ISES Europe Data Repositories and Analytics Working Group

Marissa Kosnik (Eawag, Switzerland)

ISES Europe 2024

https://ises-europe.org/

Data Repositories and Analytics Working Group members

- Marissa Kosnik
- Stylianos Kephalopoulos
- Peter Fantke
- Nicolò Aurisano
- Alberto Cusinato
- Amalia Muñoz
- Sani Dimitroulopoulou
- Jaroslav Slobodnik
- Jonathas De Mello
- Maryam Zare Jeddi
- Claudia Cascio
- Andreas Ahrens
- Yuri Bruinen de Bruin
- Lothar Lieck

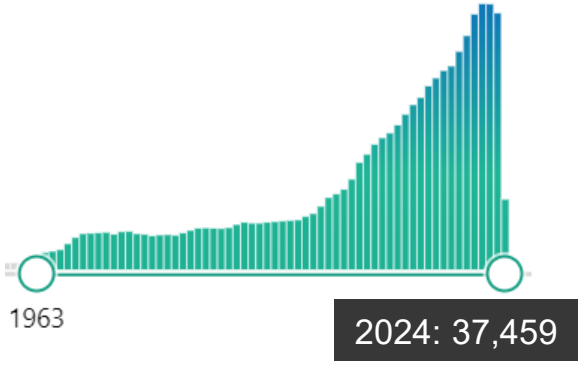
Exposure data are needed for policies and regulations

- Exposure assessment is a critical component of chemicals risk assessment and management with different types of exposure data needed
- Data are sought by various actors and stakeholders (policy makers, industry, academia, regulators, international organizations/bodies) for use across EU policies and regulations

The state of exposure data

- Difficult to find, combine, and reuse relevant information, keep up with research developments, and make use of the fast growing body of data
- With an aim to protect human health and environmental quality, improved exposure-related data generation, management, and usage would support ongoing national and international strategies

 # chemical-related publications in PubMed



Example European data repositories

efsa
European Food Safety Authority

norman
NORMAN Database System

ECHA
EUROPEAN CHEMICALS AGENCY

Goals

Strategic guidance for an integrated framework of European exposure data production, management, and usage

a) To provide an overview and map existing exposure data onto related requirements for data analytics and repositories across European regulations

b) To identify needs and ways forward for improved exposure data generation, management, and usage

c) To translate the identified needs into operational actions, increase regulatory uptake of exposure data in Europe, and outline an approach for advancing European data collection, management, and multi-use efforts with global initiatives

Goals

Strategic guidance for an integrated framework of European exposure data production, management, and usage

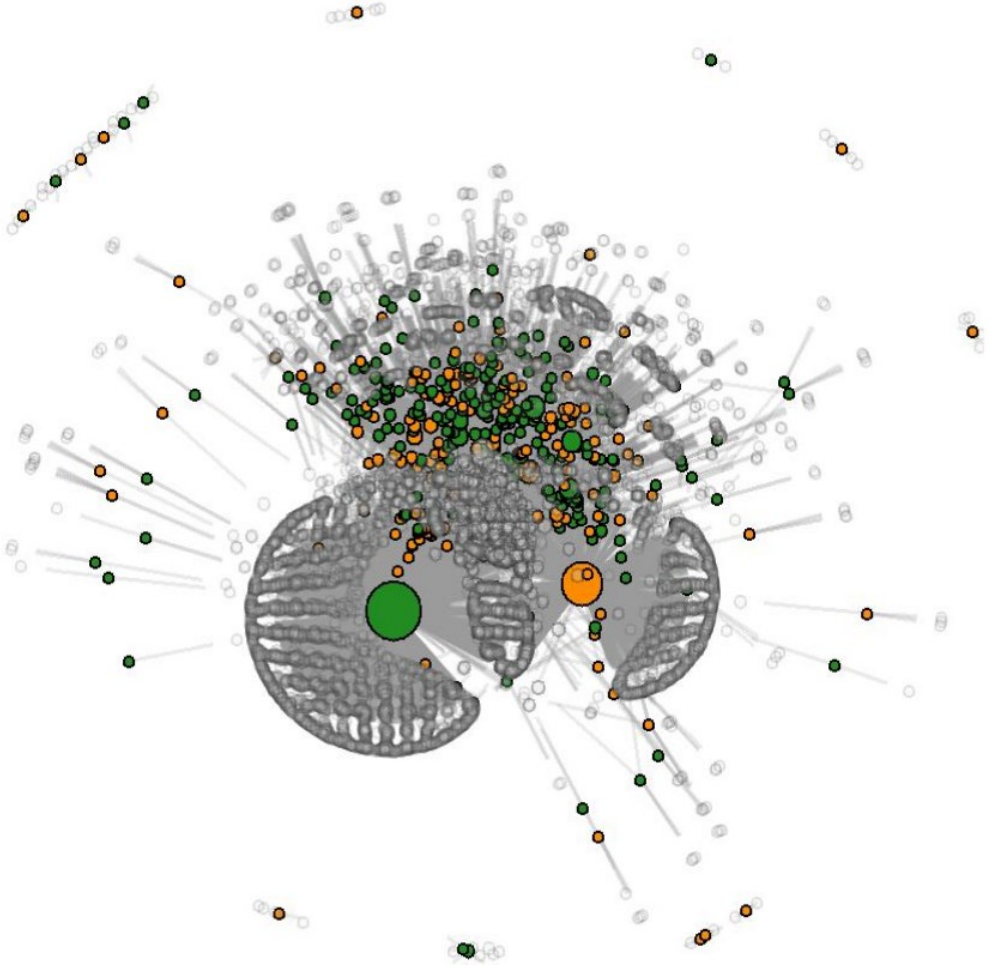
a) To provide an overview and map existing exposure data onto related requirements for data analytics and repositories across European regulations

b) To identify needs and ways forward for improved exposure data generation, management, and usage

c) To translate the identified needs into operational actions, increase regulatory uptake of exposure data in Europe, and outline an approach for advancing European data collection, management, and multi-use efforts with global initiatives

