PhD-FSTM-2023-136
The Faculty of Science, Technology and Medicine

DISSERTATION

Defence held on 06/12/2023 in Esch-sur-Alzette

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG

EN *CHIMIE*

by

Dagny AURICH
Born on 01 February 1995 in Worms (Germany)

LUXTIME: HISTORICAL EXPOSOMICS IN THE MINETT REGION

Dissertation defence committee:

Prof. Dr. Emma Schymanski, Dissertation Supervisor
Professor, Université du Luxembourg

Prof. Dr. Andreas Fickers, Chairman
Professor, Université du Luxembourg

Dr. Christophe Hissler
Luxembourg Institute of Science and Technology

Assoc. Prof. Dr. Anneli Kruve
Assoc. Professor, Stockholm University

Assist. Prof. Dr. Frederic Béen
Assist. Professor, Vrije Universiteit Amsterdam ; KWR

AFFIDAVIT

I hereby confirm that the PhD thesis entitled “**LUXTIME: HISTORICAL EXPOSOMICS IN THE MINETT REGION**” has been written independently and without any other sources than cited.

Esch-sur-Alzette, 20.12.2023

D. Aurich

Dagny Aurich

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my PI Emma Schymanski, for her invaluable guidance, patience and support throughout my PhD. Her exceptional expertise and continuous encouragement helped me not only to complete this research and write this thesis but also shaped me as a researcher and a professional. She stands as a true mentor and a role model. I am extremely grateful to Andreas Fickers, who is not only the chair of my thesis defense but also played a pivotal role in the *LuxTIME* project. His consistent support and insightful perspectives have been a source of nourishment for my academic growth, teaching me a lot about the humanities. I am deeply indebted to my CET committee members Christophe Hissler and Lars Wieneke. Their dedicated time and invaluable intellectual contributions have profoundly influenced my development as a scientist. Moreover, I want to thank my thesis committee members Anneli Kruve and Frederic Béen, for taking their time and effort and for honoring me by judging my work. This endeavor would have been an impossibility without the enabling environment provided by the Institute of Advanced Studies at the University of Luxembourg. Their provision of resources and unwavering support allowed me to carry out my research at the LCSB. I would like to extend my sincere thanks to Aida Horaniet Ibanez for being my *LuxTIME* colleague over three years, giving me moral support at all times and captivating me with the intriguing world of data science and visualization. I am also grateful to my colleagues from the ECI and *LuxTIME* team, for their support, feedback and stimulating discussions inspiring me to think outside the box, from multiple perspectives to form a comprehensive and objective critique. Special thanks to Monika Thiel, who imparted the most valuable transferable skills that proved crucial for completing my doctoral studies. Thanks should also go to the REMIX and ZeroPM team as well as the Administration de la gestion de l'eau (AGE) for their collaboration and input. The privilege of participating in two NORMAN trials has been a remarkable experience, and I am thankful for the opportunity to be part of these endeavors. I am also thankful to Paul Thiessen and Evan Bolton, who keep on improving the PubChem chemical space and made the chemical stripes possible in the first place.

Lastly, I would like to thank my family and friends, giving me the unconditional support I needed over the past three years. A special mention to my boyfriend Patrik, whose encouragement kept me going through thick and thin. Not to be overlooked, my feline companion deserves acknowledgment for providing both entertainment and, in its own unique way, emotional support.

Table of Contents

Synopsis Abbreviations

For manuscript abbreviations, see the respective chapters in the Results section.

Abbreviation	Definition
CAS	Chemical Abstracts Service
CID	Chemical Identifier
C²DH	Centre for Contemporary and Digital History
DDT	Dichlorodiphenyltrichloroethane
ECI	Environmental Cheminformatics
HRMS	High Resolution Mass Spectrometry
IAS	Institute for Advanced Studies
LCSB	Luxembourg Centre for Systems Biomedicine
LIST	Luxembourg Institute of Science and Technology
LuxTIME	Luxembourg Time Machine
m/z	mass-to-charge ratio
NTA	Non-Target Analysis
PFAS	Per- and Polyfluoroalkyl Substances
PFOS	Perfluorooctanesulfonic acid

LIST OF ALL PUBLICATIONS

Statistics updated on 20 December 2023.

I. Manuscripts

Published

- Review (first author): **Aurich D**, Miles O, Schymanski EL. *Historical Exposomics and High Resolution Mass Spectrometry*. Exposome. 2021;1(1):1-15. DOI: 10.1093/exposome/osab007. 11 x cited (ResearchGate).
- Research paper [**not included]** (co-author): Talavera Andújar B, **Aurich D**, Aho VTE, et al. *Studying the Parkinson's Disease Metabolome and Exposome in Biological Samples through Different Analytical and Cheminformatics Approaches: A Pilot Study.* Anal Bioanal Chem. 2022;414(25):7399-7419. DOI: [10.1007/s00216-022-04207-z](10.1007/s00216-022-04207-z). 11 x cited (Google Scholar).
- Viewpoint paper (co-author): Arp HPH, **Aurich D**, Schymanski EL, Sims K, Hale SE. *Avoiding the Next Silent Spring: Our Chemical Past, Present, and Future.* Environ Sci Technol. 2023;57(16):6355-6359. DOI: [10.1021/acs.est.3c01735](10.1021/acs.est.3c01735). 10 x cited (Google Scholar).
- Commentary paper (shared first author): **Aurich D**, Horaniet Ibanez A, Hissler C, et al. *Historical Exposomics: A Manifesto.* Exposome. 2023;3(1):1-35. DOI: 10.1093/exposome/osad007.
- Method paper (shared first author): **Aurich D**, Horaniet Ibanez A. *How can Data Visualization Support Interdisciplinary Research? LuxTime: Studying Historical Exposomics in Belval*. Front Big Data. 2023;6:1164885. DOI: 10.3389/fdata.2023.1164885. 1 x cited (Google Scholar).
- Research paper (first author): **Aurich D**, Diderich P, Helmus R, Schymanski EL. *Non-Target Screening of Surface Water Samples to Identify Exposome-Related Pollutants: A Case Study from Luxembourg*. Environ. Sci. Eur. 2023;35(1):94. DOI: https://doi.org/10.1186/s12302-023-00805-5.

Submitted

- Paper (shared first author): **Aurich D**, Horaniet Ibanez A., Van de Maele J. *Simulating and Visualising Data in Environmental History: Airborne Dust Concentration from the Belval Plant in Luxembourg (1911-1997).* Submitted to the Journal of Digital History.
- Encyclopaedia entry [**not included]** (shared first author): **Aurich D**, Horaniet Ibanez A. *Environmental Science and Happiness.* In: Brockmann, H. & Fernandez-Urbano, R. (Eds.). (2024). Encyclopaedia on Happiness, Quality of Life and Subjective Well-being. Edward Elgar Publishing. DOI: [10.4337/9781800889675](10.4337/9781800889675). In Press. Release 2024.

II. Posters

- 17th Annual Conference of the Metabolomics Society (co-author): Talavera Andújar B, **Aurich D**, Schymanski EL. *Studying Indoor Dust Trial Samples from the NORMAN Network using Non Targeted LC-HRMS and Open Source Data Processing Platforms.* DOI: 10.13140/RG.2.2.36286.20801.
- ICNTS 2021 - International Conference of Non-Target Screening (first author): **Aurich D**, Diderich P, Schymanski EL. *Screening of Surface Water Samples for Contaminants in an Industrialized Area of Luxembourg using Non-targeted LC-HRMS and Open Source Data Processing*. DOI: 10.5281/ZENODO.5524046. 3rd poster prize. 155 views, 70 downloads.
- 18th Annual conference of the Metabolomics Society (co-author): Talavera Andújar B, **Aurich D**, Aho VTE, et al. *Identification of Unknown Chemicals in Parkinson's Disease Biological Samples through LC-HRMS and Different Cheminformatics Approaches*. DOI: 10.13140/RG.2.2.27647.69285.

III. Presentations and Other Outcomes

- Vandalf Journal Club 04.02.2022: **Aurich D**, Miles O, Schymanski EL. *Historical Exposomics and High Resolution Mass Spectrometry.*
- IMSC - International Mass Spectrometry Conference 29.08.2022: **Aurich D**, Diderich P, Schymanski EL. *Cheminformatics and High Resolution Mass Spectrometry in Historical Exposomics of the Minette Region.* DOI: 10.5281/zenodo.7025814. 153 views, 54 downloads.
- Van de Maele J., **Aurich D**, Horaniet Ibanez A. *Pollution-Timelaps.* https://minett-stories.lu/de/document/ch8-Timelaps.
- AIFA - Artificial Intelligence and the Future of Art Conference 29.11.2022: **Aurich D**, Horaniet Ibanez A, Wieneke L, et al. *LuxTIME.* 10.5281/zenodo.10037340
- SETAC - Society of Environmental Toxicology and Chemistry 33rd Annual Meeting 04.05.2023: **Aurich D**, Arp HPH, Hale SE, Sims K, Schymanski EL. *Chemical Stripes – Visualizing Chemical Trends of the Past Influencing Today*. DOI: 10.5281/zenodo.7885031. 399 views, 235 downloads.
- Perrera J, Schymanski EL, **Aurich D**, Thiessen P. *Our Chemical Past, Present and Future.* https://vimeo.com/862087332. 487 views.
- ICNTS - International Conference of Non-Target Screening 18.10.2023: **Aurich D**, Diderich P, Helmus R, Schymanski EL. *Non-Target Screening of Surface Water Samples to Identify Exposome-Related Pollutants: A Case Study from Luxembourg.* DOI: 10.5281/zenodo.10019205. 71 views, 38 downloads.

ABSTRACT

This doctoral thesis delves into the multifaceted realm of the historical exposome, encompassing the cumulative measure of environmental exposures experienced by individuals throughout their lifespan. The Luxembourg Time Machine (*LuxTIME*) project serves as the background for this interdisciplinary exploration, incorporating elements from history, data visualization, cheminformatics, and eco-hydrology. Its focus is on researching environmental influences on the health of the population in the Minett region, located in the southwest of Luxembourg. The thesis comprises six interdisciplinary manuscripts, each contributing to the overall understanding and advancement of exposomics research. The initial focus centred on the integration of cheminformatics within the *LuxTIME* framework, examining the identification of known and unknown chemicals in the environment using high resolution mass spectrometry (HRMS) techniques. The study encompasses both external and internal exposome factors, striving to encompass a comprehensive range of determinants influencing human health. In parallel, an extensive data inventory was compiled, consisting of quantitative and qualitative data from diverse sources such as archives, newspapers, and scientific institutes. To effectively explore and communicate research findings, data visualization tools are employed, providing a visual representation of the exposome data, and facilitating interdisciplinary knowledge sharing. Next, the thesis highlights the importance of delving into the past, emphasizing the relevance of historical data in understanding the impacts of chemical exposure on humans and the environment. Case studies are conducted to simulate and visualize historical dust exposure originating from a steel factory, utilizing mathematical models and quantitative industrial and environmental data from the Luxembourgish past. Additionally, the past and present management and regulation of persistent chemicals is discussed, looking at per- and polyfluoroalkyl substances (PFAS). The chemical stripes visualization, developed in this thesis, is used to illustrate the trends of chemical use and registrations over time, emphasizing the importance of studying chemical exposure and advocating for improved management practices. An essential component of the thesis revolves around non-target analysis (NTA), involving the analysis of Luxembourgish surface water samples from diverse locations using HRMS and NTA techniques. Through classification steps and comparisons with previous studies, potential exposome-related pollutants are identified, and recommendations are made for the development of future governmental monitoring lists. In conclusion, this doctoral thesis demonstrates how exploring the exposome in an interdisciplinary manner provides valuable insights and new perspectives by adding a historical dimension to the exposome concept. By delving into the past, employing advanced analytical

techniques, and innovative data visualization methods, a more comprehensive understanding of the exposome and its implications for human health is achieved.

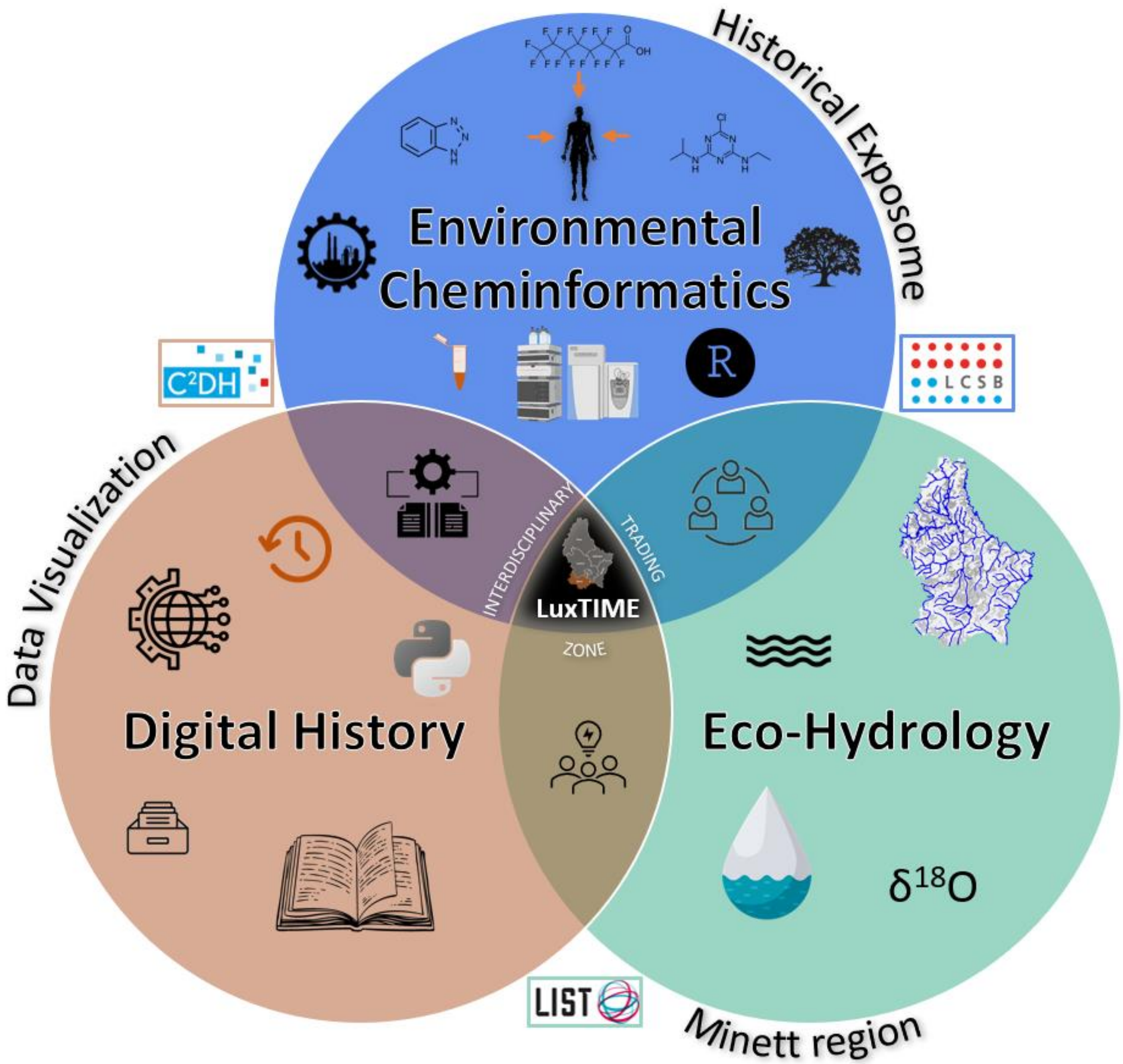


Figure 1: Interdisciplinary integration in the LuxTIME project.

Aims and Objectives

The primary objective of this doctoral thesis, conducted within the *LuxTIME* framework, was to comprehensively explore and enhance the understanding of the historical exposome, specifically investigating the influence of (past) environmental exposures on the health of individuals living in the Minett region and its surroundings. This research aimed to adopt and investigate the interdisciplinary nature of exposomics research, expanding the exposome concept to focus more on past evidence of human impact on the natural world. A particular emphasis of *LuxTIME* was on the key disciplines integrated within the *LuxTIME* team, namely environmental cheminformatics, eco-hydrology, history, and data science (see Figure 1). The underlying manuscripts contribute to the advancement of innovative methodologies and techniques for the analysis and visualization of exposome-related data. Moreover, this PhD project placed a significant focus on integrating past chemical history into exposomics workflows. Therefore, the project integrated the field of cheminformatics within the *LuxTIME* framework, utilizing HRMS and NTA techniques to identify potential exposome-related chemicals. The research in this thesis was centered on the following main objectives:

- To (re)define the concept of the historical exposome, focusing first on the chemical perspective and subsequently incorporating interdisciplinary viewpoints from the humanities and natural sciences
- To explore various ways of obtaining historical environmental data, including different sample types, archival data and analytical techniques
- To examine the use of simulations and their limitations in cases where historical data or samples were not available
- To compile and analyse qualitative and quantitative exposomics data from diverse sources, creating an extensive *LuxTIME* data inventory comprising various data from Luxembourg
- To explore and communicate research findings and map interdisciplinary knowledge using data visualisation as the main navigation tool
- To examine the relevance of exposomics research by investigating trends in chemical usage, overall chemical numbers and looking at regulatory measures, particularly for persistent chemicals
- To explore the importance of monitoring environmental pollutants in Luxembourgish surface water samples via NTA and HRMS, complementing routine targeted efforts

In general, this doctoral thesis seeks to discover new facets of the past exposome by redefining the concept, fostering interdisciplinary collaboration and providing insights into the historical and contemporary implications of chemical exposure in the context of exposomics research.

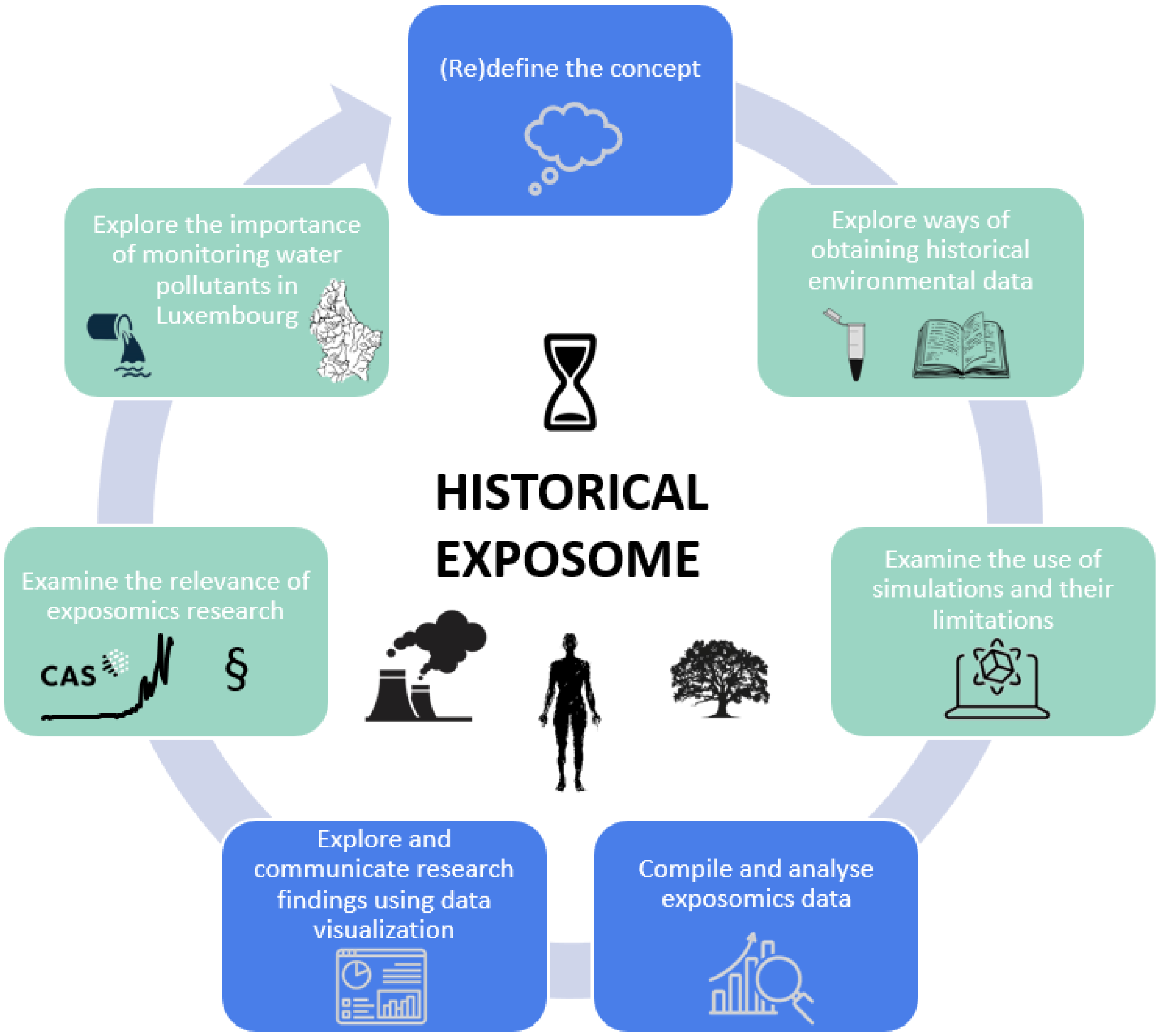


Figure 2: Research progression and objectives in the LuxTIME project.

SYNOPSIS

I. The Luxembourg Time Machine – LuxTIME

The Institute for Advanced Studies (IAS),[1] established in 2020 by the University of Luxembourg, aims to enhance interdisciplinary research efforts. Acting as a catalyst, the IAS endeavours to overcome barriers between scientific disciplines and sectors. One funding instrument offered by the IAS is the Audacity program, which supports collaborative projects at the forefront of interdisciplinary science characterized by an exploratory and audacious nature. The *Luxembourg Time Machine* project[2] named *LuxTIME* is a collaboration involving the Centre for Contemporary and Digital History (C²DH) and the Luxembourg Centre for Systems Biomedicine (LCSB) at the University of Luxembourg, along with the Luxembourg Institute of Science and Technology (LIST). The primary objective of *LuxTIME* was to investigate the feasibility of developing a national platform known as the *Luxembourg Time Machine*. This platform should provide scientists and stakeholders with the means to explore the intricate history of Luxembourg using digital tools and data from diverse disciplines and fields.

By constructing a digital dataset that incorporated information from three distinct fields and scientific perspectives—namely eco-hydrology, environmental cheminformatics, and history—*LuxTIME* aimed to employ a local case study, specifically the industrialization of the Belval/Minett region, as a testing ground for methodological and epistemological reflections on studying the impact of environmental changes on the local population's health. This investigation adopted a regional and long-term perspective. By combining historical evidence from the natural science and humanities perspective, *LuxTIME* enabled novel approaches to studying the past: through the combination of archival evidence providing contextual information and scientific evidence derived from chemical, biological, or medical investigations, the project pioneered the interpretation of 'big data of the past' within a truly interdisciplinary framework. The Belval case study served as a critical assessment of the analytical potential of a multi-layered research design that could potentially be expanded into a national case study, which could culminate in the creation of a *Luxembourg Time Machine* incorporating diverse data types from numerous institutions. The *Luxembourg Time Machine* project originated as a 'spin-off' of a larger European research endeavour called the *European Time Machine*,[3] which falls under the Future Emerging Technologies Flagship. The C²DH was among the founding members of this initiative.

Given the interdisciplinary nature of the *LuxTIME* project, there were methodological and epistemological challenges in establishing a common understanding among the research team. To address this, the project employed the concept of a *trading zone* and facilitated the development of 'interactional expertise' within the interdisciplinary team. Meetings and workshops were held to identify 'boundary objects' and foster a co-design process that clarified research questions, methodologies, and expected outcomes within this interdisciplinary context.

The initial PhD projects within *LuxTIME* specifically focused on the history of contamination and environmental impact associated with industrialization in the Minett region. This specific doctoral project – undertaken within the *LuxTIME* initiative – examined the presence of historical contamination in the environment of Luxembourg through chemical analysis under the title 'Historical Exposomics'. Non-target high resolution mass spectrometry techniques, in combination with open source cheminformatics tools, were employed to identify contaminants of particular relevance in various contexts. The outcomes were visualized and interpreted using (historical) data visualization tools in collaboration with Aida Horaniet Ibanez, the doctoral candidate from C^{2}DH. The project website can be found at https://luxtimemachine.uni.lu/.

II. The Exposome

The exposome concept refers to the cumulative measure of all environmental exposures, encompassing external and internal factors, to which an individual is subjected throughout their lifespan.[4] It takes into account various factors such as chemical substances, biological agents, physical agents, lifestyle factors, and socioeconomic factors that can impact human health. By adopting an exposome perspective, researchers aim to gain a holistic understanding of the entirety of environmental exposures and their interplay in relation to health outcomes.

The term 'exposome' was first introduced by Christopher Wild in 2005[4] to expand the focus of research beyond genetic factors and investigate the broader environmental influences on human health. Since then, the exposome framework has gained significant attention in the field of environmental health research.[5] Gary Miller and Dean Jones extended the concept in 2014[6] to its commonly used definition today:

"The cumulative measure of environmental influences and associated biological responses throughout the lifespan, including exposures from the environment, diet, behaviour and endogenous processes" [6]

There are two main components of the exposome: the external exposome and the internal exposome. The external exposome includes exposures from the external environment, such as air pollution, water quality, diet, physical activity, and occupational exposures.[6] In contrast, the internal exposome comprises the biological responses and modifications that occur within the body as a result of exposure, such as metabolic products, DNA adducts, and oxidative stress markers.[7] The exposome concept aims to capture the complexity of environmental exposures and their potential interactions, recognizing that individuals are exposed to multiple factors simultaneously and that these exposures can have cumulative effects over time. Exposure signals, which may be transient in nature, pose the possibility of health responses emerging years after the initial exposure. By considering the entirety of an individual's environmental exposures, the exposome concept provides a more comprehensive understanding of the determinants of health and disease.[7]

III. The Historical Exposome

The exposome framework aims to capture the totality of environmental exposures throughout an individual's life, including past exposures, and their potential impacts on health. The extension of the concept acknowledges that historical exposures influence health outcomes in the present and emphasizes the importance of considering long-term cumulative effects.

To delve deeper into the historical aspects of the exposome, researchers often rely on retrospective assessments, such as historical records, archival data, and biomarkers in preserved samples, to reconstruct past exposures and understand their potential health implications. By integrating historical data and environmental exposures, researchers can gain insights into the relationships between past exposures and current health outcomes.

LuxTIME aimed to study the so-called 'historical exposome' of the local population in the Minett area. The first (review) paper included in this thesis represents the first publication arising from the *LuxTIME* project, which aimed to investigate the chemical perspective of the historical exposome. Therefore, further introduction of this aspect, including state of the art literature at the time, is given in the respective manuscript (RESULTS, CHAPTER I) "*Historical exposomics and high resolution mass spectrometry*".[8] The commentary "*Historical Exposomics: A Manifesto*"[9] is the second paper of this thesis (RESULTS, CHAPTER II), which provides insights into the interdisciplinary discussion around the 'historical exposome' that took place within the *LuxTIME* framework, suggesting an interdisciplinary approach to study the Anthropocene, making use of natural and social archives. It also includes a concept criticism looking at the scientific development in exposomics research.

IV. Cheminformatics

The field of cheminformatics plays a crucial role in connecting the chemical perspective into exposome research. Cheminformatics involves the application of computational and data-driven approaches to analyse chemical data, including the identification, characterization, and prediction of chemical compounds and their properties.[10] In the context of exposome research, cheminformatics provides valuable tools and methods for analysing and interpreting the chemical exposures that individuals experience.[11] One important aspect of exposome research is the identification and quantification of chemical substances to which individuals are exposed. Cheminformatics techniques, such as chemical structure elucidation, spectral analysis, and database mining, can aid in the identification and classification of chemical compounds present in environmental samples, biological matrices, and other sources.[10] These techniques enable researchers to link specific chemical entities to exposure events and understand their potential health implications.

Cheminformatics also facilitates the integration and analysis of large-scale chemical data generated in exposome studies. With the advancements in high-throughput screening technologies and the availability of large, open chemical databases like PubChem,[12] researchers can employ cheminformatics approaches to process and analyse massive datasets containing information on chemical structures, properties, and activities. These datasets also encompass valuable information from patents and literature (used in RESULTS, CHAPTER IV and V),[13,14] presenting a significant challenge due to the substantial size of patent files and the diverse nature of literature data.

In the context of cheminformatics, the open exchange of data is of high importance as open data fosters transparency and reproducibility, allowing the scientific community to openly access and validate the underlying information, methods, and results. Moreover, open data encourages collaboration and accelerates scientific progress. This collaborative approach promotes interdisciplinary insights, fosters innovation, and enables the discovery of new patterns and correlations that may have gone unnoticed otherwise. Furthermore, open data promotes efficiency (possibility to build upon prior work) and avoids duplication of efforts, saving time and resources. Given the history of closed data systems, such as the Chemical Abstracts Service (CAS) with limitations in terms of accessibility and availability, open data initiatives like PubChem have revolutionized the field by providing a vast and openly accessible database for chemical information. This openness democratizes access to (chemical) data, enabling researchers worldwide to contribute, access, and leverage the collective knowledge, leading to more comprehensive and robust research outcomes.

The addition of the cheminformatics perspective to the *LuxTIME* team has provided significant contributions to both interdisciplinarity and exposome research, as this PhD project was carried out within the Environmental Cheminformatics (ECI) group at the Luxembourg Centre for Systems Biomedicine (LCSB). The ECI group specifically concentrates on the identification of known and unknown chemicals in the environment using open source cheminformatics tools to examine their impact on human health. This research focuses not only on the internal exposome but also on the external exposome, aiming to study a comprehensive range of factors.

V. High Resolution Mass Spectrometry and Non-Target Analysis

High resolution mass spectrometry (HRMS) is a powerful analytical technique that has gained significant prominence in cheminformatics (and in exposome research) and is the predominately used method in the ECI group. HRMS offers a comprehensive and sensitive approach for the identification and quantification of a wide range of chemical compounds, enabling the assessment of the (chemical) exposome at a high level of detail.[11] HRMS utilizes the principles of mass spectrometry to measure the mass-to-charge ratio (*m/z*) of ions generated from analytes in a sample. More details on the technique itself can be found in the RESULTS, CHAPTER I "*Historical exposomics and high resolution mass spectrometry"*.[8] The high resolution of the mass spectrometer allows for precise determination of the exact mass of ions and their fragments, enabling accurate identification and differentiation of chemical compounds. This capability is particularly valuable in exposomics research, where the complexity and diversity of environmental exposures necessitate the analysis of numerous chemical species. In the context of exposome research, HRMS can be applied to various sample types, including biological samples, environmental matrices, and archival materials, as introduced in RESULTS, CHAPTER I.[8] By coupling HRMS with advanced data analysis and (open source) cheminformatics tools, both known and unknown chemicals present in the samples can be identified and characterized. The obtained mass spectral data can be further processed using databases such as PubChem[12] and spectral libraries like MassBank[15] to aid in compound identification and annotation (for a detailed discussion of databases, see the RESULTS, CHAPTER I[8]). One of the key advantages of HRMS for a wide range of research is its ability to provide a comprehensive coverage of the growing chemical space (204 million substances in the CAS registry in 2023[16]). It allows for the detection of various classes of chemicals, including environmental pollutants, human metabolites, pharmaceuticals, and natural products. Moreover, HRMS can facilitate retrospective analysis of stored samples,[17] enabling the investigation of historical exposures and their potential long-term effects on human health, which was of interest for this project.

There are different ways to screen for chemicals using HRMS. In addition to targeted analysis, non-target analysis (NTA) is an important approach in exposomics research. Its application in *LuxTIME* can be seen in RESULTS, CHAPTER VI *"Non-Target Screening of Surface Water Samples for the Identification of Exposome-Related Pollutants: A Case Study from Luxembourg".*[18] NTA involves the detection and identification of unknown compounds present in a sample, without prior knowledge or specific targeting of particular analytes. This approach allows the exploration of the exposome beyond the known chemicals, potentially uncovering new environmental contaminants – and even their health implications – that may have been overlooked in targeted approaches.[19] This information is crucial for an improved understanding of the exposome (the complete picture is nearly impossible to obtain), for hypothesis generation (often applied in metabolomics approaches) and developing strategies for exposure reduction and risk mitigation.[20] NTA using HRMS involves the acquisition of full-scan mass spectra, capturing a wide range of ion signals across the selected mass range. These spectra contain valuable information about the molecular composition and structure of the detected compounds. Advanced data processing techniques, such as mass spectral deconvolution, peak alignment, and spectral matching algorithms, can be employed to process and analyse the complex mass spectral data.[21] NTA also poses certain challenges, as the complex interpretation and identification of unknown compounds require extensive databases, spectral libraries, and advanced cheminformatics tools.[19,22] Moreover, the workflows are neither straightforward nor standardized and there remains a need for harmonization in the field.[23,24] In addition, the high complexity and dynamic nature of environmental samples can lead to the presence of numerous interferences and matrix effects, which can affect the accuracy and reliability of compound identification.[22] For this thesis the R package *patRoon*, developed by Rick Helmus,[21,25] was mainly applied to perform NTA.[18] It offers a fully customizable workflow, integrating different NTA tools into one package to serve as an open source software platform available for a range of instruments and different study needs (see RESULTS, CHAPTER VI[18]).

VI. Data Visualisation

Data visualization plays an important role in interdisciplinary research by fostering collaborations and serving as a valuable tool for effective communication and knowledge integration. Interdisciplinary research projects like *LuxTIME* often involve complex and diverse datasets from multiple disciplines (here: cheminformatics, history, eco-hydrology). Data visualization helps to simplify complex information and improve understanding by presenting it in a visually appealing and intuitive manner. Visualizations provide a means to distil large volumes of data into meaningful patterns, trends, and relationships, enabling researchers from different disciplines to grasp and interpret the information more easily.

One illustrative example of simplifying scientific data complexity through a minimalist visualization approach are the *warming stripes* or *climate stripes* introduced by Ed Hawkins.[26,27] These stripes offer a straightforward and intuitive yet impactful means of conveying the concept of global warming, relying solely on common colour associations (red = warm; blue = cold). This visualization idea was developed further in this project, visualizing chemical patent or literature data summarized in PubChem as '*chemical stripes*'.[12] The numbers of chemicals present in the environment are on the rise as global chemical use, production and trade increases.[14] Obtaining reliable and openly accessible data on global chemical usage over time is exceedingly challenging and historical concentrations of environmental pollutants are limited to a small number of targets only. This is why patent data or data from consolidated references can serve as valuable surrogate sources of information. By quantifying the number of patents (FIGURE 3A) or references (FIGURE 3B) published for a given chemical per year, an estimate of its use and potential environmental release can be derived. The traffic light colour scheme (or alternatively the colour-blind friendly version replacing green with blue, see FIGURE 3C) was employed to signify the severity of the situation, with higher numbers depicted in red as a warning indicator (potential environmental threat). To enable the research community to generate their own *chemical stripes* for their desired chemicals, the R-package *chemicalStripes* was developed.[28] FIGURE 3 shows three examples of the *chemical stripes* for perfluorooctanesulfonic acid (PFOS) using patent and literature data. Modifications of the *chemical stripes* and further information on the visualization can also be found in RESULTS, CHAPTER IV *"How can Data Visualization Support Interdisciplinary Research? LuxTime: Studying Historical Exposomics in Belval"*,[13] CHAPTER V *"Avoiding the Next Silent Spring: Our Chemical Past, Present, and Future"*[14] and in the GENERAL DISCUSSION AND FUTURE PERSPECTIVES part.

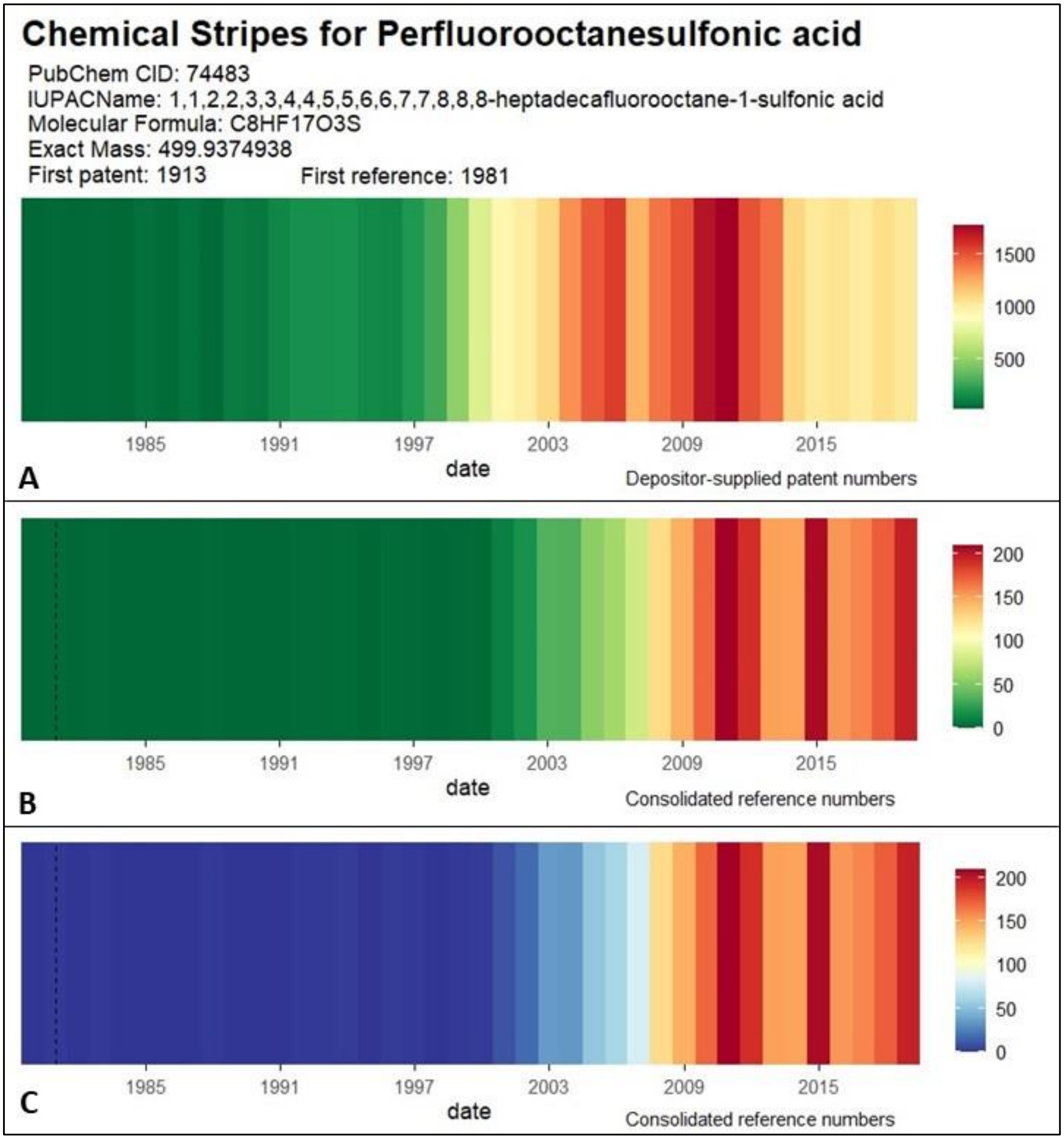


Figure 3: Chemical Stripes example for PFOS (1980-2019) generated using the 'chemicalStripes' R package[28] (CID: Chemical Identifier); (A) stripes based on patent data (B) stripes based on literature data (C) colourblind-friendly version.

Effective communication is crucial in interdisciplinary research, where researchers from different backgrounds like humanities or natural sciences may have varying levels of domain-specific knowledge. Data visualization enabled *LuxTIME* participants to present their findings in a visual format that transcends disciplinary boundaries (*e.g.* with the *chemical stripes*), which was especially important within the *LuxTIME* framework. Moreover, since collaboration and the exchange of ideas across the different disciplines was one of the main goals, data visualization served as a common language that enabled participants to communicate and share their findings, irrespective of their disciplinary backgrounds. By providing a (visual) representation of data, visualizations promoted interdisciplinary discussions and fostered mutual understanding within *LuxTIME*. Data visualization techniques also offered the opportunity to explore and analyse multidimensional datasets from multiple perspectives, promoting creativity and innovation. By visualizing data in different ways, hidden patterns, relationships, and insights that may not have been apparent through traditional analytical methods alone were uncovered.

Throughout the *LuxTIME* project, an extensive digital dataset capturing the historical aspects of Luxembourg was compiled, encompassing diverse quantitative and qualitative data obtained from sources such as archives, newspapers, and scientific institutes. Data visualization tools were employed to delve into this data repository, leverage the accompanying metadata, and effectively communicate research findings and project progress (see RESULTS, CHAPTER IV[13]). Interdisciplinary projects like *LuxTIME* offer a platform to showcase the vast array of available data visualization techniques and concepts, which are often used very sparsely within specific disciplines. Novel approaches involving the combination of visualization tools or the exploration of unconventional visualizations within a given field are seldom attempted, hindering the exploration of new frontiers in data presentation and interpretation. Therefore, CHAPTER IV *"How can Data Visualization Support Interdisciplinary Research? LuxTime: Studying Historical Exposomics in Belval"*[13] presents a data visualisation *toolbox,* suited for all disciplines and involving several different visualization tools and concepts. This toolbox was created to serve as an initial collection to be extended in future interdisciplinary efforts.

VII. Thesis Scope

This interdisciplinary cumulative thesis encompasses a wide range of subjects, including history, data visualization, cheminformatics, and eco-hydrology. It comprises six manuscripts (in the RESULTS section) that highlight the scope of the *LuxTIME* project, showcasing its progress and outcomes.

The manuscript in CHAPTER I *"Historical Exposomics and High Resolution Mass Spectrometry"*[8] is a review paper, published in 2021 in the Exposome journal. It aimed to explain the basic principles of the (chemical) historical exposome and create a foundation with state of the art literature upon which to build the latter work of this thesis.[8] The review discusses chemical approaches required to capture the complex nature of the historical exposome. Different sample types, analytes, analysis techniques, and data interpretation methods are discussed in the context of chemical exposures, with a specific emphasis on their connection to health. The paper concludes with perspectives and challenges studying the historical exposome, revealing that there were not any suitable data types for true historical exposomics research via HRMS within the Minett region, and presenting the need for alternative approaches and more interdisciplinary collaboration.

The interdisciplinary perspective of the historical exposome is then explored in CHAPTER II *"Historical Exposomics: A Manifesto"*,[9] published in 2023 in the Exposome journal. This manifesto like paper emphasizes the need for an interdisciplinary approach and introduces the historical exposomics concept to create a "trading zone" between exposome, history and data science. The possibilities and limitations of using natural and social archives to assess the historical exposome within the *LuxTIME* framework are presented. Moreover, the article discusses definitions of the environment, issues related to significance, visualization, and availability of suitable material (natural archives) and introduces ways of creating new historical narratives to support exposome research. Reconstructing history and historical exposures without available data or material seems to be extremely difficult to impossible.

The next manuscript in CHAPTER III titled "*Exploring the Simulation and Visualization of Historical Data in Environmental History: Airborne Dust Concentration from the Belval Factory*", submitted to the Journal of Digital History, showcases an effective approach to simulate past dust exposure originating from a steel factory in the Minett region using a mathematical model. This study utilized quantitative and qualitative industrial and environmental data from Luxembourg's historical records as input parameters to calculate an estimation of airborne dust concentrations during the last century.

Subsequently, an assessment was conducted to evaluate the chemical contamination and health effects associated with the simulated dust, along with a comparison to actual dust measurements conducted in the past. The resulting dust values were then visually represented through a video, illustrating the extent and temporal variation of atmospheric pollution surrounding the steel plant.

Besides visualising historical dust data, the subsequent paper in CHAPTER IV titled *"How can Data Visualization Support Interdisciplinary Research? LuxTime: Studying Historical Exposomics in Belval"*,[13] published in 2023 in Frontiers - Big Data, further extends the exploration of data visualization techniques. This study focused on visualizing the exposomics data collection, available metadata, and the progress of the *LuxTIME* project. Various visualization tools were presented to facilitate knowledge sharing, track the evolution of interdisciplinary topics, and represent the data contained within the *LuxTIME* inventory. The data inventory is a collection of different qualitative and quantitative data sources connected to the exposome.

The significance of investigating historical (chemical) developments is also stressed in CHAPTER V *"Avoiding the Next Silent Spring: Our Chemical Past, Present, and Future"*.[14] This viewpoint paper – published in Environmental Science & Technology 2023 – delves into the topic of persistent chemicals, specifically per- and polyfluoroalkyl substances (PFAS), which accumulate in the environment over time. It highlights the potential for earlier regulation, drawing parallels with the regulation of the pesticide dichlorodiphenyltrichloroethane (DDT). Furthermore, the manuscript explores the management and regulation of these substances and proposes strategies for improvement. The chemical stripes visualization (see introduction in Synopsis Chapter VI. DATA VISUALISATION and further use in the RESULTS, CHAPTER IV and V)[13,14] is also incorporated and discussed in this manuscript. The included stripes showcase the trends in (estimated) chemical use of DDT and selected PFAS, along with the overall increase in chemical registrations in the CAS registry. A timeline of (increasing) chemical use is presented, showcasing the first use (according to patent data) and the first regulations (Stockholm Convention) for persistent chemicals. The viewpoint stresses the topicality and importance of investigating (past) chemical exposure aligning with the primary objectives of the present PhD project.

Finally, this exposomics research – conducted since the inception of *LuxTIME* – is presented in CHAPTER VI titled *"Non-Target Screening of Surface Water Samples for the Identification of Exposome-Related Pollutants: A Case Study from Luxembourg"*,[18] published in Environmental Sciences Europe 2023. This research paper specifically concentrates on the utilization of cheminformatics within the *LuxTIME* framework, highlighting its role in advancing exposome research. Due to the lack of suitable historical samples (identified in CHAPTER I[8]), a proof of concept was performed on water samples as an exploration of what is possible and how current research can also influence the future. A HRMS and NTA analysis (using *patRoon*) was conducted on a total of 271 surface water samples collected from 20 diverse sampling sites across Luxembourg. Subsequently, a series of classification procedures were carried out, accompanied by comparisons to previous studies and target monitoring efforts. A special focus was on the Minett region, examining specific issues in the region. Furthermore, a list of 41 chemicals of potential significance for inclusion in governmental monitoring lists is presented. This final research paper serves as the conclusive component of the doctoral thesis, illustrating the contribution of (chemical) exposomics research as one disciplinary part within the *LuxTIME* project.

Following the six preceding manuscripts, the thesis closes with a dedicated section for overall GENERAL DISCUSSION AND FUTURE PERSPECTIVES looking at potential future steps.

RESULTS

I. Historical Exposomics and High Resolution Mass Spectrometry

Authors:

Dagny Aurich, Owen Miles and Emma L. Schymanski

Manuscript type:

Review

Contribution statement:

DA and ELS conceptualized the review. **DA** was responsible for writing (original draft preparation) of all chapters except for the 'Water' chapter, which was written by OM. **DA**, ELS and OM did the review and editing. ELS was responsible for funding acquisition and supervision. Visualizations 1 and 4 were reproduced with permission from the respective sources. **DA** modified Figure 3, indicating the original source. Figures 2 and 5 were created by **DA**.

Submission status:

Published in the Exposome journal on 31 December 2021. (DOI: 10.1093/exposome/osab007)

Short summary/ Contribution to the field:

This review examined the exposome more closely from an historical and chemical perspective, setting the focus on the chemical part. The (interdisciplinary) concept of the 'historical exposome' is described further in CHAPTER II.

Different sample types, analytes, analytical methods and data analysis techniques are presented, summarized and evaluated critically based on state of the art literature. The data analysis of HRMS data is explored, emphasizing the importance of selecting appropriate tools. Limitations in HRMS for historical exposome research are highlighted, including inconsistent interpretation and terminology. Moreover, the manuscript shows that the Minett region lacks suitable data types for true historical exposomics research using HRMS, which had a major influence on the course of the *LuxTIME* project.

Exposome, 2021, **1(1)**, osab007
https://doi.org/10.1093/exposome/osab007
Advance Access Publication Date: 31 December 2021
Review

Historical exposomics and high resolution mass spectrometry

Dagny Aurich [1,*], Owen Miles[2], Emma L. Schymanski [1]

[1]Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, 6 Avenue du Swing, Belvaux L-4367, Luxembourg
[2]Luxembourg Institute of Science and Technology (LIST), 5 Avenue des Hauts-Forneaux, Esch-sur-Alzette L-4362, Luxembourg

*To whom correspondence should be addressed: Email: dagny.aurich@uni.lu

Abstract

Awareness of the exposome and its influence on health has increased in the last decade. As past exposures can cause changes in human health many years later, delving into the past is relevant for both diagnostic and prevention purposes, but remains a challenging task. Lifestyle, diet, and socioeconomic information of the past should be well documented and compatible with modern data science methods. While chemical analysis nowadays makes use of high resolution mass spectrometry (HR-MS) for highly sensitive and comprehensive coverage of samples plus retrospective analysis, these data archives are in the very early stages. Since past measurements are often only available for a limited set of chemicals, adding to this knowledge requires careful selection of sample types and sampling sites, which may not always be available. The choice of analytes and analytical methods should be suitable for the study question which is not always clear in advance in exposomics. Data interpretation and the use of appropriate databases are indispensable for a proper exposure assessment, and as databases and knowledge grow, re-analysis of physically or digitally archived samples could enable "continuous monitoring" efforts. This review focuses on the chemical analytical approaches necessary to capture the complexity of the historical exposome. Various sample types, analytes as well as analyses and data interpretation methods are discussed in relation to chemical exposures, while the connection to health remains in focus. It ends with perspectives and challenges in assessing the historical exposome, discussing how we can "learn from the past" to build a better future.

Keywords: exposome; high resolution mass spectrometry; history; health; cheminformatics; pollutants

Background

The combination of the human genome and the exposome yields the phenotype, often represented as $G \times E = P$.[1] However, the majority of health research so far has focused on genomics, with the Human Genome Project[2] building the foundation. In fiscal year 1991, 2.7 billion US dollars were invested in this project alone.[2] However, evidence is increasing that the environment (exposome) deserves greater attention, as several studies show that just 5%–10% of cancers and other diseases, for example, cardiovascular disease (CVD) can be attributed to genetic influences.[3-6] Many diseases are primarily influenced by the environment, also known under the terms *nurture*, the exposome or the *envirome* as J.C. Anthony called environmental factors influencing human health in 1995.[7] In 2020, the European Human Exposome Network[8]—funded with over 100 million euros—was launched to answer the need for research in this area. The exposome concept as first introduced by Wild in 2005[9,10] and later revised by Miller and Jones[6] directed the focus for the first time on *nurture* not just on *nature*. For those unfamiliar with the *nature versus nurture* concept, *nature* can often be misunderstood to mean the natural environmental factors, although in this context it actually means the genetic composition, while *nurture* covers all external factors. The exposome definition expanded by Miller and Jones considers not only environmental influences by chemical exposure, diet, or behavior, but also the associated biological responses.[6] All these factors are acting on the genome since conception onward, with environmental changes being much faster than genetic ones.[11,12] Both present and past factors have to be considered in assessing the exposome, as exposures fluctuate over various timescales: minutes to hours (eg, mealtimes and daily activities), weekly (working hours versus free time), seasonally (eg, changes in sunlight and rainfall hours or chemical application of pesticides, see eg, Wang *et al.*[13]), annually or even over decades (eg, industrialization or decommissioning of activities in an area). Pre-natal exposures during pregnancy—summarized as the maternal exposome—can play a major role and influence the health outcome of a child.[14]

The environment was divided into three subsections by Bhatnagar, visualized in the first figure of his article "Environmental Determinants of CVD."[15] The *natural environment* contains geographic and ecological conditions such as sunlight exposure, altitude, or living in green spaces.[16] The next subsection includes the *social environment*, containing culture, socioeconomic status, or social networks as well as the built environment with structures of houses or cities.[15] This could also be considered as the *urban exposome* concept.[17] Senier *et al.*[18] showed the facets of the *socio-exposome* in greater detail. Lastly, the *personal environment* deals with the important factor of lifestyle as well as income, physical activity, and habits such as smoking.[15] Nutrition plays an important role when considering the

Received: August 26, 2021. **Revised:** December 06, 2021. **Accepted:** December 14, 2021

exposome. In 2017, Cifuentes[19] elaborated on the concept of the *foodome* that is being addressed by many approaches including metabolomics as shown by Borzouie.[20] Other approaches make use of machine learning and text mining algorithms, for example, "FoodMine"[21] that aims at building databases containing the complete chemical composition of food. Most scientists agree that for assessing the whole exposome, an interdisciplinary and long-term approach is required taking together all internal and external factors including "Big Data" sources.[22,23]

Taking a look at the so-called *Anthropocene* described, for example, by Karlsson[24] in 2020, planetary health comes into focus, not only human health. Beginning with industrialization, human beings noticeably changed the environment.[25] To find and prove the origin of past pollution, historically relevant samples must be found, for example, sediment cores showing historical profiles of polychlorinated biphenyls (PCBs)[26] or chlordecone (CLD)[27] or ice cores showing lead pollution thousands of years ago.[28] This is needed for prevention of exposures and the related risks for diseases. In these cases, the assessment of the internal exposome using biomonitoring or untargeted metabolomics is likely of limited applicability.[29] The source or the route of exposure as well as spatial and temporal aspects are more likely to be found in environmental samples, as Turner *et al.*[29] demonstrate. However, information regarding the internal exposome—if present—can complement external exposomics data,[30,31] as some compounds can only be found in one matrix (eg, blood) and others in another (eg, water and ice).[32] While multi-omics techniques are needed to explore the relationship between internal and external exposome and to fully understand the exposome as a whole, these are not a specific focus of this article. For both environmental and biomonitoring, challenges such as accessibility and degradation of samples have to be considered as well as sample analysis and data interpretation. While exposomics research can be performed today using sensors or wearables,[29,33] this data are not (yet) available for the past. Questionnaires as well as literature on exposure from, for example, accidents, population data and health records can help to obtain an approximate picture of the historical exposome, but can be incomplete or inaccurate (as discussed further below).

Assessing any kind of exposure leading to disease outcomes is a quite challenging task. Most diseases, for example, neurodegenerative diseases as Parkinson disease show first effects years after possible exposures.[22] In the past such research was mainly restricted to patient questionnaires, trying to get any causalities out of patient's memories, as Coggon[34] shows. However, those questionnaires are highly subjective and some chemical exposures might not be recognized as such by patients, for example, the use of "Roundup", a glyphosate-based weed killer, in the garden is not always connected with pesticide exposure by patients. In addition, questionnaires are potentially of limited use for people already dealing with cognitive defects.[22]

To join all these factors together, this review focuses mainly on the chemical and data analytical challenges associated with the external historical exposome. The internal exposome is also briefly discussed in terms of historical relevance, with the knowledge that it can only provide partial information on external exposures of the past. Different sample types and approaches used in exposomic research are presented, as well as the challenges and potential for historical research. A focus is placed here on chemical analysis using high resolution mass spectrometry (HR-MS) and its data interpretation as it is a state-of-the-art technique for the detection of environmental pollutants.[25] The limitations currently faced in assessing present-day exposures also apply to the exploration of past exposures. This is covered later in the section "Comparability and quantification issues".

Selecting samples for historical exposomics

To obtain a complete picture of the "historical exposome", one has to consider all factors that contribute to the human exposome—a near impossible task. Besides lifestyle, diet, or socioeconomic factors, chemical pollution has a major impact on the development of certain phenotypes, especially diseases.[18,23] Chemical pollution is also, in many ways, a more tangible concept than many other exposomics factors and as such, a wide array of samples and techniques is available for use. In the following section, various sample types are examined more closely for their suitability and potential for exposomics research, especially regarding their historical information content. The influence of different pollution sources, such as agriculture, industry, or medicine in deciding which sample types may be appropriate is also discussed.

Human subjects

In the past, exposures leading to diseases in humans were mainly assessed by questionnaires, which can be problematic as mentioned above.[34] It may be possible to at least partially counteract this challenge by looking for environmental samples with "historical relevance" indicating exposure, as shown in one of the next sections or just looking at human samples. Some governmental institutions collect and store human samples in biobanks over a period of time. Countries have been collecting and storing samples over 100 years without necessarily knowing the appropriate storage conditions or sample handling, especially for example, for methods in use now that were not available in the past.[35] The establishment of standards for storing different human samples began during the last 30 years, around the time when the term "biobank" was first used.[35] Using proper documentation of the samples makes dating much easier than it is for other sample types. However, for extremely old samples, such as mummies, other dating methods must be applied. Major issues of biobanks are sample degradation and missing long-time biobanks in many countries, as many efforts have just started collecting samples in the last few years.

Human samples contain a lot of information.[5] However, it should be kept in mind that due to different metabolic processes, each sample matrix also reflects a different picture of the internal exposome, which in combination can give an overall picture (discussed further below). The most common samples, such as urine or blood, provide very transient signals for non-persistent chemicals, but can provide long-term information on environmental exposure for some persistent chemicals. Moreover, they have in the meantime well-established analytical protocols and standard materials available for analytical method development. Feces is a similarly non-invasive but relatively short-term sample type with a very complex matrix of increasing relevance given the attention to the microbiome; however, standard materials are still rare. Due to differences in diet, there is a high variability in stool samples; therefore, pooled samples have to be considered to perform individual or population studies. In the field of forensic toxicology, hair is used to provide valuable evidence on many aspects such as drug consumption that occurred, for example, up to weeks before.[36] Calafat *et al.*[37] show that baby teeth, amniotic fluid, and meconium can provide information on pre-natal exposures. Frye *et al.*[38] studied the connection of pre-natal metal exposure and autism looking at baby teeth. Spinal fluid and vitreous fluid, which is often used for post-mortem analyses, are two valuable but invasive sample types to mention; other sample types such as various tissues may also be available but are even more invasive. The main challenges with human samples are accessibility in sufficient quantity and the complex matrix that contains compounds such as endogenous metabolites in much higher concentrations than environmental pollutants.[39] Appropriate sample preparation and highly sensitive analytical techniques such as HR-MS are needed to analyze such samples. In Figure 1, the challenge of low concentrations of pollutants in biological samples is demonstrated compared with concentrations of drugs, endogenous, or food compounds. Another major issue is the comparability of different samples and measurements, as there is no standardized method to perform corrections for, for example, urinary dilution.[40,41] Moreover, there is no standard for reporting concentrations based on the different routes of elimination. These issues still need to be addressed in the context of exposomics, even before historical studies are performed.

Rappaport suggested to use a top-down approach in exposomic studies defining exposures as "biologically active chemicals" in the internal environment of a human organism.[5] The original exposure, as well as its fingerprints in the form of metabolites and even detectable biomarkers are very valuable traces. Thus, even years after a smoker has stopped smoking, this past activity can be identified by certain changes in the human organism.[42,43] Unfortunately, this is not the case for all exposures. While metabolites can act as potential indicators for a certain disease or exposure (ie, functioning as biomarkers), they are not always new or sufficiently unique compounds. In many cases, exposure biomarkers may refer to elevated concentration of some compounds in the part of a population exposed to certain chemicals compared with lower levels in unexposed organisms as shown by Xu *et al.*[44] Not all exposures may have sufficiently specific biomarkers or associated changes in biochemical signals. However, the accessibility of "historical" human samples is a problem, as mentioned above: There are not always cohort studies or biobanks present that reach back to times of initial exposure; for instance, Luxembourg started to collect samples in 2009.[45] Moreover, not all types of exposures can be detected over a long time period in humans as for smokers, as metabolism is a very dynamic process such that most levels decrease over time, and metabolites are further transformed or eliminated.

Today, exposure assessment can be done efficiently by using wearables as shown by Hammel *et al.*[46] They can either function as sensors for environmental data or as monitoring devices for health data of the carrier.[29] The human organism is a complex sample itself, as it is influenced by many factors besides environmental pollutants; the biological response is a measure that can indicate such influences. Other factors as lifestyle, social factors, or other variables in the surrounding ecosystem play a major role, shown by example of cardiovascular risk factors as obesity or hypertension.[47,48]

Other organisms

Other organisms may carry useful information about the present health or pollution state of the (aquatic) environment they are living in, for example, as demonstrated in mussels.[49] However, not much is known about backdating contaminants found in other organisms to the time of exposure, while not all species live sufficiently long. Some cetacean species live for several hundred years and have been shown to be exposed to environmental pollution from persistent chemicals many times over via biomagnification.[50] However, such studies mainly reflect the status and pollution levels of marine organisms.

When looking for representative plant species, trees stand out as promising specimen when it comes to finding historically

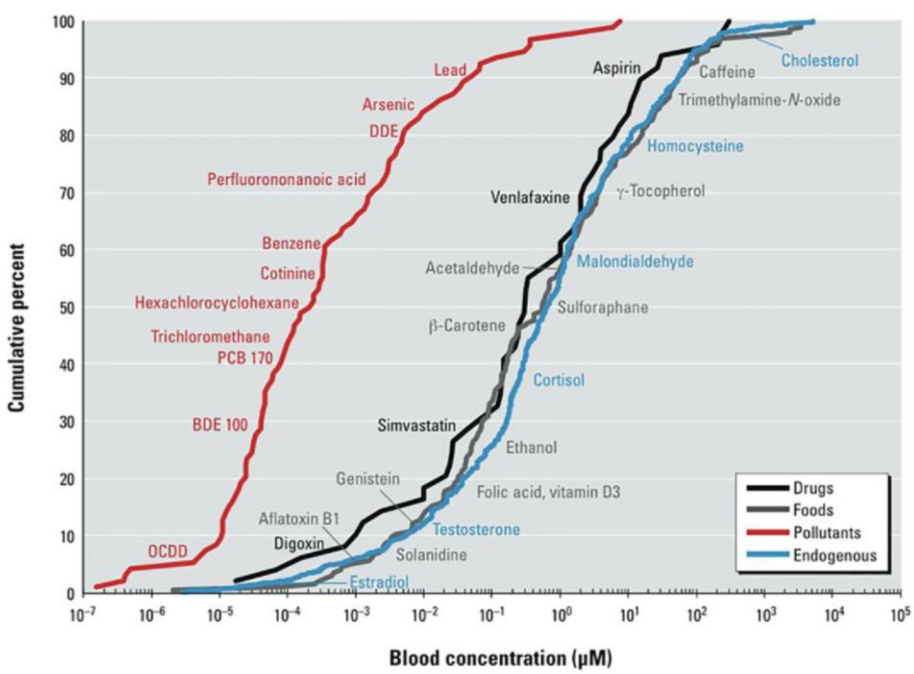


Figure 1. Concentration levels of small molecules and metals in human blood, taken from Rappaport *et al.*[39] Reproduced with permission from Environmental Health Perspectives.

relevant samples. Dendrochronologists can classify the exact age of trees via rings very accurately. It could be postulated that tree rings may yield data on air or soil pollution in chronological sections. Some studies deal exactly with this assumption, for example, Perone *et al.*[51] showed in 2018 the temporal and spatial variability of air pollution from a wide range of sources in oak tree rings looking at tree cores. However, a study by the University of Göttingen in 2006 shows how problematic this assumption is.[52] The enrichment of heavy metals in soils and the accompanying rising acidity increase the cation take-up of plants. Other factors such as growth rate and redistribution of elements determine element concentrations in each ring as well,[52] confounding the interpretation. Some studies are also investigating the use of tree needles or leaves for biomonitoring of environmental pollution.[53,54] With the help of botanical collections even spatial and temporal trends can be found.

Environmental samples

For environmental samples, so-called environmental specimen banks (ESBs) were established in many countries during the last 40 years.[55] In combination with this, there are digitally archived sample measurements available on repositories for HR-MS data such as Global Natural Products Social Molecular Networking (GNPS)[56] and NORMAN Digital Sample Freezing Platform (DSFP)[57] allowing retrospective analysis for recent years.[58] However, this will help future research, whereas historical environmental samples are most of the time not accessible any more.

The three most frequently examined sample types are soil/sediment, water, and air. Figure 2 shows the connection between these sample types and different contamination sources.

Air pollution

Outdoor

There are many different ways to detect air pollution. However, there are few studies that indicate air pollution in the past and most studies are based on theoretical models. Today there are many governmental institutions monitoring air quality of different countries in terms of ozone, particulate matter ($PM_{2.5}$; PM_{10}), nitrogen dioxide (NO_2) and other nitrogen oxides (NOx), lead (Pb) in PM_{10}, benzene (C_6H_6), sulfur dioxide (SO_2), and carbon monoxide (CO) concentrations regulated by the Ambient Air Quality Directive in Europe.[59] Many regulations still focus on atmospheric particles of a given size, although nanoparticles require attention as their toxicity is often underestimated due to the lack of data.[60] Monitoring stations usually make use of passive samplers that collect pollutants over weeks up to 1 month looking at population scale pollution.[59] Other methods also exist to detect air pollution, for example, Hissler *et al.*[61] looked at a lichen species in 2008 to examine local impact of steel production on pollutant concentration in atmospheric deposition in an industrial region of Luxembourg at community scale. Again, there is—for many countries—just data covered from the last decade due to technical facilities not being available in the past and the fact that regulations have just been introduced for many compound classes starting with air pollution acts in the 1950s.[62] In contrast, air pollution awareness and its impact on human health dates back to ancient Rome.[62]

Community scale studies on connecting industrial air pollution to diseases exist, focusing on recent years, for example, on respiratory illness in Valenti *et al.*[63] In the *Global Burden of Disease Study* by the Institute for Health Metrics and Evaluation (IHME), air pollution is shown to be a leading risk factor for diseases and death, causing an estimate of 5 million or 9% deaths in 2017 globally (see Figure 3).[64]

The *State of Global Air 2020* report by the IHME and the Health Effects Institute summarizes different burden of diseases caused by air pollution, namely fine particulate matter and ozone.[66] In 2019, 40% of chronic obstructive pulmonary disorder (COPD) deaths and 30% of lower-respiratory infection deaths were due to air pollution.[66] Air pollutants such as sulfur dioxide originating from industry were addressed in many articles, such as in the article of Calderón-Garcidueñas *et al.*[67] on air pollution causing brain damage. However, during COVID-19 shutdowns in 2020, air

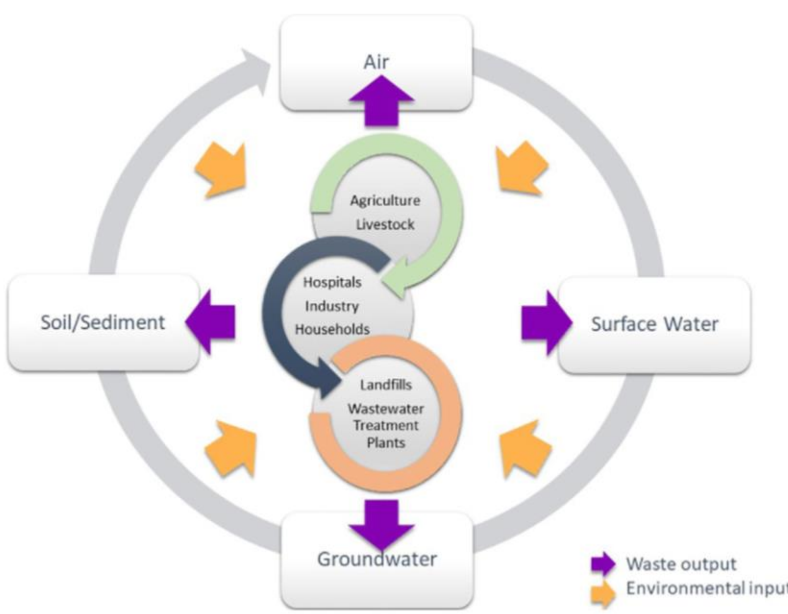


Figure 2. Interconnection of pollution sources, soil, water, and air.

