TRANSIENCE

TRANSITIONING TOWARDS AN EFFICIENT, CARBON-NEUTRAL CIRCULAR EUROPEAN INDUSTRY

Date: 26/2024

D3.1 – Machine-actionable, open data management plan

WP3 – Characterising circularity and decarbonisation opportunities – generating model inputs

Disclaimer

Copyright Message

Grant Agreement Number	101137606	**Acronym**	TRANSIENCE	
Full Title	TRANSItioning towards an Efficient, carbon-Neutral Circular European industry			
Topic	HORIZON-CL4-2023-TWIN-TRANSITION-01-36			
Funding scheme	HORIZON EUROPE, RIA – Research and Innovation Action			
Start Date	January 2024	**Duration**	48 Months	
Project URL	https://www.transience.eu/			
EU Project Advisor	Eskarne ARREGUI PABOLLET			
Project Coordinator	Institute of Communication and Computer Systems (ICCS)			
Deliverable	D3.1 - Machine-actionable, open data management plan			
Work Package	WP3 - Characterising circularity and decarbonisation opportunities – generating model inputs			
Date of Delivery	**Contractual**	30/04/2024	**Actual**	26/04/2024
Nature	Report	**Dissemination Level**	Public	
Lead Beneficiary	Institute of Communication and Computer Systems (ICCS)			
Responsible Author	Natasha Frilingou	**Email**	nfrilingou@epu.ntua.gr	
	ICCS	**Phone**	+30 210 772 3612	
Contributors	Georgios Xexakis [HOLISTIC]; Christina Tigka; Panagiotis Kokkinakos, Alexandros Nikas [ICCS]			
Reviewer(s)	Maro Baka [E3M]; Thurid Meta Lotz [Fh ISI]; Jakob Duerrwaechter [PIK]; Gergo Suto [UU]			
Keywords	Data management plan; FAIR; machine-actionable; data handling			

EC Summary Requirements

1. Changes with respect to the DoA

No changes with respect to the work described in the DoA.

2. Dissemination and uptake

This report shall serve as a guide for all consortium partners on how to handle project datasets. It can also be used by external users to understand how data in the TRANSIENCE project is collected, processed, and disseminated.

3. Short summary of results (<250 words)

This report documents the first version of the Open Data Management Plan (DMP) of TRANSIENCE. The DMP is designed to evolve alongside the project, allowing for information to be refined and updated as the implementation progresses. It currently provides a description of the data that will be used and created during the project, information on how to make data findable, accessible, interoperable, and permit the widest re-use possible (FAIR), along with details on resource allocation, data security, and ethical aspects. Besides this report, TRANSIENCE will be using the ARGOS service of OpenAIRE and EUDAT to maintain a machine-actionable data management plan (maDMP). The maDMP will be continuously updated with links and metadata for datasets generated from project activities. This DMP report will also be updated during Phase 2 (D7.1) and Phase 3 (D11.2) of the project.

4. Evidence of accomplishment

This report and the machine-actionable DMP in ARGOS (link).

Preface

The need to approach climate action, resource efficiency, and circularity performance as integrated, economy-wide, cross-cutting issues is growingly gaining attention in the policy world, stimulating the development of new industrial policies in Europe and worldwide. Currently, however, there is little progress in conceptualising the circular economy and understanding its interactions with climate action. State-of-the-art modelling capacity to capture the interplay of the two agendas and their implications for energy-intensive sectors as well as to represent the European industry's transformation in line with the region's vision for climate neutrality is not yet fully developed. TRANSIENCE will undertake a comprehensive characterisation and assessment of circularity principles and measures vis-à-vis decarbonisation, by looking at the twin transition of European industries through the lenses of global competitiveness, innovation, and holistic sustainability. It will then produce MIC3, a consistent, fully open-source model ecosystem to assess industrial circularity, decarbonisation, and sustainability. A series of interoperable modules on the socioeconomic, service and product, material, industrial, energy-system, and environmental perspectives of the transformation of European industry will be developed and integrated, building on and opening the code of leading modelling tools. MIC3 will finally be used in extensive scenario modelling to produce diverse pathways toward a material-efficient, circular, climate-neutral, sustainable European industry. Transparency, openness, and knowledge sharing will be promoted, and technical capacities will be developed in four industrial agglomerations in the EU, moving beyond stakeholder consultation, onto model co-development, continuous validation of assumptions, co-creation of scenario modelling, evaluation of the desirability and usability of the developed model and insights, and eventually co-production of science and action.

ICCS – Institute of Communication and Computer Systems	EL
CEPS – Centre for European Policy Studies	BE
E3M – E3-Modelling AE	EL
Fraunhofer – Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e.V.	DE
HOL – HOLISTIC IKE	EL
PIK – Potsdam Institut für Klimafolgenforschung e.V.	DE
PNTEC – Park Naukowo-Technologiczny Euro-Centrum Spolka Z Ograniczona Odpowiedzialnoscia	PL
TECNALIA – Fundacion Tecnalia Research & Innovation	ES
UU – Universiteit Utrecht	NL
WI – Wuppertal Institut für Klima, Umwelt, Energie gGmbH	DE
PSI – Paul Scherrer Institut	CH
UCL – University College London	UK

Funded by the European Union

Executive Summary

The extensive model development and scenario analysis activities that will take place in TRANSIENCE in the next four years, will require swathes of data as inputs and will result in large amounts of data as outputs. This report documents the Open Data Management Plan of the project, providing a summary of all data used and developed within the project. The plan also includes information on how to make data findable, accessible, interoperable, and permit the widest re-use possible (FAIR), and details on resource allocation, data security, and ethical aspects. The DMP report will be updated during Phase 2 (D7.1) and Phase 3 (D11.2) of TRANSIENCE.

In parallel, TRANSIENCE will be using the ARGOS service of OpenAIRE and EUDAT to deliver a machine-actionable version of the data management plan (maDMP). The maDMP will include an inventory with metadata for project datasets that will be generated from project activities. As with the DMP report, the maDMP will be constantly updated throughout the project, allowing for information to be refined and updated when significant changes occur.

The report begins with the working definitions of data, DMP, and machine-actionability used by TRANSIENCE (Section 1), along with a short introduction to the ARGOS service. Then, Section 2 describes the purpose of the data collection and/or generation and its relation to the objectives of the project as well as data formats, use, origin, and utility. Section 3 provides a comprehensive strategy to ensure that this data is Findable, Accessible, Interoperable, and Reusable (FAIR) and Section 4 describes the allocation of required resources to enable FAIR data. Section 5 addresses data security, recovery, and storage while Section 6 discusses ethical issues on data handling practises. The final section of this report (Section 7) describes the ARGOS service and how new datasets will be added in the maDMP and provides an outlook for its continuous update.

Contents

Table of Figures

Table of Tables

Funded by the European Union

PROCESSES4PLANET

1 Introduction

This report is the first version of the open data management plan (DMP) of the Horizon Europe TRANSIENCE project. It describes how research data are handled and establishes guidelines and protocols governing the collection, generation, processing, and dissemination of data throughout the project's duration. It also provides a transparent strategy to govern open data management, thoroughly outlining actions for provision of FAIR data (Findable, Accessible, Interoperable, and Reusable), thus ensuring effective knowledge and data exchange.

A DMP is a structured document outlining the procedures for managing research data. It includes the necessary steps, measures, and strategies to manage their entire life cycle (Gajbe et al., 2021) throughout and beyond the duration of a research project, and describes how to maintain the quality, safety, sustainability, and, where feasible, accessibility and reusability of the data.

