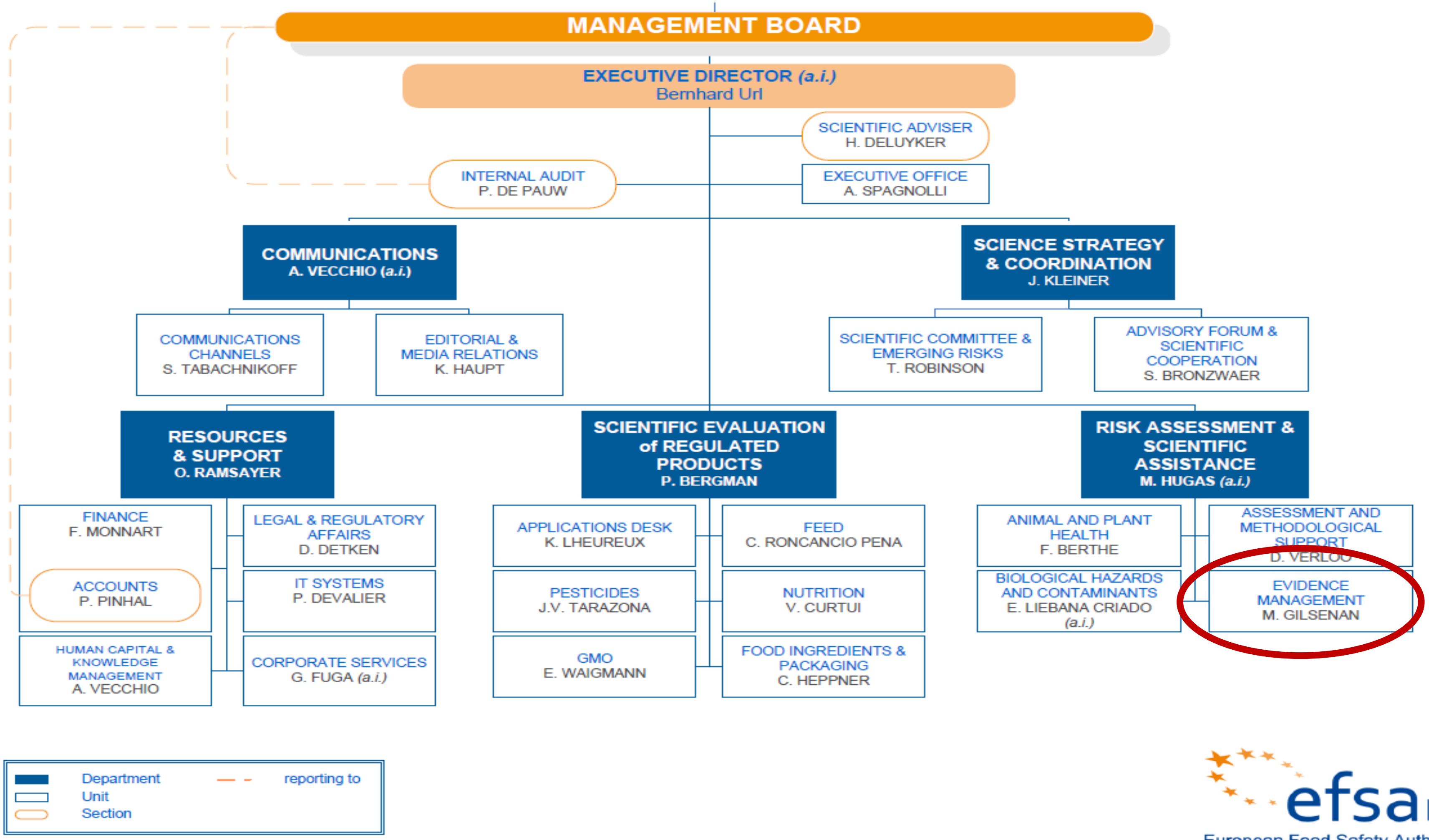
European Food Safety Authority

EFSA Data Warehouse: data collection and data access at EFSA

Data collection and information sharing in plant health
Parma, 1 - 3 April 2014

Preamble: the DATA unit



MANAGEMENT BOARD

EXECUTIVE DIRECTOR *(a.i.)*
Bernhard Url

SCIENTIFIC ADVISER
H. DELUYKER

INTERNAL AUDIT
P. DE PAUW

EXECUTIVE OFFICE
A. SPAGNOLLI

COMMUNICATIONS
A. VECCHIO *(a.i.)*

COMMUNICATIONS CHANNELS
S. TABACHNIKOFF

EDITORIAL & MEDIA RELATIONS
K. HAUPT

SCIENCE STRATEGY & COORDINATION
J. KLEINER

SCIENTIFIC COMMITTEE & EMERGING RISKS
T. ROBINSON

ADVISORY FORUM & SCIENTIFIC COOPERATION
S. BRONZWAER

RESOURCES & SUPPORT
O. RAMSAYER

FINANCE
F. MONNART

ACCOUNTS
P. PINHAL

HUMAN CAPITAL & KNOWLEDGE MANAGEMENT
A. VECCHIO

LEGAL & REGULATORY AFFAIRS
D. DETKEN

IT SYSTEMS
P. DEVALIER

CORPORATE SERVICES
G. FUGA *(a.i.)*

SCIENTIFIC EVALUATION of REGULATED PRODUCTS
