



insight into value

PATIENT-GENERATED HEALTH DATA (SOCIAL MEDIA) IS A POTENTIAL SOURCE FOR ADVERSE DRUG REACTION REPORTING

PhUSE 2017 – Paper DH01

KEYRUS
biopharma

AGENDA

Context

- Context
- Goal
- Sources

Proposed solution

- Process flow
- Extraction
- Coding
- Storage

Results & Discussion

- Results
- Discussion



AGENDA

Context

- Context
- Goal
- Sources

Proposed solution

- Process flow
- Extraction
- Coding
- Storage

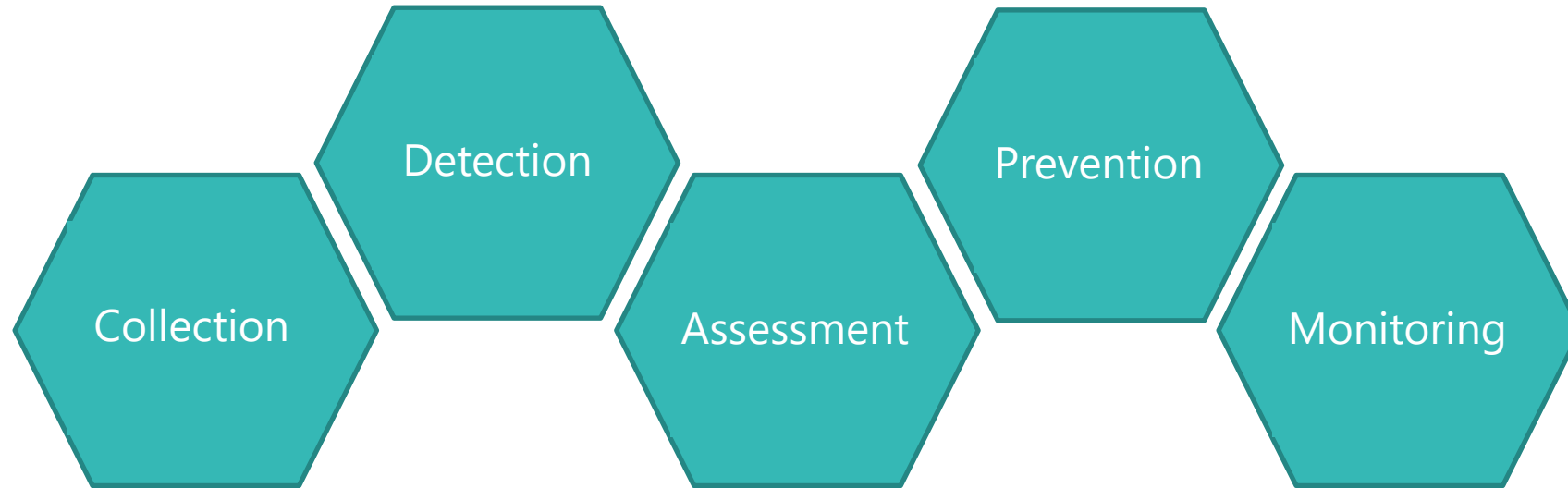
Results & Discussion

- Results
- Discussion



CONTEXT - PHARMACOVIGILANCE

Pharmacovigilance, the science related to the



of *Adverse Drug Reactions* (ADRs) with pharmaceutical products.

How to use social media as a potential data source of adverse events ?

CONTEXT - EXISTING RESEARCH

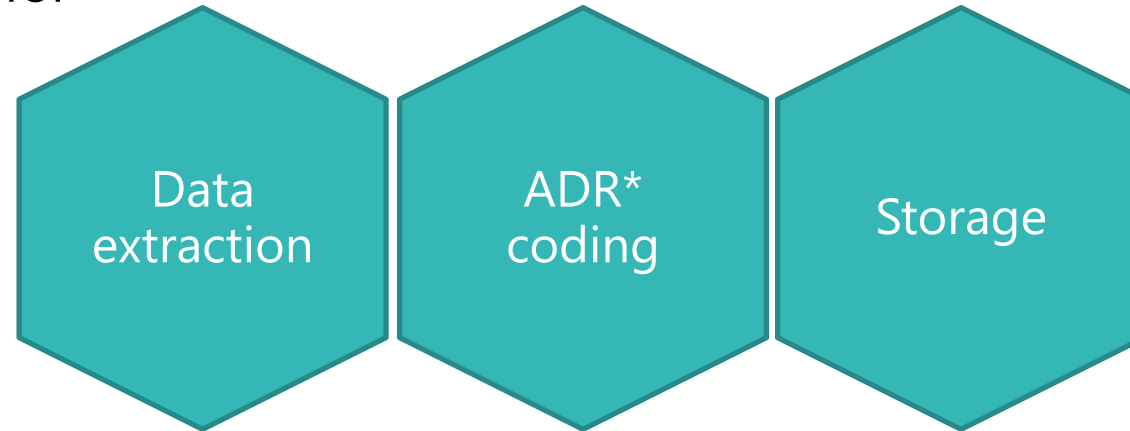
Method	Data	Limitations
Lexicon-based	Twitter®	Drug representation, Time consuming
Association rule-mining	DailyStrength®	Time consuming
Machine learning	Twitter®, DailyStrength®	Normalization



