Open and FAIR data for nanofiltration in organic media: A unified approach

Simon Van Buggenhout[1], Gergo Ignacz[2,3], Scout Caspers[1], Robin Dhondt[1,4], Marie Leneaerts[1,4], Nathalie Lenaerts[1], Sareh Rezaei Hosseinabadi[1], Ines Nulens[1], Guy Koeckelberghs[4], Yi Ren[5], Ryan Lively[5], Murielle Rabiller-Baudry[6], Ki Min Lim[7], Nazlee Ghazali[7], Joaquin Coronas[8,9], Milan Abel[10,11], Matthias Wessling[10,11], Mirko Skiborowski[12,13], Adam Oxley[14], Seok Ju Han[14], Andrew Livingston[14], Zhuan Yi[15], Congjie Gao[15], Kecheng Guan[16], Ralph Rolly Gonzales[16], Hideto Matsuyama[16], Jeffrey R. McCutcheon[17], Farzane Radmanesh[18], Nieck E. Benes[18], Akbar Asadi Tashvigh[19], Qing Fang[20,21], Kaisong Zhang[20,21], Guining Chen[22], Wanqin Jin[22], Yatao Zhang[23], Baowei Su[15,24], Chun-xu Zhang[22], Shi-Peng Sun[22,25,26], Anita Buekenhoudt[27], Chen Zhao[28], Bart Van der Bruggen[28], Jeong F. Kim[29,30], Lucas C. Condes[31], Matthew T. Webb[31], Michele Galizia[31], Banan Alhazmi[32], Lakshmeesha Upadhyaya[32], Suzana P. Nunes[32], Dae Woo Kim[33], Henrik Schröter[34], Udo Kragl[34], Sven Störtte[35], Andreas J. Vorholt[35], P. Zeynep Culfaz-Emecen[36], Marie-Alix Pizzoccaro-Zilamy[37], Louis Winnubst[37], Alexey Yushkin[38], Alexey Volkov[38], John Chau[39], Kamalesh K. Sirkar[39], Shao Lu[40], Gyorgy Szekely[2,3,**], Ivo Vankelecom[1], Rhea Verbeke[1,*]

[1] Membrane technology group (MTG), Division cMACS, Faculty of Bioscience Engineering, KU Leuven, Celestijnenlaan 200F box 2454, 3001, Leuven, Belgium

[2] Advanced Membranes and Porous Materials Center, Physical Science and Engineering Division (PSE), King Abdullah University of Science and Technology (KAUST), Thuwal, 23955-6900, Saudi Arabia

[3] Chemical Engineering Program, Physical Science and Engineering Division (PSE), King Abdullah University of Science and Technology (KAUST), Thuwal, 23955-6900, Saudi Arabia

[4] Laboratory for Polymer Synthesis, KU Leuven, Celestijnenlaan 200F, B-3001 Heverlee, Belgium.

[5] School of Chemical and Biomolecular Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

[6] Univ Rennes, CNRS, ISCR (Institut des Sciences Chimiques de Rennes) – UMR 6226, F-35000 Rennes, France

[7] Faculty of Chemical and Energy Engineering, Universiti Teknologi Malaysia (UTM), 81310, Skudai, Malaysia

[8] Instituto de Nanociencia y Materiales de Aragón (INMA), Universidad de Zaragoza-CSIC, 50018 Zaragoza, Spain

[9] Chemical and Environmental Engineering Department, Universidad de Zaragoza, 50018 Zaragoza, Spain.

[10] Chemical Process Engineering AVT.CVT, RWTH Aachen University, Forckenbeckstraße 51, 52074 Aachen, Germany

[11] DWI - Leibniz-Institute for Interactive Materials, Forckenbeckstraße 50, 52074 Aachen, Germany

[12] Laboratory of Fluid Separations, TU Dortmund University, Emil-Figge-Str. 70, 44227 Dortmund, Germany

[13] Institute of Process Systems Engineering, Hamburg University of Technology, Am Schwarzenberg-Campus 4, 21073 Hamburg, Germany

[14] School of Engineering and Materials Science, Queen Mary University of London, London, E1 4NS, United Kingdom

[15] Key Laboratory of Marine Chemistry Theory and Technology, Ministry of Education/Frontiers Science Center for Deep Ocean Multispheres and Earth System/College of Chemistry and Chemical Engineering, Ocean University of China, Qingdao 266100, China
[16] Research Center for Membrane and Film Technology, Department of Chemical Science and Engineering, Kobe University, Rokkodaicho 1-1, Nada, Kobe 657-8501, Japan
[17] Chemical & Biomolecular Engineering, University of Connecticut, 191 Auditorium Road, Unit 3222, Storrs, CT, 06269-3222, USA
[18] Films in Fluids, Department of Science and Technology, MESA+ Institute for Nanotechnology, University of Twente, P.O. Box 217, 7500 AE Enschede, the Netherlands
[19] Biobased Chemistry and Technology, Wageningen University & Research, Bornse Weilanden 9, 6708 WG Wageningen, the Netherlands
[20] College of Environmental Science and Engineering, Ocean University of China, Qingdao 266100, China
[21] Key Lab of Marine Environment and Ecology, Ministry of Education, Ocean University of China, Qingdao 266100, China
[22] State Key Laboratory of Materials-Oriented Chemical Engineering, College of Chemical Engineering, Nanjing Tech University, PR China
[23] School of Chemical Engineering, Zhengzhou University, Zhengzhou, 450001, China. zhangyatao@zzu.edu.cn.
[24] College of Chemistry & Chemical Engineering, Ocean University of China, 238 Songling Road, Qingdao, 266100, China
[25] Jiangsu National Synergetic Innovation Center for Advanced Materials, Nanjing Tech University, Nanjing, 210009, China
[26] College of Chemical Engineering, Nanjing Tech University, Nanjing, 210009, China
[27] Unit Separation and Conversion Technology, VITO N.V.—Flemish Institute of Technological Research, Boeretang 200, 2400 Mol, Belgium
[28] Department of Chemical Engineering, KU Leuven, Celestijnenlaan 200F, B-3001 Leuven, Belgium
[29] Department of Chemical Engineering, KU Leuven, Celestijnenlaan 200F, B-3001 Leuven, Belgium
[30] Innovation Center for Chemical Engineering, Incheon National University, Incheon 22012, Korea
[31] School of Sustainable Chemical, Biological and Materials Engineering, University of Oklahoma, 100 E. Boyd Street, Norman, 73019, USA
[32] Biological and Environmental Science and Engineering Division (BESE), Advanced Membranes and Porous Materials Center, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia
[33] Department of Chemical and Biomolecular Engineering, YONSEI University, Yonsei-ro 50, Seodaemun-gu, Seoul, 03722, Republic of Korea
[34] Department of Chemistry, University of Rostock, Albert-Einstein-Straße 3a, D-18059 Rostock, Germany
[35] Max-Planck-Institute for Chemical Energy Conversion, Stiftstraße 34-36, 45740 Mülheim, Germany
[36] Middle East Technical University, Chemical Engineering Department, Ankara, Turkey
[37] Inorganic Membranes, MESA+ Institute for Nanotechnology, University of Twente, 7500 AE, Enschede, the Netherlands
[38] A.V.Topchiev Institute of Petrochemical Synthesis, Russia Academy of Science (TIPS RAS), 29, Leninsky prospekt, Moscow, 119991, Russian Federation
[39] Otto York Department of Chemical, Biological and Pharmaceutical Engineering, New Jersey Institute of Technology, University Heights, Newark, New Jersey 07102, United States

[40] MIIT Key Laboratory of Critical Materials Technology for New Energy Conversion and Storage, State Key Laboratory of Urban Water Resource and Environment, and School of Chemistry and Chemical Engineering, Harbin Institute of Technology, Harbin 150001, China.

