Value of chemicals data

48th Meeting of the Management Board 14-15 December 2017

Key messages

The Management Board is invited to take note of a report on the value of chemicals data held by ECHA, and the opportunities related to the harmonisation of tools and formats.

In the constantly developing field of information technology and digitalisation, ECHA follows important Union initiatives in the area of the Digital Single Market and Data Union, including those related to the free flow of non-personal data or building the European Data Economy. ECHA strives to contribute to these priorities within its mandate.

The report builds on the information in the ECHA IT Master Plan. It complements it with a broader view on opportunities for the future to bring benefits to human health and environment and the competitiveness of European industry. It illustrates ECHA's current realisations in the field of data management, the potential of further deepening and expanding data access and knowledge cooperation with stakeholders, as well as integrating towards other chemical-, health- and environment related databases.

Background

ECHA is a de-centralised EU agency, managing EU internal market regulations in the area of industrial chemicals and biocides. In addition, ECHA has a regulator's role in the area of import and export notifications under the UN Rotterdam Convention.

In its role in implementing EU chemicals regulations[1], ECHA has become an organisation holding a wealth of scientific and regulatory data on chemicals and their uses that is unique in the world. ECHA has a reputation as pioneer e-Agency and provides its tools and expertise to all its stakeholders: industry, national Authorities and, increasingly, to other EU Agencies and OECD member countries' organisations.

Wealth of information unique in the world

REACH	CLP	BPR	PIC
Registration Evaluation Authorisation Restriction	Classification Labelling Packaging	Biocides	Prior Informed Consent
All chemicals >1 tonne per annum	All chemicals and mixtures	Active substances and biocidal products	Import/export of certain hazardous chemicals
	UN-wide standards		Rotterdam Convention

More to come?
Poison Centres: 20 million of mixture compositions by 2020
EUCLEF: Lists of chemicals informing how a substance is regulated at EU level
IPCheM: Access to chemical monitoring data collections in humans and the environment?
Common IUCLID database ECHA-CND-US-EPA?

A very large number of companies as well as over 100 stakeholder organisations acting across Europe use ECHA's services. ECHA maintains close working relations with other EU agencies and regulators in third countries.

By law ECHA has to be largely IT-based and by design it has chosen to be IT-based as much as possible, seeing IT as a key enabler of the regulatory activities and an indispensable instrument to elaborate and disseminate knowledge on chemicals. The Agency provides a range of IT tools free of charge to industry[2], Commission and Member States. ECHA's databases serve the needs of a wide range of stakeholders and its data intelligence capabilities are used by other regulators in the Member States, third countries and international organisations.

---

[1] REACH: Regulation (EC) No 1907/2006 on the Registration, Evaluation, Authorisation and Restriction of Chemicals
CLP: Regulation (EC) No 1272/2008 on the classification, labelling and packaging of substances and mixtures
BPR: Regulation (EU) No 528/2012 concerning the making available on the market and use of biocidal products
PIC: Regulation (EU) 649/2012 concerning the export and import of hazardous chemicals.
[2] 150.000 users in REACH-IT only with over 10 million dossiers received so far.

16 000	Chemicals registered
60 000	Comprehensive datasets sent by chemical companies
120 000	Chemicals classified and labelled
6 million	Classification dossiers
>2 million	Study summaries on properties and effects of chemicals
Big data?	~200 million IUCLID fields and ~1 million IUCLID attachments
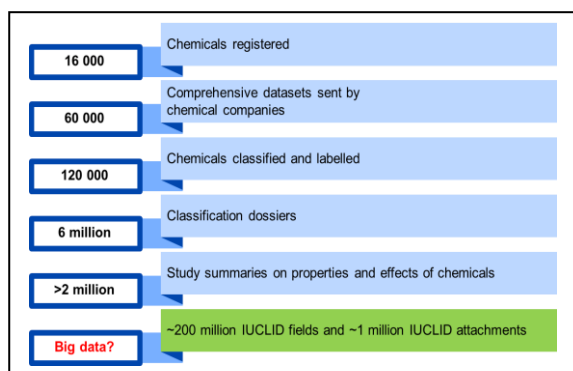
Most of the data is public in a layered and laymen friendly format. Secure connections with Commission and Member States allow efficient information exchange and work flows. For identifying substances with potential concern ECHA enriches the data with other public sources e.g. from EU agencies, U.S., Canada, open literature and applies advanced analytics. On top of this ECHA develops specific IT based methods to data mine, prioritise and assess the (un)structured information at hand. Finally, the data is developed into knowledge through the accumulation of regulatory decisions, opinions and evaluations, which are digitally linked to the data that originated them.