<http://creofire.com/spending-scientific-research/>

GOAL

An *automated* solution for



*Adverse Drug Reaction

Based on two sources:

- **Twitter®**, a micro-blogging social network



- **Reddit®**, a widely used forum



The solution should be compatible to various sources.

SOURCES



- 12th **most used** website in the **world**
- Used to send short messages (140 characters)
- Used to **share** news, pictures and **everyday life**



- 9th **most used** website in the **world**
- Used for online discussion
- Contains various **sections** from video games to **drugs usage**

Both source have mention potential ADRs

AGENDA

Context

- Context
- Goal
- Sources

Proposed solution

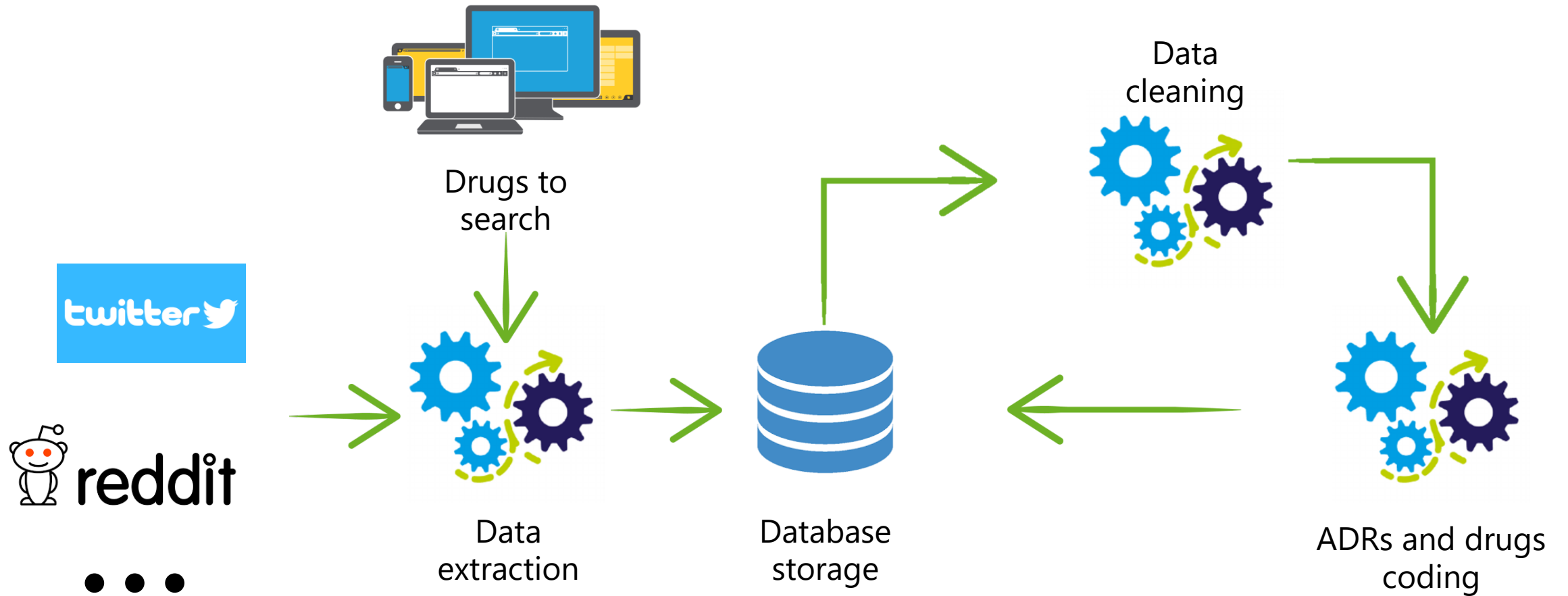
- Process flow
- Extraction
- Coding
- Storage

Results & Discussion

- Results
- Discussion



PROCESS FLOW



EXTRACTION

Twitter® & Twitter API:

- **Freely** available
- **Easy** to use
- Return **structured** JSON

Reddit® :

- **No tool** for the needed data extraction
- **HTML** structure **harder** to analyze

```
{
  "date": datetime.datetime(2017, 6, 19,
15, 57, 22),
  "retweeted": False,
  "text": u"I took Gabapentin for my bach
pain and now I have nausea...",
  "user_id": 704703338261188608,
  "user_location": u"USA & Switzerland",
  "user_name": u"erwanlc",
  "drug": u"Gabapentin"
}
```

```
43 <body <?php Body class="
44 <div id="fb-root"></div>
45 <script>(function(d, s, id) {
46   var js, fjs = d.getElementsByTagName(s)[0];
47   if (d.getElementById(id)) return;
48   js = d.createElement(s); js.id = id;
49   js.src = "//connect.facebook.net/en_US/sdk.js#xfbml=1&version=v2.6&appId=286644436211";
50   fjs.parentNode.insertBefore(js, fjs);
51   }(document, 'script', 'facebook-jssdk'));</script>
52 <div id="page" class="site">
53   <a class="skip-link screen-reader-text" href="#content"><?php esc_html_e('Skip to content', 'wordpress'); ?></a>
54   <header id="masthead" class="site-header" role="banner">
55     <div class="site-branding">
56       <div class="nav-btn pull-left">
57         <?php if(is_home()) && $xpanel['homepage-style'] == 1 { ?>
58           <a href="#" id="openMenu"><i class="fa fa-bars fa-3x"></i></a>
59           <?php } else { ?>
60             <a href="#" id="openMenu2"><i class="fa fa-bars fa-3x"></i></a>
61             <?php } ?>
62         </div>
63         <div class="logo pull-left">
64           <a href="<?php echo esc_url( home_url() ) ?>">
65             
67         </div>
68         <div class="search-box hidden-xs hidden-sm pull-left ml-10">
69           <?php get_search_form(); ?>
70         </div>
71         <div class="submit-btn hidden-xs hidden-sm pull-left ml-10">
72           <a href="<?php echo get_page_link($xpanel['submit-link']) ?>" class="header-submit-btn"><i class="fa fa-search fa-3x"></i></a>
73         </div>
74         <div class="user-info pull-right mr-10">
75           <?php
76           if ( is_user_logged_in() ) {
```

CODING - DICTIONARIES



MedDRA: MEDical Dictionary for Regulatory Affairs

- Medical Terminology Dictionary
- Used to *detect* and *code ADRs*



WHO-DDE: WHO-Drug Dictionary Enhanced

- Classification of medicines for drug identification
- Used to *detect* and *code drugs*