Apart from this report, we have also developed a machine-actionable Data Management Plan (maDMP). Machine-actionability in data management refers to the capability of DMPs to be understood, processed, and executed by computational systems without human intervention (GO FAIR, 2024). Research Data Alliance (RDA), an initiative by the European Commission, the United States Government's National Science Foundation and National Institute of Standards and Technology, and the Australian Government's Department of Innovation, focuses on the importance of machine-actionability in DMPs with the goal of "automatic exchange, integration, and validation of information provided in DMPs and facilitating the exchange of information between systems acting on behalf of stakeholders involved in the research life cycle" (Miksa et al., 2021). Machine-actionability thus enables DMPs to become "active" documents, and be updated, when necessary, as the project progresses.

We developed our maDMP using the ARGOS service of OpenAIRE and EUDAT. ARGOS is an online tool that simplifies the management, validation, and monitoring of a maDMP throughout its lifecycle. First, users initiate DMPs and input descriptions for their datasets. By default, both DMPs and datasets are created privately, but can be shared with colleagues to aid in the drafting process. Datasets are categorised within DMPs based on type and/or scientific discipline, ensuring well-organized information and easy identification of relevant data activities within each DMP record. DMPs and their associated datasets remain in draft status upon creation, allowing for updates and versioning within the ARGOS environment. Upon completion, users can validate and finalise their DMPs, and change their visibility from private to open, making the DMP accessible to all via Public Dashboards. To complete the DMP lifecycle, ARGOS integrates Zenodo, enabling users to publish their DMPs directly from the platform (OpenAIRE, 2024).

2 Data summary

2.1 Project objectives and implications for data collection and generation

TRANSIENCE will undertake a comprehensive characterisation and assessment of circularity principles and measures vis-à-vis decarbonisation, by looking at the twin transition of European industries through the lenses of global competitiveness, innovation, and holistic sustainability. Towards this goal, it will develop an open-source, user friendly, and validated modelling framework that meets these requirements and can address the questions arising in the next phase of Europe's industry transformation.

The proposed framework will be the Model for European Industry Circularity and Climate Change mitigation (MIC3) and will allow the enhanced representation and full integration of six key dimensions of the transition to a climate-neutral, sustainable, and circular industry in Europe: the socioeconomic, the product and end-use/service, the material, the industrial, the energy system, and the environmental dimensions.

To capture these dimensions, a series of interoperable modules will also be developed and integrated, focusing on energy intensive industries and process industries, by combining different modelling paradigms, including system dynamics, industrial metabolism, agent-based, bottom-up techno-economic, and macro-economic modelling, while building on and opening the code of leading modelling tools. The project will generate an open database of circular economy measures, costs, and potentials (MS10) which will be publicly shared and used for modelling activities within TRANSIENCE. Similarly, we will develop a service and product database of key components for the carbon neutrality of European economy and society, containing product and service characteristics and specifications as well as material composition and supply chain-related information (D4.2).

Our flagship output, MIC3, the 'satellite' modules comprising it, and the scenario exercises stemming from it will be co-developed and validated with relevant stakeholders from industry, policy, and civil society, as well as used to inform assessments and transition strategies, at EU and country level, at global level, and within four heterogeneous regional industry clusters in Europe, to ensure their usability and exploitation in real-world use cases.

To support the development of MIC3, open-source and FAIR data principles and established practices from the broader software development community will be followed. We will create guidelines and tools for collaborative model development and APIs (Application Programming Interfaces) to enable communication among satellite models. The APIs will be based on clear, robust rules agreed with partners, enabling the required exchanges for MIC3 operation, extensively documented in a report (D7.2) and an online service (Swagger), and explicitly designed to ensure modularity beyond the project context, allowing third party models to connect with modules (in tandem and individually). Executable containers of the modules will be provided to enable exploitation of the APIs by others. Automated error checking routines to ensure quality of data transfers between modelling tools will be built, to ensure quality of data transfers between modelling tools, create protocols to guide modelling exercises and activities of the project based on input from partners and external stakeholders, and program validation routines for evaluating the feasibility of produced scenarios against a variety of indicators (MS16). Validation diagnostics will be developed to enable consistency and robustness of results for all modules individually as well as for MIC3.

The entire model code will be open source and model development will follow an open development strategy (D3.2). We will also define diagnostics and open science protocols to ensure smooth module integration and create a clear documentation of all variables, parameters, concepts, and assumptions used in the models (D3.8). We will provide guidelines for opening our work via GitHub and Docker using established open-source practices from the wider software development community (D7.2). Apart from open practices, we will secure models' accessibility by stakeholders, providing extensive documentation and APIs to online, running versions of all modules, and promote the simplified version of MIC3 among industry actors, along with tutorials and a helpdesk service during the project and for two years afterwards (D12.4)

All input data of TRANSIENCE models, and outputs of scenario analysis, will be hosted in the data sharing platform I²AM PARIS[1], along with extensive documentation of the satellite modules and MIC3, links to model code in GitHub, and user-tailored libraries for each stakeholder group (industry, policymakers, researchers) to interactively visualise our model results. Apart from hosting detailed documentation for models and interactive visualisations to describe the results of modelling exercises, TRANSIENCE will also include fit-for-purpose information and input data, user guides and training kits, as well as explanatory components to discuss results in non-expert language, highlighting key takeaways.

Our maDMP will provide rich metadata for datasets along with their links in Zenodo and I²AM PARIS, further increasing their findability and transparency, which in turn could enable results verification, synergies between research projects, leveraging on existing work, and identification of errors or biases (Alvarez-Romero et al., 2022), and thus contribute on TRANSIENCE goal towards understanding the pathways of European industries to decarbonisation and increased circularity performance.

2.2 Types and formats of data

To achieve the projects objectives outlined in Section 2.1 requires the generation and collection of various types of data. Table 1 shows a non-exhaustive list of the data types and formats that will be used for different project activities.

Data used for modelling inputs will be mostly quantitative and based on datasets in spreadsheet format (e.g., xlsx or csv) or prebuilt data (e.g., Rda format used for the R language). New model code will be in the format of the programming languages used to develop these models. For instance, EDM-S of the WISEE-EDM suite, an MFA model with high level of detail that traces materials from manufacturing to use, waste collection and sorting, and recycling, is programmed in Python and, thus, its code will be provided in the Python file format.

Qualitative insights from existing body of knowledge and from stakeholder co-creation activities such as interviews and surveys will also be used to inform modelling. Examples of such qualitative insights could be potential barriers in the development of technologies used for industrial circularity or decarbonisation based on interviews with experts. Our shared database of policy measures and technology costs for mitigation, circularity action, and sustainability will allow qualitative analysis to proceed systematically and explore transformation opportunities, challenges, and appropriate policy packages or technologies in sectors with significant potential. The database will be provided in csv format in Zenodo and as an online interactive application. The latter will allow for different searches and filtering of outputs and potential translation into either exogenous or endogenous variables to increase flexibility of adaptation to generate

[1] https://www.i2am-paris.eu/

Funded by the European Union
PROCESSES4PLANET

input that is appropriate for the diversity of our models.

Our scenario results will also be quantitative and exported in a spreadsheet format or other formats that are easily converted to spreadsheets, and available on I²AM PARIS. These output datasets will inform the scientific and policy publications of the project which will be provided in pdf formats.