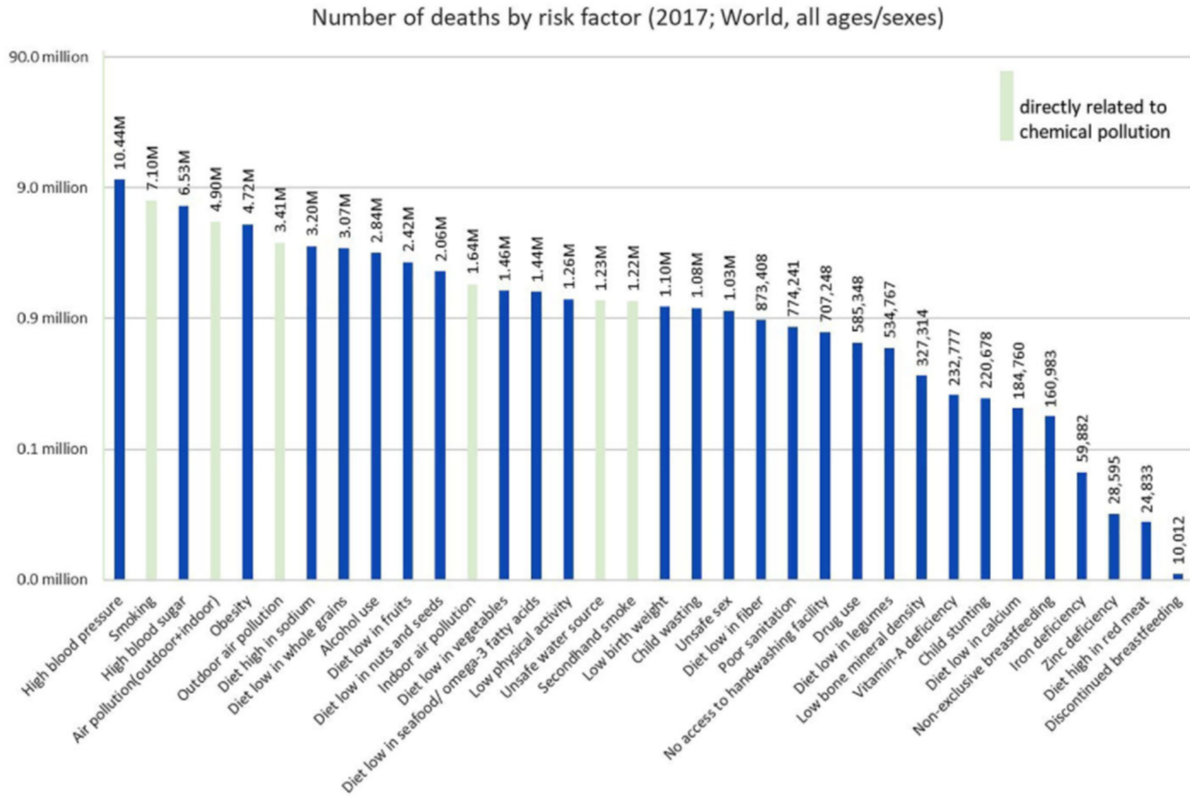


Figure 3. Number of deaths by risk factor, modified from "Our World in Data"[65].
Source: The IHME.[64] Note the logarithmic scale.

pollution levels decreased temporarily as another report of the *State of Global Air* shows. Moreover, there are indications that long-time exposure to air pollution increases the COVID-19 susceptibility as the body's immune defense is affected.[66] Most exposomics studies on air pollution are at population or community scale, presenting an overall picture of pollution issues. However, individual studies can help in case of, for example, occupational diseases to track the workers exposures directly. Personal air monitors have been gaining increased interest over the past few years, although concerns remain regarding data protection and privacy.

Indoor

Dust is a promising sample type for measuring environmental pollution, as many atmospheric pollutants accumulate in dust.[68] Through examining windowsill dust, Han *et al.*[69] showed that industrial activities lead to severe air pollution by potentially toxic metals or PTMs causing serious harm, for example, internal organ damage to a population in China. The analysis of dust on old documents would be therefore quite interesting, as old dust samples could provide information on air quality in the past. There are already quite a few studies on household dust, such as the NORMAN collaborative dust trial,[70] providing information on exposure to chemicals via different sources.[68] Those studies represent just a few possible ways to look at air pollution. Indoor dust plays a major role in chemical exposure as it is present everywhere and contact is unavoidable, especially as a lot of time is spent indoors.[71] It serves as a repository for many chemicals, such as plastic additives, pesticides, heavy metals, cigarette smoke, or personal care products coming from various sources (cooking and cleaning) and individual exposure differs.[72-74]

Water

Water can dissolve, store, and transport chemicals through the environment, making it a versatile sample type.[75] Most of the water on earth is present in the oceans as saltwater or stored in ice, and therefore unavailable for human consumption without modification.[76] The remainder is fresh liquid water and is commonly classified into groundwater and surface water. Surface water is flowing or standing at the surface, and groundwater exists in the pores between soil grains or in fractures in rock formations.

Both groundwater and surface water are consumed by humans and used in agricultural and industrial processes. In Europe, approximately 24% of water is extracted from groundwater and the remainder from surface water resources.[77] Most water is used in agriculture for irrigation and processing food (40%), while around 18% is used in industrial processes.[77]

Water in different flow systems has different residence times, such that these can be useful stores of historical environmental information.[78,79] For example, a groundwater aquifer can transport chemicals over a period of years, decades, or centuries.[78,80] Residence times for different water systems are summarized online in Table 8b-2 of Pidwirny's "Hydrologic cycle."[81]

By understanding the source of water and using different chemical signatures, the so-called "age" of water can be estimated, which is usually defined as the average time that the water entered the flow system from the atmosphere or was released from human activities.[78,79] Care needs to be taken when using the term "age," as water is mixed in the environment, so any sample of water is really a distribution of water molecules of different ages, usually referred to as an average or mean residence time, or MRT.[82]

Groundwater MRT has been studied using environmental tracers such as isotopes or chemical signatures and combining that information with studies of the flow system with hydraulic measurements, often with the aid of numerical or analytical models.[79] This can be used to determine historical contamination at a site for tens of years. Radioisotope dating of the unstable isotopes carbon-14 and tritium are most commonly used in these studies, though other tracers can be used.[78] Uncertainty does exist in the dating, as groundwater flow is mixed with preferential pathways (areas of higher flow in the soil or rock) and radiocarbon dating can be confounded by geochemical changes in the soil and rock that can alter the carbon-14 ratios.

Deeper groundwater can also be of so-called "fossil age," where the water is on the order of thousands of years old. This water can still be mixed with more modern water, making assessment of groundwater MRT very challenging in some environments. Understanding the interactions between new and old groundwater in a deep groundwater flow system are needed to interpret the MRT of groundwater samples, and ultimately the presence and concentrations of chemicals in the systems.

Groundwater can experience chemical concentration changes as the water interacts with the chemicals in the rock and soil grains, organic material, redox conditions, and biological organisms.[75,83] Chemical signatures can be changed or transformed by these chemical or biological processes, which must also be considered when analyzing groundwater samples.

Surface water tends to have MRTs of days to years depending on the flow systems and MRTs are often more easily estimated in surface water systems than groundwater systems, owing to the ability to direct observe flow. Due to the shorter MRTs, stable isotope analysis is more easily performed on surface waters.[79] Surface waters are more rapidly mixed than groundwaters and are exposed to the sunlight and the atmosphere, which can make interpretation of chemical signatures difficult in respect to concentration and transformation effects.

With both surface water and groundwater, it is important to understand the hydraulic conditions that drive the flow and mixing as well as geochemical and biological reactions when assessing water samples.

Wastewater

Urban wastewater is usually treated in treatment plants and released to surface water or groundwater by means of direct flow, injection, or infiltration. Rural wastewater is usually treated in holding tanks or ponds and allowed to infiltrate to groundwater. Wastewater treatments usually rely on filtration followed by biological treatment to reduce organic compounds. Industrial processes often have specific wastewater treatment plants to treat the specific pollutant loads from the processes. Wastewater plants can have long records of their influent and effluent samples as part of regular plant operations.

Wastewater also has great potential for pollution assessment at the community scale.[84] In the context of historical exposomics, it is particularly interesting when it comes to inferring currently relevant diseases from medicines or industrial pollution from certain substances at a population level. Wastewater-based epidemiology (WBE) uses the potential of wastewater to monitor drug consumption and abuse of, for example, narcotics,[85] lifestyle factors such as personal care products and environmental influences such as temperature change or pollutants.[84,86-89] Wastewater also serves as a repository for viruses and bacteria that can indicate the presence of such diseases in a population (a prime example being SARS-CoV-2 WBE).[90,91] Archiving wastewater samples over time would potentially provide a comprehensive picture of the health and lifestyle of a population in a multi-omic manner, avoiding data protection issues and offering a cost-efficient model of population-based monitoring.[88] Wastewater contains a lot more information than surface or groundwater, however, its matrix is inhomogeneous and thus it is not as comparable as water (but better than feces).

Sediment and soil

Sediment sampling can yield historical information about pollutants found in different layers.[92,93] Many personal care products, biocides, or additives accumulate in sediment cores,[92] representing a complementary picture to the pollutants found in water or marine organisms and providing insight into the lifestyle aspect of the exposome (Figure 4).[94,95] Pollutants getting from surface water into sediments and even further to groundwater can endanger human health via different routes of exposure (direct or indirect ingestion and dermal contact).[96] However, there is the big limitation of having suitable water bodies with stable sedimentation patterns present in the area of interest. Turbulence and bioturbation often play a decisive role in sediments, which is why lake rather than river sediments are used.[92]

Soil samples prove to be even more problematic as backdating through the different layers is complicated. Soil evolution is a complex process and there is also bioturbation and mixing of the deposition layers taking place. Natural peat bogs could be used for backdating of samples,[97] however, geographically these are rather rare to find. Sediment as well as soil contamination in general can be monitored quite efficiently using the appropriate extraction or digestion sample preparation techniques. The best technique to use depends on the analytes of interest, for example, heavy metals, persistent organic pollutants (POPs) or emerging pollutants. Han *et al.*[69] or Yang *et al.*[98] showed the great potential of analyzing soil in the context of industrial pollution in China. Industrial pollutants migrate into soil via diffuse atmospheric depositions, sediment particle deposition during flood events in alluvial areas, waste or wastewater disposal, or direct pollution events and thus threaten the soil ecosystem health.[99]

Limitations and future strategies

For all sample types, there are limitations: For water and sediment analysis, suitable water bodies are required. To look at human samples of the past, cohort studies must exist with suitable biobanks for the respective country or region. Other samples may not be sufficiently representative to monitor industrial pollution over a time period. Industrial pollution in soil, water, and air originating from past industries could be the reason for many diseases present nowadays. However, to connect those pollution events in the past to disease outcomes years later is a difficult task. Most of the time a combination of many factors plays a role in disease development, not just a handful of chemical compounds.[23] Exposomics research focuses on making these kinds of connections, which is a very challenging task that, in the context of historical exposomics, may require making the best of available information. Today, with the data and samples available presently, historical exposomics will involve estimating a reasonable exposure assessment for the past. For the future, proactively

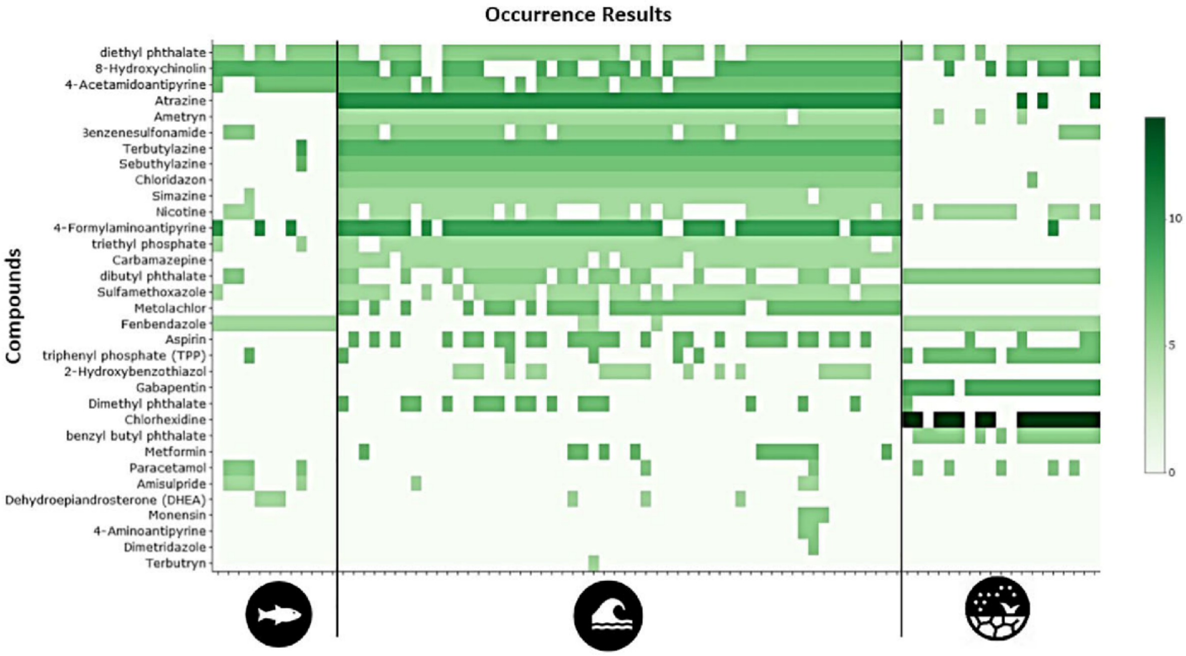


Figure 4. Heatmap showing the occurrence of different pollutants in different media (fish, water, and sediment), adapted from NORMAN-REACH DSFP[57] based on EMBLAS-II project.[95]

organized sampling campaigns could be used to gather and store all kinds of possible sample types at an early stage, to enable and simplify retrospective analyses. This type of sample retention should be in the interest of every population group, because the actual environmental cause of many diseases is often detected far too late and can then no longer be traced. If one also looks at the individual case, work-related illnesses, for example, in the military, can be traced back through previous exposure. In case one sample type is not present in a specific area, one can switch to other specimen indicating exposure, however, the comparability of different sample types is limited. Some pollutants only accumulate in certain media, like, for example, sediment, water, or fish (see Figure 4).

Analytes and analysis for historical exposomics

Environmental pollutants can be anything, from metals to macronutrients through to trace concentrations of organic compounds and organometallic compounds (often termed "micropollutants"). In the past, mostly only contaminations of a limited set of chemicals that exceeded a regulated threshold value were considered. Many other pollutants were likely present in low concentrations, but not yet regulated or monitored at that time. This was mainly due to the technical possibilities and regulatory aspects, which made analysis difficult in the past.[93] Today, techniques such as HR-MS can detect the smallest amounts of contamination in environmental samples where requirements such as ionization properties (ie, if a compound ionizes at all or how efficiently) or compensation of matrix effects and thus good sensitivity are met (see the section "Comparability and quantification issues"). Using the appropriate technique even traces at atto-gram level can be detected, such as for hydroxycholesterol, which is related to breast cancer.[100]

Metals and organometallics

Metals are typical inorganic industrial contaminants. Their toxicity depends among other things on their total concentration as well as from the speciation of elements in the system.[101] Even small amounts can bioaccumulate and have an influence on health. Many neurotoxins are metals such as aluminium (Al), arsenic (As), or mercury (Hg).[102] Lead (Pb) contamination is a major issue especially in developing countries as its use is not regulated there.[103] It is persistent, widely used (eg, paints and cars) and therefore accumulates in the environment quickly, causing serious hazards all over the world. Organometallics have a broad range of applications in plastic manufacturing or as an additive to petrol in the past.[104] Tributyltin (a biocide) and methyl mercury (MeHg, formed by microbes or as a byproduct in industry) are just two high profile organometallic environmental contaminants to be mentioned. Organometallics possess a very high toxicity (eg, Minamata disease caused by MeHg[105]), which is problematic as they are detectable in a variety of environmental samples through past or present use. An inductively coupled plasma (ICP) can be very useful as an ion source when analyzing metals via mass spectrometry (MS) and coupled to liquid chromatography (LC), even organometallics can be analyzed.[106]

Organic compounds

A larger number of organic contaminants are only relatively recently coming into focus, the so-called emerging pollutants: Pesticides like chlorpyrifos, per- and polyfluoroalkyl substances (PFAS), surfactants, pharmaceuticals or persistent, mobile, and toxic (PMT) substances in general, just to mention some groups. However, when investigating historical contamination, it is often difficult to determine the original concentration of some organic compounds and many might not be detectable any more. Other compounds, termed POPs accumulate in the environment over decades, such as the pesticide and insecticide dichloro-diphenyl-trichloroethane (DDT), which was banned in many countries in the 1970s.[107] However, regulation and therefore replacement of those chemicals often led to new emerging pollutants accumulating in nature,[107] with different transformation products that are not yet monitored (so-called regrettable substitution). Organic micropollutants at trace levels (μg/L to ng/L) have been released via anthropogenic activities to the environment over centuries and new substances are being discovered all the time. In a comprehensive annual 2020 review PFAS, replacement flame retardants, iodinated and nitrogenous dibutyl phthalates (DBPs), and antibiotic resistance genes (ARGs) were highlighted as groups of concern.[108] Many of the above mentioned analytes are associated with neurodegenerative diseases such as Alzheimer and Parkinson disease or certain cancers.[109] These connections were only found due to new technical improvements and methods.

Analysis

The choice of the right method for each class of analytes is crucial to find contaminants even at trace levels. After sample preparation, chromatographic methods such as LC or gas chromatography (GC) are usually used to separate compounds of interest. MS is the detection tool of choice in most laboratories. Using different ion sources and mass analyzer modules such as Orbitraps can increase the sensitivity many times over. Before the actual analysis, the acquisition type must be determined. A distinction is made between targeted and non-targeted (NT) analyses.[93] Targeted analyses focus on a limited set of substances to be detected, where the reference standards are available in house in advance for method development. Some instruments are generally only used for targeted analysis (eg, triple quadrupole instruments), while others can offer both targeted and NT acquisition methods. For more details on analytical methods, several recent reviews, overviews, and comparisons exist.[110-113]

Comparability and quantification issues

Exposure can be calculated based on concentration values. However, for example, for biological samples correction methods to report concentrations are not (yet) harmonized and the values are therefore often not comparable.[40] For difficult matrices such as wastewater or feces, correction for matrix effects is essential to obtain reliable results.[114] Moreover, an inter-batch correction compensating for varying signal intensities in a study is needed to compare measurements. Using signal intensities to quantify compounds measured with HRMS is highly problematic, since each compound ionizes differently (ionization efficiency can vary by up to six orders of magnitude)[115-117] and thus intensities are not directly comparable. Semi-quantification approaches such as structural similarity, parent—transformation product proximity, close eluting, ionization efficiency, or combined approaches can be used instead.[118] Another major problem lies in the comparability of pollutant concentrations found in different media, as some pollutants only accumulate in specific matrices[95] (see Figure 4) and concentrations in organisms strongly depend on different metabolic processes. For exposomics studies, it is necessary to look at all types of pollutants and samples as, for example, viruses or bacteria in wastewater mirror the health status of a community[86] and consumer products in sediments[92] reveal information about the lifestyle, each reflecting different sides of the exposome.

Using targeted methods with reference standards to quantify compounds does not necessarily cover the full range of substances in a sample, as important transformation products or pollutants that are not yet monitored may be omitted. However, HRMS is not required to perform targeted analysis. Targeted analysis on lower resolution instruments can be both cost and time efficient for routine analyses. Routine monitoring of, for example, certain rivers in targeted mode is needed to control if regulation values are met.[119] While NT analysis covers more compounds and is more conducive to retrospective screening, is also not yet sufficiently harmonized and/or standardized for routine applications. For both historical exposomics and exposomics in general, precise definitions on how to measure each sample (number of repeats, choice of internal standards, or column, etc.) are necessary. Taking a critical look at the variety of existing methods in exposomics, there is some way to go before harmonization is sufficient for current studies, let alone for implementation into past studies. Some efforts at standardization of NTS are underway, which may pave the way for future harmonization efforts.[120]

Data analysis and interpretation

For NT MS data (hereafter NT-data), there are different data analysis options to consider: Targeted, suspect and NT screening. For targeted screening of NT-data, reference standards are required and matching MS data and retention time, preferably along with fragmentation (MS/MS) data are needed for identification of a compound.[115] Suspect screening is the next potential step: A suspect list of several compounds, for example, pharmaceuticals is used and a search is made for matching MS and—if a library is used—MS/MS. However, if this list becomes too large one easily ends up in a NT approach, where peak-picking is performed, followed by identification efforts.[115]

Databases

There are many compound databases present for use in exposomics, which can be combined with spectral libraries that include spectra and thus fragmentation information for each compound for increased identification confidence.[121] The largest compound databases now contain over 100 million entries, including CAS (184 million),[122] PubChem (111 million),[123] and ChemSpider (114 million).[124] One example of a medium sized compound database that is often used in NT-HR-MS screening approaches, particularly in metabolomics is the Human Metabolome Database (HMDB), with HMDB4.0 containing 115 398 metabolite entries linked to 5702 protein sequences.[23,125] Major sources for suspect lists include the CompTox Chemical Dashboard with >300 lists and the NORMAN Suspect List Exchange with >80 lists, with individual lists containing 10 s up to >100 000 chemicals.[126-128] Using information of existing databases to generate new exposomic resources can be a useful approach to limit the number of chemicals considered in exposomics studies. The Blood Exposome Database was constructed using text mining and information from various databases, resulting in approx. 65 000 entries.[129] The Exposome Explorer, on the other hand, is a much smaller database of approx. 1000 entries.[130] Health-related databases using, for example, cohort studies or exposome databases like the Toxin-Toxin-Target Database (T3DB)[131] can help to find a connection between exposures and health or specific phenotypes. As a database, T3DB is unique in that it shows mechanisms of toxicity as well as target proteins for each toxin, thus linking toxins (3678) and toxin targets (2073).[131] PubChem are integrating many resources and presenting the interlinking of gene, protein, enzyme, disease, and chemical information as knowledge panels.[132] Building a suspect list on, for example, industrial pollutants can be done by patent search of the PubChem[123] database and reducing the overlaps between different fields. The choice of the database always depends on the study question and often databases are too big. For some databases such as PubChem,[123] subsets exist to limit the number of compounds, for example, PubChemLite for Exposomics[133] contains the most relevant and annotated subset of chemicals in PubChem for exposomics. Besides compound databases, there are spectral libraries containing either experimentally or *in silico* predicted spectra of different compounds. Examples of databases containing compound and spectral information are GNPS,[56] MassBank of North America (MoNA),[134] National Institute of Standards and Technology (NIST),[135] METLIN,[136] and MassBank[137,138] with NIST 20 having for example 1.3 million tandem spectra compared with MoNA with 200 000 spectral records. More detailed numbers can be found in other articles on the topic.[139]

A general issue lies within the use of databases: Using large databases yields many candidates per mass, providing many new ideas for possible chemicals, or leaving users juggling interpretations of various scoring terms. However, smaller databases containing just substances related to, for example, a disease or industrial use bear the risk of containing just “old knowledge” and not revealing any new knowledge. Thus, there is still a lot of work remaining until complete and comparable exposomics research is feasible, especially since harmonization and standardization is required in terms of terminology, methods, and reporting.

Software

In untargeted analysis or suspect screening, there is the challenge of peak picking or feature detection, followed by annotation efforts to decipher the identity of the chemicals causing the features. A typical feature count can be of the order of tens of thousands of features per sample using NT-HR-MS.[140] There are several tools enabling (partially) automated data analysis, for example patRoon,[141] XCMS,[142] or MS-DIAL[143] (see Figure 5) that can be used to look for specific masses or compounds present in a sample. In 2019, Wang *et al.*[140] presented PAVE, a peak annotation and verification engine for metabolomics, which includes the “cleaning” of the data and results in the matching metabolite formula. However, this approach requires stable isotope labeling, which is not feasible for most specimen in exposomics.

One speaks intentionally of peak annotation instead of identification as a full identification of a chemical can only be achieved using reference standards (ie, confirmation with target compounds).[93] However, as many standards are difficult to obtain, feature annotation using different computational tools is an alternative approach to tentatively identify chemicals of interest for further confirmation efforts. Feature annotation and compound identification are based on different parameters: The exact mass of the compound paired with its fragmentation pattern can be compared, for example, to experimental or *in silico* spectra using open source software such as MS-DIAL[143] or MetFrag.[144] In addition, retention time can improve identification, depending strongly on the method and instrument used.[145] Blaženović *et al.* used for their analysis of urinary metabolites a combination of several computational tools as CSI: FingerID[146] or NIST hybrid search[147] in order to annotate all metabolites found.[139] There are many other ways to annotate features; the software approaches provided by vendors of MS devices are also a good option for many, but are not covered in detail here. Figure 5 shows various ways of analyzing NT MS data resulting in the different identification levels.[121]

Usually, a statistical analysis follows after annotation (sometimes even before), including uni- and multivariate analysis as well as a validation of the study design and the interpretation of the results. The statistical evaluation methods will not be further elaborated here.[148] Statistics can be used to understand the

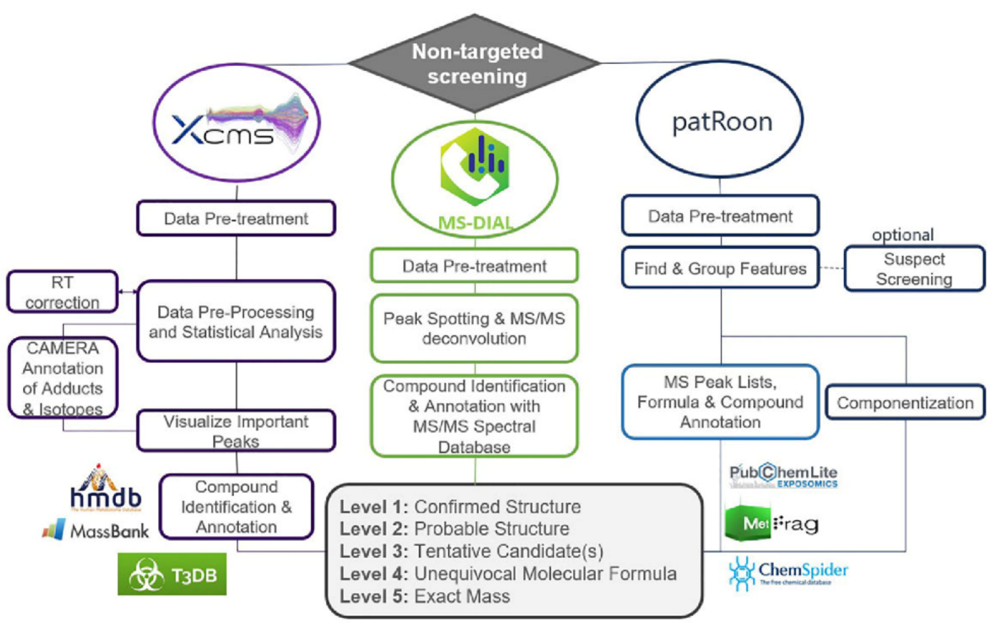


Figure 5. NT screening performed with three different (partially) automated workflows using XCMS,[142] MS-Dial,[143] and patRoon[141] with example databases resulting in reported annotations of various confidence.[121]

individual, community, and ecological exposome changes over time and the resulting health outcomes by interconnecting different data types and analyzing exposure values. This is required in order to understand environmental influences and their impact on health better (eg, for occupational diseases) and to act according to precautionary principles with regards to certain prevention measures (regulations, etc.). Typical strategies in exposomics studies include risk-based prioritization of chemicals and estimating the environmental disease burden.[149] Models for exposure risk and hazard assessment can then be applied in environmental policies and regulations.

Further exposome resources

There are computational tools in development to enter exposure data and analyze it using standard methods from, for example, Bioconductor.[150] The R-package "rexposome" can be used to connect exposures to phenotypes in exposome association studies.[151] The input data in such tools can be acquired by many disciplines, not just chemically using HR-MS. HExpMetDB was developed as a risk-prioritized human exposome database containing physiochemical properties and risk prediction with a graphical user interface (GUI) which enables searching.[152] Finding the connection to health is quite challenging. Health-related databases—as the ones mentioned above—can assist in finding this connection. Health records can be of great value when looking at past events. However, it is not just about doing an epidemiological study looking at several pollutants in connection to a health risk.

Association studies

The interconnection of exposures in chemical networks as well as other factors influencing the health of an organism have to be analyzed.[23] Therefore, environment-wide association studies and even exposome-wide association studies are appropriate ways of finding the connection between exposure and health.[23,153] Metabolome wide association studies find connections between metabolic profiles and disease risk, look for biomarkers of exposure, and even predict future disease onset. However, for all those studies, it is challenging to find relationships between thousands of molecular markers and disease phenotypes with minimal false positive associations.[154] Analytical techniques such as HR-MS or nuclear magnetic resonance enable metabolic profiling and exposure assessment. This makes association studies, metabolic pathway enrichment as well as looking at molecular networks possible.[155] Machine learning can help recognize patterns and make predictions thereafter. However, the limitations of all those approaches have to be considered: Dealing with thousands of features per sample, annotation and identification (with a certain confidence) become difficult and there is still a lack of automation. Moreover, the diversity of chemicals and chemical mixtures has to be taken into account with many unknown variables remaining.[156] Today, it is no longer a problem to measure the samples with sufficient sensitivity in a short time, but rather to draw the right conclusions from the results. Another issue is the terminology that differs between the research fields and the urgent need for harmonization. Association studies require strong international collaborations and high level networking, which is nearly impossible without establishing common terminology.[156]

Geographical information systems-based exposomics

Spatial and geographic data can be used in many ways for exposomics purposes. Geographical information systems (GIS) can connect different types of information that seem completely unrelated.[29] For exposomics it can be helpful to look at various sample types as presented above, at literature and health records to derive facts about historical exposure from those sources and link them geographically. Historical maps or aerial photographs often provide a good source of information on possible contaminated areas to establish connections to industrial sites, landfills, main traffic routes, and bigger cities as these may be more likely to have high levels of pollutants.[29] GIS helps in combining these different information sources by overlaying maps or aerial photographs, integrating data of environmental exposures and health-related data, and analyzing changes over time.[29] Distances between the source of exposure, for example, a closed landfill and affected people can be monitored,[157] as well as mobility of people and factors such as the density of grocery stores offering healthy options.[158] GIS can help in risk assessment and future planning as well as with environmental models.[159] Presenting information on environmental issues in a spatial and graphical way makes analysis easier and enables planning for future needs.

Conclusion

This review covers just a few studies from exposomic research to demonstrate how challenging a historical, retrospective study of the human exposome can be. Sample types have to be chosen carefully as their inter-comparability and suitability for the research question and their availability are limiting factors. If there is no cohort study or no representative sampling site, one has to choose a different sample type to determine exposure in the past. Moreover, analytes and finding the right method for analysis are important as well as the choice of databases for identification efforts. NT-HR-MS is often used for looking at environmental samples and their chemical composition even at trace levels. There are many possibilities to interpret experimental data and various computational tools can be applied. However, the choice of method is often a matter of availability at the institute, or personal preference. Harmonization efforts will be needed in the coming years to increase the comparability between methods.

To fully assess the human exposome, an interdisciplinary approach is required including also other efforts beyond the workflows presented here. In order to obtain an estimate of the historical exposome, a database containing environmental pollutants from different sample measurements, geographical, historical, socioeconomic, and population data as well as health records would be needed to find networks and interconnections and develop prevention strategies for the future. Changes in lifestyle, neighborhoods, and environment can decrease the risk of several diseases when the risk is recognized as such.[15] Thus, it is important not only to monitor health and pollution sources nowadays, but to pay attention to the past influencing factors on health as well. Such infrastructure is a major investment that is only possible at a high level (beyond a single institute) and the recent announcement of a dedicated European Infrastructure for the Exposome (EIRENE) is a very positive sign for this growing field.[160]

A possible topic for future research based on exposomic databases would be the implementation of an Exposome Risk Score.[23] This could be a measure indicating, for example, higher risks for CVD. All in all, as the European Human Exposome Network[8] demonstrates, research is well on the way to shifting the focus to the exposome as well, not just the genome.

Acknowledgments

We acknowledge Aida Horaniet Ibanez and Andreas Fickers from the Luxembourg Centre for Contemporary and Digital History, Reinhard Schneider from the Bioinformatics Core group of the Luxembourg Centre for Systems Biomedicine, Christophe Hissler and Laurent Pfister from the Luxembourg Institute of Science and Technology for their comments, and all LuxTime colleagues who participated in the discussions.

Funding

D.A. acknowledges funding support from the University of Luxembourg through its Institute for Advanced Studies for the project LuxTIME (Audacity-2019); E.L.S. acknowledges funding support from the Luxembourg National Research Fund (FNR) for project A18/BM/12341006.

Conflict of interest statement

None declared.

References

1. Miller GW. The exposome: Purpose, definitions, and scope. In: Washinton K, ed. Exposome. Elsevier; 2020:1–26. https://doi.org/10.1016/B978-0-12-814079-6.00001-8
2. The Human Genome Project. Genome.gov. Accessed April 20, 2021. https://www.genome.gov/human-genome-project
3. Anand P, Kunnumakara AB, Sundaram C, *et al.* Cancer is a preventable disease that requires major lifestyle changes. Pharm Res. 2008;25(9):2097–2116. https://doi.org/10.1007/s11095-008-9661-9
4. Rappaport SM, Smith MT. Environment and disease risks. Science. 2010;330(6003):460–461. https://doi.org/10.1126/science.1192603
5. Rappaport SM. Implications of the exposome for exposure science. J Expo Sci Environ Epidemiol. 2011;21(1):5–9. https://doi.org/10.1038/jes.2010.50
6. Miller GW, Jones DP. The nature of nurture: Refining the definition of the exposome. Toxicol Sci. 2014;137(1):1–2. https://doi.org/10.1093/toxsci/kft251
7. Anthony JC, Eaton WW, Henderson AS. Looking to the future in psychiatric epidemiology. Epidemiol Rev. 1995;17(1):240–242. https://doi.org/10.1093/oxfordjournals.epirev.a036182
8. European Exposome Network—Human Exposome. Accessed April 21, 2021. https://www.humanexposome.eu/
9. Wild CP. Complementing the genome with an "exposome": The outstanding challenge of environmental exposure measurement in molecular epidemiology. Cancer Epidemiol Biomarkers Prev. 2005; 14(8):1847–1850. https://doi.org/10.1158/1055-9965.EPI-05-0456
10. Wild CP. The exposome: From concept to utility. Int J Epidemiol. 2012;41(1):24–32. https://doi.org/10.1093/ije/dyr236
11. Forschungsverbund Berlin. Climate changes faster than animals adapt. *ScienceDaily*. Published 2019. Accessed April 19, 2021. https://www.sciencedaily.com/releases/2019/07/190723121912.htm
12. López-Maury L, Marguerat S, Bähler J. Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. Nat Rev Genet. 2008;9(8):583–593. https://doi.org/10.1038/nrg2398
13. Wang X, Shen Z, Zeng Y, *et al.* Day–night differences, seasonal variations and source apportionment of PM10-bound PAHs over Xi'an, Northwest China. Atmosphere. 2018;9(2):62. https://doi.org/10.3390/atmos9020062
14. Wright ML, Starkweather AR, York TP. Mechanisms of the maternal exposome and implications for health outcomes. ANS Adv Nurs Sci. 2016;39(2):E17–E30. https://doi.org/10.1097/ANS.0000000000000110
15. Bhatnagar A. Environmental determinants of cardiovascular disease. Circ Res. 2017;121(2):162–180. https://doi.org/10.1161/CIRCRESAHA.117.306458
16. Mueller W, Steinle S, Pärkkä J, *et al.* Urban greenspace and the indoor environment: Pathways to health via indoor particulate matter, noise, and road noise annoyance. Environ Res. 2020; 180:108850. https://doi.org/10.1016/j.envres.2019.108850
17. Andrianou XD, Makris KC. The framework of urban exposome: Application of the exposome concept in urban health studies. Sci Total Environ. 2018;636:963–967. https://doi.org/10.1016/j.scitotenv.2018.04.329
18. Senier L, Brown P, Shostak S, Hanna B. The socio-exposome: Advancing exposure science and environmental justice in a postgenomic era. Environ Sociol. 2017;3(2):107–121. https://doi.org/10.1080/23251042.2016.1220848
19. Cifuentes A. Foodomics, foodome and modern food analysis. TrAC Trends Anal Chem. 2017;96:1. https://doi.org/10.1016/j.trac.2017.09.001
20. Borzouie S. Application of metabolomics to measure the Alberta "Foodome". ERA https://doi.org/10.7939/R3891235D.
21. Hooton F, Menichetti G, Barabási AL. Exploring food contents in scientific literature with FoodMine. Sci Rep. 2020;10(1): 16191. https://doi.org/10.1038/s41598-020-73105-0
22. Schymanski EL, Baker NC, Williams AJ, *et al.* Connecting environmental exposure and neurodegeneration using cheminformatics and high resolution mass spectrometry: Potential and challenges. Environ Sci Process Impacts. 2019;21(9):1426–1445. https://doi.org/10.1039/C9EM00068B
23. Vermeulen R, Schymanski EL, Barabási AL, Miller GW. The exposome and health: Where chemistry meets biology. Science. 2020;367(6476):392–396. https://doi.org/10.1126/science.aay3164
24. Karlsson O, Rocklöv J, Lehoux AP, *et al.* The human exposome and health in the Anthropocene. Int J Epidemiol. 2021;50(2): 378–389. https://doi.org/10.1093/ije/dyaa231
25. Chiaia-Hernández AC, Günthardt BF, Frey MP, Hollender J. Unravelling contaminants in the Anthropocene using 2 statistical analysis of LC-HRMS non-target 3 screening data as recorded in lake sediments. Environ Sci Technol. 2017;51(21): 12547–12556. https://doi.org/10.1021/acs.est.7b03357
26. Naffrechoux E, Cottin N, Pignol C, Arnaud F, Jenny JP, Perga ME. Historical profiles of PCB in dated sediment cores suggest recent lake contamination through the "halo effect". Environ Sci Technol. 2015;49(3):1303–1310. https://doi.org/10.1021/es5043996
27. Sabatier P, Mottes C, Cottin N, *et al.* Evidence of chlordecone resurrection by glyphosate in French West Indies. Environ Sci Technol. 2021;55(4):2296–2306. https://doi.org/10.1021/acs.est.0c05207
28. Hong S, Candelone JP, Patterson CC, Boutron CF. Greenland ice evidence of hemispheric lead pollution two millennia ago by greek and roman civilizations. Science. 1994;265(5180): 1841–1843. https://doi.org/10.1126/science.265.5180.1841
29. Turner MC, Nieuwenhuijsen M, Anderson K, *et al.* Assessing the exposome with external measures: Commentary on the

state of the science and research recommendations. Annu Rev Public Health. 2017;38(1):215–239. https://doi.org/10.1146/annurev-publhealth-082516-012802
30. Egeland GM, Ponce R, Bloom NS, Knecht R, Loring S, Middaugh JP. Hair methylmercury levels of mummies of the Aleutian Islands, Alaska. Environ Res. 2009;109(3):281–286. https://doi.org/10.1016/j.envres.2008.11.004
31. Weyrich LS, Duchene S, Soubrier J, *et al.* Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus. Nature 2017;544(7650):357–361. https://doi.org/10.1038/nature21674
32. Shah SFH, Shah SHH, Ahmad L, Sherwani RAK, Aslam M. A study on various pollutants in water and their effect on blood of the consumers. Appl Water Sci. 2021;11(10):164. https://doi.org/10.1007/s13201-021-01498-y
33. Lin EZ, Esenther S, Mascelloni M, Irfan F, Godri Pollitt KJ. The fresh air wristband: A wearable air pollutant sampler. Environ Sci Technol Lett. 2020;7(5):308–314. https://doi.org/10.1021/acs.estlett.9b00800
34. Coggon D. Questionnaire based exposure assessment methods. Sci Total Environ. 1995;168(2):175–178. https://doi.org/10.1016/0048-9697(95)04621-7
35. De Souza YG, Greenspan JS. Biobanking past, present and future: Responsibilities and benefits. Aids. 2013;27(3):303–312. https://doi.org/10.1097/QAD.0b013e32835c1244
36. Kuwayama K, Miyaguchi H, Iwata YT, *et al.* Strong evidence of drug-facilitated crimes by hair analysis using LC–MS/MS after micro-segmentation. Forensic Toxicol. 2019;37(2):480–487. https://doi.org/10.1007/s11419-019-00472-3
37. Calafat AM, Ye X, Silva MJ, Kuklenyik Z, Needham LL. Human exposure assessment to environmental chemicals using biomonitoring. Int J Androl. 2006;29(1):166–171. https://doi.org/10.1111/j.1365-2605.2005.00570.x
38. Frye RE, Cakir J, Rose S, *et al.* Early life metal exposure dysregulates cellular bioenergetics in children with regressive autism spectrum disorder. Transl Psychiatry. 2020;10(1):1–12. https://doi.org/10.1038/s41398-020-00905-3
39. Rappaport SM, Barupal DK, Wishart D, Vineis P, Scalbert A. The blood exposome and its role in discovering causes of disease. Environ Health Perspect. 2014;122(8):769–774. https://doi.org/10.1289/ehp.1308015
40. Hsieh CY, Wang SL, Fadrowski JJ, Navas-Acien A, Kuo CC. Urinary concentration correction methods for arsenic, cadmium, and mercury: A systematic review of practice-based evidence. Curr Environ Health Rep. 2019;6(3):188–199. https://doi.org/10.1007/s40572-019-00242-8
41. Middleton DRS, Watts MJ, Polya DA. A comparative assessment of dilution correction methods for spot urinary analyte concentrations in a UK population exposed to arsenic in drinking water. Environ Int. 2019;130:104721. https://doi.org/10.1016/j.envint.2019.03.069
42. Haswell LE, Papadopoulou E, Newland N, Shepperd CJ, Lowe FJ. A cross-sectional analysis of candidate biomarkers of biological effect in smokers, never-smokers and ex-smokers. Biomarkers 2014;19(5):356–367. https://doi.org/10.3109/1354750X.2014.912354
43. Dickens JA, Miller BE, Edwards LD, *et al.*; Evaluation of COPD Longitudinally to Identify Surrogate Endpoints (ECLIPSE) Study Investigators. COPD association and repeatability of blood biomarkers in the ECLIPSE cohort. Respir Res. 2011;12(1):146. https://doi.org/10.1186/1465-9921-12-146
44. Xu MY, Wang P, Sun YJ, Yang L, Wu YJ. Identification of metabolite biomarkers in serum of rats exposed to chlorpyrifos and cadmium. Sci Rep. 2020;10(1):4999. https://doi.org/10.1038/s41598-020-61982-4
45. IBBL (Integrated BioBank of Luxembourg)—About Us. IBBL. Accessed April 21, 2021. https://www.ibbl.lu/about-ibbl/
46. Hammel SC, Phillips AL, Hoffman K, Stapleton HM. Evaluating the use of silicone wristbands to measure personal exposure to brominated flame retardants. Environ Sci Technol. 2018; 52(20):11875–11885. https://doi.org/10.1021/acs.est.8b03755
47. Ruiz-Castell M, Kandala NB, Kuemmerle A, *et al.* Hypertension burden in Luxembourg: Individual risk factors and geographic variations, 2013 to 2015 European Health Examination Survey. Medicine (Baltimore). 2016;95(36):e4758. https://doi.org/10.1097/MD.0000000000004758
48. Samouda H, Ruiz-Castell M, Bocquet V, *et al.* Geographical variation of overweight, obesity and related risk factors: Findings from the European Health Examination Survey in Luxembourg, 2013–2015. PLoS ONE. 2018;13(6):e0197021. https://doi.org/10.1371/journal.pone.0197021
49. Booij K, Smedes F, Van Weerlee EM, Honkoop PJC. Environmental monitoring of hydrophobic organic contaminants: The case of mussels versus semipermeable membrane devices. Environ Sci Technol. 2006;40(12):3893–3900. https://doi.org/10.1021/es052492r
50. Reijnders PJH, Aguilar A, Borrell A. Pollution and marine mammals. In: WF Perrin, B Würsig, JGM Thewissen, eds. Encyclopedia of Marine Mammals. 2nd ed. Academic Press; 2009:890–898. https://doi.org/10.1016/B978-0-12-373553-9.00205-4
51. Perone A, Cocozza C, Cherubini P, *et al.* Oak tree-rings record spatial–temporal pollution trends from different sources in Terni (Central Italy). Environ Pollut. 2018;233:278–289. https://doi.org/10.1016/j.envpol.2017.10.062
52. Ruppert H, Wischow D, Rushmer T, Gestinger A, Jackson MD. Tree rings—A questionable indicator of heavy metal pollution in air or soil. Elsevier Sci. 2006;70(18):546. https://doi.org/10.1016/j.gca.2006.06.1007
53. Kylin H, Sjödin A. Accumulation of airborne hexachlorocyclohexanes and DDT in pine needles. Environ Sci Technol. 2003; 37(11):2350–2355. https://doi.org/10.1021/es0201395
54. Ockenden WA, Steinnes E, Parker C, Jones KC. Observations on persistent organic pollutants in plants: Implications for their use as passive air samplers and for POP cycling. Environ Sci Technol. 1998;32(18):2721–2726. https://doi.org/10.1021/-es980150y
55. Koizumi A, Harada KH, Inoue K, *et al.* Past, present, and future of environmental specimen banks. Environ Health Prev Med. 2009;14(6):307–318. https://doi.org/10.1007/s12199-009-0101-1
56. Wang M, Carver JJ, Phelan VV, *et al.* Sharing and community curation of mass spectrometry data with GNPS. Nat Biotechnol. 2016;34(8):828–837. https://doi.org/10.1038/nbt.3597
57. Alygizakis NA, Oswald P, Thomaidis NS, *et al.* NORMAN digital sample freezing platform: A European virtual platform to exchange liquid chromatography high resolution-mass spectrometry data and screen suspects in "digitally frozen" environmental samples. TrAC Trends Anal Chem. 2019;115: 129–137. https://doi.org/10.1016/j.trac.2019.04.008
58. Alygizakis NA, Samanipour S, Hollender J, *et al.* Exploring the potential of a global emerging contaminant early warning network through the use of retrospective suspect screening with high-resolution mass spectrometry. Environ Sci Technol. 2018; 52(9):5135–5144. https://doi.org/10.1021/acs.est.8b00365
59. Nagl C, Spangl W. Sampling points for air quality. Public Health and Food Safety. 2019:104 (published online).

60. Bakand S, Hayes A. Toxicological considerations, toxicity assessment, and risk management of inhaled nanoparticles. IJMS. 2016;17(6):929. https://doi.org/10.3390/ijms17060929
61. Hissler C, Stille P, Krein A, *et al.* Identifying the origins of local atmospheric deposition in the steel industry basin of Luxembourg using the chemical and isotopic composition of the lichen *Xanthoria parietina*. Sci Total Environ. 2008;405(1–3): 338–344. https://doi.org/10.1016/j.scitotenv.2008.05.029
62. Mosley S Environmental history of air pollution and protection. In: Agnoletti M, Neri Serneri S, eds. The Basic Environmental History. Vol. 4. Environmental History. Springer International Publishing; 2014:143–165. https://doi.org/10.1007/978-3-319-09180-862.
63. Valenti C, Pozzi P, Busia A, *et al.* Respiratory illness and air pollution from the steel industry: The case of Piquiá de Baixo, Brazil (Preliminary report). Multidiscip Respir Med. 2016;11(1): 41. https://doi.org/10.1186/s40248-016-0077-9
64. Data Visualizations. Institute for Health Metrics and Evaluation. Accessed May 25, 2021. http://www.healthdata.org/results/data-visualizations
65. Number of Deaths by Risk Factor. Our World in Data. Accessed December 17, 2021. https://ourworldindata.org/grapher/number-of-deaths-by-risk-factor
66. State of Global Air. Accessed May 25, 2021. https://www.stateofglobalair.org/
67. Calderón-Garcidueñas L, Azzarelli B, Acuna H, *et al.* Air pollution and brain damage. Toxicol Pathol. 2002;30(3):373–389. https://doi.org/10.1080/01926230252929954
68. Moschet C, Anumol T, Lew BM, Bennett DH, Young TM. Household dust as a repository of chemical accumulation: New insights from a comprehensive high-resolution mass spectrometric study. Environ Sci Technol. 2018;52(5): 2878–2887. https://doi.org/10.1021/acs.est.7b05767
69. Han Q, Liu Y, Feng X, *et al.* Pollution effect assessment of industrial activities on potentially toxic metal distribution in windowsill dust and surface soil in central China. Sci Total Environ. 2021; 759(11):144023. https://doi.org/10.1016/j.scitotenv.2020.144023
70. Rostkowski P, Haglund P, Aalizadeh R, *et al.* The strength in numbers: Comprehensive characterization of house dust using complementary mass spectrometric techniques. Anal Bioanal Chem. 2019;411(10):1957–1977. https://doi.org/10.1007/s00216-019-01615-6
71. Butte W, Heinzow B. Pollutants in house dust as indicators of indoor contamination. Rev Environ Contam Toxicol. 2002;175: 1–46.
72. Melymuk L, Demirtepe H, Jílková SR. Indoor dust and associated chemical exposures. Curr Opin Environ Sci Health. 2020; 15:1–6. https://doi.org/10.1016/j.coesh.2020.01.005
73. Whitehead T, Metayer C, Buffler P, Rappaport SM. Estimating exposures to indoor contaminants using residential dust. J Expo Sci Environ Epidemiol. 2011;21(6):549–564. https://doi.org/10.1038/jes.2011.11
74. Ramírez N, Özel MZ, Lewis AC, Marcé RM, Borrull F, Hamilton JF. Exposure to nitrosamines in thirdhand tobacco smoke increases cancer risk in non-smokers. Environ Int. 2014;71: 139–147. https://doi.org/10.1016/j.envint.2014.06.012
75. Albergamo V, Schollée JE, Schymanski EL, *et al.* Nontarget screening reveals time trends of polar micropollutants in a riverbank filtration system. Environ Sci Technol. 2019;53(13): 7584–7594. https://doi.org/10.1021/acs.est.9b01750
76. Shiklomanov I. World fresh water resource. In: Gleick PH, ed. Water in Crisis. Oxford University Press; 1993:13–15.
77. EEA. *The European Environment: State and Outlook 2020: Knowledge for Transition to a Sustainable Europe*. Publications Office; 2019. Accessed July 5, 2021. https://data.europa.eu/doi/10.2800/96749
78. Cook PG. Introduction to Isotopes and Environmental Tracers as Indicators of Groundwater Flow. The Groundwater Project; 2020.
79. Jasechko S. Global isotope hydrogeology—review. Rev Geophys. 2019;57(3):835–965. https://doi.org/10.1029/2018RG000627
80. Cook PG, Böhlke JK. Determining timescales for groundwater flow and solute transport. In: PG Cook, AL Herczeg, eds. Environmental Tracers in Subsurface Hydrology. Springer US; 2000:1–30. https://doi.org/10.1007/978-1-4615-4557-6_1
81. Pidwirny M. *The Hydrologic Cycle*. Fundamentals of Physical Geography. 2nd ed. 2006. Accessed July 5, 2021. http://www.physicalgeography.net/fundamentals/8b.html
82. Suckow A. The age of groundwater—definitions, models and why we do not need this term. Appl Geochem. 2014;50: 222–230. https://doi.org/10.1016/j.apgeochem.2014.04.016
83. Freeze RA, Cherry JA. Groundwater. Prentice-Hall; 1979.
84. Gallé T, Koehler C, Plattes M, *et al.* Large-scale determination of micropollutant elimination from municipal wastewater by passive sampling gives new insights in governing parameters and degradation patterns. Water Res. 2019;160:380–393. https://doi.org/10.1016/j.watres.2019.05.009
85. Huizer M, ter Laak TL, de Voogt P, van Wezel AP. Wastewater-based epidemiology for illicit drugs: A critical review on global data. Water Res. 2021; 207:117789. https://doi.org/10.1016/j.watres.2021.117789
86. Lorenzo M, Picó Y. Wastewater-based epidemiology: Current status and future prospects. Curr Opin Environ Sci Health. 2019;9:77–84. https://doi.org/10.1016/j.coesh.2019.05.007
87. Bijlsma L, Botero-Coy AM, Rincón RJ, Peñuela GA, Hernández F. Estimation of illicit drug use in the main cities of Colombia by means of urban wastewater analysis. Sci Total Environ. 2016; 565:984–993. https://doi.org/10.1016/j.scitotenv.2016.05.078
88. Been F, Bastiaensen M, Lai FY, *et al.* Mining the chemical information on urban wastewater: Monitoring human exposure to phosphorus flame retardants and plasticizers. Environ Sci Technol. 2018;52(12):6996–7005. https://doi.org/10.1021/acs.est.8b01279
89. Sims N, Kasprzyk-Hordern B. Future perspectives of wastewater-based epidemiology: Monitoring infectious disease spread and resistance to the community level. Environ Int. 2020;139:105689. https://doi.org/10.1016/j.envint.2020.105689
90. Alygizakis N, Markou AN, Rousis NI, *et al.* Analytical methodologies for the detection of SARS-CoV-2 in wastewater: Protocols and future perspectives. Trends Analyt Chem. 2021;134: 116125. https://doi.org/10.1016/j.trac.2020.116125
91. Lundy L, Fatta-Kassinos D, Slobodnik J, *et al.* Making waves: Collaboration in the time of SARS-CoV-2—rapid development of an international co-operation and wastewater surveillance database to support public health decision-making. Water Res. 2021;199:117167. https://doi.org/10.1016/j.watres.2021.117167
92. Chiaia-Hernandez AC, Krauss M, Hollender J. Screening of lake sediments for emerging contaminants by liquid chromatography atmospheric pressure photoionization and electrospray ionization coupled to high resolution mass spectrometry. Environ Sci Technol. 2013;47(2):976–986. https://doi.org/10.1021/es303888v
93. Chiaia-Hernandez AC, Schymanski EL, Kumar P, Singer HP, Hollender J. Suspect and nontarget screening approaches to identify organic contaminant records in lake sediments. Anal Bioanal Chem. 2014;406(28):7323–7335. https://doi.org/10.1007/s00216-014-8166-0
94. Akoto O, Azuure AA, Adotey KD. Pesticide residues in water, sediment and fish from Tono Reservoir and their health risk

implications. SpringerPlus. 2016;5(1):1849. https://doi.org/10.1186/s40064-016-3544-z
95. Slobodnik J, Alexandrov B, Arabidze M, Guchmanidze A, Komorin V. Improving environmental monitoring in the Black Sea—Phase II (EMBLAS-II) ENPI/2013/313-169, UNDP Regional Bureau for Europe and the CIS. EMBLAS PROJECT. Published 2016. Accessed October 20, 2021. https://emblasproject.org/wp-content/uploads/2017/04/Joint-Black-Sea-Surveys-2016_16.pdf
96. US EPA O. Exposure Assessment Tools by Media—Water and Sediment. Published April 23, 2015. Accessed October 20, 2021. https://www.epa.gov/expobox/exposure-assessment-tools-media-water-and-sediment
97. Marx SK, Kamber BS, McGowan HA, Zawadzki A. Atmospheric pollutants in alpine peat bogs record a detailed chronology of industrial and agricultural development on the Australian continent. Environ Pollut. 2010;158(5):1615–1628. https://doi.org/10.1016/j.envpol.2009.12.009
98. Yang Q, Li Z, Lu X, Duan Q, Huang L, Bi J. A review of soil heavy metal pollution from industrial and agricultural regions in China: Pollution and risk assessment. Sci Total Environ. 2018;642:690–700. https://doi.org/10.1016/j.scitotenv.2018.06.068
99. Yuan J, Lu Y, Wang C, *et al.* Ecology of industrial pollution in China. Ecosyst Health Sustain. 2020;6(1):1779010. https://doi.org/10.1080/20964129.2020.1779010
100. Vehus T, Roberg-Larsen H, Waaler J, *et al.* Versatile, sensitive liquid chromatography mass spectrometry—Implementation of 10 μm OT columns suitable for small molecules, peptides and proteins. Sci Rep. 2016;6(1):37507. https://doi.org/10.1038/srep37507
101. Tchounwou PB, Yedjou CG, Patlolla AK, Sutton DJ. Heavy metal toxicity and the environment. In: A Luch, ed. Molecular, Clinical and Environmental Toxicology: Volume 3: Environmental Toxicology. Experientia Supplementum. Springer; 2012:133–164. https://doi.org/10.1007/978-3-7643-8340-4_6
102. Schofield K. The metal neurotoxins: An important role in current human neural epidemics? IJERPH. 2017;14(12):1511. https://doi.org/10.3390/ijerph14121511
103. Wani AL, Ara A, Usmani JA. Lead toxicity: A review. Interdiscip Toxicol. 2015;8(2):55–64. https://doi.org/10.1515/intox-2015-0009
104. Stone FGA, West R. Advances in Organometallic Chemistry. Vol. 41. Academic Press; 1997.
105. Harada M. Minamata disease: Methylmercury poisoning in Japan caused by environmental pollution. Crit Rev Toxicol. 1995;25(1):1–24. https://doi.org/10.3109/10408449509089885
106. Wilbur S, Wahlen R, Morton K. Organotin analysis by GC and LC ICP-MS. LC GC N Am 2004;22:33–33.
107. Rahman MM. Insecticide substitutes for DDT to control mosquitoes may be causes of several diseases. Environ Sci Pollut Res Int. 2013;20(4):2064–2069. https://doi.org/10.1007/s11356-012-1145-0
108. Richardson SD, Kimura SY. Water analysis: Emerging contaminants and current issues. Anal Chem. 2020;92(1):473–505. https://doi.org/10.1021/acs.analchem.9b05269
109. Chin-Chan M, Navarro-Yepes J, Quintanilla-Vega B. Environmental pollutants as risk factors for neurodegenerative disorders: Alzheimer and Parkinson diseases. Front Cell Neurosci. 2015;9:124. https://doi.org/10.3389/fncel.2015.00124
110. Fiedler H, Kennedy T, Henry BJ. A critical review of a recommended analytical and classification approach for organic fluorinated compounds with an emphasis on per- and polyfluoroalkyl substances. Integr Environ Assess Manag. 2021;17(2):331–351. https://doi.org/10.1002/ieam.4352
111. Salgueiro-González N, Castiglioni S, Zuccato E, Turnes-Carou I, López-Mahía P, Muniategui-Lorenzo S. Recent advances in analytical methods for the determination of 4-alkylphenols and bisphenol A in solid environmental matrices: A critical review. Anal Chim Acta. 2018;1024:39–51. https://doi.org/10.1016/j.aca.2018.02.081
112. Tadeo JL. Analysis of Pesticides in Food and Environmental Samples. 2nd ed. CRC Press; 2019.
113. Lee HB. Review of analytical methods for the determination of nonylphenol and related compounds in environmental samples. Water Qual Res J. 1999;34(1):3–36. https://doi.org/10.2166/wqrj.1999.002
114. Tisler S, Pattison DI, Christensen JH. Correction of matrix effects for reliable non-target screening LC–ESI–MS analysis of wastewater. Anal Chem. 2021;93(24):8432–8441. https://doi.org/10.1021/acs.analchem.1c00357
115. Hollender J, Schymanski EL, Singer HP, Ferguson PL. Nontarget screening with high resolution mass spectrometry in the environment: Ready to go? Environ Sci Technol. 2017;51(20):11505–11512. https://doi.org/10.1021/acs.est.7b02184
116. Kruve A, Kaupmees K, Liigand J, Leito I. Negative electrospray ionization via deprotonation: Predicting the ionization efficiency. Anal Chem. 2014;86(10):4822–4830. https://doi.org/10.1021/ac404066v
117. Oss M, Kruve A, Herodes K, Leito I. Electrospray ionization efficiency scale of organic compounds. Anal Chem. 2010;82(7):2865–2872. https://doi.org/10.1021/ac902856t
118. Malm L, Palm E, Souihi A, Plassmann M, Liigand J, Kruve A. Guide to semi-quantitative non-targeted screening using LC/ESI/HRMS. Molecules. 2021;26(12):3524. https://doi.org/10.3390/molecules26123524
119. Sacher F, Ehmann M, Gabriel S, Graf C, Brauch HJ. Pharmaceutical residues in the river Rhine—results of a one-decade monitoring programme. J Environ Monit. 2008;10(5):664–670. https://doi.org/10.1039/B800701B
120. ISO (International Organization for Standardization)—Technical Committee ISO/TC 147. ISO/DIS 21253-1:2018(E)/Water quality Multi-compound class methods—Part 1: Criteria for the identification of target compounds by gas and liquid chromatography and mass spectrometry. Published online 2018.
121. Schymanski EL, Jeon J, Gulde R, *et al.* Identifying small molecules via high resolution mass spectrometry: Communicating confidence. Environ Sci Technol. 2014;48(4):2097–2098. https://doi.org/10.1021/es5002105
122. CAS REGISTRY. CAS. Accessed August 2, 2021. https://www.cas.org/cas-data/cas-registry
123. PubChem. PubChem. Accessed April 27, 2021. https://pubchem.ncbi.nlm.nih.gov/
124. Pence HE, Williams A. ChemSpider: An online chemical information resource. J Chem Educ. 2010;87(11):1123–1124. https://doi.org/10.1021/ed100697w
125. Wishart DS, Feunang YD, Marcu A, *et al.* HMDB 4.0: The human metabolome database for 2018. Nucleic Acids Res. 2018;46(D1):D608–D617. https://doi.org/10.1093/nar/gkx1089
126. Lowe CN, Williams AJ. Enabling high-throughput searches for multiple chemical data using the U.S.-EPA CompTox chemicals dashboard. J Chem Inf Model. 2021;61(2):565–570. https://doi.org/10.1021/acs.jcim.0c01273
127. Williams AJ, Grulke CM, Edwards J, *et al.* The CompTox chemistry dashboard: A community data resource for environmental chemistry. J Cheminform. 2017;9(1):61. https://doi.org/10.1186/s13321-017-0247-6
128. NORMAN Network. NORMAN Suspect List Exchange (NORMAN-SLE). Published 2021. Accessed April 22, 2021. https://www.norman-network.com/nds/SLE/

129. Barupal DK, Fiehn O. Generating the blood exposome database using a comprehensive text mining and database fusion approach. Environ Health Perspect. 2019;127(9):97008. https://doi.org/10.1289/EHP4713
130. Neveu V, Nicolas G, Salek RM, Wishart DS, Scalbert A. Exposome-Explorer 2.0: An update incorporating candidate dietary biomarkers and dietary associations with cancer risk. Nucleic Acids Res. 2020;48(D1):D908–D912. https://doi.org/10.1093/nar/gkz1009
131. Wishart D, Arndt D, Pon A, *et al.* T3DB: The toxic exposome database. Nucleic Acids Res. 2015;43(Database issue): D928–934. https://doi.org/10.1093/nar/gku1004
132. Zaslavsky L, Cheng T, Gindulyte A, *et al.* Discovering and summarizing relationships between chemicals, genes, proteins, and diseases in PubChem. Front Res Metr Anal. 2021. https://doi.org/10.3389/frma.2021.689059
133. Schymanski EL, Bolton EE, Kondic T, Thiessen P, Zhang J. PubChemLite for exposomics. Published online October 31, 2020. https://doi.org/10.5281/zenodo.4183801
134. MassBank of North America. Accessed May 17, 2021. https://mona.fiehnlab.ucdavis.edu/downloads
135. National Institute of Standards and Technology. NIST. Accessed May 26, 2021. https://www.nist.gov/
136. Guijas C, Montenegro-Burke JR, Domingo-Almenara X, *et al.* METLIN: A technology platform for identifying knowns and unknowns. Anal Chem. 2018;90(5):3156–3164. https://doi.org/10.1021/acs.analchem.7b04424
137. Horai H, Arita M, Kanaya S, *et al.* MassBank: A public repository for sharing mass spectral data for life sciences. J Mass Spectrom. 2010;45(7):703–714. https://doi.org/10.1002/jms.1777
138. MassBank—Consortium and its Contributors. MassBank/MassBank-Data: Release Version 2021.03. Zenodo; 2021. https://doi.org/10.5281/ZENODO.4563726
139. Blaženović I, Kind T, Sa MR, *et al.* Structure annotation of all mass spectra in untargeted metabolomics. Anal Chem. 2019; 91(3):2155–2162. https://doi.org/10.1021/acs.analchem.8b04698
140. Wang L, Xing X, Chen L, *et al.* Peak annotation and verification engine for untargeted LC–MS metabolomics. Anal Chem. 2019; 91(3):1838–1846. https://doi.org/10.1021/acs.analchem.8b03132
141. Helmus R, ter Laak TL, van Wezel AP, de Voogt P, Schymanski EL. patRoon: Open source software platform for environmental mass spectrometry based non-target screening. J Cheminform. 2021;13(1). https://doi.org/10.1186/:1.020-00477-w
142. Tautenhahn R, Patti GJ, Rinehart D, Siuzdak G. XCMS online: A web-based platform to process untargeted metabolomic data. Anal Chem. 2012;84(11):5035–5039. https://doi.org/10.1021/ac300698c
143. Tsugawa H, Ikeda K, Takahashi M, *et al.* A lipidome atlas in MS-DIAL 4. Nat Biotechnol. 2020;38(10):1159–1163. https://doi.org/10.1038/s41587-020-0531-2
144. Ruttkies C, Schymanski EL, Wolf S, Hollender J, Neumann S. MetFrag relaunched: Incorporating strategies beyond in silico fragmentation. J Cheminform. 2016;8(1):3. https://doi.org/10.1186/s13321-016-0115-9
145. Oberacher H, Sasse M, Antignac JP, *et al.* A European proposal for quality control and quality assurance of tandem mass spectral libraries. Environ Sci Eur. 2020;32(1):43. https://doi.org/10.1186/s12302-020-00314-9
146. Dührkop K, Shen H, Meusel M, Rousu J, Böcker S. Searching molecular structure databases with tandem mass spectra using CSI:FingerID. Proc Natl Acad Sci USA. 2015;112(41): 12580–12585. https://doi.org/10.1073/pnas.1509788112
147. Cooper BT, Yan X, Simon Y, Tchekhovskoi DV, Mirokhin YA, Stein SE. Hybrid search: A method for identifying metabolites absent from tandem mass spectrometry libraries. Anal Chem. 2019;91(21):13924–13832. https://doi.org/10.1021/acs.analchem.9b03415
148. Uppal K, Walker DI, Liu K, Li S, Go YM, Jones DP. Computational metabolomics: A framework for the million metabolome. Chem Res Toxicol. 2016;29(12):1956–1975. https://doi.org/10.1021/acs.chemrestox.6b00179
149. Agache I, Miller R, Gern JE, *et al.* Emerging concepts and challenges in implementing the exposome paradigm in allergic diseases and asthma: A Practall document. Allergy 2019;74(3): 449–463. https://doi.org/10.1111/all.13690
150. Bioconductor. Accessed May 18, 2021. https://www.bioconductor.org/help/faq/
151. Exposome Data Analysis. Accessed April 19, 2021. https://bioconductor.org/packages/release/bioc/vignettes/rexposome/inst/doc/exposome_data_analysis.html
152. Zhao F, Li L, Chen Y, *et al.* Risk-based chemical ranking and generating a prioritized human exposome database. Environ Health Perspect. 2021; 129(4):47014. https://doi.org/10.1289/EHP7722
153. Juarez PD, Matthews-Juarez P. Applying an exposome-wide (ExWAS) approach to cancer research. Front Oncol. 2018;8:313. https://doi.org/10.3389/fonc.2018.00313
154. Chadeau-Hyam M, Ebbels TMD, Brown IJ, *et al.* Metabolic profiling and the metabolome-wide association study: Significance level for biomarker identification. J Proteome Res. 2010;9(9): 4620–4627. https://doi.org/10.1021/pr1003449
155. Walker DI, Pennell KD, Uppal K, *et al.* Pilot metabolome-wide association study of benzo(a)pyrene in serum from military personnel. J Occup Environ Med. 2016;58(8 Suppl 1):S44–52. https://doi.org/10.1097/JOM.0000000000000772
156. David A, Chaker J, Price EJ, *et al.* Towards a comprehensive characterisation of the human internal chemical exposome: Challenges and perspectives. Environ Int. 2021;156:106630. https://doi.org/10.1016/j.envint.2021.106630
157. Makuleke P, Ngole-Jeme VM. Soil heavy metal distribution with depth around a closed landfill and their uptake by datura stramonium. Appl Environ Soil Sci. 2020;2020:e8872475. https://doi.org/10.1155/2020/8872475
158. Pinho MGM, Mackenbach JD, den Braver NR, Beulens JJW, Brug J, Lakerveld J. Recent changes in the Dutch foodscape: Socioeconomic and urban–rural differences. Int J Behav Nutr Phys Act. 2020;17(1):43. https://doi.org/10.1186/s12966-020-00944-5
159. González A. GIS in environmental assessment: A review of current issues and future needs. J Environ Assess Policy Manag. 2012;14. https://doi.org/10.1142/S146433321250007X
160. ESFRI Announces new RIs for Roadmap 2021 | www.esfri.eu. Accessed August 3, 2021. https://www.esfri.eu/latest-esfri-news/new-ris-roadmap-2021

II. Historical Exposomics: A Manifesto

Authors:

Dagny Aurich, Aida Horaniet Ibanez, Christophe Hissler, Simon Kreipl, Laurent Pfister, Emma L. Schymanski, Andreas Fickers

Manuscript type:

Commentary

Contribution statement:

All authors contributed to the conceptualization, review, and editing of the commentary. **DA** was responsible for writing the introduction, 'The *LuxTIME* Project as an Interdisciplinary Framework', 'Critique of the Current Approach', 'Botanical Samples', and parts of the 'Conclusion'. AHI wrote 'Finding Significance in a Different Way' and 'Data Visualization'. LP wrote the section on 'Freshwater Bivalves'. SK wrote the 'Historical Groundwater Reserves' section, and ELS wrote the abstract. SK, **DA**, and AF contributed to the sections on 'Environmental History and the Anthropocene'. AF wrote the section on 'Definitions of the Environment', the introductory part of 'Evidence from Social and Natural Archives', 'New historical narratives', and modified the 'Conclusion'. ELS and AF were responsible for funding acquisition and supervision. Figures 1, 2, and 4 were originally created by AHI based on data collected by **DA** and AHI. **DA** created Figure 3. Figure 5 was reproduced with permission from the respective sources.

Submission status:

Published in the Exposome journal on 18 August 2023. (DOI: [10.1093/exposome/osad007](https://doi.org/10.1093/exposome/osad007))
This manuscript will be also part of the PhD thesis by co-first author Aida Horaniet Ibanez.

