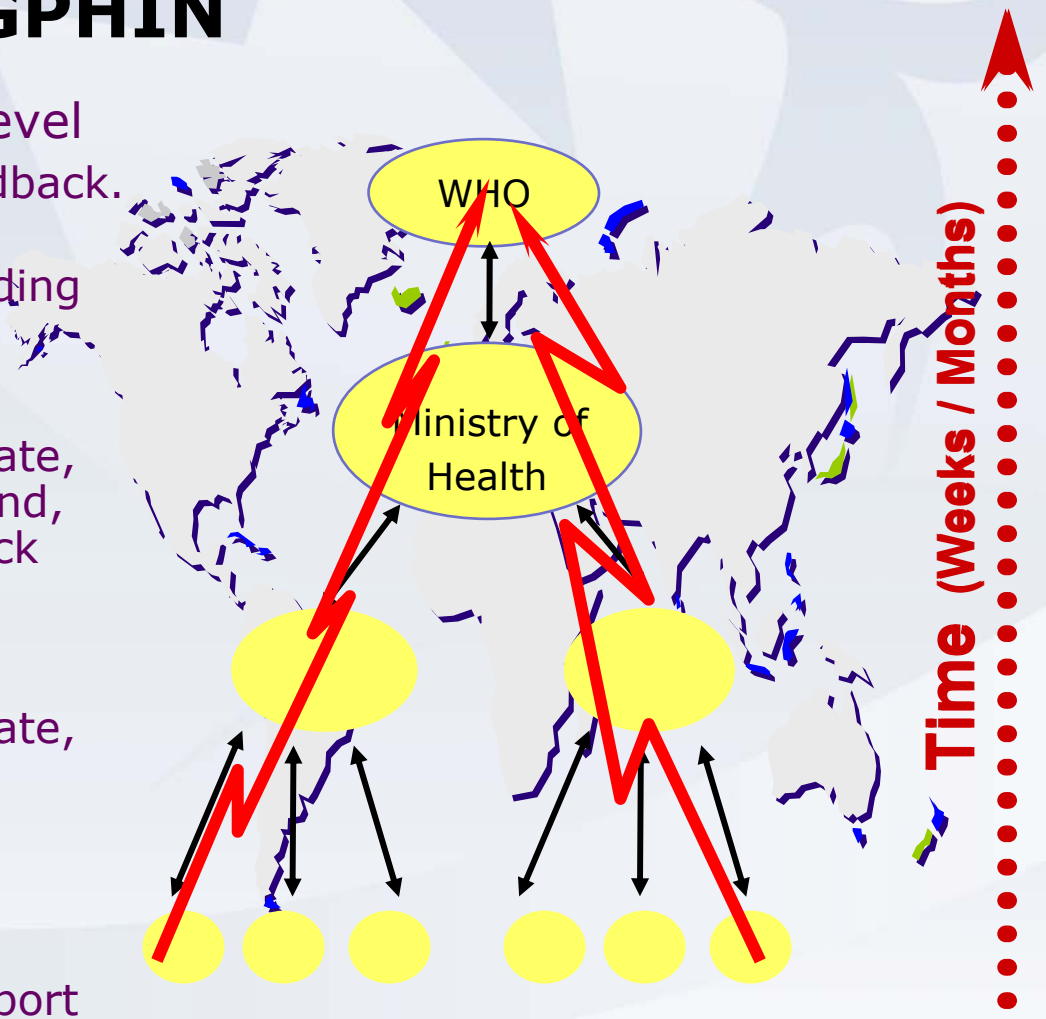
- The 'Team':
  - Multi-ethnic
  - Multi-cultural
  - Multi-lingual
  - Multi-disciplined
  - Synergy

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'.