P. BERGMAN

APPLICATIONS DESK
K. LHEUREUX

PESTICIDES
J.V. TARAZONA

GMO
E. WAIGMANN

FEED
C. RONCANCIO PENA

NUTRITION
V. CURTUI

FOOD INGREDIENTS & PACKAGING
C. HEPPNER

RISK ASSESSMENT & SCIENTIFIC ASSISTANCE
M. HUGAS *(a.i.)*

ANIMAL AND PLANT HEALTH
F. BERTHE

BIOLOGICAL HAZARDS AND CONTAMINANTS
E. LIEBANA CRIADO *(a.i.)*

ASSESSMENT AND METHODOLOGICAL SUPPORT
D. VERLOO

EVIDENCE MANAGEMENT
M. GILSENAN

Summary

Data collection and analysis in EFSA reports

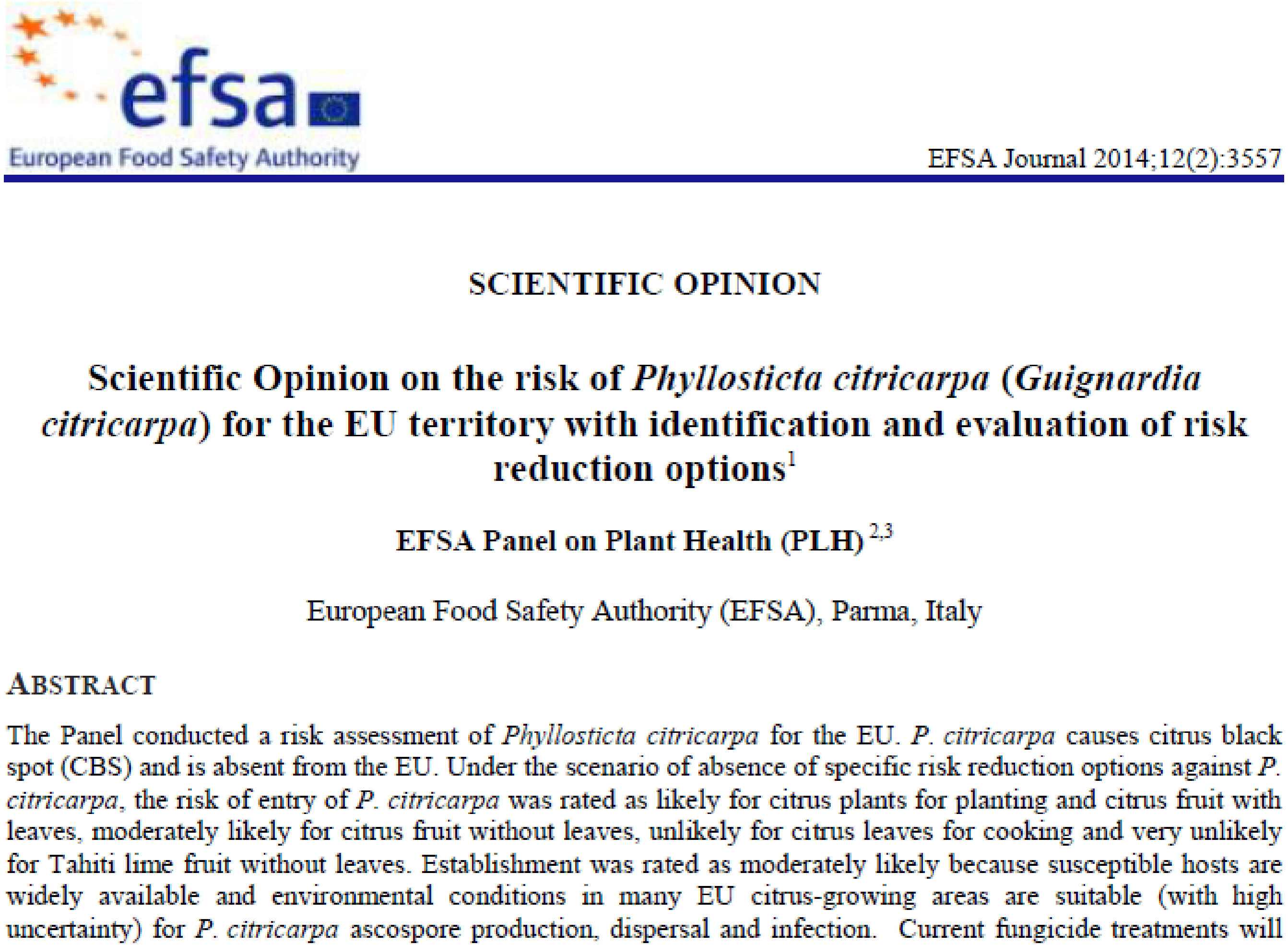
Standardised data collections

The data warehouse

Timelines

Data analysis in EFSA reports

Data are published in EFSA reports



EFSA Journal 2014;12(2):3557

SCIENTIFIC OPINION

Scientific Opinion on the risk of *Phyllosticta citricarpa* (*Guignardia citricarpa*) for the EU territory with identification and evaluation of risk reduction options[1]

EFSA Panel on Plant Health (PLH)[2,3]

