UC Irvine
UC Irvine Previously Published Works

Title

Overview of ICARUS ▫ A Curated, Open Access, Online Repository for Atmospheric Simulation Chamber Data

Permalink

https://escholarship.org/uc/item/1f80z7rt

Journal

ACS Earth and Space Chemistry, 7(6)

ISSN

2472-3452

Authors

Nguyen, Tran B
Bates, Kelvin H
Buenconsejo, Reina S
et al.

Publication Date

2023-06-15

DOI

10.1021/acsearthspacechem.3c00043

Peer reviewed

Article

Overview of ICARUS—A Curated, Open Access, Online Repository for Atmospheric Simulation Chamber Data

Tran B. Nguyen,* Kelvin H. Bates, Reina S. Buenconsejo, Sophia M. Charan, Eric E. Cavanna,
David R. Cocker, III, Douglas A. Day, Marla P. DeVault, Neil M. Donahue, Zachary Finewax,
Luke F. Habib, Anne V. Handschy, Lea Hildebrandt Ruiz, Chung-Yi S. Hou, Jose L. Jimenez, Taekyu Joo,
Alexandra L. Klodt, Weimeng Kong, Chen Le, Catherine G. Masoud, Matthew S. Mayernik, Nga L. Ng,
Eric J. Nienhouse, Sergey A. Nizkorodov, John J. Orlando, Jeroen J. Post, Patrick O. Sturm,
Bridget L. Thrasher, Geoffrey S. Tyndall, John H. Seinfeld, Steven J. Worley, Xuan Zhang,
and Paul J. Ziemann

Cite This: *ACS Earth Space Chem.* 2023, 7, 1235−1246

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Atmospheric simulation chambers continue to be indispensable tools for research in the atmospheric sciences. Insights from chamber studies are integrated into atmospheric chemical transport models, which are used for science-informed policy decisions. However, a centralized data management and access infrastructure for their scientific products had not been available in the United States and many parts of the world. ICARUS (Integrated Chamber Atmospheric data Repository for Unified Science) is an open access, searchable, web-based infrastructure for storing, sharing, discovering, and utilizing atmospheric chamber data [https://icarus.ucdavis.edu]. ICARUS has two parts: a data intake portal and a search and discovery portal. Data in ICARUS are curated, uniform, interactive, indexed on popular search engines, mirrored by other repositories, version-tracked, vocabulary-controlled, and citable. ICARUS hosts both legacy data and new data in compliance with open access data mandates. Targeted data discovery is available based on key experimental parameters, including organic reactants and mixtures that are managed using the PubChem chemical database, oxidant information, nitrogen oxide (NOx) content, alkylperoxy radical (RO$_2$) fate, seed particle information, environmental conditions, and reaction categories. A discipline-specific repository such as ICARUS with high amounts of metadata works to support the evaluation and revision of atmospheric model mechanisms, intercomparison of data and models, and the development of new model frameworks that can have more predictive power in the current and future atmosphere. The open accessibility and interactive nature of ICARUS data may also be useful for teaching, data mining, and training machine learning models.

KEYWORDS: atmospheric chamber, database, data repository, data science, atmospheric chemistry and physics
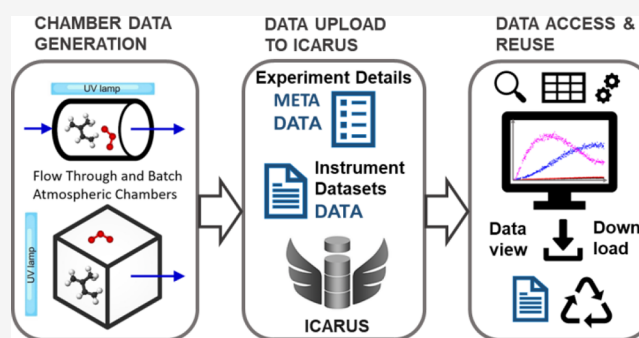
1. INTRODUCTION

Atmospheric simulation chambers (e.g., "smog" chambers, environmental chambers, flow reactors, continuously stirred reactors, etc.; Figure 1) are central to the laboratory study of atmospheric chemistry and physics.[1] These chambers serve as the critical link between "bench-top" laboratory research and ambient research by enabling scientists to study atmospheric chemistry at relevant time and length scales, but in a highly controlled manner.[2] These simulation chambers of variable or fixed gaseous volume (ranging from less than 1 m^3 to more than 200 m^3) are used to investigate reactions that encompass all phases of atmospheric matter (gaseous, particulate, aqueous, or mixed phases) and test the impacts of temperature, pressure, relative humidity, irradiation with ultraviolet or other wave-

lengths of light, oxidative exposure, and other factors on chemical reactions. Starting from the pioneering experiments of Haagen-Smit,[3] atmospheric chamber research has led to important discoveries in atmospheric chemistry, for example, new chemical mechanisms, quantification of rate coefficients and yields, and key insights into reaction dynamics. The
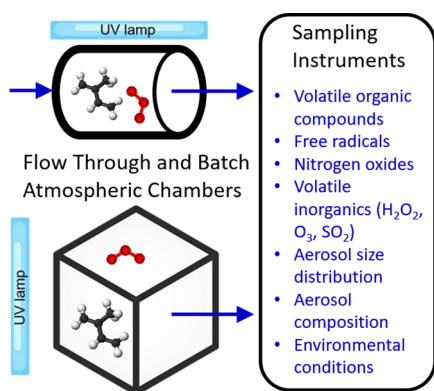
Figure 1. Simplified schematic of atmospheric simulation chambers and their measurements.

fundamental data obtained from chamber studies are routinely used in model mechanisms,[4−8] which define the known chemistry and empirical constraints in atmospheric chemical transport models that predict the chemical composition of the atmosphere and its health and climate feedbacks. See the Section 2 for a description of a general chamber reaction.

The new U.S. federal open access mandate[9] and the International Science Council's mission[10] to promote the open availability of data for research has amplified the need for open access data repositories in multiple disciplines. There are numerous atmospheric chamber facilities in the world, and more are currently under construction, yet there is no central data management and storage infrastructure for the vast amounts of highly heterogeneous information produced from chamber research within the United States and many parts of the world. Unlike observational field data and model revisions, laboratory chamber data are not routinely made accessible to the research community at large, nor to the public. Such a data infrastructure would also serve as a critical quality control mechanism and enable the facile synthesis of proprietary data across multiple chamber studies for important use applications such as model mechanism development, training machine learning algorithms, or teaching. A data repository and access hub was recently developed for the atmospheric chambers in Europe through the Eurochamp project.[11] Considerable community interest in developing such a project has been also expressed by scientists in the United States and beyond.[12,13] Finally, these efforts would shield atmospheric chamber experimental work, which started around the 1950s with the discovery of the reactions in smog,[14] from being lost when research labs retire.

ICARUS is a collaborative effort from the experimental atmospheric chemistry and physics community to build the first open access, searchable, web-based infrastructure for storing, sharing, searching, and utilizing atmospheric chamber data in North America. The overarching scientific goal of the project is to promote collaboration between atmospheric science researchers and facilitate the sharing and reuse of publicly funded data by increasing uniformity and expanding access to atmospheric chamber experiments. Furthermore, an important fundamental goal of ICARUS is to support the data principles of FAIR (findable, accessible, interoperable, and reusable).[15,16] The pilot cohort of the ICARUS project included approximately 30 chamber and flow reactor experimentalists from the California Institute of Technology, Carnegie Mellon University, Georgia Institute of Technology,

Harvard University, National Center for Atmospheric Research (NCAR), the University of California (UC) at Davis, Irvine, and Riverside, the University of Colorado at Boulder, and the University of Texas at Austin. The Scientific Steering Committee (SSC) of ICARUS is composed of the principal investigators of each research group; the SSC provided feedback and formal votes on all decisions. The repository web infrastructure was developed by the research group at UC Davis, working closely with the Data Stewardship and Engineering Team at NCAR. Each participant group in the pilot voted on project decisions and collaboratively tested each alpha version of the repository until a stable beta version emerged. The beta version was then tested by model mechanism developers (see Section 4). Comments and suggestions by all testers were implemented into the current version of the ICARUS website.