Short summary/ Contribution to the field:

This perspective examined the historical exposome concept in a manifesto-style article focusing especially on the aspect of interdisciplinarity. It discusses the use of natural and social archives to bridge the gap between exposome, history and data science, creating a 'trading zone' between apparently completely different disciplines. The main disciplines involved in *LuxTIME* are eco-hydrology, environmental chemistry and environmental history, each providing new viewpoints and approaches to the study of the exposome. This article reflects on the possibilities and limitations of current - mostly natural science based - approaches and shows the numerous possibilities to involve natural and social archives in the study of the exposome.

Exposome, 2023, **3(1)**, osad007
https://doi.org/10.1093/exposome/osad007
Advance Access Publication Date: 18 August 2023
Commentary

Historical exposomics: a manifesto

Dagny Aurich [1,*,†], Aida Horaniet Ibanez[2,*,†], Christophe Hissler [3], Simon Kreipl[4], Laurent Pfister[3,5], Emma L. Schymanski [1], Andreas Fickers[2]

[1]Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Belvaux, Luxembourg
[2]Luxembourg Centre for Contemporary and Digital History (C2DH), University of Luxembourg, Esch-sur-Alzette, Luxembourg
[3]Catchment and Eco-hydrology Research Group (CAT), Luxembourg Institute of Science and Technology (LIST), Belvaux, Luxembourg
[4]Department of Water Management, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands
[5]Faculty of Science, Technology and Medicine, University of Luxembourg, Belvaux, Luxembourg

*To whom correspondence should be addressed: Dagny Aurich: Email: dagny.aurich@uni.lu; Aida Horaniet Ibanez: Email: aida.horanietibanez@uni.lu

†These authors contributed equally to this work and share first authorship

Abstract

The exposome complements information captured in the genome by covering all external influences and internal (biological) responses of a human being from conception onwards. Such a paradigm goes beyond a single scientific discipline and instead requires a truly interdisciplinary approach. The concept of "historical exposomics" could help bridge the gap between "nature" and "nurture" using both natural and social archives to capture the influence of humans on earth (the Anthropocene) in an interdisciplinary manner. The LuxTIME project served as a test bed for an interdisciplinary exploration of the historical exposome, focusing on the Belval area located in the Minett region in southern Luxembourg. This area evolved from a source of mineral water to steel production through to the current campus for research and development. This article explores the various possibilities of natural and social archives that were considered in creating the historical exposome of Belval and reflects upon possibilities and limitations of the current approaches in assessing the exposome using purely a natural science approach. Issues surrounding significance, visualization, and availability of material suitable to form natural archives are discussed in a critical manner. The "Minett Stories" are presented as a way of creating new historical narratives to support exposome research. New research perspectives on the history of the Anthropocene were opened by investigating the causal relationships between factual evidence and narrative evidence stemming from historical sources. The concept of historical exposome presented here may thus offer a useful conceptual framework for studying the Anthropocene in a truly interdisciplinary fashion.

Keywords: exposome; environmental history; historical exposomics; digital history; social archives; natural archives

Introduction

The *nature versus nurture* debate is one of the oldest philosophical debates, dating back to ancient times (eg, Plato's Protagoras).[1] *Nature* is the genetic and biological "predetermination" of a human being, whereas *nurture* comprises all external factors influencing the individual from conception onwards. Today, this debate is—it seems—once more reframed as a scientific controversy opposing *genomics* to the concept of *exposomics*. With the highly funded Human Genome Project, next generation sequencing methods and genome-wide association studies (GWAS), genomics is both more mature and better financed than exposomics. However, interest in exposomics is increasing as many diseases cannot be traced back to genetics alone, but rather to the interplay of genetics and many other factors, which are captured under the exposome concept. The exposome covers all external influences and internal (biological) responses of a human being from conception onwards.[2] This paradigm[2] cannot be covered by just one scientific discipline; it requires a truly interdisciplinary approach. The "Luxemburg Time Machine" (LuxTIME) project is such an interdisciplinary setting—a "trading-zone" between different disciplines to find a common language between humanities and natural sciences. For this we propose the concept of "historical exposomics", aiming at interrogating the "trading zone"[3] between the exposome, history and data science to capture human influence on the natural world.

It is rare to find publications written from the perspective of both humanities and natural sciences due to the disciplinary specialization of each field. In general, both genome and exposome research is dominated by the natural sciences so far. Yet involving humanities in this research area, and more specifically adding a historical view, offers interesting perspectives for the exposome. Historical sources can help reconstruct the human exposome by revealing a multitude of historical data typically unexplored by the natural sciences. Exposomics evidence from the past is difficult to find in "natural archives" (the focus of natural sciences) present today. However, looking at "social archives"

Received: April 20, 2023. **Revised:** July 4, 2023. **Accepted:** July 21, 2023

opens new possibilities to detect historical evidence (see Evidence section with an explanation of natural and social archives). By combining "hard" and "soft" facts, a more complex understanding of past processes of environmental change is possible. This article explores this based on the example of the interdisciplinary LuxTIME project from the University of Luxembourg.[4]

The LuxTIME project as an interdisciplinary framework

The LuxTIME project aims at exploring new ways of analyzing and interpreting factual evidence of the past by building an interdisciplinary framework for the investigation of "big data" of the past. Building on the conceptual premises of the "European Time Machine" Flagship project,[5] that is to bring together High Performance Computing (HPC) facilities, the analytical capacities of Machine Learning and Artificial Intelligence, and "big" and complex historical data, LuxTIME includes information from three different fields and scientific perspectives, namely eco-hydrology, environmental chemistry and industrial history. LuxTIME uses a local case (the industrialization of Belval in the Minett region of Southern Luxembourg)[6] as a testbed for methodological and epistemological reflections on how to study the long term impact of environmental changes on the health of the local population. By mixing "contextual information" based on archival evidence with "scientific evidence" derived from chemical, biological, or medical investigations, the project explores new grounds for interpreting big data of the past in a truly interdisciplinary setting.

Environmental history and the anthropocene

LuxTIME is inspired by the rapidly growing field of environmental history, which aims at analyzing and describing the co-evolution of human society and the natural environment.[7] The interdependence of humans and other living and non-living systems on Earth is recognized, and the unfolding of history is studied in the framework of the natural world. Recently, the human influence on the development of the ecosystem is being discussed and debated under the label of the "Anthropocene"[8]—a new temporal regime and ecological period marked by the human imprint on the natural world, including the "Critical Zone", the Earth's skin which sustains nearly all terrestrial life including humanity.[9-12,13]

For most of the time since the origins of humankind, the population was too small, spatially scattered and technologically undeveloped to markedly influence the environment. With the Neolithic Revolution (~10 000 BCE) and the change from a hunting and gathering lifestyle to one of agriculture and settlement, the human species learnt new ways to modify the environment to serve their purposes. Deforestation, irrigation,14 the development of systematic sowing methods and the use of fertilizers drastically increased food production, allowing for a surge in the human population size. In 10100 BCE the population rose above 5 million. With this came the development of trading networks and complex societies.[15] Ecosystems were brought out of their natural balance.[16] Soil erosion and salinization, the depletion of nutrients or a rise in infectious diseases were some of many consequences accompanying these technological advancements.[17,18] From the Roman Era to the Middle Age, new technological development of iron metallogeny and intense mining activities also enriched in heavy metals the local atmospheric deposition all over Europe.[19,20]

Since then, the scientific (~1540s) and industrial revolution (~1780s) have resulted in countless inventions, which have left their environmental traces on Earth. This process, ideologically framed as era of "modernization" and narrated as a history of "progress", has left a long-lasting toll on the planet.[21] It is estimated that humans have transformed 20 to 100% of all land surfaces.[22,23] The release of pesticides, fertilizers and other chemical contaminants is interfering with natural ecosystems, resulting in large-scale biodiversity loss and deterioration of soil and water quality.[24] Recent extinction rates are estimated to be 100 to 1000 times higher than the average rate on geological time scales.[25] Chlorofluorocarbons (CFCs) have spread to the upper layer of the atmosphere and depleted the ozone layer, which protects the Earth from ultraviolet radiation coming from the sun.[26] Microplastics have been detected in various environments, including the air, oceans and freshwater eco-systems.[27] The aftermath of anthropogenic activities affecting the natural environment is expected to persist for long time periods (>50 000 years).

The various developments arising from the wasteful and harmful treatment of the environment by the human species stimulated the rise of preservationist and conservationist philosophies[28,29] in the late 19th and early 20th century, acknowledging the boundaries of nature and advocating a wise and efficient use of resources where humans should see themselves as an integrative part of the natural environment, rather than the conquerors thereof. This has evolved to the global environmental movement of today, which includes a wide range of activist organizations, political parties, scientific organizations, and governmental policies focused on promoting environmental values, and combatting detrimental behavior, such as the emission of greenhouse gases, deforestation and pollution.[30]

The United Nations assumed a key role in debating climate change and global environmental issues in the political sphere. The 1972 United Nations Conference on the Environment in Stockholm was the first international conference connecting economic growth and human well-being to environmental aspects such as the pollution of air, water and the oceans.[31] The 1992 United Nations Conference on Environment and Development (UNCEP), held in Rio de Janeiro, composed a declaration of 27 principles aimed at guiding countries to a sustainable future, where "human beings are entitled to live healthy and productive lives in harmony with nature".[32] The Montreal Protocol of 1987 aimed at phasing out ozone depleting substances such as CFCs.[33] It is, to date, the only environmental treaty ratified by all 198 UN member states and is regarded as a benchmark in multilateral environmental regulation. The Kyoto Protocol of 1997 and the Paris Agreement of 2015 aim at reducing greenhouse gas emissions and limiting global warming to 2°C compared to pre-industrial levels.[34,35] In 2001 the Stockholm Convention on persistent organic pollutants (POPs) was signed, a global treaty to minimize the risk posed by POPs to the environment.[36] The initial list of 12 ("the dirty dozen") compounds prepared by the intergovernmental negotiating committee (INC) is extended regularly and to date covers 23 unique compounds (including the commercial mixture of decabromodiphenyl ether, c-decaBDE) and 8 compound classes (including compounds that are listed with their salts, isomers or esters) to be eliminated or restricted.

History and the exposome: a conceptual discussion

Critique of current approach

In his famous 2005 paper, the epidemiologist Christopher Wild was one of the first to give the complementary concept to the genome a name: the exposome. According to Wild, the exposome includes "life-course environmental exposures (including lifestyle factors) from the prenatal period onwards".[37] Gary Miller and Dean Jones extended the concept of the exposome in 2014, shifting the focus of the exposome beyond solely the exposures.[38] They expanded the concept adding the aspects of diet, behavior and endogenous processes, with a particular focus on the biological responses to exposures. Even genetic and genomic alterations serve as evidence of past exposures.[39] The resulting shift in exposomics research from exposure-focused (introduced by Wild)[37] to more metabolomics focused approaches[38] to capture biological endpoints has been accompanied by several changes. Of course, the study of internal metabolic changes and the search for biomarkers in exposomics research is highly necessary, but not the sole goal of exposomics. The last years showed a trend to using increasingly expensive resources and sophisticated techniques, rather than focusing on the initial research question. Research is driven by metabolomics approaches using databases that contain mostly "known" knowledge, necessary to ensure workflow efficiency, yet at the same time trying to find new discoveries and improve the understanding of metabolic pathways. However, with the rapid pace of innovations in the age of the Anthropocene, new synthetic compounds (and their related transformation products) are appearing at an alarming rate with up to 20 million new registrations a year,[40] which can cause new problems and influence the human metabolism very differently. Although it is possible to create dynamic workflows and databases to profit from "new" knowledge generated via high throughput exposomics (eg, PubChemLite for Exposomics[41]), this requires concerted efforts at FAIR and Open data exchange,[42,43] for which resources and infrastructure are still under development. The exchange of information will be vital to support reinterpretation of older data (eg, retrospective screening[44,45]) or large community initiatives for large scale discovery such as the Global Natural Products Social Molecular Networking (GNPS) ecosystem.[46] However, while modern analytics and monitoring techniques can help capture relatively recent perturbations in the exposome, even using natural archives to an extent to investigate past pollution, there is a limit to the availability of suitable samples. The historical exposome concept can give additional perspectives to this technologically driven bio-exposome. Looking at environmental history as such, especially focusing on developments in the Anthropocene, can show many interconnections of environment and health, where expensive technological studies using biological samples of today may fail to find these connections. Historical archives can, for instance, reveal if currently perceived extreme events such as drought or flooding (relative to written records of ∼100 years) were in fact observed centuries earlier. Closer to the exposome, anecdotal, written documents (letters, newspaper articles) can also reveal the catastrophic side-effects of industrial pollution in factories, which is only partially captured in scientific literature (and if covered, often in foreign languages).[47]

Definitions of the environment

One way to approach the exposome from a more holistic perspective is to embed it into the wider framework of what has been called "the environment" since the early post-war years. As demonstrated by the three environmental historians Paul Warde, Libby Robbin, and Sverker Sörlin in their book "The Environment. A History of an Idea",[30] the term gained prominence as a political concept as a result of the success of popular science writers such as William Vogt ("Road to Survival", 1948)[48] and Rachel Carlson's best-seller "Silent Spring" (1962).[29] Although the term had been around for more than a century ("milieu" in French, "Umwelt" in German), a transformation of meaning from "a world where man was molded by environment to him being able to alter the nature of his world"[30 (p. 8)] only occurred with the "environmental revolution" in the 1970s.[49] Sparked by new academic interventions such as the "Limits to Growth" report to the Club of Rome in 1972[50] and the first "oil crisis" in 1973, ecological thinking became a mainstream concern for both scientists, environmental or ecological movements, and world politics. The problem of "the environment" became both scaled (ie, local problems of pollution or waste were interpreted as a subset of a planetary issue) and the result of a complex interplay of past, present, and future temporalities.[21]

To handle the growing complexity of issues at stake, new concepts such as the "ecosystem" were put forward to highlight the holistic nature of environmental changes. Inspired by the post-war boom of "systems theory" and "cybernetics", and long before climate science would emerge as a new interdisciplinary field, even Eugene and Howard Odum introduced the metaphor of balance between the living (organic) and nonliving (abiotic) environment in their textbook "Fundamentals of Ecology" (1953).[51] Ecosystem science became an important framework for setting up large-scale international research projects, shaping the thinking of the environment as a dynamic interplay between "nature" and "nurture" in a system called "earth" at a planetary scale.[30 (pp. 154–158)] Since the introduction of the concept of the "Anthropocene" by Paul Crutzen and Eugene Stoermer in 2000,[8] the human factor has entered the equation: according to recent scholarship, both the living and the nonliving environment has been radically affected by human activity on earth, crossing planetary boundaries.[52-54] With the human being acknowledged as an ecological factor in history (as had been suggested by historian long before),[55] the concept of environment as a tripartite dynamic system has emerged, combining the "natural" environment with the "human built" or socio-technical environment and the subjective environment of the individual.

This "holy trinity" of natural, individual, and social factors that form "the environment" suggests that the physical surrounding in which life takes place, which is comprised of the atmosphere, the biosphere, the hydrosphere and the lithosphere,[56 (p. 14)]) is fundamentally interconnected with human activities on earth. We cannot grasp the full complexity of one element without considering the impact of the others. The interconnected, dynamic relationships render "the environment" an incredibly complex system that seems hard to grasp—even from a truly interdisciplinary perspective. Yet it is this "maelstrom" of temporalities, scales, and uncertain causalities that challenges mono-disciplinary approaches and interdict mono-causal scientific explanations.[55] What is needed are new questions, concepts, and narratives aiming at forging an understanding of past and present environmental challenges, starting with a new search for evidence in and significance of existing data from both natural and social archives. Combining evidence from historical and natural sciences and thereby bridging the "two scientific cultures" is a key condition for performing historical exposomics research.

Finding significance in a different way

The search for significance is a common factor across all disciplines. Experts from the fields of humanities, natural and social sciences aim to make significant contributions in their areas of research. In a more general sense, significant contributions have the quality of being important, of being worthy of attention because they answer relevant questions and have meaningful results. In practice, the mindset, practices, and evaluation methods of assessing significance are often specific to each discipline, with a diverse range of views also present within each discipline.

One reason may be the nature of the information sources in each discipline. Focusing on the fields of history and natural sciences, a historian works more frequently with primary sources such as historical manuscripts or photographs, while a researcher in natural sciences often works with numerical data resulting from measurements. The inherent difference in the nature of the sources, as well as the *a priori* impossibility of making measurements in the past or repeating an experiment, lead to different approaches to the assessment of significance. Additionally, an apparent widespread dualism can be perceived, where natural sciences more often try to prove a theory that can be generalized for a larger population, whereas history is more frequently concerned with specific knowledge about a particular fact or occasion, with the consequent differences in the researcher's approach to significance.

Statistical hypothesis testing is an inference method widely used in natural and social sciences to determine a possible conclusion between two hypotheses. A null hypothesis is defined against an alternative hypothesis and the *P*-value is defined as the probability, calculated under the null hypothesis, that a test statistic is as extreme or more than its observed value. If the *P*-value is less than the selected significance level, then the null hypothesis is rejected. Traditionally, a $P\text{-value} < 0.05$ has been considered a reasonable significance level, but in some fields such as genomic studies, more stringent levels are adopted. The criticism of this method is extensive across fields of research,[57-60] including technical limitations such as the difference between the characteristics of the scientific data as opposed to the assumptions upon which the significance tests are defined, sampling issues regarding size and randomness, the arbitrary level of significance, the dichotomous reject/not-reject, the misinterpretation of *P*-values and the lack of reproducibility. However not only technical issues have been raised, but the actual scientific value of the tests has repeatedly been questioned, pointing out weaknesses in the use of isolated tests, the difficulty to generalize results and to integrate previous knowledge, causing erroneous scientific reasoning.

Numerous options have been proposed over decades of research, from the use of stricter *P*-values to the total abandonment of statistical significance, as well as a wide range of alternatives,[61-63] but the statistical significance in hypothesis testing is still a widely used method due to its intuitive interpretation, ease of calculation with existing tools, and facility given to the choice of the research path with yes/no questions. Bayesian approaches, older than frequentist statistics, gained popularity with the advances in computational methods. Bayesian methods depend on a prior and on the probability of the observed data, allowing the sensitivity of the experiment result to be measured for different priors.[64,65] In Bayesian inference, the uncertainty or "degree of belief" with respect to the parameters is quantified by probability distributions.

Controversy about statistics at a more general level includes the opposition of two attitudes to the question of reality, one realistic or objectivist (ie, pre-existing); and the other, relativistic or historicist (ie, constructed),[66] very present when bringing together history and natural sciences. In *Trust in Numbers*, Theodore M. Porter challenges the ubiquity of quantification in the sciences of nature, highlighting that only a small proportion of the numbers and quantitative expressions "make any pretense of embodying laws of nature, or event providing complete and accurate descriptions of the external world".[67]

There is an increasing interest in qualitative research methods across disciplines, such as research techniques that rely on non-statistical or numerical methods of data collection, analysis and evidence production. Qualitative research allows to retain complexity and nuance through data collection methods that adapt to the context and make it possible to explore emergent issues; and they present a reflexive approach that acknowledges the perspective of the researcher in the process.[68] They are used in the social sciences and the humanities, but they can also complement quantitative approaches in the natural sciences.[69] Qualitative research can be used independently to uncover topics that are not amenable to quantitative research. Furthermore, it can be used as the preliminary of quantitative research, that is to uncover ambiguities and misunderstandings regarding terms and definitions, to help understanding the reasons behind certain results, or to validate quantitative research by providing a different perspective on the same phenomena, sometimes forcing major reinterpretations of quantitative data.[70] Despite the many potential applications, qualitative methods are often evaluated according to quantitative measures of rigor and dismissed as unscientific and unreliable.[71]

Criteria of significance in history includes two main questions: what is important to learn about the past and how do historians know what they know. The first refers to deciding which events or people resulted in a change that had deep consequences. The second is about the historical interpretation based on inference made from sources: what questions turn a source into evidence, who created it, when and for what purpose, what is the historical context and w the inferences can be corroborated.[72] Partington[73] identifies three criteria of significance. First, the importance to people in the past "if we use the egalitarian principle of counting heads to establish priorities here we are in difficulty, because we have unequal access to the opinions and judgements of different social and ethnic groups [...]."[73] Second, objective criteria that includes profundity, as whether an event profoundly changed people's lives; quantity, as the number of lives impacted; and durability of the event in time. Thirdly, criterion is relevance, or how an event contributes to an increased understanding of the present.

An interdisciplinary approach to historical exposomics blurs the boundaries between discipline-specific approaches to significance and makes it possible to study the past from new perspectives. Historical sources like pollution measurements included in official reports, complaint letters from the citizens about the dust from the factories and contaminated water, newspapers publishing about new labor laws, innovative industrial techniques, or disease outbreaks, can be used to test hypotheses about pollution concentration or the impact on health in the past, while assessing at the same time for historical significance. The historical archives are filled with scientific and non-scientific, numerical, and non-numerical sources, that together with current scientific knowledge and measurements about water systems or use of chemicals, allow to simulate and validate hypothesis about the past. The dualism "specific" versus "universal" between science and history is replaced by a combination of close and distant reading from a historical and scientific point of view adapted to the source and specific analysis, which allows access to much more information, to analyze it from different points of view, and to use multiple approaches to assess significance.

Evidence from social and natural archives

A historical exposomics approach requires the combination of a great variety of historical and current data. As shown here using data collected for the research project "LuxTIME" as an example, the information gathered comes from both "natural" and "social" archives. Natural archives are physical objects which have been collected, processed and deposited in the environment under natural circumstances (ie, without the interference of the human species), and preserve information about the characteristics of the surroundings from the time and place where deposition took place.[74] Through various analysis techniques, this stored information can be extracted from the archive, enabling the reconstruction of environmental conditions far beyond the period where direct measurements have taken place. This helps determine the baseline conditions at the time, as well as spatial and temporal changes, allowing researchers to relate these to natural or anthropogenic influences.[75]

Social archives

Social archives refer to archives that are the result of a conscious and directed collection of past evidence, be it for cultural (religious, artistic), political (administrative, governmental, legislative), or economic (companies, trade) purposes. Information stored in such archives (be it textual, audio-visual, or material remains/objects) provide documentary evidence of past activities that have shaped the environment, be it through new production and consumption processes, regulatory or legal procedures, or political and cultural discourses on scientific or technological progress, nature protection or market and trade regulation. Based on methods of historical source criticism, social archival records are important to contextualize the development of science, technology, industry production, agriculture, energy consumption and all other human activities that have shaped our contemporary environment.[76] The information or data that can be extracted from social archives contains different kind of evidence, such as narrative data (eg, weather reports on newspapers), long series of quantitative proxies (eg, by procedures related to census or fiscal data), illustrative sources (photographs, drawings, maps), and instrumental measurements (eg, research diaries).[76] Such sources have been used extensively for the production of a rich historiography on Luxembourgish steel industry and the industrialization of the Minett region more generally.[77-79] Comparative studies on the emergence of coal and steel industries in various European regions have highlighted the specificities of the Luxembourgish development, both in terms of availability of natural resources and its dependencies on technology and capital transfers (especially from the Ruhr region), economic alliances (such as the adherence to the German "Zollverein" [customs union] until the end of the Great War), and political constellations (such as the creation of the European Community of Coal and Steel [ECCS] in 1951).[80-84]

The LuxTIME research efforts included consultation of many different archives in Luxembourg, including national and local archives, libraries, museums, and deposits of public administrations (see Figure 1).[85] The sources from these archives were categorized and indexed, specifying the temporal and spatial coverage as well as the exposome category that the sources are referring to (eg, physical/chemical; ecosystem; lifestyle, etc.). The analysis of this data inventory of the social archives has been turned into a treemap showing the number of datasets covering exposome categories and subcategories, see Figure 2.

By means of example (without any claim for comprehensiveness), the following sections describe several natural archives that have been studied or were considered in the framework of the LuxTIME project.

Historical groundwater reserves

Groundwater serves as an archive for historical climatic and hydrological conditions. When aquifer recharge takes place, the ambient environmental conditions are stored in the precipitated water as chemical and isotopic signals, which preserve sequential changes as the groundwater moves away from the point of recharge.[86] Extracted groundwater samples, once related to a proper chronology, for example through radiocarbon dating (^{14}C), have been found to provide indications about the temperature, air-mass circulation and rainfall intensity of the past.[87-89] In its simplest sense, the mere presence of dated groundwater can be an indicator for prolonged wet periods, while absence thereof is a sign of no recharge, possibly due to a period of drought or ice cover.[86,90] The concentration of dissolved noble gases in groundwaters is a function of the temperature and salinity at the time of recharge (Henry's Law). The concentration decreases with increasing temperature.[91] Since these concentrations are largely unaffected in groundwater bodies, this relationship can be utilized to establish historical temperature records.[87,89,92] The stable isotopes of water (H and O) are separated into heavier and lighter compounds as they move through the hydrological cycle. This creates an isotopic ratio characteristic to the air temperature and humidity at the time of groundwater recharge, as well as environmental processes such as evaporation and condensation.[93] These isotopic ratios have been utilized to reconstruct air-mass circulation and temperatures of the past.[89,92] Age dating uncertainties and mixing of different waters in the aquifer over time creates a smoothing of the signal fluctuations. Groundwater thus serves as an archive which stores an average of the conditions at glacial/interglacial time scales.[89]

Freshwater bivalves

Stable isotopes of oxygen and hydrogen in precipitation and stream water are used for conceptualizing and modeling hydrological, ecological, biogeochemical, and atmospheric processes.[94-97] However, their full potential—for example in climate change impact studies[98] or climate and earth system modeling[99]—cannot be leveraged due to short and truncated time series.[100,101] A promising way forward for reconstructing the history of flowing waters is the conceptualization of rivers as living entities.[102,103] Past changes in eco-hydrological catchment functions and their impacts on freshwater habitats may eventually be recorded in natural archives (and *vice versa*). This feature qualifies trees and freshwater bivalves as (biotic) sensors for reconstructing chronologies of past climate and flowing waters into pre-instrumental times. During the growth process, bivalves (or mollusks) record environmental data in their shells in the form of variable geochemical and microstructural properties, as well as variable increment widths, like tree rings. Controlled by biological clocks, the shell growth process occurs periodically and eventually results in the formation of growth patterns (representing periods of fast and slow growth) that can then be placed in a precise temporal context. By archiving in-stream environmental

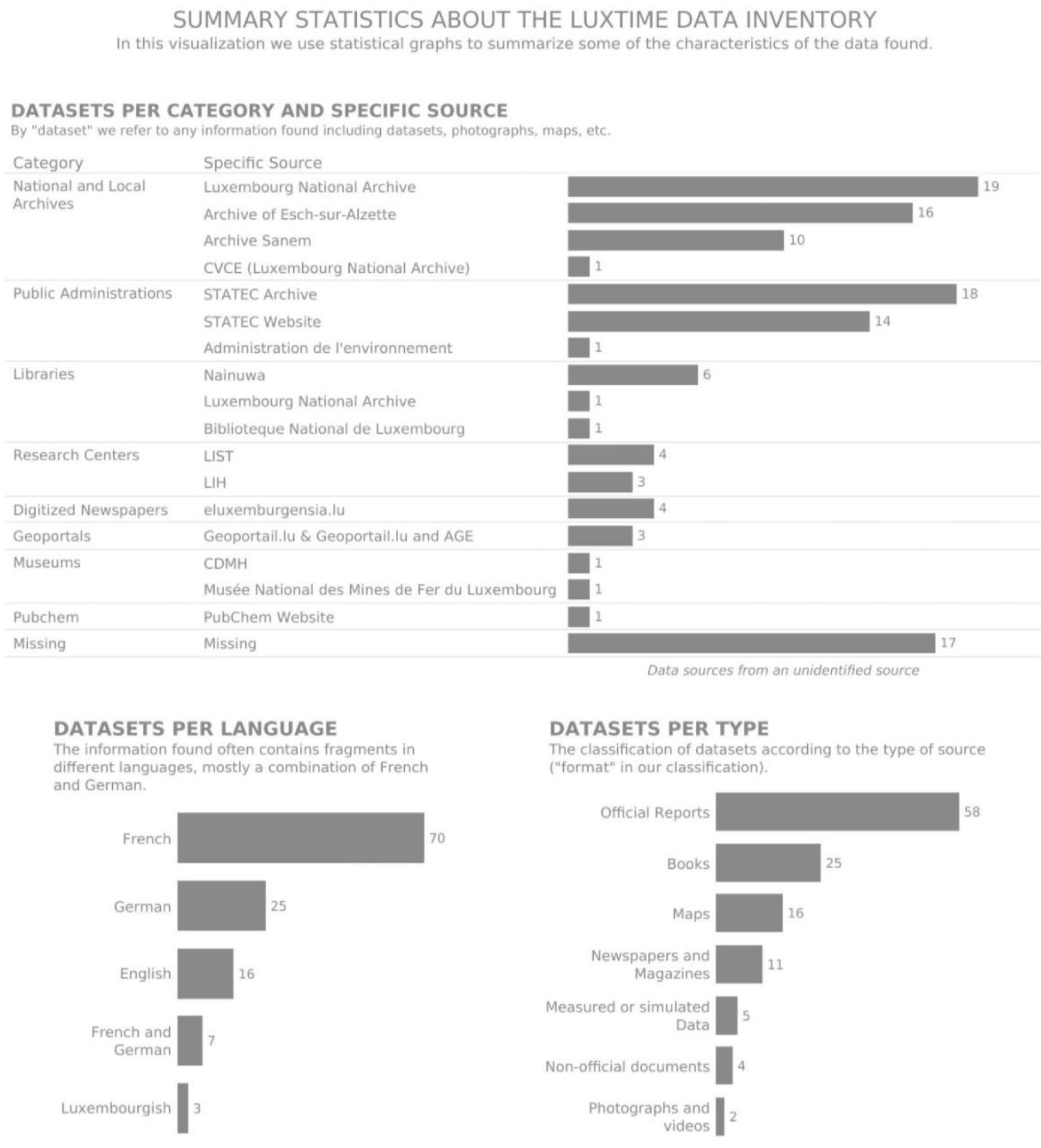


Figure 1. Summary Statistics Data Inventory for the LuxTIME project.

conditions,[104,105] freshwater bivalves thus show considerable potential for complementing stream water isotope records. By forming their shells near equilibrium with the oxygen isotope value of ambient water, changes in isotope values in the shell (represented as $\delta^{18}O$ values) can serve two different purposes. They may help in reconstructions of both stream water temperature (if water $\delta^{18}O$ values are known) and $\delta^{18}O$ values (provided water temperature during shell formation is known or can be reconstructed by other means). While stream water temperature reconstructions have been extensively used,[104] hydroclimate reconstructions have received considerably less attention.[105] Recent proof-of-concept work has demonstrated the potential for freshwater mollusks to solve the problem of limited $\delta^{18}O$ isotope records in stream water.[106,107] New analytical protocols based on Secondary Ion Mass Spectrometry have revealed the previously untapped variability in stream water $\delta^{18}O$ signatures over nearly 200 years and their connection with changes in atmospheric circulation patterns.

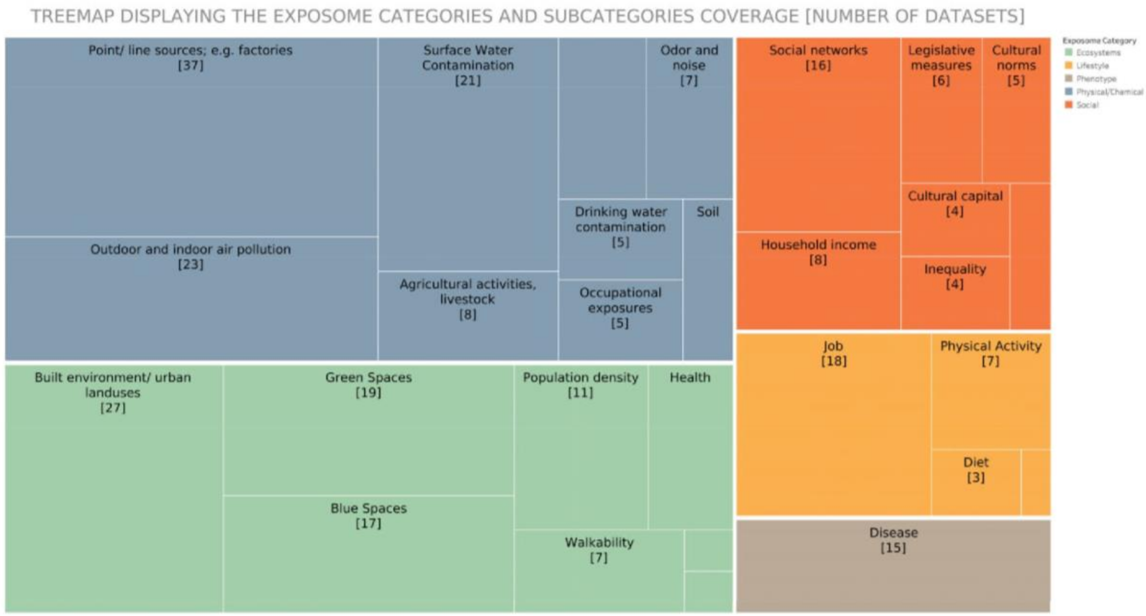


Figure 2. Treemap visualization of exposome categories and subcategories covered by the data inventory.

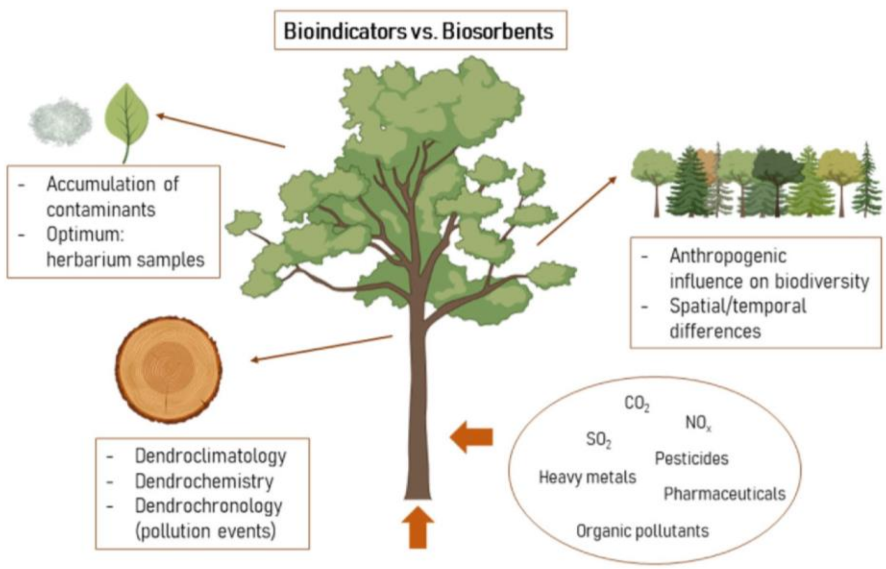


Figure 3. The function of trees as bioindicators versus biosorbents.

Botanical samples

Many studies show that botanical samples can serve as evidence for the impact of man in the age of the Anthropocene.[108-113] Plants, as sensitive organisms, respond strongly to anthropogenic influences such as air pollution. The susceptibility of plants to air pollution can be evaluated by their Air Pollution Tolerance Index (APTI), which is determined by the parameters ascorbic acid, chlorophyll, relative water content, and leaf-extract pH.[114] The APTI is often used studying the potential of plants as biosorbents (remediation activities)[115,116] or bioindicators (biomonitors)[113,117] of pollution (Figure 3).

Plants, such as trees, can be considered as evidence for historical events and thus as natural archives. Trees can provide accurate information about climatic conditions in the past,[118] using dendrochronology information (dendroclimatology). Additionally, those measurements can help to differentiate between natural and anthropogenic causes (increase of CO_2 and other gases) of changing environmental conditions.[119]

The population of specific species in certain areas is often connected to human activities. For example, tree species such as birch and alder grow well at post-mining sites, even in highly "contaminated" mine soils, and can even change the microbial properties of the soil.[120,121] As the biodiversity of an area depends directly on factors such as soil, air or water contamination, any changes in plant species may correspond directly or indirectly to an (anthropogenic) change in the direct ecosystem. Analyzing the age of those species could even provide information about the date of these environmental changes, for example, the time of reforestation in post-industrial areas. While looking at historical documents about the influence of air pollution on plants, geographical mapping of species can be found, even if no measurement data is available.[122] If one examines plant species for environmental contamination, one can—under favorable circumstances—make statements about the health of an ecosystem (even in the past). This can be done, for example by analyzing lichens,[123-125] which act as a chemical filter for air pollution, accumulating contaminants over many years. The same applies for tree bark and leaves,[126] however, the time point of pollution is not analyzed in most studies. In the optimal case, one can resort to archived plant samples from herbaria, for example, for the analysis of per- and polyfluoroalkyl substances (PFAS) in pine needles,[127] although samples need to be stored suitably to allow for this analysis. Moreover, using state of the art techniques like matrix-assisted laser desorption ionization—imaging mass spectrometry (MALDI-IMS) allows the monitoring and localization of the contaminant uptake and distribution pathways through the plant.[128,129] There are some studies using tree rings as archives of atmospheric pollution, which can be analyzed in a temporal and spatial manner (dendrochemistry).[108,112] However, there are several limitations and uncertainty factors, such as soil acidity or growth rate to be considered.[130,131] In general, looking at botanical samples always has some uncertainty factors due to plant metabolic processes altering concentrations, as well as external influences (eg, sample handling) influencing all research outcomes—but could nonetheless reveal interesting information about the past that may not otherwise be accessible.

Data visualization

Both natural sciences and humanities use data in their research, whether they are historical or modern texts, measurements from observations or experiments, elements of a photograph, time series, geographical data or many more. Data is the connecting element in interdisciplinary research. Various disciplines often share data analysis methods, but also share the same challenges, such as how to discover what is relevant for their specific research question when dealing with a large volume of data, how to validate the data, or how to communicate the results to readers according to their specific needs. They also face fundamental differences in the definition and use of data and the required communication objectives. Data visualization supports multiple scenarios including discovery and communication. It helps to navigate large volumes of data, whether to perform an initial exploratory data analysis to define the line of research, or to make discoveries that form part of the project results. It is also essential for communicating results, which allows creation of the desired user experience in each case.

Especially in the context of interdisciplinary projects, it is important to understand the different approaches to data visualization since a combination of these might be required for the analysis of data from the point of view of different disciplines. In the case of the natural sciences, data visualization is generally used as a tool to simplify a complex topic, to make data-driven decisions about the best path of research, to build upon previous scientific knowledge by integrating references and benchmarks, and to communicate a clear message to an audience. Scientific data visualization is largely based on statistical graphics which are characterized by the principles of abstraction, reduction, standardization, representation, and legibility. The user is expected to be familiar with the visual elements—the statistical charts, for example, bar charts, line charts—and to ask predefined questions to get formatted answers, to apply filters, and to drill down to predefined levels of aggregation. In the case of the humanities, the use of data visualization often aims at showing the complexity of the subject of research, to engage the user in the interpretation and to present multiple narratives. The data visualization is defined by its granularity, specificity, and full coverage.

To study historical exposomics, data visualization techniques such as concept maps can be used to understand the specific contributions of each field and the joint area of work, see Figure 4.[85]

During the exploration of the data, the use of data visualization allows identification of the types of data available—time series, geographical, texts, images—the geographical area and the period covered, the variety of sources—historical archives, sampling campaigns, web archives—the domains—air, water and soil pollution, demographics, health, industry, urbanism, green and blue areas—and the percentage of sources digitized, among others. This facilitates the identification of gaps, areas of interest and future challenges. Lastly, one of the most frequent use of data visualization is to share the research output. A combined approach to data visualization, scientific and humanistic, allows integration of representative analyses to understand the general impact of the main environmental exposures on health and exploration of the different narratives by integrating individual cases, when relevant.[85]

New historical narratives

In studying historical evidence of human traces on the environment by applying both scientific and hermeneutic methods of interpretation, LuxTIME aims at contributing to a new form of data-driven scholarship on the Anthropocene by experimenting with digital forms of historical storytelling, such as animations of dust pollution in Esch-sur-Alzette over nearly one century (1911–1996) shown in the website "minett-stories" and in Figure 5.

The animation that was produced within the virtual exhibition "Minett Stories"[132]—an interdisciplinary public history project in the framework of the "Esch22/Capital of Europe" initiative,[133] was the result of a collaboration between historians, metallurgy experts and chemists. Based on a data set including information about the production output of the Belval steel plant throughout the 20th century, the type of production process (eg, LD-AC or Thomas for steelmaking), the typical usage of filtering systems, and the prevailing wind directions in Luxembourg over time, the animation offers a plausible approximation of how the dust emissions of the Belval iron and steel complex might have affected local air quality over the decades. Even if the animated map can only offer a hypothetical approximation rather than an exact reproduction of past reality, such temporal visualizations can trigger our historical imagination and open new paths for further historical investigation.

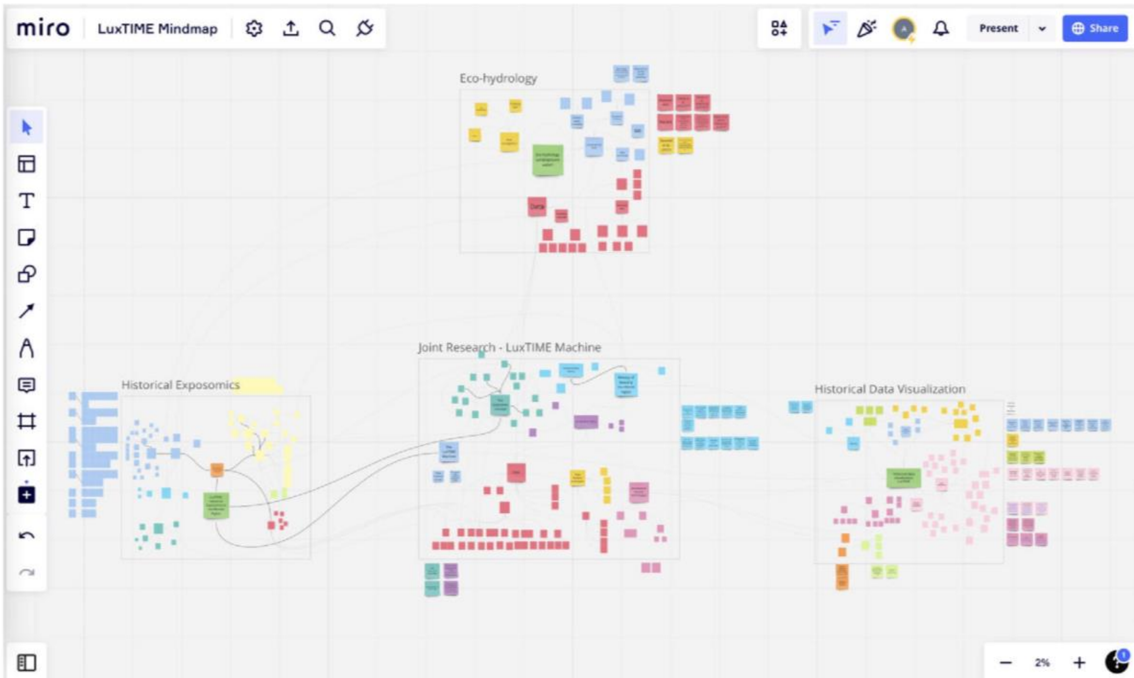


Figure 4. LuxTIME Concept Map including History, Data Visualization, Environmental Cheminformatics and Eco-hydrology.

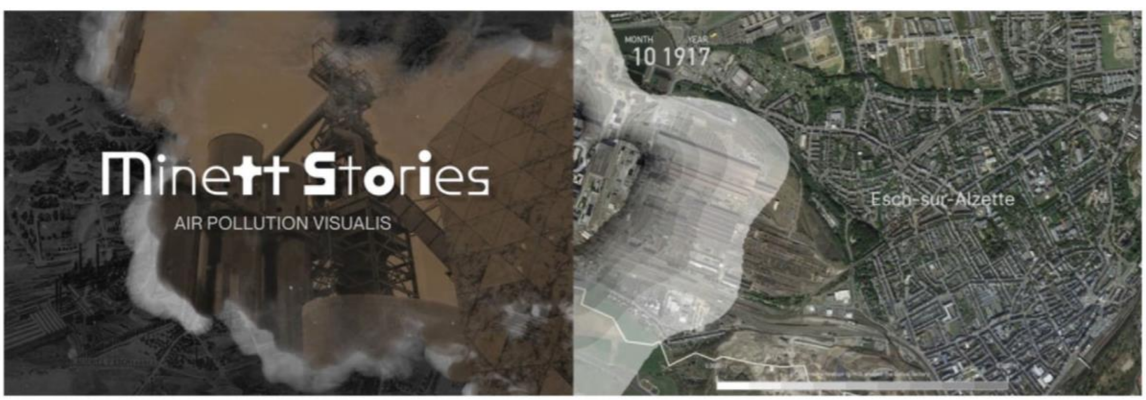


Figure 5. Screenshots from the Minett Stories Website (https://minett-stories.lu/en/story/air-pollution-visualized).

Conclusions

Bridging the gap between "nature" and "nurture" by bringing a great variety of different historical sources and datasets together is an exciting and challenging exercise in interdisciplinary collaboration. While the LuxTIME project was conceived as a testbed for data-driven scholarship in the field of environmental history with a focus on the Minett region, the interdisciplinary collaboration in the team has opened new research questions and approaches beyond the field of environmental humanities. In exploring the concept of "historical exposomics", this article hopes to make a convincing argument for enlarging the scope of current research on the exposome by enriching it by historical evidence from "social archives". In contextualizing the data from both natural and social archives and investigating the causal relationships between factual evidence from the sciences and the narrative evidence stemming from historical sources, new research perspectives on the history of the Anthropocene can be opened, challenging classical forms and formats of historical storytelling and interpretation.[134] The concept of historical exposome, such is our hypothesis, might offer a useful conceptual framework for studying the Anthropocene in a truly interdisciplinary fashion.

Future efforts in this field require research and training in "digital hermeneutics", looking critically at the use of digital tools and (big) data of all kinds (as covered by the Digital History &

Hermeneutics Doctoral Training Unit DTU of the C^2DH).[135] Interdisciplinary collaboration is required for the testing of hypotheses and scientific interpretation of evidence studying the historical exposome. Finding a common language to discuss past, present and future exposomics-related developments is required to tackle this challenging topic.

The historical exposome is presented here as an interdisciplinary approach to address the exposome. Adding value from both perspectives, natural sciences and humanities to study this highly complex paradigm opens new possibilities for present and future research. Instead of solely performing expensive cohort studies looking at the present state, looking at past digital data already present can provide valuable insights. If scientists wish to learn from the past, they must dig into the past, as physical evidence often is not present to be analyzed today. Social archives provide a wealth of soft and hard data sources of information. It is worth investing time to look at historical documents, perform simulations to estimate the past state and even go one step further and simulate future developments. In terms of environmental pollution, looking at past evidence might often prove to be more useful to find for example connections to present day diseases, than analyzing present evidence (eg, looking at past exposure to pesticides studying present day Parkinson's disease[136]). Combining it with today's state of knowledge and tools can help to improve the understanding of the exposome and may help prevent further environmental contamination and disease.

Acknowledgments

We would like to thank all the LuxTIME team members who took part in the discussion.

Author Contributions

Dagny Aurich (Conceptualization [equal], Data curation [equal], Visualization [equal], Writing—original draft [equal], Writing—review and editing [lead]), Aida Horaniet Ibanez (Conceptualization [equal], Data curation [equal], Visualization [equal], Writing—original draft [equal], Writing—review and editing [equal]), Christophe Hissler (Conceptualization [equal], Writing—original draft [supporting], Writing—review and editing [equal]), Simon Kreipl (Conceptualization [supporting], Writing—original draft [equal], Writing—review and editing [equal]), Laurent Pfister (Conceptualization [equal], Writing—original draft [equal], Writing—review and editing [equal]), Emma Schymanski (Conceptualization [equal], Funding acquisition [equal], Supervision [equal], Writing—original draft [supporting], Writing—review and editing [lead]), and Andreas Fickers (Conceptualization [equal], Funding acquisition [equal], Supervision [equal], Writing—original draft [equal], Writing—review and editing [equal])

Funding

D.A. and A.H.I. are funded by the University of Luxembourg Institute for Advanced Studies Audacity Program and S.K. has been funded by the Luxembourg Institute of Science and Technology (LIST) for the Luxembourg Time Machine (LuxTIME). E.L.S. acknowledges funding from the Luxembourg National Research Fund (FNR) for project A18/BM/12341006.

Conflict of interest statement

None declared.

Data availability

Not applicable.

References

1. Misawa K. Nature and nurture. In: Smeyers P, ed. International Handbook of Philosophy of Education. Springer International Handbooks of Education. Springer International Publishing; 2018:905–919. https://doi.org/10.1007/978-3-319-72761-5_64.
2. Canali S. What is new about the exposome? Exploring scientific change in contemporary epidemiology. Int J Environ Res Public Health. 2020;17(8):2879. https://doi.org/10.3390/ijerph17082879.
3. Galison P. Computer simulations and the trading zone. In: Galison P, Stump DJ, eds. The Disunity of Science: Boundaries, Contexts, and Power. Stanford University Press; 1996:118–157.
4. University of Luxembourg. Luxembourg Time Machine. Published 2021. Accessed March 1, 2022. https://luxtimemachine.uni.lu/.
5. Time Machine Europe. Time Machine Europe. Accessed December 8, 2022. https://www.timemachine.eu/.
6. Banken R. Early start and late breakthrough: the industrialization of the Minette region in Lorraine and Luxembourg. In: Czierpka J, Oerters K, Thorade N, eds. Regions, Industries, and Heritage. Palgrave Macmillan, UK; 2015:54–72. https://doi.org/10.1057/9781137333414_5.
7. Hughes JD. An Environmental History of the World: Humankind's Changing Role in the Community of Life. 2nd ed. Routledge; 2009.
8. Crutzen PJ, Stoermer EF. The 'Anthropocene' (2000). In: Robin L, Sörlin S, Warde P, eds. The Future of Nature. Yale University Press; 2017:479–490. https://doi.org/10.12987/9780300188479-041.
9. Waldron P. Critical zone science comes of age. Eos. 2020;101. https://doi.org/10.1029/2020EO148734.
10. Latour B. Some advantages of the notion of "critical zone" for geopolitics. Procedia Earth Planet Sci. 2014;10:3–6. https://doi.org/10.1016/j.proeps.2014.08.002.
11. Jones J. 'Critical Zones' reimagines humanity's relationship with the earth. *Stir World*. Published 2020. Accessed January 12, 2023. https://www.stirworld.com/inspire-people-critical-zones-reimagines-humanitys-relationship-with-the-earth.
12. Solnick S, Latour B, Peter W, eds. Critical zones. *e-flux criticism*. Published 2020. Accessed March 20, 2022. https://www.e-flux.com/criticism/367892/bruno-latour-and-peter-weibel-eds-critical-zones.
13. Chakrabarty D. Anthropocene time. Hist Theory. 2018;57(1):5–32. https://doi.org/10.1111/hith.12044.
14. Sojka RE, Bjorneberg DL, Entry JA. Irrigation: historical perspective. In: Encyclopedia of Soil Science, 3rd ed. CRC Press; 2017: 1264–1268. https://doi.org/10.1081/E-ESS3-120042707.
15. Turchin P, Whitehouse H, Gavrilets S, *et al.* Disentangling the evolutionary drivers of social complexity: a comprehensive test of hypotheses. Sci Adv. 2022;8(25):eabn3517. https://doi.org/10.1126/sciadv.abn3517.
16. Headrick DR. Humans versus Nature: A Global Environmental History. Oxford University Press; 2020.
17. Nentwig W. 19 Human Environmental impact in the Paleolithic and Neolithic. In: Henke W, Tattersall I. eds. Handbook of

Paleoanthropology. Springer; 2007:1881–1900. https://doi.org/10.1007/978-3-540-33761-4_62.
18. Uekötter F. Im Strudel: Eine Umweltgeschichte Der Modernen Welt. Campus; 2020.
19. Cloy JM, Farmer JG, Graham MC, MacKenzie AB, Cook GT. Historical records of atmospheric Pb deposition in four Scottish ombrotrophic peat bogs: an isotopic comparison with other records from western Europe and Greenland. Global Biogeochem Cycles. 2008;22(2):1–16. https://doi.org/10.1029/2007GB003059.
20. Stille P, Pourcelot L, Granet M, *et al.* Deposition and migration of atmospheric Pb in soils from a forested silicate catchment today and in the past (Strengbach case): evidence from 210Pb activities and Pb isotope ratios. Chem Geol. 2011;289(1–2):140–153. https://doi.org/10.1016/j.chemgeo.2011.07.021.
21. Brooke JL. Climate Change and the Course of Global History: A Rough Journey. Cambridge University Press; 2014.
22. Watson RT, Rodhe H, Oescher H, *et al.* Intergovernmental Panel on Climate Change. In: Houghton JT, Jenkins GJ, Ephraums JJ, eds. Climate Change: The IPCC Scientific Assessment. Cambridge University Press; 1990.
23. Hooke RLeB, Martín-Duque JF. Land transformation by humans: a review. GSA Today. 2012;12(12):4–10. https://doi.org/10.1130/GSAT151A.1.
24. Brühl CA, Zaller JG. Biodiversity decline as a consequence of an inappropriate environmental risk assessment of pesticides. Front Environ Sci. 2019;7:177. https://doi.org/10.3389/fenvs.2019.00177.
25. Pimm SL, Russell GJ, Gittleman JL, Brooks TM. The future of biodiversity. Science. 1995;269(5222):347–350. https://doi.org/10.1126/science.269.5222.347.
26. Molina MJ, Rowland FS. Stratospheric sink for chlorofluoromethanes: chlorine atom-catalysed destruction of ozone. Nature. 1974;249(5460):810–812. https://doi.org/10.1038/249810a0.
27. Barnes DKA, Galgani F, Thompson RC, Barlaz M. Accumulation and fragmentation of plastic debris in global environments. Philos Trans R Soc Lond B Biol Sci. 2009;364(1526):1985–1998. https://doi.org/10.1098/rstb.2008.0205.
28. Leopold A. A Sand County Almanac and Sketches Here and There. Special commemorative ed. Oxford University Press; 1989.
29. Carson R, Darling L, Darling L. Silent Spring. Houghton Mifflin; Riverside Press; 1962.
30. Warde P, Robin L, Sörlin S. *The Environment: A History of the Idea*. John Hopkins University Press. 2018.
31. UN. *Report of the United Nations Conference on the Human Environment, Stockholm, June 5–16, 1972*. 1973. Accessed March 31, 2022. https://digitallibrary.un.org/record/523249.
32. UN. *Report of the United Nations Conference on Environment and Development*. 1992. Accessed March 31, 2022. https://www.un.org/esa/dsd/agenda21/Agenda%2021.pdf.
33. UN. *The Montreal Protocol on Substances That Deplete the Ozone Layer*. Vol. 1522. 1989. Accessed November 20, 2022. https://treaties.un.org/doc/publication/unts/volume%201522/volume-1522-i-26369-english.pdf.
34. UN. *Kyoto Protocol to the United Nations Framework Convention on Climate Change*. United Nations Publication; 1998. Accessed 20 November 2022. https://unfccc.int/resource/docs/convkp/kpeng.pdf.
35. UN. *Paris Agreement*. 2015. Accessed March 31, 2022. https://unfccc.int/sites/default/files/english_paris_agreement.pdf.
36. Stockholm Convention. Stockholm Convention-Home page. Accessed March 13, 2023. http://www.pops.int/.
37. Wild CP. Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. Cancer Epidemiol Biomarkers Prev. 2005;14(8):1847–1850. https://doi.org/10.1158/1055-9965.EPI-05-0456.
38. Miller GW, Jones DP. The nature of nurture: refining the definition of the exposome. Toxicol Sci. 2014;137(1):1–2. https://doi.org/10.1093/toxsci/kft251.
39. Li S, Hursting SD, Davis BJ, McLachlan JA, Barrett JC. Environmental exposure, DNA methylation, and gene regulation. Ann N Y Acad Sci. 2003;983(1):161–169. https://doi.org/10.1111/j.1749-6632.2003.tb05971.x.
40. Arp HPH, Aurich D, Schymanski EL, Sims K, Hale SE. Avoiding the next silent spring: our chemical past, present, and future. Environ Sci Technol. 2023;57(16):6355–6359. https://doi.org/10.1021/acs.est.3c01735.
41. Schymanski EL, Kondić T, Neumann S, Thiessen PA, Zhang J, Bolton EE. Empowering large chemical knowledge bases for exposomics: PubChemLite meets MetFrag. J Cheminform. 2021;13(1):19. https://doi.org/10.1186/s13321-021-00489-0.
42. Schymanski EL, Bolton EE. FAIRifying the exposome journal: templates for chemical structures and transformations. Exposome. 2022;2(1)osab006. https://doi.org/10.1093/exposome/osab006.
43. Schymanski EL, Schymanski SJ. Water science must be Open Science. Nat Water. 2023;1(1):4–6. https://doi.org/10.1038/s44221-022-00014-z.
44. Alygizakis NA, Samanipour S, Hollender J, *et al.* Exploring the potential of a global emerging contaminant early warning network through the use of retrospective suspect screening with high-resolution mass spectrometry. Environ Sci Technol. 2018;52(9):5135–5144. https://doi.org/10.1021/acs.est.8b00365.
45. Alygizakis NA, Oswald P, Thomaidis NS, *et al.* NORMAN digital sample freezing platform: A European virtual platform to exchange liquid chromatography high resolution-mass spectrometry data and screen suspects in "digitally frozen" environmental samples. TrAC Trends Anal Chem. 2019;115:129–137. https://doi.org/10.1016/j.trac.2019.04.008.
46. Wang M, Carver JJ, Phelan VV, *et al.* Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat Biotechnol. 2016;34(8):828–837. https://doi.org/10.1038/nbt.3597.
47. Blanc PD. Fake Silk: The Lethal History of Viscose Rayon. 1st ed. Yale University Press; 2016. https://doi.org/10.12987/yale/9780300204667.001.0001.
48. Vogt W, Baruch BM, Freeman SI. Road to Survival. W. Sloane Associates; 1948.
49. Chien JP. Umwelt, milieu(x), and environment: a survey of cross-cultural concept mutations. Semiotica. 2007;2007(167):65–89. https://doi.org/10.1515/SEM.2007.071.
50. Meadows DH, Meadows DL, Randers J, Behrens WW. The Limits to Growth: A Report for the Club of Rome's Project on the Predicament of Mankind. Universe Books; 1972.
51. Odum EP, Odum HT. Fundamentals of Ecology. Saunders; 1953. Accessed November 25, 2022. https://books.google.de/books?id=EawlAAAAMAAJ.
52. Persson L, Carney Almroth BM, Collins CD, *et al.* Outside the safe operating space of the planetary boundary for novel entities. Environ Sci Technol. 2022;56(3):1510–1521. https://doi.org/10.1021/acs.est.1c04158.
53. Rockström J, Steffen W, Noone K, *et al.* Planetary boundaries: exploring the safe operating space for humanity. E&S. 2009;14(2). https://doi.org/10.5751/ES-03180-140232.
54. Steffen W, Richardson K, Rockström J, *et al.* Planetary boundaries: guiding human development on a changing planet.

Science. 2015;347(6223):1259855. https://doi.org/10.1126/science.1259855.
55. White L. The historical roots of our ecological crisis. Science. 1967;155(3767):1203–1207. https://doi.org/10.1126/science.155.3767.120.
56. Grotzinger JP, Jordan TH. Understanding Earth. 7th ed. W.H. Freeman and Company, a Macmillan Higher Education Company; 2014.
57. Halsey LG. The reign of the p-value is over: what alternative analyses could we employ to fill the power vacuum? Biol Lett. 2019;15(5):20190174. https://doi.org/10.1098/rsbl.2019.0174.
58. Dahiru T. P-value, a true test of statistical significance? A cautionary note. Ann Ib Postgrad Med. 2008;6(1):21–26. https://doi.org/10.4314/aipm.v6i1.64038.
59. Goodman S. A dirty dozen: twelve p-value misconceptions. Semin Hematol. 2008;45(3):135–140. https://doi.org/10.1053/j.seminhematol.2008.04.003.
60. Wagenmakers EJ. A practical solution to the pervasive problems of p values. Psychon Bull Rev. 2007;14(5):779–804. https://doi.org/10.3758/BF03194105.
61. Benjamin DJ, Berger JO. Three recommendations for improving the use of *p*-values. Am Stat. 2019;73(Supp 1):186–191. https://doi.org/10.1080/00031305.2018.1543135.
62. Ioannidis JPA. The proposal to lower *p* value thresholds to.005. JAMA. 2018;319(14):1429–1430. https://doi.org/10.1001/jama.2018.1536.
63. Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. Nature 2019;567(7748):305–307. https://doi.org/10.1038/d41586-019-00857-9.
64. Bayarri MJ, Berger JO. The interplay of Bayesian and frequentist analysis. Statist Sci. 2004;19(1). https://doi.org/10.1214/088342304000000116.
65. Hackenberger, BK. Bayes or not Bayes, is this the question? Croat Med J. 2020;61 (6):564–568. https://doi.org/10.3325/cmj.2019.60.50.
66. Desrosières A. La Politique Des Grands Nombres: Histoire de La Raison Statistique. La Découverte/Poche 2010.
67. Porter TM. Trust in Numbers: The Pursuit of Objectivity in Science and Public Life. Princeton University Press; 1996.
68. Ritchie J, Lewis J, McNaughton Nicholls C, Ormston R. Qualitative Research Practice: A Guide for Social Science Students and Researchers. 2nd ed. SAGE; 2014.
69. Bhangu S, Provost F, Caduff C. Introduction to qualitative research methods–Part I. Perspect Clin Res. 2023;14(1):39–42. https://doi.org/10.4103/picr.picr_253_22.
70. Pope C, Mays N. Qualitative methods in health research. In: Pope C, Mays N, eds. Qualitative Research in Health Care. Blackwell Publishing Ltd; 2006:1–11. https://doi.org/10.1002/9780470750841.ch1.
71. McAleese S, Kilty J. Stories matter: reaffirming the value of qualitative research. Qual Rep. Published online April 17, 2019. Accessed November 25, 2022. https://doi.org/10.46743/2160-3715/2019.3713.
72. Seixas P, Morton T. The Big Six: Historical Thinking Concepts. Nelson Education; 2013.
73. Partington G. What history should we teach? Oxf Rev Educ. 1980;6(2):157–176. https://doi.org/10.1080/0305498800060204.
74. Nowak Z. Theorising the natural archive. Environ Hist Camb. 2022;28(1):105–127. https://doi.org/10.3197/096734019X15740974883852.
75. Blais JM, Rosen MR, Smol JP, eds. Environmental Contaminants: Using Natural Archives to Track Sources and Long-Term Trends of Pollution. 1st ed. Imprint: Springer; 2015. https://doi.org/10.1007/978-94-017-9541-8.
76. Pfister C, Wanner H. Climate and Society in Europe: The Last Thousand Years. Haupt Verlag; 2021.
77. Trausch G. Luxemburg, ein Land vom Stahl geprägt. Banque de Luxembourg; 1995:28.
78. Barthel C, Kirps J, Feltes P. Terres rouges, histoire de la sidérurgie luxembourgeoise. Archives nationales du Luxembourg Centre d'études et de recherches européennes Robert Schuman; 2011.
79. Knebeler C, Scuto D. Belval: Passé, présent et avenir d'un site luxembourgeois exceptionnel (1911-2011). Le Phare; 2010.
80. Leboutte R, Lehners JP. Formation et muations des bassins industriels en Europe. Passé et Avenir Des Bassins Industriel En Europe. Centre Universitaire de Luxembourg; 1995.
81. Leboutte R. Vie et Mort Des Bassins Industriels En Europe. L'Harmattan; 2013.
82. Leboutte R, Puissant J, Scuto D. Un Siècle D'histoire Industrielle (1873 - 1973): Belgique, Luxembourg, Pays-Bas; Industrialisation et Société. SEDES; 1998.
83. Raphael L. Jenseits Von Kohle Und Stahl: Eine Gesellschaftsgeschichte Westeuropas Nach Dem Boom: Frankfurter Adorno-Vorlesungen 2018. Erste Auflage, Suhrkamp; 2019.
84. Brüggemeier FJ. Grubengold: Das Zeitalter Der Kohle Von 1750 Bis Heute. C.H.Beck; 2018.
85. Aurich D, Horaniet Ibanez A. How can data visualization support interdisciplinary research? LuxTIME: Studying historical exposomics in Belval. (forthcoming).
86. Edmunds WM. Groundwater as an archive of climatic and environmental change. In: Aggarwal PK, Gat J, Froehlich KFO, eds. Isotopes in the Water Cycle: Past, Present and Future of a Developing Science. Springer; 2005:341–352.
87. Andrews JN, Lee DJ. Inert gases in groundwater from the Bunter Sandstone of England as indicators of age and palaeoclimatic trends. J Hydrol. 1979;41(3–4):233–252. https://doi.org/10.1016/0022-1694(79)90064-7.
88. Gasse F. Hydrological changes in the African tropics since the Last Glacial Maximum. Quat Sci Rev. 2000;19(1–5):189–211. https://doi.org/10.1016/S0277-3791(99)00061-X.
89. Stute M, Talma AS. Glacial temperatures and moisture transport regimes reconstructed from noble gases and δ18O, Stampriet aquifer, Namibia. In: Isotope Techniques in the Study of Environmental Change: Proceedings of an International Symposium on Isotope Techniques in the Study of past and Current Environmental Changes in the Hydrosphere and the Atmosphere. Proceedings series. International Atomic Energy Agency; 1998:307–318.
90. Sonntag C, Muennich KO, Junghans C, *et al.* Palaeoclimatic information from deuterium and oxygen-18 in carbon-14-dated north Saharian groundwaters. Groundwater formation in the past. In: Agence Internationale De L'Energie Atomique. [Vienne.] Conférence. [1978, 19–23 juin. Neuherberg.], Organisation Des Nations Unies Pour L'Education, La Science Et La Culture. [Paris.] Conférence. [1978, 19–23 juin. Neuherberg.], eds. *Isotope Hydrology 1978: Proceedings of an International Symposium on Isotope Hydrology Jointly*. Proceedings series. I.A.E.A; 1979:341–352.
91. Benson BB, Krause D. Empirical laws for dilute aqueous solutions of nonpolar gases. J Chem Phys. 1976;64(2):689–709. https://doi.org/10.1063/1.432215.
92. Stute M, Deak J. Environmental Isotope Study (14-C, 13-C, 18-O, D, Noble Gases) on Deep Groundwater Circulation Systems in Hungary with Reference to Paleoclimate. Radiocarbon 1989; 31(03):902–918.