Table 1: Data types and formats per project activity

Project activity	Data collected and generated	Data formats
Develop the new modules and models	**Collect**: Code from previous versions of models, feedback from project stakeholders and communities of practice for the different models **Generate**: New model code and documentation	File formats related to the programming language of each model (e.g., py, R, h, cpp, GAMS), reports documenting the model and stakeholder feedback in pdf
Compile the input data for the new models	**Collect**: National and sectoral statistics on socioeconomic parameters, macroeconomic indicators, energy use and supply, industrial indicators, physical demand for products, policy options; technological cost and innovation data; investment costs; global material demands, supply, and circularity; global material demands, supply, and circularity; and other modelling inputs **Generate**: Curate the raw data into a format and data structure that is usable by project models	Spreadsheet-like formats, e.g., xlsx and csv, as well as other formats such as txt, json, sqlite databases etc.
Develop two open databases, one on circular economy measures and another on products and services needed for climate neutrality	**Collect**: Information about circular economy measures from policy and scientific documents, including their costs and potentials/effectiveness in terms of achieving circularity and climate neutrality; characteristics of products and services that are used for the transition towards climate neutrality, e.g., wind turbines, PV panels, EVs, etc. **Generate**: Curate the raw data into a consistent format and data structure that is usable by the project's models and can be easily understood by external users. Both databases will be accessible through online interactive interfaces in I²AM PARIS.	xlsx, csv, txt, json
Scenario analysis using the new models	**Collect**: Scenario definitions based on the stakeholder engagement activities of the project **Generate**: GHG emissions, industry relocation, material imports, circularity, sufficiency, future energy demand, fuel mix, jobs creation, resilience to shocks,	Reports documenting results and stakeholder feedback in pdf; results in xlsx, csv, txt, json. Scenario results will be also uploaded to the I²AM

Funded by the European Union
PROCESSES4PLANET

	costs, and other social, economic, and environmental indicators	PARIS platform as interactive tables.

2.2.1 Data processing tools

The main data processing tools used in the project will be the satellite modules comprising MIC3 (Table 2), and MIC3 itself. To develop a hybrid model ecosystem that helps promote the strengths and overcome the weaknesses of each of these modules, efforts will be put on creating interfaces among them to enable firm links in terms of data flows, time, structural differences, and solutions. Dedicated data exchange interfaces will be developed to couple the satellite modules, so that further features and higher fidelity modelling can be achieved, while maintaining their modularity and standalone operational capacity. This entails creating tools for processing the outputs and state-of-the-art visualisation that draws from the best available options, such as those featured by the JRC's Energy and Industry Geography Lab, a tool which maps energy, industrial and other relevant infrastructure required for a climate-neural EU industry (JRC, 2023).

Table 2: Satellite modules of MIC3

Satellite modules	Scope of modules	Responsible partner(s)
OPEN-GEM-E3[2]	computable general equilibrium model	E3M
OPEN-PROM[3]	energy system model	E3M
FORECAST-sites based on FORECAST[4]	industrial model	Fh ISI
WISEE-EDM[5]	industrial model and material flow model	WI
REMIND-MFA[6]	material flow model	PIK
MFA module based on FORECAST-materials and WISEE-EDM	material flow model	WI, Fh ISI
LCA module	life-cycle analysis model	PSI

In addition to models, other data processing tools will be utilised in the project, including (but not limited to) Microsoft Excel, for reading and editing xlsx and csv files, Microsoft Word for writing reports and manuscripts based on project outcomes, and Adobe Reader/Acrobat for generating publications in pdf format. Additional tools and scripting languages, such as Python (e.g., in WISEE EDM-S) and R, are also expected to be used for data processing and curation purposes.

The integration of different sources of input and output data in the MIC3 framework poses challenges to data consistency. In this context, methods to ensure interoperability, smooth integration, and quality of model inputs and outputs, despite heterogeneity of data sources across modules will include dedicated quality checks for all modules. All tasks share a common milestone to define interfaces for data exchange

[2] https://e3modelling.com/modelling-tools/gem-e3/
[3] https://e3modelling.com/modelling-tools/prometheus/
[4] https://doi.org/10.1016/j.esr.2018.09.005
[5] https://www.i2am-paris.eu/detailed_model_doc/wisee-edm
[6] https://rse.pik-potsdam.de/doc/remind/3.2.0/

Funded by the European Union
PROCESSES4PLANET

among satellite modules (MS5), and one (MS8) on internal data exchange and validation. Modelling teams within the consortium will serve as pilot users to identify any shortcomings and devise effective quality enhancements for the final release of the MIC3 framework, and validation feedback from stakeholders will be incorporated. Typical quality checks in this respect will include inspections of individual time series and indicators for outliers and anomalies, pre-feasibility assessments of each module separately, evaluation of the modules' accuracy based on their ability to recreate historical developments and contrasting internal consistency across scenarios. Therefore, the project will invest in developing additional tools to facilitate compatibility and consistency and ensure that key indicators are within reasonable ranges (Guivarch et al., 2022) via interfaces for automated vetting in the I^2AM PARIS platform, thus securing that the individual modules produced as part of the fully integrated MIC3 model can be used to directly feed into international scientific assessments, notably those of the IPCC AR7 cycle.

2.2.2 Data inputs and outputs for models

In terms of data inputs, we will use fully open datasets when possible or at least provide flexible options for defining inputs, allowing users to choose between open or proprietary datasets for simulations. Data outputs will be fully open access based on the FAIR data guidelines provided in this report and its updates (Section 3). These guidelines will be also added in the maDMP to facilitate the uptake and reuse of all outputs via automatic parsing of metadata. Insights from scenario modelling will be publicly available via Zenodo in the form of scientific papers for academics, guides for industrial actors, and policy briefs for policy and other stakeholders (D9.3, D12.3). Datasets will be presented via interactive visualisations in I^2AM PARIS, offering customised interfaces for all audiences and key policy- and industry-relevant insights/clarifications.

All publications and datasets will be open access via a Zenodo community while all project code will be open source via GitHub. All papers and relevant datasets will be shared at the latest upon publication, using Creative Commons licenses for reusability and DOIs for findability. Table 3 shows an indicative list of data inputs and outputs for models.

Table 3: Indicative input and output data for project models (Giarola et al., 2021; Pauliuk et al., 2017)

Data type	Category	Input/output
Population	Socioeconomic	Input
GDP	Socioeconomic	Input & output
Interest rates and exchange rates	Socioeconomic	Input
Labour participation and productivity	Socioeconomic	Input
Household income	Socioeconomic	Input & output
Macroeconomic and sectoral activity data	Socioeconomic	Input & output
CO_2 emissions	Emissions	Input
CH_4 emissions	Emissions	Input & output
N_2O emissions	Emissions	Input & output
F-gases	Emissions	Input & output
Other pollutants	Emissions	Input & output
Fuel efficiency	Energy	Input
Industrial energy demand	Energy	Input & output
Sectoral energy intensities	Energy	Output
Energy/electricity trade	Energy	Output

Vintage tracking	Industry / Energy	Input
Production volumes	Industry	Input
Material stocks and flows	Industry	Input & output
Recycling potential	Industry	Input
Waste generation and use	Industry	Input
Waste flows	Industry	Input & output
Fossil fuel prices	Prices/costs	Input & output
Technology costs (capital, operating and maintenance)	Prices/costs	Input & output

2.2.3 Data used in scenario analysis

In addition to the data inputs outlined in the previous section, TRANSIENCE modules will require inputs specific to exploratory scenarios, to ensure that MIC3 is able to respond to real-world questions supporting policy and industry decision-makers with relevant information. The desired functionality of MIC3 will be specified in close coordination with relevant stakeholders and in respect to the identified needs, developed typologies, interfaces, and data flows. We will organise four workshops, one in each cluster region, to synthesise in four cognitive maps of systemic interdependencies and feedback loops inhibiting/promoting industrial transition to decarbonisation and circularity. These maps will serve as a canvas for the open development strategy, enabling to identify entry points for MIC3 modules and the joint application of their integration to offer analysis relevant to identified challenges and provide a starting point for identifying key policy and industrial questions. We will also organise a workshop to evaluate overlaps & divergences of modelled pathways with the ones envisioned by stakeholders and identify assumptions for new pathways to reduce these divergences.

A survey (D11.5) will reach an even wider pool in Europe; survey participants will view results tailored to their context (e.g., region, sector), before assessing whether these results are informative, comprehensible, and can be used directly for their policies, industrial strategies, or research. The findings will guide the final stages of pathway modelling. Additionally, all survey and workshop participants will provide their views on the design of the simplified mode, such as on features and functionalities they would like to see on the tool, format of the results, questions they would like to answer using this tool, etc.

