

About surveys, and other data

Andrea Brandolini

Bank of Italy, DG Economics, Statistics and Research

The views expressed here are solely those of the authors and they do not necessarily reflect those of the Bank of Italy

“The use of surveys for monetary and economic policy”

SUERF-Banca d'Italia-ECB-EIB Conference

Bank of Italy, Rome, 27 April 2023

Do we need surveys? Yes, of course

Strengths

- Timeliness, filling data gaps, adaptability to users' research needs – which is not the case with administrative data that are collected for different purposes
- An interest that goes back in time at the Bank of Italy

Paolo Baffi (then Head of research department, later Governor) to Luzzatto Fegiz in 1951

A survey ... gauging prevailing opinions regarding distribution of a possible increase in income (say, 10 and 50 per cent) between consumption and savings – distinguishing the former into current consumption goods and durable goods, and the latter into direct investments (in real estate or businesses), hoarding and other forms of savings ...

... capture changes in consumer habits ...

Do we need surveys? Yes, of course

Limitations

- Meyer, Mok, Sullivan, “Household Surveys in Crisis”, 2015

Large and nationally representative surveys are arguably among the most important innovations in social science research of the last century ...

*However, the quality of data from household surveys is in decline. Households have become increasingly less likely to answer surveys at all, which is the problem of **unit nonresponse**. Those that respond are less likely to answer certain questions, which is the problem of **item nonresponse**. When households do provide answers, they are less likely to be accurate, which is the problem of **measurement error**.*

- Under the urgency to provide answers, we sometimes tend to put aside these limitations ...

Surveys vs. administrative data

A false tension

- Way forward is integrating different sources, comparing their respective pros and cons → there is no perfect source
- We know limits of existing sources: Beware the new ones!
- Paul Schreyer, “Don’t trust just any statistics!”, 2023

With the digital transformation, the process of statistical production as well as available data and techniques have changed massively.

In general, a healthy dose of caution is required when using new sources for statistical production: while new ‘big’ data sets are often extensive, they are not necessarily representative; private sources may not be sustainable; or data ownership may be unclear. NSOs have to find the right balance between innovation and quality assurance for trustworthy statistics.

Embarrassment of riches? A test

How has inequality varied in Italy since World War II?

- Compare available evidence from:

Original sources

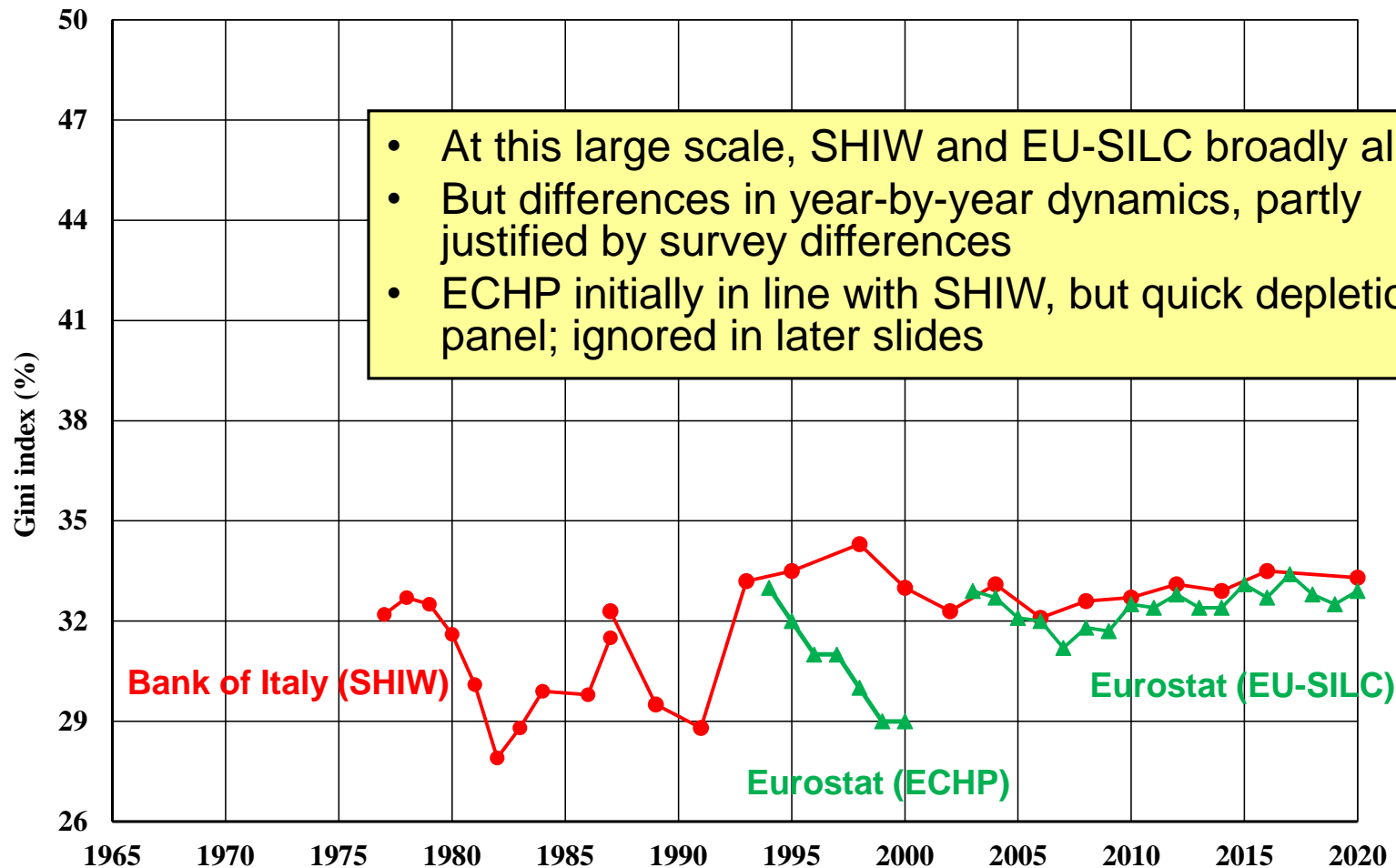
- Bank of Italy's Survey of Household Income and Wealth (SHIW)
- European Community Household Panel (ECHP), EU Statistics on Income and Living Conditions (EU-SILC)

Main international databases

- Luxembourg Income Study (LIS)
 - OECD Income Distribution Database (IDD)
 - World Bank Poverty and Inequality Platform (PIP)
 - UNU-WIDER World Income Inequality Database (WIID Companion)
 - Standardized World Income Inequality Database (SWIID)
 - World Inequality Database (WID)
 - Global Repository of Income Dynamics (GRID)
- Can we draw a coherent story?