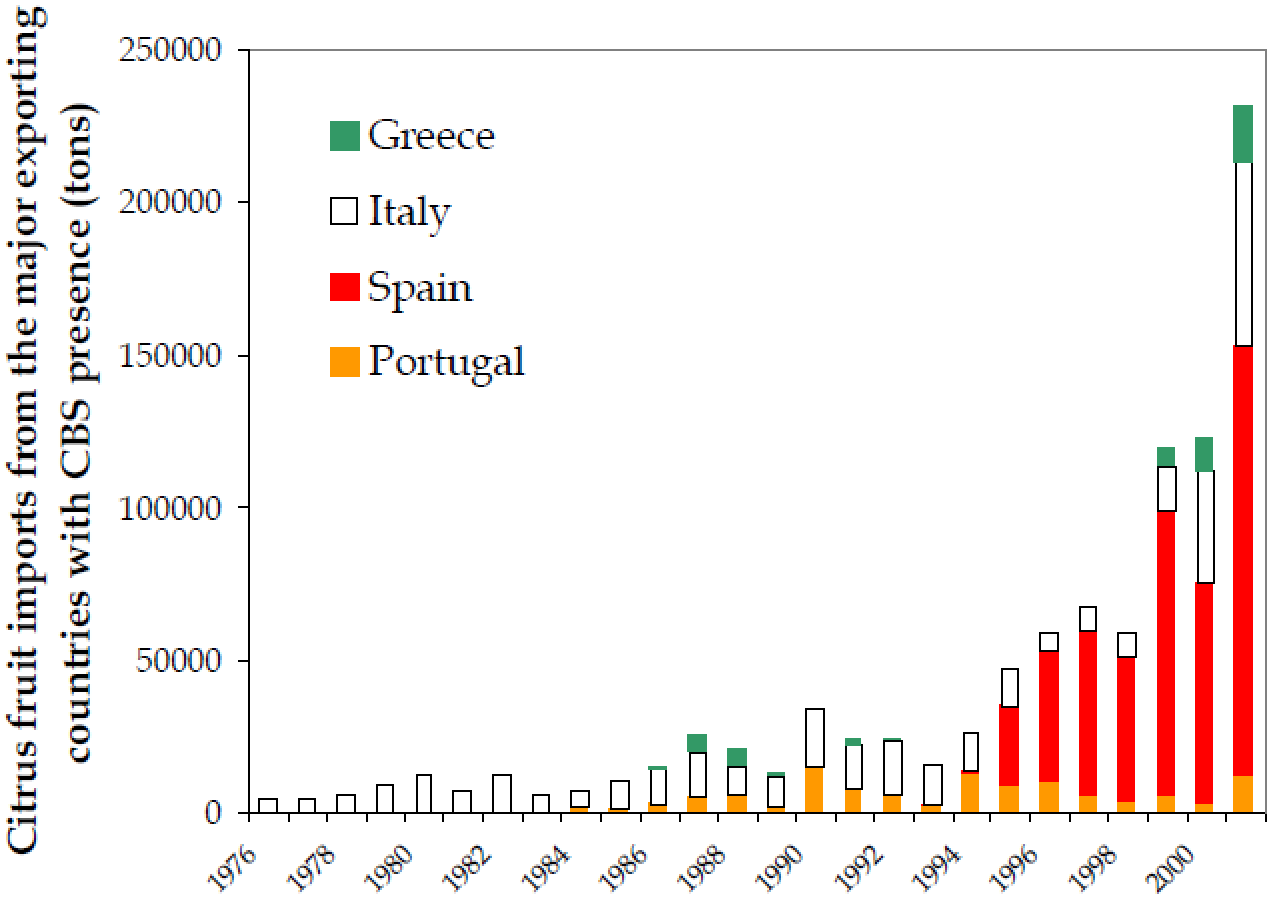
European Food Safety Authority (EFSA), Parma, Italy

ABSTRACT

The Panel conducted a risk assessment of *Phyllosticta citricarpa* for the EU. *P. citricarpa* causes citrus black spot (CBS) and is absent from the EU. Under the scenario of absence of specific risk reduction options against *P. citricarpa*, the risk of entry of *P. citricarpa* was rated as likely for citrus plants for planting and citrus fruit with leaves, moderately likely for citrus fruit without leaves, unlikely for citrus leaves for cooking and very unlikely for Tahiti lime fruit without leaves. Establishment was rated as moderately likely because susceptible hosts are widely available and environmental conditions in many EU citrus-growing areas are suitable (with high uncertainty) for *P. citricarpa* ascospore production, dispersal and infection. Current fungicide treatments will

Data analysis in EFSA reports

Data are visualised as graphs



Data analysis in EFSA reports

Data is aggregated in tables

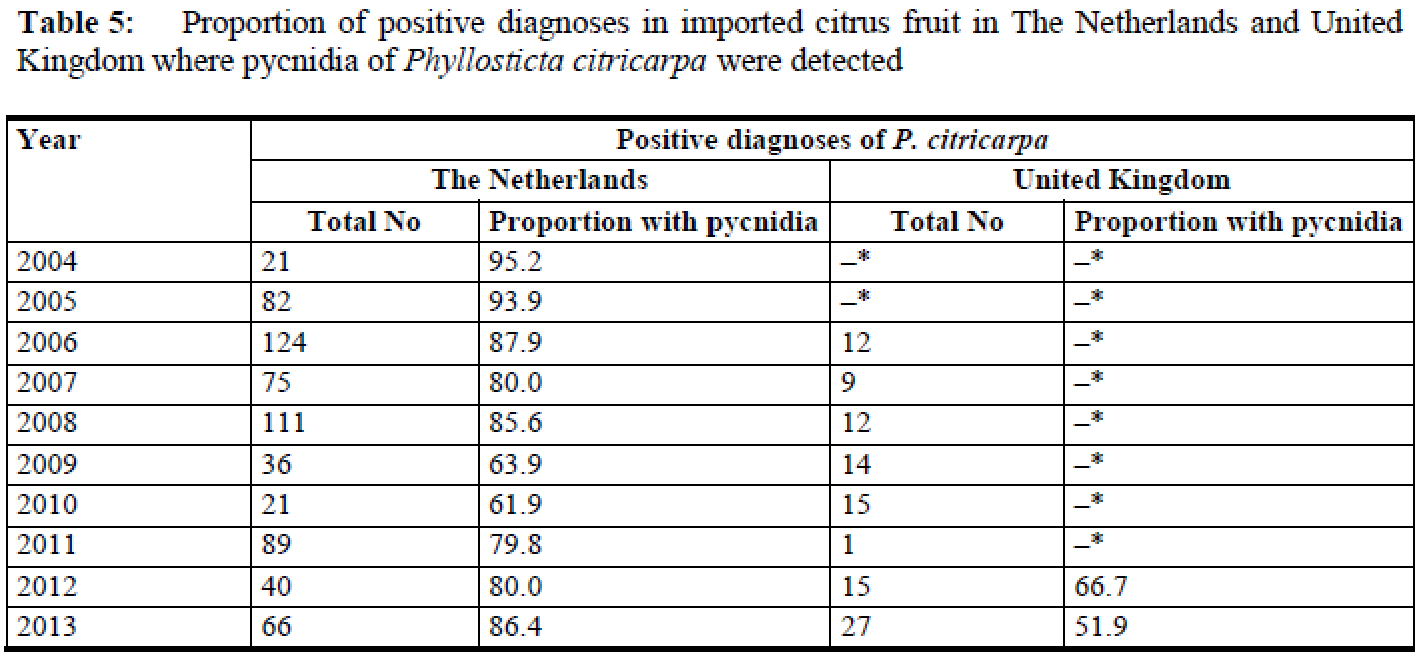
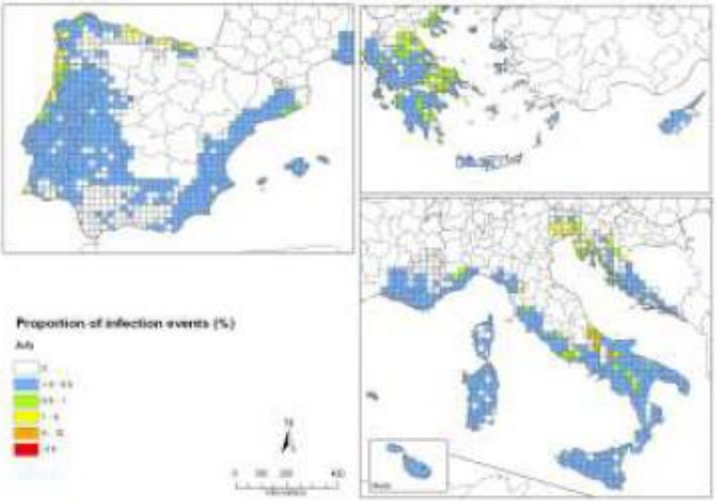