Most of these inputs will be produced during the engagement and outreach activities of the project and will relate to pertinent policy and research questions that stakeholders want to explore with the new models. An example of policies that could be included in our scenario analysis is shown in Table 4.

Table 4: Examples of policy categories and objectives used in scenario analysis (Böck et al., 2020).

Policy category	Policy objective
Climate action	Reduced GHG emissions
	Improved carbon pricing
	Carbon sequestration
Economy & Society	Green growth
	Reduced poverty and inequality
	Unemployment reduction
Energy Efficiency	Nearly Zero Energy Buildings
	Electrification

Resources & Materials	Treatment of landfill gas
	Reduced use of fertilisers
	Increased use of alternative cement
Energy generation, storage & transmission	Energy system flexibility
	Fossil fuel phase-out
	Fuel shift: Green H_2 production & use

Other inputs for explorative scenario modelling will be based on assumptions on the variables mentioned in Table 3, such as whether some technologies are available or not and various assumptions on their costs and other characteristics, and on harmonisation parameters shown in Table 5.

More details on potential scenario inputs will be provided in the Open Model Development Strategy (D3.2) and the Open database of policies & technologies (D3.5).

2.3 Origin, expected size, and utility of data

2.3.1 Origin

As discussed in the previous sections, the development of the new models will be based on the existing code of their base models while their inputs will be based on existing data sources. Strategies to achieve fully open-access data inputs will be drawn out during the project in a case-by-case basis and will aim to substitute proprietary datasets with similar open-source data. In case this is not possible, another potential solution will be to convert input data in a prebuilt, binary format which can be only read by the models and can be potentially shared openly. An indicative list of the origin of harmonisation data that may be required is adapted by the Broad Scenario Logic[7] of the IAM COMPACT project and shown in Table 5.

Table 5: Indicative harmonisation data, their origin, and format

Topic	Variable	Context	Indicative data origin	Data format
Socioeconomics	**Population**	EU27+Norway	Population data from Eurostat: Historical statistics through 2018 (Eurostat, 2023); EUROPOP2019 projections from 2019 through 2100 (Eurostat, 2022a, 2022b)	csv
		OECD+7 except EU27/Norway	Population data from OECD Economic Outlook 109 Long-term baseline projections (available for 199 through 2060), to ensure consistency with GDP projections (OECD, 2021). Extend with growth rates from UN WPP2022 after 2060 if necessary (see below).	csv
		Rest of World	UN World Population Prospects 2022 figures for all years (available from 1960 through 2100, with historical data until 2022) (United Nations, 2022)	csv
	GDP	All regions, through 2028	GDP from IMF World Economic Outlook (IMF, 2023) in constant 2017	csv

[7] https://iam-compact.eu/sites/default/files/2023-05/D4.3_Broad%20Scenario%20Logic_v1.00_SUBMITTED.pdf

Funded by the European Union

			international dollars unless otherwise indicated by the research question	
		EU27+Norway, from 2029	Extend IMF forecast with linearly interpolated real GDP growth rates from the 2021 Ageing Report (European Commission, 2021b, 2021a)	csv
		OECD+ except EU27/Norway, from 2029	Extend IMF forecast with real GDP growth rates from the OECD Economic Outlook 109 long-term baseline	csv
		Rest of World	Extend IMF forecast with GDP growth rates per working age capita from SSP2 (must be calculated using the projected population figures) multiplied by population growth rate from the harmonised population time series	csv
Techno-economics	Technology costs (for reference scenarios)	EU27 and comparable countries	The EU Reference Scenario 2020 (European Commission, 2021; European Commission et al., 2021).	csv
		Other regions, power-sector technologies	Most suitable option are cost assumptions from IEA's World Energy Outlook 2022 (International Energy Agency, 2022).	csv
		Other regions, non-power technologies	Options include adapting costs from the EUR Reference Scenario 2020.	csv
	Fossil fuel prices	Historical prices	Regional prices from IEA datasets (International Energy Agency, 2023a). If participating modellers do not have access to proprietary IEA data, data for some years and regions can be extracted from the freely available World Energy Outlook 2022 report (International Energy Agency, 2022), or if global benchmarks are sufficient, some of these are available for free from the World Bank "pink sheet" (World Bank, 2023)	csv
		Price projections	Regional price forecasts from the World Energy Outlook 2022 extended dataset (International Energy Agency, 2023), requires subscription. Participating modellers who do not have a license can extract some prices visually from charts in the World Energy Outlook 2022 report (International Energy Agency, 2022). For long-term EU-specific exercises that do not need short-term trends, price projections from the EU Reference Scenario 2020 may be used (European Commission, 2021).	csv
Energy	Energy production and consumption	Historical data	IEA World Energy Balances (International Energy Agency, 2023c)	csv
Emissions	Energy-related CO_2, CH_4, N_2O	Historical data	IEA Greenhouse Gas Emissions from Energy dataset (International Energy	csv

			Agency, 2023b), if modellers have or can acquire access. Alternatively use EDGAR v7.0 (Branco et al., 2022) which is consistent with IEA but with less detailed breakdowns. Emissions can also be calculated from energy consumption data using default Tier 1 emission factors from the 2006 IPCC guidelines for GHG inventories, which are consistent with IEA emission factors.	
	IPPU CO_2 from cement	Historical data	Production data and emission factors from (R. Andrew, 2018). Updated data available on Zenodo (R. Andrew, 2023).	csv
	F-gases and non-energy CO_2, CH_4, N_2O	Historical data	EDGAR v7.0 (Branco et al., 2022)	csv
	Other emissions	Historical data	CEDS (O'Rourke et al., 2021)	csv
	Land-use change emissions	Historical data	Check that used or generated data falls within or close to range spanned by 3 bookkeeping models in the Global Carbon Budget 2022 (Friedlingstein et al., 2022). Use an average of the three models if a single harmonised dataset is required and no other constraints are implied by the research question.	csv
	All emissions	Infilling of historical and future emissions	Use Silicone software package to infill missing emission components, if required by the modelling exercise (e.g., for climate impact assessment) (Lamboll et al., 2020, 2022b, 2022a)	csv

2.3.2 Size

Based on previous modelling-based projects such as PARIS REINFORCE[8], the total size of data that is expected to be collected, processed, and produced will be around 500GB. Most of this size is expected to result from the data inputs that will be used in the new models, and, especially, from the sheer amounts of data outputs that will be produced. Other types of data that can be relatively heavy include the audio and video recordings of consortium and stakeholder meetings, which should be at a scale of hundreds of MB for each meeting, multiplied by a few dozens of meetings that will be organised. The minutes of these meetings will be documented in a small number of reports which will not be more than a few MB each. Model publications are also expected to not take too much space, considering that they will be around 70 pdf documents. More accurate estimations will be available during the project and the next revisions of the DMP.

2.3.3 Utility

The data outputs that will be produced by TRANSIENCE are expected to be useful to all expected audiences

[8] https://paris-reinforce.eu/

Funded by the European Union
PROCESSES4PLANET

of the project. The new models will be directly useful to climate-economy modellers, circular economy modellers, and other researchers of relevant topics while they will be also indirectly useful to policymakers, industry, and civil society representatives. The results of the scenario analysis will be useful to the same target audiences and inform their policies, strategies, research, and other activities. We will especially focus our efforts to disseminate project data to the European Commission and EU agencies, national/local governments, businesses, industrial clusters, energy-intensive industries, financial institutions, and researchers of the broader climate, energy, and circularity modelling landscape.

