A Proposed Approach to Address Data Gaps Through Data Mining Initiatives

1ST GLOBAL FOOD REGULATORY SCIENCE SYMPOSIUM
15th Dubai International Food Safety Conference (DIFSC)
20 November 2021

Amine KASSOUF, PhD

Research Scientist - GFoRSS

Mrs. Joyce Haddad

Expert - GFoRSS

➤ Introduction: Importance of Data Availability for Food Risk Analysis

➤ Current Situation in the Arab Region

➤ Proposed Methodology through Data Mining

➤ Conclusion and Future Perspectives

GF⚫RSS | GLOBAL FOOD REGULATORY SCIENCE SOCIETY

Introduction

Food analysis is a key component in the healthy function and performance of a food control system.

Sound decision making relies upon the **availability** and **accuracy** of **scientific data**, to inform the risk analysis process: the foundation for establishing responsible regulatory measures.

The investment in data availability to support food regulatory measures is one the key elements of the GFoRSS mandate.

Current Situation in the Arab Region

A **3-day Codex Colloquium** for the Middle East and North Africa Region was held in Dubai, United Arab Emirates, from **12 to 14 September, 2021**.

67 participants from **12 countries** were physically present and over **120** attended virtually

Organized by

With the partnership of

- The Ministry of Industry and Advanced Technologies of the UAE
- The National Codex Committee for the United Arab Emirates (NCC)
- Abu-Dhabi Agriculture and Food Safety Authority (ADAFSA)
- The Dubai Municipality

Current Situation in the Arab Region

Aim

➢ To enhance Codex capacity in the Arab region through the development of enablers to Codex work, such as **availability of data** information sharing and coordination.
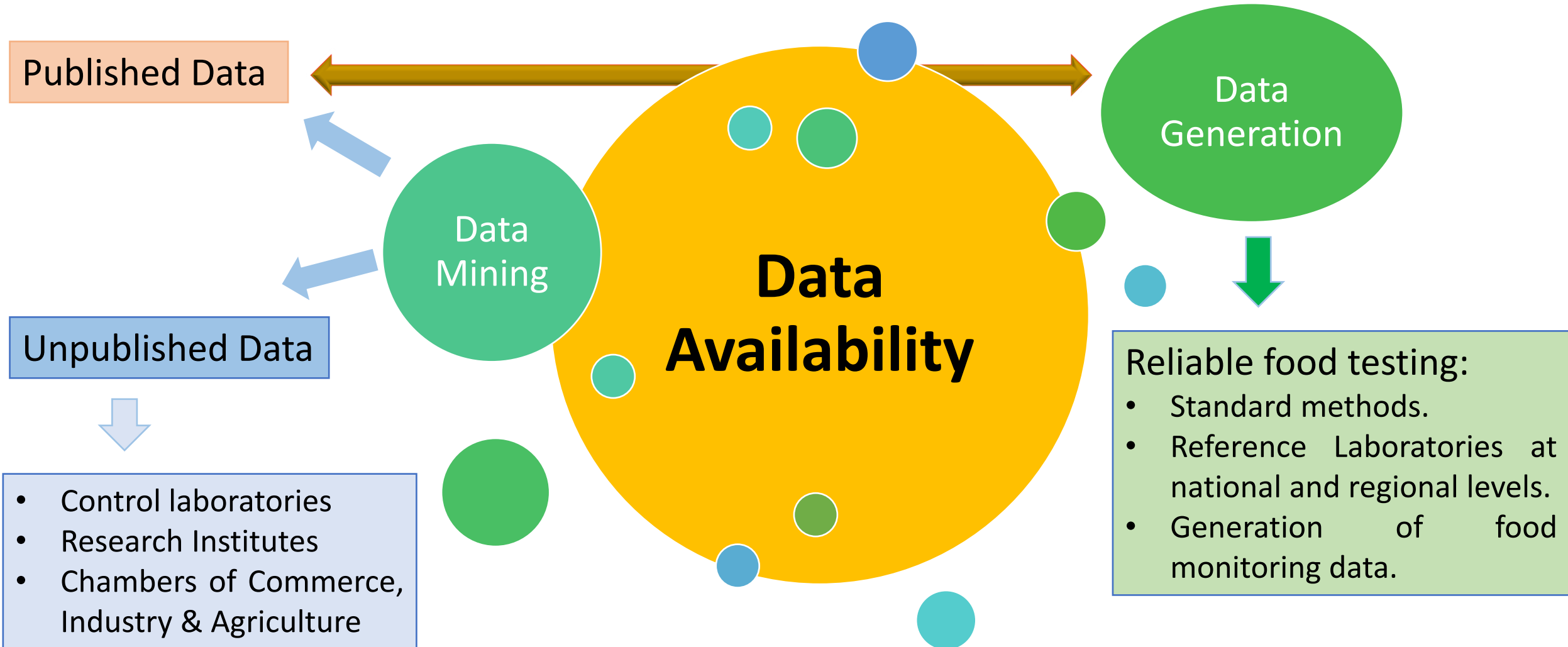
Discussions

➢ A major challenge for the region is to have access to **occurrence data**, so that competent authorities can prepare their national positions for upcoming discussions.