Developing a consistent exposure data terminology

Keywords from ~7000 “air pollution” and/or “particulate matter” articles



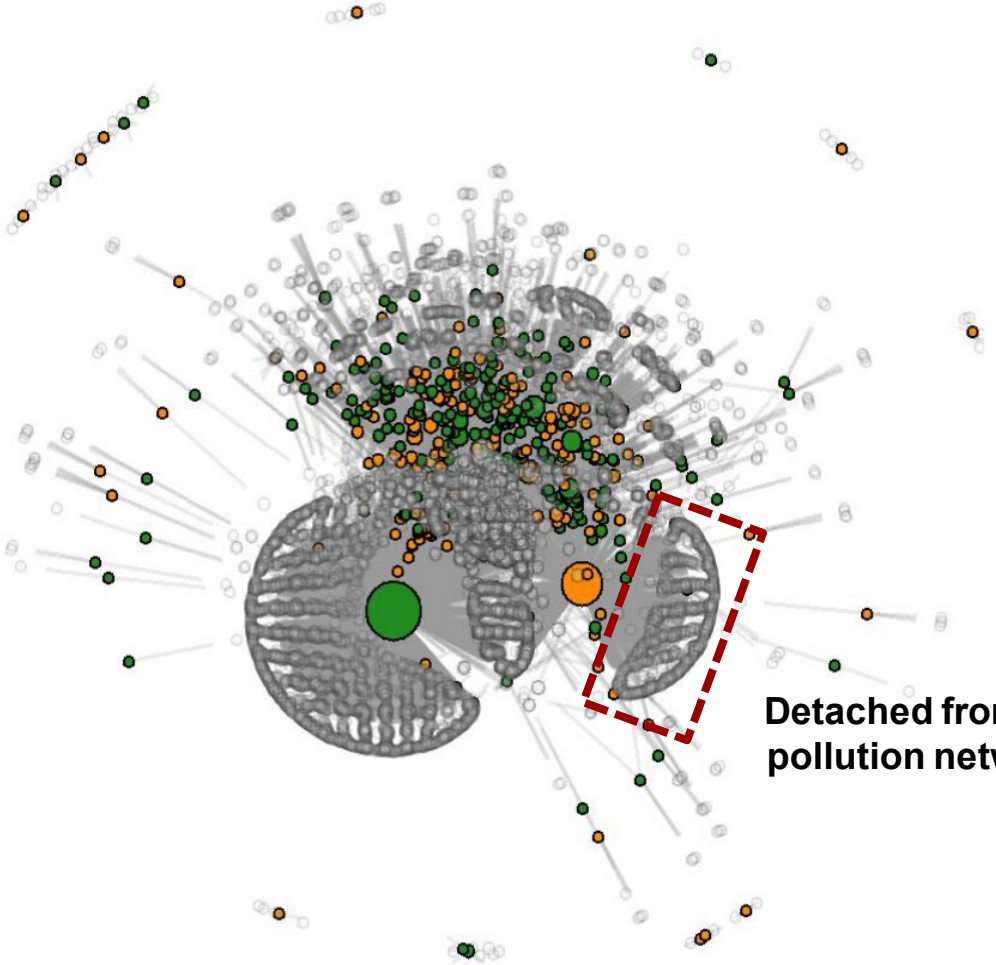
247 “air pollution” keywords

178 “particulate matter” keywords

Other keywords

Developing a consistent exposure data terminology

Keywords from ~7000 “air pollution” and/or “particulate matter” articles



247 “air pollution” keywords

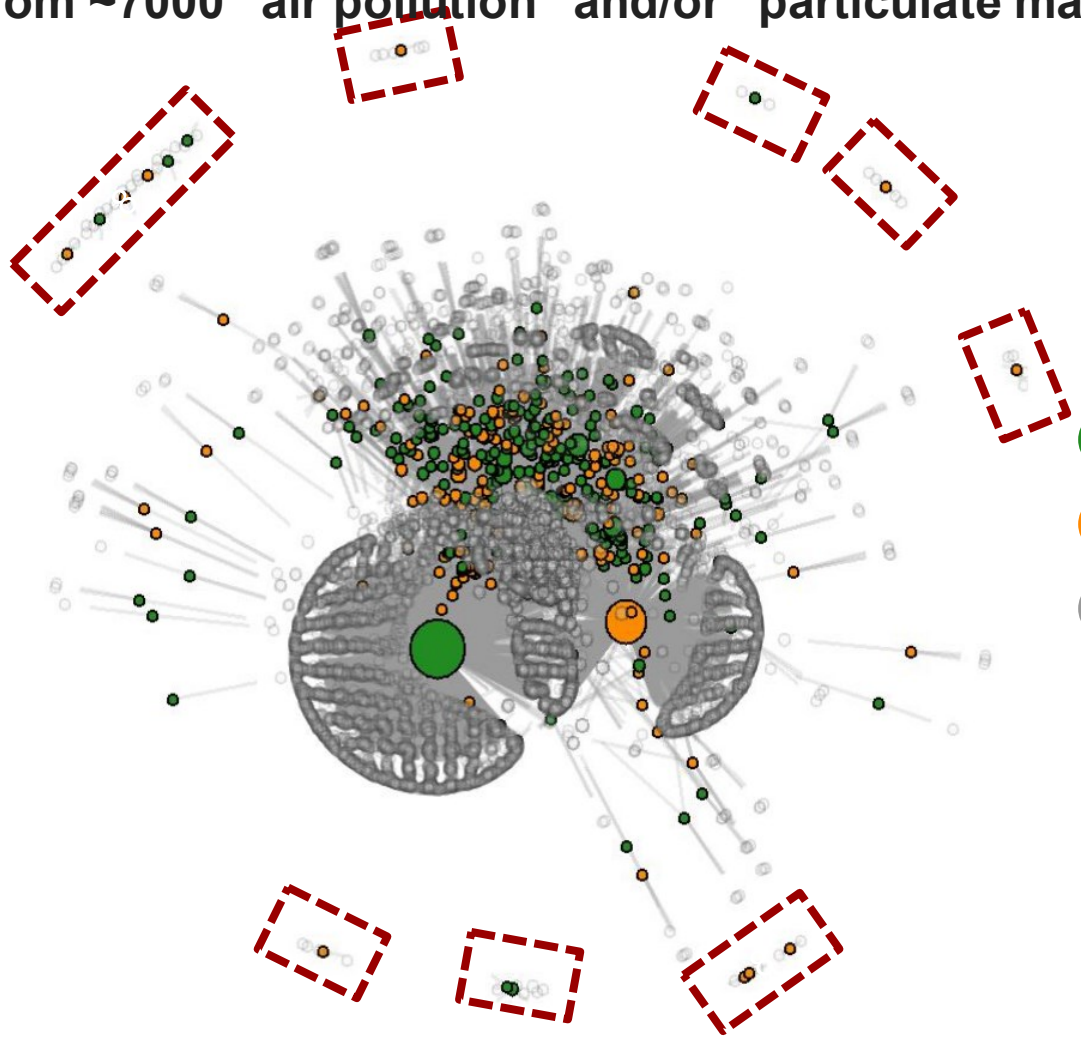
178 “particulate matter” keywords

Other keywords

Developing a consistent exposure data terminology

Keywords from ~7000 “air pollution” and/or “particulate matter” articles

Overly specific keywords



247 “air pollution” keywords

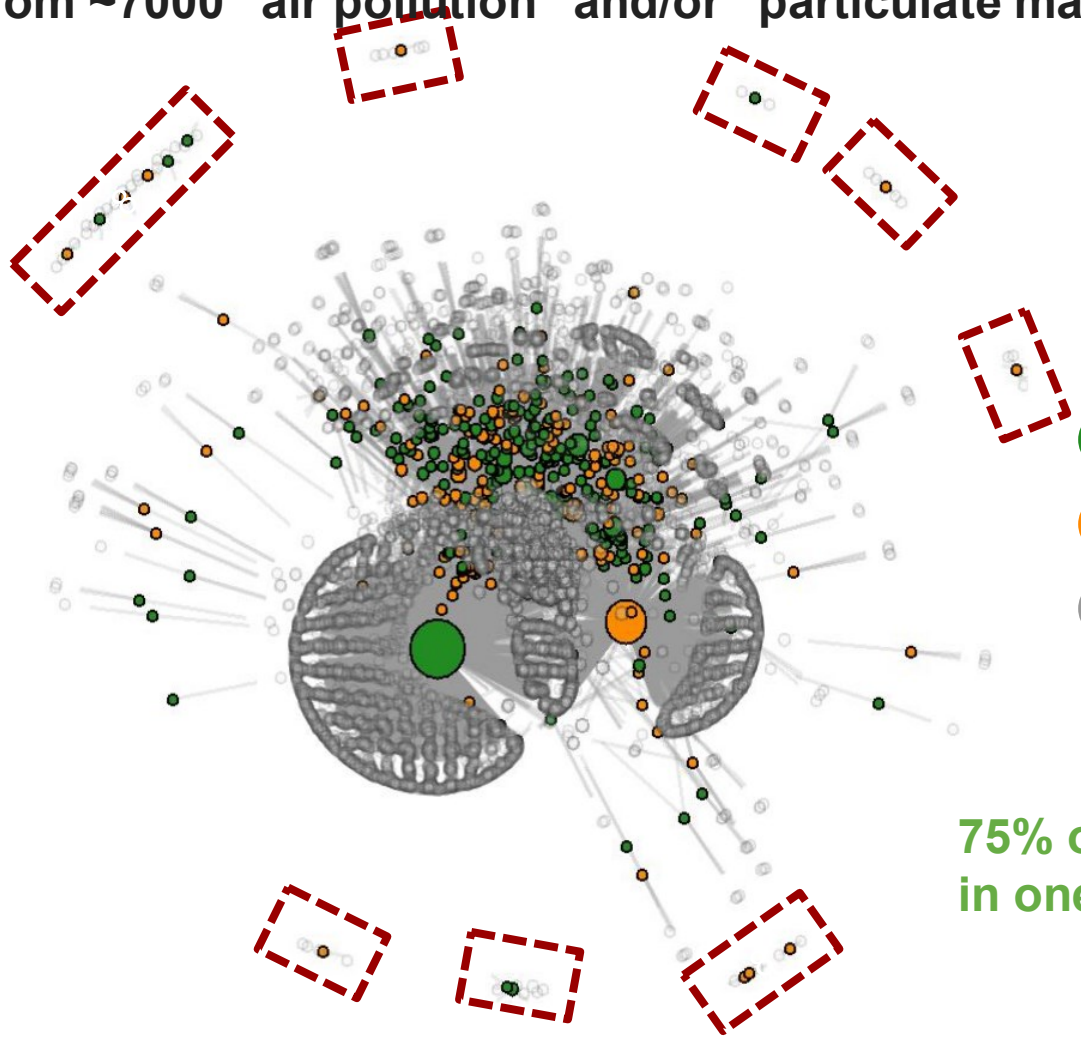
178 “particulate matter” keywords

Other keywords

Developing a consistent exposure data terminology

Keywords from ~7000 “air pollution” and/or “particulate matter” articles

Overly specific keywords



247 “air pollution” keywords

178 “particulate matter” keywords

Other keywords

75% of keywords only present in one research article