This paper provides an overview of the ICARUS data repository, including the goals of the project, functions, and features of the web interface, organizational system, data infrastructure and formats, technical considerations, and anticipated uses.ICARUS contains two interactive parts: a Data Intake portal and a Data Access portal. In this overview, we term the Data Contributor as a person who uses the Data Intake portal to provide manual entries of metadata descriptors and to upload data files on behalf of a group or organization. The Data User is defined as any person who uses the Data Access portal to view and/or download the data for reuse for any purpose. A key insight by the Eurochamp project is that ample metadata describing each chamber experiment is critically needed for Data Users to accurately interpret the data and lower the barrier to reuse.[17] ICARUS has focused on making these extensive metadata available for Data Users, while at the same time decreasing unnecessary work for the Data Contributor through automation of repetitive tasks where feasible.

2. EXPERIMENTAL SECTION

This section describes one example of a typical chamber experiment that may be hosted on ICARUS. Generally, atmospheric chambers are used to isolate reactions that would otherwise occur concurrently in the ambient atmosphere for a detailed study. Experimental methods that generate atmospheric chamber data will vary with the type of chamber, type of gaseous mixing mode (plug flow, batch, or continuously stirred reactors), experiment goals, type of reactant, and type of analytical instruments. Detailed overviews of atmospheric simulation chamber experimental methods are reported elsewhere.[18−25] A typical application of a cubical Teflon chamber (Figure 1, bottom) equipped with ultraviolet lights run in batch (stopped flow) mode is described briefly here. Consider, for example, a reaction between the volatile organic compound isoprene (C_5H_8), with the hydroxyl radical (OH) under a "non-polluted" chemical regime, where the organic peroxy (RO_2) radicals react primarily with hydroperoxy (HO_2) radicals.[26] Chemical reagents such as hydrogen peroxide (H_2O_2) and isoprene are injected at the start of the reaction at the desired concentrations. Concentrations of volatile compounds in the chamber are generally monitored as a function of reaction time by in situ chemical instruments such as gas chromatography-flame ionization detectors or chemical ionization mass spectrometers. When measuring secondary aerosol yields, seed particles (such as ammonium sulfate) are typically atomized into the chamber to serve as the
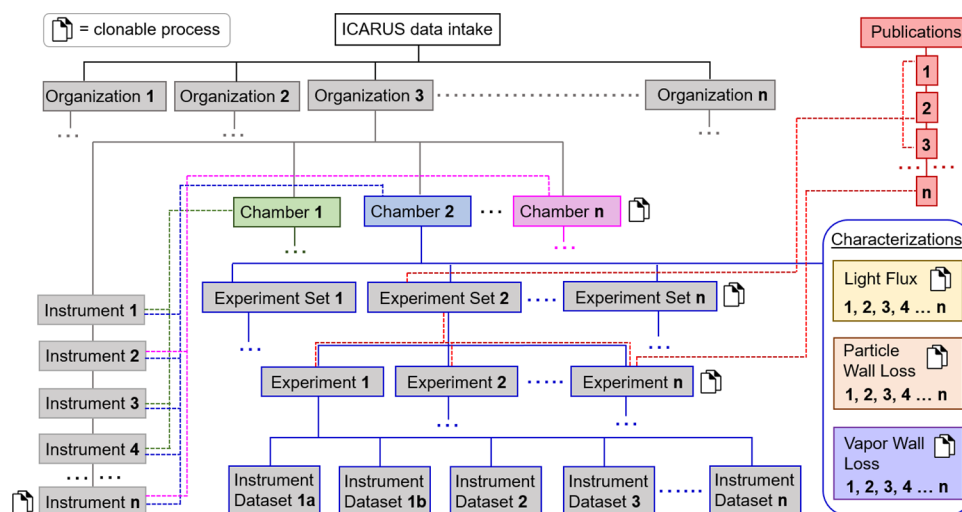
Figure 2. Schematic of how each entity relates to another in the ICARUS data ecosystem. Enumerations start from 1 and end in *n* (any number) for each entity. Publications are shown linked only to Experiment Set 2 for brevity, and Characterizations show no linkages; however, both can be linked to any Experiment Set or Experiment. Characterizations are also linked to Instruments because they require a dataset upload.

surfaces to which condensable vapors may partition. Particle concentrations and/or composition are monitored by in situ particle sizers and counters, aerosol mass spectrometers, particle into liquid samplers, or other analytical techniques. Either ultraviolet or simulated solar lights are turned on to initiate radical formation, e.g., photolysis of H_2O_2 and subsequent reactions will produce both OH and HO_2 radicals. The decay of reactants and generation of products are monitored for the duration of the experiment, often together with environmental conditions (temperature, relative humidity, and pressure). Lights are turned off after the objectives of the experiments are met, and then gases and particles might be sampled for offline analyses. Dark reactions (using ozone, nitrate radical, or other dark oxidants) may also be tested in atmospheric chambers in the absence of irradiation. The data from chemical and particle instruments are then processed for later use and may be uploaded onto ICARUS together with experimental details.

3. RESULTS: THE ICARUS FRAMEWORK

This section describes the user interfaces and technical process for the intake and access of data and metadata through the ICARUS website.

3.1. Data Intake. Data intake is the process by which the Data Contributor categorizes and describes their scientific datasets and uploads the associated data. The data intake process is access-controlled with login access from user accounts. Please see the ICARUS user guide (Section S1) for technical details about the data intake process.

3.1.1. Technical Objectives. One of the primary technical goals of ICARUS is to adopt a data format standard that unifies the various file types and structures that come from the many analytical instruments available for atmospheric research (e.g., .xlsx, .txt, .hdf, .mat, .itx, and .ict). This enables Data Users to treat all downloaded data similarly, with the same software and reader script. Similarly, ICARUS seeks to digitize experimental metadata and adopt a uniform standard for metadata reporting from different research labs. This information may include descriptions of chambers, experiment goals, instrument sampling protocols, analytical uncertainties, and other information necessary for the Data User to accurately interpret

the experiment data. In that spirit, the metadata requirements of ICARUS are high. Users are required to extensively describe each measurement, experiment, and any linked entities; however, the process for data entry and uploads is streamlined to minimize repetitive work for the Data Contributor. Machine-based quality assurance checks are built into the data entry and upload process (e.g., minimum word requirements for descriptions of experiments, data file column checks, and required fields). Archiving legacy data with high amounts of metadata is a first priority, after which focus turns to managing current and future datasets. Given the increase in FAIR-aligned data availability requirements of large scientific publishers during the data submission or publication stage,[27−31] ICARUS is best used by the Data Contributor as an integrated part of the data workflow from experimentation to publication.

3.1.2. Data Ecosystem. The ICARUS data ecosystem includes a set of entities that are related to one another via parallel and hierarchical relationships, all of which are internally tracked (Figure 2). Data Contributor-provided descriptions of these entities are stored as the repository metadata and may be updated at any time.

At the top of the hierarchy is the Organization, which describes the research group, organized research unit, or other organizational entities to which the Data Contributor belongs. Immediate descendants of the Organization are the Instruments and Chambers. Instruments describe the analytical sensors and other scientific equipment that record the data uploaded to ICARUS. These Instruments may provide either in situ or "offline" measurements that characterize the chemical reactions or physical phenomena that are tested in the atmospheric chambers (e.g., concentrations of chemicals or particles, composition of chemicals or particles, spectral irradiance or absorbance, temperature, relative humidity). Chambers are at the same hierarchical level as Instruments. Chambers describe physical characteristics and other properties of the atmospheric chambers or flow reactors (as well as surrounding infrastructure), in which the scientific experiments are performed. These characteristics of the Chamber may include size, chamber material, environmental controls, a description of the lamps, and so forth. Instruments may be

Table 1. Experimental Fields and the Associated Controlled Vocabulary[a]

Experiment category(s)	Gas phase chemical reaction, condensed phase chemical reaction, multiphase chemical reaction, volatility and partitioning, hygroscopicity and phase changes, aerosol formation, aerosol aging, instrument/chamber characterization, blank/control, custom
Reaction type(s)	Non-chemical, photooxidation, hydrolysis, oligomerization, heterogeneous oxidation, other aqueous/aerosol phase, custom
Reactant name(s)[#]	Any substance or compound from PubChem[#], custom (free form entry)
Reactant functional group(s)	Aerosols, alcohol, aldehyde, alkane, alkene, alkyne, amine, anhydride, aromatic, cycloalkane, cycloalkene, diol, ester, ether, halocarbon, hydroperoxide, imine, inorganic acid, ketone, N-heterocycle, NOx, nitrate, nitrile, nitroalkane, nitroaromatic, O-heterocycle, organic acid, peracid, peroxyacid, sulfate, sulfide, sulfoxide, others
Seed	Ammonium sulfate, ammonium nitrate, sodium sulfate, sodium chloride, SOA, N/A, custom
Oxidant	None, hydroxyl radical, ozone, nitrate radical, chlorine radical, O atom, custom
RO_2 main fate[#]	HO_2, NO, RO_2, not sure, custom

[a]Field names marked with a [#] are described in more detail in the text.

linked to multiple Chambers within ICARUS; a linkage will define the analytical datasets that will be available to upload for each experiment performed in that Chamber. The immediate descendants of Chambers are Experiment Sets, which are groupings of individual experiments based on logic defined by the Data Contributor (Section S1.D).