CODING - PROCESS

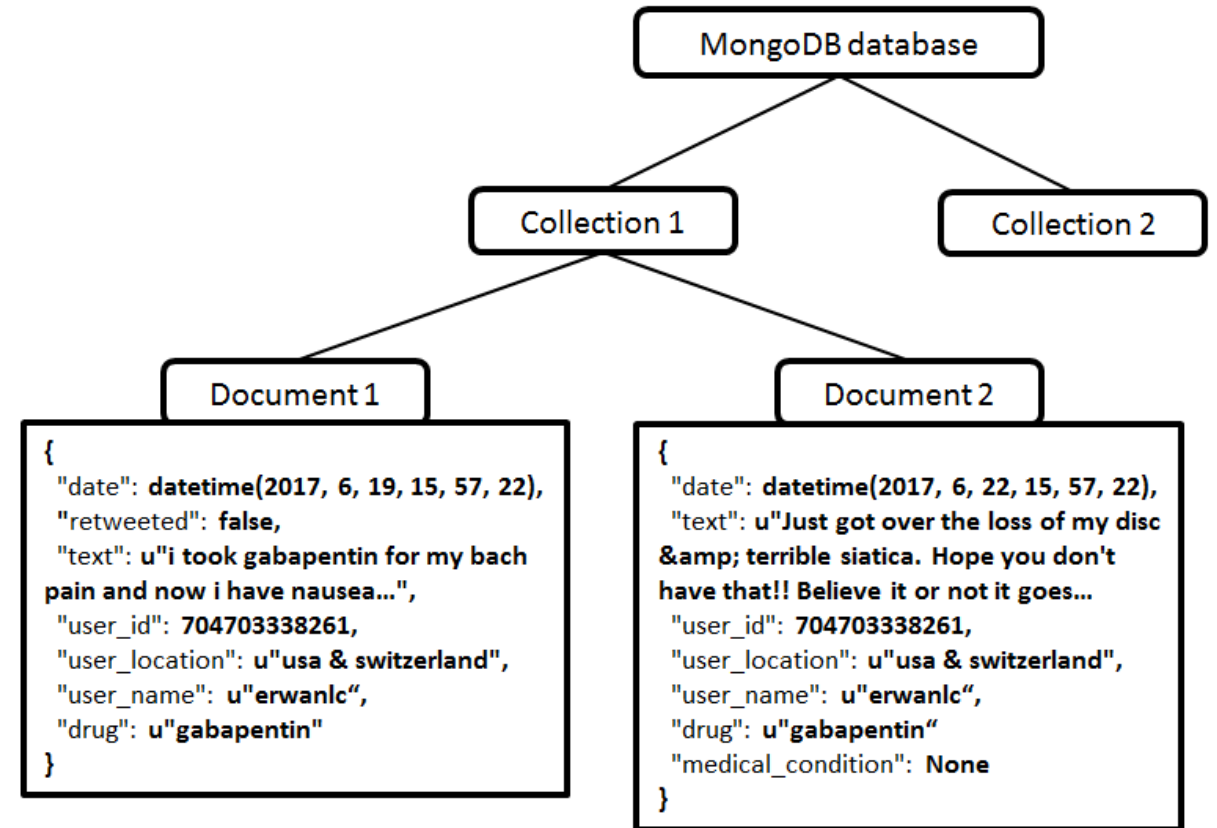


STORAGE

Twitter and Reddit can *change* their data *structure* anytime which can imply:

- Change of *data schema* and/or format
- Data *deletion*
- Data *addition*

→ Need of a *flexible structure*



MongoDB: a document-oriented database

AGENDA

Context

- Context
- Goal
- Sources

Proposed solution

- Process flow
- Extraction
- Coding
- Storage

Results & Discussion

- Results
- Discussion



RESULTS

	Twitter	Reddit
Period	3 months	1 year
# records	23,695	145,770
records / month	7,898.33	12,147.5

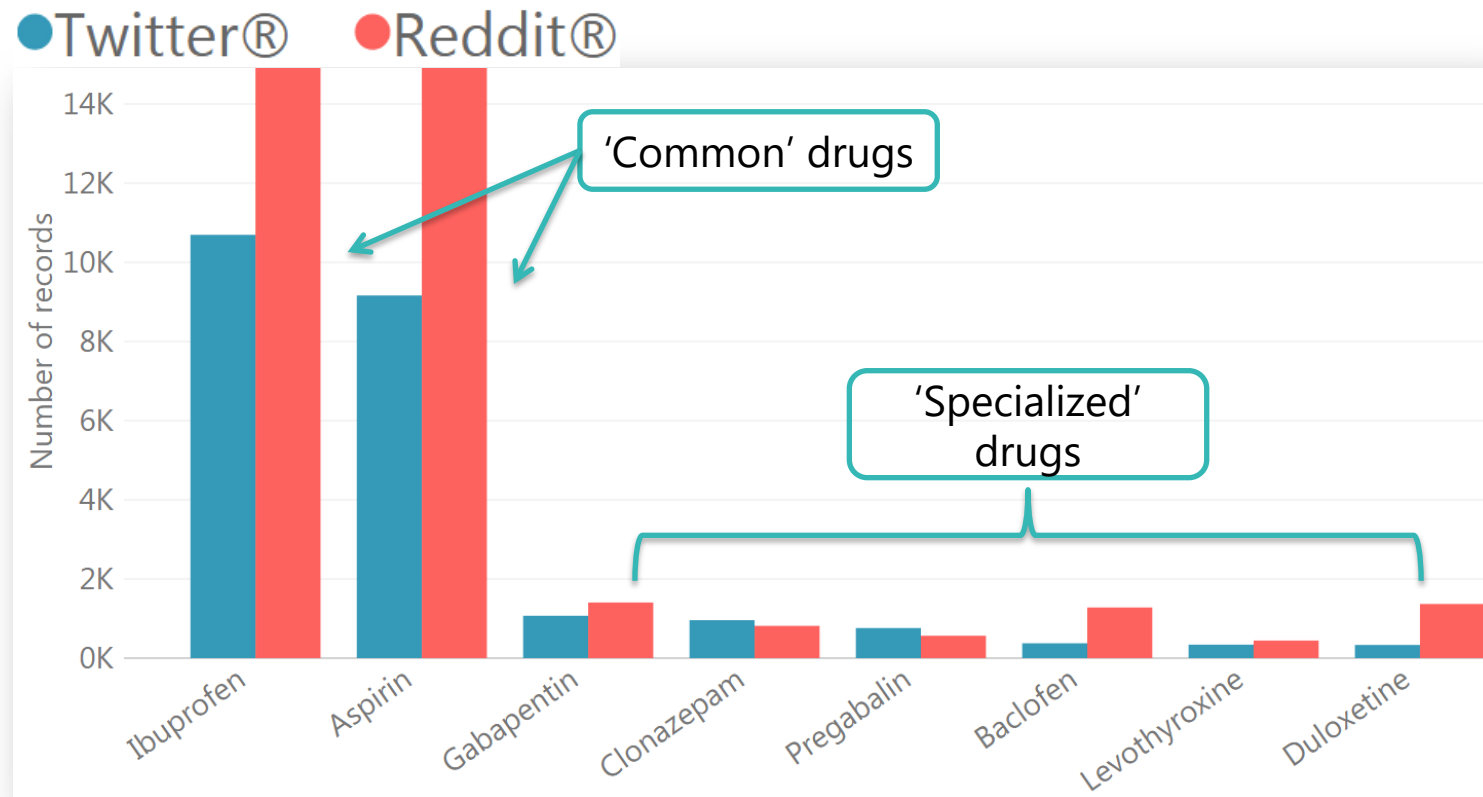
Disclaimer:

The data contained in this paper was obtained for demonstration purposes only (from Twitter and Reddit) using the techniques presented. Anyone analyzing data from social media sites, either public or membership based, should investigate the source terms and conditions relevant to data reuse and obtain any necessary permissions if required.

Detection ~40% of records with potential ADR*

*Potential ADR = ADR and other medical events

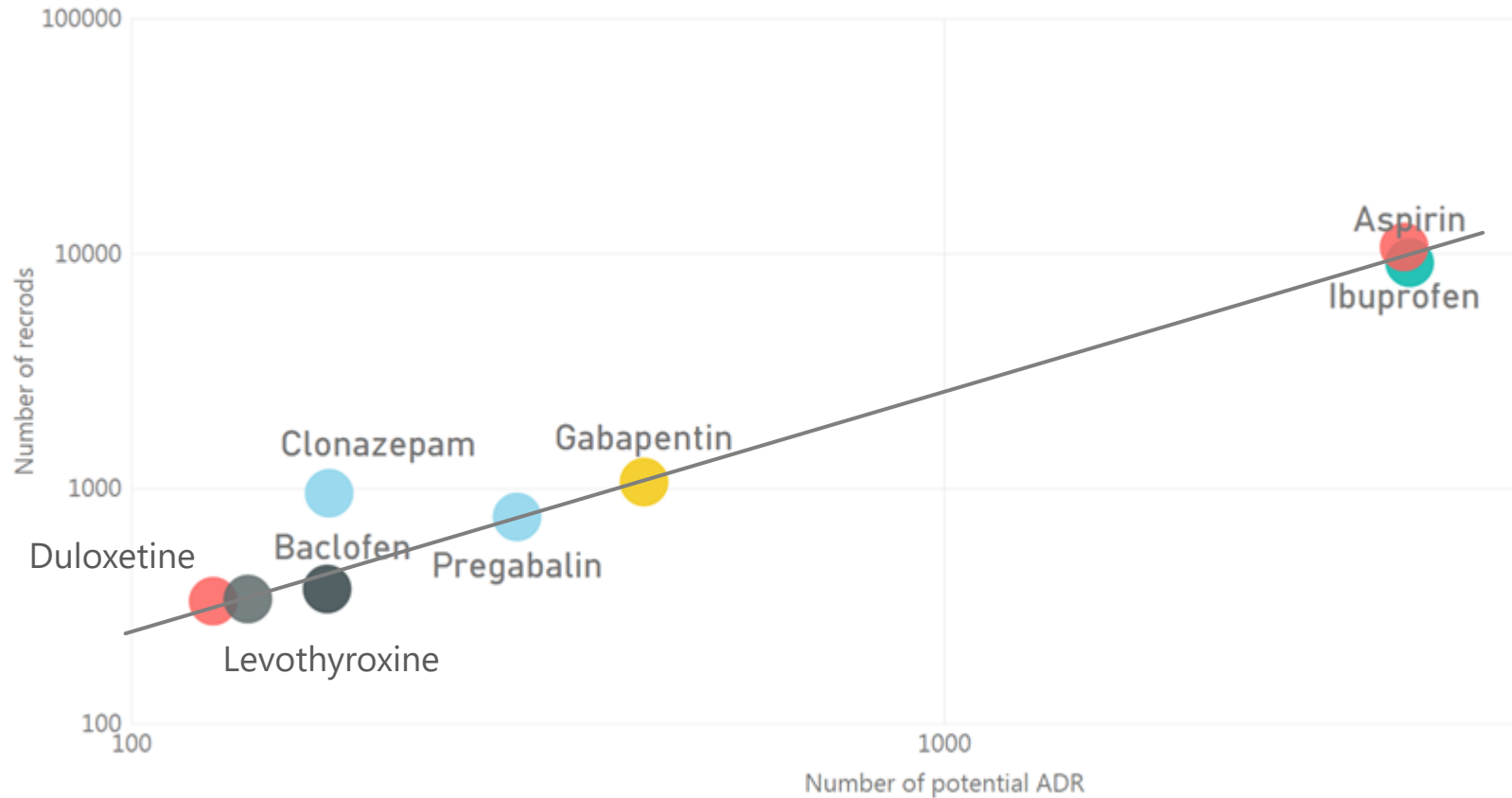
DRUG REPRESENTATION IN TWITTER® VS REDDIT®