Developing a consistent exposure data terminology

Literature challenges	Data repository challenges	Policy challenges
• Studies utilize different keywords and ontologies • Different journals have different keyword/ontology requirements	• Different resources use different data identifiers	• European policies have their own vocabulary, which is minimally employed outside of a specific policy domain

EuroVoc terms

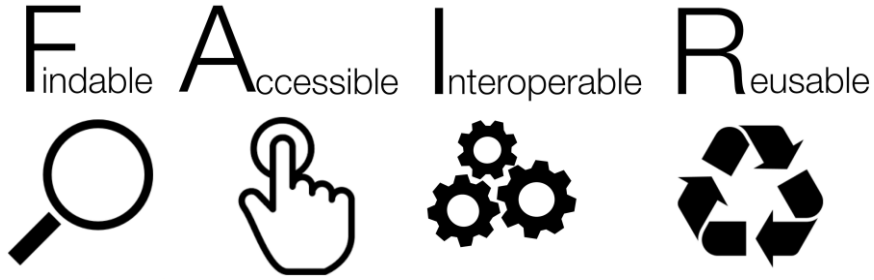
- INDUSTRY
  - industrial structures and policy
  - chemistry
    - chemical industry search
      - NT1 rubber industry search
      - NT1 glass industry search
        - NT2 glass search
        - NT2 glass fibre search
      - NT1 hydrogen production search
      - NT1 chemical product search
      - NT1 raw chemical industry search
      - NT1 chemical accident search
      - NT1 chemicals legislation search

Increased data quality, transparency, and availability

Springer Nature Data Availability Statements

Data availability statements are important because they support **validation, reuse** and **citation**

of research data.