93. Geyh M. Environmental Isotopes in the Hydrological Cycle: Principles and Applications-Vol. IV. International Hydrological Programm; 2000:200.
94. McGuire KJ, McDonnell JJ. A review and evaluation of catchment transit time modeling. J Hydrol. 2006;330(3–4):543–563. https://doi.org/10.1016/j.jhydrol.2006.04.020.
95. Rinaldo A, Benettin P, Harman CJ, *et al.* Storage selection functions: a coherent framework for quantifying how catchments store and release water and solutes: on storage selection functions. Water Resour Res. 2015;51(6):4840–4847. https://doi.org/10.1002/2015WR017273.
96. Kirchner JW. Aggregation in environmental systems - Part 2: Catchment mean transit times and young water fractions under hydrologic nonstationarity. Hydrol Earth Syst Sci. 2016; 20(1):299–328. https://doi.org/10.5194/hess-20-299-2016.
97. Rodriguez NB, Klaus J. Catchment travel times from composite storage selection functions representing the superposition of streamflow generation processes. Water Resour Res. 2019; 55(11):9292–9314. https://doi.org/10.1029/2019WR024973.
98. Rank D, Wyhlidal S, Schott K, Weigand S, Oblin A. Temporal and spatial distribution of isotopes in river water in Central Europe: 50 years experience with the Austrian network of isotopes in rivers. Isotopes Environ Health Stud. 2018;54(2): 115–136. https://doi.org/10.1080/10256016.2017.1383906.
99. Risi C, Ogee J, Bony S, *et al.* The Water Isotopic Version of the Land-Surface Model ORCHIDEE: implementation, evaluation, sensitivity to hydrological parameters. Hydrol Curr Res. 2016; 07(04):1–24. https://doi.org/10.4172/2157-7587.1000258.
100. Halder J, Terzer S, Wassenaar LI, Araguás-Araguás LJ, Aggarwal PK. The Global Network of Isotopes in Rivers (GNIR): integration of water isotopes in watershed observation and riverine research. Hydrol Earth Syst Sci. 2015;19(8):3419–3431. https://doi.org/10.5194/hess-19-3419-2015.
101. Vitvar T, Aggarwal PK. Monitoring Isotopes in Rivers: Creation of the Global Network of Isotopes in Rivers (GNIR): Results of a Coordinated Research Project 2002-2006. International Atomic Energy Agency; 2012.
102. Dooge JCI. Looking for hydrologic laws. Water Resour Res. 1986; 22(9S):46S–58S. https://doi.org/10.1029/WR022i09Sp0046S.
103. Zehe E, Ehret U, Pfister L, *et al.* HESS Opinions: from response units to functional units: a thermodynamic reinterpretation of the HRU concept to link spatial organization and functioning of intermediate scale catchments. Hydrol Earth Syst Sci. 2014; 18(11):4635–4655. https://doi.org/10.5194/hess-18-4635-2014.
104. Versteegh EAA, Vonhof HB, Troelstra SR, Kaandorp RJG, Kroon D. Seasonally resolved growth of freshwater bivalves determined by oxygen and carbon isotope shell chemistry: FRESHWATER BIVALVE INTRA-ANNUAL GROWTH. Geochem Geophys Geosyst. 2010;11(8):1–16. https://doi.org/10.1029/2009GC002961.
105. Kelemen Z, Gillikin DP, Graniero LE, *et al.* Calibration of hydroclimate proxies in freshwater bivalve shells from Central and West Africa. Geochim Cosmochim Acta. 2017;208:41–62. https://doi.org/10.1016/j.gca.2017.03.025.
106. Pfister L, Grave C, Beisel JN, McDonnell JJ. A global assessment of freshwater mollusk shell oxygen isotope signatures and their relation to precipitation and stream water. Sci Rep. 2019; 9(1):4312. https://doi.org/10.1038/s41598-019-40369-0.
107. Schöne BR, Meret AE, Baier SM, *et al.* Freshwater pearl mussels from northern Sweden serve as long-term, high-resolution stream water isotope recorders. Hydrol Earth Syst Sci. 2020; 24(2):673–696. https://doi.org/10.5194/hess-24-673-2020.
108. Alterio E, Cocozza C, Chirici G, Rizzi A, Sitzia T. Preserving air pollution forest archives accessible through dendrochemistry. J Environ Manage. 2020;264:110462. https://doi.org/10.1016/j.jenvman.2020.110462.
109. Barroso PJ, Martín J, Santos JL, Aparicio I, Alonso E. Analytical method for the evaluation of the outdoor air contamination by emerging pollutants using tree leaves as bioindicators. Anal Bioanal Chem. 2018;410(2):417–428. https://doi.org/10.1007/s00216-017-0733-8.
110. Dzierzanowski K, Popek R, Gawrońska H, Saebø A, Gawroński SW. Deposition of particulate matter of different size fractions on leaf surfaces and in waxes of urban forest species. Int J Phytoremediation. 2011;13(10):1037–1046. https://doi.org/10.1080/15226514.2011.552929.
111. Krizkova S, Ryant P, Krystofova O, *et al.* Multi-instrumental analysis of tissues of sunflower plants treated with silver(I) ions - plants as bioindicators of environmental pollution. Sensors (Basel). 2008;8(1):445–463. https://doi.org/10.3390/s8010445.
112. Perone A, Cocozza C, Cherubini P, *et al.* Oak tree-rings record spatial-temporal pollution trends from different sources in Terni (Central Italy). Environmental Pollution. 2018;233: 278–289. https://doi.org/10.1016/j.envpol.2017.10.062.
113. Forbes PBC, van der Wat L, Kroukamp EM. Biomonitors. In: Comprehensive Analytical Chemistry. Vol 70. Elsevier; 2015: 53–108. https://doi.org/10.1016/bs.coac.2015.09.003.
114. Singh SK, Rao DN, Agrawal M, Pandey J, Naryan D. Air pollution tolerance index of plants. J Environ Manage. 1991;32(1):45–55. https://doi.org/10.1016/S0301-4797(05)80080-5.
115. Kvesitadze G, Khatisashvili G, Sadunishvili T, Kvesitadze E. Plants for remediation: uptake, translocation and transformation of organic pollutants. In: Öztürk M, Ashraf M, Aksoy A, Ahmad MSA, Hakeem KR, eds. Plants, Pollutants and Remediation. Springer Netherlands; 2015:241–308. https://doi.org/10.1007/978-94-017-7194-8_12.
116. Adeniyi AG, Ighalo JO. Biosorption of pollutants by plant leaves: an empirical review. J Environ Chem Eng. 2019;7(3): 103100. https://doi.org/10.1016/j.jece.2019.103100.
117. Sahu C, Basti S, Sahu SK. Air pollution tolerance index (APTI) and expected performance index (EPI) of trees in Sambalpur town of India. SN Appl Sci. 2020;2(8):1327. https://doi.org/10.1007/s42452-020-3120-6.
118. Briffa KR, Schweingruber FH, Jones PD, *et al.* Trees tell of past climates: but are they speaking less clearly today? Philos Trans R Soc Lond B Biol Sci. 1998;353(1365):65–73. https://doi.org/10.1098/rstb.1998.0191.
119. Jacoby GC, D'Arrigo RD. Tree rings, carbon dioxide, and climatic change. Proc Natl Acad Sci U S A. 1997;94(16):8350–8353. https://doi.org/10.1073/pnas.94.16.8350.
120. Pietrzykowski M. Tree species selection and reaction to mine soil reconstructed at reforested post-mine sites: Central and eastern European experiences. Ecol Eng. 2019;142:100012. https://doi.org/10.1016/j.ecoena.2019.100012.
121. Chodak M, Sroka K, Woś B, Pietrzykowski M. Chemical and microbial properties of post-mining and post-fire soils afforested with different tree species. Appl Soil Ecol. 2022;171:104321. https://doi.org/10.1016/j.apsoil.2021.104321.
122. Diederich P. Cartographie de la pollution atmosphérique au grand-duché de Luxembourg à l'aide de bioindicateurs. Société des naturalistes luxembourgeois association sans but lucratif SNL; 1979.
123. Hissler C, Stille P, Krein A, *et al.* Identifying the origins of local atmospheric deposition in the steel industry basin of Luxembourg using the chemical and isotopic composition of

the lichen Xanthoria parietina. Sci Total Environ. 2008; 405(1–3):338–344. https://doi.org/10.1016/j.scitotenv.2008.05.029.
124. Will-Wolf S, Jovan S, Amacher MC. Lichen elements as pollution indicators: evaluation of methods for large monitoring programmes. The Lichenologist. 2017;49(4):415–424. https://doi.org/10.1017/S0024282917000299.
125. Forbes P, Wat L, Strumpher J. Comparative Perspectives on Extraction Methods for Organic Metabolites and Pollutants from Lichens. In: Yusuf M, ed. Lichen-Derived Products. 1st ed. Wiley; 2020:27–73. https://doi.org/10.1002/9781119593249.ch2.
126. El-Khatib AA, Barakat NA, Youssef NA, Samir NA. Bioaccumulation of heavy metals air pollutants by urban trees. Int J Phytoremediation. 2020;22(2):210–222. https://doi.org/10.1080/15226514.2019.1652883.
127. Kirkwood KI, Fleming J, Nguyen H, Reif DM, Baker ES, Belcher SM. Utilizing pine needles to temporally and spatially profile per- and polyfluoroalkyl substances. Pharmacol Toxicol. 2021. https://doi.org/10.1101/2021.08.24.457570.
128. Lagarrigue M, Caprioli RM, Pineau C. Potential of MALDI imaging for the toxicological evaluation of environmental pollutants. J Proteomics. 2016;144:133–139. https://doi.org/10.1016/j.jprot.2016.05.008.
129. Annangudi SP, Myung K, Avila Adame C, Gilbert JR. MALDI-MS Imaging Analysis of Fungicide Residue Distributions on Wheat Leaf Surfaces. Environ Sci Technol. 2015;49(9):5579–5583. https://doi.org/10.1021/es506334y.
130. Ferretti M, Innes JL, Jalkanen R, *et al.* Air pollution and environmental chemistry – what role for tree-ring studies? Dendrochronologia 2002;20(1–2):159–174. https://doi.org/10.1078/1125-7865-00014.
131. Ruppert H, Wischow D. Tree rings—A questionable indicator of heavy metal pollution in air or soil. Geochim Cosmochim Acta. 2006;70(18):A546. https://doi.org/10.1016/j.gca.2006.06.1007.
132. Van de Maele J. *Minett Stories - Air Pollution Visualized. Minett Stories*. Published 2022. Accessed March 31, 2023. https://minett-stories.lu/en/story/air-pollution-visualized.
133. Esch2022. *Esch2022*. Published 2022. Accessed March 13, 2023. https://esch2022.lu/de/aboutesch/.
134. Chakrabarty D, Latour B. The Climate of History in a Planetary Age. The University of Chicago Press; 2021.
135. Digital History & Hermeneutics Doctoral Training Unit (DTU). C2DH | Luxembourg Centre for Contemporary and Digital History. Published 2017. Accessed December 9, 2022. https://www.c2dh.uni.lu/de/projects/digital-history-hermeneutics-doctoral-training-unit-dtu.
136. Nandipati S, Litvan I. Environmental Exposures and Parkinson's Disease. Int J Environ Res Public Health. 2016;13(9):881. https://doi.org/10.3390/ijerph13090881.

III. Simulating and Visualising Data in Environmental History: Airborne Dust Concentration from the Belval Plant in Luxembourg (1911-1997)

Authors:

Dagny Aurich, Aida Horaniet Ibanez, Jens van de Maele

Manuscript type:

Research paper

Contribution statement:

All authors contributed to the conceptualization and the review and editing of the manuscript. **DA** was responsible for writing (original draft preparation) of Chapters 3, 5, 6, 8 and the 'Conclusion' in Chapter 9. AHI wrote Chapter 2 and 7 and JM 'Abstract', 'Introduction' and Chapter 4. The Visualizations 1, 3, 4, 6 (a, b), 9, 11 (a) and 12 were reproduced with permission from the respective sources. The simulation video was mainly created by AHI – on the code basis as indicated – and checked and modified by **DA**. AHI also inserted the screenshots shown in Figures 8, 10 (a, b) and 11 (b, c). **DA** modified Figure 5 and Table 1, indicating the original source. Figure 2, Graphs 1 and 2 were created by **DA**.

Submission status:

Submitted to the Journal of Digital History on 28 April 2023, under review.
This manuscript will be also part of the PhD thesis by co-first author Aida Horaniet Ibanez.

Short summary/ Contribution to the field:

This paper uses historical data and interdisciplinary collaboration (projects *REMIX* and *LuxTIME*) to model the airborne dust concentration generated by the Belval steelwork in Luxembourg. The research combines scientific knowledge, historical sources, and visualizations to understand the extent and reasons behind dust pollution. It explores new ways of generating and analysing historical data and discusses the impact on the environment and human health. The limitations of the model, and challenges associated with visualizing and simulating historical data are discussed.

General Remarks

The Journal of Digital History is a *non-traditional* journal.

[Red text] in the attached manuscript indicates interactive content, visible in the online journal view.

For the interactive view of the manuscript, please visit the Journal of Digital History via: https://journalofdigitalhistory.org/en/notebook-viewer-form

Jupyter notebook file in GitHub repository: https://github.com/DagnyAurich/Dust-paper/blob/main/Dust_pollution_paper.ipynb .

Simulating and visualising data in environmental history: Airborne dust concentration from the Belval plant in Luxembourg (1911-1997)

Dagny Aurich[1]*, Aida Horaniet Ibanez[2]*, Jens van de Maele[2]

[1] Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, 6 avenue du Swing, 4367, Belvaux, Luxembourg.

[2] Luxembourg Centre for Contemporary and Digital History (C²DH), University of Luxembourg, 11, Porte des Sciences, 4366 Esch-sur-Alzette, Luxembourg.

*Corresponding authors: dagny.aurich@uni.lu, aida.horanietibanez@uni.lu

To be published at: https://journalofdigitalhistory.org/en

Keywords

Dust Pollution, Environment, Data Visualization, Simulation, Steel Industry, History of Luxembourg

Abstract

Luxembourg's industrial history can be reconstructed using primary sources containing quantitative and qualitative industrial and environmental data, which can then be analysed and visualised digitally. This research combines current scientific knowledge and historical data to model the airborne dust concentration generated by the Belval steelworks in Esch-sur-Alzette between 1911 and 1997. The calculations are based on a model of atmospheric dispersion, using parameters such as production volumes, the presence of filter systems and meteorological data. Visualisation of the simulated data offers insights into the extent and variability of dust concentrations over time, along with the underlying reasons (*e.g.* wars or technical innovations). An interdisciplinary approach allowed the integration of chemical and health perspectives, which were then embedded into a historical context. Simulations, a form of data generation widely used in engineering, open new opportunities for the re*creation* of historical data, which can be compared with scientific research. These historical data can also be re-analysed and visualised on the basis of new and/or current perspectives.

This research is a collaboration between the REMIX project (an interdisciplinary project focusing on the history of the Minett region in connection with Esch-sur-Alzette – European Capital of Culture 2022) and the interdisciplinary LuxTIME project (which attempts to build and visualise historical datasets with the aim of studying the impact of environmental changes on Belval's population). The aim is to explore new ways of generating, validating, analysing and visualising past data through interdisciplinary collaboration, rather than to validate a model for the study of historical dust pollution. This work discusses (*a*) the model, (*b*) the parameters used, (*c*) the results of the dust simulation, (*d*) the impact on health, (*e*) the historical sources, and (*f*) the limitations of the model and the challenges involved in visualising and simulating historical data.

Introduction

Air pollution has a history. In today's society, this historicity is most commonly perceived in the context of global warming, where the 'weight' of billions of tonnes of carbon dioxide released in the past has become a factor of major importance for our global ecosystem. The mechanism behind this has been elegantly described by Malm, who observed that our climate has essentially become a 'product of past emissions', while the 'emissions produced by cars [today will generate] their greatest impact on generations not yet born'. For Malm, present-day emissions are effectively 'invisible missiles aimed at the future' (Malm 2016). The adjective 'invisible' is rather important here. After all, the peculiar interrelationship between past, present and future is not the only paradoxical characteristic of air pollution history (on this interrelationship, see also (Uekötter 2020)). Another paradox is the fact that the ongoing global warming is caused by a gas that, when encountered under everyday circumstances, is both invisible and odourless. This lack of perceptibility does not imply, however, that pollution problems went completely unnoticed by previous generations. Prior to the scientific and societal problematisation of the greenhouse effect, other manifestations of air pollution – such as dust, dark smoke, soot particles and stench – were particularly palpable in cities and industrial regions, and as such were a regular source of indignation (for two classic studies on this indignation, see (Brüggemeier and Rommelspacher 1992; Mosley 2001)).

The southwest of the Grand Duchy of Luxembourg, an important iron and steelmaking region from the 1870s onward, was one of those places where people were constantly faced with air pollution that was as omnipresent as the industrial plants themselves.[1] In the Minett, as Luxembourg's metallurgical region is called, the soot and smoke of the steelworks were part of a 'sensescape' that could be seen and smelled on a near-permanent basis (on the notion of sensescape, see (Pritchard and Zimring 2020)). To borrow the words of an elderly inhabitant of Dudelange, who recently reminisced about his life in Luxembourg's industrial region: day in, day out, 'the dust and dirt fell from the skies' (Back-Hoffmann 2021). In order to objectify the severity of this problem, several scientific studies were carried out in the Minett towns of Differdange and Esch-sur-Alzette between the late 1930s and the early 1970s, with the specific aim of measuring the typical quantity of airborne dust particles at or near industrial sites. Yet these measurements – which we discuss in section 8.2 – only lasted for a limited period (*e.g.* three months), and thus only provide a snapshot in time.

In this contribution, we attempt to create a *long-term* picture of past dust pollution levels by using the Belval iron and steel complex in Esch-sur-Alzette – the informal capital of the Minett (hereafter: Esch) – as a test case (Figure 1). For this purpose, historians, metallurgy experts, chemists and data visualisation specialists have collaborated to analyse a major metallurgical pollutant: dust. We present a simulation and visualisation of the diachronic evolution of local airborne dust concentrations in the vicinity of the Belval plant, using various parameters as input for our model. These include the production numbers of the plant throughout the 20th century, the production process (*e.g.* LD-AC or Thomas for steelmaking), the typical usage of filtering systems, and the prevailing wind directions over time. Based on this data set, we offer a plausible approximation of how the dust emissions of the Belval iron and steel complex might have affected the local air quality between 1911 (the first year of operation) and 1997 (when the last blast furnace was shut down).

[1] For the economic and social history of Luxembourg's metallurgical industry, see the six-volume series *Terres Rouges: Histoire de la sidérurgie luxembourgeoise,* published from 2009 to 2018 by the National Archives of Luxembourg.

Figure 1: Smoking chimneys of the Belval steel plant. Slide from around 1960. Source: Archives de la Ville d'Esch-sur-Alzette.

An important precedent for modelling in environmental history can be found in the work of Tarr, who has investigated emissions of man-made components such as pesticides, herbicides and heavy metals in the estuary of the Hudson and Raritan rivers between 1700 and 1980. To achieve this aim, Tarr worked together with physicists and environmental engineers ((Tarr 1996); see also (Tarr 2021)). Likewise, in their 2021 study on climate change and its effects on people, Pfister and Wanner attempted to reconcile the scientific cultures of the natural sciences and the humanities: 'Scientists explain how natural systems work, while (environmental) historians tell stories of people who grapple with the effects of weather and climate' (Pfister and Wanner 2021). As such, the research by Pfister and Wanner is in line with the approach outlined by Guldi and Armitage, who have discussed the role of history in the process of understanding 'the multiple pasts which gave rise to our conflicted present'. Regarding climate change, Guldi and Armitage have argued that 'history, with its rich, material understanding of human experience and institutions and its apprehension of multiple causality, is re-entering the arena of long-term discussions of time where evolutionary biologists, archaeologists, climate scientists, and economists have long been the only protagonists' (Guldi and Armitage 2014).

Similar to the way in which environmental history and the history of technology have become more entangled in the last decade (resulting in the development of a crossover field known as 'envirotech', *e.g.* (Pritchard and Zimring 2020)), interdisciplinary endeavours between the human and natural sciences are indeed essential if one wants to adequately evaluate environmental changes that have occurred in the past. This has been made explicit by Massard-Guilbaud:

> *How can we write a history of climate without contacting climatologists? A history of pollution without touching on chemistry? A history of sanitation systems or energy resources without taking an interest in technology? Let us note, moreover, that it is not only historians who are discovering this need for interdisciplinarity; specialists in other disciplines […] are also approaching historians [...] on environmental issues because [...] they feel the need to do so.* (Massard-Guilbaud 2007)[2]

The research questions that motivate our own interdisciplinary research are as follows. Throughout the 20th century, how were people living near the Belval iron and steel complex affected by dust pollution? Can this pollution from the past be quantified and displayed on a map? And how did it evolve over time? Using modelling and simulation with historical data as input, we generate new data allowing us to analyse such questions, thus uniting the points of view of exact sciences (*e.g.* the health impact of various concentration levels) and history (*e.g.* the industrial development of the region and the societal narratives about pollution). We hypothesise that the pollution generated by the Belval plant affected many (if not all) of the inhabited quarters of both Esch and the surrounding municipalities (including some on French territory, which would make Belval a case of cross-border air pollution).

It must be emphasised that our contribution offers only an *estimation* of pollution values, due to the limitations inherent in any simulation. In the current study, such limitations include the lack of adequate corporate sources on the presence and efficiency of filtering systems, forcing a reliance on assumptions (based on data in non-Luxembourgish secondary literature on ideal typical metallurgical filtering systems). Another important limitation of our model is the omission of other pollution sources, such as iron and steel complexes other than Belval (including two more plants in Esch), non-metallurgical industrial complexes, traffic and domestic heating. As such, we agree with a warning recently posited by Joy Parr concerning interdisciplinary work. While Parr endorsed collaborations between 'researchers in environmental history and the history of technology' and scientists from fields like 'metallurgy [and] bio-chemistry […]', she pointed at possible incompatibilities between historical science and the natural sciences, due to the latter's tendency to 'simplify' for the purpose of experimentation, as well as to 'derogate knowledge and reasoning that is not readily represented in symbols and signs' (Parr 2010). In our interdisciplinary contribution, we take a rather modest view on

[2] All translations throughout have been done by the authors.

this matter, by explicitly pointing at the potential of our research as well as the lacunae and possibilities for future improvement.

The structure of our article can be summarised as follows. In section 2, we discuss the terms 'modelling' and 'simulation', their history and application across disciplines, and their largely untapped potential in the humanities. Section 3 explains the general impact of dust pollution on health, in order to provide a basis for the subsequent analysis of our results. Section 4 establishes the historical context of the metallurgical industry in Luxembourg, the Minett region and Esch, while also briefly exploring the societal narratives about pollution. In section 5, we introduce steelmaking production techniques as well as their evolution and impact on dust pollution. In section 6, we explain the actual simulation process by focusing on the data generated by the dispersion model. Here, we discuss in detail the data basis and the model specifications, to explore the hypothesis presented (regarding the impact of pollution on the inhabitants of Esch and the surrounding area). In section 7, we explore the data visualisation process. Section 8 contains an analysis of the results including the validation of the model used, a basic Gaussian dispersion model that has already been proven through its use in similar scenarios (in various contexts) (Leelőssy et al. 2014); this section is focussing on our way of applying the model. Further, a comparison of the model results against available measurements is presented in section 8.2, and the limitations and potential ways to improve the model in future research in 8.3 and 8.4. For future research on historical dust pollution, the model could be used with alternative data sources for new simulations, or new models could be tested adding new variables. Our conclusions are offered in the final section (9).

Modelling and simulation

Since modelling and simulation are key elements in our research, a brief introduction of terms and usage across disciplines is appropriate. Banks defined simulation as 'the imitation of the operation of a real-world process or system over time' (such as the generation and dispersion of dust pollution of a steel plant) that can model both existing and conceptual systems (Banks 1999). As such, simulation is used for a variety of purposes including prediction, performance, training, entertainment, proof and discovery. While *induction* can be used to find patterns in data and *deduction* to find consequences of assumptions, *simulation* generates data that can be analysed inductively (Carson II 2005).

Initiated during the Second World War 'to address problems too complex for theory and too remote from laboratory materials for experiment' (Galison 1996), simulation spread through the social sciences by the end of the 1960s, where it was (and still is) often used for modelling artificial populations and investigating human behaviour (Lebherz et al. 2018). However, 'it has remained almost unknown to the humanities' (McCarty 2016), notwithstanding a few exceptions. Apart from the environmental historical research mentioned in the introduction, Lebherz et al. have presented a case of text mining by scientific workflows and computer simulation, with the aim of investigating potential influences on the author's literary productivity. In this context, they have identified four prerequisites: a solid data basis, the specification and construction of a model, the definition of hypotheses to answer research questions, and the reuse of proven models in similar scenarios (Lebherz et al. 2018). Simulation is also frequently used in certain fields of archaeological research, such as evolutionary archaeology and the study of human evolution (Lake 2014).

Champion has discussed the distinction between model ('a physical or digital representation of a product or process') and simulation ('the re-configurative use of a model to reveal new and potential aspects') (Champion 2016). In the context of the humanities, McCarty has referred to modelling as 'the analogical bridge between computing and the interpretative disciplines' that 'keeps the digital construct separate, informing humanistic research both by what it discovers and especially by what it cannot'. The defining moment of simulation occurs 'when it becomes the only way to know something or to form a coherent picture from fragmentary knowledge' (McCarty 2016).

Dust pollution from a chemical and health perspective

Pollution caused by dust (atmospheric particulate matter) is a global problem, both historically and in the present (Butte and Heinzow 2002; Han et al. 2021). Dust coming from various anthropogenic sources such as industry, households and traffic is a major source of chemical contamination of entire regions and has a negative impact on health. Dust particles can also come from natural sources, like pollen, microorganisms and the natural erosion of soils. The size, origin and chemical composition of dust particles play a major role when looking at health effects. Airborne dust comes in all shapes and sizes, which are either visible or invisible to the naked eye (aerodynamic diameter 1 to 400 µm) (Kumar and Kumar 2018).

Large dust particles settle quickly (> 100 µm) or are trapped in the nose, mouth and larynx region when inhaled (*inhalable* fraction ≤ 100 µm), as shown in Figure 2 (Wippich et al. 2020). Smaller particles stay for longer in the air, while very small particles can penetrate the respiratory system. The *respirable* dust fraction that can enter the alveolar region is defined to be below an aerodynamic diameter of 10 µm, independent of the particle's length (WHO n.d.). Small particles that remain in the gas-exchange region of the lungs can cause allergic reactions, cancer and other serious diseases or disorders (Wippich et al. 2020; WHO n.d.; Bala and Tabaku 2010).

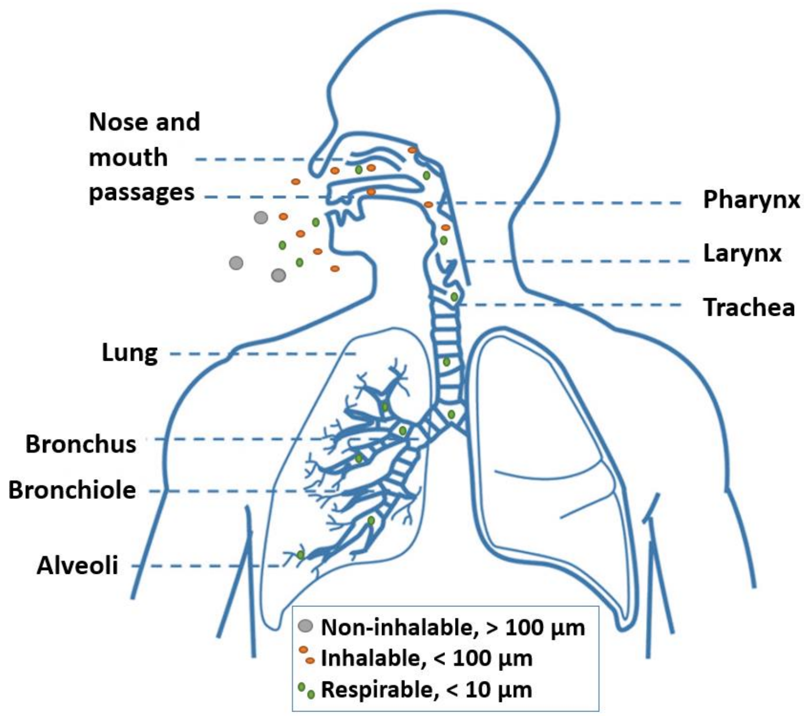


Figure 2: Respiratory system with different dust fractions and their aerodynamic diameter

Dust sources can be either point or area sources, from where particles (having long lifetimes) can spread over many kilometres (Leelőssy et al. 2014). Their chemical composition is essentially linked to their origin: during the extraction and processing of minerals, for instance, dust containing silica is often released in the air, which can result in permanent health damage (*e.g.* diseases such as silicosis) (Mlika, Adigun, and Bhutta 2022). Likewise, metallic dusts (*e.g.* lead and cadmium) from smelters cause harm to fauna and flora, as do chemical dusts created in agriculture (*e.g.* pesticides), vegetable dusts (*e.g.* wood, cotton), moulds and spores (WHO n.d.). Other air pollutants are created by the reaction of sunlight with atmospheric compounds, leading to photochemical smog (sunlight reacting with volatile organic compounds and nitrogen oxides).

The iron and steel industry is known to be a major contributor to air pollution. In the past, burning coal for steel production produced vast amounts of black coal dust, resulting in fine particles covering entire cities and regions. Such particles often contain silica and various metals; those with a diameter below 2.5 µm can easily enter the alveoli (*i.e.* the gas-exchange regions of the lungs) (Figure 2) and cause serious harm (Su, Ding, and Zhuang 2020). Early types of filtering devices based on cyclone techniques only removed larger particles, thus leaving smaller fractions uncaptured ('The Steel City and a Brief History of Dust Collection' 2019). In the Western world, with the rise of environmental regulation and technical advances from the mid-20th century onward, it became possible to significantly decreased dust emissions as long as steel companies were prepared to make the necessary investments. The decrease was mainly enabled by improved filtering techniques, including shaker bag collectors, reverse air or pulse jet baghouses, and (later) cartridge collectors ('The Steel City and a Brief History of Dust Collection' 2019). However, even today many iron and steel plants continue to rely on coal fuel (the 'dirtiest' but also cheapest option available), with little concern for the environmental effects and the availability of cleaner techniques ('Do We Really Need Coal to Make Steel?' 2020). As well as producing greenhouse gases like nitrous oxide (N_2O), steelmaking generates other airborne contaminants (Koponen et al. 1980). Metals like cobalt, lead and chromium are released into the atmosphere (Nurul, Shamsul, and Noor Hassim 2016) when not filtered appropriately, and so are concentrations of polycyclic aromatic hydrocarbons (PAHs), iron oxides and sulphur dioxide (SO_2) (Bala and Tabaku 2010). Historically as well as in the present, respirable particle emissions from the iron and steel industry pose risks to human health (Valenti et al. 2016).

Esch-sur-Alzette and the Belval metallurgical complex: a brief historical overview

Measured per capita, Luxembourg was the world's foremost iron and steel producer throughout the 20th century (for global production trends, see (Hemmer 1953)). While all of the Minett's towns were severely affected by air pollution, Esch had the specific characteristic of being surrounded by iron and steel complexes. The local historians Assa and Pagliarini have rightfully noted that 'Esch and its factories formed one unity; their evolution occurred along parallel lines' (Assa and Pagliarini 1998). From the moment the first blast furnace in Esch started operating, the urban infrastructure of this 'mushroom town' (Knebeler and Scuto 2010) indeed developed almost exclusively to serve the metallurgical industry. An engraving from the 1920s shows that iron and steel complexes could be found to the southwest of Esch (the *Terres Rouges* plant, established in 1872), to the east (the Esch-Schifflange plant, 1871) and to the west (the Belval plant, 1911). Other polluting factories, such as a cement factory, a brewery and a coal-fired power plant (which supplied electricity to the steelworks) were also present in the town throughout much of industrial era (for a general overview, see (Scuto 1993)). A scale model of the Minett, made for the 1937 World's Fair in Paris, clearly illustrates how Esch was 'embraced' by factories and, consequently, by industrial smoke and dust (on Luxembourg's participation at the 1937 World's Fair, see (Millim 2014)).

Figure 3: Engraving by G. Peltier showing the three iron and steel complexes surrounding Esch: Esch-Schifflange (left), Terres Rouges (centre) and Belval (below right). Source: *(Knebeler and Scuto 2010)*.

Figure 4: Scale model of Esch exhibited at the 1937 World's Fair in Paris, showing the three iron and steel complexes surrounding Esch: Esch-Schifflange (bottom), Terres Rouges (top left) and Belval (top right). Source: Archives Nationales de Luxembourg, Fonds ARBED, File AES-U1-54.

Contrary to present-day popular belief (*e.g.* (Moia 1998; Logelin-Simon 2006; Back-Hoffmann 2021)), the air pollution caused by the metallurgical complexes of the Minett was not uncontested by the local population. From the early 20th century onwards, inhabitants voiced their criticism about dusty factory smoke through articles in newspapers and magazines, readers' letters and interventions in parliament and municipal council meetings. This paragraph will shortly explore a few criticisms concerning the Belval complex from the period when the plant was established. A more in-depth analysis of complaints from inhabitants about air pollution in the entire Minett region is offered in a separate article, which focuses on the period of the *Trente Glorieuses* (c. 1945-1965) and conceptualises the role of newspapers, politicians and scientists as 'mediators' in the public debate on air pollution (Van de Maele forthcoming).

From around 1900, about three decades after the dawn of industrialisation in Luxembourg, the country's press began to report on the so-called 'plagues' of dust and smoke (in German: *Staubplage, Rauchplage*). These 'plagues' were reported to soil clothes and buildings, irritate lungs and eyes, and affect animal and plant life. As early as 1899, for example, the newspaper *Luxemburger Wort* published a remarkably lucid analysis of air pollution in industrial regions (for the history of this newspaper, see (*150 Jahre Und Kein Bisschen Alt: 150 Joër Wort, 1848-1998* 1998)). Despite its conservative, pro-business profile, the newspaper offered a remarkably modern-sounding, holistic approach towards both human and non-human nature, thus offering a rather bleak picture of the industrial era:

> *No more than fifteen years ago, everyone took for granted that a chimney should smoke, and every stranger who came to a factory town looked in admiration at the myriad of smoking chimneys […]. Even those who were directly harmed by the smoke considered this to be so inevitable that they hardly ever complained about it, and when they did, they were told – with a shrug – that nothing could be changed. And so for decades, the vegetation in industrial areas withered, entire forests had to be cut down prematurely […], and soot particles and poisonous gases inhaled by human lungs slowly but surely planted the germ of many a deadly disease.*[3]

Already around this time, awareness about the detrimental side effects of metallurgical dust also existed on a governmental level. This becomes evident in a 1910 investigation undertaken at the instigation of the Luxembourg Prime Minister in response to the construction request for the Belval plant submitted by the German corporation Gelsenkirchener Bergwerks AG. At that time, the corporation had recently acquired a plot of forested land just to the west of Esch. Conceived as a vertically integrated plant, the Belval complex was to include two blast furnaces (later supplemented with a third), steel converters, a sintering plant and rolling mills. In a letter to three engineers (who were most likely state-employed), the Prime Minister expressed reservations about the addition of yet another metallurgical factory in the Minett: '[…] [It] is necessary to investigate all [technical] methods to curtail noise […] and dust […].'[4] Consistent with the legal framework of the time (Parmentier 2008), the Prime Minister's reservations were mainly driven by a desire to minimise damage to the private property of adjacent landowners; a true 'environmental' awareness (as can be seen embryonically in the very *avant-garde* 1899 newspaper article mentioned above) was not yet part of the politicians' intellectual *habitus*. In response, the experts offered a highly ambiguous evaluation of the prospective plant. Although they acknowledged the existence of a pollution burden, this burden was considered to be a 'normal' phenomenon in industrial regions:

> *The inconveniences these kinds of establishment can have for certain [neighbouring] owners in terms of their enjoyment of the tranquillity and pure air of the countryside are not to be denied. These inconveniences stem from the vapours and dust that detract from the purity of the air, as well as the noise caused by the machines and the movement of workers […]. Such inconveniences are common in*

---

[3] 'Die Rauchplage und ihre Beseitigung', *Luxemburger Wort für Wahrheit und Recht*, 13 May 1899, p. 1.
[4] Archives Nationales de Luxembourg, *Ministère de la Justice* Collection, File J-090-00804, Letter from the Prime Minister, March 1910.

> *industrial centres; everyone experiences them in Esch, Differdange, Rodange, Dudelange, etc. Yet they cannot be invoked as a reason to refuse authorisation.*[5]

This discourse was typical of the professional ethos of late-19th and early-20th-century engineers, who typically saw themselves as expert organisers, both within and beyond their technical sphere of competence (on the role of engineers in Luxembourg society, see (Glesener and Wilhelm 2009)). As such, the engineers also gave advice about the *moral* economy in which the steelmaking was to take place: in their view, the imperative of economic development outweighed the time-honoured right of neighbouring inhabitants to 'pure' air. The engineers' opinion proved to be decisive: shortly after the governmental enquiry, a green light was given to the Gelsenkirchener Bergwerks AG, and in 1911 the two blast furnaces of the iron and steel complex (which would later become known as Belval) were fired up. Soon enough, newspapers would again voice concerns about the increasing dust problem in Esch. In April 1914, just three months before the start of the First World War, another conservative newspaper, the *Obermosel-Zeitung*, for instance voiced concerns about the worsening 'smoke plague':

> *If any town far and wide suffers greatly from the smoke plague, it is Esch. No fewer than 30 to 40 blast furnace chimneys in the town's immediate vicinity spew their contents on the Minett's capital day and night. If a resourceful person were to come up with an effective invention against the smoke plague, he should certainly file a patent in Esch. It is only a small consolation for the people of Esch to hear that the smoke plague is also known to be a great evil in other places, and that in many cases desperate efforts are being made to combat it […].*[6]

Still in 1914, the Esch-based architect Paul Flesch declared in a municipal council meeting that Esch should keep a corridor without industry in the north, to allow the town's 30 000 inhabitants 'to breathe some clean air for at least a few days a year'. With his call for the incorporation of public hygiene principles in urban planning policies, Flesch sought to pair economic development with a more decent quality of life for the town's inhabitants (Scuto 2005). As it turned out, the north of Esch would effectively remain industry-free – even though the 'right to produce' (and, consequently, the right to pollute) of the existing iron and steel complexes would never be called into question by local and national authorities. Using a term from German historiography (*e.g.* (Geissler 2016)), it can be said that these authorities consistently affirmed the *Ortsüblichkeit* ('geographical appropriateness') of industry. This principle was further underpinned by the powerful political and economic position of the ARBED company (*Aciéries réunies de Burbach-Eich-Dudelange*), which took over ownership of the Belval

[5] Archives Nationales de Luxembourg, *Ministère de la Justice* Collection, File J-090-00804, Letter from Noppeney, Haardt and Jaans, 6 May 1910.
[6] 'Chronik aus dem Erzbassin', *Obermosel-Zeitung,* 10 April 1914, p. 2.

complex in 1919 (Knebeler and Scuto 2010). In this constellation, the plant would thrive: between 1911 and 1997, a total of almost 80 million tonnes of iron was produced at the Belval site – or about one third of Luxembourg's total production during that period. Belval's corresponding steel production amounted to approximately 75 million tonnes – a number that is continuing to grow to this day, as the steelworks are still in operation even after the closure of the last blast furnace in 1997 (for production numbers, see (Knebeler and Scuto 2010)).

Steelmaking techniques in Luxembourg

As steelmaking techniques developed, the degree of dust pollution resulting from the steel industry in the Minett varied over time. Initially, at the onset of industrialisation in the 1870s, Luxembourg's plants used the Bessemer process, the first industrial method to produce steel from molten pig iron, which was invented in the 1850s and complemented in 1864 by the Siemens-Martin process (Knebeler 2011; Metz 1972). Impurities were removed by oxidation, *i.e.* by blowing air through the molten iron. However, as the iron ore found in the Minett is high in phosphorous, only low-quality steel could be manufactured (Knebeler 2011). This changed with the modification of the Bessemer process by Sidney Gilchrist Thomas, who used dolomite, which is basic, as the converter lining (instead of packed sand, which is acidic), thereby removing the phosphorous from the steel into the slag (Knebeler 2011; Metz 1972). In Luxembourg, the first Thomas steel was produced in 1886; the process was complemented by electric arc furnace (EAF) steelmaking after 1900. From the 1950s onwards, pure oxygen processes like the LD-AC process (*Linz-Donawitz – ARBED – Centre national de recherches métallurgiques de Liège*) were predominantly used to produce higher quality steel with fewer impurities compared to Thomas steel (Metz 1972; Knebeler 2011).[7]

The LD-AC method was reported to generate less dust ('Compte-rendu Chambre des Députés, volume 1 - 1975-1976' 1976). In a 1972 article, Metz signalled that the steelworks in the Minett used two principal techniques to perform dedusting: electro-static purification with negatively charged dust particles collected on a positive collection plate, and wet processes which retained dust with water (Metz 1972). The efficiency of these processes seems to have been limited, however. According to

[7] Other techniques replacing Thomas steel were the Lance Bubbling Equilibrium (LBE), ARBED Ladle Treatment (ALT) or LD Kaldo, and Rotovert. (Metz, Knebeler)

Hoffmann (1974), who based her findings on an undisclosed source, dust pollution levels in Esch regularly exceeded 1.8 g/m^3/day, which was well beyond the West German environmental limits of the time (0.42 g/m^2/day for urban zones and 0.85 g/m^2/day for industrial zones) (Hoffmann 1974). Yet even though she was a critic of the steel industry, Hoffmann stressed that 90% of the air pollution in Luxembourg resulted from households (oil heating) and traffic, and not from the iron and steel works – a position in line with the steel industry itself (Metz 1972). A number of measurements undertaken between the mid-1950s and the early 1970s, which we discuss below (section 8.2), indicate that this estimation was most likely too flattering for the metallurgical industry, and that the airborne iron oxide load caused by steel processes was probably much higher.

By the time Metz and Hoffmann published their articles, governmental awareness about the environmental impact of the iron and steel industry was on the rise. This rise was spurred by multiple factors, including decades of public complaints, the international emergence of the 'modern' environmental movement in the second half of the 1960s, and the ongoing scientific research on metallurgical pollution conducted in the transnational framework of the European Coal and Steel Community (Van de Maele forthcoming). As a result, in the mid-1970s, the demand for the 'polluter pays' principle became stronger, eventually requiring Luxembourg's industry to minimise its environmental impact ((Hoffmann 1974); on the worldwide rise of environmental policy-making during the 1970s, see for instance (Jarrige and Le Roux 2017; Buelens 2022; Uekötter 2020)). The iron oxide content in the Minett's atmosphere was addressed in a 1976 parliamentary debate, when an MP stated that the 'red clouds' of iron oxide were predominantly a result of Thomas steel production, while the LD-AC process (which was usually operated with pre-installed filters) was reported to generate less dust ('Compte-rendu Chambre des Députés, volume 1 - 1975-1976' 1976). In Luxembourg, the last Thomas steel was produced in 1977 (Metz 1972; Hoffmann 1974).

The simulation process

The calculation of our dust simulation model – estimating dust concentrations from 1911 to 1997 – was performed at our request by Inspyro, a technical consultancy company assisting metallurgical enterprises.[8] The various sources (and their limitations) used as input data for the Gaussian model are explained below.

There are several ways to create the atmospheric dispersion models needed to understand and predict air pollution. Gaussian, Lagrangian, Eulerian and computational fluid dynamics models all offer possibilities,[9] differing in their mathematical complexity and field of application. All models require the input of parameters like meteorological data, emission parameters and terrain information. To develop such a model, an interdisciplinary approach including disciplines like meteorology, chemistry and physics is often the best option. The output data is usually plotted on maps indicating areas of higher/lower air pollution concentrations and therefore higher/lower health risk.

In our case, a basic Gaussian dispersion model, as shown by Leelőssy et al. – known for its fast response time in the application of long-term average loads for distances between 1 and 100 km – was used (Leelőssy et al. 2014). The equation used (1), based on the Gaussian plume model for atmospheric dispersion modelling, is shown below.

$$\bar{c}\,(x,y,z) = \frac{Q}{2\pi\sigma_y\sigma_z\bar{u}}\exp\left(\frac{-y^2}{2\sigma_y^2}\right)\left(\exp\left(\frac{-(z-h)^2}{2\sigma_z^2}\right)+\exp\left(\frac{-(z+h)^2}{2\sigma_z^2}\right)\right) \qquad (1)$$

A detailed formula derivation can be found in Stockie's 'Mathematics of atmospheric dispersion modelling'. Stockie defines dispersion as the 'combination of turbulent diffusion and advection by the

[8] The metallurgical experts (civil engineers) who performed the calculation were Sander Arnout, Yannick Cryns and Cem Tekin.

[9] Gaussian models assume a normal statistical distribution (parabolic behaviour near the origin of the coordinates, "bell curve") and are typically used for buoyant air pollution plumes. This approach historically dominated dispersion models with a number of simplifications to be taken in the advection-diffusion equations. Lagrangian models use trajectories – calculated using ordinary differential equations – of air pollutants determined by *e.g.* wind field, buoyancy and turbulence (Leelőssy et al. 2014). Pollutant particles are followed in time and space along their trajectories. Eulerian models use a fixed coordinate frame providing a spatiotemporal evolution of pollutant concentration at each time step and point in the grid. For environmental and health protection measures, computational fluid dynamics models are often used in complex geometry where a fine grid resolution is required to calculate turbulence effects (Lagrangian and Eulerian: coarse grid). This model does account for flow velocities and turbulence in a complex 3D terrain, unlike the Gaussian approach (Tripathi et al. 2018). For more information and a detailed comparison see Leelőssy (Leelőssy et al. 2014).

wind' (Stockie 2011). Atmospheric contaminant concentrations can therefore be explained using the advection-diffusion equation. Using the simplified model of a Gaussian plume, it is assumed that air contaminants come unidirectionally (given the wind direction) from one point source – here one chimney of the Belval steel plant –, as outlined in Figure 5.

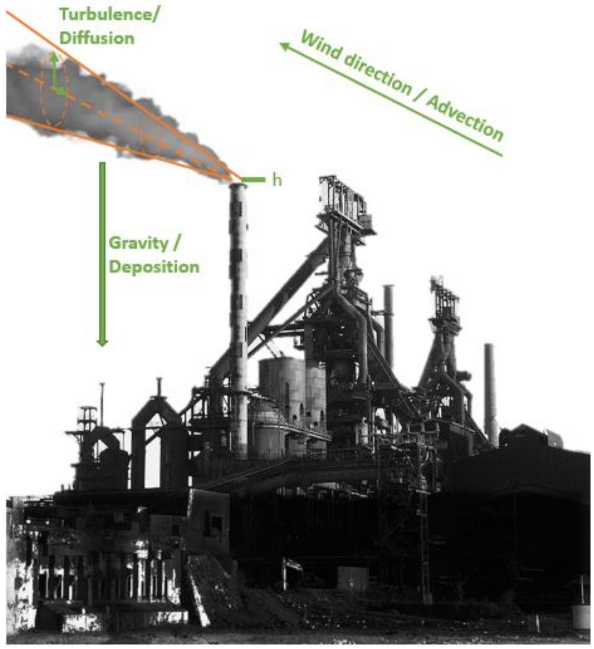


Figure 5: Schematic demonstration of the three main contributors to dust transportation: advection, diffusion and deposition at the Belval steel plant (photo taken in 2003) indicating the release height h. Photo modified by the authors. Source of the photo: www.agora.lu/

In formula (1) $\bar{c}$ is defined as the time-averaged concentration at a given position, Q represents the constant emission rate [kg/s] and (x,y,z) stand for the wind directions (downwind, crosswind, vertical direction) with the standard deviations σ_y and σ_z . $\bar{u}$ is the time-averaged wind speed at contaminant release height h (see Figure 5). Moreover, a homogeneous steady-state flow at the point source (0;0;h) is assumed (Leelőssy et al. 2014).

For the Belval case only one chimney of the steel plant was assumed to be the point source. To calculate the amount of dust produced per month, the various steelmaking processes were categorised; in relation to ironmaking, sintering and crushing of iron ore were the main processes contributing to dust production. Consequently, the average amounts of dust produced by the three steelmaking processes were used to calculate the amount of dust generated during the production of raw iron. For electric steel production (using an electric arc furnace (EAF)), the average amount of dust was multiplied by the amount of steel produced (see Figure 6). The same approach was applied to Thomas steel and LD-AC steel. Annual steel production data at the Belval plant was retrieved from a table reproduced in a study by Knebeler and Scuto (Knebeler and Scuto 2010). This table includes the production data of steel in tonnes per steelmaking technique (Thomas, LD-AC, EAF, total steel) per year, shown in Figure 6. As the simulation output shows, the dust concentrations are calculated as averages per month. (The annual production values turn out to be a limitation, since monthly data would generate a more exact output of the model.) The hypothetical average dust values generated per process (blast furnace, sintering furnace, EAF and LD-AC) were taken from Schueneman, as summarised in Table 1 (Schueneman 1963). Schueneman offered typical values for steel plants in the US in 1963; as such, our hypothetical average dust values offer only an estimation of dust-producing steel processes in Luxembourg from 1911 to 1997.

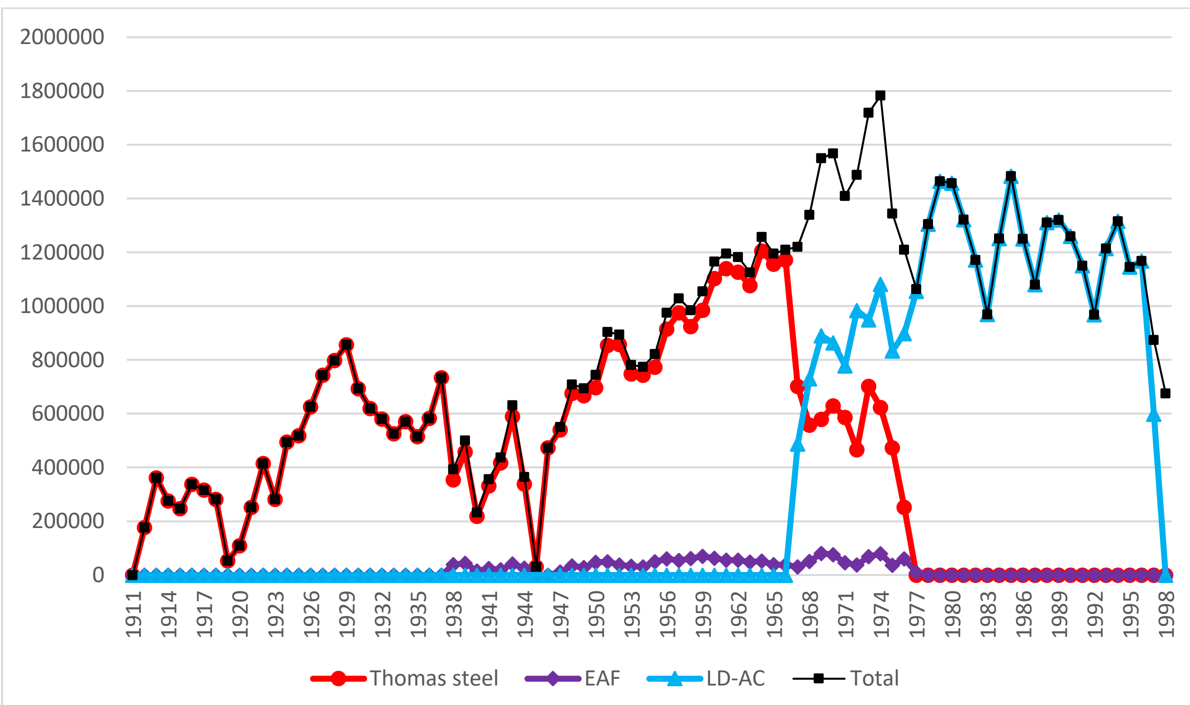


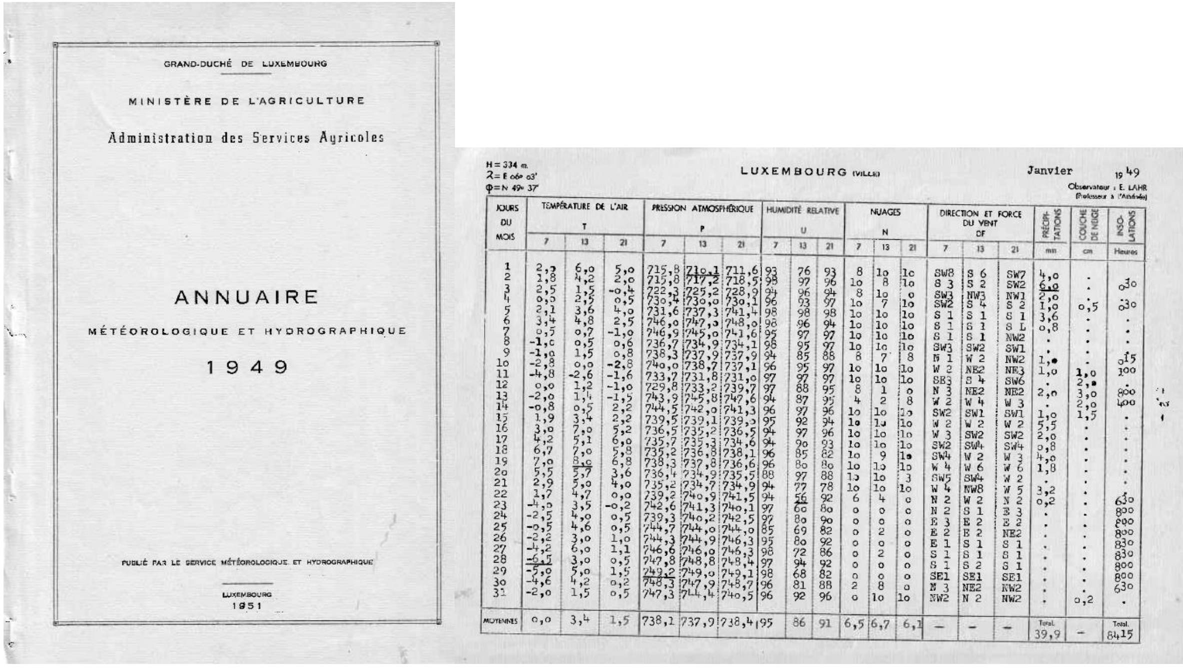
Figure 6: Steel production at the Belval plant from 1911 to 1998, in t per technique

The hypothetical installation date of filters (with varying efficiencies) was likewise taken from Schueneman (Schueneman 1963) (Table 1). Based on this hypothesis, steel production techniques in Belval after 1963 were assumed to have state-of-the-art filters installed. Again, the installation dates are typical estimates for US steel plants until 1963, which will not correspond 1:1 with our case study on Luxembourg. We used these assumptions because of the lack of literature on the historical presence and efficiency of filter systems in the European iron and steel industry. Table 1 summarises the filters installed per technique (based on Schueneman); the amount of dust caught by a filter was subtracted from the total amount produced. Afterwards the calculated amount of dust in t/year was converted into g/sec.

Table 1: Average dust produced and filters installed per technique and year, adapted from Schueneman (Schueneman 1963)

Technique	**Average dust production [g/t]**	**Installation year**	**Filter**	**Efficiency [%]**
Blast furnace	90718.5	1911	Preliminary	97
		1948	High efficiency	99.50
		1976	State of the art	99.90
Sintering furnace	9071.85	1931	Centrifugal separators	90
EAF	48080.79	1938	No filters	-
LD-AC	18143.7	1967	State of the art	99.50

To determine the wind directions and velocities per year and month, weather data from the yearly meteorological reports for Luxembourg City (about 20 km from Esch) ('Annuaire météorologique et hydrographique' 1949) were used and converted from the Beaufort scale to m/s. Since weather data for the years 1911-1949 were not available in digitised form, an average wind profile consisting of speed and direction – taken from the 10th and 20th day of each month – was created and used for all years, based on the data for the 1949-1997 period (see Figure 7).



GRAND-DUCHÉ DE LUXEMBOURG

MINISTÈRE DE L'AGRICULTURE

Administration des Services Agricoles

ANNUAIRE

MÉTÉOROLOGIQUE ET HYDROGRAPHIQUE

1949

PUBLIÉ PAR LE SERVICE MÉTÉOROLOGIQUE ET HYDROGRAPHIQUE

LUXEMBOURG
1951

H = 334 m.
λ = E 06° 03'
φ = N 49° 37'

LUXEMBOURG (VILLE) Janvier 1949

Observateur : E. LAHR (Professeur à l'Athénée)

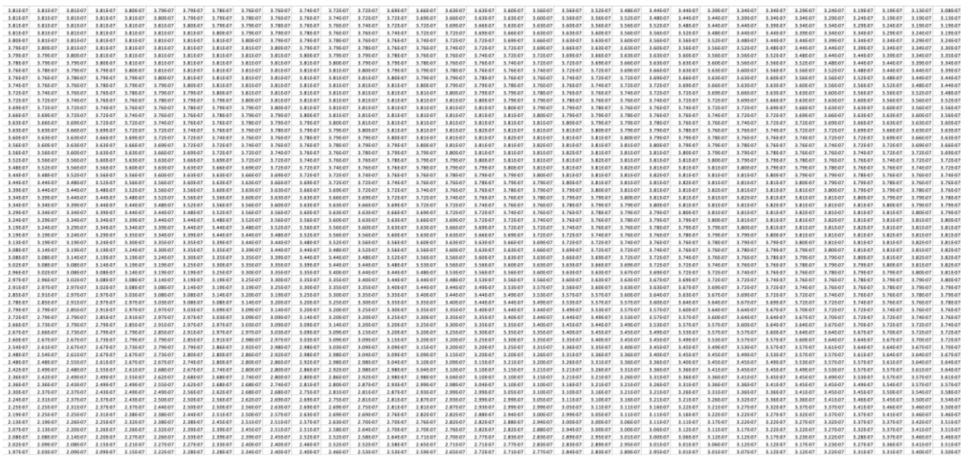
JOURS DU MOIS	TEMPÉRATURE DE L'AIR T			PRESSION ATMOSPHÉRIQUE P			HUMIDITÉ RELATIVE U			NUAGES N			DIRECTION ET FORCE DU VENT DF			PRÉCIPITATIONS	COUCHE DE NEIGE	INSOLATIONS
	7	13	21	7	13	21	7	13	21	7	13	21	7	13	21	mm	cm	Heures
1	2,9	6,0	5,0	715,8	710,1	711,6	93	76	93	8	10	10	SW8	S 6	SW7	4,0	.	
2	1,8	4,2	2,0	715,8	717,2	718,5	98	97	96	10	8	10	S 3	S 2	SW2	6,0	.	0^{30}
3	2,5	1,5	-0,4	722,3	725,2	728,9	94	96	94	8	10	0	SW3	NW3	NW1	2,0	.	
4	0,0	2,5	0,5	730,4	730,0	730,1	96	93	97	10	7	10	SW2	S 4	S 2	1,0	0,5	0^{30}
5	2,1	3,6	4,0	731,6	737,3	741,4	98	98	98	10	10	10	S 1	S 1	S 1	3,6	.	.
6	3,4	4,8	2,5	746,0	747,0	748,0	98	96	94	10	10	10	S 1	S 1	S L	0,8	.	.
7	0,5	0,7	-1,0	746,9	745,0	741,6	95	97	97	10	10	10	S 1	S 1	NW2	.	.	.
8	-1,0	0,5	0,6	736,7	734,9	734,1	98	95	97	10	10	10	SW3	SW2	SW1	.	.	
9	-1,0	1,5	0,8	738,3	737,9	737,9	94	85	88	8	7	8	N 1	W 2	NW2	1,•	.	0^{15}
10	-2,8	0,0	-2,8	740,0	738,7	737,1	96	95	97	10	10	10	W 2	NE2	NE3	1,0	1,0	300
11	-4,8	-2,6	-1,6	733,7	731,8	731,0	97	97	97	10	10	10	SE3	S 4	SW6	.	2,•	.
12	0,0	1,2	-1,0	729,8	733,2	739,7	97	88	95	8	1	0	N 3	NE2	NE2	2,0	3,0	800
13	-2,0	1,4	-1,5	743,9	745,8	747,6	94	87	95	4	2	8	W 2	W 4	W 3	.	2,0	400
14	-0,8	0,5	2,2	744,5	742,0	741,3	96	97	96	10	10	10	SW2	SW1	SW1	1,0	1,5	.
15	1,9	3,4	2,2	739,5	739,1	739,0	95	92	94	10	10	10	W 2	W 2	W 2	5,5	.	.
16	3,0	7,0	5,2	736,5	735,2	736,5	94	97	96	10	10	10	W 3	SW2	SW2	2,0	.	.
17	4,2	5,1	6,0	735,7	735,3	734,6	94	90	93	10	10	10	SW2	SW4	SW4	0,8	.	.
18	6,7	7,0	5,8	735,2	736,8	738,1	96	85	82	10	9	10	SW4	W 2	W 3	4,0	.	.
19	7,0	8,0	6,8	738,3	737,8	736,6	96	80	80	10	10	10	W 4	W 6	W 6	1,8	.	.
20	5,5	5,7	3,6	736,4	734,9	735,5	88	97	88	10	10	3	SW5	SW4	W 2	.	.	.
21	2,9	5,0	4,0	735,2	734,7	734,9	94	77	78	10	10	10	W 4	NW8	W 5	3,2	.	.
22	1,7	4,7	0,0	739,2	740,9	741,5	94	56	92	6	4	0	N 2	W 2	N 2	0,2	.	630
23	-4,0	3,5	-0,2	742,6	741,3	740,1	97	60	80	0	0	0	N 2	S 1	E 3	.	.	800
24	-2,5	4,0	0,5	739,3	740,2	742,5	97	80	90	0	0	0	E 3	E 2	E 2	.	.	800
25	-0,5	4,6	0,5	744,7	744,0	744,0	85	69	82	0	2	0	E 2	E 2	NE2	.	.	800
26	-2,2	3,0	1,0	744,3	744,9	746,3	95	80	92	0	0	0	E 1	S 1	S 1	.	.	830
27	-4,2	6,0	1,1	746,6	746,0	746,3	98	72	86	0	2	0	S 1	S 1	S 1	.	.	830
28	-6,5	3,0	0,5	747,8	748,8	748,4	97	94	92	0	0	0	S 1	S 2	S 1	.	.	800
29	-5,0	5,0	1,5	749,2	749,0	749,1	98	68	82	0	0	0	SE1	SE1	SE1	.	.	800
30	-4,6	4,2	0,2	748,3	747,9	748,7	96	81	88	2	8	0	N 3	NE2	NW2	.	.	630
31	-2,0	1,5	0,5	747,3	744,4	740,5	96	92	96	0	10	10	NW2	N 2	NW2	.	0,2	.
MOYENNES	0,0	3,4	1,5	738,1	737,9	738,4	95	86	91	6,5	6,7	6,1	–	–	–	Total 39,9	–	Total 84^{15}

Figures 7 a and b from left to right: (a) Cover of and (b) excerpt from the yearly meteorological report for Luxembourg (in this case showing the data for Luxembourg City, including the average wind speeds and directions in January 1949). Source: (*Annuaire Météorologique et Hydrographique 1949* 1951)

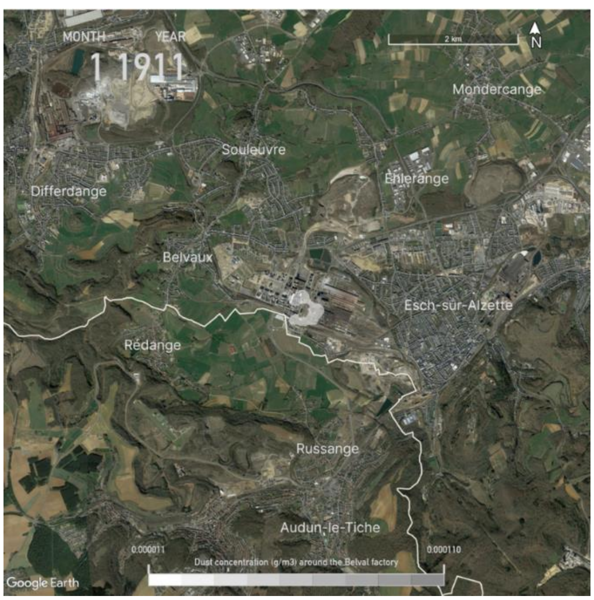
Using this input data, the dust concentration (g/m^3) of all 15m x 15m x 1m cells – in a 2.5 km radius around the Belval plant – was calculated, using a MATLAB ('MATLAB - MathWorks' 2022) script based on equation (1). The code was generated by metallurgical experts[10] and is therefore not included here. Nevertheless, we show the input, output (csv files) and further processing (data visualization) of the MATLAB script. MATLAB is a computing environment and a programming language used by engineers and scientists, aiding with data analytics, linear algebra, signal processing, *etc.* MATLAB language is a matrix-based language that includes many specialized libraries to solve engineering and scientific problems. Although it also includes plotting capabilities, we have decided to use Python for the data visualization, to have more flexibility.

[Data is included in Jupyter notebook (JDH), example matrix below just for illustration – not an actual image]

[10] The metallurgical experts (civil engineers) who performed the calculation were Sander Arnout, Yannick Cryns and Cem Tekin.



The visualisation of the model was generated using a Python in Jupyter notebook, from 1044 generated matrices of 302x302 values, saved in separate csv files. Each file corresponds to a month and a year between January 1911 and December 1997. The data is not geocoded; instead, the reference point (or source point) is one hypothetical chimney of the steelworks, located at the centre of the matrix, from which concentration values are calculated. The visualisation process of the data is explained in section 7.



[Actual visualisation is included in the Jupyter notebook (JDH), images are just for illustration – not actual images]

It can be said in advance for our study that the calculated concentration values are most likely to be underestimations, since only one emission source in southern Luxembourg was considered (thus eliminating many others). As such, our analysis serves as an experiment and can be taken as a basis for further research (*e.g.* calculating the dust concentrations from all industrial plants in the entire Minett region). Other limitations and improvements of our model are discussed in sections 8.3 and 8.4.

Data visualisation

As introduced above, the data visualisation was generated using a Jupyter notebook with Python for a multi-layered approach, which allows us to share not only the narrative layer but also the data and the code. The data was not geocoded, since explicitly assigning coordinates to each concentration point in each dataset would have increased the volume of data considerably. Instead, the data was kept in the form of multiple matrices and positioned on the map using Google Earth to calculate the distance covered by the density matrix (see Figure 8).

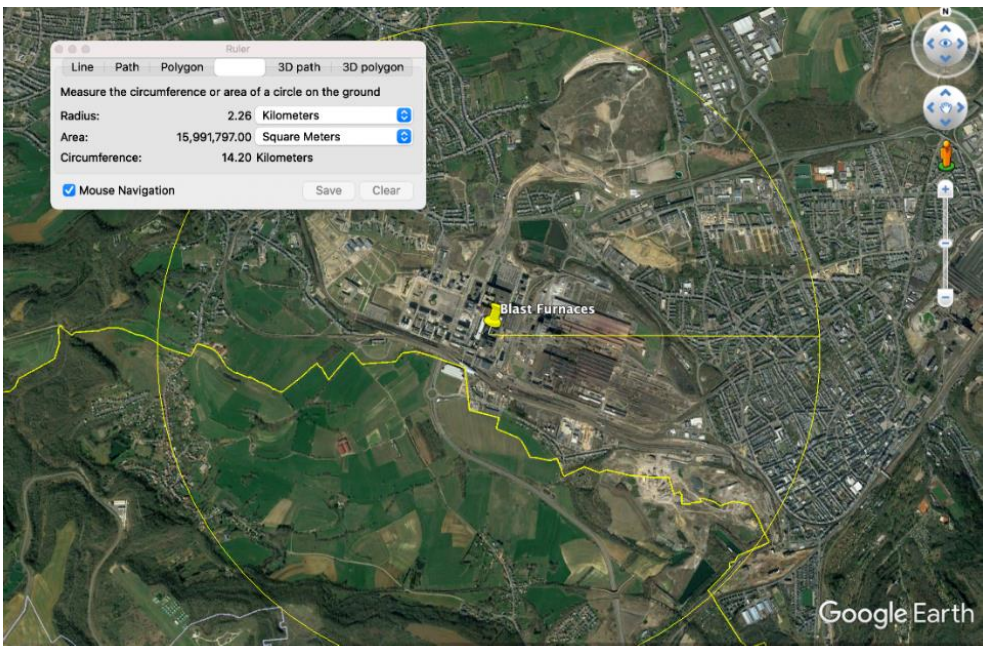


Figure 8: Screenshot of Google Earth[11] demonstrating how the distance from the Belval plant is measured.

[11] https://www.google.com/intl/de/earth/

Initially, the option of using a historical map (see Figure 9) was considered. This was discarded in favour of facilitating more accurate calculation of distances and generating a view with the steelworks at the centre.

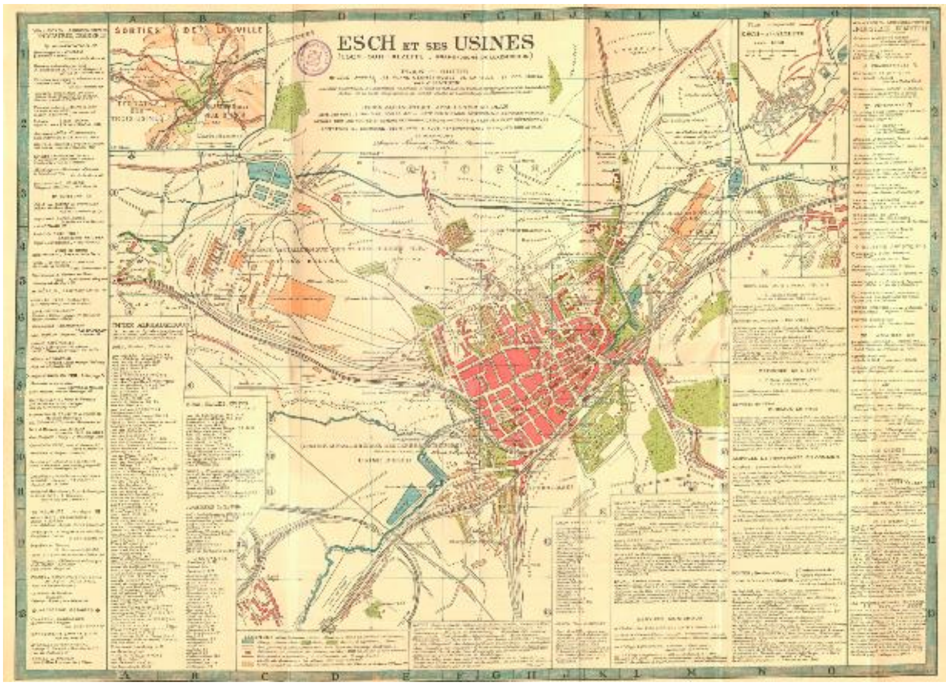


Figure 9: 1923 map of Esch. Source: (Hansen 1923)

The dust concentrations around the Belval plant (in g/m^3) were visualised using a contour map with ten predefined levels. We used the extent parameter to plot the concentrations on the map, and a customised sequential colour palette to indicate the presence of dust and emphasise a higher concentration near the plant.

[Code to generate the contour map using the extent parameter and colour palette, see JDH]

Contour maps were generated for every month and year between January 1911 and December 1997 in English, French and German. For this purpose, the names of the main localities were translated; using Figma, they were added to the base maps extracted from Google Earth (see Figures 10 a and b).

Figures 10 a and b from left to right: (a) English and French version of the map, (b) German version of the map.

The visualisations were subsequently animated to show evolution over time.

[Code to generate the animation, see JDH]

To visualise the impact of pollution in the inhabited quarters of Esch, we repeated the process with a zoomed-in map, thus highlighting the immediate vicinity of the Belval plant.