Recommendations

➢ Priority areas of investment to support Codex capacities in the region is to ensure **availability of data** from the region and any needed technical **tool** (online portal).

Proposed Methodology

Published Data

Data Mining

Data Availability

Data Generation

Unpublished Data

- Control laboratories
- Research Institutes
- Chambers of Commerce, Industry & Agriculture

Reliable food testing:
- Standard methods.
- Reference Laboratories at national and regional levels.
- Generation of food monitoring data.

Proposed Methodology:
Data Mining of Published Data

Systematic Scanning of the major scientific search engines and databases

Data Retrieval

Online Portal

Proposed Methodology: *Data Mining of Published Data*

1. ***Systematic scanning of major scientific search engines and databases***

   - Topic selection.
   - Establishment of the pool of articles related to the topic using well constructed research queries.
   - Start with research engines (Scholar google, Web of Science, etc.).
   - Scan scientific databases (ScienceDirect, Taylor and Francis, Springer, Wiley online library, MDPI, etc.).
   - Use of a reference manager to save time formatting citations and organize all this material (*e.g. Mendeley*).
   - Exclusion and inclusion criteria based on: journal properties (IF, Quartile, etc.), analytical quality of the data, etc.

Proposed Methodology: *Data Mining of Published Data*

2. *Data Retrieval:* after establishing the pool of articles related to the topic of interest, data is extracted respecting the following fields:

Reference
- Journal
- Publisher
- IF / Quartile
- Publishing year
- Country / region

Contaminants
- Heavy metals (Pb, Cd, etc.)
- Mycotoxins (AF, OCT, etc.)
- Others…

Food Commodity
- Food group
- WHO food identifier
- WHO food code
- Food origin
- State of food analyzed

Study Scheme
- Objective
- Sampling plan

Analytical Method
- Sample preparation
- Determination technique
- Analytical quality assurance (LOD, LOQ, recovery, precision RSD, etc.)

Contaminant Occurrence
- Concentration range (min-max)
- Average
- Median
- Number of samples below LOD

Proposed Methodology: *Data Mining of Published Data*

2. *Data Retrieval*: extracted data is organized based on contaminant and on food group/identifier, into an Excel sheet.

	Tag	Year	Country/Region	Contaminant	Food group	WHO Food Identifier	WHO Food Code	Food Origin	State of Food Analyzed
2	1	2013	Lebanon	Fe	Legumes and pulses	Chick-pea	VD 0524	Locally consumed	Canned (liquid+seeds)
3	1	2013	Lebanon	Sn	Legumes and pulses	Chick-pea	VD 0524	Locally consumed	Canned (liquid+seeds)
4	1	2013	Lebanon	Pb	Legumes and pulses	Chick-pea	VD 0524	Locally consumed	Canned (liquid+seeds)
5	1	2013	Lebanon	Cd	Legumes and pulses	Chick-pea	VD 0524	Locally consumed	Canned (liquid+seeds)

Study Objective	Sampling Plan	Analytical Technique	Analytical Quality Assurance	LOD	Unit	LOQ	Unit	Recovery (%)	Concentration Range-Min	Concentration Range-Max	Average	Unit	RSD (%)	Number of Samples below LOD
Migration from tinplate-coated cans into chickpeas	45 cans / Food canning industry	Microwave digestion/FAAS	Internal quality assurance only	2	mg/kg	9.5	mg/kg	96-97	<LOQ	27	N/A	mg/kg	7	0
Migration from tinplate-coated cans into chickpeas	45 cans / Food canning industry	Microwave digestion/FAAS	Internal quality assurance only	8.5	mg/kg	32	mg/kg	80-91	<LOQ	<LOQ	N/A	mg/kg	5.7	all samples
Migration from tinplate-coated cans into chickpeas	45 cans / Food canning industry	Microwave digestion/GFAAS	Internal quality assurance only	0.059	mg/kg	0.076	mg/kg	88-96	<LOQ	0.6	N/A	mg/kg	N/A	0
Migration from tinplate-coated cans into chickpeas	45 cans / Food canning industry	Microwave digestion/GFAAS	Internal quality assurance only	0.003	mg/kg	0.005	mg/kg	89-102	0.0146	0.0252	N/A	mg/kg	1.4 - 8	0