Version History

DisGeNET

May 4, 2020

- **DisGeNET Database 7.0 released**
  - All data sources were updated
  - Risk allele of the disease variant now available for ClinVar, the GWAS Catalog and GWASdb
  - Added the expansion of the disease search using semantic similarity

May 13, 2019

- **Updated platform (v1.1.0)**
  - Added Disease-Disease Associations (DDAs)
  - Added DisGeNET REST API (v1.0.0)
  - Updated disgenet2r package

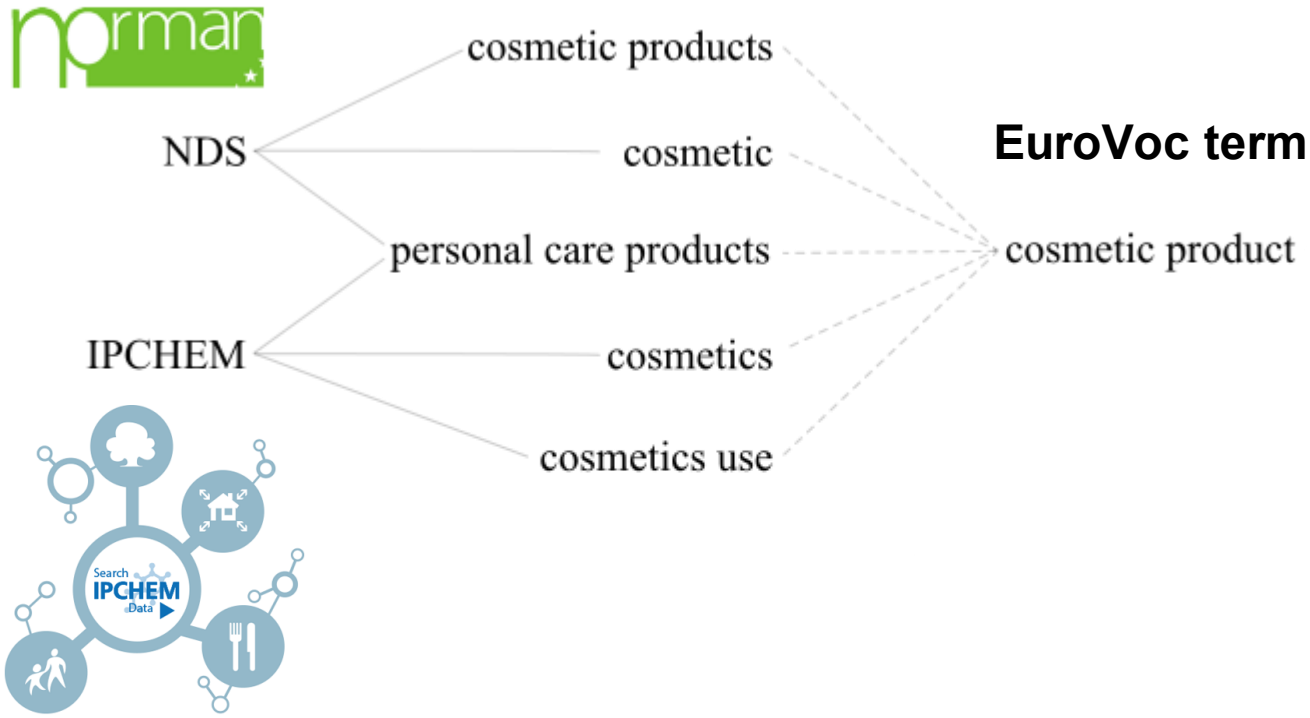
Increased data quality, transparency, and availability

Literature challenges	Data repository challenges	Policy challenges

• Authors do not always:
  • make data available on publication
  • provide adequate data descriptions to enable data reuse
  • provide data in readily usable formats
	• Not all repository data are extractable in bulk or easy to use formats
	• Data confidentiality can inhibit data availability
	• Interoperability across data platforms is low
		• Limited transparency and availability of data hampers assessments across EU policies
		• Enforcement that all data funded by the EU must be made publicly available