Funded by
the European Union

PROCESSES4PLANET

3 FAIR data

As the significance of open data continues to grow, a consortium comprising stakeholders from academia, industry, and government collaborated to formulate the FAIR Principles as a benchmark for evaluating the extent to which scientific data exhibits characteristics of being Findable, Accessible, Interoperable, and Reusable (Wilkinson et al., 2016). Embracing the FAIR principles empowers researchers to leverage and expand upon existing knowledge, fostering novel discoveries and technological advancements. Instead of prescribing specific technical requirements, these principles offer a flexible framework that supports a spectrum of enhanced reusability across diverse implementations. The core guidelines for assessing the FAIRness of research data are described in Box 1.

Box 1: FAIR principles for research data

1. **Findability:** Findability enables easy location and access of data by interested parties. To achieve findability, data must be allocated a distinct and enduring identifier known as a Persistent Identifier (PID). PIDs provide unique and long-lasting references to digital objects and serve as standardised references linked to the data, even if its location or access method changes over time. PIDs are a crucial component of findability, enabling anyone possessing the identifier to discover and access the data irrespective of its storage location. While PIDs are essential for ensuring findability, comprehensive machine-readable metadata are also important for the automated discovery of relevant datasets and services. They must clearly and explicitly include the identifier of the data they describe and be registered or indexed in a searchable resource, thus forming a vital aspect of the FAIRification process (Jacobsen et al., 2020).

2. **Accessibility:** Accessibility ensures that data are readily available and can be accessed and used by both humans and machines. The (meta)data should be accessible, even when the data are no longer available. They should also be retrievable by their identifier using a standardised communications protocol, which is open, free, universally implementable, and allows for an authentication and authorisation procedure, where necessary. Accessible does not necessarily mean open, as accessibility stands for “accessible under well-defined conditions”, which includes shielding data for personal privacy or national security reasons and assuring the proper data protection (Mons et al., 2017).

3. **Interoperability:** Ensuring interoperability is vital for seamlessly integrating and analysing scientific data across various systems and tools. Achieving this requires structuring scientific data in open and standardised formats that are easily understandable and usable by different software and applications. This approach facilitates data sharing and reuse across diverse disciplines and domains, promoting collaboration and interdisciplinary research efforts. Utilising standardized formats and protocols enables the integration of data into analysis workflows, enabling researchers to extract insights and knowledge from varied datasets (Ravi et al., 2022). This is accomplished through the exchange of data and metadata among different software packages via application programming interfaces (APIs), adhering to relevant community standards and incorporating references to other objects. As a result, metadata should use vocabularies that follow FAIR principles and include qualified references to other (meta)data (Calamai & Frontini, 2018). When effectively utilised, metadata can enable research data to function as "mobile" objects (Latour, 1987) indicating their ability to transition between diverse production contexts while maintaining significant evidential value (Pasquetto et al., 2019).

4. **Re-usability:** Reusable data should be structured and documented in a manner that facilitates its effective reuse for different purposes. This includes providing detailed descriptions of the data's

provenance, quality, and usage rights, as well as adhering to clear and standardized data formats and structures. Enhancing reusability optimises the value and impact of scientific data, given they are thoroughly documented, structured, and annotated with metadata in a way that enhances comprehensibility and usability (da Silva Santos et al., 2023). This involves providing details about the origin, quality, and accessibility of data as well as adopting transparent and standardised data structures and formats of widespread recognition and acceptance. By ensuring data's reusability over time, researchers can leverage existing knowledge, generate results more quickly, and establish connections among researchers and scientific fields.

In the context of TRANSIENCE, we will implement open science principles and establish an open pipeline for model development, opening the 'black box' of scientific assumptions, processes, and results, providing full access to the new modules and model produced (including code, interfaces, and data). Data used and produced will be FAIR, allowing to build and sustain a vibrant community of practice on industry-academia collaboration for knowledge valorisation, yielding benefits for both creators and users (Mons et al., 2017), and to document new modelling capacity for expert and non-expert audiences. The FAIRness of research data involves evaluating various aspects of the data's characteristics and infrastructure.

3.1 Making Data Findable

We will make all project outputs findable through PIDs, adequate metadata, and keywords to facilitate document retrieval via search engines. All deliverables, policy briefs, and business guides will be uploaded in the project's Zenodo community[9] and receive digital object identifiers (DOIs). In the case of scientific publications, a DOI will be provided by the publishing journal, although we will still upload publications, accepted manuscripts, or preprints on Zenodo to keep a full archive of our work there. We will also use the versioning system of Zenodo to keep track of the different versions of the documents that we upload; each version will get a separate DOI, but a top-level DOI will be also available, resolving to the latest version. Additionally, we define a consistent naming system for each type of output:

- Scientific publications: "{author(s)}_{year}" (e.g., Smith_et_al_2023.pdf)
- Policy/Business briefs: "TRANSIENCE_{title}" (e.g., TRANSIENCE_Policy_Brief_on_CE_Principles.pdf)
- Deliverable: "TRANSIENCE_DX.X_{title}" (e.g., TRANSIENCE_D3.1_Open_Data_Management_Plan.pdf)
- Datasets: "{author(s)}_{year}_{dataset_name}" (e.g., Smith_et_al_2024_Cost_Assumptions.csv)

All datasets developed during the project will also be archived in Zenodo and fitted with a DOI. After uploading a dataset in Zenodo we will also link it to the maDMP in ARGOS and describe it with rich metadata and keywords, using the Horizon Europe template provided in ARGOS (see Chapter 7). Similarly, all final model code will be publicly stored in GitHub which will be then linked to Zenodo and the maDMP.

All model documentation, inputs, and outputs will be also published in the I^{2}AM PARIS modelling platform to further promote them among the climate-economy modelling community that have been using the platform since its development in 2020. For each dataset, we will include links to all related project publications to further increase findability. Finally, both the platform and the project website will include

[9] https://zenodo.org/communities/transience

Funded by the European Union
PROCESSES4PLANET

adequate content and an optimised sitemap to ensure findability in search engines like Google and Bing.

3.2 Making Data Accessible

We will use Creative Commons licenses for all deliverables, policy briefs, business guides, and datasets produced in TRANSIENCE. Most project outputs will be published using the highly permissive CC BY license (version 4.0). Exceptionally, we may also consider less permissive licenses such as CC BY-SA to ensure that derivative works will be made available with the same open license. For instance, this will be useful for model code developed in TRANSIENCE to ensure that any derivative software will stay open and free to support industrial decarbonisation in Europe.

Scientific publications will be also published in journals offering open-access options in compliance with the Horizon Europe rules. When possible, we will publish in fully open-access journals, also considering the Open Research Europe publishing platform. In case that the available fully open-access options do not align with the scope of a publication, we will select a journal that offers a gold open-access option from the list of journals that the organisations of project partners have a publishing agreement with.

All modules that will be developed in the project, along with the integrated MIC3 framework, will be published under open-source licenses. For each module, we will explore different license options along with the respective model development team. The exact licenses will be defined by the end of Reporting Period 1, where module development will be already at a mature stage. It is noted that we may use different licenses for model code and input data. Even in case that some of the input data in a module is not open access, we will suggest alternative datasets that can be used in their place.

In terms of file formats, we will strive to use well-known formats that can be opened by freeware software such as pdf, csv, txt, mp3, and mp4 files. When this is not feasible (for instance, when we need to publish an Excel spreadsheet file featuring multiple tabs), we will accompany files in proprietary formats with links to compatible freeware software such as the Open Office suite.

We will also ensure that all open project outputs will remain available for as long as possible. All project publications and datasets will be published in established online repositories such as Zenodo and GitHub where high availability is expected for many years to come. The project website will be also kept online for at least three years after the project's end to support the dissemination and findability of project outcomes. HOLISTIC will also ensure the longevity of the I^{2}AM PARIS platform for at least four years after the project ends (till around 2032).