Corresponding authors:
* rhea.verbeke@kuleuven.be, +3216374230
** gyorgy.szekely@kaust.edu.sa, +966128082769

Abstract

Organic solvent nanofiltration (OSN), also called solvent-resistant nanofiltration (SRNF), has emerged as a promising technology for the removal of impurities, recovery of solutes, and the regeneration of solvents in various industries, such as the pharmaceutical and the (petro)chemical industries. Despite the widespread use of OSN/SRNF, the presence of scattered, non-standardized data, and the absence of openly accessible data pose critical challenges to the development of new membrane materials and processes, their comparison to the state-of-the-art materials, and their fundamental understanding. To overcome these hurdles, data from peer-reviewed research articles and commercial datasheets were curated via a standardized procedure to obtain an extensive dataset on the membrane materials, synthesis parameters, operational conditions, physicochemical properties, and performance of OSN/SRNF membranes. Thanks to a truly impressive joint effort of the OSN/SRNF community, the dataset contains, as per April 2024, 5157 entries from 294 publications for 42 solvents under several process parameters. This findable, accessible, interoperable, reproducible, and open (FAIR/O) dataset is available on both the OSN Database and the newly inaugurated Open Membrane Database for SRNF (OMD4SRNF). These databases provide multiple visualization and data exploration tools. Here, the standardized procedure applied to curate the data and the functionality of the

databases are outlined, as well as the online user interface to deposit new data by external users on the OMD4SRNF. This community-led project has been supported by all the co-authors of this work. Most importantly, they additionally agreed to systematically deposit their future peer-reviewed data on OSN/SRNF into the databases. We thereby pave the road for FAIR/O data in the field of OSN/SRNF to increase transparency, enable more accurate data analysis, and foster collaboration and innovation.

Keywords

Big data, OSN Database, Open Membrane Database, OMD4SRNF, organic solvent nanofiltration, solvent-resistant nanofiltration

Highlights

- New FAIR/O dataset on OSN/SRNF membranes, with 5157 datapoints (as per April 2024), containing synthesis parameters, operational conditions and membrane performance metrics
- A data standardization protocol is proposed, following FAIR/O data practices, to allow more accurate data comparison within the membrane community and beyond
- Inauguration of the Open Membrane Database for SRNF, called ‘OMD4SRNF’, that includes a tool to submit data by external users
- Comprehensive data visualization and exploration tools available on the OSN Database and the OMD4SRNF

- Community-led commitment to share future data on both the OSN Database and the OMD4SRNF

1 Introduction

The urgent societal and environmental challenges the world is facing require an equally urgent acceleration of solutions provided by, amongst others, research and technology. It is crucial to tackle the most pressing challenges, such as providing affordable energy and clean water, as highlighted by the United Nations Sustainable Development Goals [1]. Membrane technology has been proposed as a potential candidate to help reach these goals thanks to their high energy-efficiency, modularity, compactness, absence of waste creation, and ease of operation, compared to conventional separation techniques [2].

Membrane technology offers an alternative for energy- and solvent-intensive separations, such as distillation, extraction, and chromatography [3]. It is already extensively used in desalination [4], (waste)water treatment [5], and gas separations [6], among others, and is expected to play a major role in future energy conversion and storage devices [7]. Different types of membrane technologies exist, of which nanofiltration (NF) is of particular interest to reject molecules smaller than 1000 g mol^{-1}, such as dyes and micropollutants, down to divalent ions from contaminated water streams [8]. When NF membranes are used to treat organic media, the technology is called organic solvent nanofiltration (OSN) or solvent-resistant NF (SRNF) [2, 9]. When small organic molecules of similar size (usually less than 200 g mol^{-1}) need to be separated, organic solvent reverse osmosis (OSRO) membranes can be used [10, 11]. OSN/SRNF and OSRO are receiving increased attention in academia and in pharmaceutical, fine chemical, and oil industries, but their progress and understanding is significantly hindered by, amongst others, the data management practices in place [12, 13]. Although effort has been directed towards standardizing and optimizing the reporting and measurement practices for OSN/SRNF applications [14], the membrane synthesis and performance data currently present in literature is often scattered, non-standardized, and blocked behind paywalls. These practices endanger accurate data comparison, and hamper the development and understanding of new membrane materials and processes [12].

The recent advances in computational power and digital storage efficiency have enabled the (partial) digitalization of industrial, commercial, and research sectors. This digital transformation has given rise to repositories and large databases that allow researchers to share and analyze information [15]. When the data is additionally made findable, accessible, interoperable, reusable, and open (FAIR/O), machine-actionability, verification, replication, and reuse of data is facilitated [16]. Databases, either generic or domain-specific, play a critical role in developing better theoretical models, identifying suitable materials for specific applications, and answering key research questions [17].

While many research fields have already created and adopted large repositories in the 60s and 70s, such as the Protein Data Bank [18], the development of membrane-focused datasets and databases has only recently begun. Existing databases, such as PolyInfo [19], the Polymer Genome Project [20], the Cambridge Structural Database [21] and the Database of Zeolite Structures [22], provide information about organic and inorganic materials, but have a limited connection to membranes. In 2012, the first synthetic membrane-related database, the Polymer Gas Separation Membrane Database, was established, but it has not been maintained since 2018 [23]. Recently, a small open access dataset on the upper-bound for OSRO was revealed [24]. The first two large databases that focus on liquid separations with membranes are the OSN Database [25] and the Open Membrane Database (OMD) [26], both established in 2021.

The OSN Database is a free, open access database focusing specifically on OSN/SRNF applications and currently hosts more than 6000 entries with more than 250 000 manually curated or calculated features on solute, membrane, solvent and process parameters [27]. Besides providing freely available data and various data visualization options, the OSN Database also allows machine learning [28] and contains enantiomer separation and solute rejection prediction tools [29]. The OSN Database currently hosts performance data from commercial and custom-made membranes from literature.

The OMD is a free, crowd-sourced, and open-access platform containing FAIR/O data on membrane performance and physicochemical properties, as well as several membrane performance calculators and unit converters [30]. As the initial OMD only contained data on RO desalination membranes, it was coined 'OMD4RO' [26]. The data in the OMD is sourced from peer-reviewed journals, patents, and commercial product datasheets, and can be explored via different visualization tools. The data can also be exported as a text file for further processing. Importantly, the OMD contains an online submission tool that enables users to upload their peer-reviewed data so that the database is up to date with the newest scientific findings.