[Code to calculate the contour plot and the new extent, and the video, see JDH]

The videos were integrated into a dedicated page of the *Minett Stories* virtual exhibition, which was launched in May 2022. Produced by the Centre for Contemporary and Digital History (University of Luxembourg) in connection with *Esch-sur-Alzette – European Capital of Culture 2022*, the exhibition offers twenty-two stories, outlining various transformations of landscapes, places and people, from the region's industrialisation in the late 19th century to the crisis of the 1970s and the subsequent deindustrialisation. Instead of focusing on the often-told stories about the growth and success of the steel industry, *Minett Stories* explores everyday life in the Minett, with attention to questions such as environmental pollution.

Our visualisation helps us to answer the initially formulated questions about the geographical distribution of dust pollution. By using various historical maps of Esch (see Figures 11 a,b,c) and comparing them to the data visualisation for the same year, we can analyse which neighbourhoods were most affected over time. July has been chosen, as this is the month when the pollution typically reached the farthest in the town of Esch.

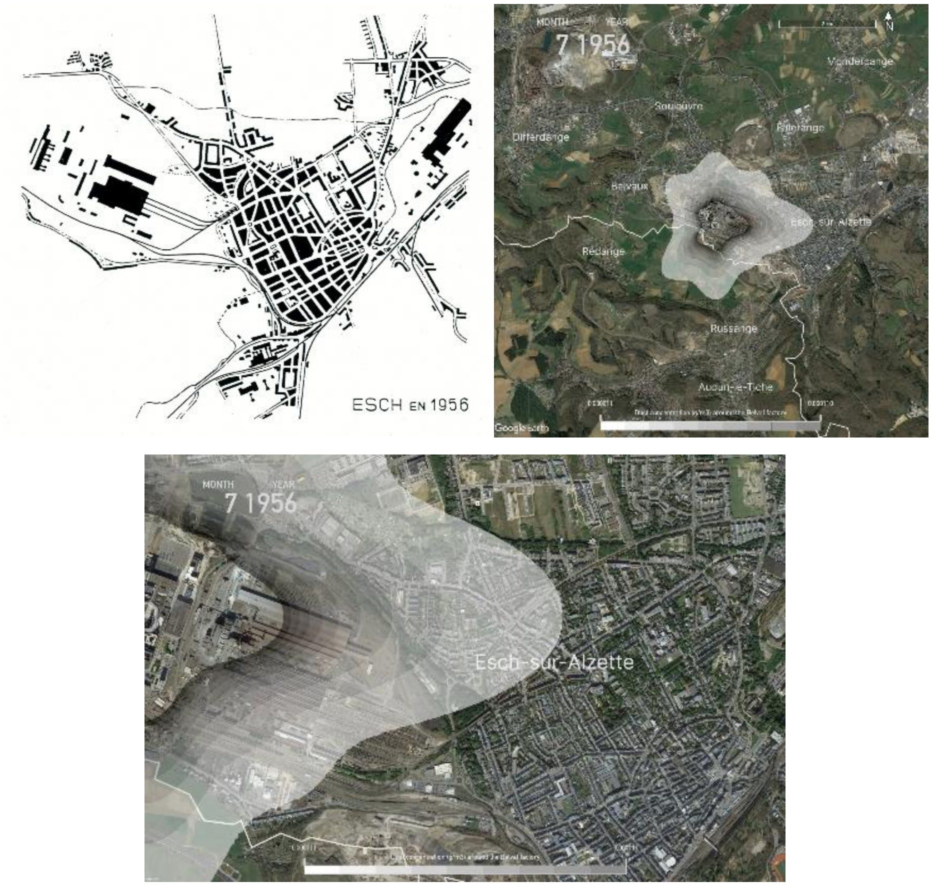


Figure 11 a,b,c: from left to right: (a) 1956 map of Esch, (b) simulation of dust pollution around the Belval plant in July 1956 showing the entire region, and (c) simulation of dust pollution for the same month and year with a focus on the centre of Esch. Source: (Buchler et al. 2020)

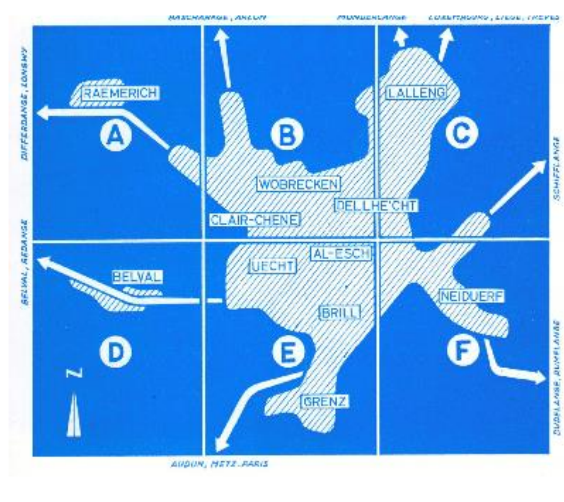


Figure 12: 1958 map of Esch, showing the location of various urban quarters. Source: ('Esch-Sur-Alzette Nouveau Plan 1958. Répertoire Des Rues et Places' 1958)

It is clear, for example, that high dust levels were often prevalent in the centre of Esch; unsurprisingly, the western part of town was the most severely affected. Using a 1958 map of Esch as a reference, the affected areas primarily included the neighbourhoods of Uecht, Clair-Chêne, Al-Esch and Wobrecken. Moreover, from the outset, Belval's dust pollution was a cross-border phenomenon, with the French villages of Rédange, Russange and Audun-le-Tiche (the latter itself the site of an ironworks from the 1880s to the 1960s) sharing the pollution burden with small Luxembourgish locations such as Soleuvre, Ehlerange and Belvaux.

Final analysis

To conclude, we perform a critical evaluation and interpretation of the dust simulation. We explore the plausibility of the calculated values, comparing them to historical and current air pollution legislation, as well as to historical measurements. Limitations and possible ways to optimise the model are also discussed.

8.1 Validation of the model

Figure 13 shows the summarised (total) amount of dust calculated by year *versus* steel production in tonnes, with dust in blue and production data in red. Important events like the (hypothetical) installation of filters are indicated. The graph further indicates a number of historical dates (such as the installation of filters, changing techniques, World War I and II, and the steel crisis from the second half of the 1970s onward), which all correspond to visible changes in the production and/or dust curves. This effect is for instance highly visible after 1965, with steel production increasing and the dust curve decreasing, owing to more intensive use of oxygen-based steelmaking and the hypothetical installation of 'state-of-the-art' filters (see section 0 and Table 1).

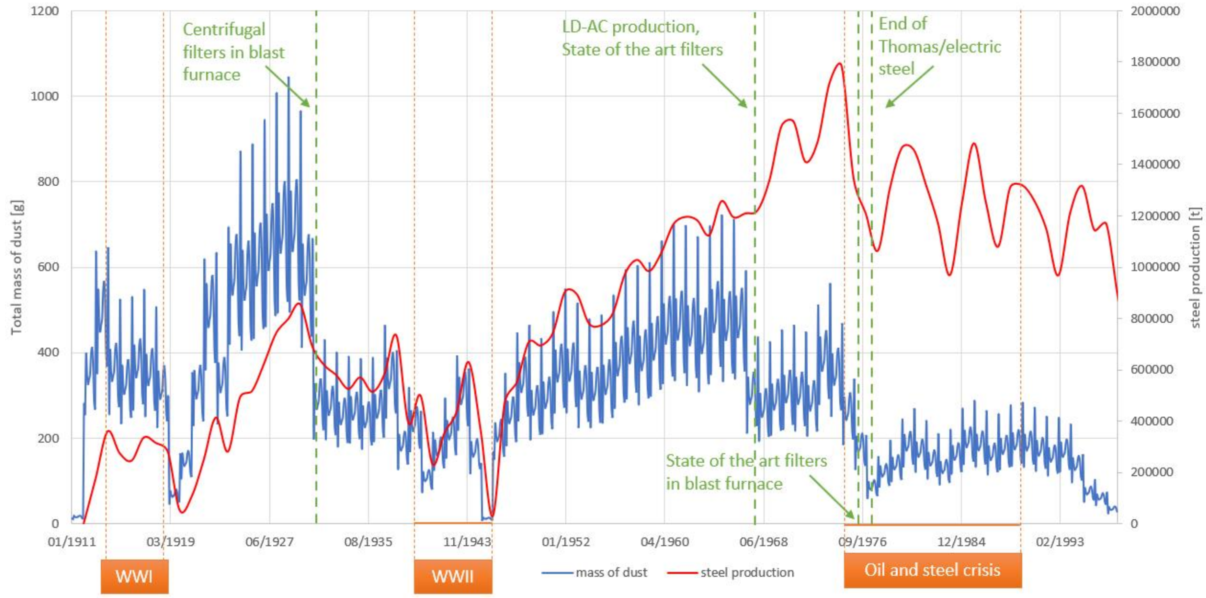


Figure 13: Total amount of dust calculated in blue; steel production numbers in red; important events indicated

As well as checking the plausibility of the data, it is necessary to look at the actual meaning of the calculated dust values (with the highest calculated value being about 3 mg/m^3 in October 1929). To understand the effects of such values on human health, we will compare them to present-day regulatory limit values.

Directives and measures regarding air quality were first introduced by the European Economic Community in the 1970s. In subsequent decades, air pollution resulting from industry decreased significantly (European Environment Agency EEA 2013). In Luxembourg, the first legislation regarding air pollution was the 'Loi du 21 juin 1976 relative à la lutte contre la pollution de l'atmosphère'. No limit values were laid down in this Act. In the more recent context of EU law, Directive 2008/50/EC (European Union 2008) postulates limit values for different atmospheric pollutants to protect human health, including values for $PM_{2.5}$ (25 µg/m^3) and metals like Cd (5 ng/m^3) (European Parliament. Directorate General for Parliamentary Research Services. 2021). Such fine-grained values not only indicate recent improvements in measuring methods; they are also indicative of an increase in awareness among legislators in recent decades.

Our model, however, does not distinguish between the various dust fractions or different pollutants. When comparing the simulation model with the current EU legislation values, the calculated maximum

(3 mg/m^3) clearly exceeds the limit related to particulate matter concentration (25 µm/m^3 for $PM_{2.5}$). Furthermore, the EU directive from 2008 is under revision following the recent proposal of more stringent values by the World Health Organization (WHO). In 2021, the WHO indeed put forward a maximum value of 5 µg/m^3 for long-term $PM_{2.5}$ exposure in order to protect human health (World Health Organization 2021).

Specific legislative values have also been established for industrial producers. Relevant EU laws include the 2001 'Directive on the limitation of emissions of certain pollutants into the air from large combustion plants (2001/80/EC)', followed by the 2010 'Directive on industry emissions (2010/75/EU)', which focuses particularly on industrial pollution. Summarising the various sector-specific directives, a dust limit concentration of 30 mg/Nm^3 for coal and other solid fuels (lower values apply for gaseous fuels and higher thermal nominal power) is valid in all EU Member States. The N in the concentration unit emphasises an important point to consider. Since dust values depend on pressure and temperature, a comparison with other values requires standardised conditions (N: 25°C, 1 atm). This is another limitation of our model, as temperature and pressure data were not included. Compared to the simulated dust concentrations, the limit value is much higher.

Decision (EU) 2017/1757 of the Council of the European Union defines separate dust limit values for the iron and steel industry: 50 mg/m^3 for sintering, 15-20 mg/m^3 for pelletising plants, 10 mg/m^3 for hot blast stoves in blast furnaces (> 2.5t/h), 30 mg/m^3 for oxygen steelmaking and casting (> 2.5 t/h), and 5 (new ovens) to -15 (existing ovens) mg/m^3 for steel production and casting using the EAF process (> 2.5 t/h) (European Union 2017). In comparison, our calculated dust amount is below these limits. These values, however, define steelmaking processes of today, not of the past; moreover, they are process specific.

Regarding workplace exposure limits measured in time-weighted averages, pollutant-specific levels can be found in legislation; in Luxembourg they are regulated by the 'Règlement Grand-Ducal du 14 novembre 2016 concernant la protection de la sécurité et de la santé des salariés contre les risques liés à des agents chimiques sur le lieu de travail' (Le Gouvernement du Grand-Duché de Luxembourg 2016).

In general, it can be said that it is difficult to compare the simulated dust values with legislative limit values. First, our values do not include information on external conditions like temperature and pressure that is necessary to perform an exact comparison. Second, and even more important, the calculated dust concentrations reveal neither the composition of dust and the type of pollutants, nor the dust generating process in steelmaking. Therefore, limit values for specific dust fractions or processes cannot be applied. Moreover, when looking at dust concentrations in legislation, it clearly makes a difference whether one looks for general or sector-specific directives, as illustrated above.

Nevertheless, there is evidence for the health impact of air pollution coming from the iron and steel industry in Luxembourg and other countries. Of course, other factors such as climate, noise, vibrations and other types of pollution (*e.g.* soil, water) had an influence, too. Regarding health effects analysed in Luxembourg, Molitor and Mosinger published a research study on the effects of atmospheric dusts on human health in 1960 (Molitor and Mosinger 1960). They found evidence for the connection of industry dust with pulmonary diseases like silicosis and siderosis. In 1975 there was a medical symposium (industrial medicine) taking place in Luxembourg, collecting several studies on the effects of industry emissions on respiratory diseases and proving the impact of dust emission on health (Comission of the European Communities ECSC 1976). A special focus of researchers was on the occupational health of steel workers, which were exposed to even higher amounts of pollution than the citizens around the steel plant (occupational medicine (Gochfeld 2005)). Other studies demonstrating the impact of atmospheric steel pollution on human health were conducted *e.g.* in England, studying mortality effects (Beach and Hanlon 2018), in China, looking at heavy metals - representing a carcinogenic risk above certain limits - (Qing 2015) or in Albania (Bala and Tabaku 2010) and Brazil (Valenti et al. 2016) studying chronic obstructive pulmonary disease.

In combination with time witness reports (‘Compte-rendu Chambre des Députés, volume 1 - 1975-1976’ 1976) and dust measurements conducted in other countries (see above) this information supports the assumption that the dust pollution in the area around Luxembourgish steel plants exceeded most legislative limit values introduced until today.

8.2. Comparison of the model results with air pollution measurements from the late 1930s to the early 1970s

To compare the simulated dust levels with real conditions, it is important to scrutinise historical documents on dust measurements in the Minett region. We were able to retrieve data on four measurements performed between the late 1930s and the early 1970s: 1939 (Schifflange, undertaken by the ARBED steel company) (‘Étude sur le dépoussiérage des gaz des hauts-fourneaux à l’usine d'Arbed-Esch' 1939), 1956-1957 (Differdange, undertaken by the government) (Molitor and Barthel 1958), 1966-1969 (Esch, undertaken by the municipality) ('Retombées de poussières: Rapport sur les mesures de retombées en poussière effectuées de 1966 à 1969 à Esch-sur-Alzette' 1972), and 1970-1971 (Belval, undertaken in an academic context) ('Les Composes du Fluor dans la pollution atmospherique' 1971). In all cases, dust was collected for a specified period through passive samplers (known as Bergerhoff devices). Both the motivation for the measurements and the methodological framework for the subsequent interpretation differed over time, as we will outline below.

In 1939, ARBED undertook dust measurements on the premises of its Schifflange iron and steel complex, located directly to the east of Esch (about 2.5 km from the Belval plant, which was owned by the same company) ('Étude sur le dépoussiérage des gaz des hauts-fourneaux à l'usine d'Arbed-Esch' 1939). Investigated parameters included the amount of dust generated by the blast furnaces, particle sizes and the iron (Fe) content of the samples. This analysis was conducted for economic reasons only, with the aim of studying the potential for reusing blast furnace gas after the removal of dust particles. The analysis pointed out that very small dust particles (ranging from 0 to 10 µm) constituted about 25% of the total emitted dust mass; as noted in section 3, such particles can penetrate the respiratory system and pose a threat to human health. In addition, around 48% of all dust released was reported to be made up of Fe: a very high iron content that was not unusual in the context of the Thomas steelmaking process. Even though the 1939 corporate investigation concluded that recycling and filtering techniques would be too expensive to implement, it is clear – in retrospect – that the installation of a dedicated filter could have resulted in significant health gains for both the local population and factory employees.

About two decades later, the scientists Molitor and Barthel conducted a government investigation of the dust concentration in the vicinity of the Differdange iron and steel complex (1956, published in 1958) (Molitor and Barthel 1958). Six passive samplers were placed for a period of about three months in a 2 km radius around the plant, allowing the amount of dust in g per m^2 per month to be calculated and a compositional analysis to be performed. Molitor and Barthel claimed that insoluble particles such as Fe_2O_3, SiO_2, CaO and SO_4 did not have harmful effects, regardless of particles sizes. However, pulmonary diseases like siderosis (caused by iron dust) were identified decades before this study was published, indicating that this claim was incorrect even at the time of publication (Doig and Mclaughlin 1936). Dust fractions below 0.03 mm (*inhalable* fraction, see section 3) were observed in the samples, including silica, which indeed poses a risk for respiratory diseases and other health issues.[12] The Differdange analysis further indicated that dust concentrations dropped rapidly as the distance from the factory increased, starting with 140 g/m²/month near the centre of pollution and falling to about 40 g/m²/month at a 2 km distance. With the decreasing dust mass, the iron content decreased as well, dropping from 75% to 38.4% (Molitor and Barthel 1958). These measurements once again show the high contribution of steelmaking to dust pollution levels in the Minett, especially near blast furnaces. However, the values measured by Molitor and Barthel cannot be compared directly to the simulation presented in our analysis, since the passive sampler was unable to capture the smallest fractions of dust (below 5 µm (Heyart 1958); in addition, the measurement units and production site are not the same. The measured values in 1956 nevertheless seem to be far higher than the simulated ones, which indicates that the dust concentrations in our simulation are probably underestimated.

In 1966-1969, Barthel was again involved in dust measurements, this time commissioned by the municipal administration of Esch (‘Retombées de poussières: Rapport sur les mesures de retombées en poussière effectuées de 1966 à 1969 à Esch-sur-Alzette’ 1972). Barthel’s new observations were publicly released in 1972, a year marked by intense worldwide interest in environmental pollution problems – not least with the publication of the Club of Rome’s report *Limits to Growth* (Buelens 2022). Compared to the Differdange analysis, the number of measurement points was increased to 52 (spread over the entire surface area of the town), resulting in 16 samplers per km^2. This allowed for a more fine-grained analysis of the accumulated dust generated by multiple iron and steel complexes (Belval, *Terres Rouges* and Schifflange). The measurements were subsequently compared to the West German limit values for atmospheric dust pollution, which were a maximum of 0.42 g/m^2/day in urban zones and 0.85 g/m^2/day in industrial zones (‘Retombées de poussières: Rapport sur les mesures de retombées en poussière effectuées de 1966 à 1969 à Esch-sur-Alzette’ 1972). Over half of the measurement points were reported to exceed the West German dust limits for *urban* regions (34 in 1966-1967, 35 in 1967-1968 and 28 in 1968-1969), while several points even showed values above the *industrial* limits (9, 6 and 7 respectively). When compared to the simulation model, the measured values are significantly higher (dimensions of measurements in g/day and for the simulation in mg/month). The dust composition showed an iron oxide content of about 65%, varying with distance,

[12] Molitor and Mosinger even published a study themselves on industry dust resulting in silicosis and siderosis in 1960 (Molitor and Mosinger 1960).
Knowledge about respiratory diseases caused by dust was even present in ancient times. See: (Rosen 1993, 459–76)

indicating a clear link with the steelworks. Even though Barthel (incorrectly) stressed that the dust concentrations had no significant effect on human health, he did call for technical improvements that could bring the pollution levels down to the West German norm for industrial zones.

Lastly, in 1971, Christiane Conter wrote a dissertation on atmospheric pollution, based on measurements performed in 1970-1971 as part of an academic research project at the Belval plant ('Les Composes du Fluor dans la pollution atmospherique' 1971). Conter's work clearly presented an environmental point of view, with a focus on various sources of pollution (steelworks, households and traffic). The dust samples at the Belval steel complex were specifically analysed with regard to fluorine compounds, known for being released during various industrial processes and for their toxic effects on vegetation and animals. Sampling stations were placed near various points in the production chain; the average dust concentration measured was 6 g/m^2/day, which exceeded not only the West German legislative limits but also the measurements undertaken by Barthel in the years before. Just like Molitor and Barthel (Differdange, 1956-1957), Conter analysed dust composition in terms of soluble and insoluble particles. However, her specific focus on fluorinated compounds yielded no concrete results, as possible trace amounts were superimposed by the dust matrix; additionally, the methods used (X-ray spectroscopy and electron microscopy) were not sensitive enough to detect fluorinated compounds ('Les Composes du Fluor dans la pollution atmospherique' 1971).

Despite the lack of comparability between the simulated values and the available measurements, our simulation allows us to isolate the dust pollution potentially generated by a single steelworks in the past and to understand how far it reached overtime. From this point, the analysis could be extended to other industrial plants and other types of pollution, which would probably bring us closer to the existing measurements. Furthermore, the lack of detail in historical measurements, which obviously cannot be measured again (*e.g.* a precise type of pollution created by a particular plant at a specific point in time), creates an opportunity for the use of modelling and simulation.

8.3. Limitations and lacunae of the dust simulation

Throughout this research, we have presented different decisions taken during the process of simulation, validation and visualization of the data. These decisions simplify and reduce the complexity of the problem, but they have been consciously made to serve a specific purpose. This was to simulate industrial dust pollution from a single factory in a given period, based on historical data; above, we have presented the results of these decisions. This simulation can change perspectives on the past, *e.g.* with looking at the seasonal changes of wind directions and therefore changes in dust distribution (which is often omitted in historical literature). In future steps, we could modify the model variables (*e.g.*, using data from other sources), or include new variables and use a different model (see section 8.4). In any case, we have proven that data simulation based on historical data offers the option not only to recreate situations in the past, but also to isolate effects that would not be possible to analyse even with real measurements.

Nevertheless, we want to present a final analysis of the main limitations and gaps of our simulation, to which we have already alluded in previous sections. Such limitations become relevant if we seek to understand the total pollution to which the analysed area was exposed. First, not all dust emitting sources in the Belval region were considered; obviously, the steel industry was only one contributor to atmospheric pollution (albeit a very important one) (Metz 1972; Hoffmann 1974; 'Compte-rendu Chambre des Députés, volume 1 - 1975-1976' 1976). Pollution from households, industrial heating and traffic (motor vehicles and trains) accounted for some of the dust particles as well. Second, regarding the mathematical formula (1) used (presented in section 6), a few parameters must be analysed critically. The choice of a Gaussian model in itself involves simplifications. The model assumes that the contaminants come unidirectionally from one point source (one of several chimneys) at a constant emission rate *Q*, with the wind direction always aligned in one axis, considering the time frame of one month (Stockie 2011; Leelőssy et al. 2014). This assumption does not cover intramonth variations and only considers one metallurgical plant (from a total of three in Esch alone). The model is not suitable for low wind speeds, and in terms of meteorological data only averages for Luxembourg City (the 10th and 20th of each month), which is about 20 km from Belval, were used. In Gaussian plume models used in the calculation of atmospheric pollution, the specific features of the surrounding terrain are usually considered (Stockie 2011). As such, our formula could be extended to involve terrain information and a greater radius could be covered to better observe long-range effects. State-of-the-art Gaussian models like AERMOD, CTDM and ADMS try to extend the basic equation to make it more accurate and include more factors like a complex terrain (Leelőssy et al. 2014). Other dispersion models like the Lagrangian, Eulerian or CFD models are more suitable for different scales and applications.

Another factor influencing the results is our use of US data from 1963 (Schueneman 1963) on average dust values for different steelmaking techniques and typical filter installation dates. Steelmaking and filter use vary between individual steel plants. The production data used covered only annual values, so data on monthly production would make the simulation more precise. To better interpret the calculated dust values with regard to regulatory limit values and to determine a possible health effect, it would be useful to identify the dust fractions and pollutants contained therein. This is complicated, however, as historical dust samples have not been preserved and documents on measurements performed in the past (see 7.2) only study a small set of pollutants. Dust pollution values are usually standardised using temperature and pressure, which presents another limitation for interpretation. Lastly, the dust values identified using the model are technique specific (*e.g.* for Thomas steel) and do not relate to dust generation processes (*e.g.* sintering) used to set limit values for current EU regulations.

To summarise, the dust simulation of the Belval complex between 1911 and 1997 is just a starting point towards modelling the operation of a real-world process: the total generation and spread of pollution to which the inhabitants of Esch and the surrounding area were exposed. In Section 8.4, we offer a few possible ways of optimising the simulation in future research.

8.4. Optimisation and improvement of the model

Some of the lacunae mentioned in section 8.3 can be improved, thereby opening possibilities for further research. Other emission types like traffic, heating and other industrial plants could be considered, as well as the terrain around the point sources. The overlap of multiple dust generating sources and the resulting plumes, however, involves the risk of making the visualisation too complex and therefore confusing. As we state above, the model and used parameters were chosen on purpose, simulating an isolated phenomena.

The model area could be expanded, and temperature and pressure data could be considered. All these changes would increase the complexity of the mathematical formula. The meteorological data for the years 1911-1948 could be digitised to access additional weather data such as the wind directions and speeds for all months. Moreover, subdividing dust particles into different pollutants and fractions would enhance accuracy, while the generation of dust for each individual ironmaking and steelmaking process could be an interesting parameter to analyse further (especially with regard to the health impact of dust). However, the latter extension of the simulation is at the limit of what is practically feasible, since past dust samples (or contemporary analyses) from the respective processes are required for a precise investigation of the dust, its fractions and its origins. For several short periods, a composition analysis of the dust particles was performed; for other periods, the amount of dust created by different processes in the plant was measured, as explained in section 8.2. Using these data points is a possible way to extend the simulation, even if only for these specific years. Finally, it would be beneficial to look for average dust values produced by Luxembourg's steel industry, monthly production data, filter installation dates and filter efficiency rates. Such data is most likely present in the corporate archives of ARBED, which are unfortunately not completely accessible for researchers. The optimisation possibilities outlined above would take our study beyond its initial scope.

8.5. Conclusion

Air pollution has a long history: it is not only a problem of the past but also an issue affecting the present and the future. Especially in industrialised regions, atmospheric pollution is not only annoying; it can also be a dangerous companion of daily life. Dust particles in the respirable fraction pose the most worrisome threats to human health, as they can lead to lung diseases. In many parts of the world, the steel industry has contributed significantly to the total mass of airborne dust. Improvements in production techniques and the installation of filters, as well as legislative measures and economic crises, have influenced air pollution over the years.

Our analysis outlines the evaluation of dust production from a single plant in terms of plausibility and health effects. The simulated dust concentrations were compared to measurements and observations from the past, representing the 'scientifically proven reality' during a specific time period. For the simulation, a Gaussian plume model was used; the inserted parameters were analysed critically in terms of their limitations and possible improvements. Given the lack of availability of data regarding the dates of filter installations in Belval, secondary literature on the US case was used.

When the results are compared with dust measurements performed in the past, a clear deviation can be seen. Measurements from the period between the 1930s and the 1970s show that the dust concentration in the Minett was much higher than the simulation shows. Even limit values for industrial zones – taken from West Germany – were exceeded. Moreover, a high iron content and smaller particle fractions were identified. However, only a partial comparison with measured values is possible, since there are differences in terms of units, techniques and measurement area. One of the challenges of modelling and simulation is to find the right balance between complexity (and therefore precision) and interpretability. In this case, given the already added complexity of the interdisciplinary approach, we decided to use a simplified model, which serves as a precedent for creating new ways of generating historical data. At the same time, it allows us to present evidence (albeit on just a fraction of the dust to which inhabitants were exposed) and to understand the impact on health in a specific historical context.

The decision to visualise the pollution as coming from just one chimney also has benefits, as adding multiple dust sources might overload the visualisation (in the latter case, it would no longer be possible to discriminate between the origins of the dust). Furthermore, it has never been possible to measure the isolated dust concentration of a single plant using samplers, since there are always other dust sources. Hence, a simulation provides a suitable heuristic tool for generating this non-existent (historical) data. In sum, our model serves as a basis for studying atmospheric pollution released from one selected source and offers the potential to be extended by looking at other contributing factors.

Acknowledgements and funding

We would like to thank Sander Arnout, Yannick Cryns and Cem Tekin from the metallurgical consultancy company Inspyro (Leuven, Belgium), who built the initial simulation. Our thanks also go to all the REMIX and *LuxTIME* team members who took part in discussions about our research, as well to Sarah Cooper for her language proofreading.

This research has been funded by *Esch-sur-Alzette – European Capital of Culture 2022* and the University of Luxembourg Institute for Advanced Studies Audacity Programme for the Luxembourg Time Machine (*LuxTIME*).

150 Jahre Und Kein Bisschen Alt: 150 Joër Wort, 1848-1998. 1998. Luxembourg: Saint-Paul.

'Annuaire météorologique et hydrographique'. 1949. Annuaire météorologique et hydrographique. Luxembourg: Service météorologique et hydrographique. https://www.agrimeteo.lu/Internet/AM/inetcntrLUX.nsf/cuhome.xsp?src=WY1BE977UP&p1=Q0BOB9JITQ&p3=9VXL1WMS7M.

Annuaire Météorologique et Hydrographique 1949. 1951. Luxembourg: Service météorologique et hydrographique.

Assa, Guy, and Luciano Pagliarini. 1998. *Esch-Uelzecht Gëscht an Haut*. Esch-sur-Alzette: Schortgen.

Back-Hoffmann, Jean. 2021. 'Allgegenwärtige ARBED!' *Forum*, no. 419: 42–45.

Bala, Silvana, and Afrim Tabaku. 2010. 'Chronic Obstructive Pulmonary Disease in Iron-Steel and Ferrochrome Industry Workers'. *Central European Journal of Public Health* 18 (2): 93–98. https://doi.org/10.21101/cejph.a3548.

Banks, Jerry. 1999. 'Introduction to Simulation'. In *Proceedings of the 1999 Winter Simulation Conference*, 7.

Beach, Brian, and W. Walker Hanlon. 2018. 'Coal Smoke and Mortality in an Early Industrial Economy'. *The Economic Journal* 128 (615): 2652–75. https://doi.org/10.1111/ecoj.12522.

Brüggemeier, Franz-Josef, and Thomas Rommelspacher. 1992. *Blauer Himmel Über Der Ruhr: Geschichte Der Umwelt Im Ruhrgebiet, 1840-1990*. Essen: Klartext.

Buchler, Georges, Jean Goedert, Antoinette Lorang, Antoinette Reuter, Denis Scuto, Christof Weber, and capybarabooks. 2020. *Esch-sur-Alzette Guide historique et architectural*.

Buelens, Geert. 2022. *Wat We Toen al Wisten: De Vergeten Groene Geschiedenis van 1972*. Amsterdam: Querido Facto.

Butte, Werner, and Birger Heinzow. 2002. 'Pollutants in House Dust as Indicators of Indoor Contamination'. *Reviews of Environmental Contamination and Toxicology* 175: 1–46.

Carson II, John S. 2005. 'Introduction to Modeling and Simulation'. In *Proceedings of the 2005 Winter Simulation Conference*, 8.

Champion, Erik Malcolm. 2016. 'Digital Humanities Is Text Heavy, Visualization Light, and Simulation Poor'. *Digital Scholarship in the Humanities*, November, fqw053. https://doi.org/10.1093/llc/fqw053.

Comission of the European Communities ECSC. 1976. 'Medical Symposium: Researches on Chronic Respiratory Deseases - Industrial Medicine'. 18. Industrial Health and Medicine Series. Luxembourg.

'Compte-rendu Chambre des Députés, volume 1 - 1975-1976'. 1976. 53. Luxembourg.

'Do We Really Need Coal to Make Steel?' 2020. *Wildsight* (blog). 1 June 2020. https://wildsight.ca/2020/06/01/do-we-really-need-steelmaking-coal/.

Doig, A.T., and A.I.G. Mclaughlin. 1936. 'X RAY APPEARANCES OF THE LUNGS OF ELECTRIC ARC WELDERS'. *The Lancet* 227 (5875): 771–74. https://doi.org/10.1016/S0140-6736(00)56868-8.

'Esch-Sur-Alzette Nouveau Plan 1958. Répertoire Des Rues et Places'. 1958.

'Étude sur le dépoussiérage des gaz des hauts-fourneaux à l'usine d'Arbed-Esch'. 1939. Written by ARBED ingenieurs. Luxembourg. Bibliotheque Nationale Luxembourg.

European Environment Agency EEA. 2013. 'Rechtsvorschriften zur Luftqualität in Europa'. https://www.eea.europa.eu/de/signale/signale-2013/artikel/rechtsvorschriften-zur-luftqualitaet-in-europa.

European Parliament. Directorate General for Parliamentary Research Services. 2021. *EU Policy on Air Quality : Implementation of Selected Legislation : European Implementation Assessment.* LU: Publications Office. https://data.europa.eu/doi/10.2861/808342.

European Union. 2008. *Richtlinie 2008/50/EG über Luftqualität und saubere Luft für Europa*. https://eur-lex.europa.eu/legal-content/DE/TXT/PDF/?uri=CELEX:32008L0050&qid=1419250736800&from=DE.

———. 2017. *Beschluss (EU) 2017/1757 Des Rates*. Vol. 2017/1757. https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=CELEX:32017D1757&from=FR.

Galison, Peter. 1996. 'Computer Simulations and the Trading Zone'. In *The Disunity of Science: Boundaries, Contexts, and Power*, edited by Peter Galison and David J. Stump, 118–57. Stanford University Press.

Geissler, Stephanie. 2016. *Wem Gehört Die Stadt? Umweltkonflikte Im Städischen Raum Zur Zeit Der Früh- Und Hochindustrialisierung in Aachen Und Duisburg*. Münster & New York: Waxmann.

Glesener, Jeanne E., and Frank Wilhelm. 2009. 'L'image de La Sidérurgie Dans Les Romans Des Luxembourgeois Francophones Joseph Leydenbach Willy Gilson et Nicolas Ries: Des Années Fastes à La Crise Des Années 1970'. In *Terres Rouges: Histoire de La Sidérurgie Luxembourgeoise*, 1:71–95. Terres Rouges: Histoire de La Sidérurgie Luxembourgeoise. Luxembourg: Archives Nationales.

Gochfeld, Michael. 2005. 'Chronologic History of Occupational Medicine': *Journal of Occupational and Environmental Medicine* 47 (2): 96–114. https://doi.org/10.1097/01.jom.0000152917.03649.0e.

Guldi, Jo, and David Armitage. 2014. *The History Manifesto*. Cambridge, United Kingdom: Cambridge University Press.

Han, Qiao, Yang Liu, Xixi Feng, Pan Mao, Ang Sun, Mingya Wang, and Mingshi Wang. 2021. 'Pollution Effect Assessment of Industrial Activities on Potentially Toxic Metal Distribution in Windowsill Dust and Surface Soil in Central China'. *Science of the Total Environment* 759: 11. https://doi.org/10.1016/j.scitotenv.2020.144023.

Hansen, J. 1923. 'Esch et Ses Usines'. Esch-sur-Alzette: Kremer-Müller.

Hemmer, Carlo. 1953. *L'économie Du Grand-Duché de Luxembourg: Deuxième Partie – La Production Secondaire (L'industrie Sidérurgique)*. Luxembourg: Joseph Beffort.

Hoffmann, Sylvie. 1974. 'Kritik in Punkto Umweltschutz - "Elo... Um Richtege Wé...?"' *Democratie 80*, 1974, N 74 edition, sec. Eine persönliche Meinung.

Jarrige, François, and Thomas Le Roux. 2017. *La Contamination Du Monde: Une Histoire Des Pollutions à l'âge Industriel*. Paris: Le Seuil.

Knebeler, Christophe. 2011. 'Wichtige technische Beiträge der ARBED zur Stahlherstellung'. *forum* 304 (ARBED): 36–38.

Knebeler, Christophe, and Denis Scuto. 2010. *Belval: Passé, présent et avenir d'un site luxembourgeois exceptionnel (1911-2011)*. Esch-sur-Alzette: Le Phare.

Koponen, M., T. Gustafsson, K. Kalliomäki, Pirkko-Liisa Kalliomäki, M. Moilanen, and L. Pyy. 1980. 'Dusts in a Steel-Making Plant'. *International Archives of Occupational and Environmental Health* 47 (1): 35–45. https://doi.org/10.1007/BF00378326.

Kumar, Dilip, and Deepak Kumar. 2018. 'Chapter 12 - Dust Control'. In *Sustainable Management of Coal Preparation*, edited by Dilip Kumar and Deepak Kumar, 265–78. Woodhead Publishing. https://doi.org/10.1016/B978-0-12-812632-5.00012-4.

Lake, M. W. 2014. 'Trends in Archaeological Simulation'. *Journal of Archaeological Method and Theory* 21 (2): 258–87. https://doi.org/10.1007/s10816-013-9188-1.

Le Gouvernement du Grand-Duché de Luxembourg. 2016. *Règlement grand-ducal du 14 novembre 2016 concernant la protection de la sécurité et de la santé des salariés contre les risques liés à des agents chimiques sur le lieu de travail.* Vol. A235. https://legilux.public.lu/eli/etat/leg/rgd/2016/11/14/n1/jo.

Lebherz, Daniel S., Christian Zeyen, Jan Hess, Ralph Bergmann, Ingo J. Timm, Thomas Burch, Vera Hildenbrandt, and Claudine Moulin. 2018. 'Towards Scientific Workflows and Computer Simulation as a Method in Digital Humanities'. https://doi.org/10.18420/INFDH2018-03.

Leelőssy, Ádám, Ferenc Molnár, Ferenc Izsák, Ágnes Havasi, István Lagzi, and Róbert Mészáros. 2014. 'Dispersion Modeling of Air Pollutants in the Atmosphere: A Review'. *Open Geosciences* 6 (3): 257–78. https://doi.org/10.2478/s13533-012-0188-6.

'Les Composes du Fluor dans la pollution atmospherique'. 1971. Doctoral Thesis. Luxembourg. 2048. Luxembourg Archives de l'Etat.

Logelin-Simon, Armand. 2006. 'Differdange: Le Décryptage d'une Âme'. *Nos Cahiers* 27 (3–4): 81–104.

Malm, Andreas. 2016. *Fossil Capital: The Rise of Steam Power and the Roots of Global Warming*. Brooklyn: Verso.

Massard-Guilbaud, Geneviève. 2007. 'Quelle Histoire Pour l'environnement?' *Responsabilité & Environnement*, no. 48: 30–36.

'MATLAB - MathWorks'. 2022. 2022. https://www.mathworks.com/products/matlab.html.

McCarty, Willard. 2016. 'Fictions of Possibility: Simulation for the Humanities from Its History in the Technosciences'. *Journal of Siberian Federal University. Humanities & Social Sciences* 9 (7): 1553–61. https://doi.org/10.17516/1997-1370-2016-9-7-1553-1561.

Metz, M. Paul. 1972. 'Recherche et dévelopment en sidérurgie'. In *Revue Technique Luxemburgeoise 1972*, 175–91.

Millim, Anne-Marie. 2014. 'Schooling the Gaze: Industry and Nation-Building in Luxembourgish Landscapewriting, 1900–1940'. *Journal of European Studies* 44 (2): 151–69.

Mlika, Mouna, Rotimi Adigun, and Beenish S. Bhutta. 2022. 'Silicosis'. In *StatPearls*. Treasure Island (FL): StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK537341/.

Moia, Nelly. 1998. *Für Die Katz'*. s.l.: Nelly Moia.

Molitor, Leon, and Josy Barthel. 1958. 'Etude sur la pollution atmosphérique au Grand-Duché de Luxembourg'. Ministère de la Santé Publique Luxembourg. Bibliotheque Nationale Luxembourg.

Molitor, Leon, and Michel Mosinger. 1960. 'Les Poussieres industrielles en pathologie humaine et expérimentale (Silicose et Silico-Anthracose, Sidérose et Sidéro-Silicose)'. 1. Plan et Première Partie 1. Bulletin de la Société des Sciences Médicales. Luxembourg: Haute Autorité de la Communauté Européenne du Charbon et de l'Acier. Bibliotheque Nationale Luxembourg.

Mosley, Stephen. 2001. *The Chimney of the World: A History of Smoke Pollution in Victorian and Edwardian Manchester*. Cambridge: White Horse Press.

Nurul, Ainun Hamzah, Bahri Mohd Tamrin Shamsul, and Ismail Noor Hassim. 2016. 'Assessment of Dust Exposure in a Steel Plant in the Eastern Coast of Peninsular Malaysia'. *Work (Reading, Mass.)* 55 (3): 655–62. https://doi.org/10.3233/WOR-162433.

Parmentier, Isabelle. 2008. *Histoire de l'environnement En Pays de Charleroi 1730-1830: Pollution et Nuisances Dans Un Paysage En Voie d'industrialisation*. Brussels: Académie Royale de Belgique.

Parr, Joy. 2010. 'Our Bodies and Our Histories of Technology and the Environment'. In *The Illusory Boundary: Environment and Technology in History*, 26–42. Charlottesville & London: University of Virginia Press.

Pfister, Christian, and Heinz Wanner. 2021. *Climate and Society in Europe-The Last Thousand Years*. Bern: Haupt.

Pritchard, Sara B., and Carl A. Zimring. 2020. *Technology and the Environment in History*. Baltimore: Johns Hopkins University Press.

Qing, Xiao. 2015. 'Assessment of Heavy Metal Pollution and Human Health Risk in Urban Soils of Steel Industrial City (Anshan), Liaoning, Northeast China'. *Ecotoxicology and Environmental Safety*, 9.

‘Retombées de poussières: Rapport sur les mesures de retombées en poussière effectuées de 1966 à 1969 à Esch-sur-Alzette’. 1972. Esch-sur-Alzette: Institut d’Hygiène et de Santé Publique, Laboratoire des Eaux. 637.3.

Rosen, George. 1993. *A History of Public Health*. Revised Expanded Edition. Baltimore: Johns Hopkins University Press.

Schueneman, Jean J. 1963. *Air Pollution Aspects of the Iron and Steel Industry*. U. S. Public Health Service, Division of Air Pollution.

Scuto, Denis. 1993. *Industriekultur in Esch: Eine Stadtgeschichtliche Wanderung Durch Die Luxemburger Minettemetropole*. Esch-sur-Alzette: Le Phare.

Scuto, Denis. 2005. ‘Paul Flesch’. In *100 Joer Esch, 1906-2006*, 60–71. Esch-sur-Alzette: Ville d’Esch-sur-Alzette.

Stockie, John M. 2011. ‘The Mathematics of Atmospheric Dispersion Modeling’. *SIAM Review* 53 (2): 349–72. https://doi.org/10.1137/10080991X.

Su, Xianbo, Rui Ding, and Xinguo Zhuang. 2020. ‘Characteristics of Dust in Coal Mines in Central North China and Its Research Significance’. *ACS Omega* 5 (16): 9233–50. https://doi.org/10.1021/acsomega.0c00078.

Tarr, Joel A. 1996. *The Search for the Ultimate Sink: Urban Pollution in Historical Perspective*. Akron: University of Akron Press.

Tarr, Joel A. 2021. ‘Land Use and Environmental Change in the Hudson-Raritan Estuary Region, 1700-1980, With an Addendum to 2018’. In *Coastal Metropolis: Environmental Histories of Modern New York City*, 10–38. Pittsburgh: Pittsburgh University Press.

‘The Steel City and a Brief History of Dust Collection’. 2019. *Imperial Systems, Inc.* (blog). 19 June 2019. https://www.isystemsweb.com/the-steel-city-and-a-brief-history-of-dust-collection/.

Tripathi, A, C Souprayen, A Stanley, and N Warrilow. 2018. ‘Computational Fluid Dynamics or Gaussian - Is There a Right Way to Model Gas Dispersion ?’ In *Proceedings of the 10th IMA International Conference on Modelling in Industrial Maintenance and Reliability*. Institute of Mathematics and its Applications. https://doi.org/10.19124/ima.2018.001.26.

Uekötter, Frank. 2020. *Im Strudel: Eine Umweltgeschichte Der Modernen Welt*. Frankfurt/Main: Campus.

Valenti, Carla, Paolo Pozzi, Alessandra Busia, Roberto Mazza, Paolo Bossi, Cinzia De Marco, Ario Alberto Ruprecht, Alessandro Borgini, and Roberto Boffi. 2016. ‘Respiratory Illness and Air Pollution from the Steel Industry: The Case of Piquiá de Baixo, Brazil (Preliminary Report)’. *Multidisciplinary Respiratory Medicine* 11 (1): 41. https://doi.org/10.1186/s40248-016-0077-9.

Van de Maele, Jens. forthcoming. ‘“We Swallow the Dirt.” Popular and Governmental Perceptions of Metallurgical Air Pollution in Luxembourg During the Trente Glorieuses’. In *Rethinking the History of Deindustrialization: Mediating the Transformation of Industrial Cities in Postwar Europe*. Routledge.

WHO. n.d. ‘Hazard Prevention and Control in the Work Environment: Airborne Dust’. WHO Publications. Accessed 21 June 2022. https://www.who.int/publications-detail-redirect/WHO-SDE-OEH-99-14.

Wippich, Cornelia, Jörg Rissler, Dorothea Koppisch, and Dietmar Breuer. 2020. ‘Estimating Respirable Dust Exposure from Inhalable Dust Exposure’. *Annals of Work Exposures and Health* 64 (4): 430–44. https://doi.org/10.1093/annweh/wxaa016.

World Health Organization. 2021. *WHO Global Air Quality Guidelines: Particulate Matter (PM2.5 and PM10), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide*. Geneva: World Health Organization. https://apps.who.int/iris/handle/10665/345329.

IV. How can Data Visualization Support Interdisciplinary Research? LuxTime: Studying Historical Exposomics in Belval

Authors:

Dagny Aurich, Aida Horaniet Ibanez

Manuscript type:

Method paper

Contribution statement:

DA and AHI contributed to the conceptualization and the review and editing of the manuscript. **DA** and AHI were responsible for writing (original draft preparation) of all chapters. AHI gave major input on the data visualization theory and **DA** helped with sections on the interdisciplinary work and visualization examples. Figures 2, 3, 4, 5, 6, 7, 10 and 11 were co-designed by both authors and were then created by AHI based on data collected by **DA** and AHI. **DA** inserted the screenshot shown in Figure 1 and created Figure 8 and 9.

Submission status:

Published in Frontiers Big Data: Visualizing Big Culture and History Data on 29 September 2023.
(DOI: 10.3389/fdata.2023.1164885)
This manuscript will be also part of the PhD thesis by co-first author Aida Horaniet Ibanez.

Short summary/ Contribution to the field:

In this article the Luxembourg Time Machine (*LuxTIME*) *data visualization toolbox* is introduced, a set of data visualization concepts and techniques from epistemologically distant disciplines. The use of such a *toolbox* is discussed through several data visualizations co-designed by the authors from the different disciplines involved. The aim was to explore and communicate the project data and metadata and to analyze data visualization as a process in itself. One example is the *chemical stripes* visualization based on chemical patent data, which demonstrates a simple yet effective way of communicating via visualization. The *data visualization toolbox* opens new disciplinary perspectives and facilitates interdisciplinary work and could therefore serve as a starting point for other interdisciplinary projects that employ data visualization as a navigation tool.

frontiers | Frontiers in Big Data

TYPE Methods
PUBLISHED 29 September 2023
DOI 10.3389/fdata.2023.1164885

Check for updates

OPEN ACCESS

EDITED BY
Florian Windhager,
University for Continuing Education
Krems, Austria

REVIEWED BY
Georgia Panagiotidou,
King's College London, United Kingdom
Reza M. Salek,
Bruker, Germany

*CORRESPONDENCE
Dagny Aurich
✉ dagny.aurich@uni.lu
Aida Horaniet Ibañez
✉ aida.horanietibanez@uni.lu

†These authors have contributed equally to this work and share first authorship

RECEIVED 13 February 2023
ACCEPTED 04 September 2023
PUBLISHED 29 September 2023

CITATION
Aurich D and Horaniet Ibañez A (2023) How can data visualization support interdisciplinary research? LuxTIME: studying historical exposomics in Belval.
Front. Big Data 6:1164885.
doi: 10.3389/fdata.2023.1164885

How can data visualization support interdisciplinary research? LuxTIME: studying historical exposomics in Belval

Dagny Aurich[1*†] and Aida Horaniet Ibañez[2*†]

[1]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg, [2]Luxembourg Centre for Contemporary and Digital History, University of Luxembourg, Esch-sur-Alzette, Luxembourg

The Luxembourg Time Machine (LuxTIME) is an interdisciplinary project that studies the historical exposome during the industrialization of the Minett region, located in the south of Luxembourg. Exposome research encompasses all external and internal non-genetic factors influencing the health of the population, such as air pollution, green spaces, noise, work conditions, physical activity, and diet. Due to the wide scope of the interdisciplinary project, the historical study of the exposome in Belval involved the collection of quantitative and qualitative data from the National Archive of Luxembourg, various local archives (e.g., the communes of Esch-sur-Alzette and Sanem), the National Library, the Library of National Statistics STATEC, the National Geoportal of Luxembourg, scientific data from other research centers, and information from newspapers and journals digitized in eluxemburgensia.[1] The data collection and the resulting inventory were performed to create a proof of concept to critically test the potential of a multi-layered research design for the study of the historical exposome in Belval. The guiding navigation tool throughout the project was data visualization. It has facilitated the exploration of the data collected (or just the data) and the metadata. It has also been a valuable tool for mapping knowledge and defining the scope of the project. Furthermore, different data visualization techniques have helped us to reflect on the process of knowledge sharing, to understand how the relevance of certain topics changed throughout the project and why, and to learn about the publication process in different journals and the experience of the participants. Data visualization is used not only as a means to an end but also to embrace the idea of *sandcastles* using a speculative and process-oriented approach to advance knowledge within all research fields involved. LuxTIME has proven to be an ideal case study to explore the possibilities offered by different data visualization concepts and techniques resulting in a *data visualization toolbox* that could be evaluated and extended in other interdisciplinary projects.

KEYWORDS

data visualization, historical data, interdisciplinary research, exposome, digital history

1 https://eluxemburgensia.lu/de: eluxemburgensia is a collection of digitized newspapers created by the Bibliothèque nationale du Luxembourg.

1. Introduction

Interdisciplinary collaboration, especially in projects involving researchers across the natural and applied sciences and the humanities, presents many challenges, including the multiplicity of research questions, the use of different methodologies, and the varying underlying assumptions based on different epistemic cultures. Bridging these epistemological and methodological differences to establish a shared understanding requires a substantial intellectual commitment from all participants. In an ideal scenario, such collaboration has the potential to produce new research questions, foster interdisciplinary approaches for the analysis of complex issues, and generate new knowledge based on interactional expertise (Fickers et al., 2022). Data visualization can play a significant role in facilitating such interdisciplinary collaboration.

Data visualization is widely used in research for exploring quantitative data, validating hypotheses, and communicating results, primarily using statistical charts such as bar charts, line charts, or scatter plots (Glivinska, 2021). It is less frequently used to study the research process itself. It is a process through which ideas are explored and a collective discourse is constructed using data visualization to critically observe different mediations (Hinrichs et al., 2019). In this article, we reflect on some of the data visualizations designed during the Luxembourg Time Machine (LuxTIME) project discussed in section 2. Based on the experience obtained during 2.5 years of interdisciplinary collaboration, working with a team of historians, hydrologists, chemists, and data visualization experts, we want to demonstrate that interdisciplinary research can benefit from an "extended" *data visualization toolbox*. We refer to a *toolbox* as a set of data visualization concepts and techniques, and we want to "extend" it, as opposed to only applying techniques frequently used within each specific field.

Specific disciplines can develop certain data visualization conventions. Considering visualization in interdisciplinary projects can help break out of these conventions and create new insights. Statistical graphs are used mostly in the natural, social, and applied sciences with the objective of exploring data to discover trends, patterns, and outliers, and to validate hypotheses or communicate results quickly for decision-making. Visualizations that emphasize esthetics and use metaphors and non-standard visual vocabularies are often found in journalistic, educational, and artistic contexts, while the study of interpretive practices and non-representative approaches predominate in the humanities. Each discipline can benefit from different practices. This becomes even more evident in the case of interdisciplinary research.

Combining these concepts and techniques predominant in different disciplines and applying them in the same context, we developed the *LuxTIME data visualization toolbox* exploring standard statistical graphs and variations thereof, concept maps, visual rhetoric, data humanism, multivariate data glyphs, non-representational approaches and visual elements of interpretation, and data storytelling. The use of different types of visualizations helped to map and exchange knowledge, thereby defining the project scope and the contributions of different disciplines, track timelines, and project deliverables. It also framed the participants' experiences along the way, inviting them to self-reflect on changes throughout the project and to explore the iterative research process. The *toolbox* aims to inspire a wide range of projects, especially those involving different disciplines. It describes and discusses the application of a set of epistemologically distant techniques and concepts from which to begin the exploration, and then add or remove tools, adapting the *toolbox* to each project. Foremost, it extends an invitation to explore and integrate other ways to visualize data, diverging from the traditional techniques commonly employed in each respective discipline.

In the current article, we start with a description of the LuxTIME project. Next, we discuss the methods, including the data collection process, the participants and their roles, the research questions addressed in the interdisciplinary context, the visualization concepts and techniques included in our project, and their application in the results section. We finish with a discussion of the frictions experienced as part of working on an interdisciplinary project with data visualization, how the idea of data visualization as a *sandcastle* (Hinrichs et al., 2019) has been present throughout the project, and finally, how other projects can benefit from our experience.

2. Case study

The Luxembourg Time Machine Project (LuxTIME) is an interdisciplinary research project funded by the Institute for Advanced Studies (IAS) of the University of Luxembourg.[2] Three research institutes engage in this so-called "Audacity Project", namely, the Center for Contemporary and Digital History (C^2DH), the Luxembourg Center for Systems Biomedicine (LCSB), and the Luxembourg Institute of Science and Technology (LIST). The main objective is to explore the potential implementation of a national platform ("Luxembourg Time Machine") that would allow scientists and stakeholders to "dive" into the complex past of this country using digital tools and data from different disciplines and fields. It is a subsidiary project of the European Time Machine project[3] adding a new dimension to the past.

By building a digital dataset including information from three very different fields and scientific perspectives, namely, eco-hydrology, environmental cheminformatics, and history, LuxTIME is using a local case (the industrialization of Belval and the Minett region) as a testbed for methodological and epistemological reflections on how to study the impact of environmental changes on the health of the local population, with a regional and long-term perspective. "Contextual information" based on archival evidence is mixed with "scientific evidence" derived from chemical, biological, or medical investigations as the project explores new grounds in interpreting "big data of the past" in a truly interdisciplinary setting. The Belval case study is a pilot project, preceding a project at the national level, the LuxTIME INITIATE.[4]

2 https://luxtimemachine.uni.lu/

3 https://www.timemachine.eu/

4 LuxTIME INITIATE aims at building a consortium with the main stakeholders of historical data in Luxembourg (archives, libraries, museums, statistical offices, research institutions, governmental bodies, and private associations) to study the past in an interdisciplinary inter-institutional setting, with both intellectual and technical impact.

The above-mentioned impact of environmental changes on human health is covered by the exposome concept (Miller, 2020). Global and local changes have severe impacts on environmental systems and their inhabitants (Karlsson et al., 2021). The analysis of those changes and their impacts is a challenging task that can, however, be of help to understand and therefore prevent potential future outcomes. This could be, for example, any case of environmental pollution happening in an area resulting in an increase in disease cases as a phenotypic response. Human phenotypes, being a set of observable characteristics or traits, are mainly influenced by two factors and their interactions: the genetic factors described by the *genome* and all non-genetic factors covered by the *exposome* concept. Measuring the genome is difficult, but it is limited to the combination of four nucleotides that are stable over time. The *exposome* of an individual, however, is the measure of environmental influences (e.g., lifestyle, diet, and behavior) and the associated biological responses (Miller and Jones, 2014), changes within the course of a lifetime, and historical developments. Consequently, the *exposome* is influenced by many factors that vary over time and influence each other. Interdisciplinary efforts and data sources are required to come as close as possible to covering the whole picture.

The focus of this article was on how the use of a variety of data visualization concepts and techniques has supported our interdisciplinary (exposome) research.

3. Methods

3.1. Data collection

In this article, we refer to *data* as any collection of values conveying information, whether to represent abstract ideas (e.g., knowledge exchange and relevance of a topic), specific measurements (e.g., concentration of chemicals and number of articles published), or statistics (e.g., census data and steel production). We collected quantitative and qualitative data, mostly from secondary data sources (e.g., statistical archives, existing research), but also included some primary sources (e.g., observations from historical sources and reflections on the process itself). The collection for the LuxTIME project was done through observation, measurement, simulation, and analysis.

The Minett region is known for its past in the iron and steel industry, which was accompanied by many regional changes in terms of environment, socioeconomy, and health (Knebeler and Scuto, 2010). Some of the initial questions of the LuxTIME members included the following: where to look for data (*e.g.*, local and national archives); where to set the geographical and time limits; which topics to focus on (*e.g.*, environmental pollution); and how to obtain scientific data. For example, in the case of performing new chemical analyses, which sample types could be used? The initial objective was to find links between environmental pollution and other influencing factors of the past and disease patterns in the population by looking at archival sources combined with information received by scientific or governmental institutes and current chemical analyses, revealing facts about past exposures. This includes information related to the *historical exposome* in Belval such as datasets, images, text, events, and maps. For each dataset, relevant metadata was also collected, including title, source, reference from the source if available, author, publication date, the team member who collected the information, language(s), description, number of files, digitization status, class, format, type, period covered, geographical area covered, access rights, and the categories and subcategories of the *exposome* covered. Furthermore, we collected information about the process, such as the relevance of the different topics, disciplines involved, period of validity and reasons for change, relation to other disciplines, the amount of information found, and its potential for further analysis. We also collected data about the project deliverables such as the number of publications, the type, the disciplines involved, the different steps in the process (e.g., work starts, submission, and publication), the time, and the experience.

All these quantitative and qualitative data allowed us to study the research questions stated in section 3.3, using two data collection methods. First, a structured and normalized *data inventory* was created, where we included all the pieces of information found through the sources. To date, this table contains 121 records from 17 different information sources, each of them registered with the metadata described above. Second, for the evolution, reflection on the process, and experience, we used the visualization directly to generate the data (that could be extracted later if necessary).

The first steps included contacting governmental and scientific institutes to access past data already collected in the area (*e.g.*, scientific measurements of parameters such as soil or air pollution resulting from industry or other anthropogenic sources). Moreover, measures taken by the industry or the government to enhance life quality and health were investigated. Historical data—not only about environmental pollution but also social and economic data—were retrieved from archives. Newspaper articles, scientific reports and books, pictures, and contemporary witness reports were included. For chemical measurements generating scientific data, sampling campaigns of surface and groundwater were discussed as well as looking at dust samples and soil or biological samples such as mussels, trees, or even teeth. Based on the research on this topic from a chemistry point of view, a review article was published in 2021 (Aurich et al., 2021), helping to plan further project work packages in terms of, e.g., sampling campaigns. One example is the discussion of analyzing human samples to get to know more about past steel-pollution exposure and its health effects. However, the sample bank located in Luxembourg does not provide samples dating back to the steel industry times as it is a fairly new facility. The outcomes of the review showed many possibilities for how to access the historical *exposome* in terms of data and chemical analyses; however, none of them was available or feasible within the project scope and timeline.

3.2. Participants and roles

The core of the research described in this article was conducted during the collaboration between a researcher in data visualization and a researcher in environmental cheminformatics, with occasional feedback and participation from the overall project team that included other researchers in environmental cheminformatics, eco-hydrology, and history. The researcher in environmental cheminformatics already had disciplinary knowledge of data visualization and functioned as a domain expert in the field of the *exposome*, together with the rest of the environmental cheminformatics team. The data visualization researcher collected and studied many techniques and concepts used across different disciplines and then proposed an initial *toolbox* to discuss and experiment with during the visualization sessions are described in section 3.5. The researcher in data visualization also participated in the work of the target domain, *historical exposomics*, and the environmental cheminformatics researcher participated in visualization research. The collaboration was based on a *design-by-immersion* approach with reciprocal immersion, where both the visualization researcher and domain expert engaged with and participated in the work of the other domain, and knowledge emerged from the experiences and interactions (Hall et al., 2020). They not only participated in each other's research approaches and practices but also did archival work, collecting and analyzing historical sources. As discussed in section 5, this approach shapes and enriches the research team and also changes both participants' perspectives on their own fields.

3.3. Research questions

With data visualization as a navigation tool, our goal is to analyze the research questions outlined below.

Question #1: How could we map knowledge and knowledge sharing to define the project scope? The first question aims at exploring the *knowledge gap*, resulting from the different backgrounds and expertise in the project, e.g., history, design, and chemistry (Van Wijk, 2006). To create a "trading zone", i.e., "a space for interactions and negotiations between different knowledge domains" (Fickers and van der Heijden, 2020), we first needed to understand the knowledge within the group and what needed to be shared to define the project scope. In this research question, we explored how to arrive at an initial space, which exists at the boundary of four disciplines in which research can begin; a space in which diverse voices can speak and be heard and differences can be examined to mutually validate diverse perspectives, creating opportunities for mutual learning (Mao et al., 2019).

Question #2: How did the project evolve in terms of scope, relevance of the topics, the information available, and blending of the disciplines in each of these topics? The scope of the project was not permanent, nor was the knowledge or interest of the different disciplines in the different topics, which blended with different intensities throughout the process; as the project progressed, new sources of information appeared, and others were discarded. In this research question, we were interested in the evolution of how, from a series of central themes defined in the previous question, the scope changed throughout the project: which topics gained importance, which ones appeared or disappeared, and why.

Question #3: How could we explore the data and metadata respecting the priorities of the different disciplines? To synthesize disparate datasets in a visualization, especially in a project involving epistemologically distant disciplines, *data frictions* emerge regarding discipline-specific interpretation of the data, methodological approaches, ways of handling uncertainty, and scale and granularity in the datasets (Panagiotidou et al., 2022). Without analyzing in detail the different causes of such frictions, in this research question, we explored how the use of visualization techniques from other disciplines reveals the different priorities, and how through co-construction and exchange throughout the process, a common space of data and metadata visualization can be reached.

Question #4: How could we monitor the project deliverables, including the different steps of the process, the contribution of the different disciplines, and the experience of the participants? In addition to the *knowledge gap* discussed in the first research question, there is often an *interest gap*, caused by the different aims of the researchers, e.g., participation in different types of conferences and publication requirements (Van Wijk, 2006). These differences not only have an impact on the deliverables and, therefore, on the project timeline but also on the experience of the participants. In this research question, we explored the representation of the different steps in the process in time vs. in the experienced temporality.

We have explored each of these questions through visual means. In section 3.4, we introduce the concepts and techniques that we have included in our *toolbox* throughout the project followed by a discussion of the visualizations in which they are applied, in section 4.

3.4. The LuxTIME data visualization toolbox

We refer to *data visualization* as the graphical representation of data, using a variety of visual encoding methods (a representational approach), as well as the use of a graphical representation, to model interpretation and generate or augment data (a non-representational approach). The applications vary (e.g., exploratory analysis, data validation, hypothesis validation, and communication), and the techniques used depend on the intended purpose (e.g., quick decision-making vs. in-depth exploration of multiple narratives). Data visualization (or Dataviz) encompasses other terms including information visualization (InfoVis), information design, scientific visualization (SciVis), information graphics (Infographics), statistical graphics, or exploratory data analysis. The difference between these terms has been largely discussed in previous research (Rhyne et al., 2003; Manovich, 2011; Lankow et al., 2012; Kim et al., 2016).

In this section, we discuss several data visualization concepts and techniques that have been fundamental to our project and, therefore, essential elements of our *data visualization toolbox*.[5] The selection of "tools" for our *toolbox* is based on an extensive literature review and the study of numerous data visualization examples from different disciplines. It aims at integrating the perspectives on the field of the different disciplines (discussed below in relation to each concept) and above all

5 The theory behind the concepts and techniques mentioned below might overlap but each of them has in some way proven to be useful during our project, and therefore, we considered it worthwhile including them all in our *toolbox*.

to experiment with concepts and techniques originating from epistemologically distant disciplines, which are rarely applied in the same context. The *toolbox* is built with an inclusive approach, integrating visualizations frequently used across disciplines (e.g., statistical charts) and potential variations, which also led to rich conversations and outcomes.

Some of these elements may be useful for other interdisciplinary projects; therefore, we present below a brief description and discussion of each of them, including examples from other projects. These methods and techniques are then applied in section 4 to answer our research questions.

3.4.1. Statistical graphs

The most frequently used data visualization techniques across disciplines are all types of statistical charts, such as bar charts to compare magnitudes, line charts to show evolution in time, or scatter plots to analyze relationships (Glivinska, 2021). The main purpose of these graphs is to summarize, validate, and communicate a message effectively. The use of statistical graphs assumes that the audience is data literate, i.e., has some basic knowledge of descriptive statistics; and that the designer[6] is aware of the extensive existing research on the subject.

Statistical graphs have a long history, and there is extensive literature about how to use them correctly to explore and summarize data effectively. However, academic and non-academic data visualization practitioners from different disciplines often fail to apply these theoretical principles. One of the frequent pitfalls is the wrong selection of charts despite the literature about graphical perception and suitability for analytical purposes (Lockwood, 1969; Cleveland and McGill, 1984, 1986; Evergreen, 2017). Other errors include the inappropriate use of colors based on color perception or cultural differences (Rogowitz and Treinish, 1998; Ware, 2004; Silva et al., 2007, 2011), the use of misleading graphs (Cairo, 2019), or poor storytelling (Nussbaumer Knaflic, 2015; Dykes, 2020). Moreover, the frequent lack of polish on elements such as axes, labels, gridlines, annotations, legends, descriptions, and titles (Schwabish, 2021) can create distractions from the core message (Tufte, 1999). These are just a few examples among many other considerations to be examined when creating statistical graphs (e.g., layout, context, transparency, accessibility, and interactivity).

Applied cases of statistical graphics to the LuxTIME are demonstrated in sections 4.3.1 and 4.3.2.

3.4.2. Variations of statistical graphs

We refer to the variation of a statistical graph when, knowing the theory mentioned in the preceding section, the designer decides not to apply one or more of these rules deliberately, e.g., to meet a specific use case requirement. Examples of such variations include duplicating the encoding (i.e., overencoding) to highlight a particular aspect of the graph (e.g., position and color encode the same information), changing the orientation, or overlapping graphs to favor a particular visual effect (e.g., crowdedness). Other options include removing axes, measures, gridlines, or titles to focus the audience's attention on the visualization; replacing predefined geometric shapes with other elements with a rhetorical value (e.g., using the representation of an object instead of rectangles in a bar chart); or any other type of visualization that starts from a statistical graph and modifies it with a purpose, beyond the one initially established for such a graph. It is often a combination of several modifications. Such variations of the conventional statistical graphs open numerous possibilities for the designer, especially in terms of communication.

The *climate stripes* by Ed Hawkins are a variation of a statistical graph, showing a simplified heatmap that has a strong influence on the climate change debate around the world. ShowYourStripes.info[7] registered 89,000 unique visitors to the site worldwide for 30 days during the summer of 2022 (Santoro and Kirkland, 2022). Hawkins emphasized the need for a range of ways of communicating the "climate crisis" because different people learn and experience in different ways and, therefore, justified the need for a range of climate visualizations to talk to different audiences. Another example is the visualization of China's overseas investments by Alberto Lucas López and Cédric Sam[8] where the bars of a bar chart are replaced by semicircles whose area represents the value of the deals and is displayed for many countries, overlapping in the layout. Laura Bronner, Anna Wiederkehr, and Nathaniel Rakich visualized the election night in 2020 *What Blue And Red "Shifts" Looked Like In Every State*[9] using simplified area charts in a tile map of the United States for FiveThirtyEight.[10] Kim Albrecht explores the randomness of success in scientific publications[11] using overlapping timelines in which only the maximums are highlighted. Another visualization example by Alberto Lucas López in collaboration with Ryan Williams and Kaya Berne is *Migration waves,*[12] where variations of area charts are used, through minimalism (no grids, no axes, and no measures), color (to emphasize positive and negative currents), and the superimposition of the graphs for the different countries.

3.4.3. Concept maps

Concept mapping is a visualization technique that uses hierarchical networks of nodes (concepts) and links (relationships) to represent visual knowledge (Romance and Vitale, 1999). The use of a concept map allows the inclusion of cross-links to map the links between concepts in different domains and represent creative leaps in knowledge production (Novak and Cañas, 2008). Joseph D. Novak and Alberto J. Cañas identified two features of concept maps that are important in the facilitation of creative thinking: the hierarchical structure and the ability to search for and characterize

6 Designer refers to the person who designs the visualization (i.e., statistical charts), independent of their main area of expertise.

7 https://showyourstripes.info/s/globe

8 https://multimedia.scmp.com/china-overseas-investments/

9 https://fivethirtyeight.com/features/where-we-saw-red-and-blue-mirages-on-election-night/

10 https://fivethirtyeight.com

11 http://sciencepaths.kimalbrecht.com/

12 https://www.nationalgeographic.com/magazine/graphics/graphic-shows-past-50-years-of-global-human-migration?sf215829698=1&sf217104276=1

new cross-links. In addition to the purpose of enhancing creativity, concept maps are used as a design tool to generate structural organization, for communication purposes to overcome the limitations of the linear nature of the text, to stimulate the learning process by making the interrelationships between concepts explicit, and as an assessment tool to identify misconceptions (Lanzing, 1998). Johannes Wheeldon and Jacqueline Faubert argued that traditional definitions of concept mapping should be expanded to include more flexible approaches to the collection of graphic representations of experience, using concept maps to gather qualitative data from research participants (Wheeldon and Faubert, 2009).