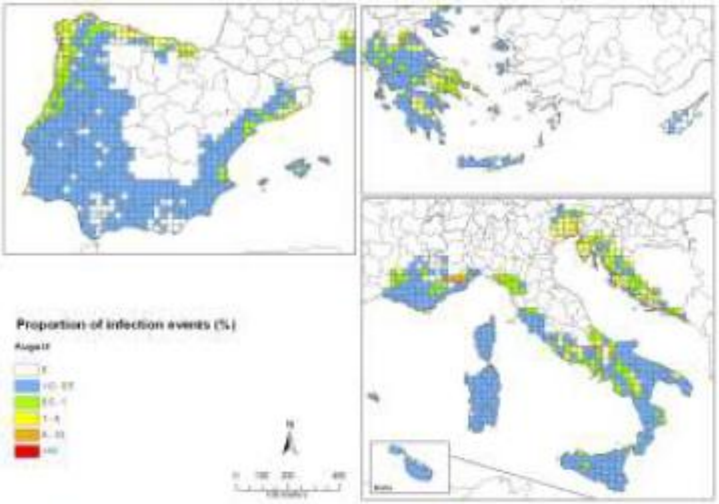
Table 5: Proportion of positive diagnoses in imported citrus fruit in The Netherlands and United Kingdom where pycnidia of *Phyllosticta citricarpa* were detected

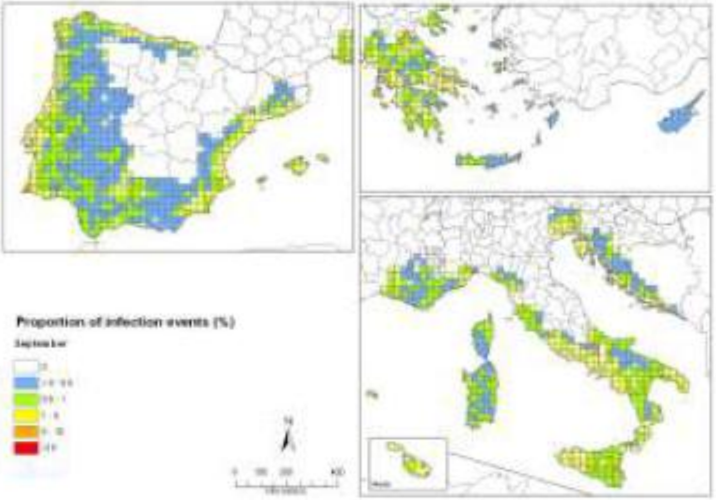
Year	**Positive diagnoses of *P. citricarpa***			
	The Netherlands		**United Kingdom**	
	Total No	**Proportion with pycnidia**	**Total No**	**Proportion with pycnidia**
2004	21	95.2	–*	–*
2005	82	93.9	–*	–*
2006	124	87.9	12	–*
2007	75	80.0	9	–*
2008	111	85.6	12	–*
2009	36	63.9	14	–*
2010	21	61.9	15	–*
2011	89	79.8	1	–*
2012	40	80.0	15	66.7
2013	66	86.4	27	51.9