Disposable income inequality in Italy

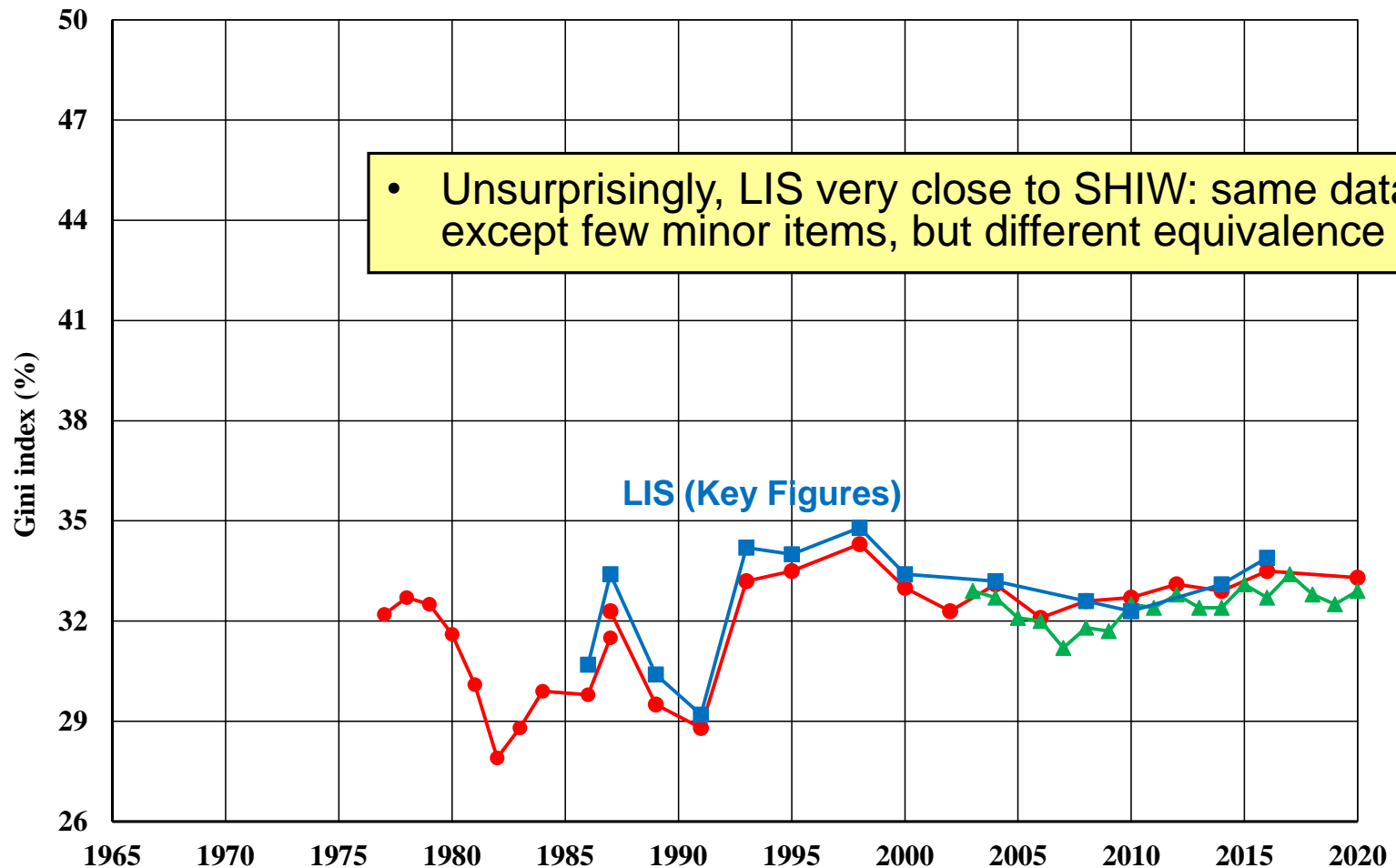
Original sources



- At this large scale, SHIW and EU-SILC broadly aligned
- But differences in year-by-year dynamics, partly justified by survey differences
- ECHIP initially in line with SHIW, but quick depletion of panel; ignored in later slides