The immediate descendants of Experiment Sets are the Experiments, including any background/control experiments, which are the records of the individual scientific investigations within the atmospheric chamber. The Experiment level has the most information available for Discovery purposes (Section 3.2) and, thus, more descriptions and supplemental information are provided at this level (Section S1.E). Besides descriptions of the Experiment categories, goals and outcomes, input of reactants, and other information about the Experiment (collectively termed the Experimental Metadata), there is a requirement for the upload of comma-separated value (csv) files: the Instrumental Dataset file(s) and Timeline file. The Instrumental Dataset files are a record of the data that have been processed to their final ready-to-use forms (e.g., raw signals that have been baseline-corrected, calibrated, and converted to concentration units). The Timeline is a two-column chronological list of experimental actions that occurred during the Experiment; it is provided so that the data can be appropriately interpreted by the Data User. For example, increases or decreases in the signal of certain compounds may be associated with the initiation of chemistry by turning on lights; similarly, discontinuities in data may be associated with instrument sampling issues.

In our experience, digitizing the experimental Timeline is a significant hurdle to the archiving of legacy atmospheric chamber data. The majority of experiment notes that exist in the community have been recorded in physical notebooks, often by a large rotating roster of academic or government personnel. Thus, the quality, quantity, and availability of such notes vary between labs and between research experimentalists within the labs. A more sustainable data management practice includes planning for the ICARUS upload in the data workflow by recording notes digitally during the experiment by using electronic laboratory notebooks or similar tools,[32,33] by using the Timeline generator tool or template available from ICARUS (Section 3.2.3), or by digitizing notes soon after an experiment.

3.1.2.1. Controlled Vocabulary. We control the vocabulary of many metadata fields through various means: for example, calendar entries for dates, radio buttons for yes and no, drop down menus with limited selections, and check boxes for multiple input terms. Table 1 provides the available terms for the non-freeform fields, where the vocabulary is controlled. These terms provide suitable keywords for sorting and searching.

Chemical reactant names and their synonyms in ICARUS are managed by the suite of controlled vocabulary from the PubChem databases,[34] the world's largest collection of freely accessible chemical information. It is necessary to manage chemical vocabulary because numerous synonyms exist for each chemical compound. The chemical synonyms result from variations in spelling or notation (e.g., HCHO, CH_2O, formaldehyde, and methanal), chemical site designations (e.g., isopropanol vs 2-propanol), and prefix symbols (e.g., b-myrcene, ß-myrcene, and beta-myrcene). This poses significant challenges for the search and discovery process, as incomplete search results will appear when Data Users query ICARUS using only one of the many existing synonyms. The usage of PubChem in ICARUS is beneficial because this allows the user to enter any known synonym for a certain compound/mixture into the data entry field, instead of needing to look up controlled names defined by other existing standards, for example, IUPAC or CAS. The user-input chemical entries will then be automatically validated by ICARUS by a rapid query with the PubChem Compound and Substance databases and updated onscreen with the "preferred" or "common" name in PubChem for that chemical or substance.

3.1.2.2. RO_2 Main Fate. In support of model mechanism development, a mechanism-focused sorting field called "RO_2 main fate" is used. Alkylperoxy radicals (RO_2) are formed in an oxidation reaction of hydrocarbons initiated by the OH, Cl, O, and NO_3 radical or by ozonolysis to various extents. Their "fate," or predominant bimolecular or unimolecular reaction, determines the chemical regime and thus product formation for that hydrocarbon precursor.[35,36] The "RO_2 main fate" describes whether most of the first-generation RO_2 radicals of that oxidation reaction react with NO and HO_2 radicals, NO_2 radicals, and other RO_2 radicals or via autoxidation (isomerization).[37,38] Custom entries are provided for experiments that study a different and/or less common fate, for example, with OH radicals or NO_3 radicals. This criterion replaces the more traditional "NOx regime" used in atmospheric chemistry, which has been noted as ambiguous.[39] The "General RO_2 Fate Estimator" program (Section 3.2.3) available on the ICARUS website uses IUPAC-recommended rates[40] to estimate RO_2 regimes based on user inputs of oxidant and hydrocarbon loadings.

Finally, each Experiment or Experiment Set can be linked to Publications and Characterizations in ICARUS. A Publication is a metadata record of published works in the scientific

Figure 3. Illustrative search and discovery table from the ICARUS website.

literature that are associated with the experimental data that were uploaded by the Data Contributor. Characterizations are experiments that are performed for the purpose of quantifying chamber wall loss coefficients of vapors and particles or the emission fluxes of chamber lights under various environmental conditions of the main Experiment. Chamber characterization data are often required for modeling the photochemistry and product formation in chamber experiments, and thus, are critical supplemental data for each Experiment or Experiment Set.

3.1.3. Task Automation and Saving. ICARUS is designed to offer both simplicity for the Data Contributor and comprehensiveness for the Data User. These goals are achieved by requiring extensive dataset descriptions and metadata as described in Section 3.1.2, while automating many repetitive tasks and saving progress to minimize lost work.

In order to increase efficiency in data entry and uploads, ICARUS provides a "Cloning" feature for duplicating Chambers, Instruments, Experiment Sets, Experiments, and Characterizations if the user seeks to make few changes to the entity (Figure 2, cloning icons). For example, a series of chamber experiments may seek to perform the same reaction but with different temperatures, relative humidities, seed particle surface area, or any other variable experimental parameter. In these cases, the Data Contributor will clone an experiment and change the field value only for the relevant parameter that was altered in the experiment. Experiments may also be moved to different Experiment Sets to allow for reorganization with the "Move" feature. Experiment names are automatically generated with the same syntax (which includes the Organization name, Experiment date, and other information) to save time and to increase uniformity. Data Contributors may also save their progress and return to data entry due to the Draft mode in ICARUS, which does not publish the data and ignores any errors from incomplete fields.

3.1.4. Data Format. When an Experiment or a series of Experiments are downloaded, ICARUS writes the data into a ASCII-based file format called YAML (Section S1.I) that is meant to be read into data processing programs written in, for example, R, Matlab, Python, or another scripting language, for further use by Data Users. An example of the data format is shown in Figure S2. A downloaded data packet (compressed .zip file) for one single Experiment contains the Experimental Metadata, all associated Instrument Datasets, all associated Characterization files, and a manifest file that describes the content of the folder and the data location on the ICARUS website. Multiple unrelated Experiments can be downloaded together in a single .zip file organized by Organization name.

The download data format of ICARUS is unique, which better serves the needs of the discipline, but may necessitate new data reader tools to be written. Fortunately, ICARUS can leverage the many free and open source data readers available for YAML, so we do not consider this to be a big disadvantage. Adoption of existing earth science data and metadata format standards for ICARUS proved challenging due to the lack of laboratory-specific, and particularly chamber-specific, fields and keywords. For example, the ICARTT format[41] from the National Aeronautics and Space Administration (NASA) is widely used for atmospheric field measurements. However, ICARTT caters to continuous measurements of the natural environment instead of experiment-based chamber research performed at intermittent frequency. The Global Change Master Directory (GCMD) includes a number of earth-science related keywords; however, insufficient laboratory-related keywords were available. The ICARUS data format integrates some GCMD keywords with other controlled vocabulary specific to atmospheric chamber experiments and with chemical vocabulary from PubChem.