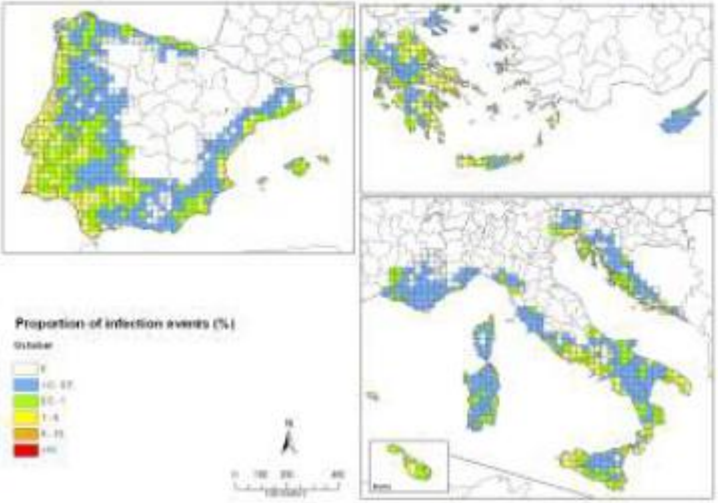
Data analysis in EFSA reports

Maps





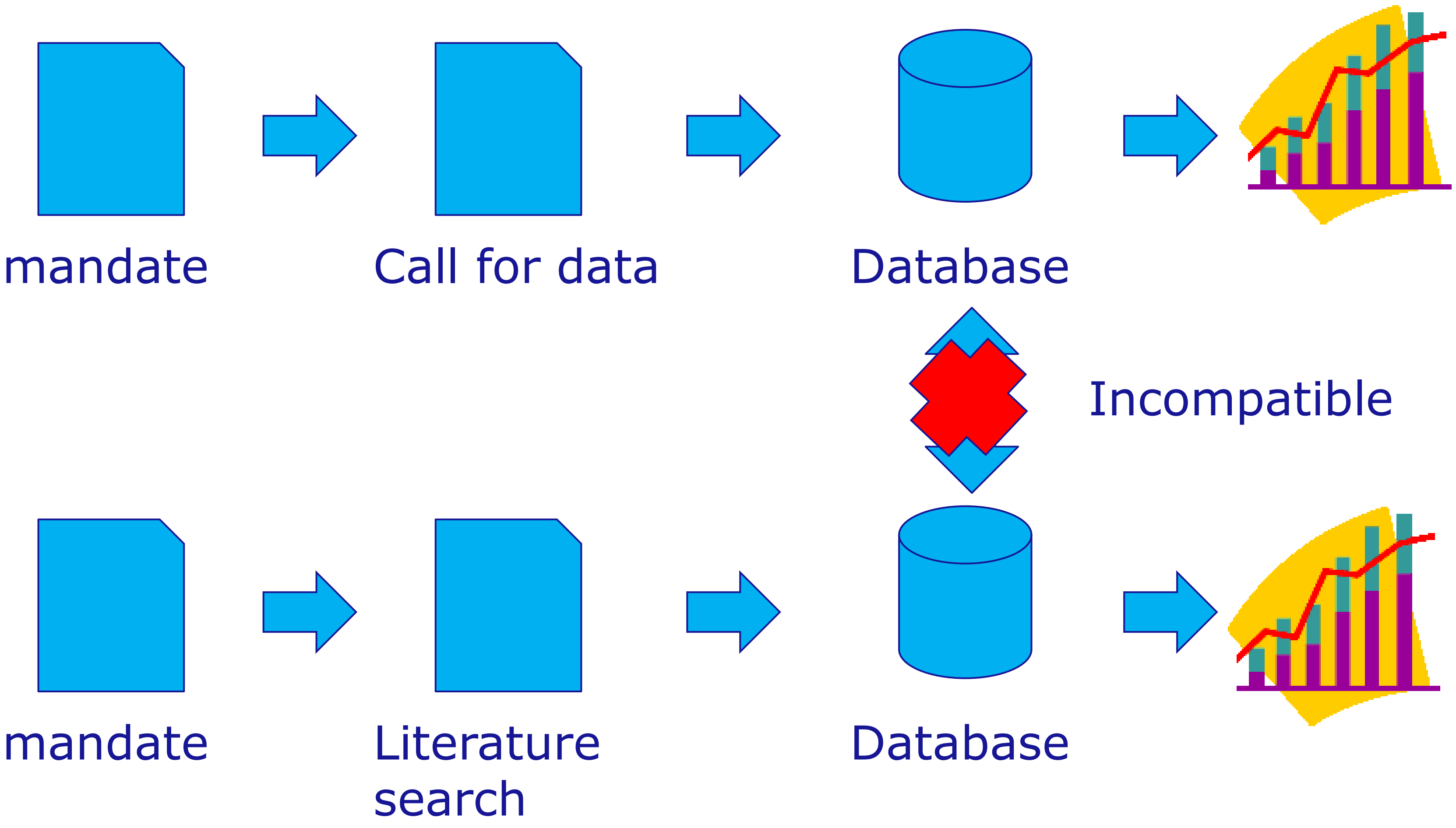




Limits of data analysis in EFSA reports

- Graphs, tables and maps are fixed
  - No further data analyses
    - Change the exclusion/selection criteria
    - Challenge the data
    - Verify whether data is robust
    - No additional analyses can be performed
- Graphs, tables and maps are produced by data specialists rather than science specialists
- Low possibility of misinterpreting data by readers

Ad-hoc data collections



mandate

Call for data

Database

Incompatible

mandate

Literature search

Database

Limits of ad-hoc data collections

- Ad-hoc data collection make data “disposable”
  - Data collected for single use only
  - Very difficult to use for further re-use in future
- Data collected is isolated, no possibility to compare or link the collected data with other data to perform additional analysis (e.g. pesticide residues data)

Summary

Data collection and analysis in EFSA reports

Standardised data collections

The data warehouse

Timelines

Standardised data collections

- Analysis of data elements of interest for a certain area e.g. plan health
- Identification of the standard data elements, their data type and their controlled terminologies
- Publication of the standard data collection formats

What is standardised data collection?

1. A list of data elements that are standardised and can be conveniently used by both data providers and data receivers to fully describe samples and analytical parameters for assessment purposes.

2. Includes controlled terminologies and validation rules to guarantee data quality (in data export, transmission and storage)

A model harmonising the collection of a wide range of data collected in several domains on the “same subject” of EFSA activity

An example:
Standard Sample Description

Currently implemented for:

- ✓ Chemical contaminants
- ✓ Pesticide residues
- ✓ Additives
- ✓ Food contact materials

Standard Sample Description ver. 2.0

- ✓ Antimicrobial isolate based data
- ✓ Data on microbiological contaminants at single sample level
- ✓ Data on zoonotic agents at single sample or flock/herd level
- ✓ Full support of the new EFSA food classification system (FoodEx2)
- ✓ Improvements and practical user experience

Moving to plant heath data…

Comprehensive data model for the collection of data on pest and pathogens of animal and plants from literature