A new, large dataset containing the initial OSN data from the OSN Database expanded with a new dataset from literature is now revealed. The applied data sourcing, curation, and standardization procedures to obtain this dataset are outlined. The so-called OMD4SRNF, the sister database of the OMD4RO, is hereby inaugurated and contains an online tool for submission of peer-reviewed data by external users. The datasets will be shared between the OSN Database and OMD4SRNF, and both databases will be maintained and kept open access. The data standardization proposed herein follows FAIR data practices to allow more accurate data comparison and meta-analyses within the membrane community and beyond. As all co-authors have committed to deposit their peer-reviewed data of subsequent related research into both databases, the sustainable digitalization of the OSN/SRNF field via FAIR/O data is ensured, together with the sincere hope that every researcher in the field will contribute their data as well.

2 Data sourcing and curation

2.1 General constraints

The OSN Database and the OMD4SRNF focus on mainly NF membrane performance, membrane materials, synthesis parameters, and operational testing conditions in organic media. In procuring data from literature, certain guidelines were followed to safeguard the scope of the new dataset and respect the intellectual integrity of the reports.

First, only NF membranes tested in organic media are included with a molecular weight cut-off (MWCO) value set between 200 – 1000 g mol^{-1} as the main demarcation criterion [2]. However, some membranes with a MWCO outside of this range are also included, provided that they are part of a series of which at least one membrane possesses a MWCO inside the set range, either within a series of synthesis parameters that were screened, or in at least one solvent or with one solute. These exceptions are made to enable larger meta-analyses in the future. In addition, largely all data available in literature regarding the separation of binary mixtures (mostly based on toluene/ 1,3,5-triisopropylbenzene), which falls in the OSRO regime, have also been included.

However, since OSRO typically focuses on the separation of different solvents or other small molecules (instead of solvent/solute separation more typical of OSN/SRNF), incorporation of more OSRO data into the OMD4SRNF will be addressed in a future database. Furthermore, in some OSRO applications there is no clear distinction between solvent/solute, thus many OSRO membranes are characterized using a separation factor or C_p/C_r (concentration in permeate/ concentration in retentate) ratio, rather than with a rejection value. Multicomponent or non-binary OSRO separations data will be therefore addressed in future versions of the database.

Second, all membrane types (*i.e.*, thin-film composite (TFC), thin-film nanocomposite (TFN), integrally skinned asymmetric (ISA), dense, commercial, and inorganic) are included in the dataset. Third, only filtrations performed with solutes in either pure solvents or binary solvent mixtures are incorporated. Membranes tested exclusively in water and water-solvent mixtures are not included at this point to distinguish from aqueous NF and solvent-tolerant NF [31], respectively. These fields are considered separate domains with their own set of challenges and optimization requirements, therefore requiring a separate database. However, aqueous filtrations were included in the database if the same membrane was also tested in organic media.

To enable a better understanding of the membrane synthesis–structure–performance relationships of OSN/SRNF membranes, detailed information on the membrane material, structure, physicochemical properties, synthesis parameters, testing conditions, and membrane performance is also collected. Overall, the aim is to strike a balance between the number of collected parameters, the associated workload, and the database power, while sustaining sufficient future submissions of novel data. The full list of the documented parameters is presented in the Supporting Information. A more detailed 'readme' file, aimed at guiding users through the data submission process, is available online and in the Supplementary Information. Below, a general overview of the collected parameters and the rationale behind the selection are given in more detail. Note that, despite our sustained effort to avoid erroneous data from entering the databases, some errors might be present. Any suspicious datapoints can be flagged by external users on the OMD4SRNF, resulting in temporarily removal from the dataset. The founding team will then review these datapoints and resolve or permanently remove them.

2.2 Collected membrane parameters

The membrane parameters that are collected for this dataset include the membrane material and structure, the synthesis method, and the modifications applied during and after synthesis, similar to the OMD4RO [26]. Membrane physicochemical properties, such as thickness, water contact angle, and roughness, are optionally included as well.

Six common membrane categories are considered: ISA, TFC, TFN, inorganic (single-material or composite), and free-standing (*i.e.*, dense, backing-free) materials, and are defined as follows [32]. ISA membranes possess a dense skin supported by a porous substructure, prepared from a single composition via phase inversion (alternatively called phase separation) [33]. TFC membranes consist of a thin, selective layer (either fully organic or fully inorganic) on top of a more open support, typically an ultrafiltration or microfiltration membrane [34]. TFN membranes are structurally similar to TFC membranes but contain a nanomaterial (*e.g.,* metal organic frameworks (MOFs), zeolites, graphene oxide) embedded in the selective layer [35]. For the purpose of this work, the presence of nanomaterials in the top layer is required to qualify as a TFN membrane. Nanomaterials present in other membrane parts are regarded as a modification strategy instead. Inorganic membranes are mainly comprised of an inorganic material, such as ceramics [36], and are subdivided here based on their structure. Single-material inorganic membranes are comprised of a single inorganic material, while inorganic composite membranes consist of a thin layer of an inorganic material, deposited on top of an inorganic support of a different composition. Lastly, dense, free-standing membranes, were also considered. Although not used in the industrial practice, these membranes are of crucial importance to understand fundamental transport aspects and develop structure-property correlations needed to design membranes and processes exhibiting enhanced permeability, rejection, and long-term durability [37, 38]. Note that for certain membrane preparation methods, appointing a membrane category is ambiguous *e.g.,* a selective layer grafted onto an ISA membrane. For ambiguous cases, the classification presented in the source material is preferentially followed. Alternatively, a membrane classification guide is also available in the 'readme' file.

A set of synthesis parameters was collected depending on the used membrane synthesis method, including phase inversion and interfacial polymerization (IP) (**Figure 1**). Although the phase inversion process is known to be very sensitive to a plethora of parameters [32, 33], not all synthesis conditions are always reported by default in literature [14]. Therefore, a careful selection of relevant synthesis parameters was made, including polymer chemistry, polymer concentration and solvent (mixture) of the dope solution, non-solvent (mixture), additive nature and concentration, and casting thickness. For inorganic membranes, only the chemistry was included. For membranes derived from IP, information on the aqueous and organic phase, and the monomer nature and concentration were collected. This is to, among other objectives, further guide the efforts directed towards the synthesis of top layers to achieve a better performance than obtainable with conventional polyamide-based top layers. The monomers and additives were specified via their simplified molecular-input line-entry specification (SMILES) string (Supporting Information). Next to IP, other thin-film deposition methods like spin coating, dip coating, grafting, spray coating, kiss coating, layer-by-layer deposition, electrospray coating, casting, pressurized filtration, and vacuum filtration were documented as well. Additional parameters that are collected for TFN membranes were the embedded nanomaterial name, its concentration and position during synthesis (*i.e.*, its dispersion either in the water phase or in the organic one when using IP). For TFC and TFN membranes, support layer parameters are collected as well, according to their support structure (*i.e.*, ISA, inorganic, fibrous or commercial). Irrespective of the membrane structure, information on the membrane post-treatment strategies, including the SMILES string of the chemical agents and their concentration, was also documented.