Currently, ECHA is taking on new areas of work based on changes in EU legislation or delegations from the European Commission, for example in the area of Poison Centres, Nano Materials or by preparing an EU Chemicals legislation finder.
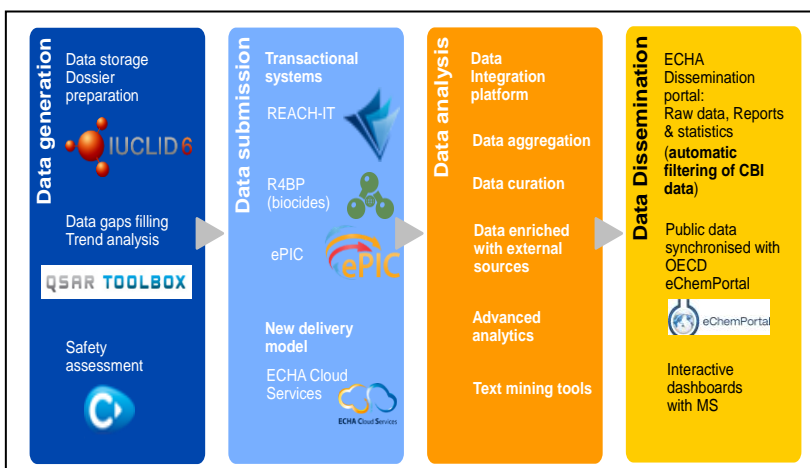
Rationale

ECHA's data value chain

The underlying regulatory framework for ECHA foresees that companies provide any data on substances in an electronic form. ECHA is responsible for defining the format and has the technical means to do so. ECHA, in cooperation with the OECD, has developed data formats to ensure interoperability of systems worldwide and facilitate electronic data exchange.

The result is that all industry IT tools provided by ECHA are fully compatible with OECD standard formats. Ontologies and controlled vocabularies are developed jointly to ensure that data are available in a structured format, are translated, ensure system interoperability worldwide and facilitate electronic data exchange in full. The backbone for this is an IT platform called IUCLID[3].

Beyond the formats, ECHA has developed tools, freely available, that enable third parties to manage their data on chemicals, giving them access to state-of-the-art risk assessment methodologies. Therefore, ECHA has become the reference provider of such capabilities for a large community of users.

The regulation of chemicals foresees a strong, harmonised cooperation between Member State Competent Authorities and ECHA. Shared central databases and IT systems allow the Member States and the Agency to collaborate directly on the data in a secure way.

Data format standardisation is a strength of the EU regulatory system and a key driver for data exchange across the EU and beyond. The standards are a model for other sectoral EU regulations and offer unique opportunities for further synergies. The European Parliament Committee for Environment Public Health and Food Safety recently recognised the potential of harmonised data and tools and set-up a working group on data harmonisation involving ECHA.

