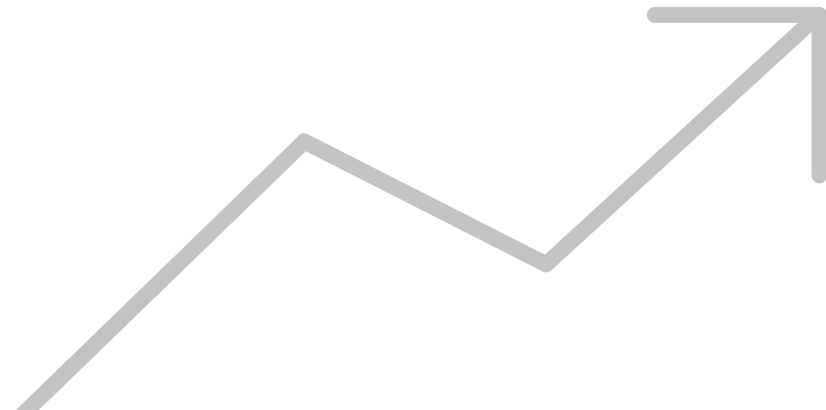


# VGR Modernisierung im Statistischen Bundesamt

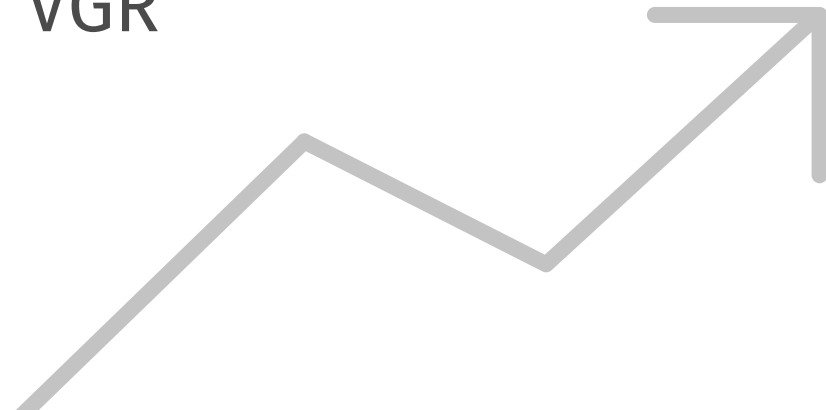
Valène Junique, Statistisches Bundesamt

D15 – Vermögensrechnung, Internationale VGR Methodik



# Agenda

1. Ausgangssituation
2. R-Modernisierung in den VGR
  - Warum R?
  - Zeitplan
  - R-Austausch in den VGR
3. Neue Datenorganisation in den VGR
  - EDO
  - Datenbank



# 1. Ausgangssituation

- Die VGR im Statistischen Bundesamt:
  - 2 Gruppen und 13 Referate
  - Unterschiedliche Technologien: **Excel**, VBA, SAS, R, etc.
  - Datenhaltung größtenteils Bestandteil der Rechenprozesse
  - Für den Datenaustausch zwischen den Referaten fehlten bisher gemeinsame Standards
- R im Statistischen Bundesamt:
  - R Server seit 2018 mit Update in 2025
  - Einführung von GitLab in 2026
  - R als zunehmend wichtiges Instrument, das vom statischen Verbund gefördert wird (z. B. LA OPTIKO in 2025: R als führendes Werkzeug für statistischen Auswertungsprogrammierungen im SDWH)
- Ziel für die VGR:
  - Rechenprozesse mit R zu automatisieren und vereinheitlichen
  - Standards für den Datenaustausch in den VGR zu definieren und anzuwenden

## 2. R-Modernisierung in den VGR

### Warum R?

- **Verbreitung & Ressourcen:** Weit verbreitet, viele R-Ressourcen und E-Learning-Angebote
- **Benutzerfreundlichkeit:** Einfach zu lernen
- **Know-how:** Neue Kolleginnen und Kollegen bringen oft R-Erfahrung mit
- **Internationale Nutzung:** Wird in anderen Ländern eingesetzt
- **Flexibilität:** Vielseitig einsetzbar und effiziente Verarbeitung großer Datenmengen
- **Unterstützung:** Abstimmung mit IT-Abteilung und Amtsleitung

## 2. R-Modernisierung in den VGR

### Zeitplan

**Ziel:** Bis etwa 2028 möglichst viele Prozesse mit R modernisieren

- **Schwerpunkte:**
  - Priorisierung zeitaufwändiger Berechnungen, bei denen R erhebliche Arbeitersparnis bietet
  - Neue Bausteine (soweit möglich) direkt mit R entwickeln
  - Schulung der Kolleginnen und Kollegen
- **Nach 2028: Fokus auf die Vorbereitung der Revision 2030, um Probleme durch parallele Modernisierung von Berechnungen und Revision zu vermeiden.**



## 2. R-Modernisierung in den VGR

### R-Austausch in den VGR

- Was ist das?
  - 2023 gegründet
  - Gruppe von 16 Kolleginnen und Kollegen aus unterschiedlichen VGR Referaten mit guten R Kenntnissen
  - Keine Entwickler!
  - Bindeglied zu IT Abteilung
- Was macht er?
  - R Ansprechpartnerinnen und -partner für die VGR
  - Entwickeln VGR-spezifischer R Funktionen und Code Beispiele
  - Definieren von „R-Standards“ (z. B. „R Style Guide“)
  - Und viel mehr...