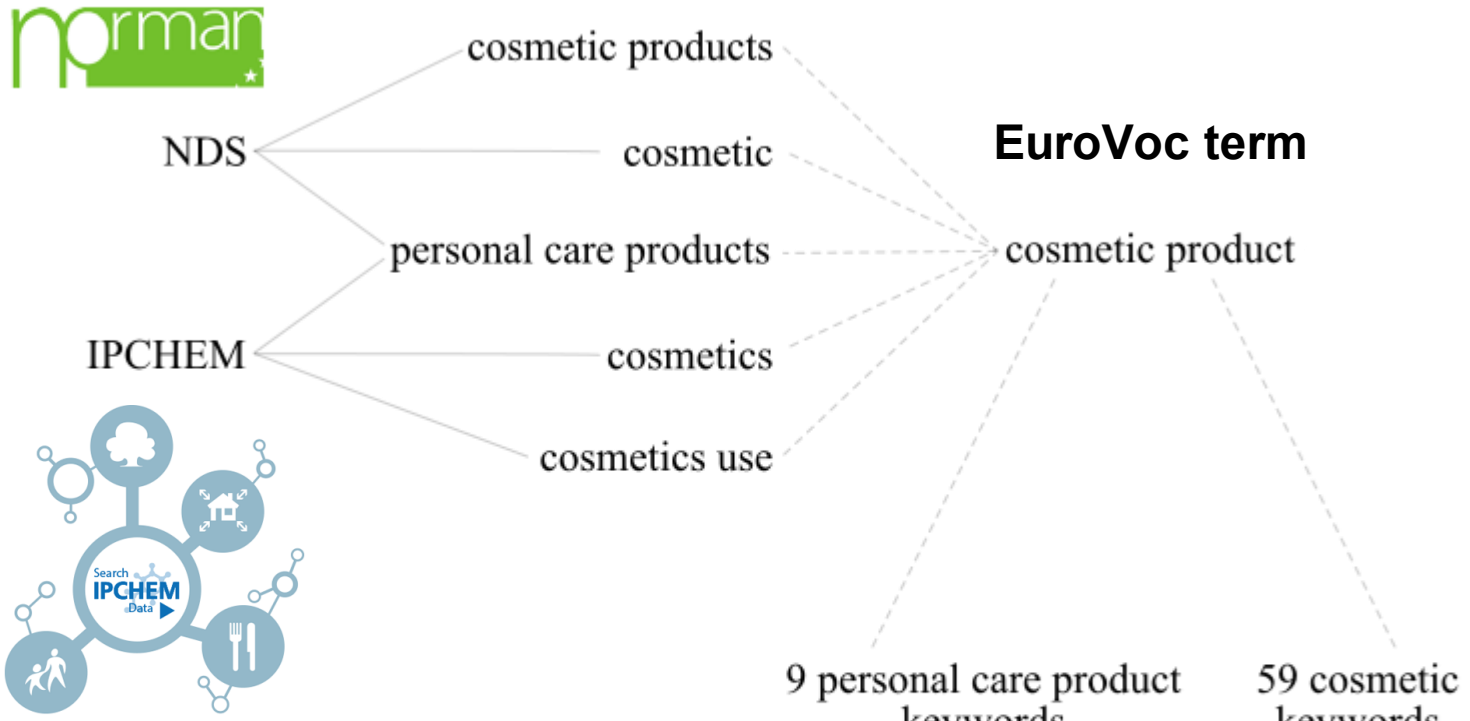
Increased data integration

Data repository information



Increased data integration

Data repository information



Literature information

287 literature articles

Increased data integration

Data repository information

NDS

cosmetic products

cosmetic

personal care products

IPCHEM

cosmetics

cosmetics use

EuroVoc term

cosmetic product

32 policies/directives

Regulation (EC) No 1223/2009
Commission Regulation (EU) 2016/1143
Commission Regulation (EU) 2019/831
Commission Regulation (EU) 2018/978
Commission Regulation (EU) 2019/1966
Fifth Commission Directive 93/73/EEC
Council Directive 2011/84/EU
Commission Directive 2008/42/EC
Directive 2003/15/EC
Council Directive 93/35/EEC
...

9 personal care product keywords

59 cosmetic keywords

Literature information

287 literature articles

Increased data integration

Literature challenges	Data repository challenges	Policy challenges
• Data quality can vary across studies • Different methodologies may be used across studies for data generation/analysis	• Few repositories use interoperable technologies with other data repositories • Meeting requirements for a repository may be more challenging than generating a new repository • Few tools for combining heterogeneous data (e.g., human biomonitoring, health and toxicity data)	• Cross-utilization of data across policies is challenging (missing metadata, data are not FAIR, missing a framework and tools for data exchange among policy makers)

Other strategic objectives

- Increased automation in data management
- Enhanced data storage and related infrastructure
- Innovative tools for improved data retrieval, handling, and analysis

EU and AI

High-throughput screening

EUROPEAN OPEN SCIENCE CLOUD

Goals

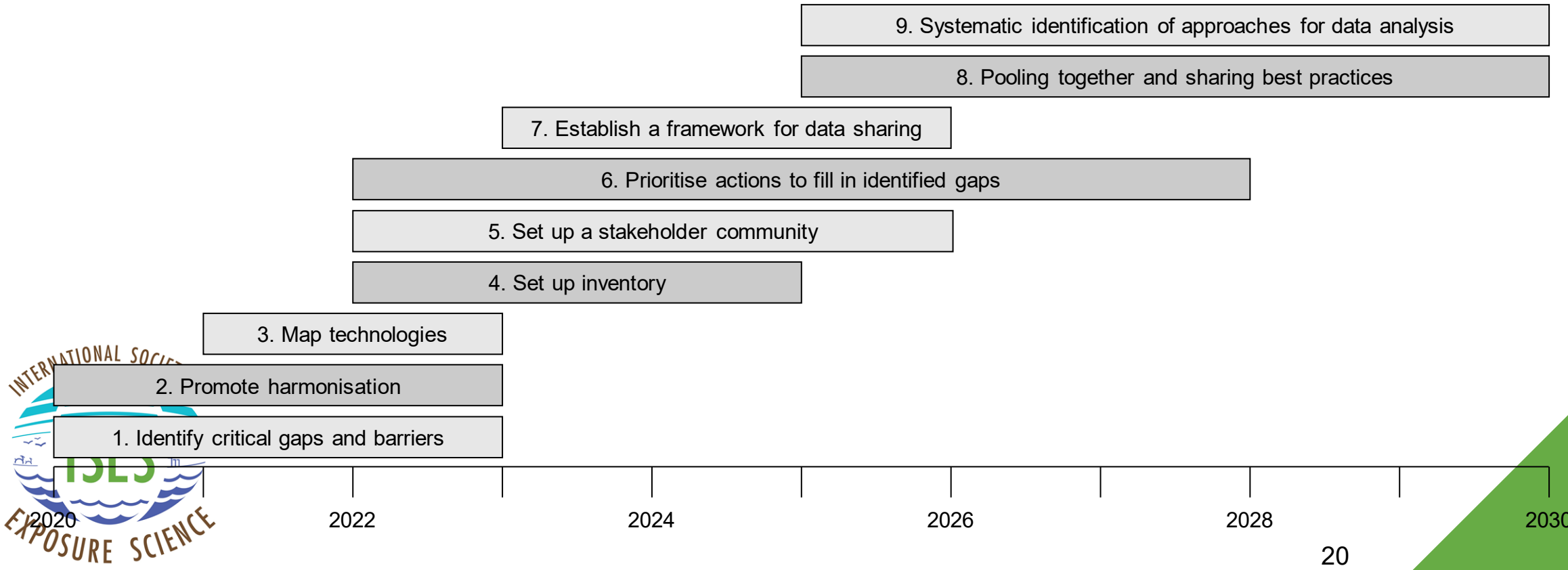
Strategic guidance for an integrated framework of European exposure data production, management, and usage

a) To provide an overview and map existing exposure data onto related requirements for data analytics and repositories across European regulations