3.3 Making Data Interoperable

We will achieve high interoperability of project data by using adequate data formats and providing informative metadata. As suggested in Section 3.2, all project datasets and reports of the project will be shared through widespread formats such as pdf, csv, and txt, avoiding proprietary formats when possible. The documentation of all new modules along with the results of the scenario analysis will be formatted based on the IPCC AR6/AR7 reporting templates, ensuring that the wider climate-economy modelling community can use them, while also achieving interoperability with other relevant software of the community such as the pyam Python package. However, we will also explore other formats used in the field of industrial ecology, to ensure that our results would be readily usable by circular economy researchers. Metadata for all project datasets will be added in TRANSIENCE's maDMP in ARGOS, using the format template of Horizon Europe. As all information in ARGOS is machine actionable, all metadata can be potentially converted to another format template, further ensuring the interoperability of project datasets.

3.4 Making Data Reusable

As suggested in the previous section on open access, by releasing all deliverables and datasets through CC BY license we will support their uptake by a wide range interested parties. Similarly, all scientific papers will be published under open-access licenses and will be made available to the research community directly after acceptance by the journals. We will also ensure the reusability of datasets by using the Horizon Europe metadata scheme in the maDMP of the project. For each dataset, the scheme will provide a short description of the data, links with publications and other datasets, and guidelines for the specific dataset related to FAIR practices, allocation of resources, and security and ethical aspects. This scheme will be thus similar to the format of this report, although it will provide more specific information for each dataset. Lastly, all modelling documentation and results will be formatted using the reporting templates of IPCC AR6 and, potentially, relevant templates from industrial ecology research, ensuring reusability by the wider modelling community of climate change mitigation and circular economy.

Funded by
the European Union

4 Allocation of resources

Most of the FAIR practices described in Section 3 do not require any costs from the project. All deliverables and datasets will be uploaded in Zenodo which is free to use, and we will also use the free version of GitHub to store project code. Similarly, the documentation of datasets in the maDMP in ARGOS is also free of charge.

Additional activities for the implementation of our open data management plan and require resources have been considered in the project's Grant Agreement. The extension and hosting of I^2AM PARIS requires funds that have been budgeted under WP3, while funds on the development and maintenance of the project website have been considered in the budget of WP1 and the defined purchase costs for ICCS. We have also earmarked a part of the budget for publishing in open access journals. Most of this budget is managed by the project coordinator ICCS while all partners have some funds available for individual open access publications related to their work in the project. Suggested options for publishing in open access journals are shown in Table 6.

Access type	Funder	Fees	License
Publish in a fully Open Access journal	TRANSIENCE Grant	Article Processing Charges; ranging between ~200€ (e.g., Elsevier Societal Impacts[10]) to over 10,000€ (e.g., Nature[11])	CC BY 4.0 CC BY-NC-ND 4.0
Publish in a journal that has the option of Gold Open Access	Publishing agreements between the organisations of project partners and the publisher		

On data curation, storage, and preservation, the project coordinator and quality manager will be responsible for data management and quality control. ICCS and HOLISTIC will jointly take up associated costs and put together (and frequently update) the present DMP report.

All project partners will be responsible for correct data handling and curation based on the guidelines of the DMP, including that model code is frequently uploaded in the TRANSIENCE community in GitHub. As also mentioned above, HOLISTIC will be responsible for keeping the website and the platform online and all related project data available for at least three years after the end of the project. For the platform, both ICCS and HOLISTIC are exploring ways to further extend its lifetime.

[10] https://www.elsevier.com/about/policies-and-standards/pricing
[11] https://www.nature.com/nature/for-authors/publishing-options

Funded by the European Union

5 Data security

Data assets collected, processed, or stored during the research project have been identified in Section 2, are classified based on sensitivity and importance, as they require different levels of protection. All measures to ensure the security of all project data are taken through robust data storage and secure platforms for communication and data exchange.

ICCS has established a dedicated workspace within its enterprise version of Microsoft Teams to facilitate internal communication, including video calls and chats among project partners, as outlined in Milestone 3. Access to this platform is restricted to authorised users, namely consortium members, ensuring confidentiality. Additionally, this system is seamlessly integrated with a secure instance of Microsoft SharePoint, serving as the exclusive data exchange platform for the project. The management and security of these systems are overseen by ICCS administrators and the Data Protection Officer (DPO), with servers located within the EU (Greece), ensuring compliance with GDPR and relevant EU regulations. This adherence to GDPR standards is particularly vital as SharePoint will also store contact details of project stakeholders, necessitating robust security measures. Likewise, personal data of newsletter subscribers will be stored in HOLISTIC's MailerLite account, which is GDPR-compliant.

Apart from the SharePoint, other data storage systems used in the project include the databases of the project website and the I^{2}AM PARIS platform. For both databases, HOLISTIC and ICCS have implemented disaster recovery and backup policies to ensure that the data is safe from loss caused by a disaster such as a critical systems failure, fire, theft, or natural disaster. A similar process is followed by ICCS for the SharePoint system while there is also a versioning system in place that protects the users from accidentally deleting or modifying data.

The project's communities in Zenodo and GitHub will be also used to store data during the process, and, most importantly, to preserve all created datasets and publications after the end of the project. For each dataset, only the minimum amount of data necessary for achieving the intended purpose of the dataset will be released. The likelihood of experiencing data loss within these repositories is minimal, given that all files and documents are stored across multiple online servers to guarantee redundancy. Furthermore, the prospect of these repositories ceasing operations is highly unlikely. However, in such an unlikely event, they have contingency plans in place to migrate all content to appropriate archives, such as the servers managed by the Software Heritage Foundation and Internet Archive.

6 Ethical aspects

All data collection and management activities within the project will adhere to the EU GDPR regulation and the national privacy and data protection laws of each partner country. This entails implementing appropriate measures to protect the privacy and confidentiality of personal data collected or processed during the research, obtaining consent for data processing where required, and ensuring that data is stored and handled securely. For most activities related to model development, ethical and legal considerations are minimal, primarily focused on honouring the licenses of databases serving as sources for modelling inputs. Conversely, ethical considerations take precedence in all co-creation activities of the project (WP2 – Phase 1), where diverse perspectives of project stakeholders will be gathered through workshops, interviews, and surveys.

During workshops, the Chatham House Rule (Chatham House, 2024) will be applied, ensuring anonymity of speakers and participants. Workshop minutes will strictly adhere to this rule, avoiding any linkage between individuals and specific statements. Similar procedures will be followed for documenting interviews, while surveys will only inquire about personal details necessary for research purposes, such as academic or policymaking affiliation. Explicit and clear informed consent will be sought from all participants in engagement activities, with details on ethical aspects included in WP2 deliverables.

As outlined in Section 5, all contact details and feedback from project stakeholders will be securely stored within the project's SharePoint instance, while newsletter subscriber contact details will be kept in HOLISTIC's MailerLite account. Both platforms are hosted within the EU and are GDPR compliant. Measures will be taken to prevent bulk exchange of contact details through insecure channels like emails. No transfer of personal data will occur between EU and non-EU countries. The sole information collected will be lists of research questions voiced by project stakeholders, maintaining anonymity to ensure anonymisation.

Finally, TRANSIENCE will consider the broader societal implications of our research and strive to maximise its positive impact on society. This involves engaging with stakeholders to ensure that research outcomes are socially beneficial and address pressing societal challenges.

Funded by the European Union

PROCESSES4PLANET

7 maDMP in ARGOS

DMPs serve as awareness tools, aiding researchers in effectively handling their data, and guaranteeing its quality, accessibility, and reusability beyond the project's completion. A DMP outlines digital research methods that will evolve throughout a project's duration. Hence, to remain a valuable resource for researchers and other stakeholders, the content of the DMP must be regularly updated to reflect the methods employed and the data generated. While maintaining a human-readable narrative remains essential, there is growing recognition of the potential for enhancing the DMP with thematic, machine-actionable details to provide added value for all stakeholders (Michener, 2015).