3.2. Search and Discovery. ICARUS is also a search and discovery platform for atmospheric chamber data. The data search, browse, and sort functions in ICARUS are optimized for model mechanism development. Data Users can find data in two ways: a keyword search on the home page and the search results page, as well as browsing all data ("Show All Experiments") and then utilizing the filter, sorting, and category-specific search functions on the results page (Figure 3).
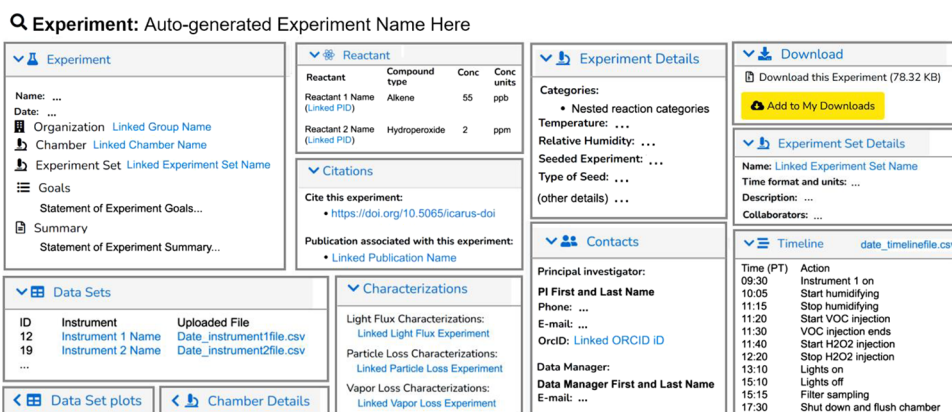
Figure 4. Simplified diagram of the Experiment View webpage from the ICARUS website. Certain noted entities are linked to their respective webpages with persistent url. Each csv file can be downloaded directly by clicking on its name. Most panels are abridged for brevity. Chamber Details and Data Set Plots panels are shown collapsed. The configuration of open/collapsed panels is automatically saved for each Data User. An expanded Data Set Plot panel is shown in Figure 5. A representative example of an Experiment page in ICARUS is shown in Figure S1 for Experiment number 691.
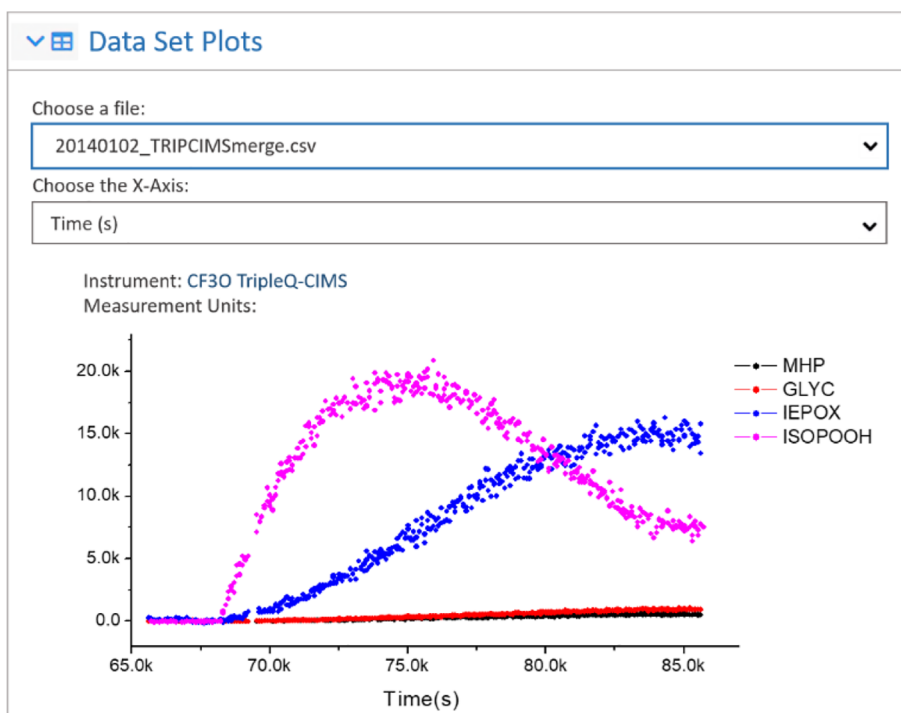

Figure 5. Representative data visualization panel results from the ICARUS website. Available data files are selectable with a drop down menu, and within that data file, all columns are plotted. Users may choose their ideal x-axis for plotting and choose data columns to show or hide. Nomenclature and units are defined by Data Contributors in the Instrument metadata for each dataset.

From the search and discovery table, the Data User can customize their viewing experience and further refine their search. The search and discovery table shows 12 default columns that are categories from which the Data User can narrow down their queries. Each category column is searchable with its own search box and sortable alphabetically. The default columns are chosen to represent categories that would appeal to the majority of users, for example, we expect most Data Users would want to search for keywords within the Experiment Name, PI name, Date, Experimental Category, Reaction Type, Reactant Name(s), Seed Name(s), Oxidant Name(s), RO_2 Main Fate, Temperature, Humidity, and whether Supplemental Information is available for that Experiment. In total, there are 22 categories that can be selected/deselected to show on the results table by clicking on the "Select columns to display" button, and these user preferences are retained on the website regardless of login status.

It is challenging to view many columns of metadata in a traditional scrolling format; thus, we included an "Expand" button (the blue plus sign, Figure 3) that shows hidden or collapsed metadata as an expanded list in the vertical direction without the need to scroll. This expansion function makes the search and discovery process friendly for mobile devices.

3.2.1. Viewing Experiments. Visitors to the website may download data through multiple mechanisms (Section S1.J) or click on an experiment for additional details before deciding to download the data. This takes them to the Experiment View

page (Figures 4 and S1), which shows all of the associated metadata, uploaded Timelines and Dataset, links to parent entities (Organization, Chamber, and Experiment Sets), links to associated Publication and Characterizations, links to Reactant PubChem ID (PID), and a link to download the Experiment, each as separate collapsible panels. The Experiment View page also allows direct downloads of individual Timeline and Dataset .csv files by clicking on their names. As Data Users may have different needs for the information available, the ICARUS website will remember the configuration of collapsed or expanded panels that were set up, without the need for a login, in order to optimize the unique viewing experience of the Data User.

Data Users can also visualize each dataset within an experiment with the interactive Data Plots feature (Figure 5) prior to deciding to download the Experiment. Upon selection of any available dataset, all data columns from that dataset are plotted. Large datasets are automatically down-sampled to display more quickly. Users can select their preferred independent (x-) axis from a drop-down list of column names, which is useful in situations where there are multiple potential independent axes. Columns of data can be turned on and off by clicking on the legend, which removes unwanted data on the plotted axis. The data acronyms are defined in the Instrument and/or Experimental metadata.

3.2.2. Shared Tools and Documentation. A number of freeware tools that were contributed by the community are also available to download on the repository website. These tools help Data Contributors with various steps in the data upload workflow. For example, Timeline generator tools, Timeline templates, RO_2 fate estimator tool,[42] and Oxidation flow reactor exposure estimator tool[43] help generate necessary files and metadata information for input into forms throughout the Data Intake process. A 30-min video guide is available to document the process of Data Intake and Access. The video and its transcribed script in English are available through the "Help" button on the front page and the search results page (https://icarus.ucdavis.edu/help). This video describes the entire process from signing up for an account to submitting data. Additional help is available by contacting the ICARUS team.

3.2.3. Mirroring and Indexing. All data in ICARUS are automatically discoverable and version-tracked at the Geoscience Data Exchange repository (GDEX, https://gdex.ucar.edu/search.html?q=icarus) that is operated by the National Center for Atmospheric Research (NCAR). All data from ICARUS experiments are mirrored within GDEX and receive a DOI for persistent identification, location, and citation. ICARUS is synced with GDEX daily. Changes to data files and metadata from ICARUS experiments are tracked within GDEX as separate versions. Prior versions are retained and available upon direct request to GDEX, but are not exposed to data users to prevent confusion. ICARUS data, including the versions mirrored at GDEX, are indexed on Web of Science and Google for discovery through general searches.