- For *Aspirin* and *Ibuprofen*: ~116,000 records
- For *other* drugs: ~3,500 records

Several sources needed

RECORDS AND POTENTIAL ADR RATIO ON TWITTER®

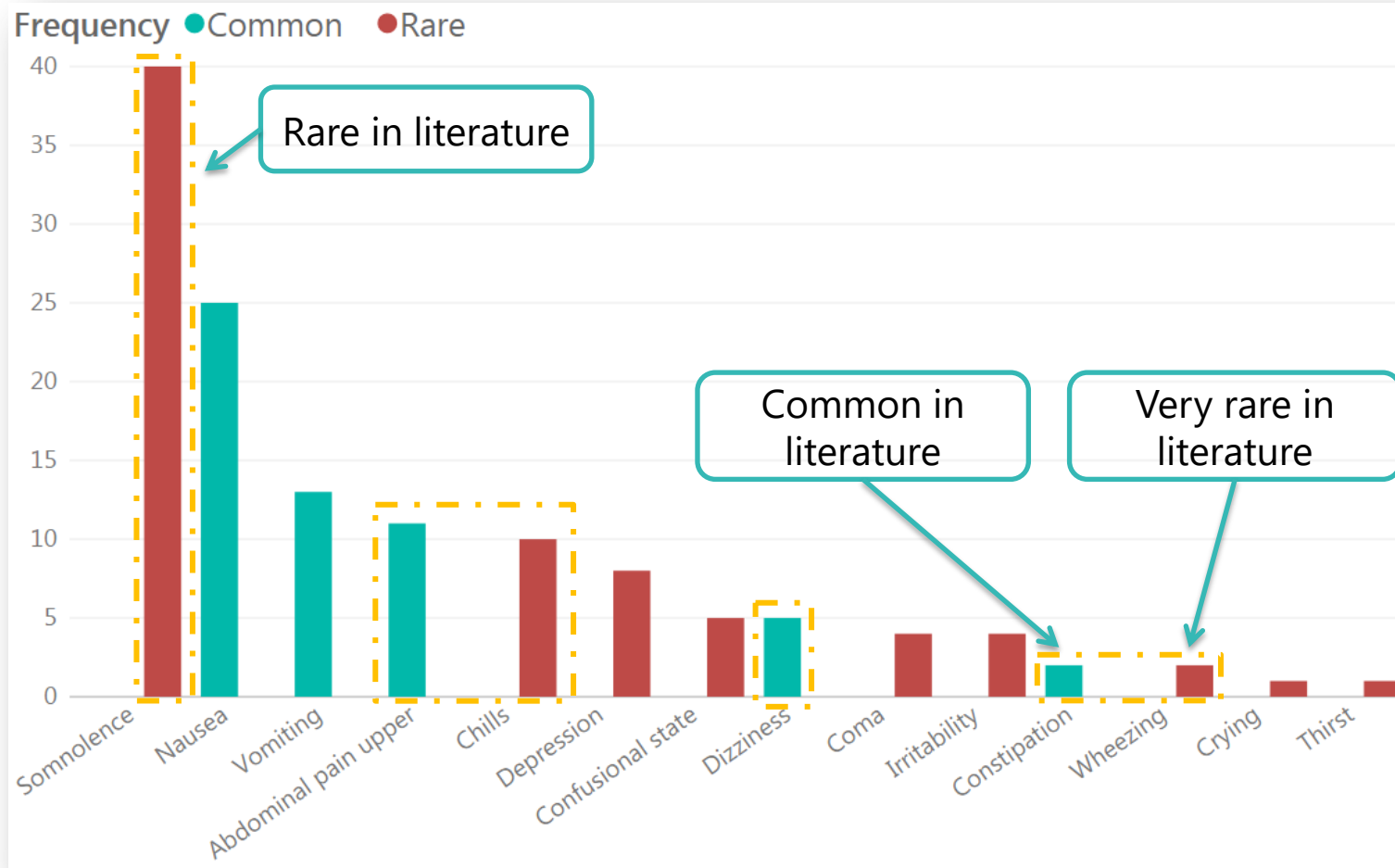


Ratio (potential ADR/records)
constant. ~40% for all drugs

- Dictionary size
- Expressions recognition
- Self learning

Method limitation ?