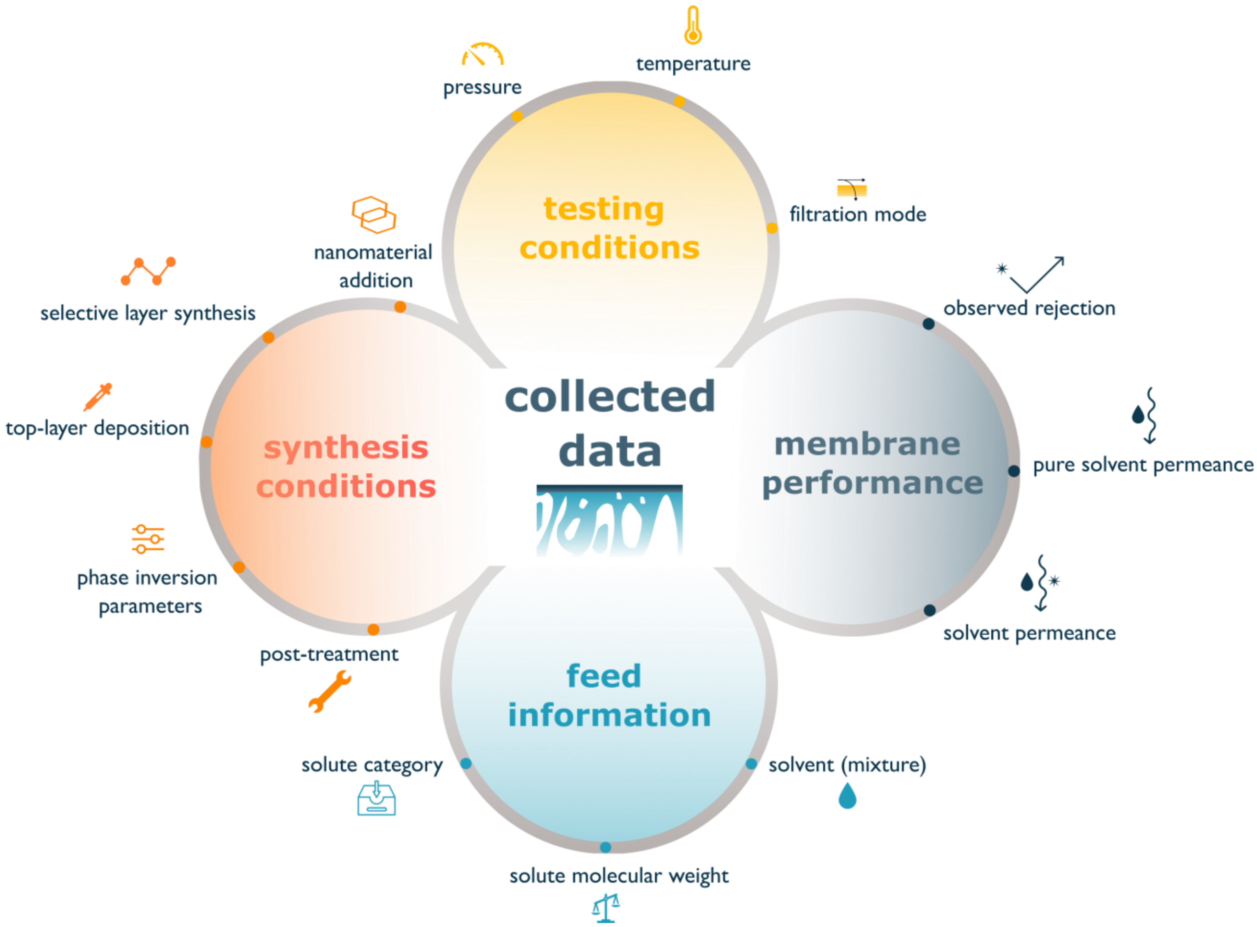

Figure 1. Overview of the different categories for which data was collected as input for the dataset. The main categories are synthesis conditions, feed information, testing conditions, and membrane performance. The most important parameters per category are also shown.

2.3 Collected testing parameters

Other membrane applications, like RO or gas separations, are defined by a limited set of relevant feed compositions, allowing the compilation of trade-off plots and the accompanying upper bound relationship for membrane performance, of which both fields have greatly benefitted [26, 39, 40]. Such convenient benchmarking tools have thus far been absent for the field of OSN/SRNF due to the quasi-infinite variety of used solute(s) and solvent(s), and the complexity of the resulting solute–solvent–membrane interactions. This complex interplay hinders direct comparison of membrane performance between different solvents, membrane chemistries, different solutes of similar MW, and even between different operational conditions (*i.e.,* solute concentration, temperature, transmembrane membrane pressure, detection methods) [14, 41]. In an attempt to enable more accurate membrane comparison, detailed information on the membrane testing conditions was collected. Parameters related to the membrane rejection for a specific solute (mixture) used in the feed solution include the solute concentration, MW and category, as defined in reference [41]. Solute structures were represented by their unique SMILES string, which is the default database representation for molecular structural data. SMILES also allows to calculate various molecular properties using software packages like Mordred and RDKit Python packages, as currently done in the OSN Database [28]. Several molecular descriptors and physicochemical properties related to the solute and solvent of interest are also available on the OMD4SRNF (Supporting Information). These descriptors, such as molar volume and dipole moment, were included to help better understand certain observations, but are by no means meant to be exhaustive. Next to solute information, data on other important experimental conditions, such as filtration mode (*i.e.*, cross-flow and dead-end), temperature, transmembrane pressure, and the nature of the solvent (mixture) were also collected.

2.4 Collected membrane performance parameters

To characterize membrane performance, observed solute rejection (R_{obs}, further denoted as R, **Eq. S1**), pure solvent permeance (*i.e.*, measured in the absence of solutes), solvent permeance (*i.e.*, measured in the presence of solutes), and MWCO values were collected, only if explicitly reported in the source material (**Eq. S2**). Note that R_{obs} can differ from the real solute rejection, R_{real}, due to concentration polarization and fouling. However, due to the limited availability of

R_{real} in the literature, only R_{obs} is documented. Despite recognizing the potential significance of fouling in some OSN/SRNF applications, no parameters related to these phenomena are collected at this point due to the very limited number of available studies [42-44].

3 Database content and functionalities

3.1 Database content

The OSN Database and the OMD4SRNF are continuously growing databases, with the data on both platforms being coupled. Both will remain separately accessible free of charge to exploit the data via the different tools that exist on each platform. The OSN Database hosts existing datasets (*i.e.*, Datasets 1–4) [25, 45-47], and an entirely new dataset (*i.e.*, Dataset 5) curated for the purpose of this work and for the inauguration of the OMD4SRNF (**Table 1**). Now all 5 datasets are available on both databases. Dataset 5 consists of 294 peer-reviewed articles, published between 1999 and 2024, and compiled from a literature search across 7 search engines (Supporting Information, **Table S1**). Some additional peer-reviewed manuscripts that fulfill the general dataset requirements but that were not found through the literature search, were provided by the authors and were also incorporated. The datasets on the OSN Database and the OMD4SRNF fall under a CC-BY-4.0 and a CC BY-NC-4.0 license, respectively.

Table 1. Information on the different datasets hosted by the OSN Database and the OMD4SRNF. A datapoint is one individual value either collected from literature, or calculated from literature data (*e.g.*, solute MW calculation from solute structure, permeance calculated from pressure and flux). A dataset is a compilation of several entries. A database is a collection of datasets. Dataset 5 is based on available data as per April 2024.