4. DISCUSSION: APPLICATIONS FOR THE REUSE OF ICARUS DATA

Some of the major applications of ICARUS involve modeling, for example, evaluation and revision of current model mechanisms, intercomparison of model mechanisms, and development of new frameworks for modeling atmospheric chemistry. In order to optimize the search and discovery

experience and validate data/metadata quality for modeling, ICARUS underwent a technical review and revision process with 13 international model mechanism developers from Harvard University, NCAR Atmospheric Chemistry Observations & Modeling group, Columbia University, UC Riverside, the US Environmental Protection Agency, Colorado State University, the University of Texas at El Paso, NASA Goddard Flight Center, the University of Cambridge, the University of York, and the National Oceanic and Atmospheric Administration (NOAA). The data in ICARUS were tested against the following models: GAMMA,[44] F0AM,[45] MusicBox,[46] SOM-TOMAS,[47] AtChem with MCM,[48] WRF-Chem,[49] GECKO-A,[50] BoxMox[51] and various custom models in R, Python, and Matlab. Modelers provided ratings and specific feedback regarding the repository website, including ease of access and navigation, usability of the search and browse functions, quality of data and metadata (including relevancy, completeness, etc.), quantity of metadata (including if there was enough information in the files and names to support the use application), display aesthetics, and multiple other considerations. Feedback from the scientific community of both Data Contributors and Data Users were incorporated into the current version of ICARUS.

ICARUS can also be used by instructors who wish to integrate public data to teaching atmospheric chemistry. Using data to teach may be helpful for classrooms that do not have the infrastructure to conduct these sophisticated experiments, but desire an active-learning model for teaching abstract concepts in atmospheric science such as photochemistry, heterogeneous reactions, kinetics, or surface deposition. Many classical experiments are available, such as VOC-NOx-O_3 reactions and the photo- and dark oxidations of hydrocarbons. Many of these experimental data are accompanied by published scientific articles that can be used for literature review by students. In particular, the interactive data plotting feature of ICARUS offers a valuable hands-on experience for students that can be compatible with guided discovery models of pedagogy. Based on the experimental conditions and timeline, a student will be able to interact with each data product that had been collected and understand the cause and effects of experimental actions, helping to solidify their learning. Students will be able to download data to practice quantitiative concepts in data analysis, for example, extracting kinetic constants from reactant decay plots or calculating the timescale of particle loss and coagulation.

Regarding the development of new model frameworks in research, one potential research need is a way to parameterize the oxidation chemistry of volatile organic compounds (VOC) in the atmosphere to relevant current and future scenarios. Such scenarios may include (1) increases in temperature due to exacerbation of the climate crisis (with related changes to relative humidity) that impact air pollutant concentrations;[52] (2) transitions to green energy in cities that amplify the air quality importance of VOC emissions from more diverse sources (e.g., a variety of biogenics, volatile chemical products (VCPs), cooking emissions, fire emissions, and other sources) at the expense of emissions from combustion engines;[53] and (3) further reductions in NOx emissions that increase the lifetime and change the fate distributions of the RO_2 radical intermediate in the atmosphere toward autoxidation.[54]

Regarding scenario 1, ICARUS has categorized data explicitly by temperature and relative humidity, which supports community efforts to advance the understanding of atmos-

pheric aerosol and gas-phase processes in a warmer atmosphere. Regarding scenario 2, the integration of the ever-expanding PubChem registry of chemicals and substances in ICARUS supports the reporting of data from new chamber experiments using highly diverse VOCs (such as those in the VCP family of compounds), VOC mixtures (such as fire emissions), and other chemicals of emerging importance. Regarding scenario 3, current model mechanisms do not represent autoxidation well due to NOx-dependent parameterizations [which generally refer to dependences of nitric oxide (NO) instead of the total NOx; Figure 6A]; these were
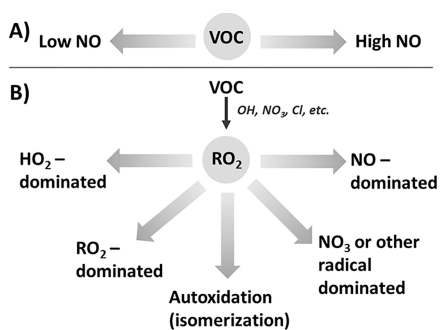

Figure 6. Simplified diagram of model parameterizations of the dependence of atmospheric oxidation chemistry on nitric oxide emissions: (A) Dichotomy of low and high NO regimes; each one may incorporate more than one distinctive reactive pathways for the volatile organic compound (VOC) precursor; and (B) more explicit representation of fates of the alkylperoxy (RO_2) radical intermediate from VOC oxidation with branching toward more diverse chemical regimes that exist in the current and future atmosphere.

developed because other RO_2 fates were not yet recognized to be important historically.[55] The lower availability of data and parameterizations for autoxidation and $RO_2 + RO_2$ reactions have also precluded the widespread inclusion of these processes in simplified SOA models; however, more RO_2-focused data are becoming available from chamber experiments to support mechanism revisions.[56−59] It has been shown that understanding the RO_2 reactivity is important for accurately representing the formation of secondary organic aerosol, organic nitrogen species, and other products in some chemical systems.[60,61] ICARUS enables the data reporting and targeted discovery of data based on RO_2 fate, which supports development of alternative model frameworks that can capture more diverse reactions pathways in the current and future atmosphere (Figure 6B). These examples represent areas in which ICARUS can grow with the needs of the community.

Data mining and model development activities related to machine learning[62−64] will benefit from the increasing availability of high-quality data as provided by ICARUS. In particular, machine learning algorithms gain more information each time there is a new "phase change" or time discontinuity with distinct characteristics that can be observed to effect a measurable system impact.[65] These phase changes occur quite often in each experiment, for example, each time lights are turned on or off, a reactant introduced, or there is a temperature or humidity change that will lead to a measurable impact in the data observations. Thus, datasets from chamber experiments may be excellent tools on which to train machine learning frameworks to simulate complex processes such as secondary organic aerosol formation, gas phase oxidation, or

heterogeneous chemistry, as a complement to traditional modeling tools.

5. DATA AVAILABILITY

Data described in this work are freely available on the ICARUS website. The reuse of ICARUS shared data in any publication requires notification of the PI and other terms as specified by the Data Use Policy (Section S2). Disseminated reuse of ICARUS data require citation to the data using digital object identifiers (DOIs) of the experiments and to this publication. Data DOIs are assigned by DataCite (https://datacite.org).[66] All Experiment DOIs associated with a single Experiment Set are shown on the webpage of the Set ID for ease of tracking. The repository DOI (http://doi.org/10.17616/R31NJN8W) is assigned through the re3data registry (https://re3data.org).[67]

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsearthspace-chem.3c00043.

> ICARUS user guide; ICARUS data policy; example of experiment timeline; example of the experiment view page on the ICARUS website; and example of the ICARUS data format (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Tran B. Nguyen − *Department of Environmental Toxicology, University of California Davis, Davis, California 95616, United States;* ⓞ orcid.org/0000-0001-9206-4359; Email: tbn@ucdavis.edu

Authors

Kelvin H. Bates − *Department of Environmental Toxicology, University of California Davis, Davis, California 95616, United States; Center for the Environment, Harvard University, Cambridge, Massachusetts 02138, United States;* ⓞ orcid.org/0000-0001-7544-9580

Reina S. Buenconsejo − *Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, United States*

Sophia M. Charan − *Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, United States;* ⓞ orcid.org/0000-0002-2023-6403

Eric E. Cavanna − *Information and Educational Technology, University of California Davis, Davis, California 95616, United States*

David R. Cocker, III − *Department Chemical and Environmental Engineering, University of California Riverside, Riverside, California 92507, United States;* ⓞ orcid.org/0000-0002-0586-0769

Douglas A. Day − *Cooperative Institute for Research in Environmental Sciences and Department of Chemistry, University of Colorado Boulder, Boulder, Colorado 80309, United States;* ⓞ orcid.org/0000-0003-3213-4233

Marla P. DeVault − *Cooperative Institute for Research in Environmental Sciences and Department of Chemistry, University of Colorado Boulder, Boulder, Colorado 80309, United States;* Present Address: Now at Minnesota

Department of Health, 625 Robert Street N. P.O. Box 64975, St. Paul, Minnesota 55164, United States (M.P.D.)

Neil M. Donahue − *Department of Chemistry, Department of Chemical Engineering, and Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States;* orcid.org/0000-0003-3054-2364

Zachary Finewax − *Cooperative Institute for Research in Environmental Sciences and Department of Chemistry, University of Colorado Boulder, Boulder, Colorado 80309, United States;* Present Address: Now at NOAA Chemical Sciences Laboratory, 325 Broadway R/CSL, Boulder, Colorado 80305, United States (Z.F.)