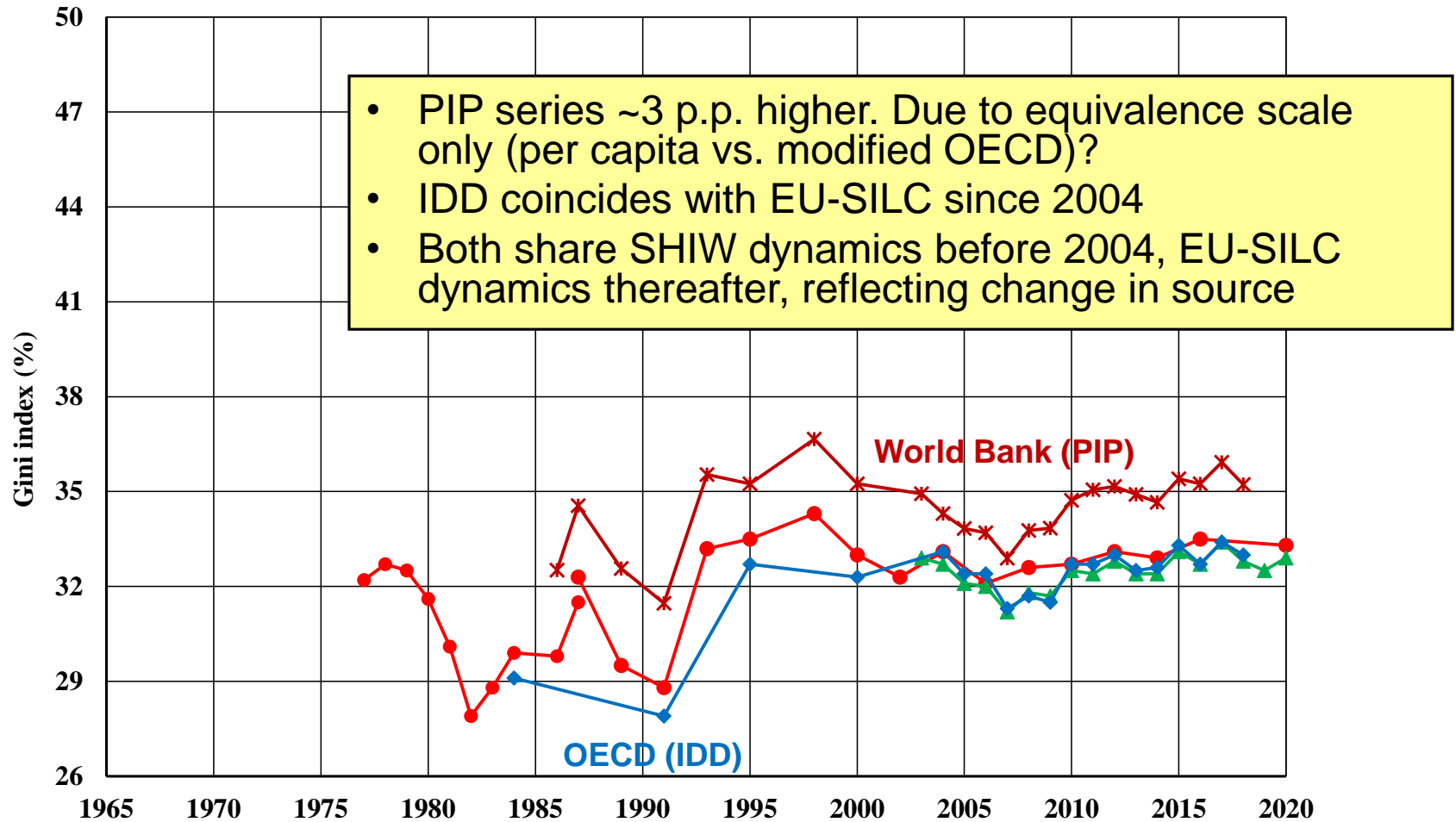
Source: websites of Bank of Italy, Eurostat.