Database	Dataset	Membrane types	Nr° of datapoints	Membrane chemical structure	Launch year	Ref.
OSN Database	Dataset 1	Commercial	44 342	No	2021	[47]

	Dataset 2	Commercial	2 840	Yes	2022	[25]
	Dataset 3	Commercial	9 336	Yes	2023	[46]
	Dataset 4	Commercial, tailor-made ISA	15 504	Yes	2023	[45]
	Total (1-4)	Commercial, tailor-made ISA	72 022	Partial	2021–2023	[27]
OMD4SRNF & OSN Database	Dataset 5	Commercial, tailor-made TFC, tailor-made ISA	5 157	Yes	2024	This work
	Total (1-5)	Commercial, tailor-made ISA, tailor-made TFC	149 201	Yes	2024	This work

3.2 Online submission tool & database maintenance

The OMD4SRNF contains an online submission tool that allows external users to deposit their peer-reviewed data in the database in a step-by-step fashion, and free of charge. The required input follows the same structure as outlined above, with a distinction made between mandatory and optional fields. Only submissions possessing a valid DOI that are published in a peer-reviewed journal are eligible for submission to the OMD4SRNF [26]. All data deposited by external users is checked for any errors by the OMD4SRNF team and then published on both the OMD4SRNF and the OSN Database. Additionally, a GitHub mirror repository of the dataset and the code base is available as an open-access backup.

The sustainability of the OMD is currently directly dependent on the involvement of the membrane community by uploading their data. While the commitment of all co-authors to supply their future work is a step in the right direction, on-boarding the entire field of OSN/SRNF will require a shift in the academic structures in place so that contributing to the common good through, amongst others, FAIR/O data management is supported and domain-specific databases are valued [17, 48]. Several concrete actions are currently undertaken to advocate for and accelerate this shift to ultimately ensure database sustainability. While the maintenance of the OMD and the OSN Database is currently ensured by the founding teams of the respective databases, an international call for so-called OMD ambassadors will be launched to further involve the field. These ambassadors will be asked to encourage data supply and revise uploaded data, similar to associate editors of journals. In addition, the OMD is advocating to partner up with journals related to membrane technology, as they have the leverage to enforce data publication in an open repository. The OMD can then be populated from such a repository by data discovery tools, ensuring the longevity of the OMD. Lastly, sustained involvement of the membrane societies would also be befitting as the aim of the OMD is to become a database-hub for the entire field of membrane technology. Action is currently being undertaken to include more detailed synthesis information on ceramic membranes and hollow fiber membranes, and to expand the OMD from RO and OSN/SRNF to, amongst others, gas separation and ion-exchange membranes.

3.3 Data processing interface

Dataset 1-5 can be explored on the OMD website through an interactive chart featuring various search functionalities (**Figure 2**). Users can manipulate the chart layout to compare membrane properties, synthesis parameters, filtration conditions, and membrane performances. The x and y axes (in log-normal or linear scale) can be freely selected from a range of numerical properties (*e.g.,* (pure) solvent permeance, R_{obs}, MWCO, polymer concentration in dope solution, filtration pressure and temperature, report year). The legend display can be altered between a selection of quantitative (*e.g.,* report year, (pure) solvent permeance, R_{obs}) and non-quantitative (*i.e.,* selective layer chemistry and membrane structure) properties. At the time of writing, 21 different filter types can be applied simultaneously to narrow down the displayed data, allowing facile data analysis. Numerical filters (*e.g.,* solute MW) allow specification of a minimum-maximum value range, while non-numerical filters (*e.g.,* filtration mode) permit selection of the desired value from a list. Highlighting individual datapoints prompts a pop-up tag to display, where applicable, the datapoint name, the membrane chemistry, solute name, database ID, digital object identifier (DOI) of the source document, and the x and y coordinates. Alternatively, lasso or rectangle selection tools allow highlighting multiple datapoints at once. All selected points are tabulated at the bottom of the page together with additional information that can be selected by the users via the displayed columns. The database plot and the data table can be exported as an image (.png) and a data file (.csv), respectively. Note that the displayed or exported information can potentially consist of a smaller number of datapoints compared to the entire dataset, as not all collected parameters are always disclosed in the source material nor are they applicable to all membrane categories.

