

2025-03-31

econdataverse: A Universe of Packages to Work Seamlessly with Economic Data

Christoph Scheuch 
Tidy Intelligence

Teal Emery 
Teal Insights

Christopher C. Smith 
Promptly Technologies

Signatories

The core project team combines technical expertise with domain knowledge in economic policy, software engineering, and product management to ensure that the project will meet the practical needs of both technical users and policymakers:

- **Project Lead:** Christoph Scheuch, Founder of Tidy Intelligence, holds a PhD in Finance from Vienna Graduate School of Finance and is the co-author of “Tidy Finance with R” and “Tidy Finance with Python.” Previously, he served as Director of Product and Head of Business Intelligence & Data Science at wikifolio.com and has extensive experience transforming complex financial data into elegant, efficient solutions.
- **Lead Developer:** Christopher C. Smith, President of Promptly Technologies, holds a PhD in Religion from Claremont Graduate University and is a seasoned R and Python developer. He revamped the `imfr` package, a critical tool for R users accessing International Monetary Fund data. His background in data engineering and his expertise in API and SDK development will be essential for building reliable economic data pipelines.
- **Lead Analyst:** Teal Emery, Founder of Teal Insights, advises countries and institutions on sovereign debt and climate finance issues while building open-source tools to improve policy analysis. He is an Adjunct Lecturer in International Finance at Johns Hopkins SAIS and previously worked at the World Bank and Morgan Stanley Investment Management researching emerging markets.

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. Research shows this is a pervasive

challenge, with data scientists spending approximately 45% of their time on data preparation tasks, including 26% specifically on data cleaning (Anaconda 2020). 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.

This challenge has real-world consequences. Decision-makers in developing countries often lack the resources to process complex economic data effectively, forcing them to:

1. Pay tens of thousands of dollars annually for commercial data platforms that merely provide better interfaces to freely available data
2. Hire expensive external consultants to perform basic data integration tasks
3. Make critical policy decisions with incomplete or outdated information

The financial impact of these inefficiencies is substantial. A 2002 Data Warehousing Institute report estimated based on survey data that poor quality customer data cost U.S. businesses \$611 billion annually in postage, printing, and staffing costs (Eckerson 2002). English (2009) catalogued corporate disasters from poor quality business information amounting to approximately \$1.25 trillion (English 2009). A closed-source estimate by IBM, published in a 2016 infographic, suggested poor data quality costs the U.S. economy approximately \$3.1 trillion per year (IBM 2016).

Notably, academic studies using open data and methodologies to quantify these costs are thin on the ground, which ironically underscores the need for better tools to facilitate precisely this sort of analysis.

According to research on digital development programs, key challenges include merging data from different sources, validating accuracy, and extracting data from non-standard formats (ICTworks 2022). For example, in sovereign debt sustainability analysis, policymakers must integrate World Bank debt statistics with IMF economic forecasts to evaluate sustainability. A task that should take minutes can consume hours or days of analyst time reconciling inconsistent country codes, standardizing formats, and validating calculations.

These challenges not only waste resources but also significantly impair the ability of governments, particularly in resource-constrained environments, to respond to pressing economic, climate, and social challenges with evidence-based policies. Indeed, NBER research by Nagaraj and Tranchero (2023) demonstrates that improved data access for economists leads to an 18.3% increase in high-quality publications and a 25-35% increase in citation impact, underscoring how data access directly facilitates effective economic research and policy.

The Proposal

Overview

The `econdataverse` initiative was conceived as a unified ecosystem of packages for economic data access and analysis, applying the proven principle that big things get built by creating strong, durable building blocks that can be reliably stacked together. By enforcing consistent function naming, tidy data formats, and cross-source compatibility, we significantly reduce the time spent on data acquisition and preparation and facilitate the creation of reproducible workflows.

This initiative emerged directly from our work developing tools for economic policymakers in developing countries who are navigating critical climate investment decisions during periods of high debt stress. The project addresses a fundamental gap in the data infrastructure needed to support evidence-based economic policy.

Audience

The econdataverse will serve two distinct groups:

1. **Direct users** of R packages: Technical staff at international organizations, researchers, economists, and data analysts in developing country economic policymaking institutions who need efficient access to economic data
2. **Indirect beneficiaries**: High-level policymakers without R programming skills who will gain access to insights through Shiny applications, parameterized reports, and other interfaces powered by econdataverse packages

By building rock-solid data infrastructure components with comprehensive unit tests, continuous integration, and detailed documentation, the econdataverse will create a foundation for both technical analysis and accessible policy tools.