In an interdisciplinary context where mapping knowledge of the different stakeholders is key to understanding the possibilities and developing the project, concept maps are a data visualization technique that helps to navigate the complexity, in terms of the variety of topics and subtopics and how they interrelate (see applied for LuxTIME in section 4.1).

3.4.4. Visual rhetoric

Rhetoric is the study of the communication techniques used to inform, persuade, or motivate a given audience in a particular situation, modifying their conceptions and attitudes toward the object of communication. Visual rhetoric (as an artifact) is the purposeful production or arrangement of colors, forms, and other elements to communicate with an audience (Hill and Helmers, 2009). One case of visual rhetoric in the context of data visualization is mapping rhetoric, which refers to "manipulating the information presentation via the data-to-visual transfer function, the constraints that determine how a piece of information will be translated to a visual feature" (Hullman and Diakopoulos, 2011).

A metaphor is a rhetorical figure, which refers to the cognitive process humans engage in when they reconceptualize a concept from a target domain in terms of another; and, therefore, when visual language is used to perform these functions, it is a visual metaphor (Steen, 2018). Despite the predominance of a minimalist data visualization approach that favors high data-ink ratios,[13] the data visualizations that use visual rhetoric—to engage, communicate, and be memorable using visual metaphors, and elements of embellishment—are still very present, especially in the fields of journalism, information design, or data art. Often these types of visualizations use organic forms, associated with nature, such as plants or rocks. Lima (2014) devoted an entire book to the study of trees in visualization, where he analyzes how throughout history, their trunks, branches, leaves, and fruits have served to represent connections between entities, through different domains of knowledge (e.g., family trees, systems of law, and biological species like trees, see application to LuxTIME in section 4.2).

An example of mapping rhetoric is *What's Cookin?*[14] by Sarah Emery Clark, where she used the metaphor of preparing a meal to explore and analyze the state of the data visualization industry. Other examples of visual rhetoric include the visualization *Apparel Exports to the US* by Liz Bravo, where she visualized trends in the clothing industry using area charts shaped as sewing patterns; One *Angry Bird* by Periscopic,[15] displaying emotional arcs of the past 10 U.S. presidential inaugural addresses; or *The Great War* by Valentina D'Efilippo,[16] visualizing the fatalities during World War I as a poppy field. In *A View on Despair,*[17] Sonja Kuijpers visualized suicide in the Netherlands in 2017, representing the different categories in a landscape.

3.4.5. Data humanism

In *Data Humanism: The Revolutionary Future of Data Visualization,* Giorgia Lupi advocated for the connection of numbers to knowledge, behaviors, and people as data represent real life; making data unique, contextual, and intimate (Lupi, 2017). To do so, she promotes embracing a certain level of visual complexity, i.e., high-density data visualizations containing multiple attributes; and moving beyond standards, away from conventional graphics, and out-of-the-box solutions, to expand the "data-drawing vocabulary". She stated that "data is a tool that filters reality in a highly subjective way" and, therefore, it is important to reclaim a personal approach to how data are captured, analyzed, and displayed, as data are imperfect. She urged a paradigm shift to "always sneak context in", in which data visualization embraces imperfection and approximation, "allowing ways to use data to feel more empathetic, to connect with ourselves and others at a deeper level".

In their book *Data Feminism*, Catherine D'Ignazio and Lauren F. Klein stated that "refusing to acknowledge context is a way to assert authoritativeness and mastery without being required to address the complexity of what the data actually represent" (D'Ignazio and Klein, 2020). Based on the concept of "situated knowledge", initially raised by Donna Haraway in the 1980s, they stated that the responsibility of ensuring that the situatedness of data is considered is with the person evaluating the knowledge or building upon it. Yanni A. Loukissas referred to errors in data collection as "signifiers taken out of their original interpretative texts" (Loukissas, 2019). Hannah Schwan, Jonas Arndt, and Marian Dörk (Schwan et al., 2022) identified key aspects of disclosure, i.e., the aspiration to be conscious of the potential effects of the designer's assumptions. They invited "the viewer into exchanges with the designer, reflections about the visualization, and engagement with an issue" (Dörk et al., 2013) and proposed several representation forms to integrate the disclosure information into the visualizations.

Some examples of visualizations that use information-rich designs and custom visual vocabularies, and highlight the relevance of details and the imperfection of the data include *Data Items:*

13 The data-ink ratio is a concept introduced by Tufte (1999) as the proportion of ink that is used to present actual data compared to the total amount of ink (or pixels) used in the entire display.

14 https://www.sarahemeryclark.com/work/whats-cookin

15 https://emotions.periscopic.com/inauguration/

16 http://poppyfield.org

17 http://www.studioterp.nl/a-view-on-despair-a-datavisualization-project-by-studio-terp/

A Fashion Landscape,[18] visualizing the role of fashion connecting people and cultures, and *Bruises—The Data We Don't See,*[19] depicting a sensorial picture of a personal journey with a disease, both by Giorgia Lupi. In *Trending seeds,*[20] Valentina d'Efilippo and Lucia Kocincova analyzed and visualized the Twitter social movement #MeToo to understand if social media could become a vehicle to foster social change and reshape traditional views. In *Data Selfi,*[21] Kadambari Komandur explored intersectional feminism. These are just a few examples among many others. Moreover, the concept of data humanism has been discussed in research articles over the last few years. Kim et al. (2018) presented an interface that enables designing and personalizing visual vocabulary to represent data, and they explored how to enable people to determine the representation of their data based on the *Dear Data* project (Kim et al., 2019).[22] Cordell advocated for exploratory, iterative, and dialogic data humanism to foster humanistic engagement with data in an academic context (Cordell, 2019). This concept is applied in several visualizations in section 4.

3.4.6. Multivariate data glyphs

A data glyph is a visual representation of data where the attributes of the graphical entity are defined by the attributes of the data record. It is a visualization technique often used for multivariate data because patterns involving more than two dimensions can often be perceived more easily in this manner (Ward, 2008). Multivariate or multidimensional data consist of a list of records, with multiple columns (variables), which may be either numerical or categorical values. The encoding can map one data attribute to one single graphical attribute; or use redundant mappings (e.g., using tone and shape to display the same variable) to facilitate the interpretation or reinforce a message. Small multiple data visualizations contain several small graphs arranged on a grid, where every representation follows the same structure; and leverages visual constancy, economy of perception, and uninterrupted visual reasoning (Chuah and Eick, 1998). Small multiples can be based on conventional graphs, but also on information-rich glyphs that encode data attributes using customized visual vocabularies. Glyphs can be displayed in a layout based on data variables, data structure (e.g., time and hierarchies), or any other layout (e.g., predefined shape and screen size). The use of glyphs allows the designer to define the level of aggregation, where each glyph is the level of detail selected. They are often used to explore details (e.g., people, objects, and cases) as they allow us to visualize multiple characteristics about each subject (see applied in section 4.3.1).

Take a walk down Fifth Avenue[23] by Molly Morgan uses data glyphs to represent the physical characteristics and the ecological benefits of every tree along Fifth Avenue in New York. Alberto Lucas López in *Our daily faces*[24] depicts every page of the South China Morning Post newspaper over a year, encoding the subjects covered or the length in small glyphs. *Representation of women in politics*[25] by Frederica Fragapane visualizes the top 40 countries in the world by political parity score, using a combination of graphical attributes to represent multiple variables for each country, such as the percentage of seats held by women in local government bodies, in lower and upper houses of national legislatures; the number of female candidates in the most recent elections; the number of elected or appointed heads of state or the geographical area.

3.4.7. Non-representational approaches and interpretation

Inspired by the work of Thrift (2008), Johanna Drucker defined non-representational approaches to modeling interpretation in a graphical environment "as the use of graphical means as a primary method of modeling human-authored interpretation rather than to display preexisting data sets" (Drucker, 2018). In contrast to representational approaches, the existence of data or other representations is not assumed before the interpretative work; the relationship between data and visualization is not unidirectional. Visualization can be the starting point, where we add, for example, connections or annotate reflections and use any visual vocabulary to encode high-level concepts (e.g., contradiction and comparison). This could be later captured as data. As already discussed in previous sections, situated, experiential, and embodied forms of knowledge have been largely researched (in contrast to observer-independent empirical approaches). "A humanistic approach is centered in the experiential, subjective conditions of interpretation" and, therefore, it requires a shift toward non-standard metrics, where the challenge is to design graphical expressions that display interpreted phenomena (Drucker, 2011). Time, for example, is modeled not to imitate its physical dimension but to provide a model that reflects the phenomena under consideration to support a given set of analyses (Aigner et al., 2011). Conventional approaches to timelines are linear, unidirectional, continuous, and structured with a single standard metric unit because they take their structure from the temporal models used in the natural sciences. In humanities, temporality is experienced as asynchronous, variable, broken, and heterogeneous (Drucker, 2021).

There are several examples of data visualizations being used as the starting point to collect and visualize data, for example, in participatory projects using street data walls such as the *Mood Test*[26] by Domestic Data Streamers to analyze people's attitudes toward life; or in data physicalization projects such as the *Data Badges* (Panagiotidou et al., 2020) that invited participants of a conference to make their own customized expressions of their academic profiles. However, examples of interpretive exercises through visualization where graphical attributes are designed to

18 http://giorgialupi.com/data-items-a-fashion-landscape-at-the-museum-of-modern-art

19 http://giorgialupi.com/bruises-the-data-we-dont-see

20 http://metoomentum.com/trending.html

21 https://kadambarik.myportfolio.com/data-selfi

22 http://www.dear-data.com

23 https://www.5avestreettrees.com

24 https://www.lucasinfografia.com/Front-pages-analysis

25 https://www.behance.net/gallery/138862771/Women-in-politics?tracking_source=search_projects%7Cinformation%20visualization%20data%20art

26 https://domesticstreamers.com/projects/the-mood-test/

show subjective conditions of interpretation are practically non-existent, beyond the theoretical study presented above and some tool prototypes such as the 3DH project.[27]

When we want to approach analysis from a more humanistic perspective, for example, to analyze an experience, the use of a non-representational approach to data visualization using standard and non-standard metrics, as required, can facilitate the interpretative exercise. In the case of the LuxTIME project, we use this technique to reflect on the evolution of the project's topics in section 4.2 and to understand the participants' experience in section 4.4.

3.4.8. Data storytelling

Data storytelling is a powerful mechanism for sharing insights that involve data, narrative, and visuals to explain, i.e., narrative couples with data, explain; visual couples with data, enlighten; and narrative coupled with visuals, engage (Dykes, 2020). Storytelling makes data interesting, facilitates the understanding of complex subjects, encourages action, and is memorable (Vora, 2019). Brent Dykes identifies nine main tasks as part of the storytelling process: identifying key insights, being aware of biases, having extensive background knowledge, understanding the audience, curating the information, assembling the story, providing narration, choosing the visuals, and adding credibility (Dykes, 2020). Edward Segel and Jeffrey Heer developed a framework of design strategies for narrative visualization in the context of journalistic storytelling, where they placed the visualizations along a spectrum of *author-driven,* i.e., linear structure, heavy messaging, and no interactivity, and *reader-driven* approaches, i.e., highly interactive with no clear path to the story (Segel and Heer, 2010). "Narrative information visualizations rely on rhetorical techniques, to convey a story to users as well as exploratory, dialectic strategies aimed at providing the user with control over the insights gained from interaction." (Hullman and Diakopoulos, 2011).

Framing Luxembourg,[28] a timeline tracing the history of public statistics in Luxembourg, is an example of data storytelling, where the narrative is divided into different chapters (e.g., migration, family, and employment), and *scrollytelling*[29] is used to facilitate the navigation through text, images, and charts.

3.5. Data visualization and feedback sessions

The visualizations presented in section 4 have been co-created by the two main participants, integrating the feedback from the other participants. The work has been developed over more than 2 years of the project, during monthly work sessions. The different types of sessions were not initially defined but were designed during the monthly meetings as the project progressed. We have identified retrospectively 5 types of sessions that took place during the project and have named them to facilitate their discussion and future application.

- *Individual and collective preparation sessions*, where the participants reviewed the data available to date and validated the initial research questions (or reformulated them based on available information). Major advances in the data collection process were discussed in a meeting with the entire team, followed by individual preparation sessions and a final group session to agree on the data and the variables to be used and, in some cases, to reformulate the research questions.
- *Data visualization toolbox discovery sessions*: The data visualization researcher introduced the less-known concepts and techniques during these sessions. Additionally, possible improvements or alternatives were discussed when the *session* was used to review visualizations already created during the *sketching* and *pilot sessions* (e.g., the use of statistical graphs).
- *Sketching sessions*: During these sessions, both participants experimented with different visualization ideas, using R, Python, Excel, or Tableau for more standard charts, and paper and markers to design tailor-made visual vocabularies. Miro was also used for concept maps and brainstorming.
- *Pilot session*: In these sessions, the selected ideas were refined and completed. The result was the data visualizations that were ready to share with the rest of the team. The tools used remain the same as for standard charts, and Adobe Illustrator was used for static visualizations with custom visual vocabularies.
- *Feedback sessions*: Feedback from the rest of the team was collected during these sessions. These exchanges also brought to light the friction between the different disciplines, as we will discuss in section 5.

All these sessions contributed to the creation and application of our *data visualization toolbox*. The types of data available and the research questions that were developed during the *individual and collective preparation sessions* were the basis for the search for suitable visualization techniques. Once these techniques had been collected by the researcher in data visualization, the *data visualization toolbox discovery sessions* allowed the presentation and discussion of these techniques with other researchers. During the *sketching* and *pilot* sessions, the selected tools were applied in a practical way to the project. Finally, the results were discussed during the *feedback sessions* with the entire team. These steps were part of an iterative process that moved back and forth as new data became available, new techniques were added to the toolbox, or new feedback was received.

4. Results

4.1. Working in an interdisciplinary team

Working as an interdisciplinary team and learning how to exchange ideas, data, and experiences were at the core of this research. In this section, we discuss how, with the support of data visualization, we laid the foundations for dialog and defined the main themes of the project and the role of the different disciplines.

27 https://threedh.net

28 https://www.framingluxembourg.lu, developed by the C2DH in collaboration with STATEC.

29 *Scrollytelling*, from "scrolling" and "storytelling", is a way to display content that unfolds as the user scrolls.

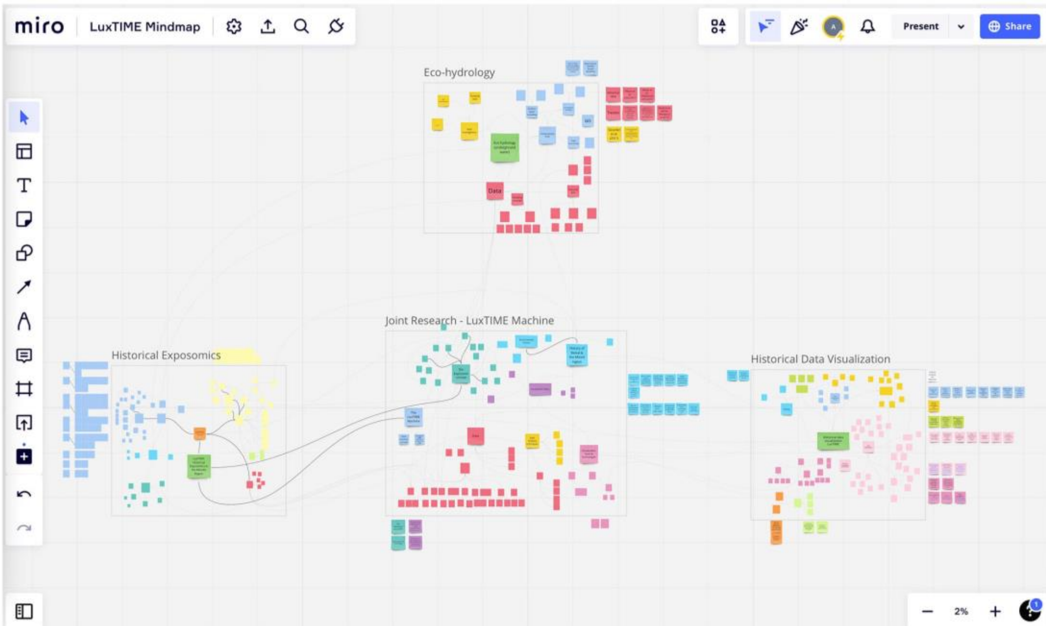


FIGURE 1
LuxTIME project concept map developed using the Miro online visual collaboration platform.

We, therefore, focus in this section on research *question #1: How could we map knowledge and knowledge sharing to define the project scope?*

Our first interdisciplinary challenge was to map the knowledge in the team to understand the skills of each team member and how they could contribute to the project. The initial team consisted of three PhD students, three supervisors, and five supporting researchers from the different research centers involved. The foundation of any interdisciplinary project is the interest in understanding other disciplines and finding out what level of learning and collaboration is required to produce new "joint knowledge" [*trading zone* concept, see (Kemman, 2021)].

To support this process, we used a *concept map*, where we mapped the knowledge of the different disciplines and the "joint knowledge" and how they related to each other. The concept map was used to add literature about the different topics, as a starting point for the other participants to familiarize themselves with the topics, and to collect the references to the first datasets. This visualization also served as both an internal and external communication tool. The interactive online platform Miro[30] was used for this purpose. Figure 1 shows an example of this collaborative platform, which was also used in interactive workshops such as the *UniTalks LuxTime Machine: Back to the future,* where the participants (e.g., researchers from other areas, librarians, and other university stakeholders) were encouraged to complete the concept map by adding terms and links, after reading a list of guiding questions such as "what exposures would you consider understanding the *exposome* of the population in the Minett region over the last 200 years?" and "where would you look for information?". The result of the workshop was an enriched version of the concept map that helped us to identify new ideas and points of view.

The final version of the concept map is shown in Figure 2. In the initial versions, we only mapped three disciplines: history, environmental cheminformatics, and eco-hydrology, displayed in different colors; and the interdisciplinary project "overlap", in gray. In later versions, we decided to map *data visualization* as a fourth discipline, since it is not only a tool that helps to achieve a technical task but a branch of knowledge that also contributes at a theoretical level. Certain topics moved from outside "specific knowledge" to inside "joint knowledge", such as "the exposome", which originated in the domain of environmental cheminformatics but became the central focus of the project; or the industrial history of the Minett region, which initiated exclusively under the expertise of the historians and evolved to become a major area of joint knowledge.

4.2. Analyzing the process: topics, information available, and participation

As discussed above, the project involves several disciplines, with a broad area of joint overlap. Once the contributions of the

30 https://miro.com

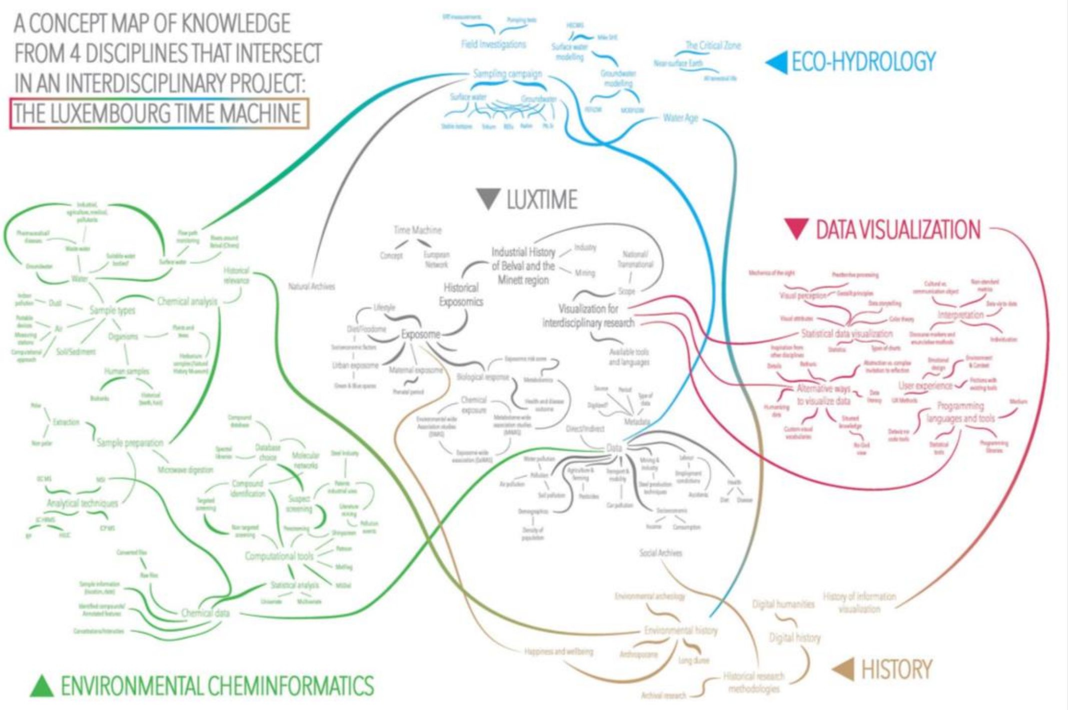


FIGURE 2
Final LuxTIME project concept map.

participants were understood, the project scope was defined and revised throughout the project. In this section, we reflected on the process: which topics were the most relevant at each moment and why, which disciplines participated in different topics, which topics had sufficient information for further development, and which ones fell out of the project scope. Thus, here, we focus on research *question #2: How did the project evolve in terms of scope, relevance of the topics, available information, and blending of the different disciplines in each of these topics?*

The concept map allowed us to see independent snapshots of the topics at different points in time in the project, but it did not allow us to compare their evolution throughout the project in the same view, in terms of the appearance and disappearance of certain subjects, to what extent they had been integrated into the different disciplines, how relevant they were considered at each time, and what were the reasons behind these changes. The data to do this analysis were not collected during the project in a structured way, since, in this case, we wanted to use the visualization directly as a reflection tool (which could generate data later, if necessary).

The result of this exercise was the visualization in Figure 3, where we created a timeline with three checkpoints (January 2021, January 2022, and January 2023), around which we randomized the themes. We included a separation line "inside/outside" of the project, which allowed us to visualize which topics and around what time had been excluded. Each circle represents a topic, and the size (3 levels) indicates how important it was considered at a certain point of time in the project. The circles have a black border at the last check point before being excluded from the project scope. The color represents discipline (this color is kept consistent throughout the different visualizations in which the discipline variable is displayed). In addition, we added annotations on the lines linking topics over time to explain how a topic is integrated into different disciplines, why a topic is excluded, or how it is merged into a different topic.

In the visualization, we observed that at the beginning of the project, there were many different discipline-specific topics with lower relevance, except for the data and the tools that were common to all disciplines from the beginning. In the second checkpoint, topics such as the field investigation integrated both eco-hydrology and environmental cheminformatics with increased relevance, but then, it was excluded from the scope due to changes in the team. We also saw how the concept of historical exposomics became more important, integrating the historical and hydrological perspectives, thanks to the research seminars, literature review, and writing workshops.[31] At the last checkpoint,

31 LuxTIME Seminar Series organized throughout the project to facilitate the exchange of knowledge and stimulate the discussions within and outside the working group: https://luxtimemachine.uni.lu/#1620717521789-01362bff-416e.

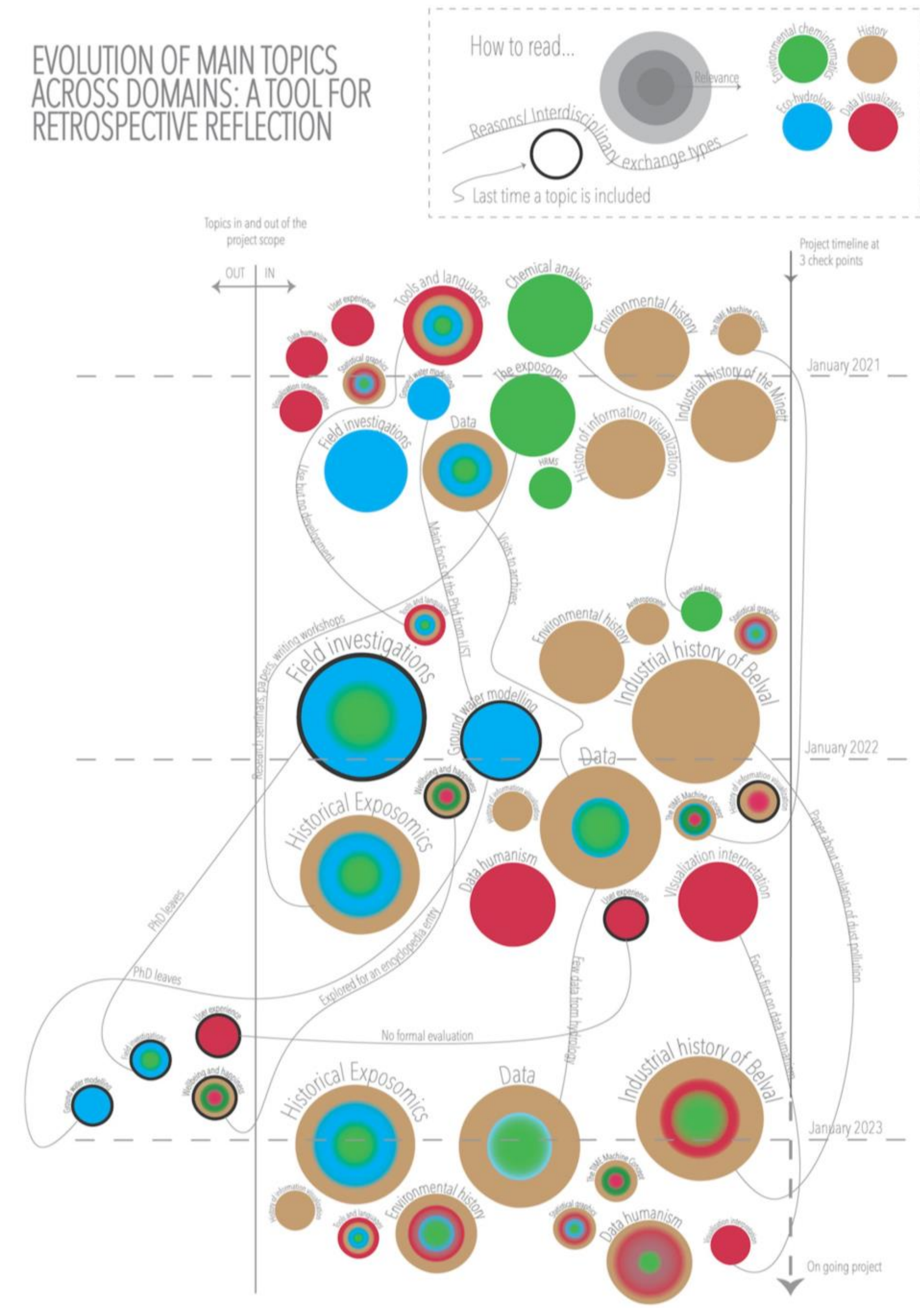


FIGURE 3
Evolution of main topics of LuxTIME.

only a few topics remained, with higher relevance and a blend of history, environmental cheminformatics, eco-hydrology, and data visualization. Overall, we saw a prioritization of topics, knowledge transfer, and shifting contributions among the project members, representing the functioning interdisciplinarity aspect of the project.

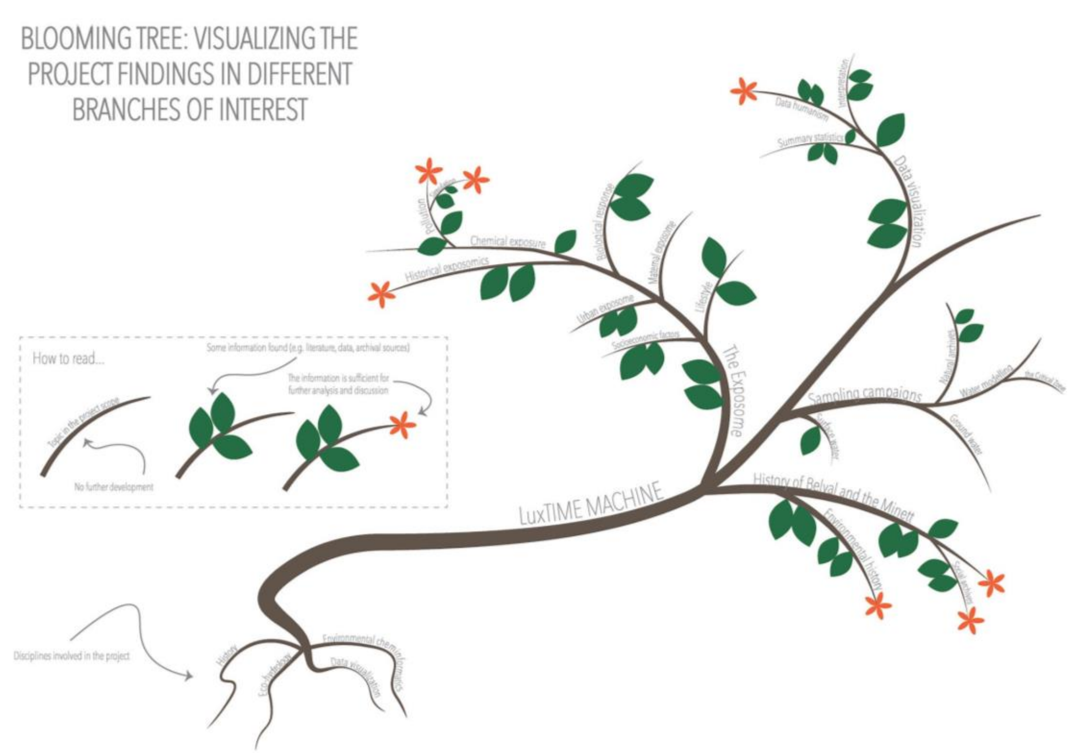


FIGURE 4
LuxTIME tree visualizing the project developments.

This visualization applied the principles of a non-representative approaches are explained in section 3.4.7, where data are not collected before the visualization is created. We used graphical features such as shapes (e.g., circles and lines), size, color, and annotations; to reflect on the evolution of the project. We drew a line connecting circles when we noticed a connection and increased or decreased the size of a circle after discussing the relative relevance of a topic at a given point in time. The existence of these elements can be registered as data, but such data did not exist beforehand, and the exercise started with the interpretative process.

After defining the topics of interest, we collected all the information found through public entities, research centers, libraries, historical archives, and other sources presented in section 3.1. In the visualization in Figure 4, the objective was to display the areas with the most potential based on the information found, i.e., the "flourishing branches" (see visual rhetoric is described in section 3.4.4 at a given moment). We defined three levels: The first was just the branch, which stated that the topic had been considered and researched; the second level was illustrated with a leafy branch, indicating that some information had been found (e.g., historical sources mentioning the topic, existing previous research, and literature available); and the third one, a flowering branch, depicted the possibility of further analysis to extend the research (e.g., data can be extrapolated to the Minett region, it concerns the period of interest, or a projection can be done). The tree, at the same time, allowed us to separate the branches into themes and sub-themes and to represent the roots of the project, the four disciplines. As we can see in Figure 4, the most developed branches, at the time of visualization, were pollution, data humanism, historical exposomics, environmental history, and the discovery of social archives about the history of the Minett region. Other topics for which some information had been found but was not yet sufficient for further analysis included other areas of the *exposome* (e.g., urban exposome, lifestyle, and biological responses) or the sampling campaigns. This visualization could be annotated to explain why these areas had been further developed, which types of data had been found, where, and what analyses they allowed. It could also be combined with a second visualization showing the details, an animation (e.g., showing a tree whose leaves appear, the flowers bloom and then wilt, and the leaves fall off), or adding interactivity that has been neglected in this first visualization phase.

The use of the tree allowed us to use multiple visual metaphors, the roots representing the disciplines, support, and

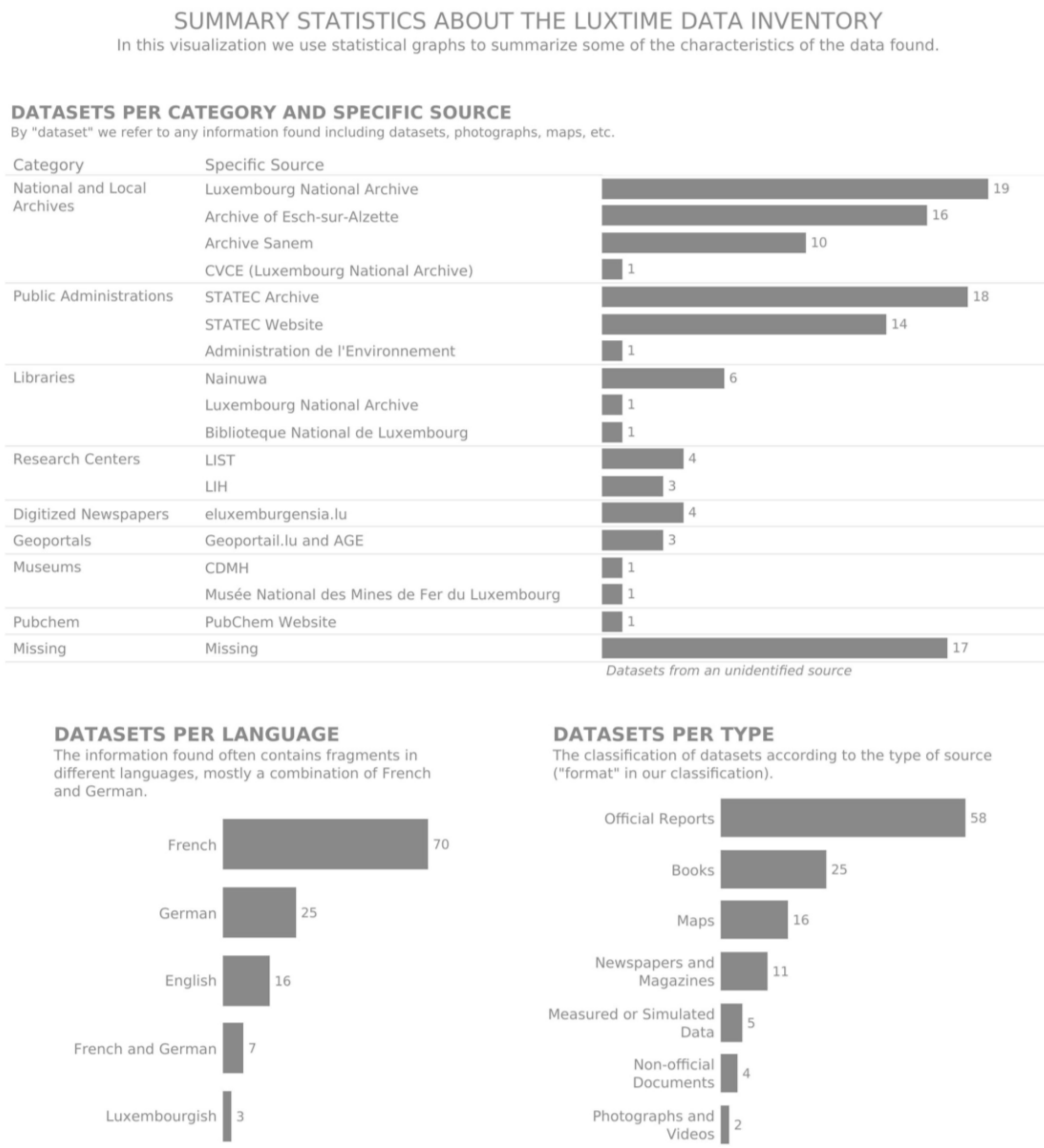


FIGURE 5
Summary statistics about the LuxTIME data inventory.

nourish the project; the branches grow and branch out as the project progresses, and the leaves and flowers bloom due to multiple factors. This visualization showed the progress of the project without using any conventional numerical or graphical charts and could be used to communicate with all kinds of audiences. It also showed which branches of the project were "blooming" at first glance, as a starting point to discuss how to move forward.

4.3. Exploring data and metadata

One of the most challenging aspects of the project was the data collection. In addition to the complexity due to the wide range of topics, there were also different types of sources (e.g., texts, images, and maps) and archives. The inventory of the information continues, as well as the analysis of the different datasets. This section focuses on research question #3 using selected examples:

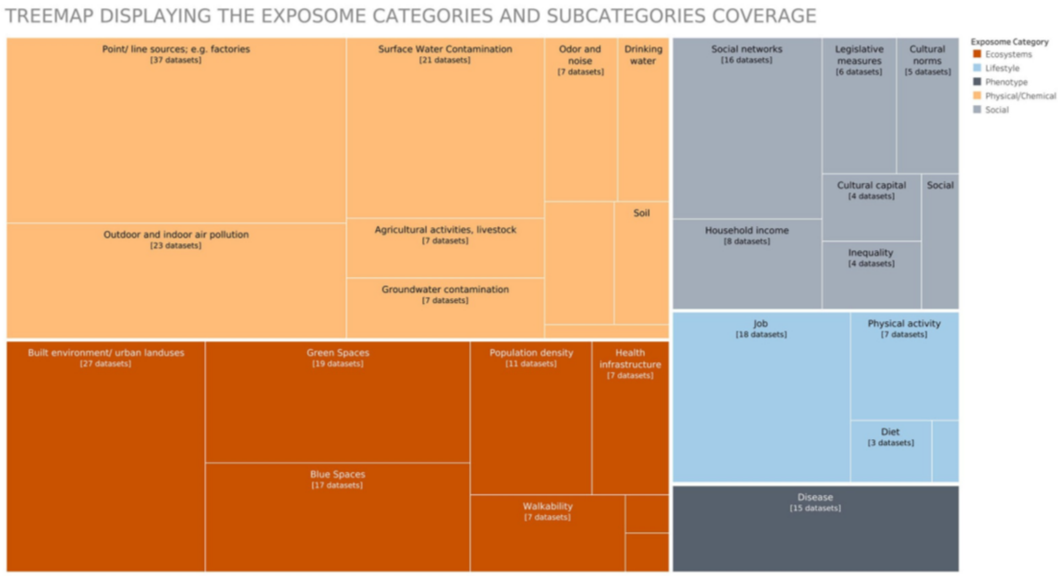


FIGURE 6
Treemap displaying the data inventory by exposome category and subcategory.

How could we explore the data and metadata respecting the priorities of the different disciplines?

4.3.1. Visualizing metadata of the information collected

The use of statistical graphs, such as bar charts allowed us to explore individual variables of the inventory, answering questions such as *How many datasets per source did we have?* This question could be easily answered using summary statistics (Figure 5), as the inventory contains 21 datasets from the Luxembourg National Archive, 18 from STATEC, 16 from the Archive of Esch-sur-Alzette, etc.

How many datasets covered every category and subcategory of the exposome? We did a subdivision of the *exposome* into five categories (Ecosystems, Lifestyle, Social, Physical/Chemical, and Phenotype) and 43 subcategories. In Figure 6, we visualized this using a treemap, with the number of datasets related to the different categories and subcategories. Most datasets belonged to several classification categories, which are not displayed in the treemap. The analysis of this overlap would require a different visualization. In the graph, we could see that the categories with the most information were physical/chemical and ecosystems, followed by social and lifestyle.

The visualizations in Figures 5, 6 allowed us to summarize the metadata, an analysis that could be extended to all the collected metadata in the inventory (e.g., time period, geography, and authors). However, we could not include multiple variables in the same view and still see the details of each of the collected datasets.

For this purpose, we used multivariate data glyphs (see Figure 7). One of the principles of *data humanism* (Lupi, 2017) already introduced in section 3.4.5 is the use of dense and unconventional data visualizations to promote exploration, as it requires the reader to become familiar with the visual encoding, and it layers multiple visual narratives for the readers to follow their own interest "since clarity does not need to come all at once".

As we see in Figure 7, we selected a series of graphical attributes customized to the visualization. The color of the outer pentagon represents the source. In the inventory, we collected 121 datasets from 17 sources, which we grouped for visualization into national and local archives, museums, libraries, research centers, the geoportal, public administrations, and eluxemburgensia. The inner pentagon, in gray, represents the areas of the *exposome* covered in the dataset. The outer square with a thick gray border indicates that the source has been digitized, and each corner of the inner square indicates in which language it is available. At the bottom of the pentagon, the period covered by the data was indicated, with the base coinciding with the period of interest of the project, between 1,850 and 1,950. The concentric triangles at the top right indicate from the inside out the geographical area covered by the data, from the Minett (smallest triangle) to a global scope (largest). This visualization, despite requiring more time to explore, would allow us to see different aspects at the same time, without losing sight of the granularity of the dataset. Further iterations of glyphs and layouts will be evaluated before final implementation.[32]

32 All visualizations are prototypes developed throughout the project, at the end of which, a final version will be integrated into the project website https://luxtimemachine.uni.lu/.

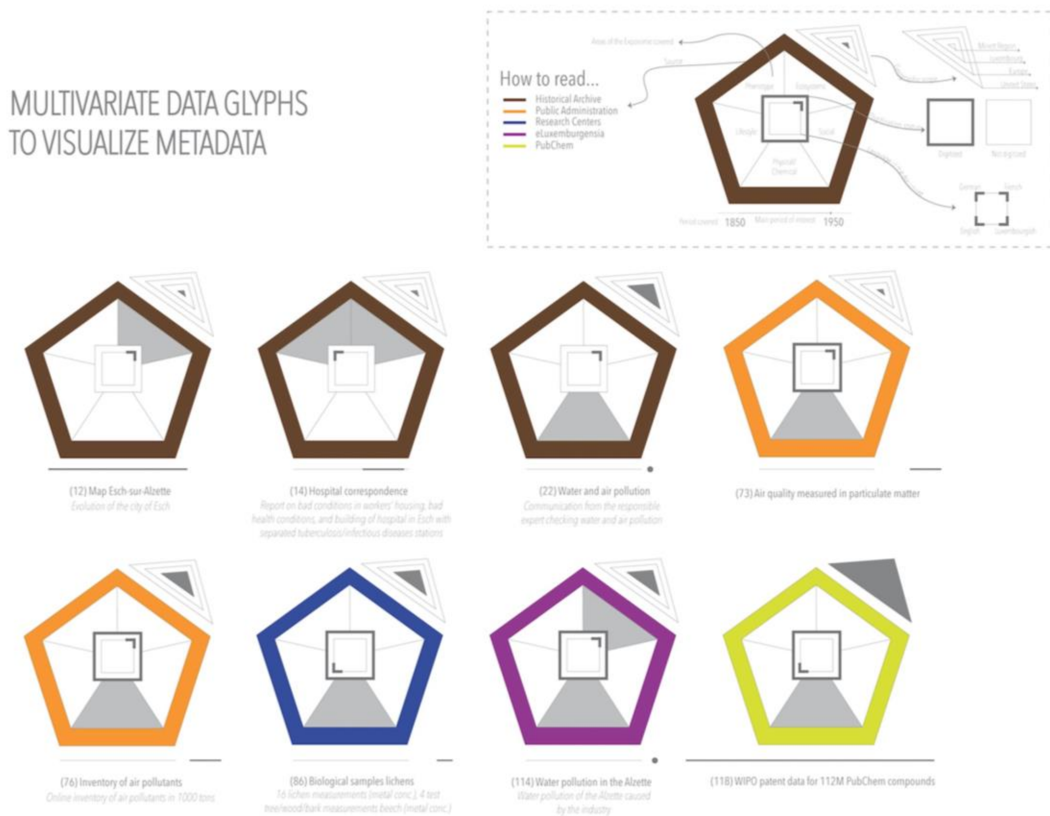


FIGURE 7
Multivariate data glyphs.

4.3.2. Exploring a dataset: number of chemicals registered over time

After having explored the metadata of the data inventory, we started to analyze the datasets. Given the variety of datasets in the inventory, each required a particular analysis to define the required data visualizations. We have chosen, as an example, a set of data about the number of chemicals registered in the Chemical Abstracts Service (CAS) registry[33] over time.

Figure 8 shows a classical representation of quantitative (chemical) data using a line chart. As an alternative, in Figure 9, *chemical stripes* (Arp et al., 2023) (inspired by the *warming* or *climate stripes* discussed in section 3.4.2) are shown, presenting trends of chemical registrations in the CAS registry since 1965, with low numbers in light red and high numbers in dark red. In Figure 9, the color hue was chosen to alert about the situation of increasing chemical numbers, while the color values (from light to dark) allowed the use of red and made it possible for readers with color deficiencies to see the differences.

The numbers of chemicals registered in the CAS registry do not correspond directly to the number or amount of chemicals in use. The stripes could be created for different compound classes or a group of several chemicals, having the registration numbers for these individual requests. The general trend remains the same, irrespective of the view or data: The number of chemicals in use and present in the environment increases along with the number of chemicals. The total number of registered unique substances in the CAS registry lies over 200 million substances currently, with 10–20 million new entries added per year. As in the case of the *climate stripes*, this simplified heatmap conveyed a clear message about the increasing number of chemicals. As discussed in section 3.4.2, this is an example of a variation of a statistical graph, a heatmap, where the graph elements (e.g., axes, legend, and tick marks) are removed to draw attention to a single message, the increasing or decreasing values through the color hue. The objective is not to provide exact values to the reader but to show a noticeably clear trend through minimalism in the visualization.

33 https://www.cas.org/cas-data/cas-registr

FIGURE 8
Line chart: Number of unique substances registered in the CAS registry per year. Source: CAS registry.

FIGURE 9
Stripes: Number of unique substances registered in the CAS Registry per year. Source: CAS registry.

4.4. Monitoring project deliverables and experience

Publications are a fundamental part of the results of a research project, including the various parts of the publication process and the experience of the participants. However, the analysis of this process was not only to optimize the logistics of the project but to embrace a hermeneutic analysis based on interpretation, to understand how the publication processes differ among disciplines, what kind of publications predominate and why, and how it changes in more or less interdisciplinary, individual, or shared publications. We used data visualization not only for the purpose of quantitative analysis (e.g., how many papers had been published) but as a tool for close reading, to develop a deeper understanding of the process itself. In this last section of the results, we focus on *research question #4: How could we monitor the project deliverables, including the different steps of the process, the contribution of the different disciplines, and the experience of the participants?*

In previous examples, we have discussed two time-based visualizations: the evolution of the project scope around three checkpoints throughout our project (Figure 3), and a simplified heatmap to visualize the change of a variable (e.g., CAS registration numbers; Figure 9) over time. Next, we were interested in monitoring the progress of planned publications during the project. We wanted to know the start date of the work, the date of publication, and also the intermediate steps (e.g., when it is submitted and the progress toward acceptance). We also wanted to know which disciplines participated in each publication, whether authors from outside the project team were involved, and what type of publication it was (e.g., article, poster). The visualization technique most frequently used to represent time intervals is Gantt charts. However, a simple Gantt chart did not allow us to represent all the variables of interest in this analysis.

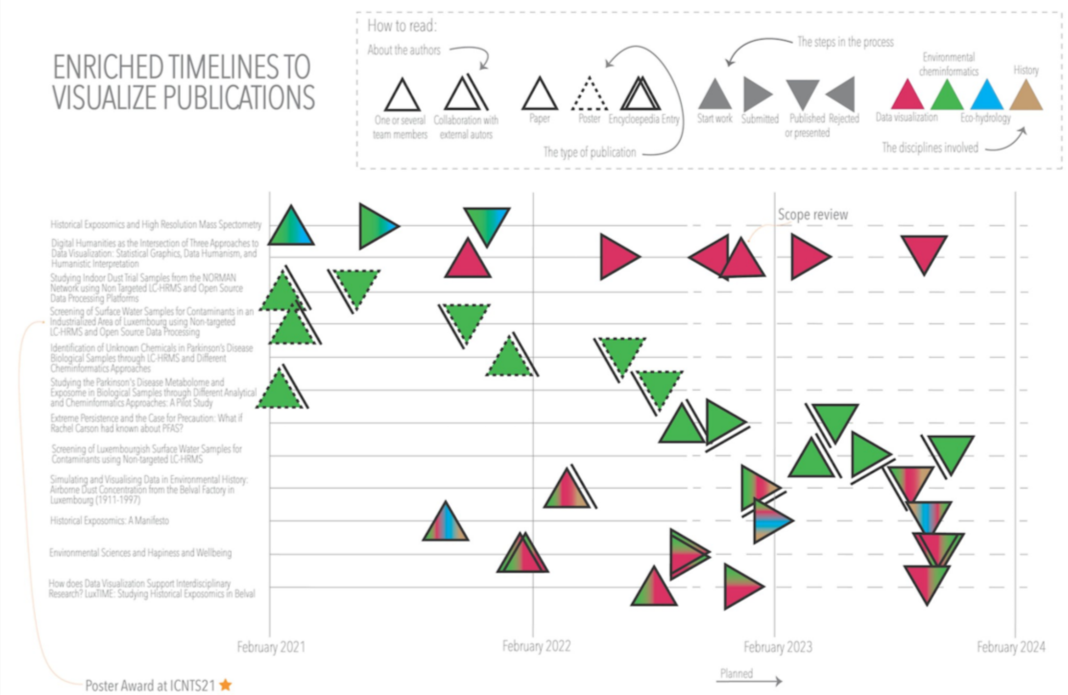


FIGURE 10
Team publications.

In Figure 10, we can see how we could enrich the information shown in a temporal diagram. By combining multivariate data glyphs and placing them at specific points on the timeline, we could divide the process into different steps between the time of starting the work and the publication. We displayed a timeline for each publication, with a triangle at each stage of the process, rotating to the right when the process moves forward (e.g., submission), and to the left when it "goes backwards" (e.g., rejection/revision leading to a restart of the submission process). The color of the triangles indicates that there are individual publications for each discipline, as well as combinations of two or more disciplines. Publication types for environmental cheminformatics (green) include three posters (dashed triangles), seven articles, and an encyclopedia entry. We can also see that in five publications, authors from outside the project are involved. This visualization allows us to see at which moments work accumulates and why. The rotation effect of the triangle helped us to understand parts of the process, for example, the effect of "going backwards" even as time moved forward, due to a rejection that required a restart. At the same time, the rotation of the triangle prevented us from seeing when the event occurred with precision (i.e., vertex or center), but as in this visualization, we did not need exact dates; we chose to keep the triangular shape and take advantage of the rotation effect.

The time scale could be visualized in more detail (e.g., with monthly details), adding more steps to the process (e.g., differentiate acceptance and publication), or factoring in the time perception of different participants, using non-standard intervals. For example, one of the authors might have perceived the initial time spent working on a publication differently from the time spent working after the rejection of the initial version (see Figure 11). The visualization in Figure 11 highlights the difference between time and temporality, the latter being relational. It opens the visualization techniques to graphical methods that represent experiential temporality, a subjective experience that depends on many psychological and physiological factors. If different participants of the project were to repeat the visualization for different publications, the challenges related to the different points of view of the multiple sources would need to be accounted for as discursive temporality. The visualization of time—based on experience—allows us to explore how the different participants experience the project (e.g., which moments are perceived as most stressful). This visualization technique, where the timeline is not standard, is probably the one that takes epistemological differences the furthest.

5. Discussion

In this article, we have presented the LuxTIME *data visualization toolbox*, including several data visualization concepts and techniques from epistemologically distant disciplines.

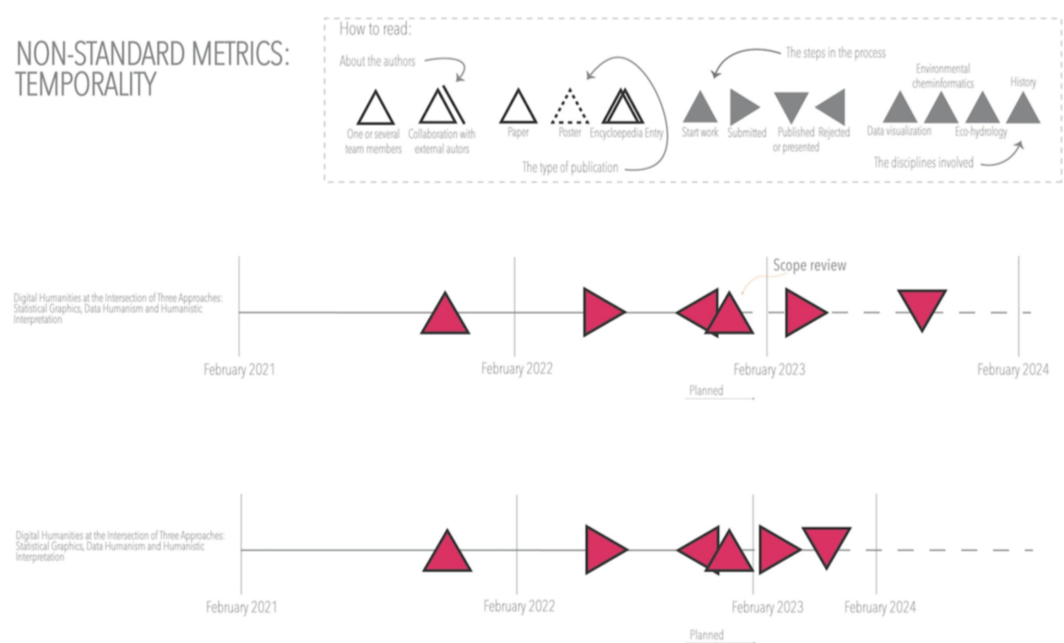


FIGURE 11
Non-standard time intervals.

This *toolbox* has facilitated an interdisciplinary collaboration among researchers in history, environmental cheminformatics, eco-hydrology, and data visualization, learning and applying concepts and techniques that reflect the paradigms of the other disciplines (e.g., integrating the use of rhetoric or temporality in natural sciences). Through the data collection process, the numerous exchanges on data visualization, the sketching and prototype development sessions, and the feedback collection and implementation, we have experienced and learned about the frictions inherent in interdisciplinary work.

First, the different perspectives on the level of granularity are worth noting, a friction already described by Panagiotidou et al. (2022). Especially, in the visualization process, we had several discussions about how important the details were in the visualizations (i.e., how we reached the results through several iterations with positive and challenging experiences: failure, team changes, and learning opportunities) vs. just visualizing the result. Second, by integrating the ideas of *data humanism* and *interpretation* into our *toolbox*, we increased the time needed to explore some of the visualizations, which often triggered the "discomfort" of not being able to immediately arrive to a clear conclusion. For example, in multivariate visualizations, there are several levels of information that cannot be extracted at first glance, but they allow us to show many facets of the metadata in a single view if we are willing to spend more time in the exploration. Third, the most accepted concepts were the integration of rhetorical mapping, and the variations of statistical graphics, notably the minimalist approach (i.e., no axes, tick marks, or grids), probably because it already had a strong precedent with the climate stripes.

Finally, one of the elements that we added to the *toolbox* at a later stage, because the need emerged, was the correct use of statistical graphics. Although statistical graphics were often used in all the disciplines involved, the visualizations generated did not always respect established theories on data visualization research. These discussions (e.g., color theory) highlighted the need to collaborate with researchers in data visualization so as not to perpetuate errors within the disciplines. Moreover, the fact of including concepts and techniques that serve different objectives (e.g., drawing quick conclusions using perception theory vs. exploring in-depth multiple narratives through multivariate visualizations with custom visual vocabularies) in our *toolbox*, created a series of discussions about how blurred the lines are between such "tools" and the paradigms they come from.

Throughout the research, the idea of data visualization as a *sandcastle* has been very present, especially the use of data visualization as part of the *speculative process*, and not just to present the results once the work is done (Hinrichs et al., 2019). This research is an attempt to use data visualization not only to explore and communicate about the domain data and metadata but also to collect and explore forms of thinking and creating knowledge through visual means. To make this visual means more specific, we have collected a series of existing concepts and techniques, our *toolbox*, to encourage participants to rethink what they already know but also to defamiliarize themselves with the usual methods and use data visualization as an *aesthetic provocation* (Hinrichs et al., 2019) to open up new perspectives for their own disciplines and the interdisciplinary work at LuxTIME. Throughout the process, visualization in its role as a *mediator* promotes an open and critical discourse (Hinrichs et al., 2019).

All the visualization concepts and techniques discussed in this article are just a proposal for a *data visualization toolbox* suitable for many research fields. Such concepts and techniques are not new but are rarely combined across disciplines or within the framework of a single project. In LuxTIME, we wanted to experiment with different data visualization techniques in a practical way, through our own process of interdisciplinary learning and exchange, during the definition of the project, and to explore the data and metadata found in relation to our main theme: historical exposomics in the Minett region. The aim was to extend the "go-to" *data visualization toolbox* and to explore, validate, and communicate, benefiting from techniques researched and applied across different disciplines. The use of a variety of techniques allowed us to look at data from different perspectives, analyzing the process quantitatively, qualitatively, and in an interpretative manner (e.g., combining Gantt charts with more flexible and detailed views, using non-standard timelines to express the experience of participants), and alternating and combining the use of statistical graphs with the use of metaphors or other graphic elements, whose aim is not necessarily to communicate quickly and accurately but to foster emotions.

This combination of statistical graphics and their variations, the use of information design elements generally present in the so-called *data art*, as well as the integration of interpretative elements gave us an *extended data visualization toolbox* to navigate our project. Probably because of the interdisciplinary nature of the project, this combination was more obvious, but we believe it could be useful for projects of all kinds. Such a *playground* is necessary to be able to formally evaluate the combined use of these techniques in future. This exploratory practice refers to the notion of "thinkering", composed of the verbs tinkering and thinking that describes the action of playful experimentation with digital tools for the interpretation and presentation of history (Fickers and van der Heijden, 2020). There is no one-size-fits-all toolbox for every research project, as the toolbox concept is built on the idea of flexible rearrangements of tools depending on the research questions, needs, and aims of a project. The toolbox is therefore the result of a "co-design" process based on situated knowledge practices.

The visualizations presented in this article are prototypes that will evolve further toward the end of the project, incorporating other techniques (e.g., interactivity, direct visualization, and enhanced ways of storytelling). Moreover, after having experimented with several types of visualizations separately to address different research questions, future steps will include a review of the connections between different visualizations and how they are linked in the overall narrative.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: Data will be published at the end of the project in 2024. In this article we just discuss the data visualizations. Requests to access these datasets should be directed to aida.horanietibanez@uni.lu and dagny.aurich@uni.lu.

Ethics statement

Ethical approval was not required for the study involving human participants in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants in accordance with the national legislation and the institutional requirements.

Author contributions

All authors participated in the conception and design of the study, performed the data collection, data analysis, data interpretation, drafting of the article, contributed to the manuscript revision, and read and approved the submitted version.

Funding

This research has been funded by the University of Luxembourg Institute for Advanced Studies Audacity Program for the Luxembourg Time Machine (LuxTIME).

Acknowledgments

The authors would like to thank Emma Schymanski and Andreas Fickers for the detailed feedback on this study and the LuxTIME team members (Christophe Hissler, Laurent Pfister, Reinhard Schneider, and Lars Wieneke) and Johanna Drucker who took part in the discussion on our research, as well as Jennifer Yang for providing the history of the CAS registry numbers.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Aigner, W., Miksch, S., Schumann, H., and Tominski, C. (2011). *Visualization of Time-Oriented Data*. London: Springer London.

Arp, H. P. H., Aurich, D., Schymanski, E. L., Sims, K., and Hale, S. E. (2023). Avoiding the next silent spring: our chemical past, present, and future. *Environ. Sci. Technol.* 57, 6355–6359. doi: 10.1021/acs.est.3c01735

Aurich, D., Miles, O., and Schymanski, E. L. (2021). historical exposomics and high resolution mass spectrometry. *Exposome* 1, osab007. doi: 10.1093/exposome/osab007

Cairo, A. (2019). *How Charts Lie: Getting Smarter About Visual Information*. 1st ed. New York, N.Y.: W.W. Norton and Company.

Chuah, M. C., and Eick, S. G. (1998). Information rich glyphs for software management data. *IEEE Comput. Graph. Appl.* 18, 24–29. doi: 10.1109/38.689658

Cleveland, W. S., and McGill, R. (1984). Graphical perception: theory, experimentation, and application to the development of graphical methods. *J. Am. Stat. Assoc.* 79, 531–554. doi: 10.1080/01621459.1984.10478080

Cleveland, W. S., and McGill, R. (1986). An experiment in graphical perception. *Int. J. Man-Mach. Stud.* 25, 491–500. doi: 10.1016/S0020-7373(86)80019-0

Cordell, R. (2019). "Teaching humanistic data analysis," in *Digital Scholarship, Digital Classrooms: New International Perspectives on Research and Teaching: Proceeding of the Gale Digital Humanities Day at the British Library*.

D'Ignazio, C., and Klein, L. F. (2020). *Data Feminism*. Cambridge, MA: The MIT Press.

Dörk, M., Feng, P., Collins, C., and Carpendale, S. (2013). "Critical InfoVis: exploring the politics of visualization" in *CHI'13 Extended Abstracts on Human Factors in Computing Systems* (Paris France: ACM), 2189–2198.

Drucker, J. (2011). *Humanities Approaches to Graphical Display*. Boston, MA: DHQ. p. 5.

Drucker, J. (2018). Non-representational approaches to modeling interpretation in a graphical environment. *Digit. Scholarsh. Humanit.* 33, 248–263. doi: 10.1093/llc/fqx034

Drucker, J. (2021). *TimeCapsule Report on Research 2001-2021*. UCLA, Department of Information Studies and Design.

Dykes, B. (2020). *Effective Data Storytelling: How to Drive Change With Data, Narrative and Visuals*. Hoboken, NJ: John Wiley and Sons, Inc.

Evergreen, S. D. H. (2017). *Effective Data Visualization: The Right Chart for the Right Data*. Los Angeles, CA: SAGE.

Fickers, A., Tatarinov, J., Fickers, A., Tatarinov, J., and van der Heijden, T. (2022). "Digital history and hermeneutics – between theory and practice: an introduction," in *Digital History and Hermeneutics: Between Theory and Practice* (Berlin: De Gruyter).

Fickers, A., and van der Heijden, T. (2020). *Inside the Trading Zone: Thinkering in a Digital History Lab*. Boston, MA: DHQ. p. 14. Available online at: http://digitalhumanities.org/dhq/vol/14/3/000472/000472.html (accessed September 05, 2023).

Glivinska, A. (2021). *The 25 Best Data Visualizations of 2023 [Examples]*. Visme. Available at: https://visme.co/blog/best-data-visualizations/ (accessed January 13, 2023).

Hall, K., Wm., Bradley, A. J., Hinrichs, U., Huron, S., Wood, J., et al. (2020). Design by immersion: a transdisciplinary approach to problem-driven visualizations. *IEEE Trans. Vis. Comput. Graph.* 26, 109–118. doi: 10.1109/TVCG.2019.2934790

Hill, C. A., and Helmers, M. (2009). *Defining Visual Rhetorics*. Reprinted from Lawrence Erlbaum ed. New York, NY: Routledge.

Hinrichs, U., Forlini, S., and Moynihan, B. (2019). In defense of sandcastles: Research thinking through visualization in digital humanities. *Digit. Scholarsh. Humanit.* 34, i80–i99. doi: 10.1093/llc/fqy051

Hullman, J., and Diakopoulos, N. (2011). Visualization rhetoric: framing effects in narrative visualization. *IEEE Trans. Vis. Comput. Graph.* 17, 2231–2240. doi: 10.1109/TVCG.2011.255

Karlsson, O., Rocklöv, J., Lehoux, A. P., Bergquist, J., Rutgersson, A., Blunt, M. J., et al. (2021). The human exposome and health in the Anthropocene. *Int. J. Epidemiol.* 50, 378–389. doi: 10.1093/ije/dyaa231

Kemman, M. (2021). *Trading Zones of Digital History*. Berlin: De Gruyter.

Kim, M. C., Zhu, Y., and Chen, C. (2016). How are they different? A quantitative domain comparison of information visualization and data visualization (2000–2014). *Scientometrics* 107, 123–165. doi: 10.1007/s11192-015-1830-0

Kim, N. W., Im, H., Riche, N. H., Gajos, K., and Pfister, H. (2018). *Fostering Data Humanism With DataPortraits: Empowering People to Create a Personalized Visual Vocabulary*. IEEE VIS Poster (VIS).

Kim, N. W., Im, H., Henry Riche, N., Wang, A., Gajos, K., and Pfister, H. (2019). "DataSelfie: empowering people to design personalized visuals to represent their data" in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow Scotland Uk: ACM), 1–12.

Knebeler, C., and Scuto, D. (2010). *Belval: Passé, présent et avenir d'un site luxembourgeois exceptionnel (1911-2011)*. Esch-sur-Alzette: Le Phare.

Lankow, J., Ritchie, J., and Crooks, R. (2012). *Infographics: The Power of Visual Storytelling*. Hoboken, N.J: John Wiley and Sons, Inc.

Lanzing, J. (1998). Concept mapping: tools for echoing the minds eye. *J. Vis. Lit.* 18, 1–14. doi: 10.1080/23796529.1998.11674524

Lima, M. (2014). *The Book of Trees: Visualizing Branches of Knowledge*. New York, NY: Princeton Architectural press.

Lockwood, A. (1969). *Diagrams: A Visual Survey of Graphs, Maps, Charts and Diagrams for the Graphic Designer*. London; New York, NY: Studio Vista; Watson-Guptill.

Loukissas, Y. A. (2019). *All Data are Local: Thinking Critically in a Data-Driven Society*. Cambridge, MA: The MIT Press. doi: 10.7551/mitpress/11543.001.0001

Lupi, G. (2017). *Data Humanism: The Revolutionary Future of Data Visualization*. PRINT Mag. Available online at: https://www.printmag.com/article/data-humanism-future-of-data-visualization/ (accessed January 10, 2023).

Manovich, L. (2011). What is visualisation? *Vis. Stud.* 26, 36–49. doi: 10.1080/1472586X.2011.548488

Mao, Y., Wang, D., Muller, M., Varshney, K. R., Baldini, I., Dugan, C., et al. (2019). How data scientists work together with domain experts in scientific collaborations: to find the right answer or to ask the right question? *Proc. ACM Hum.-Comput. Interact.* 3, 1–23. doi: 10.1145/3361118

Miller, G. W. (2020). *The Exposome - A New Paradigm for the Environment and Health*. Cambridge, MA: Academic Press.

Miller, G. W., and Jones, D. P. (2014). The nature of nurture: refining the definition of the exposome. *Toxicol. Sci.* 137, 1–2. doi: 10.1093/toxsci/kft251

Novak, J. D., and Cañas, A. J. (2008). *The Theory Underlying Concept Maps and How to Construct and Use Them. Tech. Rep. IHMC CmapTools.* Available online at: http://cmap.ihmc.us/Publications/ResearchPapers/TheoryUnderlyingConceptMaps.pdf (accessed September 05, 2023).

Nussbaumer Knaflic, C. (2015). *Storytelling With Data: A Data Visualization Guide for Business Professionals*. Hoboken, NJ: Wiley.

Panagiotidou, G., Gorucu, S., and Vande Moere, A. (2020). Data badges: making an academic profile through a DIY wearable physicalization. *IEEE Comput. Graph. Appl.* 40, 51–60. doi: 10.1109/MCG.2020.3025504

Panagiotidou, G., Poblome, J., Aerts, J., and Vande Moere, A. (2022). Designing a data visualisation for interdisciplinary scientists. How to transparently convey data frictions? *Comput. Support. Coop. Work CSCW.* 31, 633–667. doi: 10.1007/s10606-022-09432-9

Rhyne, T., Tory, M., Munzner, T., Ward, M., Johnson, C., and Laidlaw, D. H. (2003). "Information and scientific visualization: separate but equal or happy together at last" in *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control* (Seattle, WA, USA: IEEE), 611–614.

Rogowitz, B. E., and Treinish, L. A. (1998). Data visualization: the end of the rainbow. *IEEE Spectr.* 35, 52–59. doi: 10.1109/6.736450

Romance, N. R., and Vitale, M. R. (1999). Concept mapping as a tool for learning: broadening the framework for student-centered instruction. *Coll. Teach.* 47, 74–79. doi: 10.1080/87567559909595789

Santoro, C., and Kirkland, C. (2022). The warming stripes - Inspiring a movement. *Nightingale Mag.* 2022, 52–60.

Schwabish, J. A. (2021). *Better Data Visualizations: A Guide for Scholars, Researchers, and Wonks*. New York, NY: Columbia University Press.

Schwan, H., Arndt, J., and Dörk, M. (2022). Disclosure as a critical-feminist design practice for Web-based data stories. *First Monday* 27, 11. doi: 10.5210/fm.v27i11.12712

Segel, E., and Heer, J. (2010). Narrative visualization: telling stories with data. *IEEE Trans. Vis. Comput. Graph.* 16, 1139–1148. doi: 10.1109/TVCG.2010.179

Silva, S., Madeira, J., and Santos, B. S. (2007). "There is more to color scales than meets the eye: a review on the use of color in visualization" in *2007 11th International Conference Information Visualization (IV'07)* (Zurich, Switzerland: IEEE), 943–950.

Silva, S., Sousa Santos, B., and Madeira, J. (2011). Using color in visualization: a survey. *Comput. Graph.* 35, 320–333. doi: 10.1016/j.cag.2010.11.015

Steen, G. (2018). *Visual Metaphor: Structure and Process*. Amsterdam; Philadelphia: John Benjamins Publishing Company.

Thrift, N. J. (2008). *Non-Representational Theory: Space, Politics, Affect*. 1st ed. London: Routledge.

Tufte, E. R. (1999). *The Visual Display of Quantitative Information*. 17th ed. Cheshire; Conn: Graphics Press.

Van Wijk, J. J. (2006). Bridging the gaps. *IEEE Comput. Graph. Appl.* 26, 6–9. doi: 10.1109/MCG.2006.120

Vora, S. (2019). *The Power of Data Storytelling*. Thousand Oaks, CA: SAGE Publications India Pvt Ltd.