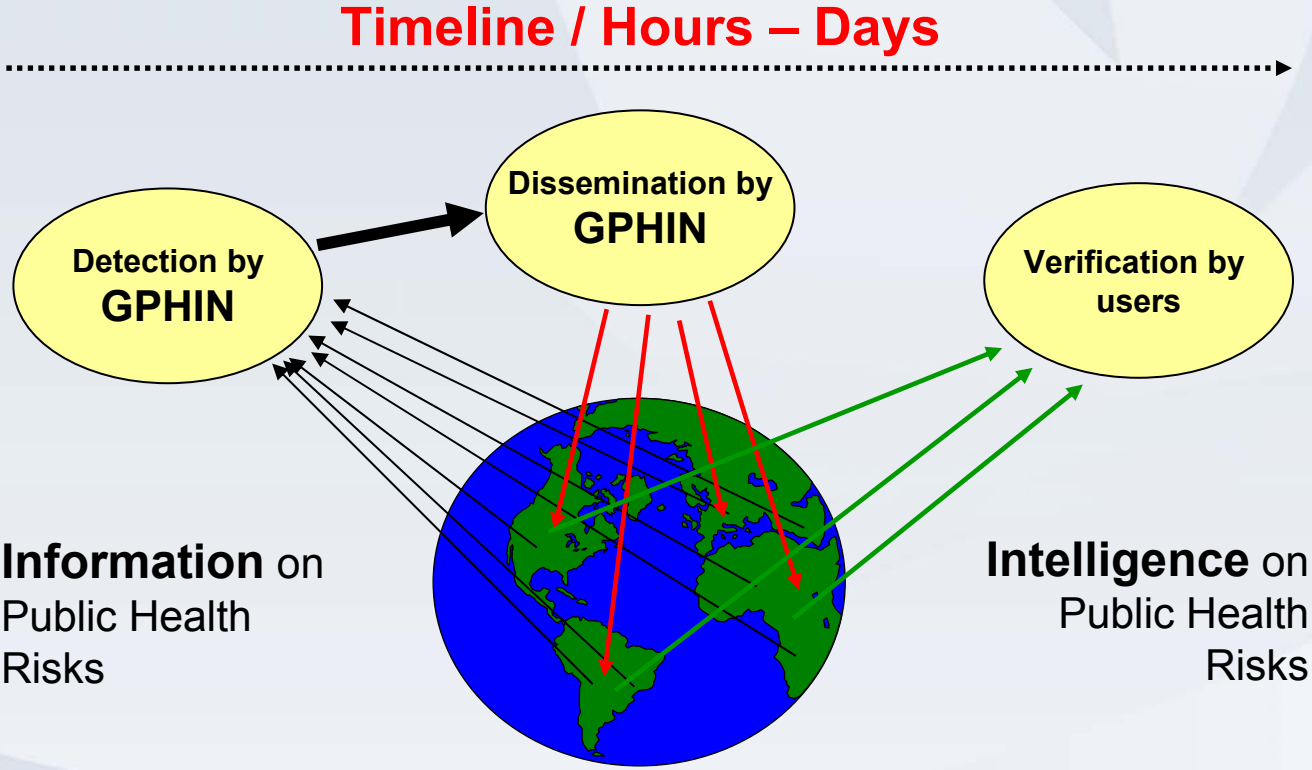
Traditional PH Surveillance Paradigm and GPHIN

- 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

GPHIN's Modus Operandi



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

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
- ☑ Unsafe products – e.g. Prozac, NSAID
- ☑ Natural Disasters – e.g. tsunami, 26 Dec 04

Public Health Organizations at the International Level

- WHO – World Health Organisation
- PHAC – Public Health Agency of Canada
- CDC – Center for Disease Control
- Others

Background

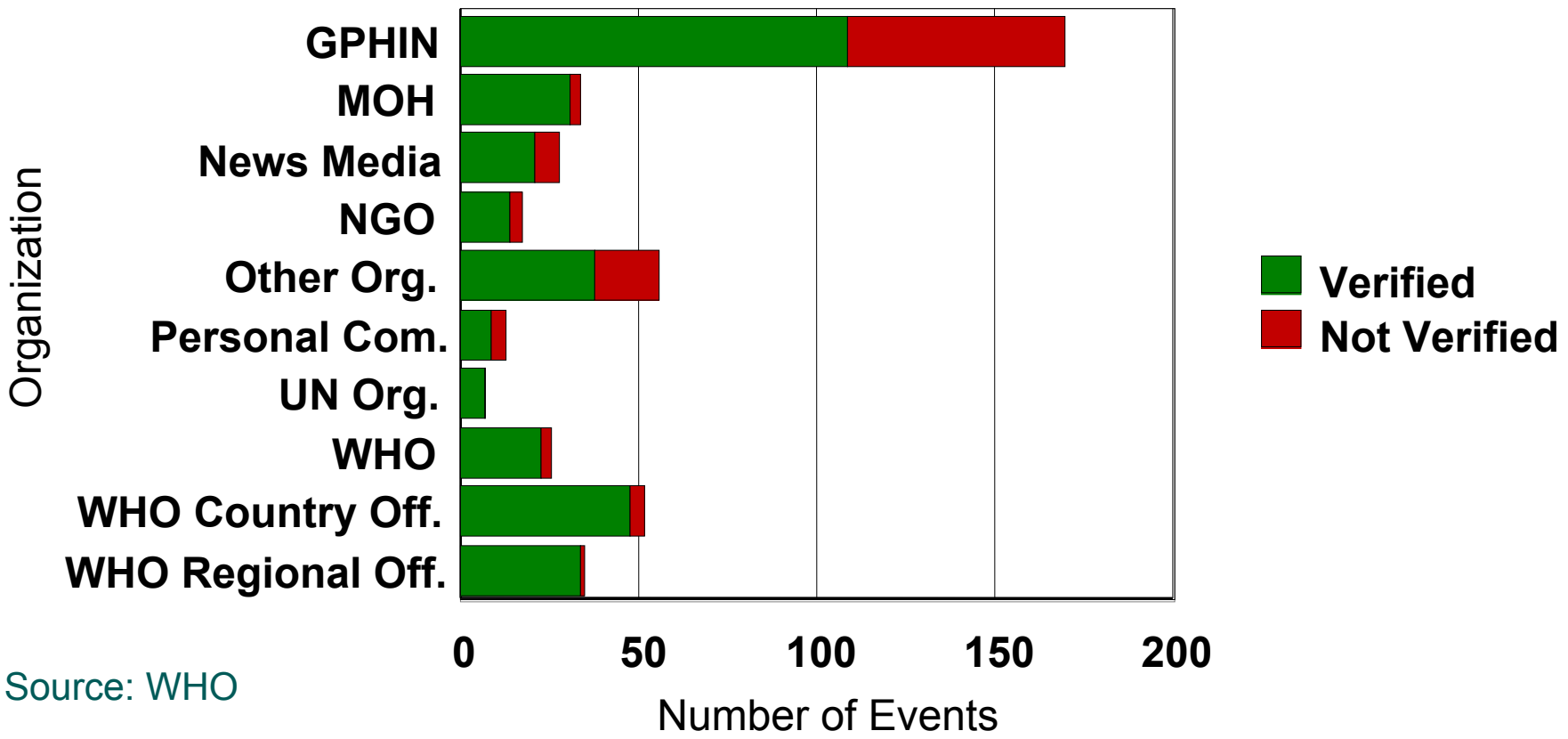
- GPHIN's Role and management
  - GPHIN was developed by the Public Health Agency of Canada
  - Managed by the Agency's Centre for Emergency Preparedness and Response
- GPHIN's history
  - 1997 – Idea of two Health Canada physicians
  - 1998/99 – Prototype developed by Health Canada in partnership with 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