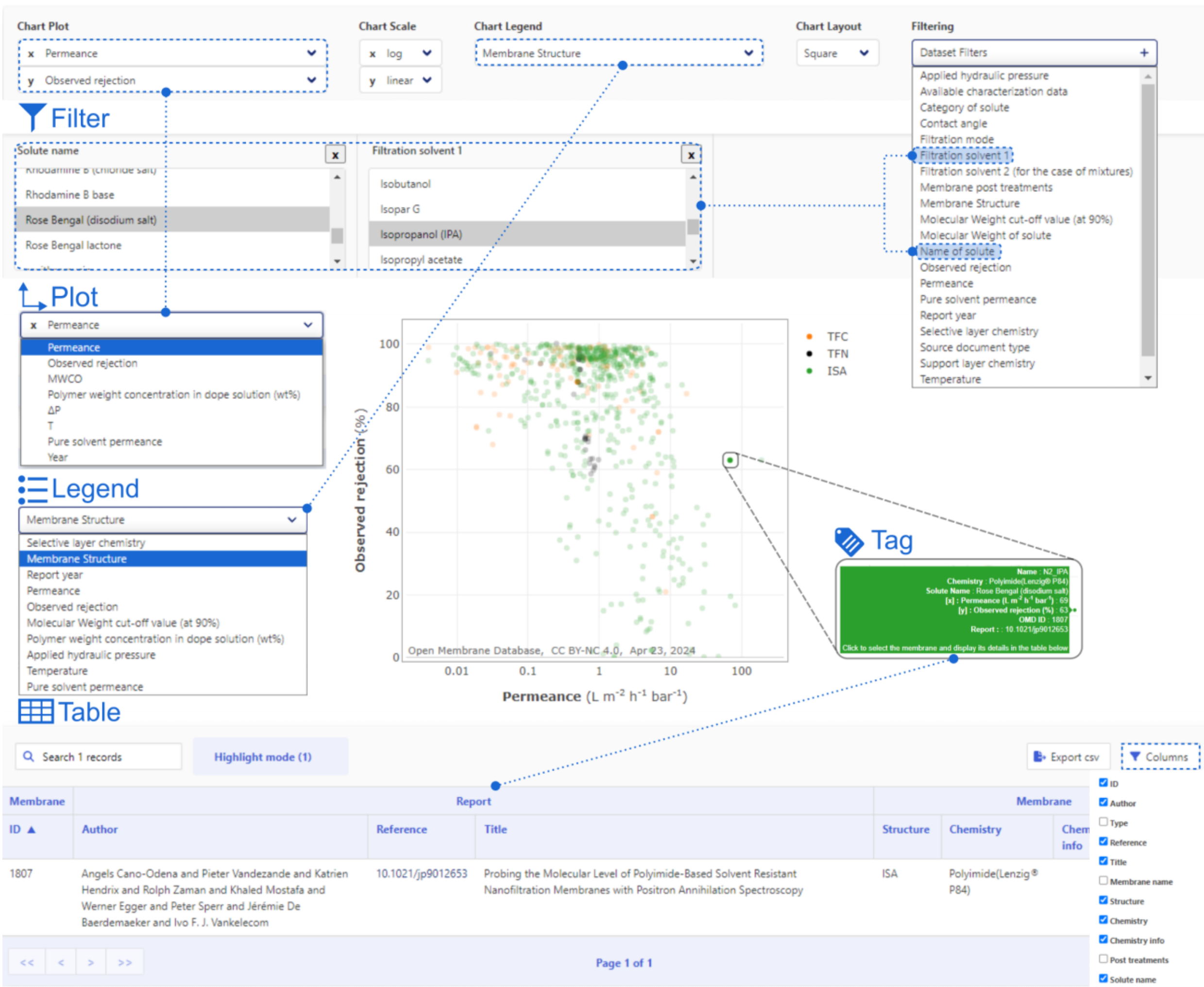

Figure 2. Visual representation of the user interface for data exploration on the Open Membrane Database (OMD). Different fields used for chart manipulation, filtering, and data retrieval are highlighted. 'Plot' and 'Legend' fields are used to manipulate the chart layout. Data can be extensively filtered using the 'Filter' fields. Selecting a datapoint from the chart will prompt a pop-up 'Tag' to display additional information. Selected datapoints are displayed in a filterable 'Table' at the bottom of the page, where different properties can be selected. The data table can be exported as a csv file. For illustration purposes, this image has been modified.

3.4 Brief analysis of collected data

As per April 2024, the newly compiled dataset (*i.e.*, Dataset 5, **Table 1**) contains data gathered from 294 publications from 2000 to 2024, adding up to 5157 datapoints (**Figure 3a**). A brief analysis of the dataset is given here to highlight some general observations about the field of OSN/SRNF and to spur the field to conduct more in-depth meta-analyses with *e.g.*, machine learning tools. More than 7000 crossflow and 4500 dead-end filtration measurements are included (**Figure 3b**), of which more than one third has a rejection value of > 90% (**Figure 3c**). The dataset also consists of a wide range of solutes, as demonstrated by the high solute MW coverage (**Figure 3d**), and a large set of chemistries, both for the active layer as well as for the support layer (**Figure 3e,h**)**.** Polyamides dominate the active layer chemistry while polyacrylonitrile is the most common support chemistry. The most prevalent membrane structure in the dataset is ISA, followed by TFC and commercial membranes (**Figure 3f**). By showcasing the monomers used during IP for the synthesis of TFC membranes in a tree manifold approximation and projection (TMAP) plot, the molecular structural similarity of the used monomers is discernible (**Figure 3g**). The mostly used monomers are depicted as well. The diverse and comprehensive dataset opens up the possibilities for future data exploration and analysis to achieve a better understanding of membrane synthesis-structure-performance relationships and to accelerate the development of novel membrane materials. The filtration data can also be used to design filtration systems for specific separations, for example by using the Open Membrane System Design Tool [49] or PROSYN® Membranes [50].

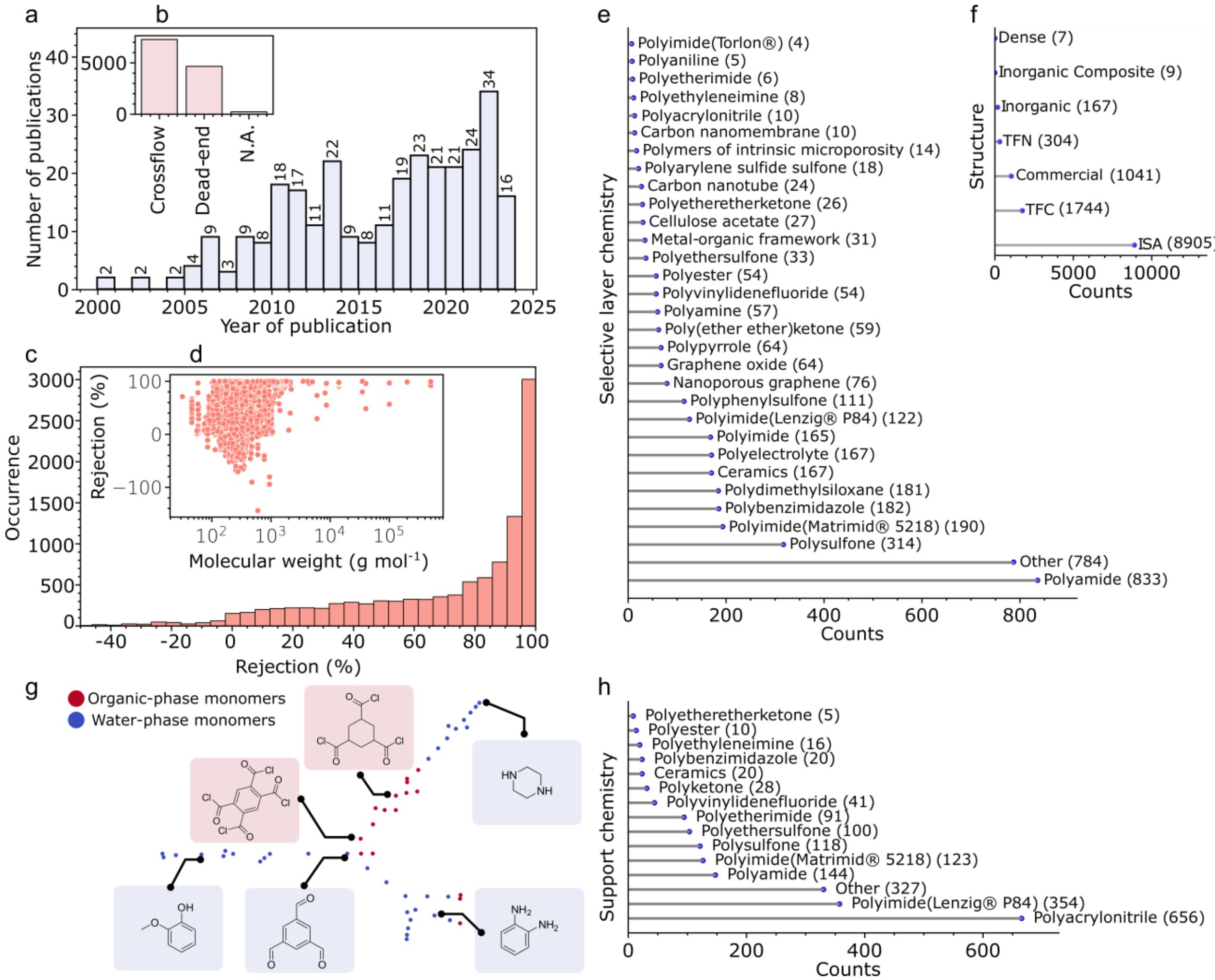

Figure 3. High-level summary of dataset 5 curated for this work. **a**) The number of publications over the years (2000 – 2024) with their yearly count. **b**) The crossflow versus the dead-end filtration ratios. N.A. indicates that the information was not available in the source document. **c**) Rejection histogram. **d**) Rejection as a function of the molecular weight. **e**) Different membrane structure types and their occurrence. 'Other' indicates that the chemistry was not available in the predefined options and had to be specified by the user. **f**) Different membrane structures and their occurrence. **g**) TFC monomer types visualized on a TMAP diagram. Structurally similar monomers are closer to each other. **h**) Different support structures and their occurrence.