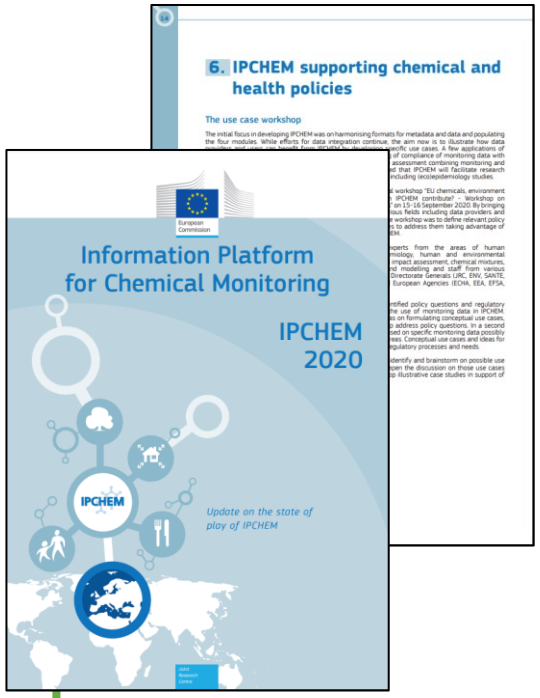
b) To identify needs and ways forward for improved exposure data generation, management, and usage

c) To translate the identified needs into operational actions, increase regulatory uptake of exposure data in Europe, and outline an approach for advancing European data collection, management, and multi-use efforts with global initiatives

Action plan overview



Action plan overview



Data covering various required:

- Timeframes
- Media
- Geographic areas

Lack of data due to:

- Legal/ethical barriers to data sharing
- Discontinued monitoring efforts
- Biases in substances monitored

9. Systematic identification of approaches for data analysis

8. Pooling together and sharing best practices

7. Establish a framework for data sharing

6. Prioritise actions to fill in identified gaps

5. Set up a stakeholder community

4. Set up inventory

3. Map technologies

2. Promote harmonisation

1. Identify critical gaps and barriers

1. Identify critical data-related gaps and barriers to their use

2020 2022 2024 2026 2028 2030

Action plan overview

www.nature.com/jes Journal of Exposure Science & Environmental Epidemiology

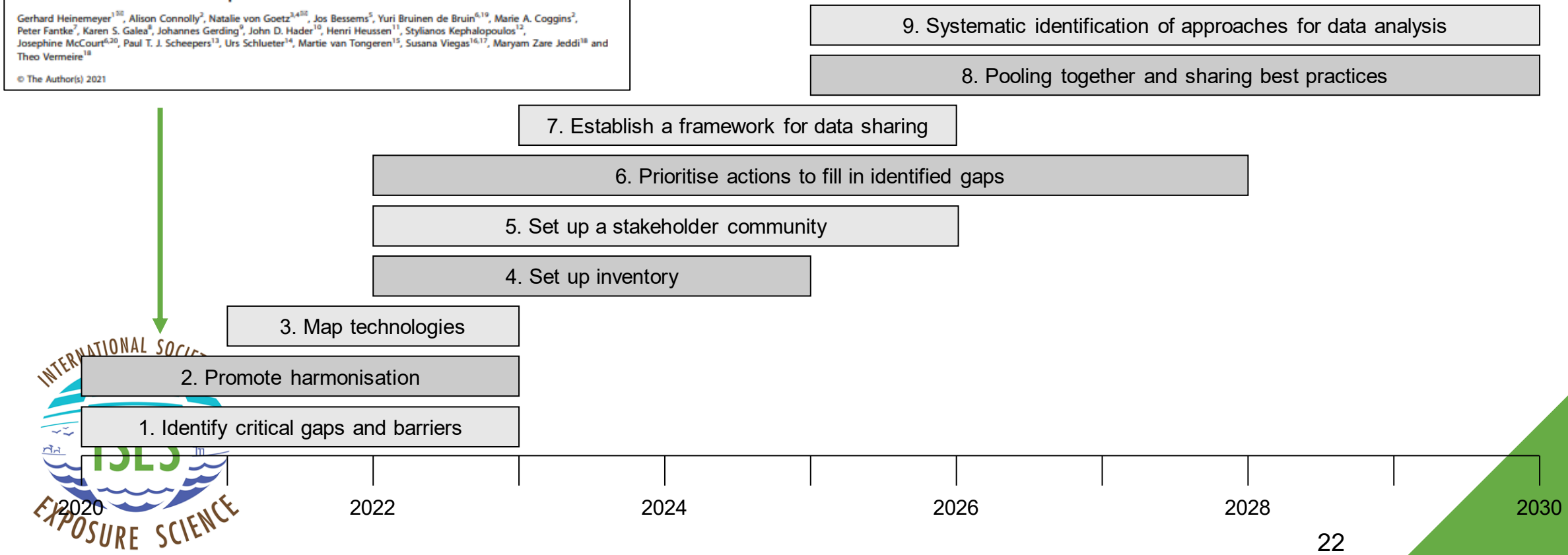
COMMENT OPEN

Towards further harmonization of a glossary for exposure science—an ISES Europe statement

Gerhard Heinemeyer[1✉], Alison Connolly[2], Natalie von Goetz[3,4✉], Jos Bessems[5], Yuri Bruinen de Bruin[6,19], Marie A. Coggins[2], Peter Fantke[7], Karen S. Galea[8], Johannes Gerding[9], John D. Hader[10], Henri Heussen[11], Stylianos Kephalopoulos[12], Josephine McCourt[6,20], Paul T. J. Scheepers[13], Urs Schlueter[14], Martie van Tongeren[15], Susana Viegas[16,17], Maryam Zare Jeddi[18] and Theo Vermeire[18]