Digitalisation and interoperability promote not only automation, data mining and data analytics but also the re-use of common components across systems. It enhances data and knowledge
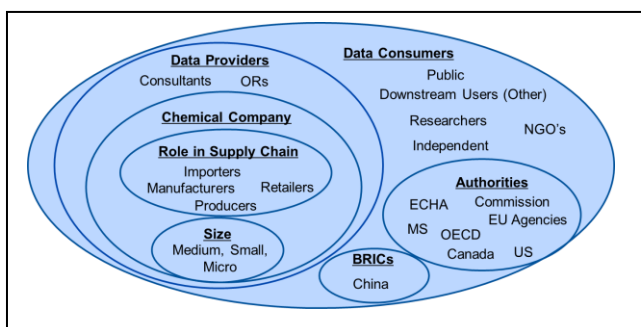
---

[3] IUCLID (International Uniform ChemicaL Information Database) is a software application to record, store, maintain and exchange data on intrinsic and hazard properties of chemical substances.

sharing and makes the location of the users through remote access to data less relevant, ensuring at the same time secure data exchanges between ECHA, Industry and Member States.

### *A brief look inside the data in ECHA's possession*

All data submitted to ECHA is natively in digital format. This ensures accessibility and automation in the further processing of that data. ECHA IT systems are built to process tens of thousands of registrations at deadline time and perform automated checks, enabling full compliance with the legal deadlines regardless of the input workload. ECHA uses already today advanced algorithms to support the submission process, for example through automated completeness check, fee calculation, etc. ECHA is also able to apply text mining by extracting plausible information from Chemical Safety Reports, free text documents, etc.

In implementing its regulatory strategy to identify chemicals of concern requiring further

regulatory action, ECHA invests significant efforts in data analytics, contributing to scientific approaches (e.g. grouping of substances, read-across methodologies etc.). To this end, data is aggregated (from dossier to chemical substance) for data mining, dissemination, etc. and also enriched with other public data e.g., from EU agencies, U.S., Canada. The success rate of ECHA's advanced algorithms is already high today: over 80% of substances identified have been later selected for regulatory action by Member States.

A physical outcome of the common screening and ranking of substances for regulatory action is the "map of the chemical universe**"**. Essentially the objective of this exercise is to use available information from various regulatory processes, as well as from external sources, to put substances into categories, with the ultimate aim to be able to decide if further regulatory action is needed or not, or whether further information needs to be generated first.
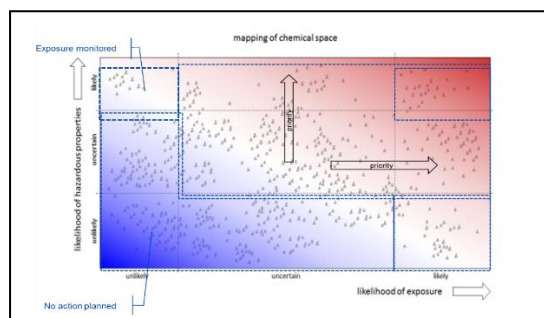

Finally, ECHA makes non-confidential information on substances accessible to the public: key information is extracted from large datasets and confidential data are automatically filtered out. The data is aggregated by chemical substance and put into an easy to understand format (info cards and brief profiles). Information on 120 000 chemicals is publicly available. The Dissemination Portal supports Industry (e.g. for Safety Data Sheet preparation, R&D & innovation), NGOs (in their information and watchdog role) and the public in getting a view on the chemicals they are exposed to in their daily lives. Dissemination of the data also feeds into the OECD eChemPortal, developed by ECHA in collaboration with the OECD.

ECHA started increased cooperation with other regulators for sharing tools and methods, aiming at benefits for all to identify quicker chemicals of potential concern.

How to move forward? - *Data Value Discovery study*

ECHA wants to better understand where it stands with its data management practices and in what direction it would best develop. Therefore, the Agency commissioned in 2016 a study by Gartner to clarify this and to compare it with the best practices in the market. The study looked also into the use cases that similar organisations are working on, as well as into how a high-level roadmap could look like to bring ECHA to the next level.