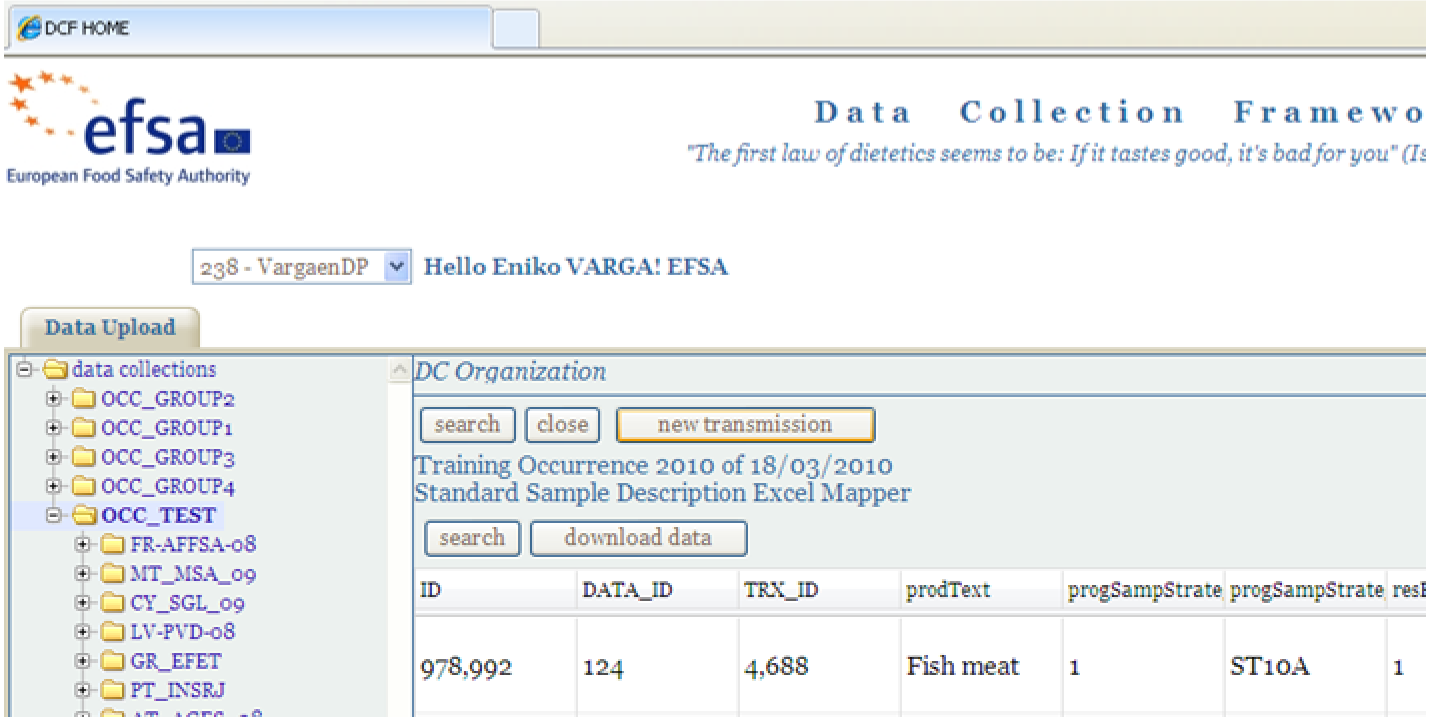
Structured data about:

- Scientific source
- Occurrence and interactions of pest/pathogen, vector (where applicable) and host
- Geographic distribution
- Environmental parameters
- Biology of the involved actors

Set of standard terminologies to support the content standardisation

Data reporting: the DCF system

❖Data Collection Framework (DCF) can be used by Data Providers to submit data.



Data Collection Framewo

"The first law of dietetics seems to be: If it tastes good, it's bad for you" (Is

238 - VargaenDP

Hello Eniko VARGA! EFSA

- data collections
  - OCC_GROUP2
  - OCC_GROUP1
  - OCC_GROUP3
  - OCC_GROUP4
  - OCC_TEST
    - FR-AFFSA-08
    - MT_MSA_09
    - CY_SGL_09
    - LV-PVD-08
    - GR_EFET
    - PT_INSRJ

DC Organization

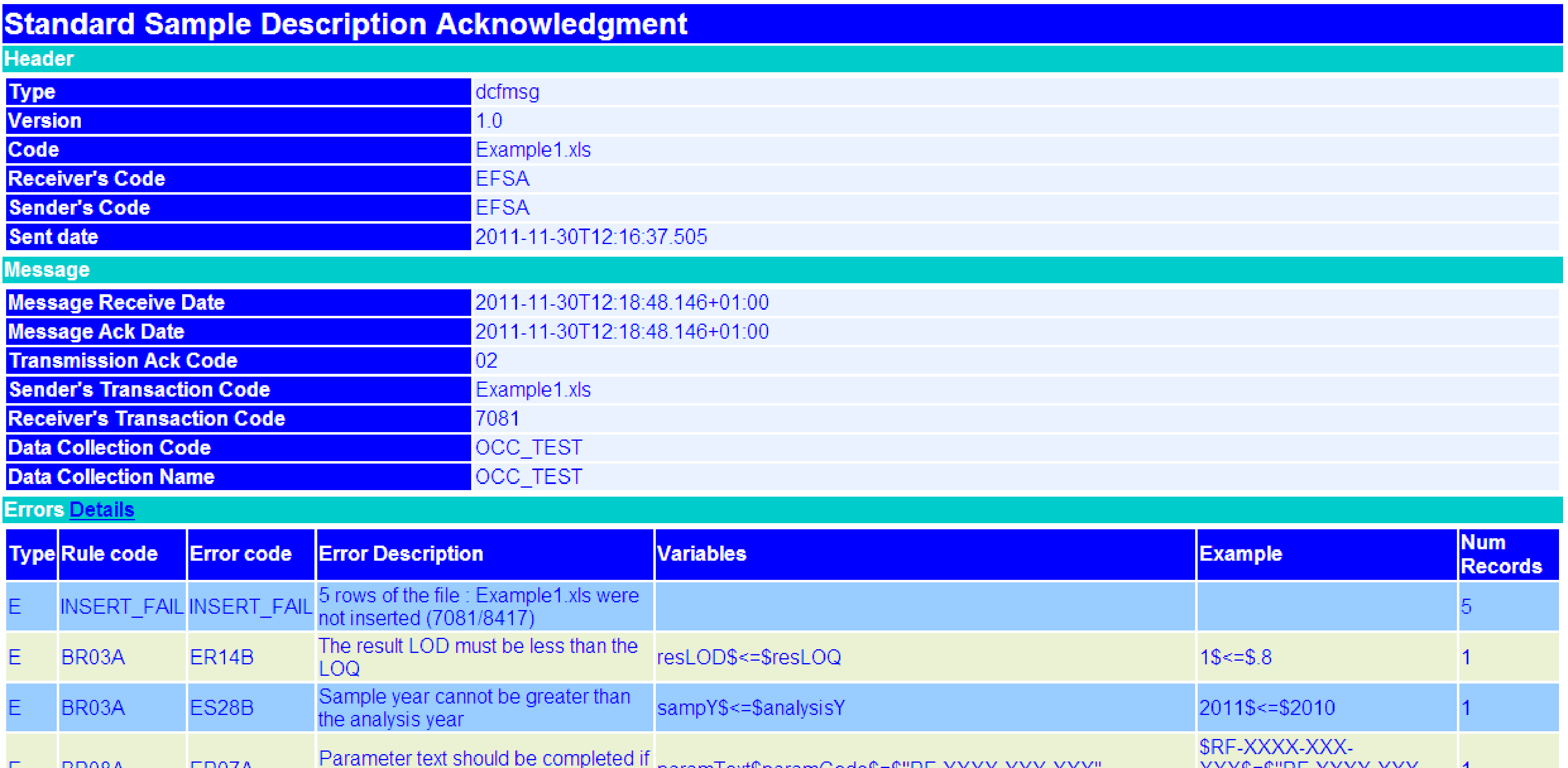
Training Occurrence 2010 of 18/03/2010
Standard Sample Description Excel Mapper