Motivation

- Supporting reproducible research with standardized access to economic data
- Providing programmatic access to novel data sources
- Lowering the learning curve for working with economic data sources
- Creating a scalable foundation for advanced economic data analysis
- Democratizing access to critical economic data for policymakers in developing countries
- Enabling faster, more informed responses to economic and climate challenges
- Saving organizations thousands of dollars in commercial data subscription costs

Real-World Impact: Tools for Economic Policymakers

The econdataverse initiative is already demonstrating practical impact through the [Debt Path Explorer](#), the first of several tools we’re developing for economic policymakers from climate-vulnerable countries (V20). This web application, powered by our [imfweo](#) package, helps policymakers understand how sustainability targets could affect their debt trajectories without expensive data subscriptions or consultant fees. It transforms complex economic data into accessible insights and will be presented to policymakers for feedback at the spring IMF/World Bank Meetings, enabling rapid refinement based on real user needs.

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** (prototyped on GitHub): IMF World Economic Outlook (WEO), global economic projections and country-level economic performance

- **imfifs** (planned): IMF International Financial Statistics (IFS), country-level financial stability data
- **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:

- **econid** (released to CRAN on 2025-03-18): standardization and conversion utilities for country, region, and institution identifiers used in economic datasets
- **econtools** (planned): 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 (IDS, WDI, OWID, UIS, WEO, ODA, IFS)
- Core packages for combining and analyzing economic data (**econid**, **econtools**)
- A unified meta-package ensuring seamless cross-source access (**econdataverse**)
- Articles that combines multiple data sources for modeling and visualizations
- Compliance with the CRAN Repository Policy

Architecture

The **econdataverse** employs a modular architecture inspired by the principle that “big things get built by making strong, durable building blocks.” Each package features robust CI/CD pipelines and comprehensive unit tests that quickly identify and isolate potential issues. Users can selectively load individual packages rather than the entire suite, eliminating unnecessary dependencies and optimizing resource utilization.

This architectural approach enables both immediate practical applications and future expansion, allowing the components to be recombined in ways we haven’t yet imagined to tackle emerging economic challenges.

Assumptions

- Data sources won’t undergo major breaking changes with respect to accessibility
- 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 additional user feedback
- Work on documentation for core packages and the **econdataverse** package
- September 2025:
 - Release stable version of **econdataverse** package to CRAN

Other aspects

- Announce the release of each package 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 2026 UseR!, posit::conf(), and EARL

Requirements

People

Our team possesses all necessary skills to execute this project successfully. We remain open to welcoming additional contributors throughout the development process.

Processes

The project requires a clear code of conduct that provides guidelines for contributors to existing packages and developing new packages, as well as succession plans should any maintainer need to transition away from the project.

Tools & Tech

All required tools and technologies are established and readily accessible::

- GitHub for version control, issue management, and collaboration
- GitHub Actions for automated testing via **testthat** and code coverage analysis with **covr**
- GitHub Pages for hosting comprehensive documentation generated through **pkgdown**

Funding

Financial resources will support developer and maintainer compensation, ensuring dedicated time for package development and documentation. To secure the project team's commitment during the 4-month timeline, we require:

- \$7000 for development activities
- \$3000 for documentation efforts

Summary

Currently, our only constraint is securing funding for development and documentation. As our team comprises independent developers and researchers, financial support is essential to enable the allocation of sufficient resources to ensure project success.

Success

Definition of done

- The `econdataverse` meta-package and its underlying core packages published to CRAN
- Function documentation, vignettes, and articles available via `pkgdown` websites
- 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 using economic data
- Create educational materials for economics courses using R
- Implement advanced features like automatic data updating and versioning

Key risks

- Lack of user engagement
- Unexpected API changes and data access restrictions
- Difficulty maintaining packages long-term due to maintainers becoming unavailable

References

- Anaconda. 2020. “2020 State of Data Science: Moving from Hype Toward Maturity.” Anaconda, Inc. <https://www.anaconda.com/state-of-data-science-2020>.
- Eckerson, Wayne W. 2002. “Achieving Business Success Through a Commitment to High Quality Data.” TDWI Report Series: Data Quality and the Bottom Line. The Data Warehousing Institute. <https://www.dw-institute.com/dqreport/>.
- English, Larry P. 2009. *Information Quality Applied: Best Practices for Improving Business Information, Processes and Systems*. Wiley Publishing.
- IBM. 2016. “Extracting Business Value from the 4 Vs of Big Data.” IBM Big Data Hub. 2016. <https://web.archive.org/web/20200102183321/https://www.ibmbigdatahub.com/infographic/extracting-business-value-4-vs-big-data>.
- ICTworks. 2022. “3 Key Challenges to Data Cleaning in Digital Development Programs.” ICTworks. March 30, 2022. <https://www.ictworks.org/data-cleaning-digital-development/>.
- Nagaraj, Abhishek, and Matteo Tranchero. 2023. “How Does Data Access Shape Science? The Impact of Federal Statistical Research Data Centers on Economics Research.” NBER Working Paper 31372. National Bureau of Economic Research. <http://www.nber.org/papers/w31372>.