## 2. R-Modernisierung in den VGR

### R-Austausch in den VGR

- **vgR-Schulung:** Interne R-Schulung mit Fokus auf VGR-spezifische Anwendungen (z. B. Arbeit mit Zeitreihen oder VGR-Prozesse/Funktionen in R)
- **R-Tandem:** Zusammenarbeit zwischen einem Mitarbeiter, der seine Berechnungen mit R modernisieren möchte, und einem „R-Experten“ aus dem R-Austausch
  - Lernen durch Praxis („*Learning by doing*“)
  - Wissenstransfer
- **Newsletter:** Kolleginnen und Kollegen über neue Entwicklungen informieren



Quelle: "Employee training illustration – in person training", Digits.co.uk, licensed under CC BY 2.0

## 3. Neue Datenorganisation in den VGR

### EDO = Effiziente Daten Organisation

Zielbild der Projekt Gruppe EDO aus dem Frühjahr 2021:

*„Alle Transferdateien folgen derselben Datensatzbeschreibung und verwenden einheitliche Schlüssel. Der Datentransfer erfolgt ausschließlich über ein hierarchisch aufgebautes System von Ordnern, das ausschließlich für Berechnungsdateien genutzt wird und organisatorisch von ansonsten notwendigen Dateitransfers abgegrenzt ist.“*

Ergebnis der Projekt Gruppe EDO:

- Prototyp eines standardisierten Austauschformats zwischen den Referaten, so-genannte **EDO-Datei** mit folgenden Eigenschaften:
  - Feste Datensatzbeschreibung (orientiert an SDMX von Eurostat)
  - Eindeutige Dateibezeichnungen und einheitliche Codes
  - Vollständige Merkmalsbeschreibungen
  - Trennung Datenhaltung und Datenverarbeitung
  - „Daten ins Schaufenster stellen“ - keine kleinteiligen Lieferdateien

# 3. Neue Datenorganisation in den VGR

## EDO = Effiziente Daten Organisation

- Datenhaltung im Sinne einer relationalen Datenbank
  - Flache Datentabellen
  - Metadaten und Werte trennen
  - Quartals- und Jahreswerte trennen
- EDO-Dateien grenzen den Umfang der Wertetabellen ein
  - Trennen von laufender Berechnung und Langzeitwerten
  - Redundanzen, wenn möglich, vermeiden
    - Nur die untere Informationsebene vollständig ablegen
    - Verzicht auf individuelle Zwischensummen, verkettete Volumen oder Indizes
- Tabellen/Werte werden über eindeutige Schlüssel (Keys) verknüpft
  - Schlüssel sind keine Informationsträger
  - Standardisierte Eurostat-Codes verwenden

# 3. Neue Datenorganisation in den VGR

## Datenbank

- **Ziel:** Einführung einer internen Datenbank für den VGR und die Konjunkturstatistik bis Ende 2026
  - Erste Tests für die VGR bereits im Gange
  - Tools für vereinfachte Interaktion mit der Datenbank (z.B. Wrap-Up-Funktionen) sind in Planung
- **Datenkonsistenz:** Einheitlichkeit zwischen Datenbank und EDO-Dateien (z. B. gleiche Keys)
- **Plattform:** Cloudera
- **Datenzugriff:** Abruf und Speicherung der Datenbankwerte erfolgt über den R-Server

**Zentrales System  
für die VGR und die Konjunkturstatistik,  
für standardisierten und effizienten  
Datenaustausch**

**Zentrale  
Datenhaltung**

**Zentraler  
Datenaustausch**

**Feste Daten-  
struktur**

**Hohe Verfügbarkeit  
des Systems**

**Archivierung**

**Datenintegrität**

# Kontakt

Statistisches Bundesamt  
Postanschrift  
65180 Wiesbaden

[www.destatis.de](http://www.destatis.de)

[www.destatis.de/kontakt](http://www.destatis.de/kontakt)

Ansprechpartnerin  
Valène Junique  
[valene.junique@destatis.de](mailto:valene.junique@destatis.de)  
Telefon +49 611 75-2497

