



Net0Pathways through impRoved IAMs across Sectors, diMensions, and scAles

Deliverable 1.2: A detailed nomenclature

*Streamlining comparison of IAM scenario
assumptions and reference data*

Deliverable
July 28, 2023



"Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Climate, Infrastructure and Environment Executive Agency (CINEA). Neither the European Union nor the granting authority can be held responsible for them."

This project has received funding from the European Union's Horizon Europe programme under grant agreement No 101081604.

Deliverable Title	Detailed nomenclature (list of variables including definitions and recommended units) for comparison of IAM scenario assumptions and reference data
Deliverable number	D1.2
Brief Description	Processes and tools to establish a standardized reporting template, integrating work from previous model-comparison projects
WP number	WP7
Lead Beneficiary	IIASA
Author(s)	Daniel Huppmann
Deliverable Due Date	30/06/2023
Actual Delivery Date	28/07/2023
Nature of the Deliverable	R – Document, report
Dissemination Level	PU - Public
KeyWords	Nomenclature, Region, Variable, Codelists, Definitions

Date	Version	Contributors	Pages
28/07/2023	01	Daniel Huppmann	



1 Background

Over the past decade, the data format for scenario timeseries data established by the Integrated Assessment Modeling Consortium (IAMC, see box below) was used in numerous model comparison projects and the contributions by WG3 to the IPCC reports. This has resulted in partly diverging lists of variables and region naming conventions across different projects, leading to confusion among scenario users and additional maintenance effort for modellers.

In the PRISMA project, we re-integrate these parallel, partly incompatible developments into a consistent and comprehensive “nomenclature”, i.e., a list of variables including definitions and related units. This deliverable presents the starting point for this new nomenclature based on two previous projects NAVIGATE and ENGAGE, as well as several tools and processes to manage future additions to this nomenclature as new requirements arise in the modelling community.

Our ambition is to implement all processes and tools in PRISMA in line with best practice of open science and transparent, reproducible research. The process for discussing proposals for extending the variable template and comparison-region definition are managed via a publicly accessible GitHub repository released under a CC0-license (copyright waiver) for simple reuse and transparency. The technical implementation is built on two related open-source software packages developed in Python.

THE IAMC DATA FORMAT AND THE “NOMENCLATURE” CODELISTS

The Integrated Assessment Modeling Consortium (IAMC, <https://www.iamconsortium.org>) developed a simple tabular data format to exchange results and input assumptions from models to facilitate scenario analysis.

The format consists of five “index dimensions” and timeseries data in wide format.

	A	B	C	D	E	F	G	H	
1	Model	Scenario	Region	Variable	Unit	2005	2010	2015	
2	MESSAGE	CD-LINKS 400	World	Primary Energy	EJ/y	462.5	500.7	...	

Read more about the IAMC format at <https://pyam-iamc.readthedocs.io/en/stable/data.html>

Each model-comparison project defines a “**template**” or “**nomenclature**” specifying an agreed list of *variables* (including definitions and required units) and *regions* for facilitating comparability between models with bespoke regional or sectoral resolution. In line with conventions established by **SDMX** (a global initiative to improve Statistical Data and Metadata eXchange, <https://sdmx.org>), such agreed list of terms are called “codelists”.

IIASA, in consultation with the broader community, developed an open-source Python package **nomenclature** for management and validation of such project templates and scenario results. See <https://nomenclature-iamc.readthedocs.io> for further information.

2 A common GitHub repository for *definitions* across projects

Numerous integrated-assessment and energy-systems modelling projects currently use the IAMC format, including the Horizon 2020 projects ENGAGE, NAVIGATE and ECEMF. WG3 of the Intergovernmental Panel on Climate Change (IPCC) used the data format for compiling a scenario ensemble of pathways to support the quantitative assessment of mitigation options. The EU Scientific Advisory Board on Climate Change (ESABCC) also based its advice to the European Commission on a scenario ensemble compiled and hosted by IASA in that format (Byers et al., 2023; <https://data.ece.iiasa.ac.at/eu-climate-advisory-board>). However, each project defined the codelists independently, leading to divergent conventions and causing frustration, confusion and potential for errors among modellers and scenario users.

To remedy this situation and provide a standardized set of codelists going forward, the PRISMA project set up a **new GitHub repository** under the umbrella of the IAMC to hold an agreed set of codelists that be used in full or in part by PRISMA and future projects with a similar scope. The repository is available under a **CC0-License** (copyright waiver) at <https://github.com/iamconsortium/common-definitions> to facilitate reuse and adoption by other projects or communities.