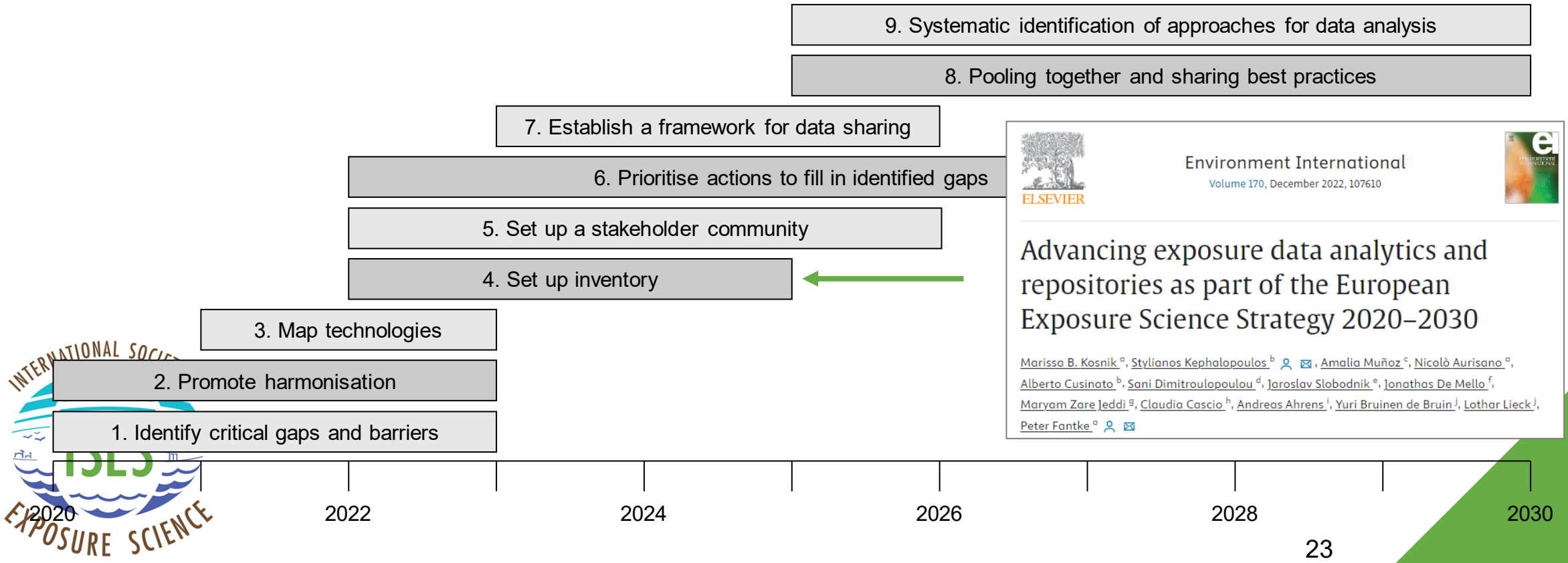
© The Author(s) 2021

2. Promote harmonisation of terms, templates for data reporting, standardised data collection methods, and indicators for data quality



Action plan overview

4. Set up an inventory of existing data sources and already undertaken case studies linking chemicals and health data



ELSEVIER

Environment International

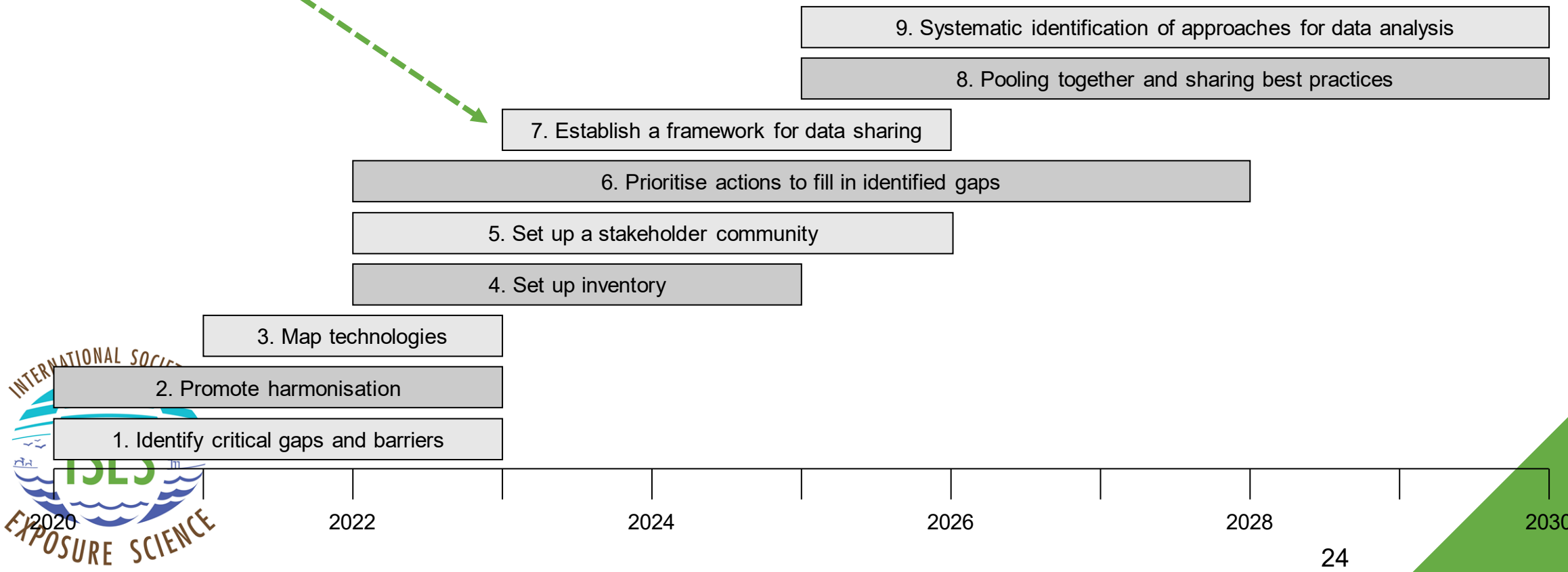
Volume 170, December 2022, 107610

Advancing exposure data analytics and repositories as part of the European Exposure Science Strategy 2020–2030