Disposable income inequality in Italy LIS



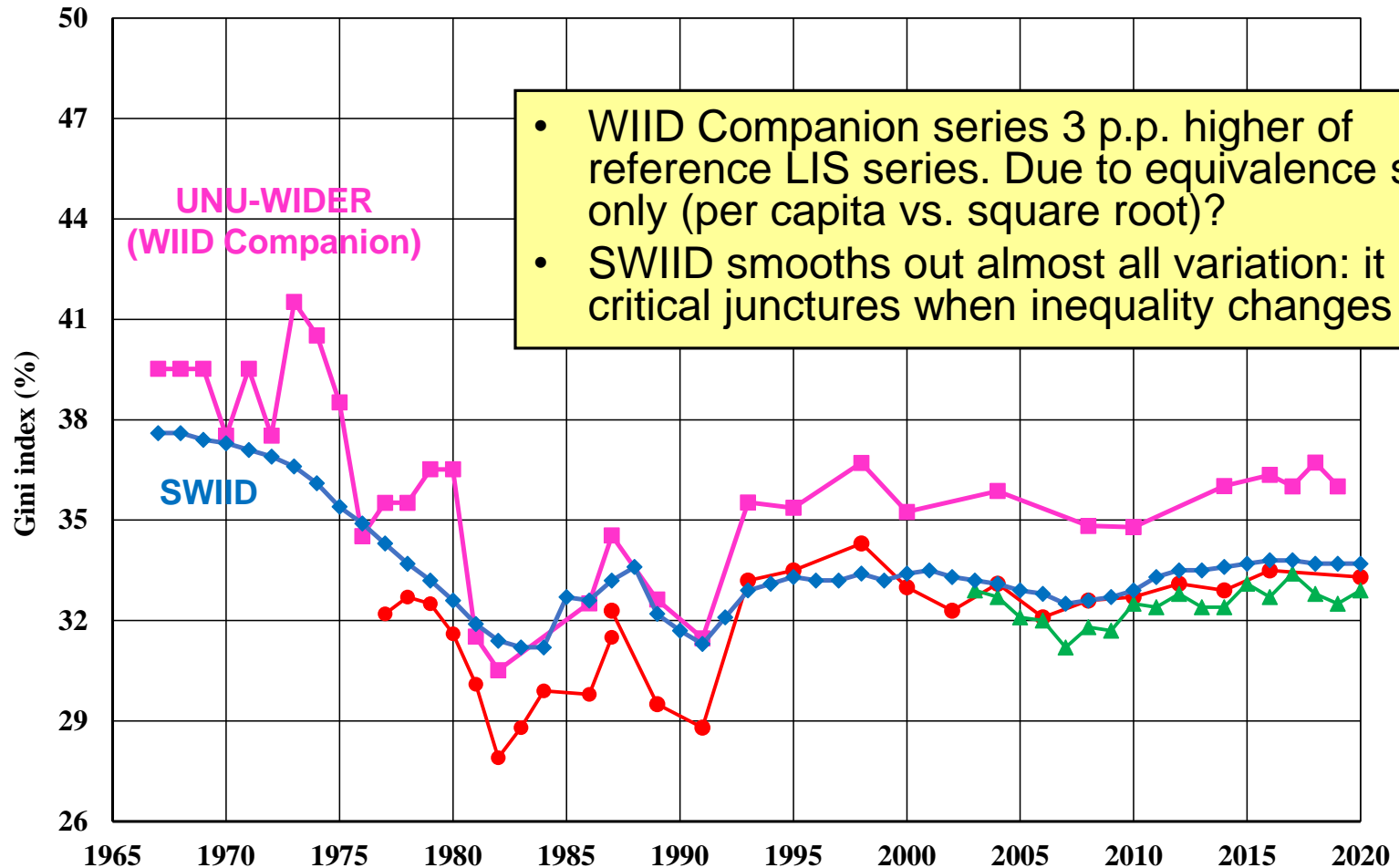
Source: websites of Bank of Italy, Eurostat, LIS.