Ward, M. O. (2008). "Multivariate Data Glyphs: Principles and Practice," in *Handbook of Data Visualization Springer Handbooks Comp.Statistics (Berlin, Heidelberg: Springer Berlin Heidelberg)*, 179–198.

Ware, C. (2004). *Information Visualization: Perception for Design*. San Francisco, CA: Morgan Kaufman.

Wheeldon, J., and Faubert, J. (2009). Framing experience: concept maps, mind maps, and data collection in qualitative research. *Int. J. Qual. Methods* 8, 68–83. doi: 10.1177/160940690900800307

V. Avoiding the Next Silent Spring: Our Chemical Past, Present, and Future

Authors:

Hans Peter H. Arp, **Dagny Aurich**, Emma L. Schymanski, Kerry Sims, and Sarah E. Hale

Manuscript type:

Viewpoint

Contribution statement:

HPA, SEH and ELS conceptualized the viewpoint and were responsible for writing (original draft preparation). **DA**, ELS, KS, HPA and SEH did the review and editing. ELS, HPA and SEH were further responsible for funding acquisition and supervision. **DA** contributed Figure 1 to the manuscript, inspired by the original suggestion of KS.

Submission status:

Published in Environmental Science & Technology on 13 April 2023.

(DOI: 10.1021/acs.est.3c01735)

Short summary/ Contribution to the field:

This viewpoint paper – published 60 years after the release of Rachel Carson's 'Silent Spring' – discusses the importance of using precaution and improving the management of persistent chemicals like per- and polyfluoroalkyl substances (PFAS). Looking at past developments of (reactionary) regulatory measures and the rapid increase of the chemical space and chemical use, the need for action is emphasized. The lack of understanding about the potential health effects of new chemicals in the past has had a noticeable impact on the present. Presently, we are confronted with challenges related to the persistence of these chemicals in the environment, in addition to the health effects. Global efforts to improve experimental designs, *in silico* approaches and to promote regulatory measures are suggested to avoid possible new 'Silent Springs'.

The 'chemical stripes' visualization in Figure 1 of this viewpoint attracted considerable attention on social media platforms and within the scientific community overall (paper is in the top 5% of all research outputs scored by Altmetric). The resulting sonification of the stripes by Jamie Perrera can be accessed via https://vimeo.com/862087332 (487 views). To increase accessibility, an R package was developed by **DA** (available on GitLab https://gitlab.lcsb.uni.lu/eci/chemicalstripes). More information on the stripes can also be found in the SETAC23 presentation slides on Zenodo (DOI: 10.5281/zenodo.7885031, 399 views, 235 downloads).

Environmental Science & Technology

pubs.acs.org/est

Viewpoint

Avoiding the Next Silent Spring: Our Chemical Past, Present, and Future

Hans Peter H. Arp,* Dagny Aurich, Emma L. Schymanski,* Kerry Sims, and Sarah E. Hale

Cite This: https://doi.org/10.1021/acs.est.3c01735

Read Online

ACCESS | Metrics & More | Article Recommendations

KEYWORDS: *precautionary principle, persistent, chemicals, patents, zero pollution, circular economy, regulation*

Rachel Carson's *Silent Spring*,[1] published just more than 60 years ago, outlined how the indiscriminate use of dichlorodiphenyltrichloroethane (DDT), a potent, environmentally persistent insecticide, was damaging the world's ecosystems, animals, and food supply. There were many other chemicals more persistent than DDT accumulating in the environment when Carson was writing, including per- and polyfluoroalkyl substances (PFAS). While man-made, PFAS were not intended to cause harm, contrary to pesticides such as DDT. Today, ambient PFAS levels are contaminating rain, soil, and drinking water resources worldwide to such an extent that they have caused substantial, irreversible health and environmental damage.[2] Like DDT, PFAS had long been in use by the time Rachel Carson was writing *Silent Spring* (see Figure 1). However, their environmental presence went unnoticed by Carson and other contemporary environmental researchers. PFAS were entering the environment under the radar, except to those who were manufacturing and emitting them.[3]

■ WHY WERE PFAS NOT CONSIDERED BY RACHEL CARSON?

When Rachel Carson was writing *Silent Spring*, the field of environmental chemistry was in its infancy, particularly in terms of the ability to detect synthetic organic substances in the environment. Carson's case against excessive DDT use was triggered mainly by visible toxicological and ecological observations. Analytical data that proved ubiquitous exposure and accumulation in the food chain were lacking. Ultimately, it was James Lovelock's development of electron capture detection and its coupling with gas chromatography[4] that enabled other scientists to confirm the omnipresence and bioaccumulation of DDT. Lovelock reflected that the use of his technology to demonstrate the "ubiquitous distribution of pesticides throughout the global environment did much to fuel the environmental revolution which followed. [This] lent veracity to the otherwise unprovable statements of that remarkable book by Rachel Carson".[4]

A few years after *Silent Spring*'s publication, Søren Jensen identified polychlorinated biphenyls (PCBs) in white-tailed eagle samples for the first time, while analyzing for DDT with this new technique.[5] PCBs were later confirmed to be as ubiquitous in the environment and food chain as DDT. This discovery was rapidly accompanied by the detection of several other persistent organic pollutants (POPs), ultimately leading to the first "dirty dozen" POPs appearing in the United Nations Stockholm Convention, which was adopted in 2001, almost 40 years after Rachel Carson's book was first published. By 2009, the most well-known substance of the PFAS family was added to the Stockholm Convention [perfluorooctanesulfonic acid (PFOS)]. Other PFAS have since followed, including perfluorooctanoic acid (PFOA) in 2019 and perfluorohexanesulfonic acid (PFHxS) in 2022 (Figure 1).

■ WHAT IF RACHEL CARSON HAD MENTIONED PFAS IN *SILENT SPRING*?

If Rachel Carson had known about PFAS and included them in *Silent Spring*, it is probable that the rapid global policy and industry action to manage DDT and PCBs would also have been applied to PFAS. One or more PFAS may even have been added to the original "dirty dozen" in 2001. Without the regulation or stewardship activities instigated by *Silent Spring*, there is little doubt that emissions of DDT, PFAS, and many other persistent pollutant groups would have been worse. This is evident in Figure 1, as the colored stripes present the relative number of filed patents for DDT and selected PFAS over time. The numbers of patents increased at an exponential rate, with patents for DDT and PFOS continuing to increase irrespective of regulatory efforts such as the Stockholm Convention listing dates shown in Figure 1. One exception to this is the most recent decrease in the number of patents for PFOA, which may be a sign of industry responding to its inclusion in the Stockholm Convention.

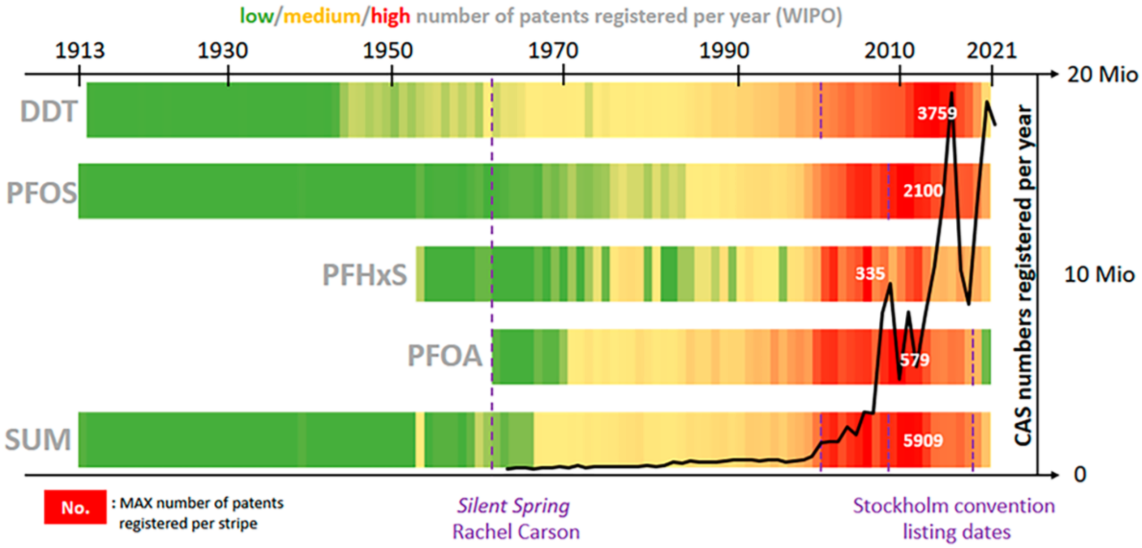


Figure 1. Chemical stripes for DDT and various PFAS. The colored stripes show the distribution of patents registered per year for DDT, PFOS, PFHxS, PFOA, and the sum of all four chemicals. Superimposed in black is the number of chemicals registered in the Chemical Abstracts Service (CAS) Registry each year. Purple dashes denote key publications and regulatory dates. Sources: World Intellectual Property Organisation (WIPO) patent numbers extracted from PubChem;[6] CAS registration data provided by CAS.

■ USING PRECAUTION TO PREVENT FUTURE SILENT SPRINGS

A shortcoming of the implementation of a chemical regulation like the Stockholm Convention is that it is reactionary and not precautionary. Substances are added to the Stockholm Convention only after exposure and ecological harm has been demonstrated through environmental and laboratory observations, often long after the first awareness of red flags. But what of the other unknown environmentally persistent substances that are out there, or that may be in future?

Rachel Carson wrote in *Silent Spring*, "the new chemicals come from our laboratories in an endless stream; almost five hundred annually find their way into actual use in the United States alone. The figure is staggering and its implications are not easily grasped—500 new chemicals to which the bodies of men and animals are required somehow to adapt each year, chemicals totally outside the limits of biologic experience."[1] Shortly after *Silent Spring* was written, the number of chemicals present in the Chemical Abstract Services (CAS) Registry was 211 934 (in 1965). In March 2023, the total has reached 204 million chemicals, 3 orders of magnitude higher. In the past several years, the number of new CAS registrations has increased to 10−20 million per year (black line in Figure 1), 5 orders of magnitude higher than the rate of 500 chemicals per year quoted by Rachel Carson. Among the new CAS registrations are likely multiple extremely persistent substances, plausibly ranging from the hundreds to hundreds of thousands. Many of those being registered now could turn out to be the next DDT or PFAS.

The premise by the Renaissance physician Paracelsus, "the dose makes the poison", is an irresponsible axiom for managing persistent substances that accumulate in the environment. In 1962, Carson did not know that PFAS could be a poison, and in 2023, scientists are still researching the dose.[2,7] Because the number of chemicals registered per year (Figure 1) is now in the millions, it is clearly not possible or desirable to perform a detailed risk assessment for all substances. It is likely that there are already extremely persistent substances in the environment that are causing harm and for which we have little knowledge or data. So how can we stop this pattern of reoccurring silent springs?

■ IMPROVING THE MANAGEMENT OF PERSISTENT SUBSTANCES

A precautionary approach is the only way forward when it comes to managing new and existing, extremely persistent substances with a clear exposure pathway to humans and the environment. This precautionary approach must be applied to a future of chemical innovation that is centered around concepts such as the "circular economy" and "safe and sustainable by design" (SSbD), which consider the diverse impacts of chemicals over their entire life cycle.[8] To enable this, research and innovation must shift toward making substances with lifetimes designed for their intended use within the circular economy. Such substances should degrade naturally or be triggered to do so at the end of their useful life cycle. As an illustrative example, some oxo-polymers are completely mineralizable in agricultural soils but can be persistent upon reaching marine environments.[9] Such oxo-polymers are potentially safer replacements to extremely persistent pesticides and plastics on soils for which they were designed; however, containment to prevent marine emissions of these oxo-polymers at their end of life would become a management priority. Similarly, persistent substances found in reusable products would need to be managed such that they are either retained in the circular economy without emissions or designed for technical or natural degradation at the end of life.

To improve such management of persistent substances, there are three fronts that require further attention: improving experimental testing, developing *in silico* methods, and strengthening regulatory options. To improve experimental testing, simplified protocols that can be applied to several

https://doi.org/10.1021/acs.est.3c01735
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

substances simultaneously would be highly valuable. This could include the development and implementation of "benchmarking" approaches, in which substances with unknown half-lives are placed in the same simulation system (or mesocosm) as those with well-known half-lives, and the degradation rates of the unknown substances are benchmarked to the known substances over time.[10,11]

Developing *in silico* methods will be necessary to strengthen and bridge the experimental and regulatory approaches to persistence. Considering the large number of chemicals on the global chemical market,[12] *in silico* methods are the only feasible way to assess all of them, though they remain highly inaccurate due to large data gaps.[13] Nevertheless, high-quality *in silico* approaches, supported by additional experimental data, remain an aspiration as they require substantially less time and resources than experimental testing. In addition, *in silico* approaches could be used in the chemical design and synthesis phase to identify new and novel replacements for persistent substances for testing or development. Increasing the availability and digitization of high-quality experimental half-life and transformation data, coupled with advances in cheminformatics and machine learning tools, will increase the accuracy of *in silico* assessment of environmental persistence based on molecular structure.[14]

Finally, to improve regulatory options over the whole life cycle of chemicals, regulators could require and if necessary act upon information related to degradation conditions of substances, targeting chemical uses with pathways to the environment. Inspiration to improve such regulatory options can be found in a recent examination of approaches to persistence assessments[11] and the European Union's Chemicals Strategy for Sustainability (CSS).[15] The CSS includes several initiatives to develop more precautionary approaches for extremely persistent substances, including a broad group restriction of PFAS, and the introduction of new hazard categories, including persistence [i.e., persistent, bioaccumulative, and toxic (PBT); very persistent and very bioaccumulative (vPvB); persistent, mobile, and toxic (PMT); and very persistent and very mobile (vPvM)].[15]

Expansion of work in these areas is required on a global scale to truly avoid the next silent spring, alongside the evolution and widespread adoption of approaches like SSbD and the circular economy. Improving our scientific understanding of environmental persistence, along with developing *in silico* methods, will encourage better, greener innovation and regulation that will result in the accumulation of fewer persistent substances in the environment. Learning from our past and present to improve a precautionary approach to persistent substances will ultimately allow humankind to foresee and forestall a future consigned to transgressing planetary boundaries and recurring silent springs.[1]

■ AUTHOR INFORMATION

Corresponding Authors

Hans Peter H. Arp – *Norwegian Geotechnical Institute (NGI), 0806 Oslo, Norway; Department of Chemistry, Norwegian University of Science and Technology (NTNU), 7491 Trondheim, Norway;* orcid.org/0000-0002-0747-8838; Email: Hans.Peter.Arp@ngi.no

Emma L. Schymanski – *Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, 4367 Belvaux, Luxembourg;* orcid.org/0000-0001-6868-8145; Email: emma.schymanski@uni.lu

Authors

Dagny Aurich – *Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, 4367 Belvaux, Luxembourg;* orcid.org/0000-0001-8823-0596

Kerry Sims – *Environment Agency, Bristol BS1 5AH, U.K.;* orcid.org/0000-0001-6273-4083

Sarah E. Hale – *Norwegian Geotechnical Institute (NGI), 0806 Oslo, Norway;* orcid.org/0000-0002-7743-9199

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.est.3c01735

Notes

The authors declare no competing financial interest.

Biographies

Hans Peter H. Arp, NGI, is an environmental chemist interested in how fundamental aspects of physical chemistry can be utilized as applied tools for understanding and preventing pollution exposure. His projects focus on designing solutions through policy mechanisms, chemical properties, interdisciplinary collaboration, and sustainable technologies to enable the circular economy and help create a zero-pollution society. He holds a Ph.D. from ETH Zürich (2008) and a professorship at the Norwegian University of Science and Technology (since 2018).

Dagny Aurich is a Ph.D. student in the Environmental Cheminformatics (ECI) Group at the Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, working on the interdisciplinary Luxembourg Time Machine Project (LuxTIME). In 2018, she successfully completed her bachelor's degree in forensic science at the Bonn-Rhein Sieg University of Applied Sciences in Rheinbach, Germany, and continued her research in toxicological forensics at the Legal Medicine in Mainz (Mainz, Germany). Then, she pursued a master's degree in analytical chemistry and quality assurance at the same university and gained practical experience in the Luxembourgish industry in 2020. During this time, she became aware of the University of Luxembourg, where she is now completing her Ph.D. in Historical Exposomics focusing mainly on cheminformatics, high-resolution mass spectrometry, nontarget analysis, environmental history, and data visualization.

Associate Professor Emma Schymanski is head of the Environmental Cheminformatics (ECI) Group at the Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg. In 2018, she received a Luxembourg National Research Fund (FNR) ATTRACT Fellowship to establish her group in Luxembourg, following a 6 year postdoc at Eawag, the Swiss Federal Institute of Aquatic Science and Technology, and a Ph.D. at the Helmholtz Centre for Environmental Research (UFZ) in Leipzig, Germany. Before undertaking her Ph.D., she worked as a consulting environmental engineer in Perth, Australia. She is involved in many collaborative efforts, with more than 100 publications and a book. Her research combines chemoinformatics and computational (high-resolution) mass spectrometry approaches to elucidate the unknowns in complex samples, primarily with nontarget screening, and relate these to environmental causes of disease. An advocate for open science, she is involved in and organizes several European and worldwide activities to improve the exchange of data, information, and ideas among scientists to push progress in this field, including NORMAN Network activities (e.g., NORMAN-SLE), MassBank, MetFrag, and PubChemLite for Exposomics.

Dr. Kerry Sims has a Ph.D. in environmental science. She is currently a Senior Advisor focusing on emerging substances with 16 years of experience at the Environment Agency. Based in the Chemicals Surveillance and Emerging Risks team, she leads on the Environment Agency's Prioritisation and Early Warning System for chemicals of emerging concern, which consolidates environmental monitoring data and horizon scanning work. She is also the Environment Agency scientific lead for international work with the NORMAN network and the Partnership for the Assessment of Risk from Chemicals.

Dr. Sarah Hale has a Master's in green chemistry and a Ph.D. in environmental organic chemistry from the U.K. She is currently a senior researcher at the Norwegian Geotechnical Institute where her work focuses on the fate and transport of PFAS in soil, water, and biota, the remediation of PFAS in the environment, and the way in which this substance class can be regulated. She coordinates the Horizon 2020 Research and Innovation Project ZeroPM: Zero pollution of persistent, mobile substances (Grant Agreement 101036756). She has vast experience with persistent, mobile, and toxic (PMT) substances, and in ZeroPM, she is working with prevention, prioritization, and removal strategies to protect the environment and human health from persistent and mobile substances.

■ ACKNOWLEDGMENTS

The authors acknowledge Jennifer Yang (CAS Registry numbers) and Paul Thiessen and Evan Bolton (PubChem, WIPO patent numbers) for providing the data and/or tips to retrieve the data used in Figure [1] and Egon Willighagen (Maastricht University, Maastricht, The Netherlands) and Jeremy Frey (University of Southampton, Southampton, U.K.) for helping stimulate these discussions. Funding from the European Union Horizon 2020 research and innovation programme under Grant Agreement 101036756, ZeroPM, the Luxembourg National Research Fund (FNR) for Project A18/BM/12341006, and the Institute of Advanced Studies at the University of Luxembourg for project LuxTIME is acknowledged.

■ REFERENCES

(1) Carson, R. *Silent Spring*; Houghton Mifflin: Boston, 1962.

(2) Cousins, I. T.; Johansson, J. H.; Salter, M. E.; Sha, B.; Scheringer, M. Outside the Safe Operating Space of a New Planetary Boundary for Per-and Polyfluoroalkyl Substances (PFAS). *Environ. Sci. Technol.* **2022**, *56* (16), 11172−11179.

(3) Prevedouros, K.; Cousins, I. T.; Buck, R. C.; Korzeniowski, S. H. Sources, Fate and Transport of Perfluorocarboxylates. *Environ. Sci. Technol.* **2006**, *40* (1), 32−44.

(4) Lovelock, J. E. Chapter 1 The Electron-Capture Detector - A Personal Odyssey. *J. Chromatogr. Libr.* **1981**, *20*, 1−11.

(5) Jensen, S. The PCB Story. *Ambio* **1972**, 123−131.

(6) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; Zaslavsky, L.; Zhang, J.; Bolton, E. E. PubChem 2023 Update. *Nucleic Acids Res.* **2023**, *51*, No. D1373.

(7) U.S. Environmental Protection Agency. Technical Fact Sheet: Drinking Water Health Advisories for Four PFAS (PFOA, PFOS, GenX Chemicals, and PFBS). EPA 822-F-22-002; 2022; p 7.

(8) European Commission. Chemicals Strategy for Sustainability Towards a Toxic-Free EnvironmentEuropean Commission (EC). COM(2020) 667 Final; 2020; p 24.

(9) Zumstein, M. T.; Narayan, R.; Kohler, H.-P. E.; McNeill, K.; Sander, M. Dos and Do Nots When Assessing the Biodegradation of Plastics. *Environ. Sci. Technol.* **2019**, *53* (17), 9967−9969.

(10) Hale, S. E.; Neumann, M.; Schliebner, I.; Schulze, J.; Averbeck, F. S.; Castell-Exner, C.; Collard, M.; Drmač, D.; Hartmann, J.; Hofman-Caris, R.; Hollender, J.; de Jonge, M.; Kullick, T.; Lennquist, A.; Letzel, T.; Nödler, K.; Pawlowski, S.; Reineke, N.; Rorije, E.; Scheurer, M.; Sigmund, G.; Timmer, H.; Trier, X.; Verbruggen, E.; Arp, H. P. H. Getting in Control of Persistent, Mobile and Toxic (PMT) and Very Persistent and Very Mobile (VPvM) Substances to Protect Water Resources: Strategies from Diverse Perspectives. *Environ. Sci. Eur.* **2022**, *34* (1), 22.

(11) Davenport, R.; Curtis-Jackson, P.; Dalkmann, P.; Davies, J.; Fenner, K.; Hand, L.; McDonough, K.; Ott, A.; Ortega-Calvo, J. J.; Parsons, J. R.; Schäffer, A.; Sweetlove, C.; Trapp, S.; Wang, N.; Redman, A. Scientific Concepts and Methods for Moving Persistence Assessments into the 21st Century. *Integr. Environ. Assess. Manag.* **2022**, *18* (6), 1454−1487.

(12) Persson, L.; Carney Almroth, B. M.; Collins, C. D.; Cornell, S.; de Wit, C. A.; Diamond, M. L.; Fantke, P.; Hassellöv, M.; MacLeod, M.; Ryberg, M. W.; Søgaard Jørgensen, P.; Villarrubia-Gómez, P.; Wang, Z.; Hauschild, M. Z. Outside the Safe Operating Space of the Planetary Boundary for Novel Entities. *Environ. Sci. Technol.* **2022**, *56* (3), 1510−1521.

(13) Zheng, Z.; Peters, G. M.; Arp, H. P. H.; Andersson, P. L. Combining in Silico Tools with Multicriteria Analysis for Alternatives Assessment of Hazardous Chemicals: A Case Study of Decabromodiphenyl Ether Alternatives. *Environ. Sci. Technol.* **2019**, *53* (11), 6341−6351.

(14) Singh, A. K.; Bilal, M.; Iqbal, H. M. N.; Raj, A. Trends in Predictive Biodegradation for Sustainable Mitigation of Environmental Pollutants: Recent Progress and Future Outlook. *Sci. Total Environ.* **2021**, *770*, No. 144561.

(15) European Commission. ANNEXES to the Commission Delegated Regulation Amending Regulation (EC) No 1272/2008 as Regards Hazard Classes and Criteria for the Classification, Labelling and Packaging of Substances and Mixtures; Classification, Labelling and Packaging. Annex Ares(2022)6485391-19.12.2022; 2022; p 34. https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13578-Hazardous-chemicals-updated-rules-on-classification-labelling-and-packaging_en (accessed 2023-03-30).

https://doi.org/10.1021/acs.est.3c01735
Environ. Sci. Technol. XXXX, XXX, XXX−XXX

VI. Non-Target Screening of Surface Water Samples to Identify Exposome-Related Pollutants: A Case Study from Luxembourg

Authors:

Dagny Aurich , Philippe Diderich, Rick Helmus, Emma L. Schymanski

Manuscript type:

Research paper

Contribution statement:

DA and ELS conceptualized the study. RH contributed the changes to the software and **DA** developed the methodology. **DA** conducted the formal analysis, main investigation, visualization (all figures) and full writing of the original draft. PD provided sources (water samples) and did parts of the investigation. All authors were responsible for review and editing. ELS was supervising, and acquired funding and provided resources.

Submission status:

Published in Environmental Sciences Europe on 07 November 2023.
(DOI: 10.1186/s12302-023-00805-5)

Short summary/ Contribution to the field:

This research article focuses on non-target screening of surface water samples to identify temporal patterns in exposome-related pollutants. In the study 271 samples from 20 sites in Luxembourg were analyzed using high resolution mass spectrometry and an open source R-package called *patRoon*. By employing scoring terms and utilizing supplementary database information, tentative identifications of chemical compounds were prioritized based on spectral similarity. Moreover, potential threats posed by these chemicals were assessed using *PubChemLite* annotation categories and *classyFire* classification software, enabling prioritization for future confirmation and data analysis. The article closes by recommending 41 chemicals (of the 378 tentative identifications) for further confirmation and potential inclusion in routine monitoring efforts via targeted approaches by Luxembourgish governmental institutions. Further confirmation was beyond the scope of the article, since the targeted analysis is performed within the remit of AGE and not LCSB.

Aurich *et al. Environmental Sciences Europe* (2023) 35:94
https://doi.org/10.1186/s12302-023-00805-5

Environmental Sciences Europe

RESEARCH

Open Access

Non-target screening of surface water samples to identify exposome-related pollutants: a case study from Luxembourg

Dagny Aurich[1*], Philippe Diderich[2], Rick Helmus[3] and Emma L. Schymanski[1*]

Abstract

Background Non-target screening of surface water samples collected over an extended period can reveal interesting temporal patterns in exposome-related pollutants. Additionally, geographical data on pollution sources close to the sampling sites, chemical classification data and the consideration of flow paths can provide valuable information on the origins and potential threat of tentatively identified chemical compounds. In this study, 271 surface water samples from 20 sampling sites across Luxembourg were analysed using high-resolution mass spectrometry, complementing routine target monitoring efforts in 2019–2022. Data analysis was performed using the open source R-package *patRoon*, which offers a customizable non-target workflow. By employing open source workflows featuring scoring terms, like spectral match and applying identification levels, tentative identifications can be prioritized, e.g. based on spectral similarity. Furthermore, by utilizing supplementary database information such as *PubChemLite* annotation categories and classification software such as *classyFire*, an overall assessment of the potential threats posed by the tentatively identified chemicals was conducted, enabling the prioritization of chemicals for future confirmation through targeted approaches.

Results The study tentatively identified 378 compounds associated with the exposome including benzenoids, organoheterocyclic compounds, and organic phosphoric acids and derivatives (11 *classyFire* superclasses, 50 subclasses). The classification analysis not only revealed temporal variations in agrochemicals, with the majority of identifications occurring in May to July, but also highlighted the prevalence of pharmaceuticals such as venlafaxine in surface waters. Furthermore, potential sources of pollutants, like metallurgic industry or household products were explored by considering common uses and geographical information, as commercial uses of almost 100% of the identified chemicals are known. 41 chemicals were suggested for potential inclusion to governmental monitoring lists for further investigation.

Conclusions The findings of this study complement existing knowledge on the pollution status of surface water in Luxembourg and highlight the usefulness of non-target screening for identifying temporal and spatial trends in pollutant levels. This approach, performed in a complementary manner to routine monitoring, can help to tentatively identify chemicals of concern for potential inclusion in target monitoring methods following additional confirmation and quantification efforts.

*Correspondence:
Dagny Aurich
dagny.aurich@uni.lu
Emma L. Schymanski
emma.schymanski@uni.lu
Full list of author information is available at the end of the article

Keywords Surface water, High-resolution mass spectrometry, Non-target analysis, Exposome, Cheminformatics, Luxembourg

Background

The variety and number of chemicals of concern in the environment continue to rise. The synthesis and registration of new chemicals happens regularly [1] and their presence in the environment and potential impact are often only recognized at a late stage. Furthermore, transformation products (TPs) of well-known chemical compounds frequently go unrecognized and are often not subject to routine monitoring. While they often exhibit properties similar to their parent, they can be even more harmful or persistent. This is for example true for the pesticide dichlorodiphenyltrichloroethane (DDT) and its TPs dichlorodiphenyldichloroethylene (DDE) and dichlorodiphenyldichloroethane (DDD) with all three compounds having carcinogenic properties and DDE being even more potent than its parent [2, 3]. In general, the origins of pollutants vary, as they can be side products of industrial, agricultural or medical applications or chemicals resulting from households ending up in the environment. Determining the source of each pollutant is therefore a highly challenging task but can be achieved in several ways, including looking at specific samples and sources or through evaluation of long-term temporal or geographical patterns.

Environmental monitoring covers just a small fraction of concerning chemicals, such that many even well-known chemicals ('known unknowns' [4, 5]) remain undiscovered in routine analysis. Therefore, it is important to take a step back from trusting solely targeted approaches and to apply and improve existing non-target (NT) methods. State-of-the-art chemical analysis methods combined with optimized cheminformatics workflows can provide a more comprehensive picture of the chemicals contained in a variety of environmental samples. The chemical load and the potential effect on human health or wildlife are of major interest when it comes to, e.g. chemical exposure assessment and exposomics research. Non-target analysis (NTA) allows for the monitoring of chemical exposure by identification of potentially toxic and persistent chemicals, even retrospectively. Looking at the full range of chemicals in sample measurements done for several years, time trends can be discovered (e.g. seasonal changes [6, 7]) and information could be found on the source and pattern of the pollution. Suitable sample types to perform such an analysis can range from soil, water and air to biological specimens, reviewed in more detail previously [8]. Looking at the availability of samples and the sampling or analysis requirements, monitoring the state of water bodies presents a convenient way to measure environmental pollution. Especially looking at different flow paths contributing to a spread of pollutants, the input of wastewaters and meteorological phenomena such as flooding events, water becomes a very interesting and versatile sample type.

The small country Luxembourg, with an industrial past, is covered by 102 natural (or nearly unmodified) surface water bodies, displayed in Fig. 1 with different river catchments and flow directions. Pollution sources are numerous, in addition to industry, agricultural activities, traffic and household waste contribute to the country's chemical load [9]. Chemicals such as pesticides, flame retardants like polybrominated diphenyl ethers (PBDEs) or pharmaceuticals are contaminating water bodies [10–12], resulting in poor quality evaluations [13]. In 2022, the report on water quality showed that no river in the country is in a good condition (according to European "one-out-all-out" criteria [13, 14]) with

Fig. 1 Surface water bodies in Luxembourg with different catchments and flow directions. Modified from: geoportail.lu

concentration values exceeded for chemicals like perfluorooctanesulfonic acid (PFOS), metazachlor or anthracene [12]. Moreover, only half of the groundwater bodies were in good status in 2022 [9]. In these evaluations, the ecological condition of surface water bodies was evaluated based on biological, physical, chemical and hydromorphological parameters [9, 13]. The chemical status was based on the analysis of priority hazardous substances and substance classes of greatest (EU)-concern, specified in the Water Framework Directive Annex VIII of Directive 2000/60/EC (Annex X) [14]. The directive was transposed into Luxembourgish law by the amended Water Directive of 2008 (Directive 2008/105/EC), known as the Environmental Quality Standards or Priority Substances Directive, setting environmental quality standards for surface water pollutants [15]. In 2013 an additional European watchlist mechanism was established, setting a list of substances to be monitored by all EU members [16, 17], which is updated regularly (newest version 2022 [18]). In 2022, the EU commission adopted a proposal to revise the list of priority substances, including 25 additional substances, e.g. per- and poly-fluorinated compounds (PFAS), bisphenol A, silver and several pesticides and pharmaceuticals [16, 19]. Besides looking at priority substances, a list of catchment-specific pollutants was considered, looking at the main catchments Mosel, lower Sûre, upper Sûre, Wiltz, Our, Alzette and Chiers (see Fig. 1) [12]. Cross-border rivers, such as the Alzette (arising from France), were particularly polluted [9], already containing a high chemical load when crossing the border (see Fig. 1, blue arrows indicating the flow direction).

This analysis can be expanded looking not only at the specified compounds for routine governmental monitoring, but also performing a full NT workflow to obtain an overview of additional chemicals that may be present, but are not a part of routine monitoring efforts yet. Performing this additional screening in parallel to the routine target analysis can help to identify risks posed by new or undiscovered contaminants in Luxembourg already at an early stage. This can be then used to guide legislative decisions, e.g. to expand the list of chemicals used for target screening in routine governmental laboratories (in Luxembourg the L'Administration de la Gestion de l'Eau (AGE)) or to improve wastewater treatment filtering systems. One step further, NTA can help support determining the (geographical) origin of contamination through differences between samples (e.g. influent/effluent of wastewater treatment plants (WWTPs) [20, 21]). Even the description of processes or transformation pathways may be possible, when looking at these measurements [22].

Currently, there is not one unified workflow to perform NTA of surface water samples. Several tools are available nowadays to screen for unknowns, but there is no one-fits-all solution. High-resolution mass spectrometry (HRMS) was the method of choice to perform the analysis of the 271 Luxembourgish surface water samples collected at 20 sampling sites in Luxembourg between April 2019 and April 2022. In addition to the routine target monitoring conducted by AGE, a NT data analysis was performed allowing for retrospective analysis, screening, for e.g. previously not discovered chemicals. In this article, the open source R-package called *patRoon* [23, 24] was used, to perform the NTA. A detailed description of the package and its functionalities can be found in Helmus et al. 2021 and 2022 [23, 24]. The package combines functionalities of many tools like XCMS [25–27] or MetFrag [28] in one 'ready to use' package to harmonize and simplify the workflow of data processing of HRMS data in environmental sciences. The utilization of this open source tool offers a potential solution for facilitating collaboration among researchers, considering the existence of various instrument types and the consequent use of different and often incompatible software. It works with the open mzML format, for which a conversion exists for each vendor.

The workflow used in this study presents only one possible combination of steps (see Methods) to perform a NTA and can be expanded using, e.g. a suspect list to screen for specific substances. A novelty of *patRoon* 2.0 (compared to *patRoon* 1.0) was the possibility to perform a 'Sets' workflow processing positive and negative analyses at once, which was applied here [23]. The package workflow included peak picking, selecting relevant features, blank correction, removing the irrelevant ones and peak alignment. For the feature finding (via XCMS [25–27]) an optimization of the input parameters was performed. As shown by Libiseller et al. in 2015 [29], Albòniga [30] or Tostengard and Smith [31], optimizing for the parameters *ppm* and *peakwith* can significantly improve the results when using XCMS and the integrated *centWave* algorithm. *CentWave* is a feature detection algorithm integrated in the XCMS package applying continuous wavelet transformation (addition of Gauss-fitting is possible) to detect features even when they are partially overlapping [26]. After feature finding componentization could be performed identifying features belonging to one compound (adducts, isotopes, in-source fragments) as well as the generation of potential chemical formulas based on accurate mass and isotope patterns. For interpretation and possible identification of features, substance and spectral databases such as *PubChem* [32] and *MetFrag* [28] are required. In this study, a subset of *PubChem* called *PubChemLite* [33, 34] was used, focussing especially on exposomics-related compounds.

The resulting features were categorized by identification level [35] using the *individualMoNAScore* (spectral similarity of the candidate structure in the MassBank of North America (MoNA) [36]). Moreover, the compounds were classified using the tool *classyFire* [37] whose primary function is to classify chemical compounds based on their structural features and properties and assign them to specific chemical classes and subclasses. *ClassyFire* is widely used and can provide valuable information about the chemical composition, functional groups, and potential biological activities of a compound. In addition, *PubChem* annotation content [34] was used to estimate the environmental effect and to determine possible sources of the chemicals like agriculture, households or industry. *PubChemLite for exposomics* makes use of selected categories available in the *PubChem* Table of Contents Classification Browser [38]. These categories can help categorizing the tentative identifications. Categories used in this study included agrochemical use (*agroChemInfo*), drug and medication information (*drugMedicInfo*), associated disorders and diseases (*disorderDisease*) and use and manufacturing (*knownUse*) [34].

A comparison to prior studies of the Environmental Cheminformatics (ECI) group at the University of Luxembourg—focussing at pesticides and pharmaceuticals in surface water [10, 11]—was conducted, looking at shared identifications to check for plausibility. These studies proved already the presence of high pesticide (even banned compounds and transformation products) and pharmaceutical load in surface waters and complemented target monitoring efforts of AGE [10, 11] but did not look beyond these classes. For the river Chiers, located in an industrial region in the south-west of Luxembourg, an additional comparison was performed looking at compound findings from a 2022 sampling campaign, see Fig. 2. The sampling was performed from May to June 2022 at the inlet of a WWTP in Petange, located prior to the sampling point of this study. The study results were then evaluated based on the geographical information on industry, households and hospitals located in the region. Finally, the AGE target list was compared to the study findings, looking as well at catchment-specific pollutants, and discussing possible candidates to include in routine monitoring.

The primary aim of this article is to employ NTA as a complementary approach to routine target monitoring, with the objective of offering tentative insights into chemicals of concern that are currently not under surveillance to form recommendations for future target monitoring efforts. To achieve this, an open source and adaptable NT workflow is proposed as an alternative to conventional vendor software. By adopting this approach, it becomes feasible to accommodate a wide range of instruments, thereby facilitating collaboration among researchers and authorities/regulators. This applies specifically to the Luxembourgish case as AGE uses a Sciex

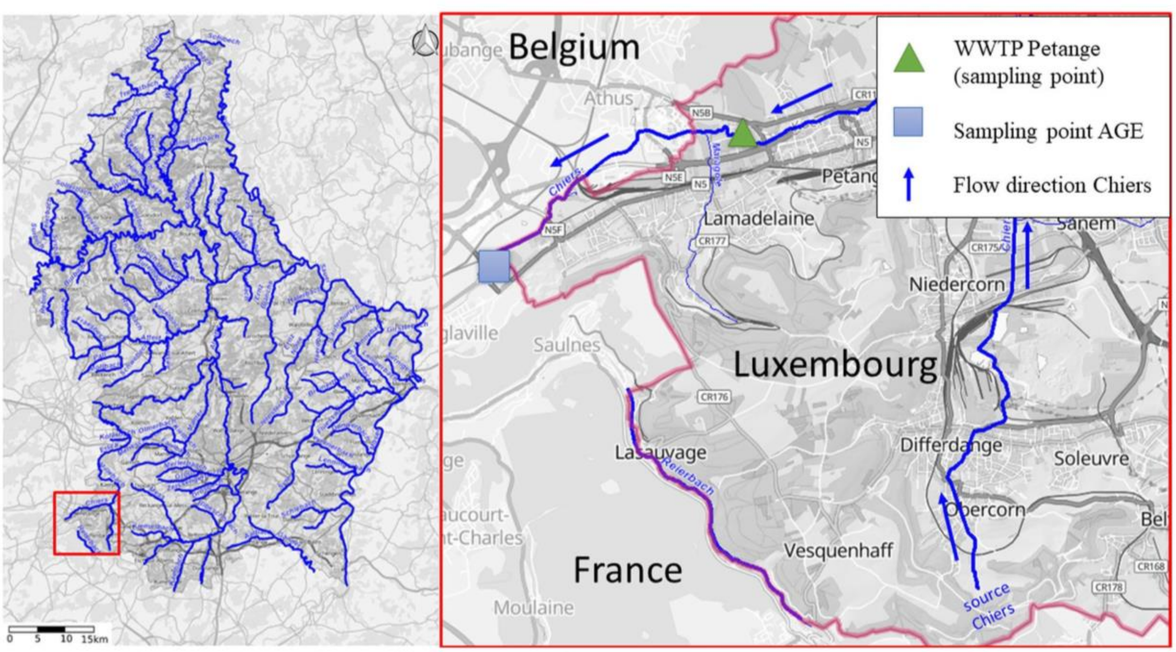


Fig. 2 Two sampling sites at the river Chiers. Modified from: geoportail.lu

and the University of Luxembourg a Thermo Fisher Orbitrap device. The combination of various tools discussed herein represents a singular, potential method for processing HRMS data via NTA. Furthermore, this study aims to compare NT-HRMS measurements obtained from samples of Luxembourgish surface water in order to identify temporal and/or spatial patterns and to classify chemicals found using a variety of tools. Based on the results, the article aims to explore potential sources of pollutants, as well as estimate potential impacts on both the environment and human health.

Methods

Sample preparation and analysis

Two hundred and one water samples were collected every 4 weeks by AGE at the sampling sites indicated in Fig. 3 and in Additional file 2: Table S1. The 3-year sampling analysed in this article took place between April 2019 and April 2022 with each year having varying sampling points on a rotational basis, spread throughout Luxembourg, selected by AGE. The same four sampling points were analysed every year, the remaining river locations vary in a 3-year cycle to cover different geographic regions (or catchments) in the country (see Fig. 1). The four constant rivers—displayed as black squares in Fig. 3—were Chiers, located in the south-west (Chiers catchment), Syr in the east (Mosel catchment), and in the centre Sûre (upper Sûre catchment) plus the Alzette near Ettelbruck, hereafter 'Alzette_E' (Alzette catchment). Some measurements were unavailable due to differing reasons, *e.g.* meteorological circumstances (excluded months: November 2019–March 2020). For March 2021 and 2022 there was an additional sampling performed at the end of the month, indicated as 'Mar_end'. Due to method and instrument instabilities, several months of 2021 were remeasured in 2022, resulting in increased feature and identification numbers for the remeasured analyses 'Mar_end-21, Apr-21, May-21'.

The 271 surface water samples were extracted (solid phase extraction, SPE) as described in Krier et al. [10] using the Atlantic® HLB SPE Disks (Horizon, Salem, NH, USA) with a 47 mm diameter and the SPE-DEX 47900 system (Horizon). The filtered extracts were spiked with a 100 ppb mix of 10 internal standards: Melamine-13C3-15N3, Carbendazim-D4, Sucralose-D6, 5-Methyl, Benzotriazol-D6, Neotame-D3, Metolachlor-D6, 5-Fluorouracil-15N2C13, Torsemide-D7, Triclosan-D3, Carbamazepine-D10 purchased from Santa Cruz Biotechnology, Heidelberg. Then they were analysed via Reversed Phase LC-HRMS using the Waters Acquity UPLC BEH C18 column and the Thermo Q Exactive HF Orbitrap Mass Spectrometer. Further analytical details

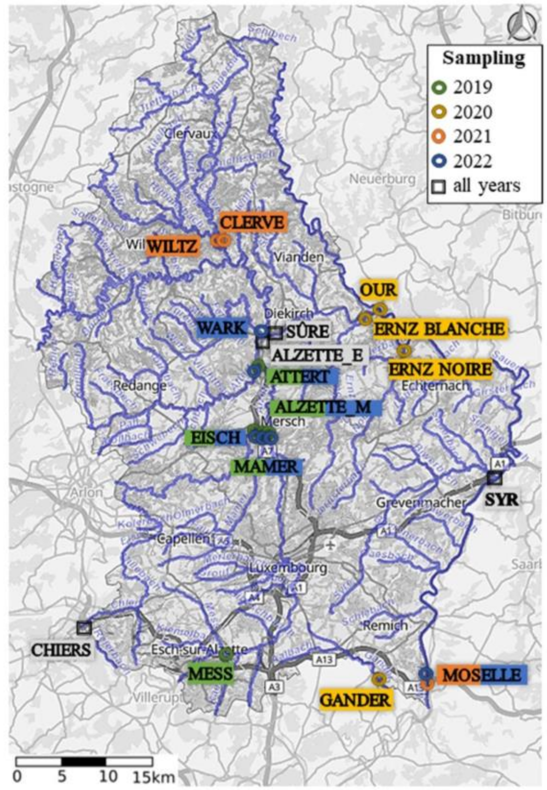


Fig. 3 Sampling sites between 2019 and 2022 (exact location in Additional file 2: Table S1). Mixed shading corresponds to sampling done in both years. Modified from: geoportail.lu

including QA/QC procedures are given in Krier et al. 2022 [10].

Data analysis

Several R-packages were used to perform the HRMS data processing and the following data analysis steps. The version of R and all installed dependencies of the open source package *patRoon* and other used packages are listed in the LCSB GitLab repository of the Environmental Cheminformatics Group (ECI). Moreover, the R script used for optimization and the full NT script can be found in a subfolder of the repository, while the raw data are available on GNPS (https://doi.org/10.25345/C55X25P62).

This study presents an NTA workflow of 271 LC–HRMS surface water measurements (in positive and negative mode), making use of a modifiable data processing workflow established by using the R package *patRoon*. The data files were converted to mzML via *ProteoWizard's MSConvert* (version 3.0.21075) [39, 40] using a peak picking (centroiding) filter. Therefore, the pre-treatment step implemented in *patRoon* was not applied here, starting directly with the finding and grouping of features (functions findFeatures, groupFeatures) in 'Sets' mode. The package *XCMS* [25–27] was used to perform feature finding and later grouping. To receive best results for the feature dataset a feature optimization step was performed (as explained above) with the *patRoon*-integrated Isotopologue Parameter Optimization (IPO) [29] algorithm [function optimizeFeatureFinding resulting in an iterative process using Design of Experiments (DoE)]. The feature grouping using *XCMS* was followed by a basic rule-based filtering operation (filter) applying a blank, intensity and replicate group filter. To annotate those features, tables of averaged mass spectra (MS and MSMS) for each feature—so called *MS peak lists*—were created (generateMSpeaklists) using functionalities of the *mzR* package [39–43]. Those lists were filtered thereafter (filter) limiting the results to the top 25 MSMS peaks. Compounds were then generated using *MetFrag* [28] and the *PubChemLite for exposomics* library [33, 34] (generateCompounds). For each feature group possible candidate compound structures were identified and then ranked, *e.g.* based on the matching fragmentation (MSMS) data. Several scoring parameters can be set in this step, including the *individualMoNAScore*, which was used in the next step to determine the level of identification. For the simplicity of analysis, three levels were chosen: a good MSMS library match, i.e. level 2 scored at least 0.9, a fair match, i.e. level 3a lay between 0.7 and 0.9 and level 3b was defined to be between 0.4 and 0.7 (adapted scheme from the NTA study by Talavera Andújar et al. [44]). For each feature, the tentatively identified candidate with the highest score was selected for the final scoring and reporting. The following data analysis involved a classification of chemicals using the web interface *classyFire* [37] and four classification categories of the *PubChemLite* database (*agroChemInfo, drugMedicInfo, disorderDisease, knownUse*), which are available in the database file and online for each chemical record. The inter- and intra-year occurrence of compounds and compound classes was then analysed. The results were evaluated looking at the 3 years of measurements, presented in the Results, followed by a critical discussion and evaluation of the used tools, in the Discussion.

Results

This section includes the summarized results from the NT workflow of the 271 Luxembourgish surface water samples analysed that are of solely qualitative nature. The workflow started with the optimization of the *ppm* and *peakwidth* parameters to perform feature finding, as described above. An example of different DoEs visualized by perspective plots can be found in Additional file 1: Figure S1 for the samples of April 2020 (in negative mode). In addition, a visualization of the best parameters determined for positive and negative mode for the same month is shown in Additional file 1: Figure S2. The full list of optimized feature finding parameters for *ppm* and *peakwidth* can be found in Additional file 2: Table S2.

After optimizing the feature finding parameters, the actual NT analysis of the measured samples was performed. Figure 4 shows the applied *patRoon* workflow with data collected for the ten April 2020 samples. In total 75,263 positive and 43,697 negative features were found in the first step of the workflow, totalling to 118,960 features. After feature grouping and filtering, using the inbuilt patRoon functionality [23, 24] (see above), the number was reduced to 24,005 features in 7,581 feature groups. After the generation and filtering of MS peak lists, 15,140 positive compounds and 12,546 negative compounds could be assigned to the feature groups (see Fig. 4). Applying the identification scheme explained in the Data Analysis section [44], 76 positive and 73 negative compounds could be identified at levels 2, 3a and 3b, of which 93 were unique compounds and 56 were overlapping (i.e. they were tentatively identified in both positive and negative mode).

Most of the rivers are interconnected in Luxembourg and therefore the same compounds appear in several measurements. There are catchment-specific pollutants—monitored by AGE—appearing mainly in the regions indicated in Fig. 1 [12]. Figure 5A shows overlapping features (using a Venn diagram) for the four rivers monitored regularly. The most feature groups were detected for the river Syr, which overlapped most with the surface water from Chiers and Alzette_E (975). However, all four rivers are located in different catchments with different, region-specific influences and therefore the overlap is not 100%. In Fig. 5B, a Chord plot for all feature groups in all rivers in April 2020 is presented. All rivers showed several overlapping feature groups with clear overlaps of some rivers belonging to one catchment, *e.g.* Gander and Mosel. However, this is not always the case, looking, *e.g.* at the two rivers in the Lower Sûre catchment or the large overlap between Alzette_E and Chiers.

The analysis steps presented in Fig. 4 were accordingly performed for all 34 months and the resulting tables can be found in the GitLab repository. In Additional file 2: Tables S3 and S4 the number of 2, 3a and 3b identifications for positive and negative mode, their sum and the number of tentatively identified unique compounds per level can be found. There was a majority in level 2 identifications compared to the level 3 numbers, *e.g.* for the April 20 samples there were 58 level 2s, 22 3as and 17 3bs. The total number of positive, negative and unique identifications

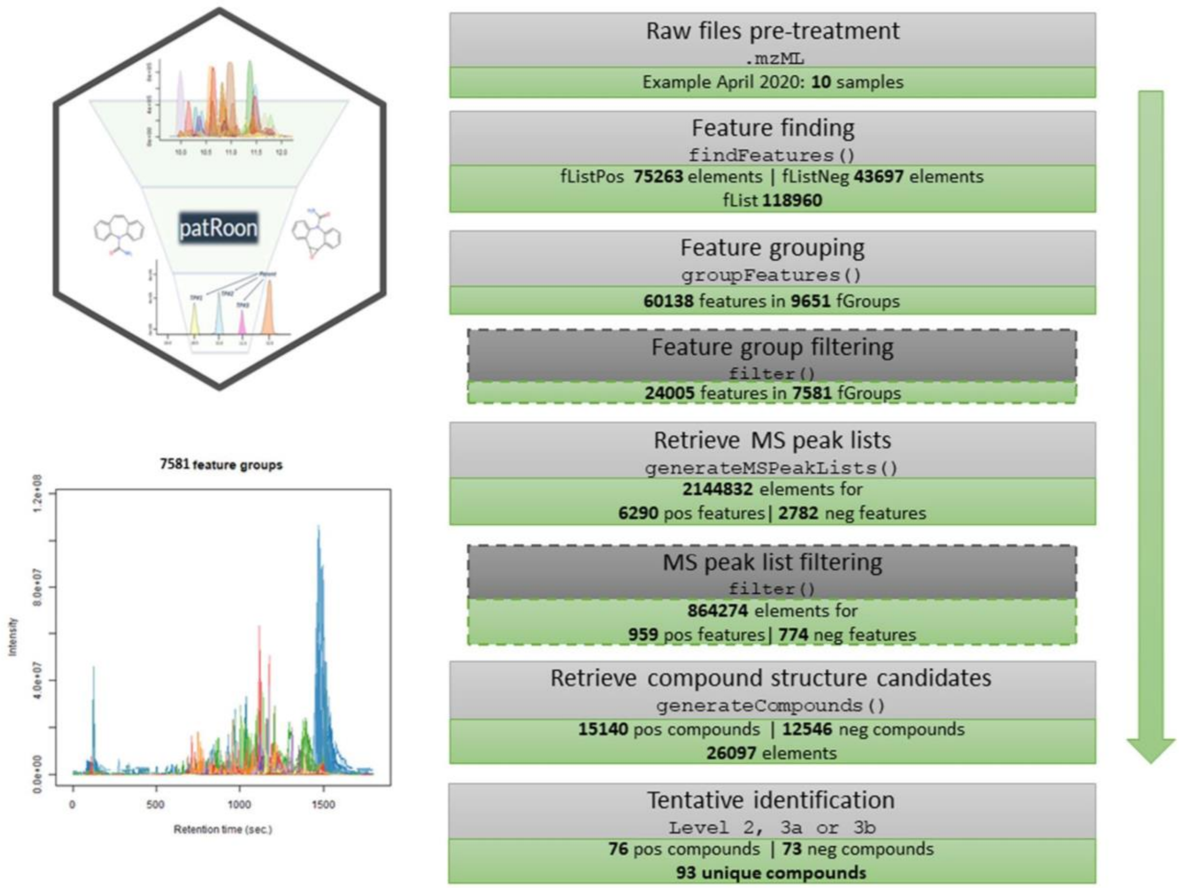


Fig. 4 patRoon workflow (workflow step terms described in [23, 24]) with exemplary values and feature groups plot for the April 2020 analyses, resulting in level 2, 3a or 3b identifications

(without discriminating between levels) is demonstrated in Fig. 6 based on Additional file 2: Table S4. The numbers of positive and negative (unique) identifications are presented in yellow and green and a black bar shows the total number of unique compounds. The count of identifications in positive mode is generally higher than the negative count and their overlap is shown in blue. Overall, a total of 2479 compounds were annotated. After deduplication, 378 unique chemicals remained with level 2, 3a and/or 3b. The chemicals identified per month and in total can be found in Additional file 2: Table S5 (further details are available in the 'Output_summary_patRoon' folder uploaded on GitLab), including tentative identifications of pharmaceuticals like valsartan or metformin, agrochemicals like 4,6-dinitro-o-cresol (DNOC) or their TPs like Flufenacet ESA and industrial chemicals like benzotriazoles, methylbenzenesulfonamide or bisphenol S.

Classification

To get a better overview of and group/interpret the tentative identifications, classification steps were performed.

ClassyFire

First, an 'interannual' (April results of all years) and an 'intraannual' (2021 results of all months) comparison was performed, looking at the number of identified compounds per *classyFire* class and parent class (superclass). The month April was the only one measured in all years and 2021 was the only year where samples were available for each month. In general, for the interannual and intraannual comparison, 11 main parent classes (superclasses) could be identified: organic oxygen compounds,

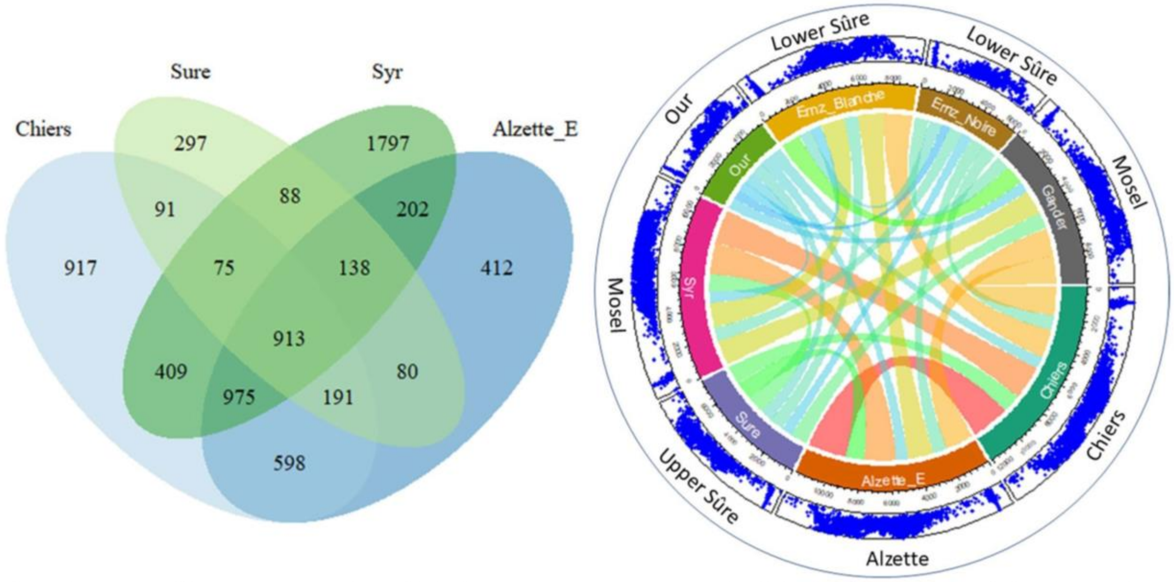


Fig. 5 **A**: Venn plot of feature groups in April 2020 for the four rivers monitored all years; **B**: chord plot for all feature groups of all April 2020 analyses with river catchments

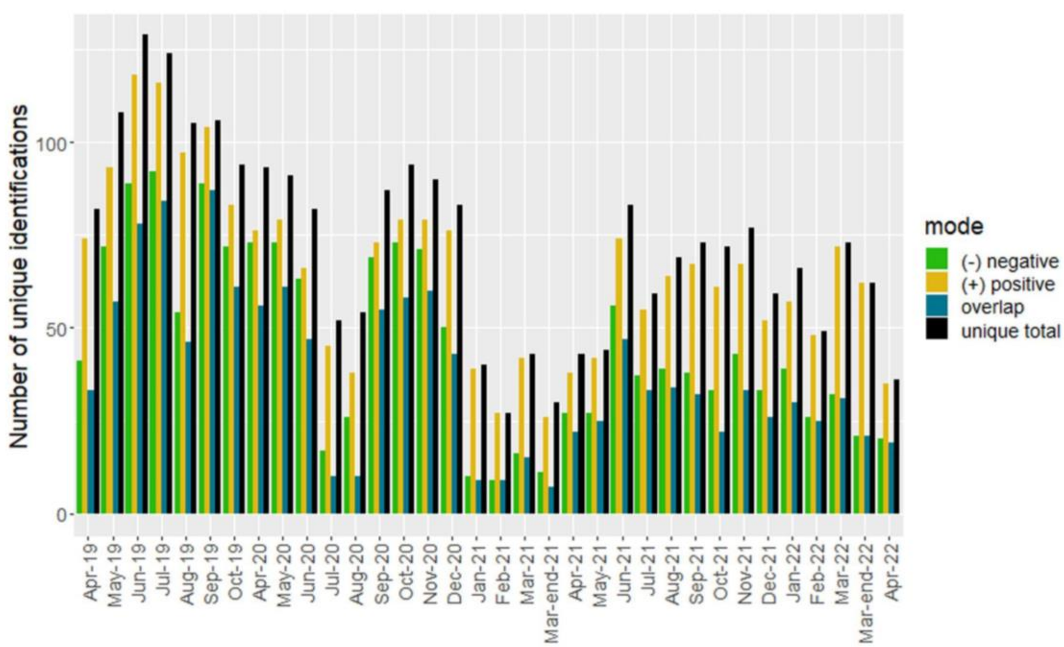


Fig. 6 Number of unique positive, negative, overlapping and total identifications per month

organohalogen compounds, nucleosides, nucleotides, and analogues, organic nitrogen compounds, organosulfur compounds, lipids and lipid-like molecules, alkaloids and derivatives, benzenoids, phenylpropanoids and polyketides, organic acids and derivatives and organoheterocyclic compounds. 50 unique subclasses of those very general superclasses could be assigned (46 in 2021), giving a more detailed picture. The underlying data (total numbers and percentage of compounds found per class and superclass in the inter- and intra-annual comparison) are included in Additional file 2: Table S6. An overview of those compound classes can be seen in Fig. 7 using the summarized identification numbers of all analysed months in 2021.

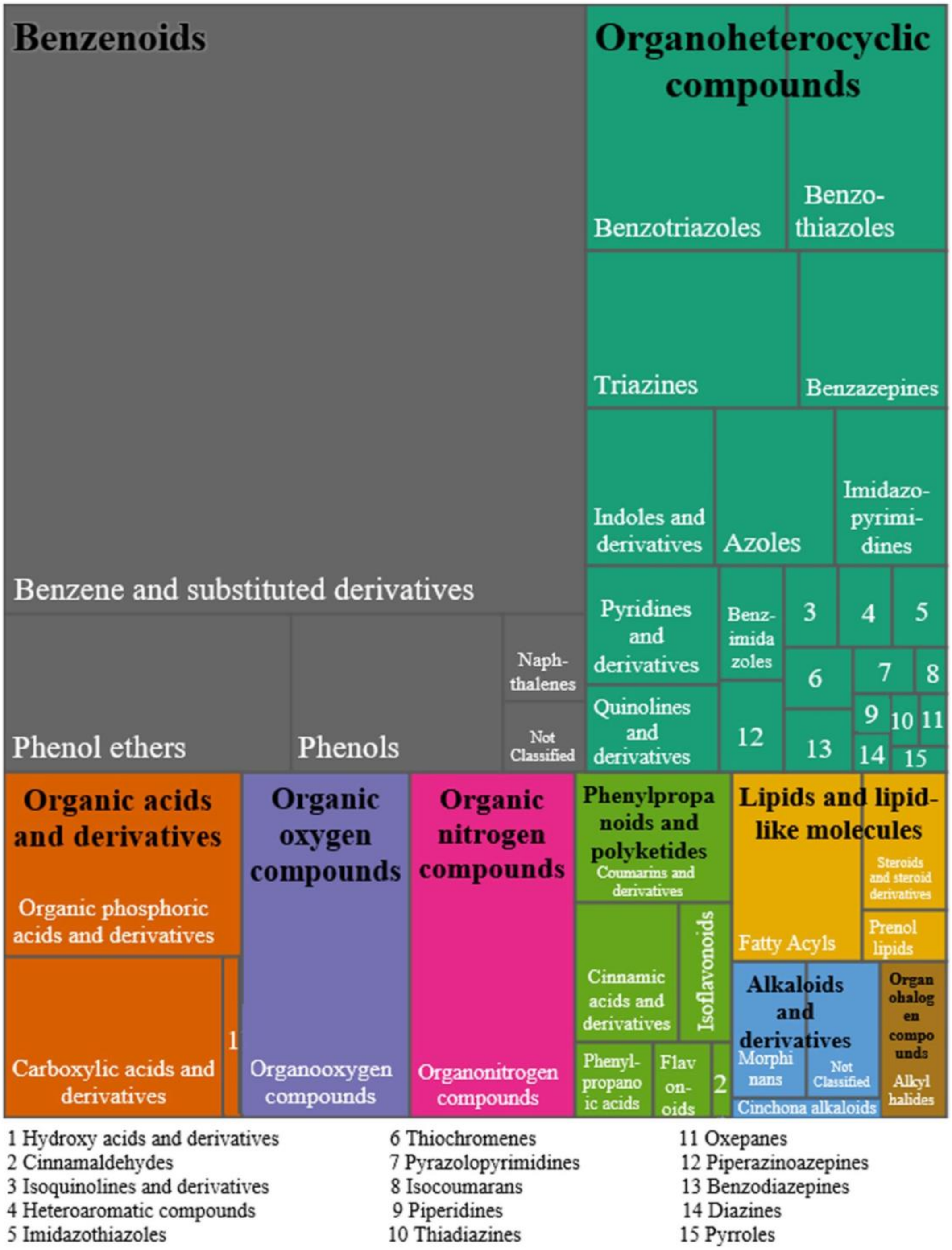


Fig. 7 Treemap of classyFire classes and subclasses using the summarized identifications of 2021

The treemap in Fig. 7 shows that nine superclasses with several subclasses could be identified for the intraannual comparison of measurements in 2021. Most of the chemicals were categorized as benzenoids (43%) followed by organoheterocyclic compounds (26%) and organic acids and derivatives (8%). Comparing the intraannual results of 2021 with the interannual comparison of the month April between 2019 and 2022, additional chemical classes were observed. One purine nucleoside, one sulfoxide and one compound belonging to the pteridines and derivatives class were tentatively detected in 2019. Purine nucleosides are generally not considered to be harmful to the environment or human health, as they are essential components of normal cellular functioning. Some sulfoxides have been shown to have toxic effects (e.g. dimethyl sulfoxide, DMSO), particularly when they are not properly disposed of or when they enter the water supply [45]. However, looking at the measurement results of April 2019, the compound was sulforaphane (in positive mode) at the sampling points Alzette_E, Syr, Mess, Mamer, Attert and Alzette_M (Alzette sampling point Mersch-Berschbach), which is a naturally occurring compound that is safe for human consumption and is used in cancer treatment. The same applies for pteridines and derivatives, some chemicals of this class have been shown to have toxic effects (e.g. atrazine), but the identified compound was in this case riboflavin, also known as vitamin B2. Overall, these examples show (and it is important to remember) that the toxicity of a chemical is complex and context-dependent, and should be evaluated on a case-by-case basis. Generally, the toxicity assessment in terms of environmental and health hazards is difficult, as the toxicity of a chemical can depend on a variety of factors, including its chemical structure and specific chemical properties, concentration, mode and duration of exposure, and the susceptibility of the organism or sensitivity of the ecosystem exposed. As the concentration is not measured in this study, little can be said about the toxicity of the annotated chemicals and the chemical class as such gives only limited to no information about the environmental or health impact. Additionally, different chemical classes can have different toxicities for different organisms, and different endpoints (such as acute toxicity, chronic toxicity, carcinogenicity, mutagenicity, and reproductive toxicity) may also be relevant. It has to be considered that some compounds may have multiple classifications, and their potential impact on the environment and human health may vary depending on the specific application. The use of *classyFire* is examined further in the Discussion.

PubChemLite categories

To identify possible sources and estimate the environmental impact of the exposome related chemicals, a classification of the compounds in the inter- and intraannual comparison was performed, using the annotation content available in *PubChem* for each chemical (via the *PubChemLite* categories described in [34]). The categories *agroChemInfo* and *drugMedicInfo* were chosen to evaluate trends of agrochemical and pharmaceutical use in 1 year and over 3 years. Moreover, information about possible disorders and diseases related to a compound and known commercial uses were analysed using the *disorderDisease* and *knownUse* categories. The resulting total and percentage trends are visualized using four line charts in Fig. 8 and the raw numbers are summarized in Additional file 2: Table S7. It has to be considered that the categories identified are not exhaustive, and there may be some overlap between them (multiple uses per chemical).

Looking at the intraannual comparison of all months in 2021 an overall increase of total numbers in all categories could be monitored, but the overall percentage (relative to total numbers) stayed roughly the same. A majority of the chemicals had associated disorders and diseases content in PubChem (between 67 and 79%), while 53% (July) to 72% (March) of the chemicals were assigned to the class of pharmaceuticals and the percentage of agrochemicals was between 7% (March) and 32% (May). This corresponds to the usual 'spraying rhythm' of farmers who increase pesticide and herbicide spraying in May to lay a foundation for the harvest. Almost all identified chemicals (93–100%) had a documented use, with multiple matches per compound when looking at the individual case in *PubChem*. The interannual values showed a sharp decrease of total identifications in drugs, disorders and diseases and known use, either due to effects of the COVID pandemic or due to measurement variations (less likely as the agrochemical curve stayed more or less constant). The percentage values (% of total identifications) showed a constant trend between the years with nearly all annotated compounds having a known use, 75–79% being associated to disorders and diseases, ~ 70% being drugs and 12–22% agrochemicals according to the PubChemLite classification.