Effectiveness evaluation criteria

- **Usefulness** (specificity): reports' value
- **Timeliness**: speed reports are available
- **Sensitivity**: relevancy of reports
- **Flexibility**: ease to make modifications
- **Stability** (robustness): downtime, staffing
- **Cost**: sustainability

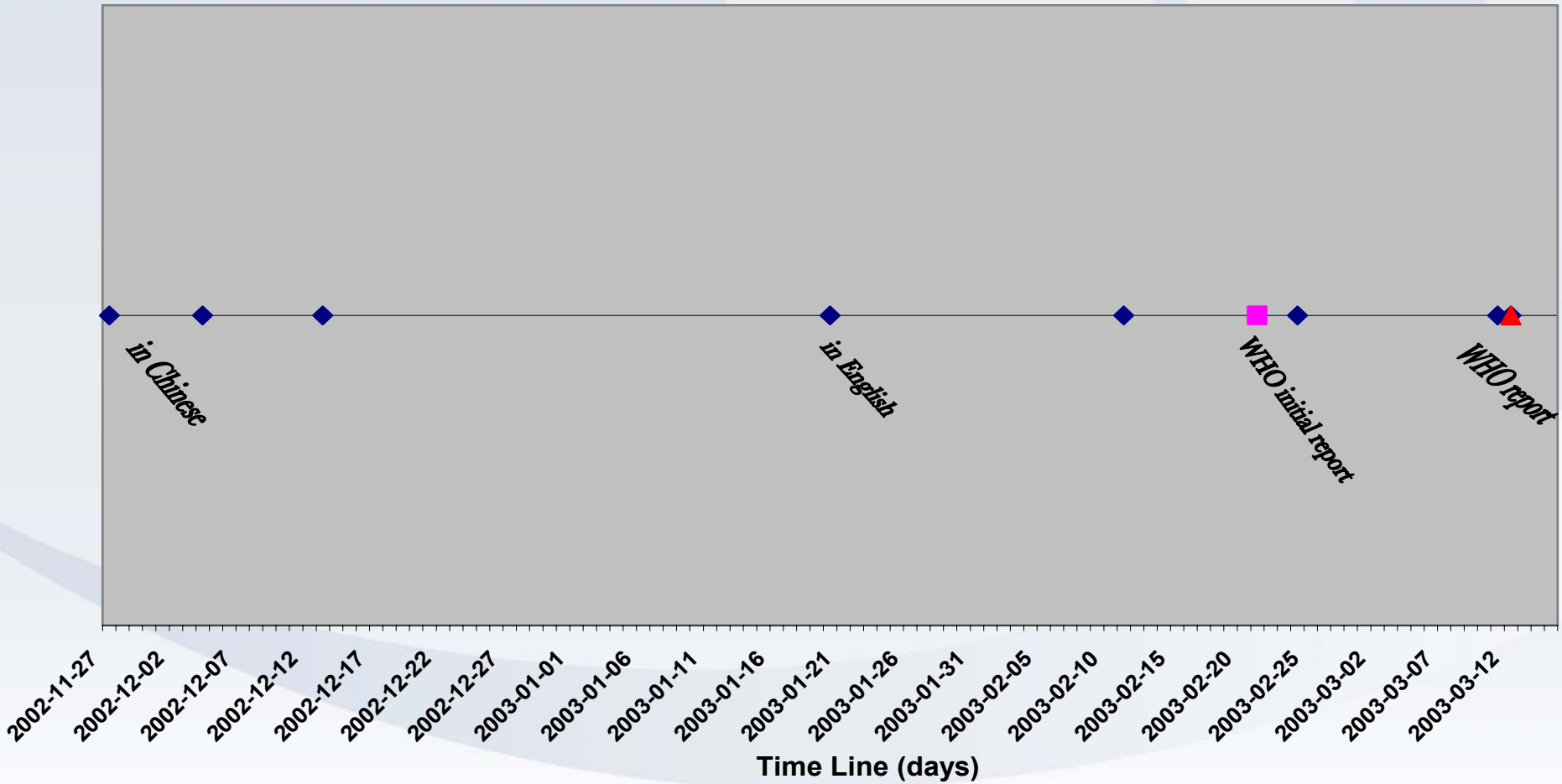
Source of Initial Reporting of WHO Reported Events of Potential PH Concern