Disposable income inequality in Italy OECD and World Bank



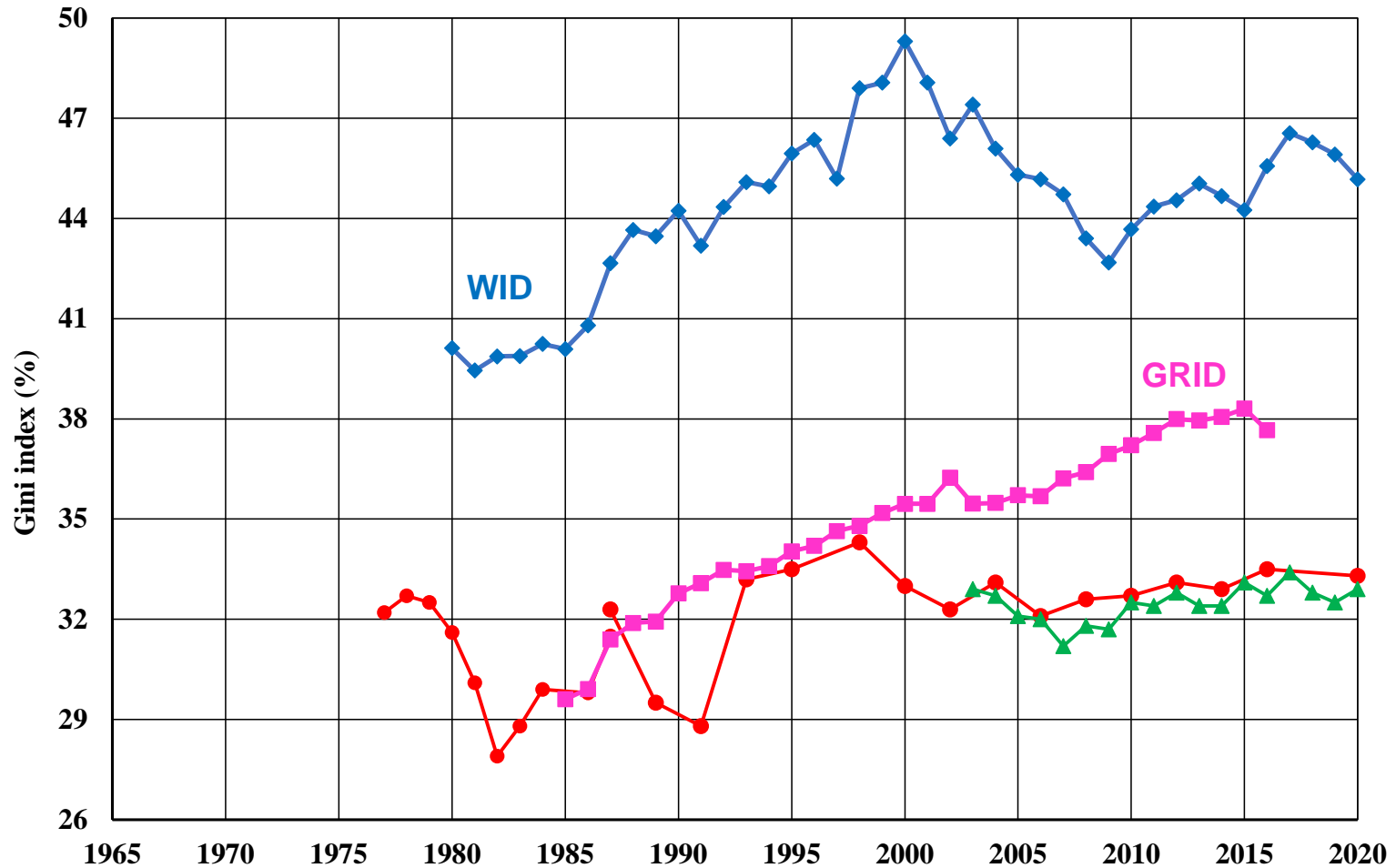
Source: websites of Bank of Italy, Eurostat, OECD, World Bank.