Proposed Methodology: *Data Mining of Published Data*

3. Online Portal: data is imported to an **online portal,** accessible at data entry level (Admin) or data consultation (User).
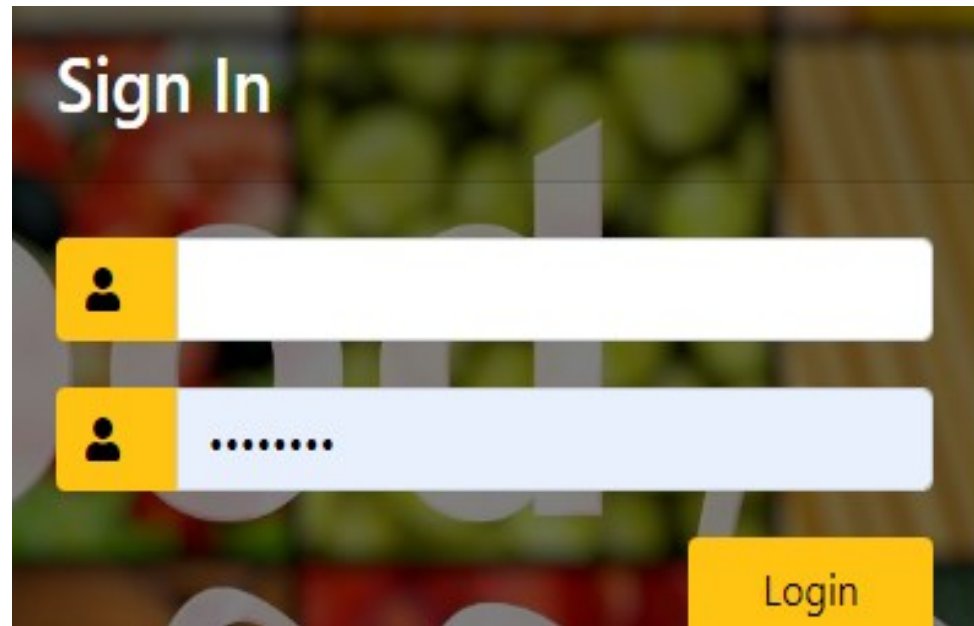
Search Criteria
Country
Contaminant
Food Category
Food Name
Date

Data treatments

Proposed Methodology: *Data Mining of Published Data*

3. *Online Portal: login*

Login Window
Credentials based on role: Admin or User

Proposed Methodology: *Data Mining of Published Data*

3. Online Portal: data entry – for admins only

Data Entry by Direct Upload

Data Entry Manually

Proposed Methodology: *Data Mining of Published Data*

3. *Online Portal: data access – for users*

Short-term Strategy

Screening of the Arab region on a country basis

Algeria
Bahrain
Comoros
Djibouti
Egypt
Iraq
Jordan
Kuwait
Lebanon
Libya

Mauritania
Morocco
Oman
Palestine
Qatar
Kingdom of Saudi Arabia
Somalia
Sudan
Syria
Tunisia
United Arab Emirates
Yemen

Short-term Strategy

Screening for heavy metals and mycotoxins to follow Codex standards development on these contaminants

Heavy metals
mainly Pb and Cd

Mycotoxins
mainly aflatoxins and ochratoxin A

ML for cadmium in cocoa powder containing or declaring 100% total cocoa solids

COP for the Prevention and Reduction of Cadmium Contamination in Cocoa Beans

Draft MLs for lead in various food categories

Maximum Levels for Aflatoxins (AFT) in Certain Cereals and Cereal-Based Products Including Food for Infants and Young Children

Maximum Level for Total Aflatoxins in Ready-To-Eat Peanuts and Associated Sampling Plan

Maximum Levels for Total Aflatoxins and Ochratoxin A in Nutmeg, Dried Chili and Paprika, Ginger, Pepper, and Turmeric and Associated Sampling Plans

Conclusion

The Arab region needs more accessible data to support food regulatory decisions and to prepare for Codex discussions and position taking.

GFoRSS will actively contribute to make data available and more accessible through its ARAB CODEX INITIATIVE.

The data mining will tackle any new contaminant of interest for the region.

This data mining project will also help to highlight gaps in data so regulatory control, research, funding, universities, etc. can be oriented towards addressing these topics in future works.

GFORSS | GLOBAL FOOD REGULATORY SCIENCE SOCIETY