01 Jan 2001 to 31 Dec 2002. (n=439)

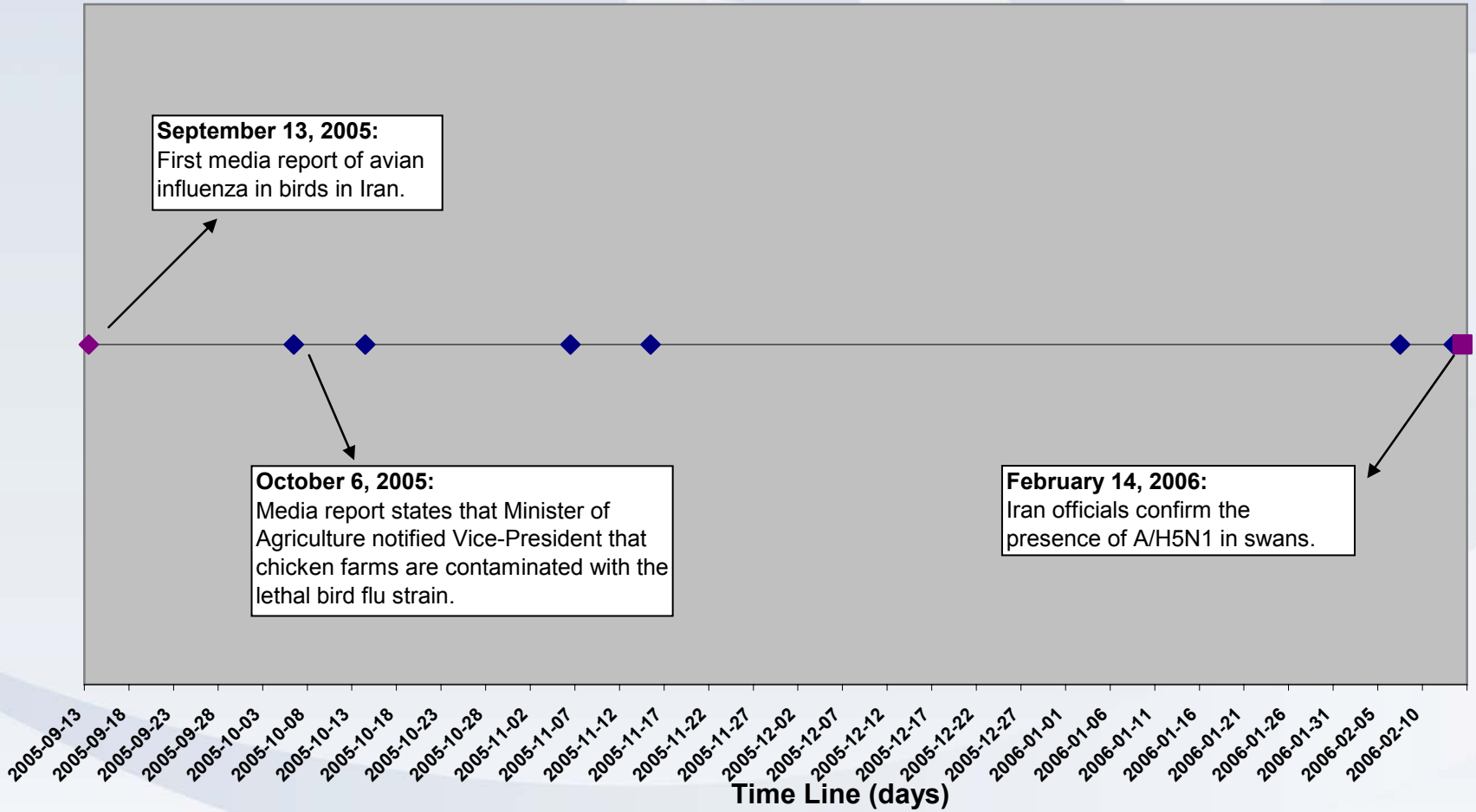


Source: WHO

GPHIN's Reporting Suggestive of the Early Detection of an Unusual Event (SARS?) in China