Disposable income inequality in Italy WIID Companion and SWIID



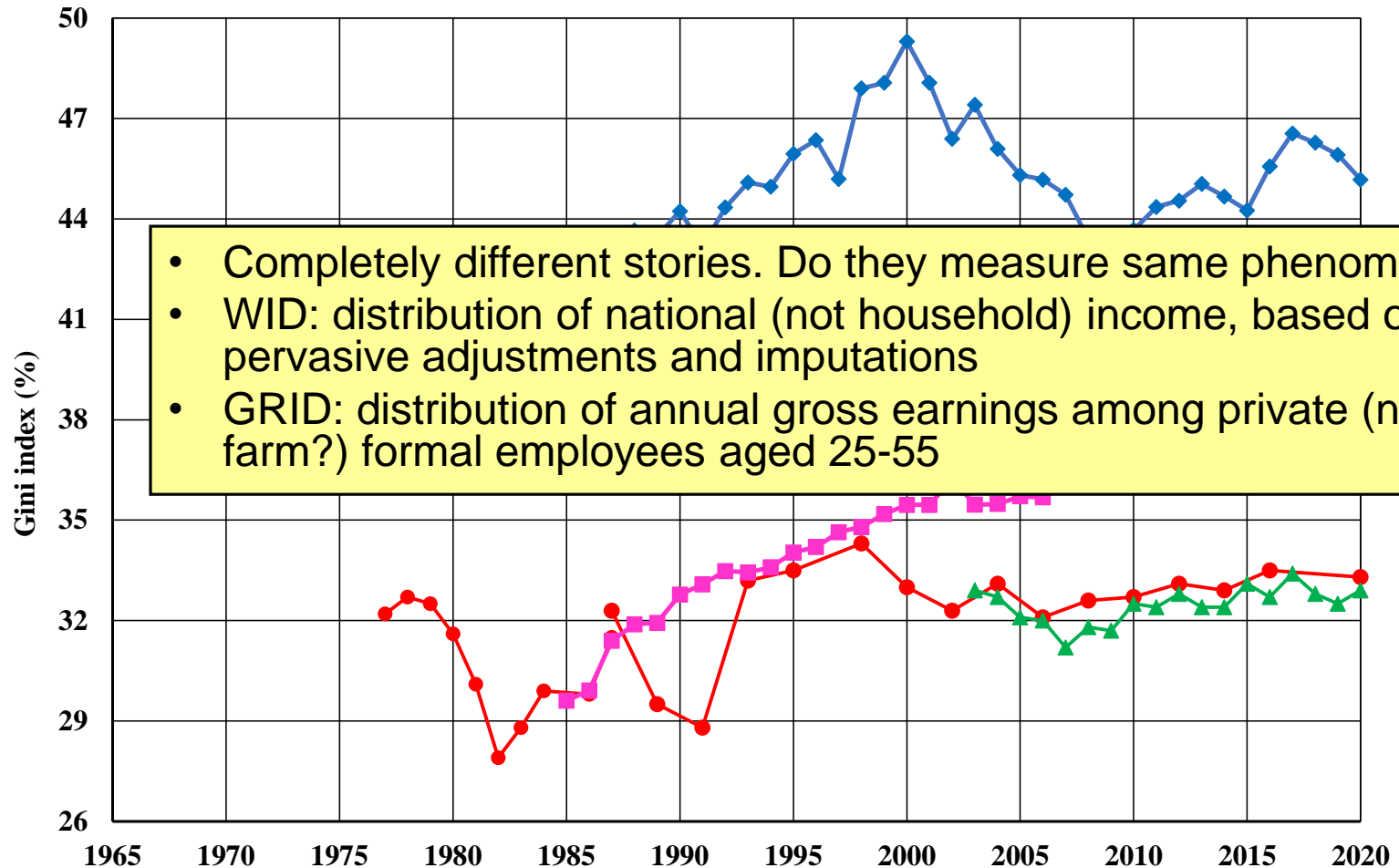
Source: websites of Bank of Italy, Eurostat, UNU-WIDER, SWIID.

Disposable income inequality in Italy WID and GRID



Source: websites of Bank of Italy, Eurostat, WID, GRID.