For this reason, the TRANSIENCE maDMP in ARGOS[12] will provide a detailed overview of all datasets generated, curated, or managed during the project. The maDMP will be updated whenever a project dataset is created or modified during the project and a summary of its contents will be provided in the updates of this report (D7.1, D11.2). Personnel from ICCS and HOLISTIC will be responsible for creating entries for new datasets in the maDMP while all project partners will be responsible for checking that the metadata provided for their datasets in the maDMP is correct. The full DMP lifecycle in ARGOS is shown in Figure 1.

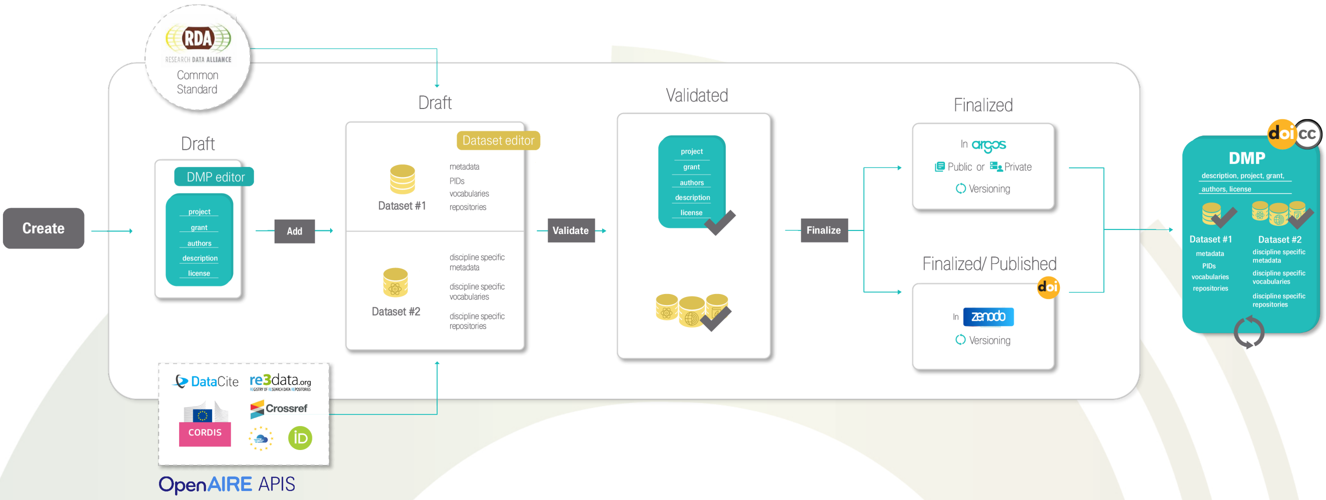

Figure 1: DMP publication lifecycle (OpenAIRE, 2024)

The following process will be used for creating a new dataset in maDMP of TRANSCIENCE:

- The project partner(s) that created the dataset (henceforth called "data creators") will share it with ICCS and HOLISTIC ("data managers").
- The data managers will then upload the dataset on Zenodo.
- The data managers will also create a new dataset in the maDMP and prefill it by searching the name of the dataset through the search engine of ARGOS.
- The data managers will complete the metadata for the dataset based on guidance from this report and its updates and will also link the dataset to relevant deliverables or scientific publications.
- The data creators will be invited to check these metadata and ensure their accuracy.

[12] https://argos.openaire.eu/plans/overview/c556435c-d68d-4db8-901d-1f257d396f0e

Funded by the European Union

- The data managers will update the maDMP with the new dataset.

A similar process will be used for new model code, where links will be created between the model repositories in GitHub and Zenodo to ensure that the code base of the models receives a unique DOI. This Zenodo entry will be then linked to the maDMP which will be documented with adequate metadata as above. All these processes will be evaluated based on the experience of data managers and creators and may be adapted and optimised further during the duration of the project.

Bibliography

Alvarez-Romero, C., Martínez-García, A., Sinaci, A. A., Gencturk, M., Méndez, E., Hernández-Pérez, T., Liperoti, R., Angioletti, C., Löbe, M., Ganapathy, N., Deserno, T. M., Almada, M., Costa, E., Chronaki, C., Cangioli, G., Cornet, R., Poblador-Plou, B., Carmona-Pírez, J., Gimeno-Miguel, A., … Parra Calderón, C. L. (2022). FAIR4Health: Findable, Accessible, Interoperable and Reusable data to foster Health Research. *Open Research Europe*, *2*. https://doi.org/10.12688/OPENRESEUROPE.14349.2

Andrew, R. (2023). *Global CO2 emissions from cement production (Version 230428) [Data set]*. https://doi.org/10.5281/ZENODO.7875557

Andrew, R. (2018). Global CO2 emissions from cement production. *Earth System Science Data*, *10*(1), 195–217. https://doi.org/10.5194/ESSD-10-195-2018

Böck, E., Eggler, L., Rohrer, M., Papagianni, S., Christodoulaki, R., & Taxeri, E. (2020). *D8.1 Review of Policy options to drive societies towards sustainability*. https://www.locomotion-h2020.eu/resources/main-project-reports/?cp=2

Bose, C. (2012). Principles of management and administration. In *PHI Learning Pvt. Ltd.* https://books.google.gr/books?id=AoFGD39Uqr4C&dq=Management+can+be+defined+as+the+process+of+planning,+organising,+directing,+and+controlling+resources+to+achieve+specific+objectives+efficiently+and+effectively.+&lr=&source=gbs_navlinks_s

Branco, A., Crippa, M., Guizzardi, D., Banja, M., Solazzo, E., Muntean, M., Schaaf, E., Pagani, F., Monforti-Ferrario, F., Olivier, J. G. J., Quadrelli, R., Grassi, G., Rossi, S., Oom, D., San-Miguel, J., & Vignati, E. (2022). Emissions Database for Global Atmospheric Research (v7.0_FT_2021) [Data set]. *European Commission, Joint Research Centre (JRC)*.

Calamai, S., & Frontini, F. (2018). FAIR data principles and their application to speech and oral archives. *Journal of New Music Research*, *47*(4), 339–354. https://doi.org/10.1080/09298215.2018.1473449

Chatham House. (2024). *Chatham House Rule*. https://www.chathamhouse.org/about-us/chatham-house-rule

da Silva Santos, L. O. B., Burger, K., Kaliyaperumal, R., & Wilkinson, M. D. (2023). FAIR Data Point: A FAIR-Oriented Approach for Metadata Publication. *Data Intelligence*, *5*(1), 163–183. https://doi.org/10.1162/DINT_A_00160

European Commission. (2021). *EU Reference Scenario 2020*. https://energy.ec.europa.eu/data-and-analysis/energy-modelling/eu-reference-scenario-2020_en

European Commission. (2021). *EU reference scenario 2020 : energy, transport and GHG emissions : trends to 2050.* https://doi.org/10.2833/35750

European Commission. (2021a). *Ageing Report 2021 Data [Data set]*. https://data.europa.eu/data/datasets/ageing-report-2018?locale=en

European Commission. (2021b). *The 2021 ageing report: Economic & budgetary projections for the EU Member States (2019 2070)*. https://op.europa.eu/en/publication-detail/-/publication/8b1015a6-ead6-11eb-93a8-01aa75ed71a1/language-en

Eurostat. (2023). *Demography, population stock and balance (demo) [Data set]*. https://ec.europa.eu/eurostat/databrowser/explore/all/popul?lang=en&subtheme=demo

Eurostat. (2022a). *EUROPOP2019 - Population projections at national level (2019-2100) (proj_19n; Short-term update 2022-09-28) [Data set]*. https://ec.europa.eu/eurostat/databrowser/explore/all/popul?lang=en&subtheme=proj.proj_19n