GPHIN's Early Reporting of the Presence of A/H5N1 Virus in Iran

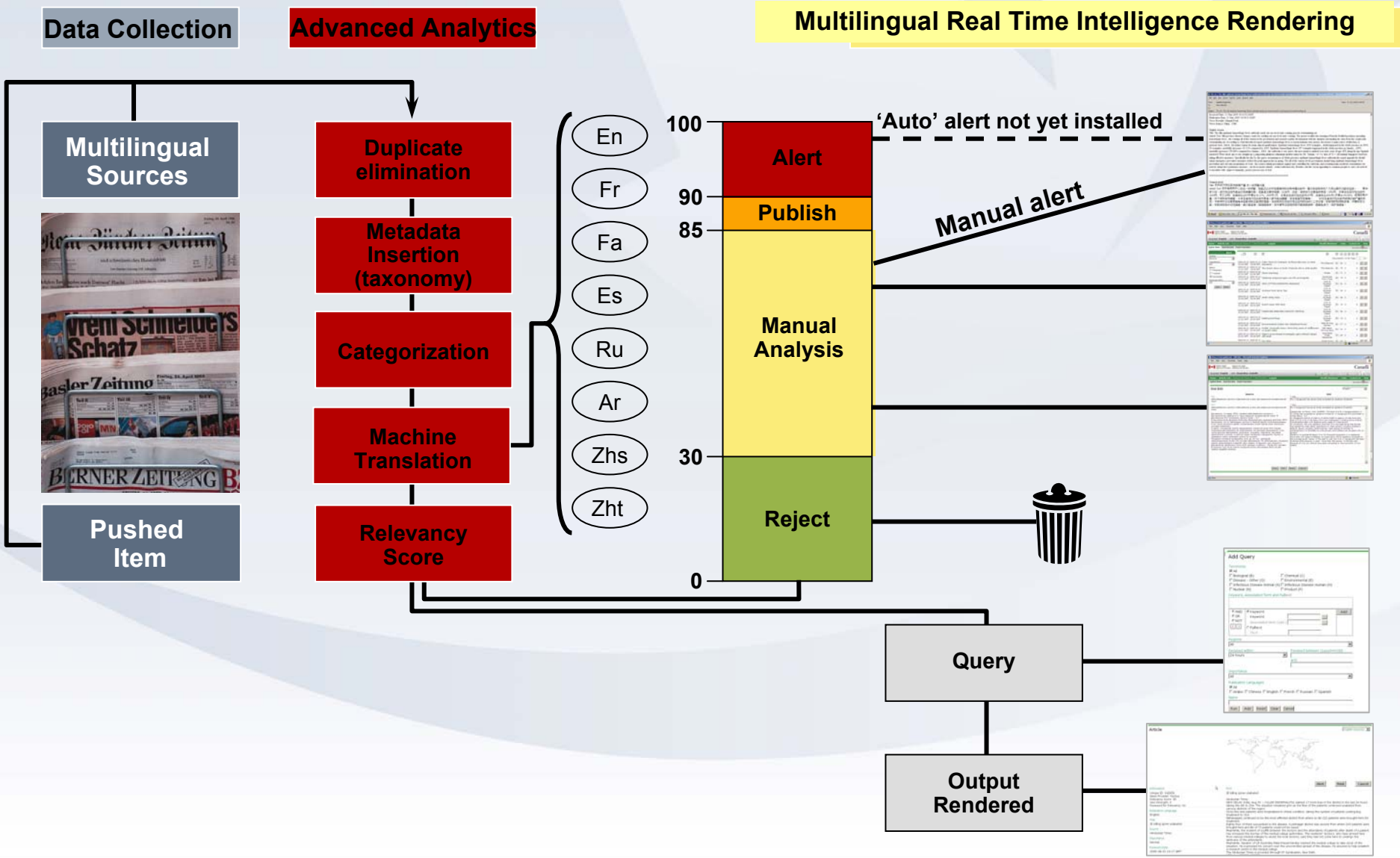


- **Initial report about the presence of avian influenza in birds in Iran - not confirmed by Iranian officials**
- **Media reports about the presence of avian influenza in birds in Iran - not confirmed by Iranian officials**
- **Media report about the presence of avian influenza in birds in Iran - confirmed by Iranian officials**