Comparison to other studies

Besides using classification workflows, data from other water studies can be used to determine possible sources of exposome related chemicals. Former studies looking at Luxembourgish surface waters provided evidence that there are more pharmaceuticals and agrochemicals entering the environment than those included in the target monitoring by AGE that could potentially cause harm

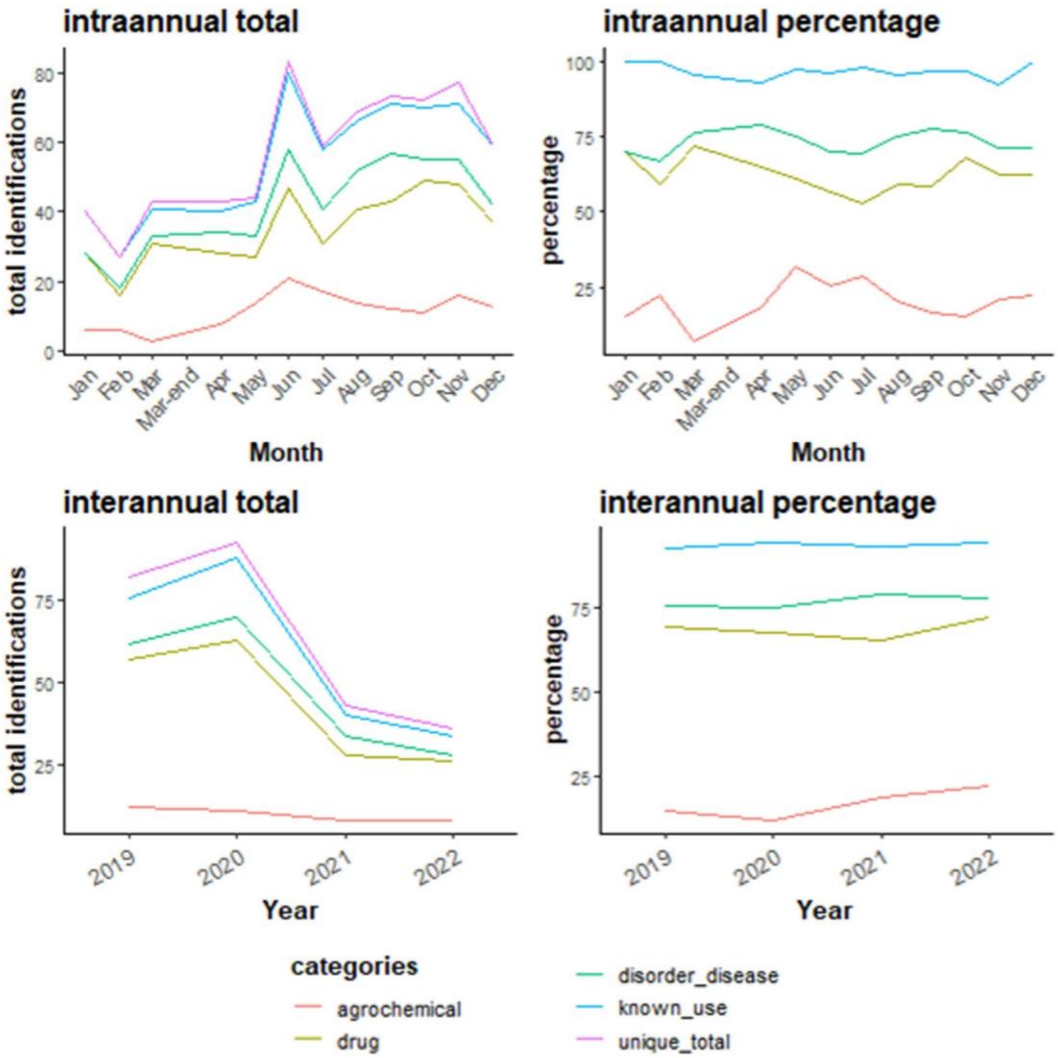


Fig. 8 Classification of tentatively identified compounds in the inter- and intra-annual comparison, according to PubChem

[10, 11]. Regarding agrochemical compounds in Luxembourgish rivers, a suspect and related transformation product screening study was conducted by Krier et al. [10] with the same instrumental methods and a subset of the data used here. The study identified 162 pesticides and 96 TPs in the water samples (several chemicals not allowed in Luxembourg). 31 chemicals were confirmed at level 1 [10]. Comparing these results to this study an overlap of 36 agrochemicals was seen, listed in Additional file 2: Table S9. Since that study focused exclusively on pesticides, which are often present in lower concentrations than, *e.g.* pharmaceuticals and industrial chemicals, it is likely that several compounds identified by Krier et al. may have other top-ranked candidates in the current study as they have been prioritized using different scoring terms (shown to improve ranking results [34]) that were not available to Krier et al. at the time. Singh et al. [11] performed a suspect screening, identifying 94 pharmaceuticals, adding quantification steps later. The AGE monitoring however, included just five pharmaceuticals (list of AGE from 2019 and 2020 in Additional file 2: Table S8): carbamazepine, diclofenac, ibuprofen, ketoprofen, and lidocaine. All five chemicals were identified in the work of Singh et al. as well as in this study. Of the 232 pharmaceuticals tentatively identified in this study, 58 were also confirmed in the results of Singh et al. [11], including the 5 covered in the AGE monitoring. The compared lists and overlapping identifications

are summarized in Additional file 2: Table S10. Singh et al. also registered the trend of decreasing pharmaceutical load looking at the years 2019 and 2020, explaining it with the reduction of medical treatments due to the COVID pandemic and lower precipitation [11].

A combination of geographical information, information on flow paths and additional measurement data from the inlet of a WWTP was used to analyse possible sources of the chemicals found in the river Chiers. The river is located in the south-west area of Luxembourg, at the border to France, with exposure to a set of different sources of pollution (see Fig. 9). Its source is in Obercorn, it passes the WWTP in Petange (green) and 6 km later the sampling point of this study (blue), located at the border to France.

The measurements at the WWTP took place between May and June 2022, resulting in 409 tentatively identified chemicals. Comparing those findings to this study (all results from Chiers in 2019, 2020, 2021 and 2022), an overlap of 178 chemicals could be identified. Figure 10 shows the number of overlapping chemicals per month compared to the total identifications (AGE sampling point). Those chemicals were probably coming from the WWTP with sources before this sampling point and result from incomplete filtering or there was chemical input between the WWTP and the AGE sampling point. Other chemicals were effectively filtered by the WWTP system or could not be identified at the later sampling point. For the overlapping chemicals, the four *PubChemLite* categories analysed above were examined as well, resulting in 36 agrochemicals, 130 pharmaceuticals, 143 compounds associated with disorders and disease and 170 known uses. Consequently, the same trend with dominating identifications of pharmaceuticals (73%) could be observed here, even after the filtering of the WWTP. Persistent synthetic chemicals, like the PFAS perfluorooctanoic acid (PFOA) or perfluorobutanesulfonic acid (PFBS) were found before the WWTP and downstream of the Chiers. However, without having quantitative data on their concentration, little can be said about their environmental effects.

Besides the overlapping chemicals, it is interesting to analyse chemicals identified only at the border to France and not at the WWTP inlet (in total 165 chemicals). Looking for example at the results from April 2022 (1 month before the Petange WWTP sampling), drugs

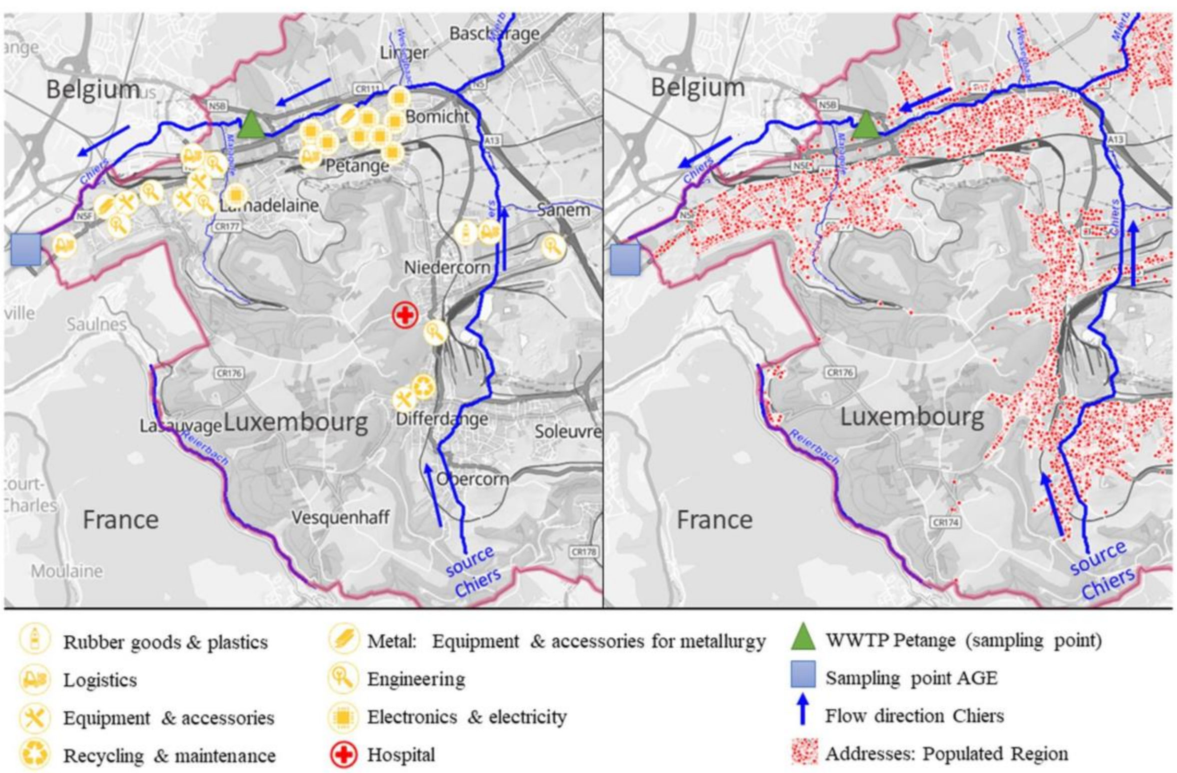


Fig. 9 **A**: Industry and hospitals located next to the sampling points at the river Chiers; **B**: populated region around the river Chiers. Modified from: geoportail.lu

Fig. 10 Number of chemicals overlapping with the Petange WWTP inlet per month compared with the total identifications at the AGE sampling point

like pregabalin (antiepileptic), tramadol (analgesic) and its TP n-desmethyltramadol (with high aquatic toxicity) were found besides other compounds like 1H-benzotriazole (anticorrosive). Again, most of the unique tentative identifications at the border to France were pharmaceuticals (56%) and related to disorders and diseases (65%). The detection of the antiepileptic pregabalin could indicate the medication being used in this area (see population distribution in Fig. 9). Industrial chemicals, like benzotriazoles, only identified after the WWTP, could result from activity in the equipment and accessories, electronics, engineering or metal industry located in the area between WWTP and border. Other PFAS identified solely at the later sampling point were perfluorononanoic acid (PFNA) and perfluoroheptanoic acid (PFHpA), known for their use as surfactants, in fire fighting foams, for the manufacturing of plastics and in the semiconductor industry. These substances are now being phased out in many applications due to their persistence in the environment and potential adverse health effects. The list of 409 chemicals compared to the Chiers results of this study can be found in Additional file 2: Tables S11 and S12.

Lastly, a comparison to the chemicals covered by the governmental target monitoring (AGE) was performed, using the screening lists from 2019 and 2020 and the published results from the water quality report in 2022 [12] (Additional file 2: Table S8). A total of 40 identified chemicals were overlapping with the target monitoring of AGE, including eight (of 16) catchment-specific pollutants: carbamazepine, metolachlor, terbuthylazine, chlorotoluron, tebuconazole, flufenacet, metolachlor ESA and metazachlor OXA. Among the eight not detected chemicals were, e.g. metolachlor OXA and metazachlor ESA, both being TPs of metolachlor and metazachlor, just as metolachlor ESA and metazachlor OXA. 338 chemicals not covered by target monitoring remained and were ranked based on their frequency of occurrence (number of months out of 34) in the Luxembourgish rivers between 2019 and 2022. The top 54 chemicals identified, occurring in at least 13 months, were listed with their common use, the *PubChem* Chemical Identifier (CID),

Table 1 List of chemicals with high occurrence in the 34 months analysed, not currently monitored by AGE and of interest to add to future monitoring lists

Synonym	Use	Parent name	PubChem CID	No of occurrences
Irbesartan	Pharmaceutical	–	3749	32
1H-Benzotriazole	Industry	–	7220	32
4- or 5-methyl-1H-benzotriazole (co-elution)	Industry	–	8705 or 122499	31
Amisulpride	Pharmaceutical	–	2159	31
Telmisartan	Pharmaceutical	–	65999	26
Celiprolol	Pharmaceutical	–	2663	25
Fluconazole	Pharmaceutical	–	3365	25
Trimethoprim	Pharmaceutical	–	5578	24
4-Acetamidoantipyrine	Pharmaceutical	Metamizole	65743	23
Codeine	Pharmaceutical	–	5284371	22
Desvenlafaxine	Pharmaceutical	Venlafaxine	125017	22
Tramadol	Pharmaceutical	–	33741	22
4-NP	Industrial	–	980	22
Valsartan	Pharmaceutical	–	60846	21
Fexofenadine	Pharmaceutical	–	3348	21
Salicylic acid	Industrial, pharmaceutical	–	338	20
Flecainide	Pharmaceutical	–	3356	20
Tiapride	Pharmaceutical	–	5467	20
TCEP	Flame retardant	–	8295	19
TCPP	Flame retardant	–	26176	19
Flufenamic acid	Pharmaceutical	–	3371	19
Aspirin	Pharmaceutical	–	2244	19
Sitagliptin	Pharmaceutical	–	4369359	19
Triethyl phosphate	Industrial	–	6535	18
Adipic acid	Industrial	–	196	18
Losartan	Pharmaceutical	–	3961	17
4-Formylaminoantipyrine	Pharmaceutical	Aminopyrine	72666	17
Sulfamethoxazole	Pharmaceutical	–	5329	16
Carbamazepine-10,11-epoxide	Pharmaceutical	Carbamazepine	2555	15
Bicalutamide	Pharmaceutical	–	2375	14
Dibutyl phthalate	Industrial	–	3026	14
PFOA	Industrial	–	9554	14
2-Hydroxycarbamazepine	Pharmaceutical	Carbamazepine	129274	14
Sulisobenzone	Consumer products	–	19988	13
Sulpiride	Pharmaceutical	–	5355	13
3-Hydroxypyridine	Industrial	–	7971	13
1-Methylbenzotriazole (possible co-elution)	Industrial	1H-Benzotriazole	25902	13
D617	Pharmaceutical	Verapamil	93168	13
Ensulizole	Consumer products	–	33919	13
Tributylamine	Industrial	–	7622	13
Hydrochlorothiazide	Pharmaceutical	–	3639	13

The full list of 54 compounds (including, e.g. food/natural products) is available in Additional file 2: Table S13

the number of occurrences and additional information like the CID of the parent compound (in case of TPs). Chemicals without an environmental or health effect according to *PubChem* data were excluded, *e.g.* natural products, food additives or ubiquitous compounds like caffeine, reducing the list to 41 entries (Table 1).

26 of 41 chemicals tentatively identified are classified as pharmaceuticals (20) or their TPs (six), predominating the analysis results. Two TPs of carbamazepine (parent included in AGE monitoring) were tentatively identified: carbamazepine-10,11-epoxide in 15 months and 2-hydroxycarbamazepine in 14 months, whereas the parent compound was only found in 11 months of this study. Desvenlafaxine, a TP of the antidepressant venlafaxine, was detected in 22 months; the parent compound was already detected in the study by Singh et al. [11] and in 12 months of this study, having a known impact on aquatic environments even at low levels [46]. Out of the 26 pharmaceuticals, seven parent and five TP chemicals were not covered by Singh et al. [11] as well as three parents of the five TPs: metamizole, aminopyrine and verapamil. They might be measured after the study by Singh et al. was conducted or have been missed due to variations in the identification approach. The example of pharmaceuticals proves that it might be worth adding (more) TPs of monitored chemicals to routine target monitoring, as the parent is sometimes not visible and their TPs cause risks as well. Furthermore, 16 chemicals were listed covering uses in industry, consumer products and as flame retardants. The omnipresent benzotriazole class was detected in nearly all measurement months (32) with four different chemicals or TPs identified in this study. However, due to technique limitations (LC-HRMS) like insufficient separation, not all isomeric species can be correctly distinguished with the chromatographic method used, resulting in multiple possible identifications. Besides, two organophosphate flame retardants (OPFRs), namely tris(2-chloroethyl) phosphate (TCEP) and tris(1-chloro-2-propyl) phosphate (TCPP) were detected in 19 months, both known for environmental and toxic effects [47–49]. OPFRs serve as a substitution for brominated flame-retardants (BFRs) such as polybrominated diphenyl ethers (PBDEs), which have been found to cause adverse health effects in many samples recently [47, 50]. The three industrial chemicals 4-nitrophenol or 4-NP (22 months), dibutyl phthalate (DBP) and PFOA (both 14 months) were frequently detected in Luxembourgish rivers as well. 4-NP, used in many industrial applications, is known to have severe environmental and human health effects [51], just as DBP, a plasticizer with high aquatic toxicity [52] and PFOA, already listed for elimination in the Stockholm convention [53]. The chemicals in question were not previously subjected to monitoring by governmental institutions in Luxembourg. Although PFOA, other PFAS or DBP are not on the official monitoring list, these are being measured at AGE using specialized target methods and we recommend these efforts continue. 12 compounds found in this study are already included in the 2022 European watchlist [18], e.g. fipronil, fluconazole or venlafaxine (see Additional file 2: Table S14). Further compounds, e.g. the identified PFAS are included in the 2022 proposal for a directive, amending the Water Framework Directive [19]. At this stage, the lists of tentatively annotated compounds (Table 1 and Additional file 2: Table S13) have been provided to AGE for further confirmation and quantification efforts (since targeted analysis is within their remit) to determine which compounds to include in future monitoring lists, due to their known environmental and health implications.

Discussion

This 3-year investigation was conducted in Luxembourg to examine chemical pollutants in surface water, encompassing various sources and types of contaminants within a single medium, building on previous work. However, it is essential to also consider factors such as flow paths and meteorological phenomena, as they primarily influence the LC–HRMS peak intensities. This can lead to missing identifications when a sample is highly diluted or—during droughts—to a concentration of analyte, making it detectable at all. Other factors, such as sewer overflow due to high precipitation, lead to notably higher analyte signals. Studies show that the latter effect overweighs the dilution effects, *i.e.* rainfall could lead to increased detection of pollutants [54]. Lower precipitation was one of the reasons Singh et al. [11] cited for the decrease of pharmaceuticals in rivers, which was true for the years 2019 and 2020. This study shows the decreasing trend continuing for the following 16 months (2021 and 2022), although the average precipitation was higher in 2021 compared to prior years [55]. Consequently, it seems more likely now that the effect of the pandemic, as explained in Singh et al. [11], might be the reason for a decrease in the numbers of identified pharmaceuticals and river pollutants in general. Analytical causes were not suspected to be the cause here, as the performance of column, device and method were monitored and internal standard signals did not decrease in later analyses.

Data analysis

The open source R package *patRoon* was used to perform the NT data analysis of the 3-year sampling of Luxembourgish surface waters with a tailor-made workflow (Fig. 4) designed for this purpose, as presented in the Data Analysis section in the Methods chapter. *PatRoon* offers the possibility to perform a componentization step, grouping related features in so called *components* based on different similarities, such as chromatographic behaviour. This step was intentionally omitted since it increased the data processing time, and several componentization algorithms lead to false associations. The current state of componentization, particularly with the presence of numerous false positives, is acknowledged to be far from optimal. It is worth noting that the limitations of componentization predominantly stem from the algorithms available in *patRoon*, including popular ones like *CAMERA* [56], *RAMClustR* [57], and *cliqueMS* [58]. Although some tools demonstrate some improvement, most suffer from significant computational inefficiency and further developments seem necessary before these are applicable to environmental/exposomics analyses, which consider a broader range of elements than metabolomics. Moreover, *patRoon* includes the functionality to calculate chemical formulae for the feature groups, based on accurate mass and other data. Depending on the elements chosen in this step (default is C, H, N, O, P), processing time can be extremely long, and possible candidates can be excluded. As a consequence, this step was omitted since the likely presence of e.g. fluorinated compounds in the samples was already clear. Using the *GenForm* [59] algorithm can be a fast way to generate formulae. Nevertheless, when confronted with a large number of candidates, the algorithm's efficiency diminishes as the data size increases, while the use of exact mass to retrieve candidates instead of formula does not increase the number of candidates sufficiently to warrant this step. In order to address this, *patRoon* utilizes 'timeouts' to interrupt the formula generation process. Overall, the tentative identification of features and associated compounds was straightforward (as it was limited to candidates within *PubChemLite* [33, 34] and *MoNA* [36]), reducing the number of features, feature groups and MS peak lists to a minimal extent, demonstrated in the Results section. However, some features might be omitted or not identified based on experimental, algorithm or basic filtering conditions (which is always a risk). The efforts to look for (new) compounds of interest will be therefore continued, looking not only at chemicals included in the *PubChemLite for exposomics* database. However, extending the database always bears the risk of getting many candidates per mass, making data analysis even more difficult.

Classification

In the classification steps of the tentatively identified chemicals, the widely employed tool *classyFire* was used to classify chemicals and obtain supplementary information. However, it is a purely structure-based application, giving just an overview performing a general grouping of chemicals in classes. This information is helpful in terms of establishing a structure in datasets, but does not provide much insight into chemical properties. Therefore, when evaluating *classyFire* compound classifications, it is important to acknowledge that the results may not encompass all information regarding the chemicals under study. Certain compounds may belong to multiple classes, which could potentially result in incomplete classification. Moreover, there are multiple uses per chemical class and a distinction can only be made looking at the individual compound. A way to figure out possible sources of chemicals was the use of *PubChem* metadata for all tentatively identified chemicals and looking at temporal trends (see Fig. 8). Four categories from the *PubChemLite* database, namely *agroChemInfo*, *drugMedicInfo*, *disorderDisease* and *knownUse* were examined to get an estimate of the count of agrochemicals, drugs, compounds associated with disorders and diseases and of those for which commercial use is known. Other categories like the *ToxicityInfo* were not considered, as they are less specific (*e.g.* it indicates whether information is available, but the availability of information alone does not say whether it is toxic or not). Several pharmaceuticals are associated with disorders and disease as the annotation also includes treatments. Again, the environmental effect of the tentatively detected exposome-related compounds can only be fully accessed looking at the individual case considering their concentrations. It is worrying that a majority of compounds found in the Luxembourgish environment were of pharmaceutical origin because these compounds are designed to have biological effects, which means they may have unintended effects on non-target organisms and ecosystems. Even at low concentrations, pharmaceutical compounds as venlafaxine [46] can accumulate in organisms and can have toxic effects, especially over time. They can also promote antibiotic resistance, alter gene expression, and disrupt endocrine systems (for more information in the Luxembourgish context, see Singh et al. [11]). Additionally, some pharmaceuticals are known to persist in the environment for long periods of time and can travel long distances, leading to contamination of remote areas and cross-border contamination (Germany, France, Belgium). The presence of pharmaceuticals in Luxembourgish rivers highlights the potential to enhance wastewater treatment and disposal systems, thereby safeguarding public health. Luxembourg's biological and mechanical WWTPs are not always designed to remove pharmaceuticals from wastewater, and some compounds are not effectively degraded by current treatment processes [60–62]. As a result, these compounds can end up in surface waters (as shown here), groundwater, and even drinking water sources.

Comparison to other studies

Besides focussing on the temporal variations of chemicals and their classification, it can be highly valuable to include geographical information, especially to analyse possible origins of chemicals (and thus possible origins of the potential threat). Therefore, the measurement results of one sampling point (Chiers) over 3 years were compared to the results of the inlet of a WWTP located ~ 6 km upstream. As shown in the Results section, an overlap of nearly 200 chemicals was identified and a majority of compounds identified in this study was found before in the WWTP (see Fig. 10). This is either a result of incomplete treatment in the WWTP or chemical input between the WWTP and the later sampling point. A comparison between inlet and outlet of the WWTP could clarify this. However, the identifications in both studies are of tentative nature: in this study, the aim was to use NTA to propose chemicals for further target verification at AGE (beyond the scope of this article). The verification efforts in the other study are still ongoing (likewise beyond the scope of this article, as different partners are involved).

Moreover, a comparison to former published studies on Luxembourgish surface water was performed, looking at pesticides [10] and pharmaceuticals [11]. This was done to perform a plausibility check of the results (chemical identifications and temporal trends), as both compared studies involved a target approach after screening for suspects. The list of 378 tentatively identified chemicals was then compared to the monitoring list used by the water administration in Luxembourg, showing not only the overlap of findings, but also nominating potential chemicals for future inclusion to this list (see Table 1). Several of the annotated chemicals, *e.g.* venlafaxine are already listed in the 2022 EU watchlist [18] or the 2022 proposal to revise the priority substances [19]. This shows that there is already work in progress to regulate these chemicals in the EU.

Monitoring and regulations are necessary steps to track contaminant distribution, identify potential sources and restrict or ban their use to stop their release to the environment. Adding NTA to track regularly potential water contaminants can help with this, as this study shows. The NTA workflow established via *patRoon* could be used and improved further by AGE, as it is also compatible with the instrument type used in their laboratories.

Future steps should involve the target analysis of the tentatively identified chemicals, with a special focus on emerging pollutants and persistent chemicals (Table 1). Giving a list of those substances to governmental institutions like AGE could improve future monitoring efforts. A quantification of these compounds, using reference standards could then help to estimate their environmental threat. Overall, the impact of these compounds on the environment and human health underscores the importance of monitoring and regulating their use and disposal to minimize their potential impact.

Conclusions

Based on the results presented in the article, the analysis of the 3-year sampling of Luxembourgish surface waters using the *patRoon* R package tentatively identified 378 chemicals associated with the exposome. 40 chemicals were already included in routine water analysis performed by AGE. The results were plausibility checked using former studies by looking at overlapping identifications and general trends, with many of the tentative identifications here matching the confirmed identifications in the previous studies. The identified chemicals were classified to get a general overview, and the results showed that benzenoids, organoheterocyclic compounds, and organic phosphoric acids and derivatives were among the most identified classes (*classyFire*). A temporal analysis between the 3 years and in 1 year was shown, looking at the classifications obtained not only by *classyFire*, but also by the *PubChemLite* categories on agrochemicals, pharmaceuticals, drugs and diseases and known uses. Most of the chemicals detected had a known use, were classified as pharmaceuticals and are associated with disorders and diseases. A decrease in identifications was observed in 2021 and 2022. Trends in the use of agrochemicals could be seen in the monthly comparison, with May to July showing the highest number of identifications. The study examined not only temporal, but also geographical variations of chemicals to analyse their possible origins and potential threats. Chemical measurements from one sampling point were compared to those from an inlet of a WWTP located 6 km away. Overlapping and differing chemicals were identified, and the study found that pharmaceuticals dominated again the identifications. Some persistent synthetic chemicals were also found on both sampling sites. Further investigations will be done, looking at the concentrations and sources of these pollutants. A list of 41 chemicals—not yet included in the AGE monitoring—is presented with some chemicals already highlighted in former studies, *e.g.* venlafaxine and its TP desvenlafaxine. The study suggests adding more TP compounds to the AGE monitoring list, as many parent compounds are not detected, but rather their (active) TPs. Moreover, industrial chemicals like 4-NP and flame retardants like OPFRs were detected in a majority of the analysed months and should be considered to be added after confirmation by target efforts. The aim of including NTA lists like these to governmental monitoring is to eliminate harmful chemicals from the environment, not only by searching for their sources, but also by implementing (WWTP) treatment technologies to reduce their presence in the environment. Biological treatments and monitoring regulations are also necessary in Luxembourg to track contaminant distribution, identify potential sources, and restrict or ban their use. All in all, this study of surface waters has shown how NTA can be used to complement routine (target) monitoring programmes and help to broaden the focus to new emerging chemicals. Future efforts will involve extending the target monitoring and working on implementing this open source workflow into AGE monitoring routines directly.

Abbreviations

4-NP	4-Nitrophenol
AGE	L'Administration de la gestion de l'eau
BFR	Brominated flame-retardant
CID	Chemical Identifier
DBP	Dibutyl phthalate
DDD	Dichlorodiphenyldichloroethane
DDE	Dichlorodiphenyldichloroethylene
DDT	Dichlorodiphenyltrichloroethane
DMSO	Dimethyl sulfoxide
DNOC	4,6-Dinitro-o-kresol
DoE	Design of Experiment
ECI	Environmental Cheminformatics Group
HRMS	High-resolution mass spectrometry
LC	Liquid chromatography
MoNA	MassBank of North America
NTA	Non-target analysis
NT	Non-target
OPFRs	Organophosphate flame retardants
PBDE	Polybrominated diphenyl ether
PFAS	Per- and poly-fluorinated compounds
PFBS	Perfluorobutanesulfonic acid
PFHpA	Perfluoroheptanoic acid
PFNA	Perfluorononanoic acid
PFOA	Perfluorooctanoic acid
PFOS	Perfluorooctanesulfonic acid
TCPP	Tris(1-chloro-2-propyl) phosphate
TCEP	Tris(2-chloroethyl) phosphate
WWTP	Wastewater treatment plant

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12302-023-00805-5.

Additional file 1: Figure S1. Design of Experiments (DoEs) of feature optimization workflow, with DoE3 giving best results for the April 2020 analyses. The iterative process involved the testing of parameters in three DoEs; the best slices for each DoE were visualized using perspective plots. **Figure S2.** Results of feature optimization workflow, with the best results in positive and negative mode for the April 2020 analyses.

Additional file 2: Table S1. Sampling sites with weblinks. **Table S2.** Feature finding optimization results. **Table S3.** Number of 2, 3a and 3b identifications. **Table S4.** Number of positive and negative identifications. **Table S5.** Tentatively identified chemicals per month. **Table S6.** ClassyFire results. **Table S7.** PubChemLite classifications. **Table S8.** AGE monitoring comparison. **Table S9.** Krier et al. (pesticides) comparison. **Table S10.** Singh et al. (pharmaceuticals) comparison. **Table S11.** Chiers WWTP Petange comparison (SMILES). **Table S12.** Chiers results (SMILES) all months. **Table S13.** List of chemicals with high occurrence in the 34 months analysed, and not currently monitored by AGE. **Table S14.** European watchlist 2022 comparison.

Acknowledgements
The authors gratefully acknowledge the contribution of Parviel Chirsir for the 2022 sampling data results and discussions with others in the LuxTIME and ECI teams.

Author contributions
DA and ELS did the conceptualization of the study. RH contributed the changes to the software and DA developed the methodology. The formal analysis, main investigation, visualization and writing of the original draft were conducted by DA. PD provided the resources (water samples) and did parts of the investigation. All authors were responsible for review and editing. ELS was supervising and performed funding acquisition and resources.

Funding
DA acknowledges funding support from the University of Luxembourg through its Institute for Advanced Studies for the project LuxTIME (Audacity-2019); ELS acknowledges funding support from the Luxembourg National Research Fund (FNR) for project A18/BM/12341006.

Availability of data and materials
All data, code, figures and supplementary files can be found in the 'data_luxwater_nt_paper_da' repository of the ECI group in GitLab, via https://gitlab.lcsb.uni.lu/eci/data_luxwater_nt_paper_da under license Artistic 2.0. The required sample identifiers to link the measurement to location and date can be found in this GitLab file. The measurement files in mzML format can be found as dataset MSV000092221 from the [GNPS MassIVE repository] (https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp), accessible via ftp://massive.ucsd.edu/MSV000092221/ and to be cited with https://doi.org/10.25345/C55X25P62. All data are also accessible via https://doi.org/10.17881/5dk0-jv81.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, 6 Avenue du Swing, 4367 Belvaux, Luxembourg. [2]Administration de La Gestion de L'eau (AGE), Avenue du Rock'n'Roll 1, 4361 Esch-Sur-Alzette, Luxembourg. [3]Institute for Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands.

Received: 3 July 2023 Accepted: 29 October 2023
Published online: 07 November 2023

References

1. Arp HPH, Aurich D, Schymanski EL et al (2023) Avoiding the next silent spring: our chemical past, present, and future. Environ Sci Technol 57:6355–6359. https://doi.org/10.1021/acs.est.3c01735
2. Agency for Toxic Substances and Disease Registry, U.S. Department of Health and Human Services (2022) Toxicological Profile for DDT, DDE, and DDD. In: https://www.atsdr.cdc.gov/toxprofiles/tp35.pdf. Accessed 23 May 2023
3. Cocco P, Kazerouni N, Zahm SH (2000) Cancer mortality and environmental exposure to DDE in the United States. Environ Health Perspect 108:1–4. https://doi.org/10.1289/ehp.001081
4. Schymanski EL, Williams AJ (2017) Open science for identifying "Known Unknown" chemicals. Environ Sci Technol 51:5357–5359. https://doi.org/10.1021/acs.est.7b01908
5. Little JL, Cleven CD, Brown SD (2011) Identification of "Known Unknowns" utilizing accurate mass data and chemical abstracts service databases. J Am Soc Mass Spectrom 22:348–359. https://doi.org/10.1007/s13361-010-0034-3
6. Wang X, Shen Z, Zeng Y et al (2018) Day-night differences, seasonal variations and source apportionment of PM10-Bound PAHs over Xi'an, Northwest China. Atmosphere 9:62. https://doi.org/10.3390/atmos9020062

7. Stamatis N, Hela D, Triantafyllidis V, Konstantinou I (2013) Spatiotemporal variation and risk assessment of pesticides in water of the lower catchment basin of Acheloos River, Western Greece. Sci World J. https://doi.org/10.1155/2013/231610
8. Aurich D, Miles O, Schymanski EL (2021) Historical exposomics and high resolution mass spectrometry. Exposome 1:1–15. https://doi.org/10.1093/exposome/osab007
9. Lickes J-P, L'Administration de la gestion de l'eau (AGE), Luxembourg (2022) Vorstellung des 3. Bewirtschaftungsplans. In: Httpsgouvernementludam-Assetsdocumentsactualites202209-Sept.-Gest.-Dist.-Hydrogr.--3-Wasserbewirtschaftungsplanpdf. https://gouvernement.lu/dam-assets/documents/actualites/2022/09-septembre/22-gestion-districts-hydrographiques/vorstellung-des-3-wasserbewirtschaftungsplan.pdf. Acessed 26 May 2023
10. Krier J, Singh RR, Kondić T et al (2022) Discovering pesticides and their TPs in Luxembourg waters using open cheminformatics approaches. Environ Int 158:14. https://doi.org/10.1016/j.envint.2021.106885
11. Singh RR, Lai A, Krier J et al (2021) Occurrence and distribution of pharmaceuticals and their transformation products in Luxembourgish surface waters. ACS Environ Au 1:58–70. https://doi.org/10.1021/acsenvironau.1c00008
12. L'Administration de la gestion de l'eau (AGE), Luxembourg (2022) Anhang 14-Überschreitungen der UQN für prioritäre und flussgebietsspezifische Stoffe. In: Httpseaugouvernementludam-Assetsadministrationdocuments3-Cycleelaboration--3e-Plan--Gest.-Doc.-Final.-14-Uberschreitungen--Uqn--Prioritare--Flussgebietsspezifische-Stoffepdf. https://eau.gouvernement.lu/dam-assets/administration/documents/3-cycle/elaboration-du-3e-plan-de-gestion-document-final/anhang-14-uberschreitungen-der-uqn-fur-prioritare-und-flussgebietsspezifische-stoffe.pdf. Accessed 23 May 2023
13. L'Administration de la gestion de l'eau (AGE), Luxembourg (2022) Elaboration du 3e plan de gestion. In: Httpeaugouvernementlufradministrationdirectives Directive-Cadre-Sur-Leau3e-Cycle-2021-2027elaboration--3e-Plan--Gest.-Doc.-Final. http://eau.gouvernement.lu/fr/administration/directives/Directive-cadre-sur-leau/3e-cycle-(2021-2027)/elaboration-du-3e-plan-de-gestion-document-final.html. Accessed 23 May 2023
14. European-Parliament, Council of the European Union (2000) Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy
15. European-Parliament, Council of the European Union (2008) Directive 2008/105/EC of the European Parliament and of the Council of 16 December 2008 on environmental quality standards in the field of water policy, amending and subsequently repealing Council Directives 82/176/EEC, 83/513/EEC, 84/156/EEC, 84/491/EEC, 86/280/EEC and amending Directive 2000/60/EC of the European Parliament and of the Council
16. Directorate-General for Environment (2023) Surface water. In: Eur. Com. - Energy Clim. Change Environ. https://environment.ec.europa.eu/topics/water/surface-water_en. Accessed 11 Oct 2023
17. European-Parliament, Council of the European Union (2013) Directive 2013/39/EU of the European Parliament and of the Council of 12 August 2013 amending Directives 2000/60/EC and 2008/105/EC as regards priority substances in the field of water policy Text with EEA relevance
18. European-Parliament, Council of the European Union (2022) Commission Implementing Decision (EU) 2022/1307 of 22 July 2022 establishing a watch list of substances for Union-wide monitoring in the field of water policy pursuant to Directive 2008/105/EC of the European Parliament and of the Council (notified under document C(2022) 5098) (Text with EEA relevance)
19. Directorate-General for Environment (2022) Proposal amending Water Directives. In: Eur. Com. - Energy Clim. Change Environ. https://environment.ec.europa.eu/publications/proposal-amending-water-directives_en. Accessed 12 Oct 2023
20. Blum KM, Andersson PL, Renman G et al (2017) Non-target screening and prioritization of potentially persistent, bioaccumulating and toxic domestic wastewater contaminants and their removal in on-site and large-scale sewage treatment plants. Sci Total Environ 575:265–275. https://doi.org/10.1016/j.scitotenv.2016.09.135
21. Wang X, Yu N, Qian Y et al (2020) Non-target and suspect screening of per- and polyfluoroalkyl substances in Chinese municipal wastewater treatment plants. Water Res 183:12. https://doi.org/10.1016/j.watres.2020.115989
22. Suman T-Y, Kim S-Y, Yeom D-H, Jeon J (2022) Transformation products of emerging pollutants explored using non-target screening: perspective in the transformation pathway and toxicity mechanism—a review. Toxics 10:22. https://doi.org/10.3390/toxics10020054
23. Helmus R, van de Velde B, Brunner AM et al (2022) patRoon 2.0: improved non-target analysis workflows including automated transformation product screening. J Open Source Softw. https://doi.org/10.21105/joss.04029
24. Helmus R, ter Laak TL, van Wezel AP et al (2021) patRoon: open source software platform for environmental mass spectrometry based non-target screening. J Cheminformatics. https://doi.org/10.1186/s13321-020-00477-w
25. Smith CA, Want EJ, O'Maille G et al (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal Chem 78:779–787. https://doi.org/10.1021/ac051437y
26. Tautenhahn R, Böttcher C, Neumann S (2008) Highly sensitive feature detection for high resolution LC/MS. BMC Bioinform. https://doi.org/10.1186/1471-2105-9-504
27. Benton HP, Want EJ, Ebbels TMD (2010) Correction of mass calibration gaps in liquid chromatography–mass spectrometry metabolomics data. Bioinformatics 26:2488–2489. https://doi.org/10.1093/bioinformatics/btq441
28. Ruttkies C, Schymanski EL, Wolf S et al (2016) MetFrag relaunched: incorporating strategies beyond in silico fragmentation. J Cheminform. https://doi.org/10.1186/s13321-016-0115-9
29. Libiseller G, Dvorzak M, Kleb U et al (2015) IPO: a tool for automated optimization of XCMS parameters. BMC Bioinformatics 16:118. https://doi.org/10.1186/s12859-015-0562-8
30. Albóniga OE, González O, Alonso RM et al (2020) Optimization of XCMS parameters for LC–MS metabolomics: an assessment of automated versus manual tuning and its effect on the final results. Metabolomics 16:14. https://doi.org/10.1007/s11306-020-1636-9
31. Tostengard AR, Smith R (2021) A review and evaluation of techniques for improved feature detection in mass spectrometry data. Grad Stud Theses Diss Prof Pap Univ Mont https://scholarworks.umt.edu/etd/11679:41. Accessed 23 March 2023
32. Kim S, Chen J, Cheng T et al (2023) PubChem 2023 update. Nucleic Acids Res 51:D1373–D1380. https://doi.org/10.1093/nar/gkac956
33. Bolton E, Schymanski E, Kondic T, et al (2023) PubChemLite for Exposomics. https://doi.org/10.5281/zenodo.7576412
34. Schymanski EL, Kondić T, Neumann S et al (2021) Empowering large chemical knowledge bases for exposomics: PubChemLite meets MetFrag. J Cheminformatics 13:19. https://doi.org/10.1186/s13321-021-00489-0
35. Schymanski EL, Jeon J, Gulde R et al (2014) Identifying small molecules via high resolution mass spectrometry: communicating confidence. Env Sci Technol 48:2097–2098. https://doi.org/10.1021/es5002105
36. Fiehnlab (2018) MassBank of North America (MoNA). In: https://mona.fiehnlab.ucdavis.edu/. https://mona.fiehnlab.ucdavis.edu/. Accessed 20 Jun 2023
37. Djoumbou Feunang Y, Eisner R, Knox C et al (2016) ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. J Cheminformatics 8:61. https://doi.org/10.1186/s13321-016-0174-y
38. NCBI (2023) PubChem Classification Browser TOC. In: https://pubchem.ncbi.nlm.nih.gov/classification/#hid=72. https://pubchem.ncbi.nlm.nih.gov/classification/#hid=72. Accessed 26 Jun 2023
39. Kessner D, Chambers M, Burke R et al (2008) ProteoWizard: open source software for rapid proteomics tools development. Bioinformatics 24:2534–2536. https://doi.org/10.1093/bioinformatics/btn323
40. Chambers MC, Maclean B, Burke R et al (2012) A cross-platform toolkit for mass spectrometry and proteomics. Nat Biotechnol 30:918–920. https://doi.org/10.1038/nbt.2377
41. Martens L, Chambers M, Sturm M et al (2011) mzML - a community standard for mass spectrometry data. Mol Cell Proteomics 10(R110):000133. https://doi.org/10.1074/mcp.R110.000133
42. Pedrioli PGA, Eng JK, Hubley R et al (2004) A common open representation of mass spectrometry data and its application to proteomics research. Nat Biotechnol 22:1459–1466. https://doi.org/10.1038/nbt1031

43. Keller A, Eng J, Zhang N et al (2005) A uniform proteomics MS/MS analysis platform utilizing open XML file formats. Mol Syst Biol 1:8. https://doi.org/10.1038/msb4100024
44. Talavera Andújar B, Aurich D, Aho VTE et al (2022) Studying the Parkinson's disease metabolome and exposome in biological samples through different analytical and cheminformatics approaches: a pilot study. Anal Bioanal Chem 414:7399–7419. https://doi.org/10.1007/s00216-022-04207-z
45. Awan M, Buriak I, Fleck R et al (2020) Dimethyl sulfoxide: a central player since the dawn of cryobiology, is efficacy balanced by toxicity? Regen Med 15:1463–1491. https://doi.org/10.2217/rme-2019-0145
46. Best C, Melnyk-Lamont N, Gesto M, Vijayan MM (2014) Environmental levels of the antidepressant venlafaxine impact the metabolic capacity of rainbow trout. Aquat Toxicol Amst Neth 155:190–198. https://doi.org/10.1016/j.aquatox.2014.06.014
47. Maddela NR, Venkateswarlu K, Megharaj M (2020) Tris(2-chloroethyl) phosphate, a pervasive flame retardant: critical perspective on its emissions into the environment and human toxicity. Environ Sci Process Impacts 22:1809–1827. https://doi.org/10.1039/D0EM00222D
48. Wu H, Zhong M, Lu Z et al (2018) Biological effects of tris (1-chloro-2-propyl) phosphate (TCPP) on immunity in mussel Mytilus galloprovincialis. Environ Toxicol Pharmacol 61:102–106. https://doi.org/10.1016/j.etap.2018.05.022
49. Ji C, Lu Z, Xu L et al (2020) Global responses to tris(1-chloro-2-propyl) phosphate (TCPP) in rockfish Sebastes schlegeli using integrated proteomic and metabolomic approach. Sci Total Environ. https://doi.org/10.1016/j.scitotenv.2020.138307
50. Zhang Z-N, Yang D-L, Liu H et al (2023) Effects of TCPP and TCEP exposure on human corneal epithelial cells: oxidative damage, cell cycle arrest, and pyroptosis. Chemosphere. https://doi.org/10.1016/j.chemosphere.2023.138817
51. Naushad Mu, Ahamad T, Rizwan Khan M (2022) Remediation of wastewater containing 4-nitrophenol using ionic liquid stabilized nanoparticles: Synthesis, characterizations and applications. Chemosphere. https://doi.org/10.1016/j.chemosphere.2022.135173
52. Lin W, He Y, Li R et al (2023) Adaptive changes of swimming crab (Portunus trituberculatus) associated bacteria helping host against dibutyl phthalate toxification. Environ Pollut. https://doi.org/10.1016/j.envpol.2023.121328
53. Fiedler H, Kennedy T, Henry BJ (2021) A critical review of a recommended analytical and classification approach for organic fluorinated compounds with an emphasis on Per- and polyfluoroalkyl substances. Integr Environ Assess Manag 17:331–351. https://doi.org/10.1002/ieam.4352
54. Zhu L, Jiang C, Panthi S et al (2021) Impact of high precipitation and temperature events on the distribution of emerging contaminants in surface water in the Mid-Atlantic, United States. Sci Total Environ. https://doi.org/10.1016/j.scitotenv.2020.142552
55. statista (2023) Luxembourg: annual rainfall in Luxembourg-City 2021. In: Httpswwwstatistacomstatistics584864annual-Rainfall--Luxemb.-City. https://www.statista.com/statistics/584864/annual-rainfall-in-luxembourg-city/. Accessed 1 Jun 2023
56. Kuhl C, Tautenhahn R, Böttcher C et al (2012) CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. Anal Chem 84:283–289. https://doi.org/10.1021/ac202450g
57. Broeckling CD, Afsar FA, Neumann S et al (2014) RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. Anal Chem 86:6812–6817. https://doi.org/10.1021/ac501530d
58. Senan O, Aguilar-Mogas A, Navarro M et al (2019) CliqueMS: a computational tool for annotating in-source metabolite ions from LC-MS untargeted metabolomics data based on a coelution similarity network. Bioinformatics 35:4089–4097. https://doi.org/10.1093/bioinformatics/btz207
59. Meringer M, Reinker S, Zhang J, Muller A (2011) MS/MS data improves automated determination of molecular formulas by mass spectrometry. Match Commun Math Comput Chem 65:259–290
60. Venditti S, Arenz-Leufen M, Köhler C, et al Treatment of pharmaceutical wastewater by O3 and O3/H2O2 processes: a pilot scale study in Luxembourg. https://www.researchgate.net/publication/260697444. Accessed 19 October 2023
61. Krein A, Pailler J-Y, Guignard C et al (2012) Determination of estrogen activity in river waters and wastewater in Luxembourg by chemical analysis and the yeast estrogen screen assay. Environ Pollut 1:p86. https://doi.org/10.5539/ep.v1n2p86
62. Pailler J-Y, Krein A, Pfister L et al (2009) Solid phase extraction coupled to liquid chromatography-tandem mass spectrometry analysis of sulfonamides, tetracyclines, analgesics and hormones in surface water and wastewater in Luxembourg. Sci Total Environ 407:4736–4743. https://doi.org/10.1016/j.scitotenv.2009.04.042

Publisher's Note

General Discussion and Future Perspectives

The interdisciplinary initiative *LuxTIME* studied the exposome, offering new insights and perspectives by incorporating a historical dimension. Through collaboration between natural scientists and humanities scholars, a shared vocabulary was developed, drawing input from all participating disciplines. The introduction of the historical exposome as an independent paradigm not only critiques prevailing trends in modern exposomics research, but also delves into the past to extract information preceding the present state. In the context of the Anthropocene era, wherein the undeniable influence of human activities on the environment is observed, equal attention must be given to the reciprocal influence of the environment on human health. This becomes especially relevant when examining the chemical and health perspectives, given the increasing presence of environmental chemicals and the consequent health and environmental concerns.

By constructing a digital inventory using diverse data sources across Luxembourg, historical exposomics data on ecosystem, lifestyle, social and physical-chemical factors was discovered and catalogued. The project focussed not only on the external or the internal exposome but tried to build an overall picture of the exposome. This endeavour proved challenging considering different data or sample types from both social and natural archives, each possessing distinct characteristics. Notably, Luxembourg faced limitations in accessing natural archives, primarily due to the absence of environmental sample banks and lack of suitable natural archive systems such as lake sediments. Consequently, in the absence of historical exposure samples, alternative sources such as social archives, simulations (utilizing mathematical models, *e.g.,* for analysing atmospheric pollution), or a combination thereof were used instead.

To effectively communicate the *LuxTIME* findings and facilitate navigation through the inventory, data visualization techniques were employed. These visualizations not only conveyed the research outcomes but also presented the metadata of the *LuxTIME* inventory, serving as a navigation tool. The creation of visual representations aimed to explore and communicate project data to an interdisciplinary audience, making use of tools and methodologies from various disciplines. Thus, a visualization *toolbox* was formed, leveraging and enhancing pre-established concepts, such as the utilization of the *chemical stripes* visualisation (developed code accessible as an R package). This *toolbox* serves as a starting point to collect (interdisciplinary) data visualization techniques that can be extended and modified.

Through the course of this exploration (presented in CHAPTER IV), it became evident that certain visualizations proved more efficacious than others, each carrying its own set of advantages and disadvantages. For instance, the utilization of visual metaphors and modifications of traditional statistical graphs demonstrated their utility in enhancing data representation within the natural sciences. On the other hand, multivariate data glyphs emerged as a valuable approach for conveying extensive datasets with many details, yet it was apparent that this approach could potentially compromise ease of comprehension due to the complexity of the resulting figures, demanding more time and effort from the audience. The exploration of visualizations involving non-standard metrics, particularly temporality, brought to light certain challenges. The depiction of varying perceptions of time within scientific visualizations poses significant implementation difficulties. This highlighted the fact that while creativity in visual representation is advantageous, it must be applied within the bounds of practicality and accessibility, *e.g.* in the context of conveying scientific findings.

The integration of environmental cheminformatics in the *LuxTIME* exposomics studies covered mainly the analysis of known and unknown chemicals in Luxembourg using HRMS and NTA. Various NTA workflows were tested to establish a method for analysing surface water samples. Different types of pollutants were considered, including persistent chemicals from industrial, medical or agrochemical applications, representing the anthropogenic influence on the natural world.

Future efforts following this PhD project may involve continuing with the NTA workflow of Luxembourgish surface water samples, as established in this work. Conducting long-term monitoring of environmental samples using the established NTA workflow can help to track changes in chemical composition and identify (new) emerging contaminants. This would provide further insights into temporal trends and enable the assessment of potential environmental and health risks over a longer time frame. The sample preparation and HRMS method utilized in this study do not encompass the whole range of environmental pollutants, with certain substances like PFAS being inadequately extracted. To enhance sensitivity across various compound classes, such as the group of persistent, mobile, and toxic (PMT) compounds, alternative extraction techniques and chromatographic columns, like HILIC (Hydrophilic interaction liquid chromatography), could be employed.

Further investigations are necessary to confirm the tentatively identified chemicals, and quantification efforts will be required to assess their environmental and health implications. Once confirmed, an inclusion of these chemicals in routine governmental monitoring lists (Administration de la gestion de l'eau, AGE) could be considered, underscoring the importance of integrating NTA into regular monitoring practices. Opting for the open source R package *patRoon* to conduct the non-target analysis aligns well with these endeavours, given its compatibility with various instrument types, including the SCIEX instrument used by AGE. Future steps could further involve proposing regulatory measures to mitigate identified risks or influencing the development of guidelines for exposure assessment.

The *LuxTIME* data inventory can be further expanded by integrating further exposomics data from multiple sources, including environmental, biological, and lifestyle factors, to gain a holistic understanding of the complex interactions between various exposures and their cumulative effects on human health. This could involve developing data integration frameworks and methodologies to analyse large-scale datasets, expanding the scale from the Minett region to the whole of Luxembourg. However, it is essential to acknowledge that the feasibility of extending the *LuxTIME* data inventory may be challenging due to limited access to certain documents and data from various Luxembourgish institutions. Some institutions might not grant unrestricted access to their resources, which could hinder the inclusion of specific datasets within the inventory, which was already a problem at an early stage of *LuxTIME*. Overcoming these access restrictions and ensuring comprehensive data integration may require extensive collaboration and negotiations with relevant stakeholders to establish data-sharing agreements and adhere to data privacy and security regulations (this may be feasible with the newly-established Luxembourg National Data Service, LNDS). Presently, the established data catalog is accessible only within the project, and permission must be granted to publish the collection on the *LuxTIME* website or other platforms. However, efforts are underway to update the website with visualizations based on the data inventory, which will be made available before the conclusion of the project.

Moreover, there is scope for further refinement and expansion of the presented data visualization *toolbox* (RESULTS, CHAPTER IV), which serves as a foundational collection of tools that can be extended and evaluated in diverse (interdisciplinary) projects. Novel visualization approaches can be explored, and their effectiveness can be evaluated in conveying complex exposome data to diverse stakeholders. Additional efforts are needed to enhance awareness regarding the availability of various visualization tools beyond those typically employed within a single discipline. The proper use of visualisations and the exploration of a wide range of visual options could be effectively conveyed through workshops or lectures, as there remains a deficit in this area among many researchers. Tools for retrospective reflection, like the visualization of the evolution of topics over time (Figure 3 of RESULTS, CHAPTER IV), or visual metaphors, like the tree in Figure 4 of RESULTS, CHAPTER IV, should be applied more often in future applications. Multivariate data glyphs, as shown in Figure 7, allow users to explore several aspects of a dataset at the same time, which can be very effective in many scientific applications (considering disadvantages presented above).

Concerning the *chemical stripes* visualization, ongoing investigations into the extensive dataset that underlies these stripes promise to be of significant interest to the environmental community, as indicated by the feedback received thus far. Further research is already in progress (master student) to differentiate between various types of patents, such as those pertaining to industrial applications, bioremediation, or monitoring methods. Moreover, the source of patents is of interest, as well as the 'decoding' of the patent identifier and the study of trends per chemical class or country over time. Additionally, alternative data sources may be explored to supplement patent or literature data. In RESULTS, CHAPTER IV and V only patent data was used to generate the stripes. The subsequently developed R package can generate both literature and patent stripes (see FIGURE 3 in the SYNOPSIS Chapter), using the consolidated reference numbers and depositor supplied patent numbers listed in PubChem. Looking at certain chemicals, like the pharmaceutical compound Gallopamil (PubChem CID 1234), the trend in patent and literature numbers can be directly correlated with its use: First studies showed the antihypertensive effect of Gallopamil in 1970 (see literature data, FIGURE 4A) resulting in increased patent numbers in the following decades (FIGURE 4B). As testing continued and common side effects were discovered, the pharmaceutical's literature counts increased around 1989 (FIGURE 4A) and the product was taken of (several) markets in 2012, resulting in low to none literature and patent numbers.

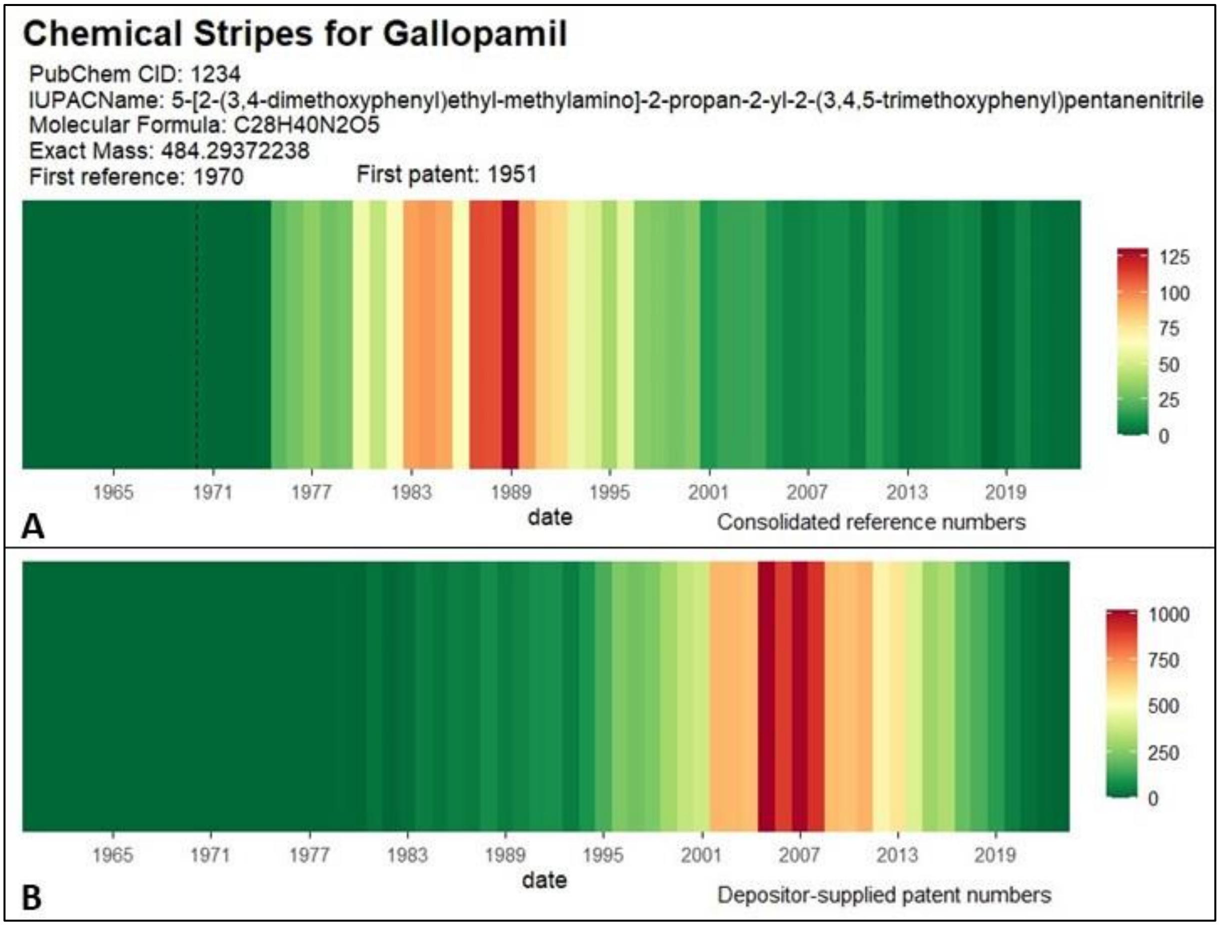


Figure 4: Chemical Stripes for Gallopamil (1960-2023) using (A) literature and (B) patent data.

Moreover, the R-package includes the functionality of creating so-called 'sum stripes' cumulating patent or literature numbers for a list of up to 300 given chemicals. This addition enables users to generate the stripes for a whole list of compounds, *e.g.* 33 compounds from the 'ZeroPM Box 1' substance list (NORMAN SUSPECT LIST EXCHANGE), see FIGURE 5. The pattern shows an increase in patents around 2010, like many stripes do, which still needs future explanation by studying the patent data. The creation of the *sum stripes* function became only feasible through our feedback, prompting PubChem to implement modifications that greatly streamlined data retrieval for us.

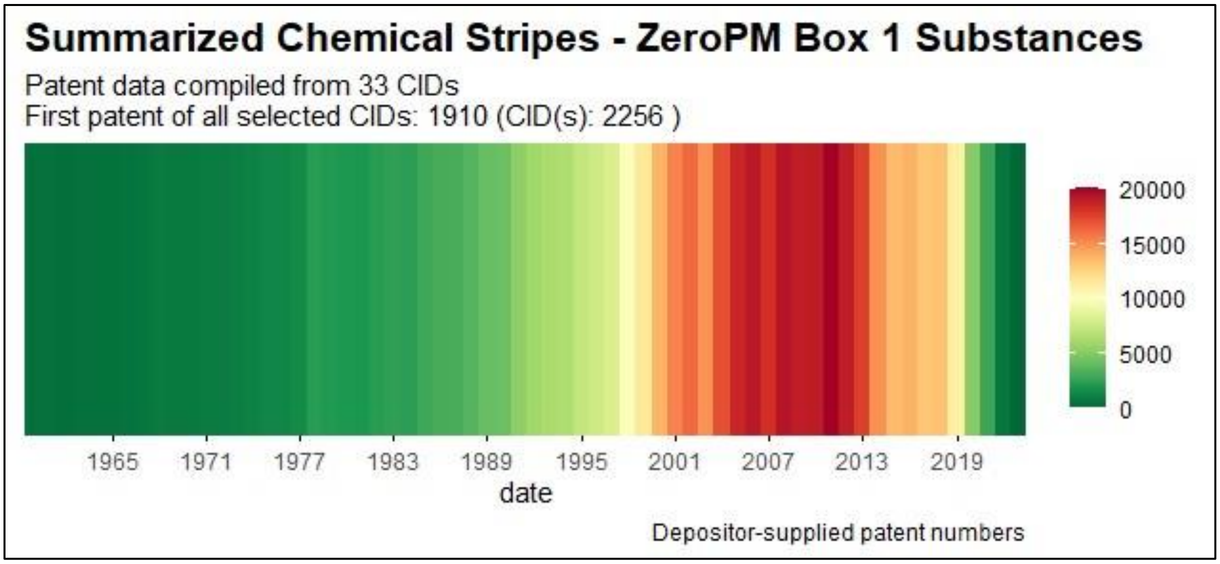


Figure 5: Summarized Chemical Stripes for 33 compounds from the Zero PM Box1 list (NORMAN Suspect List Exchange) using patent data from 1960-2023.

The interdisciplinary concept behind *LuxTIME* undoubtedly offers promising opportunities for advancing exposome research and understanding the historical interactions between environmental exposures and human health. The collaboration between experts from different fields can foster innovative methodologies and diverse perspectives, enriching the research process and broadening the scope of investigation. One of the key advantages of interdisciplinarity (in *LuxTIME*) is the potential to overcome traditional disciplinary boundaries (creating a 'trading zone'), enabling researchers to explore complex and multifaceted exposomics data comprehensively. However, the interdisciplinary nature also poses challenges and limitations. Integrating various disciplines necessitates effective communication, collaboration, and coordination among team members possessing diverse expertise. This integration can pose challenges in establishing a shared understanding and vocabulary, particularly evident during *LuxTIME's* early stages when such understanding was just beginning to develop, and even the definition of 'deliverables' remained elusive. Disparities in methodologies, data formats, and analytical techniques from different disciplines may also hinder seamless data integration and harmonization. These disparities became notably pronounced in the context of joint publications, reflecting differences in processes between the humanities and natural sciences. Similarly, the criteria and requirements that each PhD student had to meet varied depending on their respective fields, adding another layer of complexity to the interdisciplinary collaboration.

Furthermore, resource allocation and funding for interdisciplinary projects can be demanding. The diversity of expertise and research requirements may require additional financial and time investments, which could potentially limit the scalability of such projects in the future.

Additionally, as *LuxTIME* expands to include data from various sources, ensuring data privacy and ethical considerations become paramount. Handling sensitive historical and contemporary data in an interdisciplinary setting necessitates strict adherence to data protection regulations and consent requirements. Despite the challenges, *LuxTIME* can serve as a model for future interdisciplinary research initiatives. The challenging task of establishing a common 'vocabulary' (knowledge ground, metaphor) between humanities and natural sciences was addressed, showcasing the potential for successful collaboration. With ongoing collaborative efforts and a persistent commitment to refining interdisciplinary approaches within Luxembourg, *LuxTIME* could serve as a starting point for expanding its scope beyond the Minett region, encompassing the entirety of Luxembourg (covered *e.g.* by the one year *LuxTIME* INITIATE project 2022).

All in all this thesis contributes to the expanding body of knowledge on the historical exposome, highlighting the interdisciplinary nature of exposomics research. Consequently, it is expected to foster increased (international) collaboration across disciplines within this field. Some of the ideas presented within may have elicited mixed responses within the community, but this was part of the intent behind the manifesto (CHAPTER II)—to challenge prevailing thinking for the overall benefit of the field as it continues to evolve. The future steps presented here aim to advance the field of exposomics research, contribute to evidence-based decision-making, and ultimately enhance our understanding of the historical exposome.

SYNOPSIS REFERENCES

1. Luxembourg U du. IAS. Université du Luxembourg. Published 2020. Accessed June 6, 2023. https://wwwfr.uni.lu/ias

2. Luxembourg U du. Luxembourg Time Machine. Université du Luxembourg. Published 2019. Accessed June 6, 2023. https://wwwfr.uni.lu/ias/running_audacity_projects/2019_audacity_projects/luxembourg_time_machine

3. Time Machine Europe. Time Machine Europe. Accessed April 21, 2021. https://www.timemachine.eu/

4. Wild CP. Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol. 2005;14(8):1847-1850. doi:10.1158/1055-9965.EPI-05-0456

5. Vrijheid M. The exposome: a new paradigm to study the impact of environment on health. Thorax. 2014;69(9):876-878. doi:10.1136/thoraxjnl-2013-204949

6. Miller GW, Jones DP. The Nature of Nurture: Refining the Definition of the Exposome. Toxicol Sci. 2014;137(1):1-2. doi:10.1093/toxsci/kft251

7. Rappaport SM. Genetic Factors Are Not the Major Causes of Chronic Diseases. Scott RJ, ed. PLOS ONE. 2016;11(4):e0154387. doi:10.1371/journal.pone.0154387

8. Aurich D, Miles O, Schymanski EL. Historical Exposomics And High Resolution Mass Spectrometry. Exposome. 2021;1(1):1-15. doi:10.1093/exposome/osab007

9. Aurich D, Horaniet Ibanez A, Hissler C, et al. Historical Exposomics: A Manifesto. Exposome. Published online August 18, 2023:1-35. doi:10.1093/exposome/osad007

10. Engel T. Basic Overview of Chemoinformatics. J Chem Inf Model. 2006;46(6):2267-2277. doi:10.1021/ci600234z

11. Ljoncheva M, Stepišnik T, Džeroski S, Kosjek T. Cheminformatics in MS-based environmental exposomics: Current achievements and future directions. Trends Environ Anal Chem. 2020;28:e00099. doi:10.1016/j.teac.2020.e00099

12. Kim S, Chen J, Cheng T, et al. PubChem 2023 update. Nucleic Acids Res. 2023;51(D1):D1373-D1380. doi:10.1093/nar/gkac956

13. Aurich D, Horaniet Ibañez A. How can data visualization support interdisciplinary research? LuxTIME: studying historical exposomics in Belval. Front Big Data. 2023;6:1164885. doi:10.3389/fdata.2023.1164885

14. Arp HPH, Aurich D, Schymanski EL, Sims K, Hale SE. Avoiding the Next Silent Spring: Our Chemical Past, Present, and Future. Environ Sci Technol. 2023;57(16):6355-6359. doi:10.1021/acs.est.3c01735

15. Horai H, Arita M, Kanaya S, et al. MassBank: a public repository for sharing mass spectral data for life sciences. J Mass Spectrom. 2010;45(7):703-714. doi:10.1002/jms.1777

16. ACS. CAS REGISTRY. CAS. Published 2023. Accessed June 8, 2023. https://www.cas.org/cas-data/cas-registry

17. Lai A, Singh RR, Kovalova L, Jaeggi O, Kondić T, Schymanski EL. Retrospective non-target analysis to support regulatory water monitoring: from masses of interest to recommendations via in silico workflows. Environ Sci Eur. 2021;33(1):43. doi:10.1186/s12302-021-00475-1

18. Aurich D, Diderich P, Helmus R, Schymanski EL. Non-target screening of surface water samples to identify exposome-related pollutants: a case study from Luxembourg. Environ Sci Eur. 2023;35(1):94. doi:10.1186/s12302-023-00805-5

19. Hollender J, Schymanski EL, Singer HP, Ferguson PL. Nontarget Screening with High Resolution Mass Spectrometry in the Environment: Ready to Go? Environ Sci Technol. 2017;51(20):11505-11512. doi:10.1021/acs.est.7b02184

20. Hollender J, van Bavel B, Dulio V, et al. High resolution mass spectrometry-based non-target screening can support regulatory environmental monitoring and chemicals management. Environ Sci Eur. 2019;31(1):42. doi:10.1186/s12302-019-0225-x

21. Helmus R, ter Laak TL, van Wezel AP, de Voogt P, Schymanski EL. patRoon: open source software platform for environmental mass spectrometry based non-target screening. J Cheminformatics. 2021;13(1). doi:10.1186/s13321-020-00477-w

22. Schymanski EL, Singer HP, Slobodnik J, et al. Non-target screening with high-resolution mass spectrometry: critical review using a collaborative trial on water analysis. Anal Bioanal Chem. 2015;407(21):6237-6255. doi:10.1007/s00216-015-8681-7

23. Alygizakis N, Lestremau F, Gago-Ferrero P, et al. Towards a harmonized identification scoring system in LC-HRMS/MS based non-target screening (NTS) of emerging contaminants. TrAC Trends Anal Chem. 2023;159:116944. doi:10.1016/j.trac.2023.116944

24. Caballero-Casero N, Belova L, Vervliet P, et al. Towards harmonised criteria in quality assurance and quality control of suspect and non-target LC-HRMS analytical workflows for screening of emerging contaminants in human biomonitoring. TrAC Trends Anal Chem. 2021;136:116201. doi:10.1016/j.trac.2021.116201

25. Helmus R, Velde B van de, Brunner AM, Laak TL ter, Wezel AP van, Schymanski EL. patRoon 2.0: Improved non-target analysis workflows including automated transformation product screening. J Open Source Softw. 2022;7(71). doi:10.21105/joss.04029

26. Hawkins E. 2018 visualisation update, Climate Lab Book. Climate Lab Book. Published 2018. Accessed October 25, 2022. https://www.climate-lab-book.ac.uk/2018/2018-visualisation-update/

27. Santoro C, Kirkland C. The warming stripes - Inspiring a movement. Nightingale Mag. 2022;Winter 2022(02):52-60.

28. Aurich D, Schymanski EL, Thiessen PA. GitLab repository “Environmental Cheminformatics / chemicalstripes.” GitLab. Published June 6, 2023. Accessed June 12, 2023. https://git-r3lab.uni.lu/eci/chemicalstripes

Appendices

Link to the *LuxTIME* website:

https://luxtimemachine.uni.lu

Link to the *LuxTIME* inventory

Coming soon

Link to the *chemicalStripes* R package:

https://gitlab.lcsb.uni.lu/eci/chemicalstripes

The accompanying data sources for all included manuscripts can be found in the respective chapters and are listed below:

Chapter IV:

GitHub repository:

https://github.com/DagnyAurich/Dust-paper

Chapter VI:

GitLab repository:

https://gitlab.lcsb.uni.lu/eci/data_luxwater_nt_paper_da

GNPS MassIVE repository:

MSV000092221, accessible via ftp://massive.ucsd.edu/MSV000092221/

Closing Page

A compilation of SOME of the WORST chemistry jokes of my career:

Why can you never trust an atom? – They make up literally everything.

Want to hear a Potassium joke? – K!

Carbon and hydrogen went on a date. I heard they really bonded.

If you're not part of the solution – you're part of the precipitate.

What emotional disorder does chromatography suffer from? – Separation anxiety.

What do you call iron blowing in the breeze? Febreeze!

How often should you tell a chemistry joke? – Periodically.

I once told a chemistry joke. There was no reaction.

To end with….

Think like a proton. Always positive.