A key conclusion from the study is that there is a choice to be made in terms of a long-term strategy in order to adequately make technological choices and set the right foundation. The study presents three different ways to enhance the value of the data to the benefits of first ECHA and EU institutions, then trusted partners, or finally a broader audience e.g. academia, industry, public society. These options correspond to different ambition levels for a possible future, that all comply with ECHA's mission. Each option allows for delivery of additional benefits through the use of data and knowledge (e.g. analysis and interpretation tools) in different degrees and with a different focus:
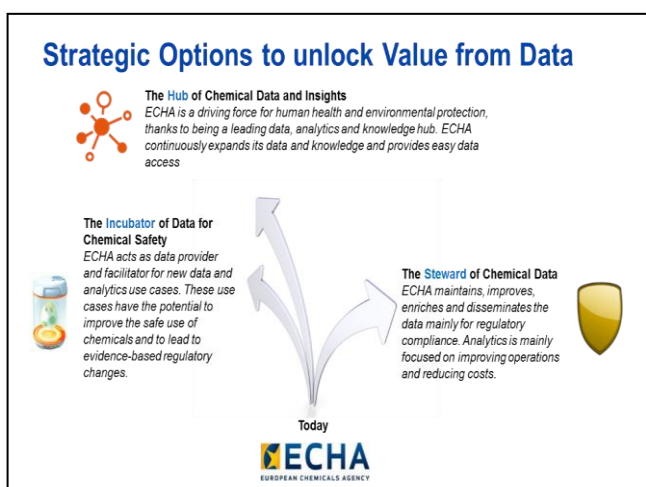
1) **Steward Option**, focusing on improving internal operations, reducing costs while still providing better advice on chemicals safety. ECHA continues to interact mostly with its current stakeholders.

2) **Incubator Option** to experiment, innovate and grow knowledge on chemicals safety, e.g. by sharing data or trying new technologies such as machine learning. ECHA opens data and its know-how to other trusted partners who themselves bring additional information and knowledge to the benefit of all by identifying and exploring synergies.

3) **Hub Option** to become a critical component for the eco-system in which it operates and the main source of knowledge and advice on chemicals. ECHA opens data[4], possibly integrated with other data sources, and its know-how to a broader audience so that the information can be re-used to the benefit of human health and environment protection.

The study found that ECHA already has a good foundation to develop any of these options.

Examples of the **Steward** role are the current identification and prioritisation of substances of potential concern, within the set of substances registered by industry, or further dissemination activities such as data reliability information.

Examples of the **Incubator** role are the exchange of data with the JRC or with other Agencies, which increase the data availability in support to opinions and decisions delivered by the various Authorities. Another example is the potential re-use of data and tools to support other chemical-based legislation.

Examples of the **Hub** role are the uses of ECHA's publicly disseminated data for consumer awareness, scientific research, development of new products, substitution of hazardous chemicals, all of which ultimately feedback into scientific progress, impact on protection of human health and environment, and innovation.

Our reflections

The Agency already achieved some concrete results:

- ✓ Stakeholders consider ECHA the most natural body for becoming a 'one-stop-shop' on information and knowledge on chemicals[5]. ECHA is a recognised source of information, intelligence and knowledge to authorities, helping them achieve more effective and

---

[4] To the extent possible: intellectual property rights considerations
[5] Outcome of the Feasibility study on the EU Chemicals Legislation Finder

efficient implementation of regulatory activities related to chemicals management

✓ A tangible example for ECHA being the 'one-stop-shop' on chemicals is the development, on behalf of the Commission, of an EU Legislation Finder which will improve the business environment for EU companies and SMEs in particular with regard to access to information on regulation applicable to a given chemical substance.

✓ IUCLID is increasingly recognised as the international platform to store/share data on chemicals substances collected by regulatory bodies worldwide.

✓ ECHA has already moved from business intelligence to analytics for decision making, e.g. by using prioritisation algorithms or mapping the chemical universe.

✓ ECHA is extending its data universe by using other data (lists of chemicals & unstructured data) and sharing its tools.