Luke F. Habib − *Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States*

Anne V. Handschy − *Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, Colorado 80309, United States*

Lea Hildebrandt Ruiz − *McKetta Department of Chemical Engineering, The University of Texas at Austin, Austin, Texas 78712, United States;* orcid.org/0000-0001-8378-1882

Chung-Yi S. Hou − *Data Stewardship Engineering Team, National Center for Atmospheric Research, Boulder, Colorado 80307, United States;* Present Address: Now at Ronin Institute, 127 Haddon Pl, Montclair, New Jersey 07043, United States (C.-Y.S.H.)

Jose L. Jimenez − *Cooperative Institute for Research in Environmental Sciences and Department of Chemistry, University of Colorado Boulder, Boulder, Colorado 80309, United States;* orcid.org/0000-0001-6203-1847

Taekyu Joo − *School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, Georgia 30332, United States;* Present Address: Now at Yale University, P.O. Box 208236, New Haven, Connecticut 06520, United States (T.J.);* orcid.org/0000-0002-8252-4232

Alexandra L. Klodt − *Department of Chemistry, University of California Irvine, Irvine, California 92697, United States;* orcid.org/0000-0002-3558-972X

Weimeng Kong − *Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, United States;* Present Address: Now at the California Air Resources Board, 1001 I Street, Sacramento, California 95814, United States (W.K.)

Chen Le − *Department Chemical and Environmental Engineering, University of California Riverside, Riverside, California 92507, United States*

Catherine G. Masoud − *McKetta Department of Chemical Engineering, The University of Texas at Austin, Austin, Texas 78712, United States*

Matthew S. Mayernik − *Data Stewardship Engineering Team, National Center for Atmospheric Research, Boulder, Colorado 80307, United States*

Nga L. Ng − *School of Earth and Atmospheric Sciences, School of Chemical and Biomolecular Engineering, and School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, United States;* orcid.org/0000-0001-8460-4765

Eric J. Nienhouse − *Data Stewardship Engineering Team, National Center for Atmospheric Research, Boulder, Colorado 80307, United States*

Sergey A. Nizkorodov − *Department of Chemistry, University of California Irvine, Irvine, California 92697, United States;* orcid.org/0000-0003-0891-0052

John J. Orlando − *Atmospheric Chemistry Observations and Modeling, National Center for Atmospheric Research, Boulder, Colorado 80305, United States*

Jeroen J. Post − *Information and Educational Technology, University of California Davis, Davis, California 95616, United States*

Patrick O. Sturm − *Air Quality Research Center, University of California Davis, Davis, California 95616, United States;* Present Address: Now at University of Southern California, 3470 Trousdale Parkway, Los Angeles, California 90089, United States (P.O.S.);* orcid.org/0000-0001-7361-9590

Bridget L. Thrasher − *Data Stewardship Engineering Team, National Center for Atmospheric Research, Boulder, Colorado 80307, United States;* orcid.org/0000-0002-3961-1971

Geoffrey S. Tyndall − *Atmospheric Chemistry Observations and Modeling, National Center for Atmospheric Research, Boulder, Colorado 80305, United States*

John H. Seinfeld − *Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, United States; Division of Engineering and Applied Science, Calif. Institute of Technology, Pasadena, California 91125, United States;* orcid.org/0000-0003-1344-4068

Steven J. Worley − *Data Stewardship Engineering Team, National Center for Atmospheric Research, Boulder, Colorado 80307, United States*

Xuan Zhang − *Atmospheric Chemistry Observations and Modeling, National Center for Atmospheric Research, Boulder, Colorado 80305, United States;* Present Address: Now at UC Merced, 5200 Lake Road, Merced, California 95343-5001, United States (X.Z.)

Paul J. Ziemann − *Cooperative Institute for Research in Environmental Sciences and Department of Chemistry, University of Colorado Boulder, Boulder, Colorado 80309, United States;* orcid.org/0000-0001-7419-0044

Complete contact information is available at:
https://pubs.acs.org/10.1021/acsearthspacechem.3c00043

Author Contributions

T.B.N. conceived the project, directed the project, supervised project efforts, and wrote the manuscript. K.H.B. supervised the technical validation. E.E.C. and P.O.S. built the initial version of the repository. J.J.P. built the final version of the repository. C.-Y.S.H., M.S.M., E.J.N., B.L.T., and S.J.W. advised on data management decisions. E.J.N. and J.J.P. developed GDEX mirroring and citation mechanisms. D.R.C., N.M.D., L.H.R., J.L.J., N.L.N., S.A.N., J.J.O., G.S.T., J.H.S., P.J.Z., and T.B.N. obtained the funding and casted final votes on project decisions as members of the Scientific Steering Committee. All authors contributed to project activities, decision making, written protocols, shared tools, technical assessments, and the submission of data and metadata. All authors read and revised the manuscript.

Notes

The authors declare no competing financial interest.

◆◆Now an independent consultant.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Schwantes, R. H.; McVay, R. C.; Zhang, X.; Coggon, M. M.; Lignell, H.; Flagan, R. C.; Wennberg, P. O.; Seinfeld, J. H., Science of the environmental chamber. In *Advances in atmospheric chemistry*; World Scientific: 2017; 1−93.

(2) Abbatt, J.; George, C.; Melamed, M.; Monks, P.; Pandis, S.; Rudich, Y. New Directions: Fundamentals of atmospheric chemistry: Keeping a three-legged stool balanced. *Atmos. Environ.* **2014**, *84*, 390−391.

(3) Haagen-Smit, A. J. Chemistry and Physiology of Los Angeles Smog. *Ind. Eng. Chem.* **1952**, *44*, 1342−1346.

(4) Burkholder, J. B.; Sander, S. P.; Abbatt, J.; Barker, J. R.; Huie, R. E.; Kolb, C. E.; Kurylo, M. J.; Orkin, V. L.; Wilmouth, D. M.; Wine, P. H. *Chemical Kinetics and Photochemical Data for Use in Atmospheric Studies, Evaluation No. 18*; JPL Publication 15−10 2015.

(5) Atkinson, R.; Baulch, D. L.; Cox, R. A.; Crowley, J. N.; Hampson, R. F.; Hynes, R. G.; Jenkin, M. E.; Rossi, M. J.; Troe, J.; Subcommittee, I. Evaluated kinetic and photochemical data for atmospheric chemistry: Volume II - gas phase reactions of organic species. *Atmos. Chem. Phys.* **2006**, *6*, 3625−4055.

(6) Saunders, S. M.; Jenkin, M. E.; Derwent, R. G.; Pilling, M. J. Protocol for the development of the Master Chemical Mechanism, MCM v3 (Part A): tropospheric degradation of non-aromatic volatile organic compounds. *Atmos. Chem. Phys.* **2003**, *3*, 161−180.

(7) Jenkin, M. E.; Saunders, S. M.; Wagner, V.; Pilling, M. J. Protocol for the development of the Master Chemical Mechanism, MCM v3 (Part B): tropospheric degradation of aromatic volatile organic compounds. *Atmos. Chem. Phys.* **2003**, *3*, 181−193.

(8) Herrmann, H.; Tilgner, A.; Barzaghi, P.; Majdik, Z.; Gligorovski, S.; Poulain, L.; Monod, A. Towards a more detailed description of tropospheric aqueous phase organic chemistry: CAPRAM 3.0. *Atmos. Environ.* **2005**, *39*, 4351−4363.

(9) Marcum, C. S.; Donohue, R., *New Guidance to Ensure Federally Funded Research Data Equitably Benefits All of America*; White House Office of Science and Technology Policy: OSTP BLOG; https://www.whitehouse.gov/ostp/news-updates/2022/05/26/new-guidance-to-ensure-federally-funded-research-data-equitably-benefits-all-of-america/ , 2022.

(10) National Research Council, *Open access and the public domain in digital data and information for science: Proceedings of an international symposium*; National Academies Press: 2004.

(11) Alvarez, E. G.; Vazquez, M.; Munoz, A.; Hjorth, J.; Pilling, M.; Saathoff, H.; Brauers, T., The Eurochamp Chamber Experiment Database: Goals And Uses. Present And Future Potential Benefits. In *Simulation and Assessment of Chemical Processes in a Multiphase Environment*; Springer: 2008; 71−82.

(12) Orlando, J. J. In *ACCORD Workshop Summary*; Atmospheric Chemistry Center for Observational Research and Data (ACCORD), Boulder, CO: Boulder, CO, 2015; https://www2.acom.ucar.edu/sites/default/files/accord/ACCORD_Workshop_Summary__Oct.2015_Update.pdf.