4 Conclusions

With this initiative, the OSN/SRNF community stands united and presents a data standardization and sharing approach that enables the sustainable digitalization of the field. A new dataset was curated, compiling 294 peer-reviewed articles ranging from 2000 – 2024, and is available on the OSN Database, as well as on the newly inaugurated OMD4SRNF. Membrane synthesis parameters (including the membrane material, structure, support and top layer chemistry, monomers, solvents, additives, and modifications), testing conditions, and membrane performance are documented in detail. This large and diverse dataset, consisting of more than 5157 datapoints, allows investigating the synthesis-structure-performance relationship of OSN/SRNF membranes from a variety of angles, enabling *e.g.*, meta-analyses, machine actionability, and statistical data interpretation. External users are encouraged to submit their latest peer-reviewed data via the online submission tool available on the OMD4SRNF, which is then directly linked to the OSN Database. The commitment of all authors to supply their future work to the databases is already a significant acceleration in the direction of large and FAIR/O data, which paves the road for increased understanding and innovation in the field of OSN/SRNF.

CRediT author statement

SVB: Conceptualization, Methodology, Validation, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Supervision, Project administration. **GI**: Conceptualization, Software, Writing - Original Draft, Writing - Review & Editing, Visualization. **GZ**: Conceptualization, Resources, Writing - Review & Editing, Funding acquisition. **IVK**: Conceptualization, Resources, Writing - Review & Editing, Funding acquisition, Project administration. **RV**: Conceptualization, Methodology, Validation, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Supervision, Project administration. **All other co-authors**: Data Curation, Writing - Review & Editing

Acknowledgments

The website and back-office development of the OMD, and by extension the OMD4SRNF, is performed with the help of Thomas Stassin and Pierre Biezemans from codefathers.be. The research reported in this publication was supported by funding from King Abdullah University of Science and Technology (KAUST) and KU Leuven. This work was supported by the Research Foundation Flanders (FWO) through R.V.'s junior postdoctoral fellowship (1216222N), S.C. and M.L.'s doctoral fellowships (1SF9823N and 1S93122N) and project grant G0C6623N. This work was supported by the Flemish government through Industrieel Onderzoeksfonds (IOF) grants

(VTI-23-00181, C3/21/066 and IOFM/17/008) and by VLAIO through the RENOVATE 2 program (HBC.2022.0536). Y.R. and R.P.L. were supported by the National Science Foundation (DMREF-1921873).

MG gratefully acknowledges financial support from the US National Science Foundation (NSF-CBET) under the CAREER Program (#2043648)

Declaration of competing interest

The authors declare that they do not have competing financial interests.

References

[1] U. Nations, Transforming Our World: The 2030 Agenda for Sustainable Development, (2015).
[2] P. Marchetti, M.F.J. Solomon, G. Szekely, A.G. Livingston, Molecular Separation with Organic Solvent Nanofiltration : A Critical Review, Chem. Rev. 114 (2014) 10735-10806. https://doi.org/10.1021/cr500006j.
[3] D.S. Sholl, R.P. Lively, Seven chemical separations to change the world, Nature 532(7600) (2016) 435-437. https://doi.org/https://doi.org/10.1038/532435a.
[4] M. Elimelech, W.A. Phillip, The Future of Seawater Deaslination: Energy, Technology, and the Environment, Science 333 (2011) 712-718.
[5] E. Obotey Ezugbe, S. Rathilal, Membrane Technologies in Wastewater Treatment: A Review, Membranes (Basel) 10(5) (2020). https://doi.org/10.3390/membranes10050089.
[6] M. Galizia, W.S. Chi, Z.P. Smith, T.C. Merkel, R.W. Baker, B.D. Freeman, 50th Anniversary Perspective: Polymers and Mixed Matrix Membranes for Gas and Vapor Separation: A Review and Prospective Opportunities, Macromolecules 50(20) (2017) 7809-7843. https://doi.org/10.1021/acs.macromol.7b01718.
[7] S.P. Badwal, S.S. Giddey, C. Munnings, A.I. Bhatt, A.F. Hollenkamp, Emerging electrochemical energy conversion and storage technologies, Front Chem 2 (2014) 79. https://doi.org/10.3389/fchem.2014.00079.
[8] M. Paul, S.D. Jons, Chemistry and fabrication of polymeric nanofiltration membranes: A review, Polymer 103 (2016) 417-456. https://doi.org/10.1016/j.polymer.2016.07.085.
[9] P. Vandezande, L.E.M. Gevers, I.F.J. Vankelecom, Solvent resistant nanofiltration: separating on a molecular level, Chemical Society Reviews 37(2) (2008) 365-405. https://doi.org/10.1039/B610848M.
[10] R.P. Lively, D.S. Sholl, From water to organics in membrane separations, Nature Materials 16(3) (2017) 276-279. https://doi.org/10.1038/nmat4860.
[11] D.-Y. Koh, B.A. McCool, H.W. Deckman, R.P. Lively, Reverse osmosis molecular differentiation of organic liquids using carbon molecular sieve membranes, Science 353 (2016) 804-807. https://doi.org/10.1126/science.aaf1343.
[12] P.-J. Piccard, P. Borges, B. Cleuren, J. Hooyberghs, A. Buekenhoudt, Organic Solvent Nanofiltration and Data-Driven Approaches, Separations 10(9) (2023). https://doi.org/10.3390/separations10090516.