Eurostat. (2022b). *EUROPOP2019—Population projections at regional level (2019-2100) (update 2022-09-28)*. .

https://ec.europa.eu/eurostat/databrowser/explore/all/popul?lang=en&subtheme=proj.proj_19r

Friedlingstein, P., O'sullivan, M., Jones, M. W., Andrew, R. M., Gregor, L., Hauck, J., Le Quéré, C., Luijkx, I. T., Olsen, A., Peters, G. P., Peters, W., Pongratz, J., Schwingshackl, C., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Alkama, R., … Zheng, B. (2022). Global Carbon Project. (2022). Supplemental data of Global Carbon Budget 2022 (Version 1.0) [Data set]. *Earth System Science Data*, *14*(11), 4811–4900. https://doi.org/10.5194/ESSD-14-4811-2022

Gajbe, S. B., Tiwari, A., Gopalji, & Singh, R. K. (2021). Evaluation and analysis of Data Management Plan tools: A parametric approach. *Information Processing & Management*, *58*(3), 102480. https://doi.org/10.1016/J.IPM.2020.102480

Giarola, S., Mittal, S., Vielle, M., Perdana, S., Campagnolo, L., Delpiazzo, E., Bui, H., Kraavi, A. A., Kolpakov, A., Sognnaes, I., Peters, G., Hawkes, A., Köberle, A. C., Grant, N., Gambhir, A., Nikas, A., Doukas, H., Moreno, J., & van de Ven, D. J. (2021). Challenges in the harmonisation of global integrated assessment models: A comprehensive methodology to reduce model response heterogeneity. *Science of The Total Environment*, *783*, 146861. https://doi.org/10.1016/J.SCITOTENV.2021.146861

GO FAIR. (2024). *FAIR Principles*. https://www.go-fair.org/fair-principles/

Guivarch, C., Kriegler, E., Portugal-Pereira, J., Bosetti, V., Edmonds, J., Fischedick, M., Havlík, P., Jaramillo, P., Krey, V., Lecocq, F., Lucena, A., Meinshausen, M., Mirasgedis, S., O'Neill, B., Peters, G. P., Rogelj, J., Rose, S., Saheb, Y., Strbac, G., … Zhou, N. (2022). Annex III: Scenarios and modelling methods. *IPCC, 2022: Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, 1841–1908. https://doi.org/10.1017/9781009157926.022

IMF. (2023). *World Economic Outlook Database*. https://www.imf.org/en/Publications/WEO/weo-database/2023/October

International Energy Agency. (2023). *World Energy Outlook 2022 Extended Dataset [Data set]*. https://www.iea.org/data-and-statistics/data-product/world-energy-outlook-2022-extended-dataset

International Energy Agency. (2022). *World Energy Outlook 2022*. https://www.iea.org/reports/world-energy-outlook-2022

International Energy Agency. (2023a). *Energy Prices [Data set]*. https://www.iea.org/data-and-statistics/data-product/energy-prices

International Energy Agency. (2023b). *Greenhouse Gas Emissions from Energy [Data set]*. https://www.iea.org/data-and-statistics/data-product/greenhouse-gas-emissions-from-energy

International Energy Agency. (2023c). *World Energy Balances [Data set]*. https://www.iea.org/data-and-statistics/data-product/world-energy-balances

Jacobsen, A., Azevedo, R. de M., Juty, N., Batista, D., Coles, S., Cornet, R., Courtot, M., Crosas, M., Dumontier, M., Evelo, C. T., Goble, C., Guizzardi, G., Hansen, K. K., Hasnain, A., Hettne, K., Heringa, J., Hooft, R. W. W., Imming, M., Jeffery, K. G., … Schultes, E. (2020). FAIR Principles: Interpretations and Implementation Considerations. *Data Intelligence*, *2*(1–2), 10–29. https://doi.org/10.1162/DINT_R_00024

JRC. (2023). *Energy and Industry Geography Lab*. https://energy-industry-geolab.jrc.ec.europa.eu/

Lamboll, R. D., Nicholls, Z., & Kikstra, J. (2022a). *Silicone documentation*. Readthedocs.Io. . https://silicone.readthedocs.io/en/latest/search.html

Lamboll, R. D., Nicholls, Z., & Kikstra, J. (2022b). *Silicone GitHub repository*. https://github.com/GranthamImperial/silicone

Lamboll, R. D., Nicholls, Z. R. J., Kikstra, J. S., Meinshausen, M., & Rogelj, J. (2020). Silicone v1.0.0: An open-source Python package for inferring missing emissions data for climate change research. *Geoscientific*

Model Development, *13*(11), 5259–5275. https://doi.org/10.5194/GMD-13-5259-2020

Latour, B. (1987). *Science in Action: How to follow scientists and engineers through society*. Cambridge, MA: Harvard University Press. https://www.hup.harvard.edu/books/9780674792913

Merriam-Webster. (2024). *Data Definition & Meaning*. https://www.merriam-webster.com/dictionary/data

Michener, W. K. (2015). Ten Simple Rules for Creating a Good Data Management Plan. *PLOS Computational Biology*, *11*(10), e1004525. https://doi.org/10.1371/JOURNAL.PCBI.1004525

Miksa, T., Walk, P., Neish, P., Oblasser, S., Murray, H., Renner, T., Jacquemot-Perbal, M. C., Cardoso, J., Kvamme, T., Praetzellis, M., Suchánek, M., Hooft, R., Faure, B., Moa, H., Hasan, A., & Jones, S. (2021). Application profile for machine-actionable data management plans. *Data Science Journal*, *20*(1). https://doi.org/10.5334/DSJ-2021-032

Mons, B., Neylon, C., Velterop, J., Dumontier, M., Da Silva Santos, L. O. B., & Wilkinson, M. D. (2017). Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. *Information Services & Use*, *37*(1), 49–56. https://doi.org/10.3233/ISU-170824

OECD. (2021). *Economic Outlook No 109—October 2021 - Long-term baseline projections (EO109_LTB) [Data set].* . https://stats.oecd.org/Index.aspx?DataSetCode=EO109_LTB

OpenAIRE. (2024). *ARGOS - a collaborative workspace for delivering DMPs*. https://www.openaire.eu/argos-guide

O'Rourke, P. R., Smith, S. J., Mott, A., Ahsan, H., McDuffie, E. E., Crippa, M., Klimont, Z., McDonald, B., Wang, S., Nicholson, M. B., Feng, L., & Hoesly, R. M. (2021). *CEDS GitHub repository. Joint Global Change Research Institute*. https://github.com/JGCRI/CEDS

Pasquetto, I. V., Borgman, C. L., & Wofford, M. F. (2019). Uses and Reuses of Scientific Data: The Data Creators' Advantage. *Harvard Data Science Review*, *1*(2), 2019. https://doi.org/10.1162/99608F92.FC14BF2D

Pauliuk, S., Arvesen, A., Stadler, K., & Hertwich, E. G. (2017). Industrial ecology in integrated assessment models. *Nature Climate Change 2017 7:1*, *7*(1), 13–20. https://doi.org/10.1038/nclimate3148

Ravi, N., Chaturvedi, P., Huerta, E. A., Liu, Z., Chard, R., Scourtas, A., Schmidt, K. J., Chard, K., Blaiszik, B., & Foster, I. (2022). FAIR principles for AI models with a practical application for accelerated high energy diffraction microscopy. *Scientific Data 2022 9:1*, *9*(1), 1–9. https://doi.org/10.1038/s41597-022-01712-9

United Nations. (2022). *World Population Prospects 2022 [Data set]*. https://population.un.org/wpp/

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. 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., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data 2016 3:1*, *3*(1), 1–9. https://doi.org/10.1038/sdata.2016.18

World Bank. (2023). *Commodity Markets*. https://www.worldbank.org/en/research/commodity-markets

Funded by
the European Union