(13) Kaduwela, A., Future Atmospheric Chemical Mechanism Development Needs In *Atmospheric Chemical Mechanisms*, Davis, CA, 2014.

(14) Pitts, J. Anthropogenic ozone, acids and mutagens: half a century of pandora's NOx. *Res. Chem. Intermed.* **1993**, *19*, 251−298.

(15) Wilkinson, M. D.; Dumontier, M.; Aalbersberg, I. J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L. B.; Bourne, P. E.; Bouwman, J.; Brookes, A. J.; Clark, T.; Crosas, M.; Dillo, I.; Dumon, O.; Edmunds, S.; Evelo, C. T.; Finkers, R.; Gonzalez-Beltran, A.; Gray, A. J. G.; Groth, P.; Goble, C.; Grethe, J. S.; Heringa, J.; 't Hoen, P. A. C.; Hooft, R.; Kuhn, T.; Kok, R.; Kok, J.; Lusher, S. J.; Martone, M. E.; Mons, A.; Packer, A. L.; Persson, B.; Rocca-Serra, P.; Roos, M.; van Schaik, R.; Sansone, S. A.; Schultes, E.; Sengstag, T.; Slater, T.; Strawn, G.; Swertz, M. A.; Thompson, M.; van der Lei, J.; van Mulligen, E.; Velterop, J.; Waagmeester, A.; Wittenburg, P.; Wolstencroft, K.; Zhao, J.; Mons, B. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **2016**, *3*, 160018.

(16) Stall, S.; Yarmey, L.; Cutcher-Gershenfeld, J.; Hanson, B.; Lehnert, K.; Nosek, B.; Parsons, M.; Robinson, E.; Wyborn, L., *Make scientific data FAIR*; Nature Publishing Group: 2019.

(17) Bell, D.; George, C.; McFiggans, G.; Muñoz, A.; Picquet-Varrault, B.; Saathoff, H.; Wiesen, P., *Deliverable D6. 8: Final scientific highlights from EUROCHAMP-2020*.

(18) Cocker, D. R.; Flagan, R. C.; Seinfeld, J. H. State-of-the-Art Chamber Facility for Studying Atmospheric Aerosol Chemistry. *Environ. Sci. Technol.* **2001**, *35*, 2594−2601.

(19) Wu, S.; Lü, Z.; Hao, J.; Zhao, Z.; Li, J.; Takekawa, H.; Minoura, H.; Yasuda, A. Construction and characterization of an atmospheric simulation smog chamber. *Adv. Atmos. Sci.* **2007**, *24*, 250−258.

(20) Wang, J.; Doussin, J.-F.; Perrier, S.; Perraudin, E.; Katrib, Y.; Pangui, E.; Picquet-Varrault, B. Design of a new multi-phase experimental simulation chamber for atmospheric photosmog, aerosol and cloud chemistry research. *Atmos. Meas. Tech.* **2011**, *4*, 2465−2494.

(21) Roberts, G. C.; Nenes, A. A continuous-flow streamwise thermal-gradient CCN chamber for atmospheric measurements. *Aerosol Sci. Technol.* **2005**, *39*, 206−221.

(22) Huang, Y.; Coggon, M. M.; Zhao, R.; Lignell, H.; Bauer, M. U.; Flagan, R. C.; Seinfeld, J. H. The Caltech Photooxidation Flow Tube reactor: design, fluid dynamics and characterization. *Atmos. Meas. Tech.* **2017**, *10*, 839−867.

(23) Che, D. L.; Smith, J. D.; Leone, S. R.; Ahmed, M.; Wilson, K. R. Quantifying the reactive uptake of OH by organic aerosols in a continuous flow stirred tank reactor. *Phys. Chem. Chem. Phys.* **2009**, *11*, 7885−7895.

(24) Nguyen, T.; Crounse, J.; Schwantes, R.; Teng, A.; Bates, K.; Zhang, X.; St Clair, J.; Brune, W.; Tyndall, G.; Keutsch, F. Overview of the Focused Isoprene eXperiment at the California Institute of Technology (FIXCIT): mechanistic chamber studies on the oxidation of biogenic compounds. *Atmos. Chem. Phys.* **2014**, *14*, 13531−13549.

(25) Carter, W. P.; Cocker, D. R., III; Fitz, D. R.; Malkina, I. L.; Bumiller, K.; Sauer, C. G.; Pisano, J. T.; Bufalino, C.; Song, C. A new environmental chamber for evaluation of gas-phase chemical mechanisms and secondary aerosol formation. *Atmos. Environ.* **2005**, *39*, 7768−7788.

(26) Paulot, F.; Crounse, J. D.; Kjaergaard, H. G.; Kurten, A.; St. Clair, J. M.; Seinfeld, J. H.; Wennberg, P. O. Unexpected epoxide formation in the gas-phase photooxidation of isoprene. *Science* **2009**, *325*, 730−733.

(27) American Chemical Society ACS Research Data Policy. https://publish.acs.org/publish/data_policy.

(28) The Royal Society Open data policy. https://royalsociety.org/journals/ethics-policies/data-sharing-mining/.

(29) Atmospheric Chemistry and Physics Data Policy. https://www.atmospheric-chemistry-and-physics.net/policies/data_policy.html.

(30) Taylor and Francis Earth, space and environmental sciences journals introduce a FAIR data policy. https://authorservices.taylorandfrancis.com/data-sharing-policies/fair-data-copdess/.

(31) Nature Portfolio Reporting standards and availability of data, materials, code and protocols. https://www.nature.com/nature-portfolio/editorial-policies/reporting-standards.

(32) Zaki, Z. M.; Dew, P. M.; Haji, M. H.; Lau, L. M.; Rickard, A.; Young, J. In A user-orientated electronic laboratory notebook for retrieval and extraction of provenance information for EURO-CHAMP-2, 2011 IEEE Seventh International Conference on eScience; IEEE: 2011, 371−378.

(33) Bird, C. L.; Willoughby, C.; Frey, J. G. Laboratory notebooks in the digital era: the role of ELNs in record keeping for chemistry and other sciences. Chem. Soc. Rev. 2013, 42, 8157−8175.

(34) Kim, S.; Thiessen, P. A.; Bolton, E. E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B. A.; Wang, J.; Yu, B.; Zhang, J.; Bryant, S. H. PubChem substance and compound databases. Nucleic Acids Res. 2016, 44, D1202−D1213.

(35) Orlando, J. J.; Tyndall, G. S. Laboratory studies of organic peroxy radical chemistry: an overview with emphasis on recent issues of atmospheric significance. Chem. Soc. Rev. 2012, 41, 6294−6317.

(36) Boyd, C. M.; Sanchez, J.; Xu, L.; Eugene, A. J.; Nah, T.; Tuet, W. Y.; Guzman, M. I.; Ng, N. L. Secondary organic aerosol formation from the β-pinene+ NO 3 system: effect of humidity and peroxy radical fate. Atmos. Chem. Phys. 2015, 15, 7497−7522.

(37) Crounse, J. D.; Nielsen, L. B.; Jørgensen, S.; Kjaergaard, H. G.; Wennberg, P. O. Autoxidation of organic compounds in the atmosphere. J. Phys. Chem. Lett. 2013, 4, 3513−3520.

(38) Jokinen, T.; Sipilä, M.; Richters, S.; Kerminen, V. M.; Paasonen, P.; Stratmann, F.; Worsnop, D.; Kulmala, M.; Ehn, M.; Herrmann, H.; Berndt, T. Rapid Autoxidation Forms Highly Oxidized RO2 Radicals in the Atmosphere. Angew. Chem., Int. Ed. 2014, 53, 14596−14600.

(39) Wennberg, P. O. Let's abandon the "high NOx" and "low NOx" terminology; IGAC News, 2013.

(40) Atkinson, R.; Baulch, D. L.; Cox, R. A.; Crowley, J. N.; Hampson, R. F.; Hynes, R. G.; Jenkin, M. E.; Rossi, M. J.; Troe, J. Evaluated kinetic and photochemical data for atmospheric chemistry: Volume III - gas phase reactions of inorganic halogens. Atmos. Chem. Phys. 2007, 7, 981−1191.

(41) Aknan, A. C.; Chen, G.; Crawford, J.; Williams, E., ICARTT File Format Standards V1.1. NASA, Ed. 2013; https://www-air.larc.nasa.gov/missions/etc/IcarttDataFormat.htm.