ID	DATA_ID	TRX_ID	prodText	progSampStrate	progSampStrate	resE
978,992	124	4,688	Fish meat	1	ST10A	1

DCF data transmission supports the following formats:
XML, Excel, CSV

The DCF system automatic validation

- An automatic feedback is sent to data providers.



Standard Sample Description Acknowledgment

Header

Type	dcfmsg
Version	1.0
Code	Example1.xls
Receiver's Code	EFSA
Sender's Code	EFSA
Sent date	2011-11-30T12:16:37.505

Message

Message Receive Date	2011-11-30T12:18:48.146+01:00
Message Ack Date	2011-11-30T12:18:48.146+01:00
Transmission Ack Code	02
Sender's Transaction Code	Example1.xls
Receiver's Transaction Code	7081
Data Collection Code	OCC_TEST
Data Collection Name	OCC_TEST

Errors Details

Type	Rule code	Error code	Error Description	Variables	Example	Num Records
E	INSERT_FAIL	INSERT_FAIL	5 rows of the file : Example1.xls were not inserted (7081/8417)			5
E	BR03A	ER14B	The result LOD must be less than the LOQ	resLOD$<=$resLOQ	1$<=$.8	1
E	BR03A	ES28B	Sample year cannot be greater than the analysis year	sampY$<=$analysisY	2011$<=$2010	1
E	BR08A	ER07A	Parameter text should be completed if	paramText$paramCode$=$"RF-XXXX-XXX-XXX"	$RF-XXXX-XXX-XXX$=$"RF-XXXX-XXX	1

Summary

Data collection and analysis in EFSA reports

Standardised data collections

The data warehouse

Timelines

A data warehouse for data analysis

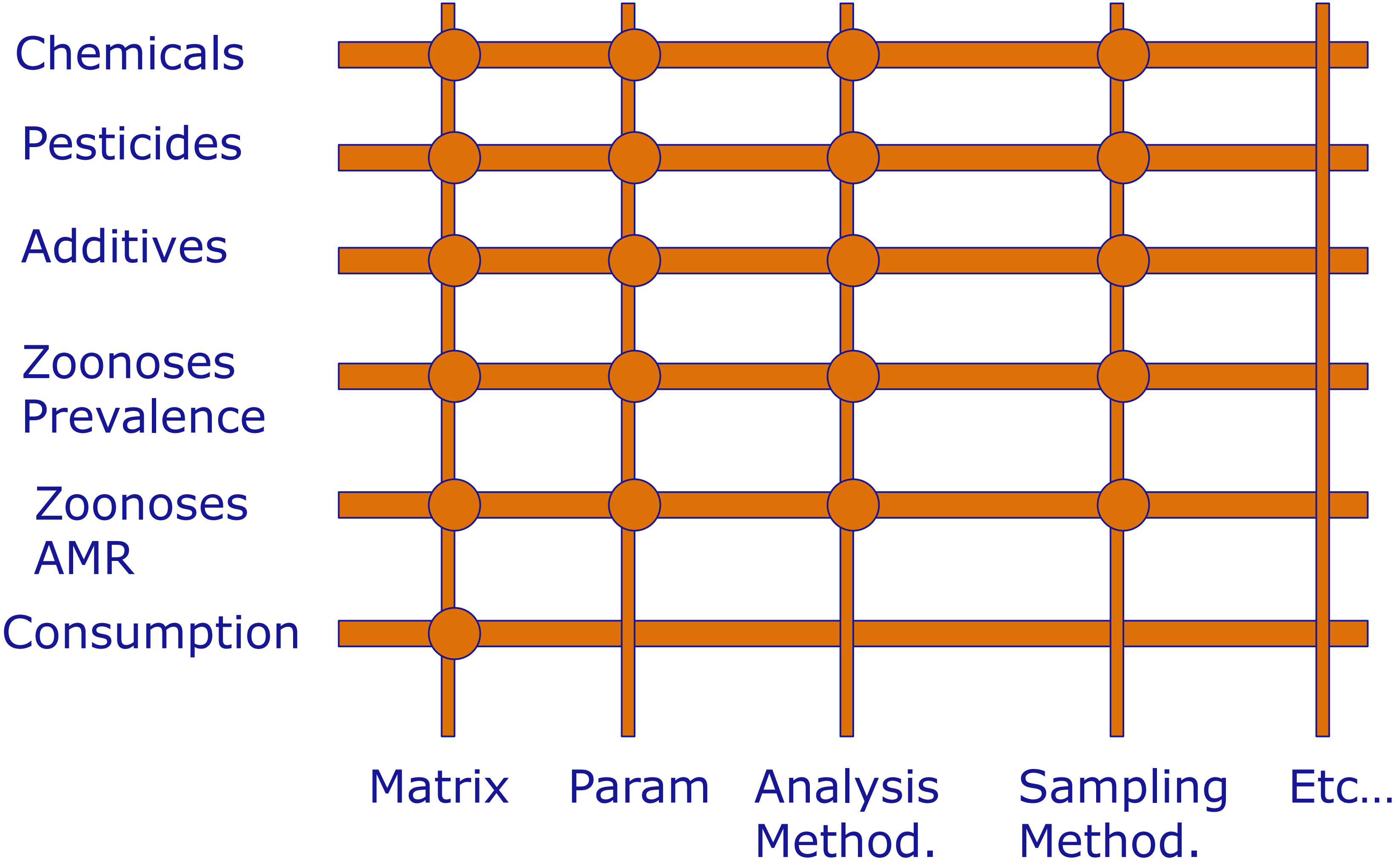
Bill Imnom definition:

- **Subject oriented**: Not organised ad-hoc but per area
- **Integrated**: using controlled terminology
- **Persistent**: Store data for years

Collection of data

- **With history management**: Capable of analysing data as available in the past

The DWH schema



	Matrix	Param	Analysis Method.	Sampling Method.	Etc…
Chemicals	●	●	●	●	
Pesticides	●	●	●	●	
Additives	●	●	●	●	
Zoonoses Prevalence	●	●	●	●	
Zoonoses AMR	●	●	●	●	
Consumption	●				

Web reporting tool

- Data warehouse is a database
- Web reporting tool or analysis tool is necessary to access the data warehouse
- Prepare a series of standard reports (graphs, tables) to access the data in different area

Access to data

- Data warehouse access policy under approval
- Identify key actors
- Provide specific access conditions for all stakeholders, e.g.:
  - Data providers (access data they submitted entirely)
  - Working group members and panel members (access data under their working scope entirely)
  - Other External stakeholders (access data at certain level of aggregation)

Summary

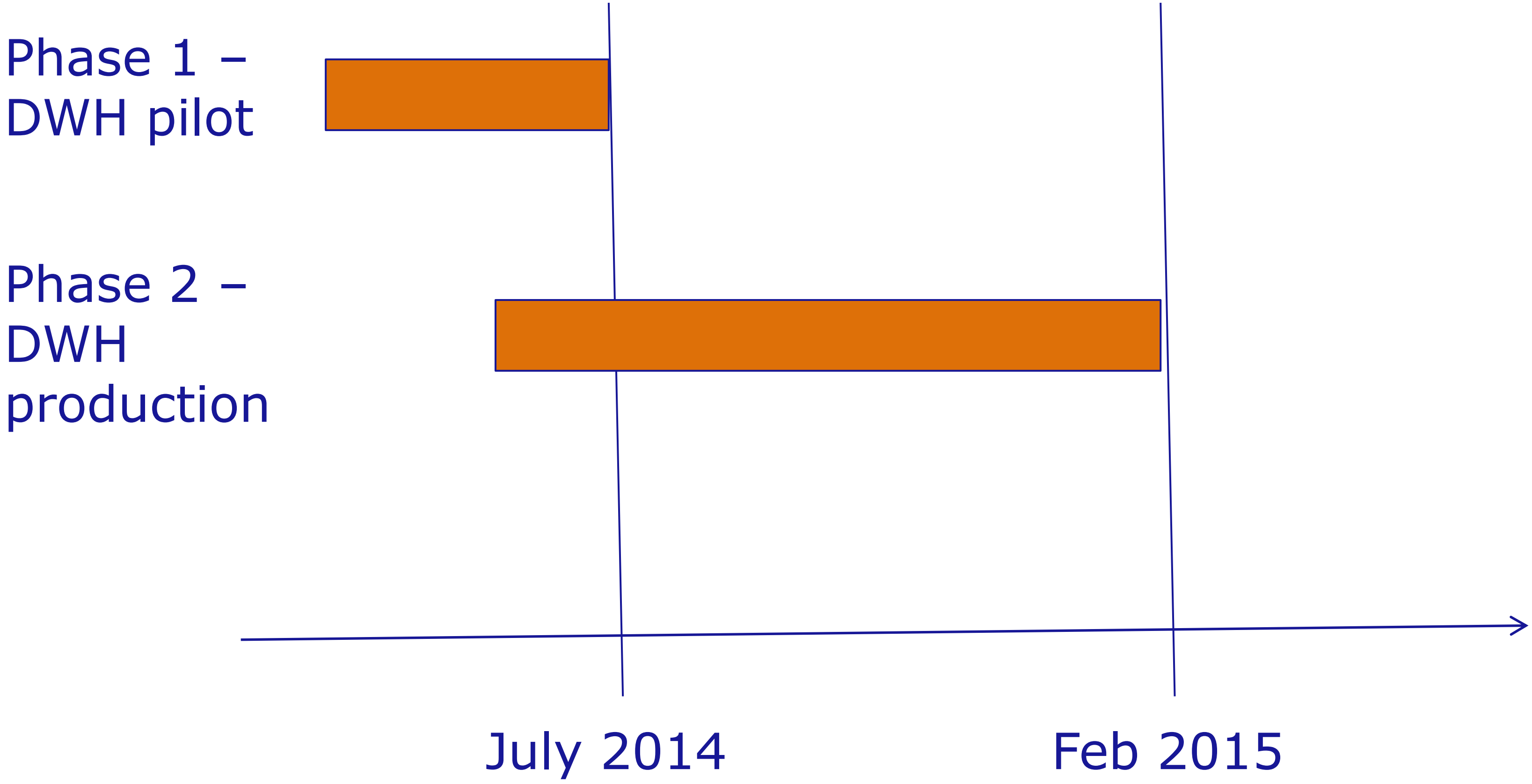
Data collection and analysis in EFSA reports

Standardised data collections

The data warehouse

Timelines

Timelines



Phase 1 – DWH pilot

Phase 2 – DWH production

July 2014

Feb 2015

Access from ouside

- After the completion of phase 2 the DWH will be initially used to perform data reporting on data delivered by data providers (High priority: Zoonoses data)
- Support the EC in risk management activities providing access to collected data (High priority: pesticide residues data and contaminants)
- Gradually opening to all stakeholders indicated in the data warehouse access policy during 2015 for available subjects

Conclusions

- Standardisation of the collected data is a fundamental component of building a data warehouse:
  - availability of standards to collect the data
  - standards implemented in the data warehouse for supporting data analysis
- The availability of a data warehouse:
  - simplifies the data analysis and the integration of the data collected
  - boosts transparency in the analysis performed
  - allow further analysis of data (e.g. when new data available)

Thank you!

Questions?

Stefano.Cappe@efsa.europa.eu

Francesca.Riolo@efsa.europa.eu