2.1 Integration with the existing open-source infrastructure

The IAMC data format is the foundation of a larger ecosystem developed by IIASA and other institutions in the energy and climate modelling communities, including the interactive, web-based **IIASA Scenario Explorer** used by numerous projects for scenario data dissemination (see <https://data.ece.iiasa.ac.at>) and the open-source Python package **pyam** (Huppmann et al., 2021; <https://pyam-iamc.readthedocs.io>) for scenario processing and analysis. The newly created **common-definitions** repository was designed to be seamlessly integrated within that ecosystem for model development and scenario analysis.

In conjunction with the release of the common-definitions GitHub repository, we extended the open-source Python package **nomenclature**, which facilitates the management of project templates and codelists (<https://nomenclature-iamc.readthedocs.io>). Release v0.11 of the package adds a feature to import codelists and region mappings from the common-definitions repository to a project-specific workflow repository. This allows to merge codelists and definitions that are used consistently across all projects in the integrated-assessment and energy-systems domain with sector-specific variables that are relevant only for a particular project.

For the PRISMA project, we have set up the **first use case** for this hierarchical nesting of the common-definitions repository at <https://github.com/iiasa/prisma-workflow>. This instance will serve as a prototype for future projects.

2.2 Harmonizing variable and region codelists from previous projects

Based on discussions in the first six months of the PRISMA project, the Horizon 2020 projects ENGAGE and NAVIGATE were identified as the most relevant starting points. As a result, the region and variable codelists from these two projects for the key indicators GDP, population, final Energy, primary Energy and emissions (including the most relevant subcategories and sectors) were harmonized and added to the common-definitions repository. Refer to the repository at <https://github.com/iamconsortium/common-definitions> for the full list of variables and regions.