✓ ECHA is considering a shared data platform with third countries in the form of a project with US and Canadian authorities.

There is still discussion needed on a data strategy implementation roadmap, in particular on identifying the use cases that should be implemented first to better contribute to the objectives of the legislation. For example, sharing a platform for the pilot project with U.S. and Canadian authorities is more strategic for our regulatory work by bringing additional knowledge on chemicals of interest than adding dissemination features, even though a shared platform falls under the "Incubator" role which corresponds to the 2nd ambition level. It is probable that the roadmap, after detailed analysis, will incorporate some use cases that belong to any of the roles described above.

It is however important to set the long-term direction so that adequate IT technological choices are made that enable a possible evolution towards a more ambitious level, such as the 'hub' role. Similarly technical choices should enable ECHA to also address calls from the European Parliament to explore synergies and re-use of data with EMA and EFSA.

It would also be interesting to investigate whether partnerships could be built and whether ECHA could benefit from actions put in place at European level under the Commission's Digital Agenda e.g. in the field of standardisation and interoperability or by using shared EU-level resources. This would allow to benefit from experience and possible budget support in order to prepare for the long-term evolution of ECHA's data strategy and digital platform.

Conclusion

Against the background of the strategic considerations above, ECHA included as priority into its draft Programming Document 2019-23[6]:

Priority 3: Maximising the use of data and competences for the benefit of human health and the environment

---

In particular, ECHA will analyse what strategic opportunities the implementation of the EU digital agenda can provide, and how ECHA can contribute to it. ECHA thereby plans to take a proactive stance by making data and knowledge more readily available, enabling access to its growing databases so that additional value can also be created by other interested parties while respecting the applicable laws. This is not least to ECHA's own benefit as the Agency will not achieve everything by itself.

ECHA aspires to become a leading data, analytics and knowledge centre for chemicals safety:
✓ By continuously expanding the data and knowledge and providing easy access to relevant data
✓ By sharing its analytical tools and knowledge to enable others to process the relevant data so that additional value can be created e.g. by supporting the implementation of EU legislations related to the safe use of chemicals or by leading to evidence-based regulatory changes

---

[6] See agenda item 6.2

    ✓ By identifying with privileged partners and stakeholders possible synergies and efficiencies created by the further interaction and integration.

A phased approach for this is envisaged, taking into account the constraints in terms of budget and human resources. As part of the implementation of its next strategic 5-year plan ECHA plans to prepare a data strategy roadmap and yearly milestones.

Drawbacks

All IT-related aspects of the work of the Agency bring challenges in development, implementation and maintenance: information richness and diversity make data manipulation and data provision complex, requiring interpretation of the data and quality checks. At all stages, ECHA has to respect industry's ownership of the core data, while enabling rich access and collaboration with Authorities and needed stakeholders. In ECHA's very dynamic 'business environment', there are also constant changes in the underlying ICT technologies and in the potential new uses of data analytics.

Regarding future developments, there is however an absolute need to manage, develop and stay up to date in the evolving field of data management. Indeed, it is essential to 'unlock' the ECHA's databases to make REACH and CLP data available to third parties and ensure its best use. Moreover, the ambitious EU goals for digitalisation and open data as well as reusability of data ('Once Only' principle) confirm the e-Agency approach and encourage us to follow the path of strong data management engagement. This includes the work on format standardisation and data provision as well as the new areas of artificial intelligence and machine learning that will become more relevant in the next decade and could become beneficial for some regulatory tasks, to make the outcome more pertinent and save resources, for example in the area of (manual) completeness check.

Nevertheless, this data strategy, which on one hand fits well in ECHA's mandate and is a key pillar for implementing successfully its regulatory obligations, needs to take into account resources constraints. Therefore indicators and/or and success criteria could be set to validate the merits of the approach and justify the resources.

For questions: christel.musset@echa.europa.eu with copy to mb-secretariat@echa.europa.eu