GPHIN's Versatility – Information for Action

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

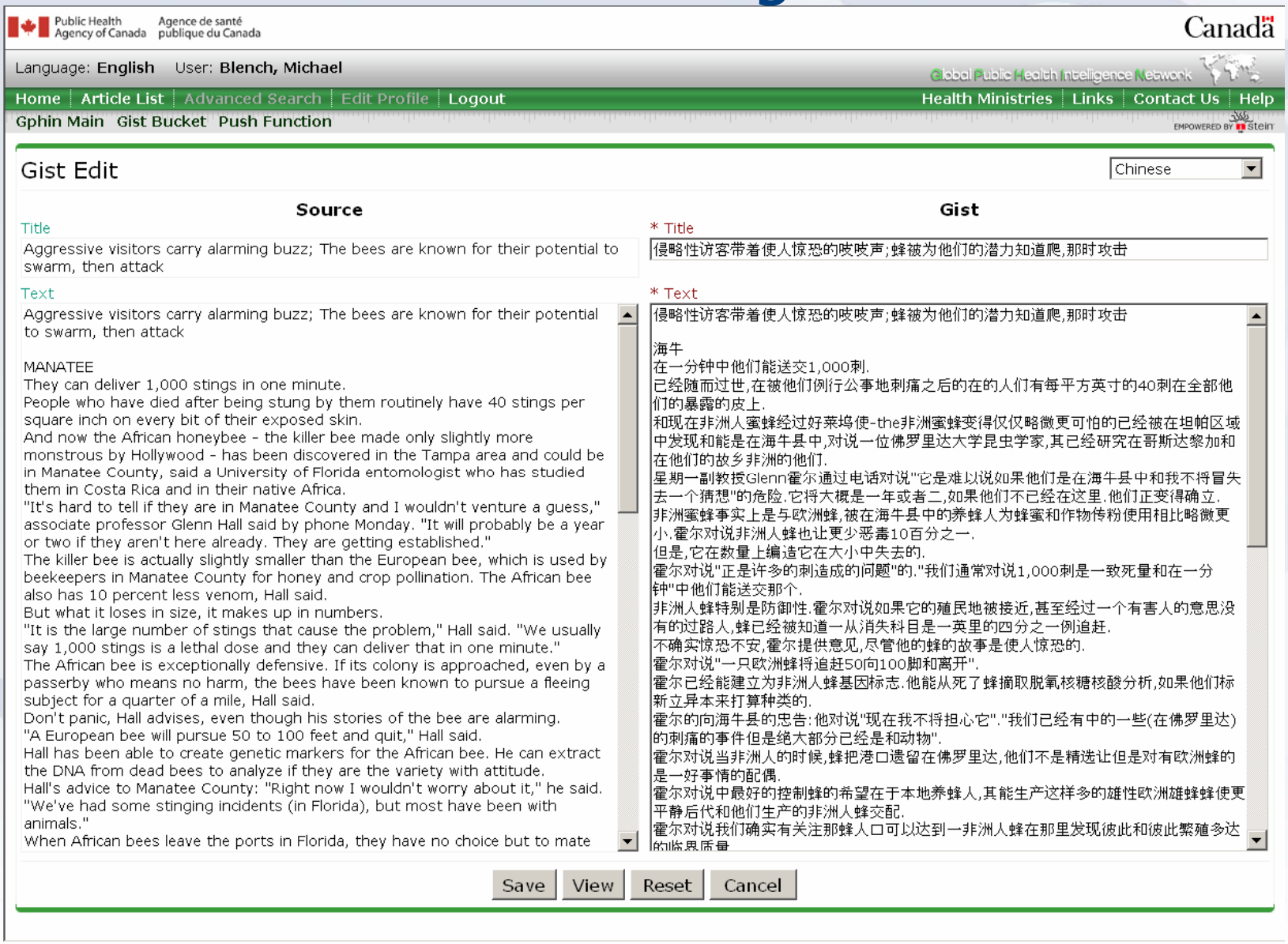
GPHIN Workflow



Machine Translation

- Gatekeeper / Analysts with topical expertise plus linguistic skills provide 'quick' post editing 'GISTS' to improve 'comprehensibility'
- "Best-of-Breed" – currently 5 MT engines
- Dictionaries / lexicons
  - Corpora
  - Taxonomy
    - Root term, synonyms
    - Spelling variants of root terms and synonyms
    - Acronyms, abbrev., colloquial terms
- Meta Data – trans-lingual text mining

GPHIN's Post Editing Mechanism



GPHIN News Feeds: unverified, 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 (French)
  - "No beef in cattle bailout"
  - "Aucun boeuf dans le saut du bétail en parachute"
- Example #2 (Spanish)
  - "Injured bats in rabies alert"
  - "Bates heridos en alert de rabia" vs "Murcielagos heridos en alerta de rabia"

Poetic licence or polysemic terms cont'd

- Example #3 (Farsi):
  - دختر۱۲ ساله به دست برادرش پرپر شد
  - "A 12 year old year girl was 'plucked' by her brother"
  - Writer meant: "A 12 year old girl was killed by her brother"
- Example #4 (Farsi):
  - دو قلم مواد بهداشتي در سمنان غيرقابل مصرف اعلام شد
  - 2 pens of health products are prohibited for usage in Semnan.
  - 2 kinds of health products are prohibited for usage in Semnan.

Deliberate / accidental mis-use of words

- Example #1
  - "China pesa demasiado" = "China weighs too much"
- Example #2
  - "Aggressive visitors carry alarming buzz …"
- Example #3 (Arabic)
  - الزراعة ترصد مع مكتب الاوبئة الدولي ظهور حالات جنون البقر = "**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-----肺結核
    Local term: Tuberculosis------癆 病
  - Proper term: Chickenpox--------水痘
    Local term: Chickenpox-------禽痘

Style of prose

Bad Example:

- **疑他人举报投毒泄愤夜黑风大伤及大量“无辜”国有林场大面积绿草 幼树被毒死**

Translates to:

- The child sets up green grass of the night reshahar doubting that others reports on a poisoning giving vent to self's anger is hurt heavily and in large amount "innocent" state-owned forest farm large area being killed with poison

Style of prose, cont'd

Good Example:

- 据悉，本田ＦＣＸ符合美国政府所有的安全标准。它的驾驶方式和普通汽车一样，噪音比普通车小，燃料箱里装的是压缩氢气而不是汽油，排放完全是水，加满氢气一次能跑３００公里。

Translates to:

- By report, Honda FCX accords with all USA government safety standards. It Driving way is same as average automobile , noise is smaller than ordinary vehicle, in the fuel cell being pretending compresses a hydrogen but being not that the gasoline , blowoff are complete is water, Jiaman hydrogen one. Can run 300 kilometres time.