(42) Peng, Z.; Lee-Taylor, J.; Orlando, J. J.; Tyndall, G. S.; Jimenez, J. L. Organic peroxy radical chemistry in oxidation flow reactors and environmental chambers and their atmospheric relevance. Atmos. Chem. Phys. 2019, 19, 813−834.

(43) Peng, Z.; Jimenez, J. L. Radical chemistry in oxidation flow reactors for atmospheric chemistry research. Chem. Soc. Rev. 2020, 49, 2570−2616.

(44) McNeill, V. F.; Woo, J. L.; Kim, D. D.; Schwier, A. N.; Wannell, N. J.; Sumner, A. J.; Barakat, J. M. Aqueous-Phase Secondary Organic Aerosol and Organosulfate Formation in Atmospheric Aerosols: A Modeling Study. Environ. Sci. Technol. 2012, 46, 8075−8081.

(45) Wolfe, G. M.; Marvin, M. R.; Roberts, S. J.; Travis, K. R.; Liao, J. The framework for 0-D atmospheric modeling (F0AM) v3. 1. Geosci. Model Dev. 2016, 9, 3309−3319.

(46) NCAR Music-Box Model. https://github.com/NCAR/music-box.

(47) He, Y.; Akherati, A.; Nah, T.; Ng, N. L.; Garofalo, L. A.; Farmer, D. K.; Shiraiwa, M.; Zaveri, R. A.; Cappa, C. D.; Pierce, J. R.; Jathar, S. H. Particle Size Distribution Dynamics Can Help Constrain the Phase State of Secondary Organic Aerosol. Environ. Sci. Technol. 2021, 55, 1466−1476.

(48) Sommariva, R.; Cox, S.; Martin, C.; Borońska, K.; Young, J.; Jimack, P. K.; Pilling, M. J.; Matthaios, V. N.; Nelson, B. S.; Newland, M. J.; Panagi, M.; Bloss, W. J.; Monks, P. S.; Rickard, A. R. AtChem (version 1), an open-source box model for the Master Chemical Mechanism. Geosci. Model Dev. 2020, 13, 169−183.

(49) Peckham, S. E., WRF/Chem version 3.3 user's guide. 2012.

(50) Mouchel-Vallon, C.; Lee-Taylor, J.; Hodzic, A.; Artaxo, P.; Aumont, B.; Camredon, M.; Gurarie, D.; Jimenez, J. L.; Lenschow, D. H.; Martin, S. T.; Nascimento, J.; Orlando, J. J.; Palm, B. B.; Shilling, J. E.; Shrivastava, M.; Madronich, S. Exploration of oxidative chemistry and secondary organic aerosol formation in the Amazon during the wet season: explicit modeling of the Manaus urban plume with GECKO-A. Atmos. Chem. Phys. 2020, 20, 5995−6014.

(51) Knote, C.; Tuccella, P.; Curci, G.; Emmons, L.; Orlando, J. J.; Madronich, S.; Baró, R.; Jiménez-Guerrero, P.; Luecken, D.; Hogrefe, C.; Forkel, R.; Werhahn, J.; Hirtl, M.; Pérez, J. L.; San José, R.; Giordano, L.; Brunner, D.; Yahya, K.; Zhang, Y. Influence of the choice of gas-phase mechanism on predictions of key gaseous pollutants during the AQMEII phase-2 intercomparison. Atmos. Environ. 2015, 115, 553−568.

(52) Nussbaumer, C. M.; Cohen, R. C. The Role of Temperature and NOx in Ozone Trends in the Los Angeles Basin. Environ. Sci. Technol. 2020, 54, 15652−15659.

(53) Zhao, Y.; Li, Y.; Kumar, A.; Ying, Q.; Vandenberghe, F.; Kleeman, M. J. Separately resolving NOx and VOC contributions to ozone formation. Atmos. Environ. 2022, 285, No. 119224.

(54) Praske, E.; Otkjær, R. V.; Crounse, J. D.; Hethcox, J. C.; Stoltz, B. M.; Kjaergaard, H. G.; Wennberg, P. O. Atmospheric autoxidation is increasingly important in urban and suburban North America. Proc. Nat. Acad. Sci. U. S. A. 2018, 115, 64−69.

(55) Porter, W. C.; Jimenez, J. L.; Barsanti, K. C. Quantifying Atmospheric Parameter Ranges for Ambient Secondary Organic Aerosol Formation. ACS Earth Space Chem. 2021, 5, 2380−2397.

(56) Goldman, M. J.; Green, W. H.; Kroll, J. H. Chemistry of Simple Organic Peroxy Radicals under Atmospheric through Combustion Conditions: Role of Temperature, Pressure, and NOx Level. J. Phys. Chem. A 2021, 125, 10303−10314.

(57) Wennberg, P. O.; Bates, K. H.; Crounse, J. D.; Dodson, L. G.; McVay, R. C.; Mertens, L. A.; Nguyen, T. B.; Praske, E.; Schwantes, R. H.; Smarte, M. D.; St Clair, J. M.; Teng, A. P.; Zhang, X.; Seinfeld, J. H. Gas-phase reactions of isoprene and its major oxidation products. Chem. Rev. 2018, 118, 3337−3390.

(58) Day, D. A.; Fry, J. L.; Kang, H. G.; Krechmer, J. E.; Ayres, B. R.; Keehan, N. I.; Thompson, S. L.; Hu, W.; Campuzano-Jost, P.; Schroder, J. C.; Stark, H.; DeVault, M. P.; Ziemann, P. J.; Zarzana, K. J.; Wild, R. J.; Dubè, W. P.; Brown, S. S.; Jimenez, J. L. Secondary Organic Aerosol Mass Yields from NO3 Oxidation of α-Pinene and Δ-Carene: Effect of RO2 Radical Fate. J. Phys. Chem. A 2022, 126, 7309−7330.

(59) Boyd, C. M.; Sanchez, J.; Xu, L.; Eugene, A. J.; Nah, T.; Tuet, W. Y.; Guzman, M. I.; Ng, N. L. Secondary organic aerosol formation from the β-pinene+NO₃ system: effect of humidity and peroxy radical fate. Atmos. Chem. Phys. 2015, 15, 7497−7522.

(60) Peng, W.; Le, C.; Porter, W. C.; Cocker, D. R. Variability in Aromatic Aerosol Yields under Very Low NOx Conditions at Different HO2/RO2 Regimes. Environ. Sci. Technol. 2022, 56, 750−760.

(61) Bates, K. H.; Burke, G. J. P.; Cope, J. D.; Nguyen, T. B. Secondary organic aerosol and organic nitrogen yields from the nitrate radical (NO3) oxidation of alpha-pinene from various RO2 fates. Atmos. Chem. Phys. 2022, 22, 1467−1482.

(62) Sturm, P. O.; Manders, A.; Janssen, R.; Segers, A.; Wexler, A. S.; Lin, H. X. Advecting Superspecies: Efficiently Modeling Transport of Organic Aerosol With a Mass-Conserving Dimensionality Reduction Method. J. Adv. Model. Earth Syst. 2023, 15, No. e2022MS003235.

(63) Kelp, M. M.; Jacob, D. J.; Kutz, J. N.; Marshall, J. D.; Tessum, C. W.; Stable, T. General Machine-Learned Models of the Atmospheric Chemical System. J. Geophys. Res. 2020, 125, No. e2020JD032759.

(64) Schreck, J. S.; Becker, C.; Gagne, D. J.; Lawrence, K.; Wang, S.; Mouchel-Vallon, C.; Choi, J.; Hodzic, A. Neural Network Emulation of the Formation of Organic Aerosols Based on the Explicit GECKO-A Chemistry Model. J. Adv. Model. Earth Syst. 2022, 14, No. e2021MS002974.

(65) Awad, M.; Khanna, R.; Awad, M.; Khanna, R. Machine learning in action: Examples. *Efficient learning machines: Theories, concepts, and applications for engineers and system designers*; Apress, 2015, 209 −240.

(66) Brase, J. In DataCite-A global registration agency for research data, *2009 fourth international conference on cooperation and promotion of information resources in science and technology*; IEEE: 2009, 257−261.

(67) Pampel, H.; Vierkant, P.; Scholze, F.; Bertelmann, R.; Kindling, M.; Klump, J.; Goebelbecker, H.-J.; Gundlach, J.; Schirmbacher, P.; Dierolf, U. Making research data repositories visible: the re3data. org registry. *PLoS One* **2013**, *8*, No. e78080.