# Bringing EFSA Scientific Data Warehouse to the next level

A.Ciccarelli,
EU Technology Liaison, DTS

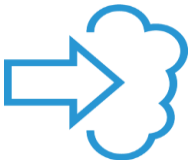
*70th Advisory Forum meeting*
*Vienna, Austria, 28-29.11.2018*

efsa
European Food Safety Authority

# DATA ANALYTICS: FACING NEW CHALLENGES

- Nowadays we are all facing the following problems in the Data Collection and Data Analysis processes:

| Problems | Needs |
| --- | --- |
| Scalability of 'on-premises data collection and data analysis solutions' | Flexibility of computational and storage capacity: add servers when users need more speed and remove them when they are not needed; |
| Increasing costs for 'storage and computational' capacity | More autonomy and empowerment of data managers/ analysts: Perform their duties in self-service mode Implement Data Management and Data Analysis as a Service |
| Performance issues | Provide and share with member States and other Agencies the Data Collections and Data Analytics services (i.e., computing, storage resources) bringing mutual collaboration and value creation to an higher level |
| Need of more autonomy in the configuration and usage pf the IT environments dedicated to Data Scientists | |

***Solution***  ***Move the Scientific Data Warehouse environment to the Cloud***

# SOME KEY FACTS ABOUT THE CLOUD

- Migration to the Cloud is today a **general trend** in Government and private industry: the question is not **If** but **When**
- Different types of clouds:
  - **Public** -> Amazon, Microsoft Azure -> shared resources
  - **Private or Community** -> similar to hosting/outsourcing -> dedicated and segregated infrastructure c/o EU certified providers
- Cloud environments are today **more secure** than on-premise environments: number of security incidents **in cloud environments** is **lower** respect to incidents recorded **on-premise** (true also for EFSA)
- **Cost efficiencies**, specifically in terms of scalability and infrastructural maintenance/renewal
- **Framework contracts** established and regularly used at **EU Commission** and **Agencies** since 2015 (> 3 years)

# WHAT IS THE EU AGENCIES COMMUNITY CLOUD

- IT infrastructure segregated and dedicated to the 22 EU Agencies participating to the EFSA Interagency Cloud Framework Contract

- Hosted in the Cancom Data Centres located in Germany (Hamburg and München)

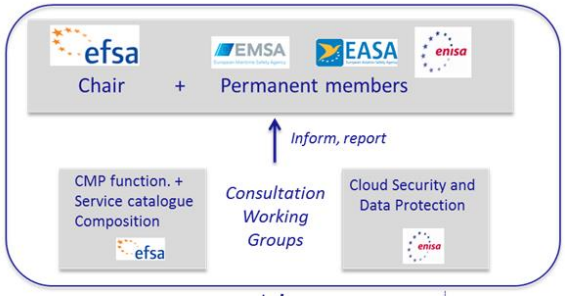
- Compliant and certified datacentre infrastructure and processes: ISO 27001, ITIL, BDSG, C5


- Governed by a specific Framework contract, a Governance Committee and a Security Working Group involving ENISA, European Data Protection Supervisor, and EC-DIGIT

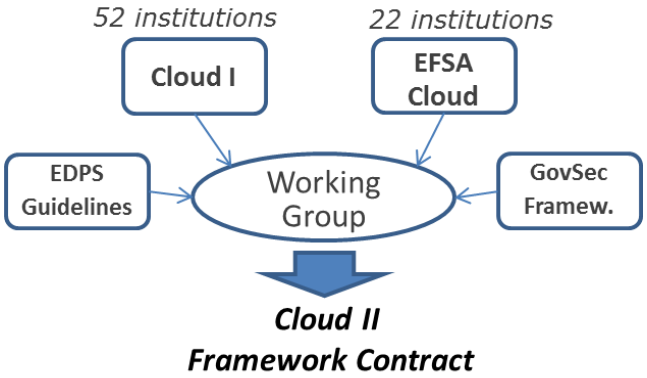
# GOVERNANCE OF THE EU COMMUNITY CLOUD

- Interagency FWC used by 22 agencies


- Governed by a body composed by EFSA, EMSA, ENISA, EASA and in consultation with EDPS




- Direct collaboration with EC (DIGIT) and the EDPS to define the 2020-2027 EU Cloud Framework (Cloud II)



