2025-03-14

econdataverse: A universe of packages to work seamlessly with economic data

Christoph Scheuch
Independent

Signatories

Project team

The core project team consists of:

- Project Lead: Christoph Scheuch, Founder of Tidy Intelligence, with extensive experience in economic data analysis and software development
- Lead Developer: Christopher Smith, President of Promptly Technologies, a seasoned R and Python developer with a background in data engineering
- Lead Analyst: Teal Emery, Founder of Teal Insights, a research consultant with 10+ years of experience working with international development data

Contributors

TODO: ask & add contributors

Consulted

TODO: ask & add ISC members

The Problem

Economic data is essential for research and policy analysis, yet it remains highly fragmented, inconsistently formatted, and difficult to access efficiently through R. While some data is available through public APIs, a significant portion exists in static formats such as spreadsheets and reports, requiring time-consuming manual processing. Analysts and researchers working with multi-source economic data face inefficiencies due to disparate tools with varying designs, syntaxes, and usability.

This challenge is particularly evident in sovereign debt analysis, where crucial datasets - such as the World Bank's International Debt Statistics - are publicly available but require extensive cleaning and transformation before use. Organizations often spend tens of thousands of dollars annually on commercial data platforms that primarily provide better interfaces to freely available data. Existing tools for accessing economic data are fragmented and lack standardization, leading to redundant efforts and inefficiencies in data workflows.

The proposal

Overview

The `econdataverse` initiative was conceived as a unified ecosystem of packages for economic data access and analysis, applying modern software engineering principles to streamline workflows and enhance reproducibility. By enforcing consistent function naming, tidy data formats, and cross-source compatibility, it will significantly reduce the time spent on data acquisition and preparation and facilitate the creation of reproducible workflows.

This initiative will directly benefit the R community by:

- Supporting reproducible research with standardized access to economic data
- Lowering the learning curve for working with economic data sources
- Creating a scalable foundation for advanced economic data analysis

Detail

The project will develop modular R packages, each targeting major economic data sources that are frequently used in economic analysis but historically difficult to access due to API inconsistencies or unavailability of APIs. The currently released or planned packages include:

- `wbids` (released to CRAN on 2024-11-15): World Bank International Debt Statistics (IDS) API, critical for sovereign debt sustainability analysis
- `wbwdi` (released to CRAN on 2025-02-25): World Bank World Development Indicators (WDI) API, a large number of country or region-level indicators for various contexts
- `owidapi` (released to CRAN on 2025-02-27): Our World in Data (OWID) API, open-source data for long-term economic trends and social indicators
- `uisapi` (released to CRAN on 2025-03-06): UNESCO Institute of Statistics (UIS) API, education and research data relevant for policy analysis
- `imfweo` (in development): IMF World Economic Outlook (WEO), global economic projections and country-level economic performance
- `oecdoda` (planned): OECD Official Development Assistance (ODA), aid flow and development finance tracking

Additional supporting tools to address cross-source compatibility and ease of use are in development:

- `econid` (in development): standardization and conversion utilities for country, region, and institution identifiers used in economic datasets
- `econtools` (in development): common economic data analysis utilities

Minimum Viable Product

For the initial release of `econdataverse`, we will focus on:

- Core packages for the primary data sources (WDI, IDS, WEO, ODA, UIS, OWID)
- Core packages for combining and analyzing economic data (`econid`, `econtools`)
- A unified meta-package ensuring seamless cross-source access (`econdataverse`)
- Documentation and vignettes for quick adoption
- Compliance with CRAN guidelines

Architecture

The `econdataverse` follows a modular architecture to maximize efficiency and maintainability while ensuring compatibility across data sources. Each package is endowed with dedicated CI / CD pipelines and unit tests to detect and isolate issues. Moreover, users who don't need the full suite of packages can load individual packages and avoid excessive dependencies.

Assumptions

- Data formats from sources won't undergo major breaking changes
- The R community values consistent interfaces and tidy data approaches

Project plan

Start-up phase

June 2025:

- Set up a dedicated GitHub organization with clear contribution guidelines
- Migrate the existing website to the new organization
- Initialize the `econdataverse` package and collect open issues
- Outline a roadmap with milestones and meeting schedule

Technical delivery

July - August 2025:

- Resolve issues in existing core packages based on user feedback
- Release missing core packages to CRAN and collect user feedback
- Work on documentation for core packages and the `econdataverse` paclage

September 2025:

- Release `econdataverse` package to CRAN

Other aspects

- Announce the release of each packags on LinkedIn and BlueSky
- Create blog posts for individual packages releases (e.g. tidy-intelligence.com) and include them in R Weekly newsletters
- Submit the `econdataverse` project for the UseR! 2026 conference, posit::conf(2026), and EARL 2026

Requirements

People

We currently have all the skills to create the project in the team. However, we are more than welcome to add additional contributors along the way.

Processes

The project requires a clear code of conduct that provides guidlines for contributors to existing packages or new packages as well as handover plans in case any existing maintainer wants to stop working on a project.

Tools & Tech

All tools & technologies to deliver this project are readily available and established:

- GitHub for code management, issue tracking, and collaboration
- GitHub Actions for automated testing based on `testthat` and code coverage checks using `covr`
- GitHub Pages for comprehensive documentation via `pkgdown`

Funding

Funding will be used to compensate developers and maintainers to commit time in creating packages and documentation. To get the full commitment of the project team for over the outlined 4-month period, the detailed funding requirements are:

- 10,000 USD for development
- 5,000 USD for documentation

Summary

The only bottleneck at the moment is funding development and documentation. Since the team consists of independent developers and researchers, it is important that the so that the team can commit resources to the project.

Success

Definition of done

- The `econdataverse` meta-package and its underlying core packages published to CRAN
- Function documentation and vignettes available via `pkgdown` websites
- Test Coverage of 90%+ test coverage for all released packages
- At least 1,000 CRAN downloads within three months of release for `econdataverse`

Measuring success

- User adoption: number of CRAN downloads of econdataverse packages using cranlogs

Future work

- Expand support for additional economic data sources

- Develop Shiny apps for interactive data visualization
- Create educational materials for economics courses using R
- Implement advanced features like automatic data updating and versioning

Key risks

- Unexpected API changes or data access restrictions
- Lack of community engagement (users or contributors)
- Difficulty maintaining packages long-term due to maintainers becoming unavailable