IBUPROFEN (MIXED STRENGTH): ADR DETECTION VS LITERATURE



Not expected data repartition:

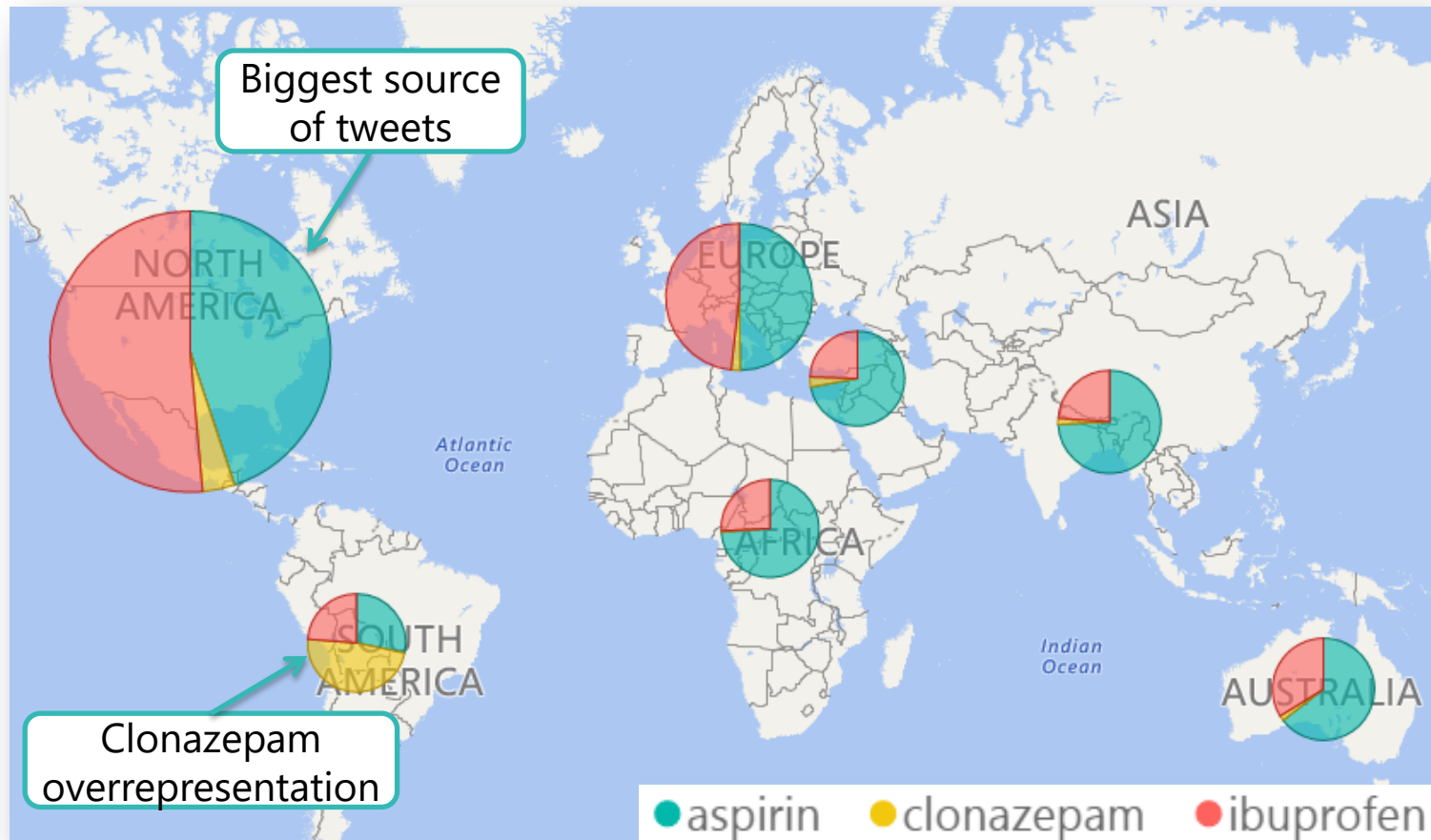
- Somnolence
- Constipation
- Wheezing

→ Social media **embarrassment**



Detection of all ADRs kind thus social media can be sources of new ADRs

WORLDWIDE DISTRIBUTION



Aspirin / Ibuprofen ratio:

- Occidental countries: ~**50 / 50**
- Africa, Middle East and Asia: ~**75 / 25**
- Oceania: ~ **34 / 66**
- South America: ~**55% / 45 %**

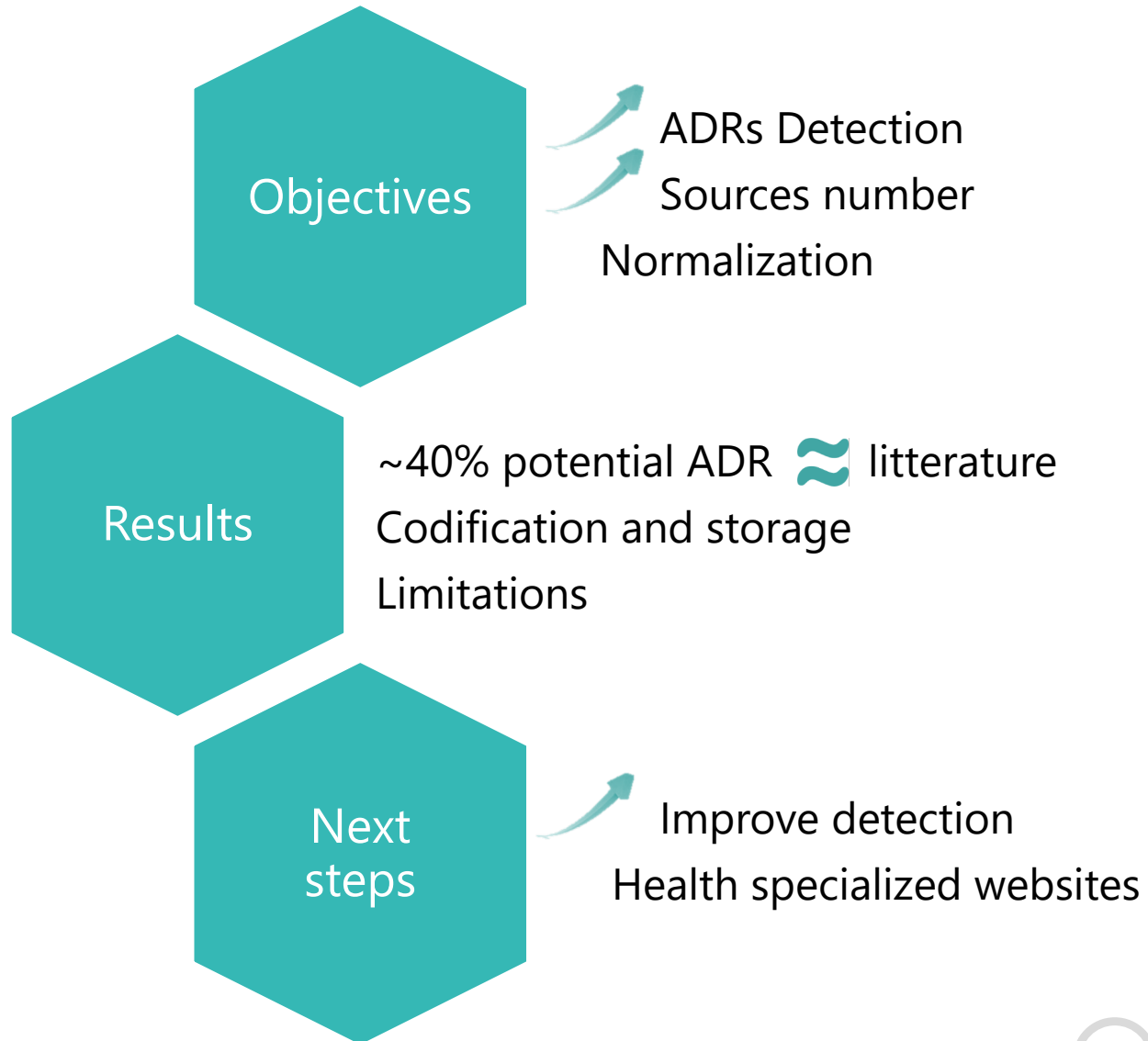
→ Aspirin **cheaper** than Ibuprofen ?

Clonazepam overrepresentation on South America:

→ **Recreational** use ?

Regions specificities

CONCLUSION



QUESTIONS



Erwan Le Covec

Stéphane Chollet

Essam Ghanem

**Keyrus
Biopharma**