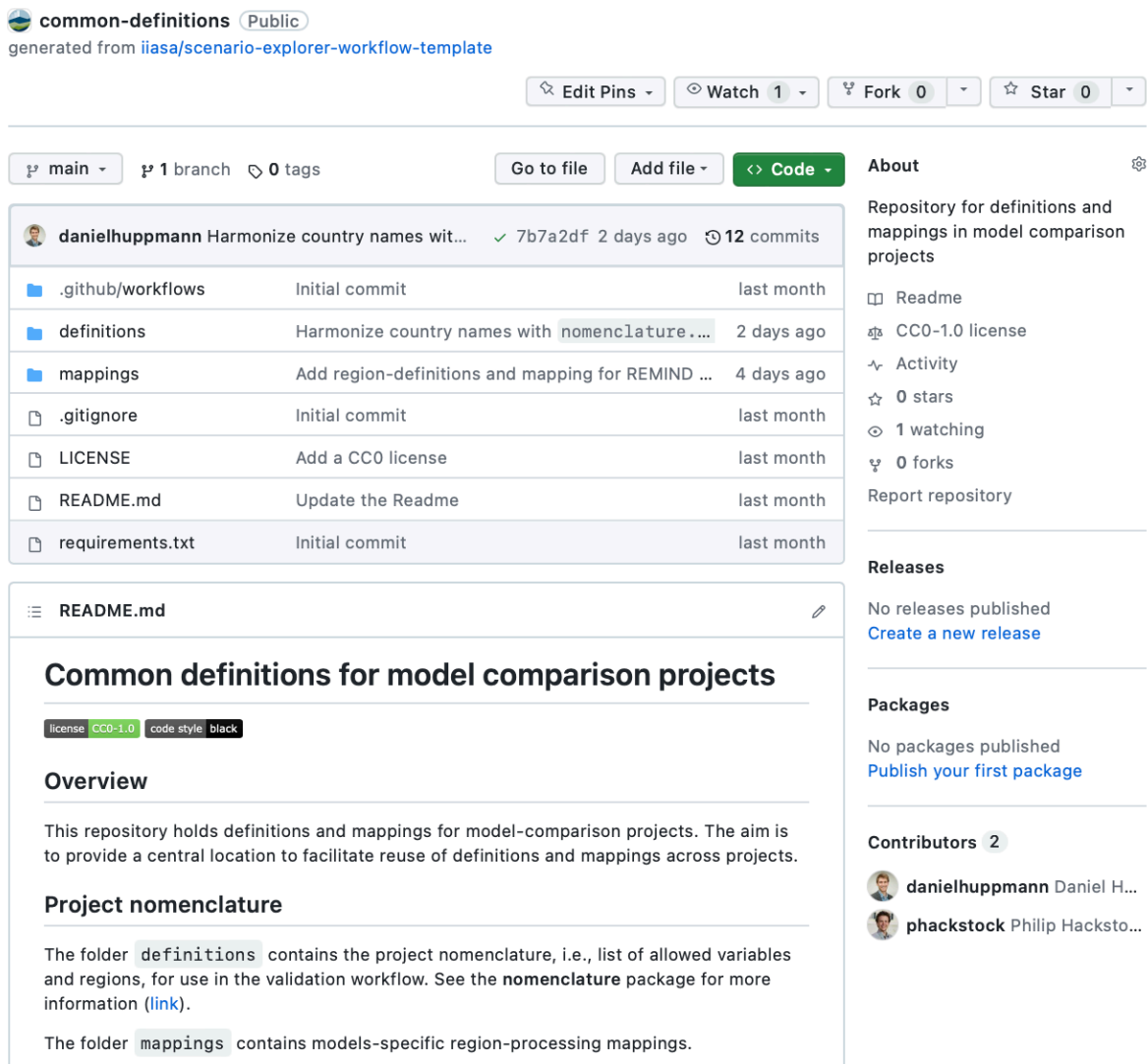


Fig 01 – Screenshot of the landing page of the common-definitions repository, see <https://github.com/iamconsortium/common-definitions>

This integrated and harmonized repository can now serve as a starting point for integrating other relevant indicators, including technical parameters such as the agriculture sector, installed capacity, trade flows. We have established a community process where modellers can propose additional indicators to be added to the definitions, in line with requirements for their projects. A steering group will then discuss how to incorporate these suggestions such that they do not (again) lead to conflicts or inconsistencies across previous or ongoing projects.

The process for these discussions will be implemented directly in the open-source GitHub repository, using the tools provided by GitHub for discussions and version control of any code

changes. This will facilitate traceability of decisions and foster a sense of an open community where any modelling team can contribute and share their ideas.

3 A standardized naming convention for country names

A related issue for frustration and complications in energy systems modelling, in particular for working with reference data for model development and calibration, are the inconsistent use of country names between different institutions and statistical agencies, including the UN, Eurostat, US Energy Information Agency (EIA), etc.

As a remedy, one common suggestion is to use ISO 3166-1 alpha-3 codes (also known as ISO3 codes). This works in some cases, but fails when working with data for Kosovo (not included in ISO 3166-1) or with reference data where small countries are included in larger, neighbouring countries (e.g., Liechtenstein treated as part of Switzerland for statistical purposes). It can also lead to confusion when it is not clear whether the region name NAM refers to Namibia or North America (which, depending on the model or data source, may include Mexico or only refer to the US and Canada).

Hence, as a step towards harmonization, we developed a „**countries**“-module to the Python package **nomenclature**, with the aim of establishing an agreed convention within the energy and climate modelling community. The implementation builds on the Python package **pycountry**, which in turn is based directly on ISO-3166 country names. However, in line with established practices and conventions in the IAM community, we added some simplifications and alterations to the ISO standard, like using “Bolivia” instead of “Bolivia, Plurinational State of”.

These simplifications of names serve the purpose of making it easier for users to work with scenario data. Also, by removing all “,-”-signs from the country names makes it much easier working with country names in csv or yaml file formats.

The package also enables simple translation between country-names and ISO2/ISO3 codes, following the API of the **pycountries** package. The package also adds mappings for the alternative ISO2-codes used by the European Commission, *EL* and *UK* instead of *GR* and *GB* for Greece and the United Kingdom, respectively.

Refer to <https://nomenclature-iamc.readthedocs.io/en/stable/api/countries.html> for a usage example and more information about this feature.

Scientific references

Edward Byers, Elina Brutschin, Fabio Sferra, Gunnar Luderer, Daniel Huppmann, Jarmo Kikstra, Robert Pietzcker, Renato Rodrigues, and Keywan Riahi. *Scenarios processing, vetting and feasibility assessment for the European Scientific Advisory Board on Climate Change*, **International Institute for Applied Systems Analysis, Laxenburg**, 2023.

url: <https://pure.iiasa.ac.at/18828>

Daniel Huppmann, Matthew Gidden, Zebedee Nicholls, Jonas Hörsch, Robin Lamboll, Paul Natsuo Kishimoto, Thorsten Burandt, Oliver Fricko, Edward Byers, Jarmo Kikstra, Maarten Brinkerink, Maik Budzinski, Florian Maczek, Sebastian Zwickl-Bernhard, Lara Welder, Erik Francisco Álvarez Quispe, and Christopher J. Smith. *pyam: Analysis and visualisation of integrated assessment and macro-energy scenarios*. **Open Research Europe**, 2021.

doi: [10.12688/openreseurope.13633.2](https://doi.org/10.12688/openreseurope.13633.2)