Style of prose, cont'd

Yellow Fever
18 November 2003
Tampa Tribune

An **epidemic** of ***penalties*** have thwarted many drives, resulting in a three-***game*** losing streak and essentially leaving the Bucs' season on **life support**.

Yellow fever grips Oasis
24 February 2006
Kalgoorlie Miner

AN **OUTBREAK** of **yellow fever struck** the Oasis last Friday when John Paul College held its annual ***swimming*** carnival.

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
  - РОС**СЕЛЬХОЗ**НАДЗОР ГАРАНТИРУЕТ БЕЗОПАСНОСТЬ ..
    translates to: ÐÎÑÑÅËÜÕÎÇÍÀÄÇÎÐ 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
  - 新华社信息石家庄 5 月 1 3 日电 (13 May) = " .. Shijiazhuang, **May 1** XinHua .."

Capitalization (in place names)

- Example #1
  - АРМЕНИЯ (Armenia) = ÀÐÌÅÍÈß
- Example #2
  - ЕРЕВАН (Yerevan, capital of Armenia) = ÅÐÅÂÀÍ

Chinese grammar rules

Quirks and Twists

- Example #1
  - AIDS: ----愛滋病 ------ “love growing disease”
- Example #2
  - حديدة Hodeida: A city in Yemen
  - حديدة Hadida: Iron in Arabic
    The reader has to understand the sense of the phrase to decide how to pronounce the word
- Example #3
  - **حالتا** وفاة بالحمى المجهولة وأزمة في اسعاف المصابين
    translates to:
    Death **cases** by the unknown fever .. When it should translate to **TWO cases**