Marissa B. Kosnik [a], Stylianos Kephalopoulos [b], Amalia Muñoz [c], Nicolò Aurisano [a], Alberto Cusinato [b], Sani Dimitroulopoulou [d], Jaroslav Slobodnik [e], Jonathas De Mello [f], Maryam Zare Jeddi [g], Claudia Cascio [h], Andreas Ahrens [i], Yuri Bruinen de Bruin [j], Lothar Lieck [j], Peter Fantke [a]

Action plan overview

7. Establish common principles and **a framework for data** sharing and use/reuse among stakeholders



Action plan overview

A common open platform on chemical safety data

Will build on:

- OECD Harmonized Templates
- QSAR Toolbox
- IPCHEM
- Other widely used tools & platforms

Streamline data flows/assessments:

- ECHA
- EFSA
- EEA

9. Systematic identification of approaches for data analysis

8. Pooling together and sharing best practices

7. Establish a framework for data sharing

6. Prioritise actions to fill in identified gaps

5. Set up a stakeholder community

4. Set up inventory

3. Map technologies

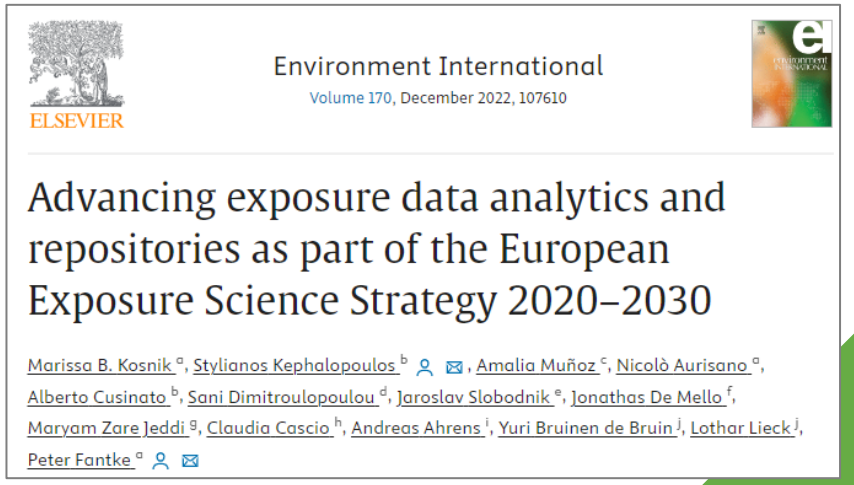
2. Promote harmonisation

1. Identify critical gaps and barriers

2020 2022 2024 2026 2028 2030

Conclusions

- Exposure-related data are essential for different decision contexts
- To ensure that exposure science meaningfully informs EU policies and regulations, need to align efforts for the generation, collection and use/reuse of exposure-related data
- The ISES working group identified strategic objectives:
  - Consistent vocabularies
  - Increased data transparency and availability
  - Enhanced data storage
  - Increased automation in data management
  - Increased data integration
  - Advanced tools for innovative data analysis

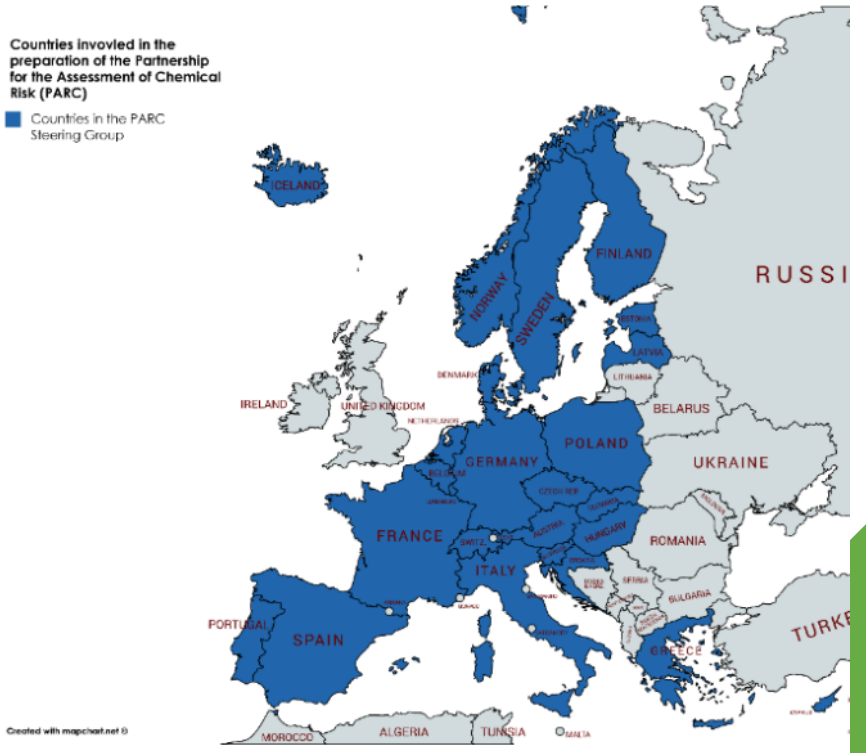


Many opportunities for involvement in the WG – join us tomorrow afternoon!

European partnerships

- Partnerships between various stakeholders allow for more efficient use of existing resources, knowledge, data, and best-practice transfer between countries and EU organizations
- PARC is a joint research and innovation programme under development to strengthen the scientific basis for chemical risk assessment in the EU
  - Brings risk assessors and managers together with scientists to accelerate methods development and data/knowledge generation and sharing
  - Data collection will be interoperable with main data repositories of EC and EU Agencies
  - Will develop tools and methods for combining heterogeneous data from available databases

European Partnership for the Assessment of Risks from Chemicals



Action plan overview

A common open platform on chemical safety data

PARC

8. Pooling together and sharing best practices for data integration and interpretation, standardised datasets and data formats

9. Systematic identification, application, and evaluation of novel methodologies, technologies, and **approaches for data analysis**