# WHO IS USING IT

| Item | Stage |
|---|---|
| AHP DaaS Assessment | Finalized |
| DRaaS Assessment | Finalized |
| GDPR Assessment | Running |

| Item | Stage |
|---|---|
| Office 365 | Live |
| Lotus Notes to Exchange Online Migration Assessment | Ordered |

| Item | Stage |
|---|---|
| Azure Direct Usage | Live |
| Job Vacancy Analytics Consulting | POC |

| Item | Stage |
|---|---|
| Shared Electronic Platform for Initial Airworthiness Certification | Live |
| Community Cloud Managed Services | Live |
| DRaaS Assessment | Finalized |
| Identity and Access Management Vendor Selection | Finalized |
| Security Operation Center | POC |
| Community Cloud | Live |

| Item | Stage |
|---|---|
| Veeam BaaS | Live |
| AWS Direct Usage | Live |

| Item | Stage |
|---|---|
| CMP | Live |
| Community Cloud | Live |
| DRaaS | Live |
| WSG Genonma Sequencing Analytics | Live |
| AHP DaaS | POC |
| Office 365 | Live |
| Azure Direct Usage | Live |

| Item | Stage |
|---|---|
| Data Center Relocation Consulting | Finalized |
| AHP DaaS | Live |

| Item | Stage |
|---|---|
| Azure Direct Usage | Live |
| AWS Direct Usage | Live |
| Automatic Behaviour Monitoring | POC |
| Cloud Checkr | POC |

| Item | Stage |
|---|---|
| CMP | Live |
| Community Cloud | Live |
| OSS Safety certification and vehicle authorization Consulting | Live |

| Item | Stage |
|---|---|
| AWS Direct Usage | Live |
| Azure Direct Usage | Live |

| Item | Stage |
|---|---|
| CMP | POC |
| Community Cloud | POC |

| Item | Stage |
|---|---|
| Data Center Relocation Consulting | Finalized |
| AHP DaaS | Live |

***Data Centre relocation due to Brexit***

# WHY CLOUD IS SAFER THAN THE EFSA DATA CENTRE

| Security Aspect | EFSA (On-prem) | Community Cloud |
|---|---|---|
| *Physical security (e.g., target visibility, flood)* | *High exposure* | *Low exposure* |
| *External (cyber) threats* | *High probability* | *Low probability* |
| *Internal threats* | *Low-Medium prob.* | *Low-Medium prob.* |
| *Security Monitoring* | *8x5* | *24x7* |
| *Hardware upgrade* | *Every 4 years (aver.)* | *Continuous* |
| *System patching* | *Monthly* | *Continuous* |
| *Back-up, Disaster Recovery* | *Ad-hoc* | *Native, standard service (SLA)* |
| *Certification* | *On-going (ISO-27001)* | *ISO-27001, IEC-62443, BSI-C5, ITIL* |

# HOW ADDITIONAL RISKS ARE MINIMIZED

| Cloud Risk (*) | Mitigation factor |
|---|---|
| *Vendor lock-in: increasing costs* | *FWC based on a discount model with respect to market prices* |
| *Vendor lock-in: portability to a different provider* | *Broker model, contractual portability requirements, Disaster Recovery c/o a different provider* |
| *Network accessibility* | *Multi-redundant networks with different carriers + GEANT (planned)*<br>*EFSA connectivity with 1 major ISP + GEANT as back-up network* |
| *GDPR compliancy* | *Both primary and secondary DC are GDPR compliant and ISO-27001, C5 certified* |
| *Data geographical location (also for back-up, disaster recovery)* | *EU geo-location of data specifically defined as contractual requirement* |
| *Data portability and erasure* | *Data portability and erasure (after termination of the service/contract) specifically defined as contractual requirement* |
| *Protocol of privileges and immunities* | *Special EU institutions data protection protocol defined as specific contractual requirement* |
| *Outsourcing of staff* | *The use of agency cloud does not outsource staff or change EFSA operational staff responsibilities* |

*(*) ref.: BVL assessment report on Cloud computing*

# DAMA (DATA ANALYTICS MANANGEMENT) - PROJECT SCOPE

- Achieve **flexibility** and **scalability** in the EFSA scientific computational environments
  - Moving the Scientific Data Warehouse, DCF and R4EU to the Cloud (EU Agencies Cloud)
  - Allocating separate computational environments for test, editorial, production
  - Allowing on-demand scalability in terms of computational power and storage
  - Enabling future evolution for possible EU Agencies & Stakeholder involvement

- Ensure **independence** of Data Scientists & Data Managers from IT specialists
  - Allowing self-definition and self-managing of the environments
  - Introducing automation and managed services around the Cloud resources

- **Rationalization** and **decommissioning** of on premise infrastructure
  - Architecture rationalization and decommissioning of the on premise infrastructure
  - Revision of the software licensing model with the different software vendors

# WHERE WE ARE AND NEXT STEPS

- Project started in November
- Set-up of the Cloud environment on-going
- Testing activities will start in December (Performance, Usability)
- Environment migration in Q1 2019

# Q&A

## QUESTION TIME