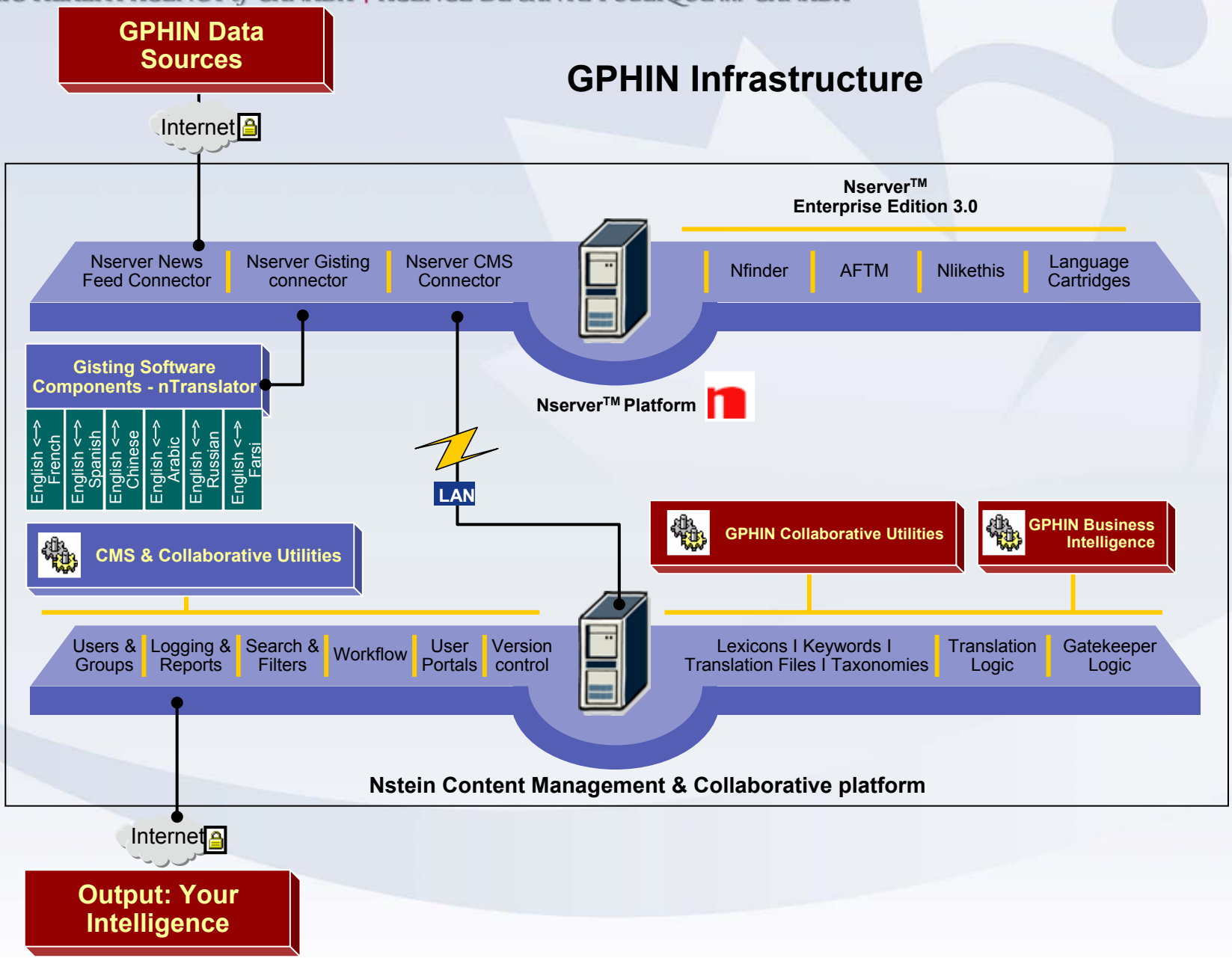
Integration Issues

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

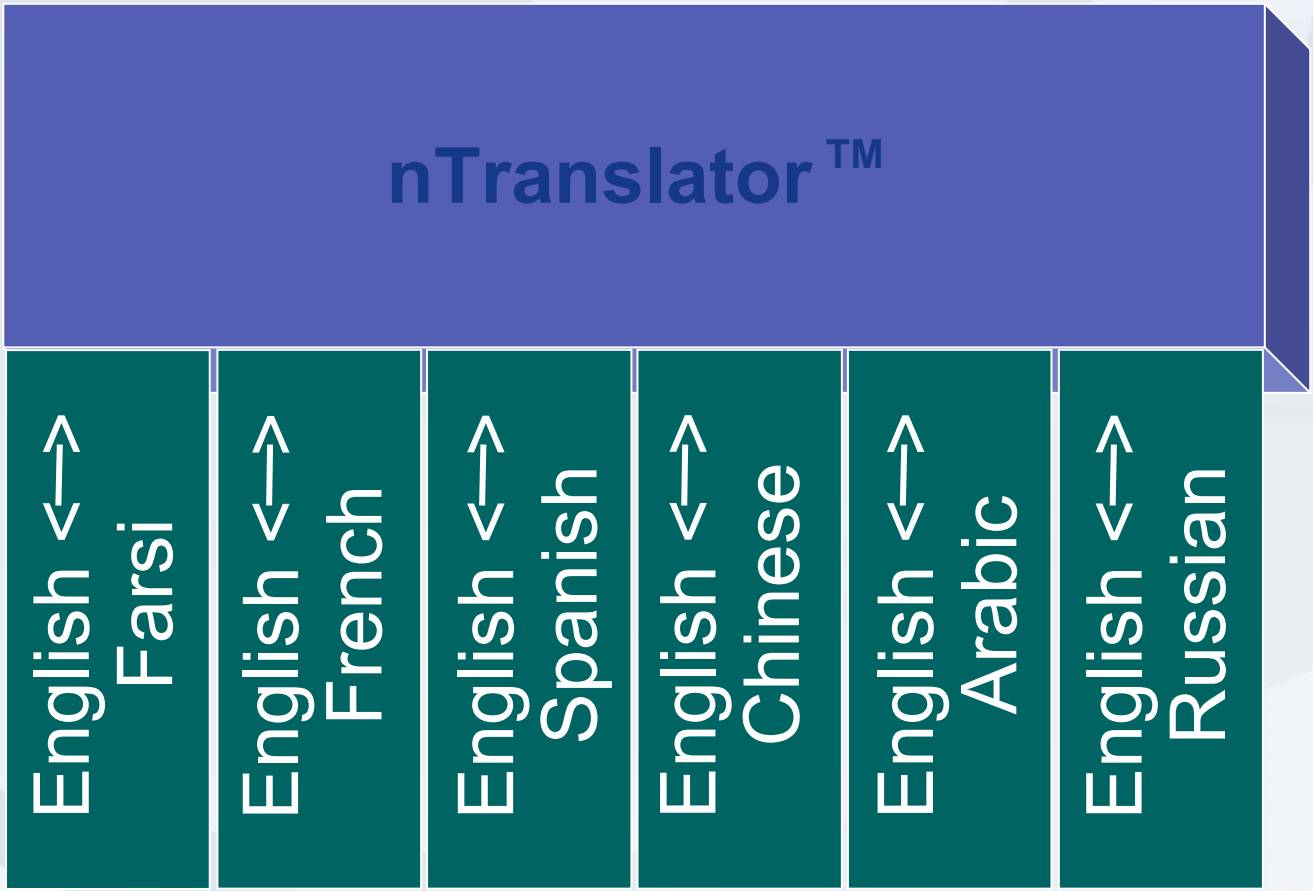
Operational Issues

- Inherently unstable code
- Bugs
- Crashes
- Unpredictable performance

GPHIN Infrastructure



nTranslator ™



nTranslator ™

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

Recommendations

- Translation memory
- Transliteration
- Tighter code to improve hardware efficacy
- Standardization
- Hybridization

Questions ??