[13] S.L. Aristizábal, R.P. Lively, S.P. Nunes, Solvent and thermally stable polymeric membranes for liquid molecular separations: Recent advances, challenges, and perspectives, Journal of Membrane Science 685 (2023). https://doi.org/10.1016/j.memsci.2023.121972.
[14] H.A. Le Phuong, C.F. Blanford, G. Szekely, Reporting the unreported: the reliability and comparability of the literature on organic solvent nanofiltration, Green Chemistry 22(11) (2020) 3397-3409. https://doi.org/10.1039/d0gc00775g.
[15] R. Kitchin, The data revolution: Big data, open data, data infrastructures and their consequences, Sage2014.
[16] (!!! INVALID CITATION !!! [16]).
[17] A.N. Quay, P.S. Fiske, M.S. Mauter, Recommendations for Advancing FAIR and Open Data Standards in the Water Treatment Community, ACS ES&T Engineering 2(3) (2021) 337-346. https://doi.org/10.1021/acsestengg.1c00245.
[18] H. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, Weissig, Helge , I.N. Shindyalov, P.E. Bourne, The Protein Data Bank, Nucleic Acids Research 28(1) (2000) 235-242.
[19] M. Tanifuji, A. Matsuda, H. Yoshikawa, Materials Data Platform - a FAIR System for Data-Driven Materials Science, 2019 8th International Congress on Advanced Applied Informatics (IIAI-AAI), 2019, pp. 1021-1022.
[20] C. Kim, A. Chandrasekaran, T.D. Huan, D. Das, R. Ramprasad, Polymer Genome: A Data-Powered Polymer Informatics Platform for Property Predictions, The Journal of Physical Chemistry C 122(31) (2018) 17575-17585. https://doi.org/10.1021/acs.jpcc.8b02913.
[21] C.R. Groom, I.J. Bruno, M.P. Lightfoot, S.C. Ward, The Cambridge Structural Database, Acta Crystallogr B Struct Sci Cryst Eng Mater 72(Pt 2) (2016) 171-9. https://doi.org/10.1107/S2052520616003954.
[22] C.B.L.B. McCusker, Database of Zeolite Structures, 1996.
[23] A.W. Thornton, B.D. Freeman, L.M. Robeson, Polymer Gas Separation Membranes. https://membrane-australasia.org/polymer-gas-separation-membrane-database/, 2012.
[24] R. Lively, OSRO Upper Bound. https://lively.chbe.gatech.edu/osro-upper-bound, 2022 (accessed 16 June 2023).
[25] G. Ignacz, C. Yang, G. Szekely, Diversity matters: Widening the chemical space in organic solvent nanofiltration, Journal of Membrane Science 641 (2022) 119929. https://doi.org/10.1016/j.memsci.2021.119929.
[26] C.L. Ritt, T. Stassin, D.M. Davenport, R.M. DuChanois, I. Nulens, Z. Yang, A. Ben-Zvi, N. Segev-Mark, M. Elimelech, C.Y. Tang, G.Z. Ramon, I.F.J. Vankelecom, R. Verbeke, The open membrane database: Synthesis–structure–performance relationships of reverse osmosis membranes, Journal of Membrane Science 641 (2022). https://doi.org/10.1016/j.memsci.2021.119927.
[27] OSN Database. https://osndb.kaust.edu.sa/, 2021.
[28] G. Ignacz, G. Szekely, Deep learning meets quantitative structure–activity relationship (QSAR) for leveraging structure-based prediction of solute rejection in organic solvent nanofiltration, Journal of Membrane Science 646 (2022) 120268. https://doi.org/10.1016/j.memsci.2022.120268.
[29] A.K. Beke, G. Szekely, Enantioselective nanofiltration using predictive process modeling: Bridging the gap between materials development and process requirements, Journal of Membrane Science 663 (2022) 121020. https://doi.org/https://doi.org/10.1016/j.memsci.2022.121020.
[30] The Open Membrane Database. https://openmembranedatabase.org/, 2021.

[31] A. El Fadil, R. Verbeke, M. Kyburz, P. E M Aerts, I.F.J. Vankelecom, From academia to industry: Success criteria for upscaling nanofiltration membranes for water and solvent applications, Journal of Membrane Science 675 (2023). https://doi.org/10.1016/j.memsci.2023.121393.
[32] M. Mulder, Basic Principles of Membrane Technology. Second Edition, 1997.
[33] A.K. Hołda, I.F.J. Vankelecom, Understanding and guiding the phase inversion process for synthesis of solvent resistant nanofiltration membranes, Journal of Applied Polymer Science 132(27) (2015) n/a-n/a. https://doi.org/10.1002/app.42130.
[34] J.R. Werber, C.O. Osuji, M. Elimelech, Materials for next-generation desalination and water purification membranes, Nature Reviews Materials 1(5) (2016). https://doi.org/10.1038/natrevmats.2016.18.
[35] S. Ali, I.A. Shah, I. Ihsanullah, X. Feng, Nanocomposite membranes for organic solvent nanofiltration: Recent advances, challenges, and prospects, Chemosphere 308(Pt 2) (2022) 136329. https://doi.org/10.1016/j.chemosphere.2022.136329.
[36] R.B. Merlet, A. Nijmeijer, L. Winnubst, Hybrid ceramic membranes for organic solvent nanofiltration : State-of-the-art and challenges, Journal of Membrane Science 599(January) (2020) 117839-117839. https://doi.org/10.1016/j.memsci.2020.117839.
[37] A.M. Tandel, W. Guo, K. Bye, L. Huang, M. Galizia, H. Lin, Designing organic solvent separation membranes: polymers, porous structures, 2D materials, and their combinations, Materials Advances 2(14) (2021) 4574-4603. https://doi.org/10.1039/d1ma00373a.
[38] M. Galizia, K.P. Bye, Advances in Organic Solvent Nanofiltration Rely on Physical Chemistry and Polymer Chemistry, Front Chem 6 (2018) 511. https://doi.org/10.3389/fchem.2018.00511.
[39] Z. Yang, H. Guo, C.Y. Tang, The upper bound of thin-film composite (TFC) polyamide membranes for desalination, Journal of Membrane Science 590 (2019). https://doi.org/10.1016/j.memsci.2019.117297.
[40] L.M. Robeson, The upper bound revisited, Journal of Membrane Science 320(1-2) (2008) 390-400. https://doi.org/10.1016/j.memsci.2008.04.030.
[41] R. Verbeke, I. Nulens, M. Thijs, M. Lenaerts, M. Bastin, C. Van Goethem, G. Koeckelberghs, I.F.J. Vankelecom, Solutes in solvent resistant and solvent tolerant nanofiltration: How molecular interactions impact membrane rejection, Journal of Membrane Science 677 (2023). https://doi.org/10.1016/j.memsci.2023.121595.
[42] A. Oxley, A.G. Livingston, Anti-fouling membranes for organic solvent nanofiltration (OSN) and organic solvent ultrafiltration (OSU): graft modified polybenzimidazole (PBI), Journal of Membrane Science 662 (2022). https://doi.org/10.1016/j.memsci.2022.120977.
[43] S.B. Sigurdardóttir, M.S.M. Sueb, M. Pinelo, Membrane compaction, internal fouling, and membrane preconditioning as major factors affecting performance of solvent resistant nanofiltration membranes in methanol solutions, Separation and Purification Technology 227 (2019). https://doi.org/10.1016/j.seppur.2019.115686.
[44] G.M. Shi, M.H. Davood Abadi Farahani, J.Y. Liu, T.S. Chung, Separation of vegetable oil compounds and solvent recovery using commercial organic solvent nanofiltration membranes, Journal of Membrane Science 588 (2019) 117202-117202. https://doi.org/10.1016/j.memsci.2019.117202.
[45] G. Ignacz, A.K. Beke, G. Szekely, Data-driven investigation of process solvent and membrane material on organic solvent nanofiltration, Journal of Membrane Science 674 (2023) 121519. https://doi.org/https://doi.org/10.1016/j.memsci.2023.121519.

[46] G. Ignacz, N. Alqadhi, G. Szekely, Explainable machine learning for unraveling solvent effects in polyimide organic solvent nanofiltration membranes, Advanced Membranes 3 (2023) 100061. https://doi.org/10.1016/j.advmem.2023.100061.
[47] J. Hu, C. Kim, P. Halasz, J.F. Kim, J. Kim, G. Szekely, Artificial intelligence for performance prediction of organic solvent nanofiltration membranes, Journal of Membrane Science 619 (2021) 118513. https://doi.org/10.1016/j.memsci.2020.118513.
[48] R. Verbeke, FAIR and Open Data requires proper incentives and a shift in academic culture, Nature Water 1(1) (2023) 7-9. https://doi.org/10.1038/s44221-022-00012-1.
[49] F.J. Aschmoneit, C. Hélix-Nielsen, OMSD – An open membrane system design tool, Separation and Purification Technology 233 (2020). https://doi.org/10.1016/j.seppur.2019.115975.
[50] P.D. Center, PROSYN Membranes. https://www.process-design-center.com/prosyn-membranes/, 2024 (accessed 15 May 2024).