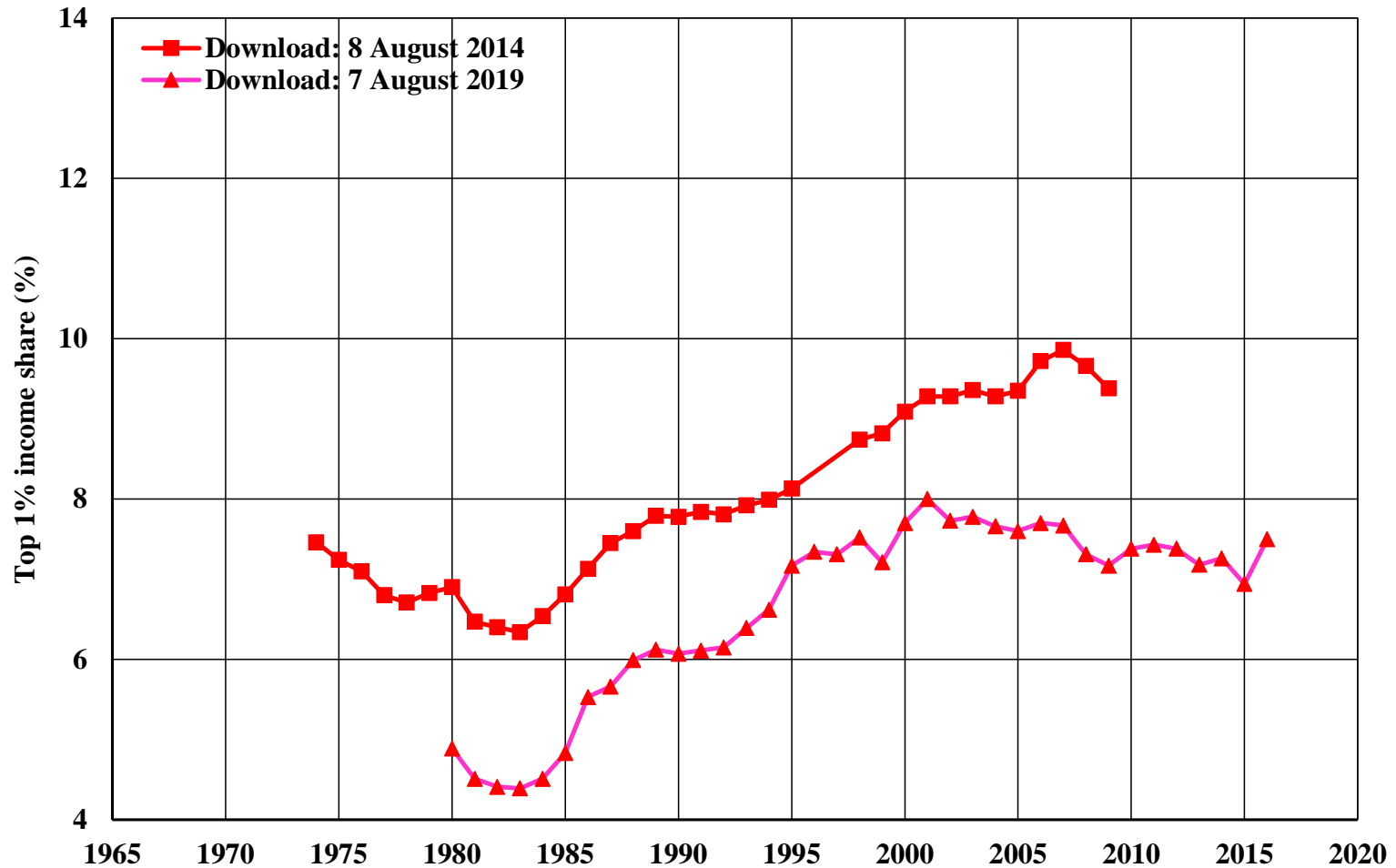
Disposable income inequality in Italy WID and GRID



- Completely different stories. Do they measure same phenomena?
- WID: distribution of national (not household) income, based on pervasive adjustments and imputations
- GRID: distribution of annual gross earnings among private (non-farm?) formal employees aged 25-55

Pre-tax income inequality in Italy

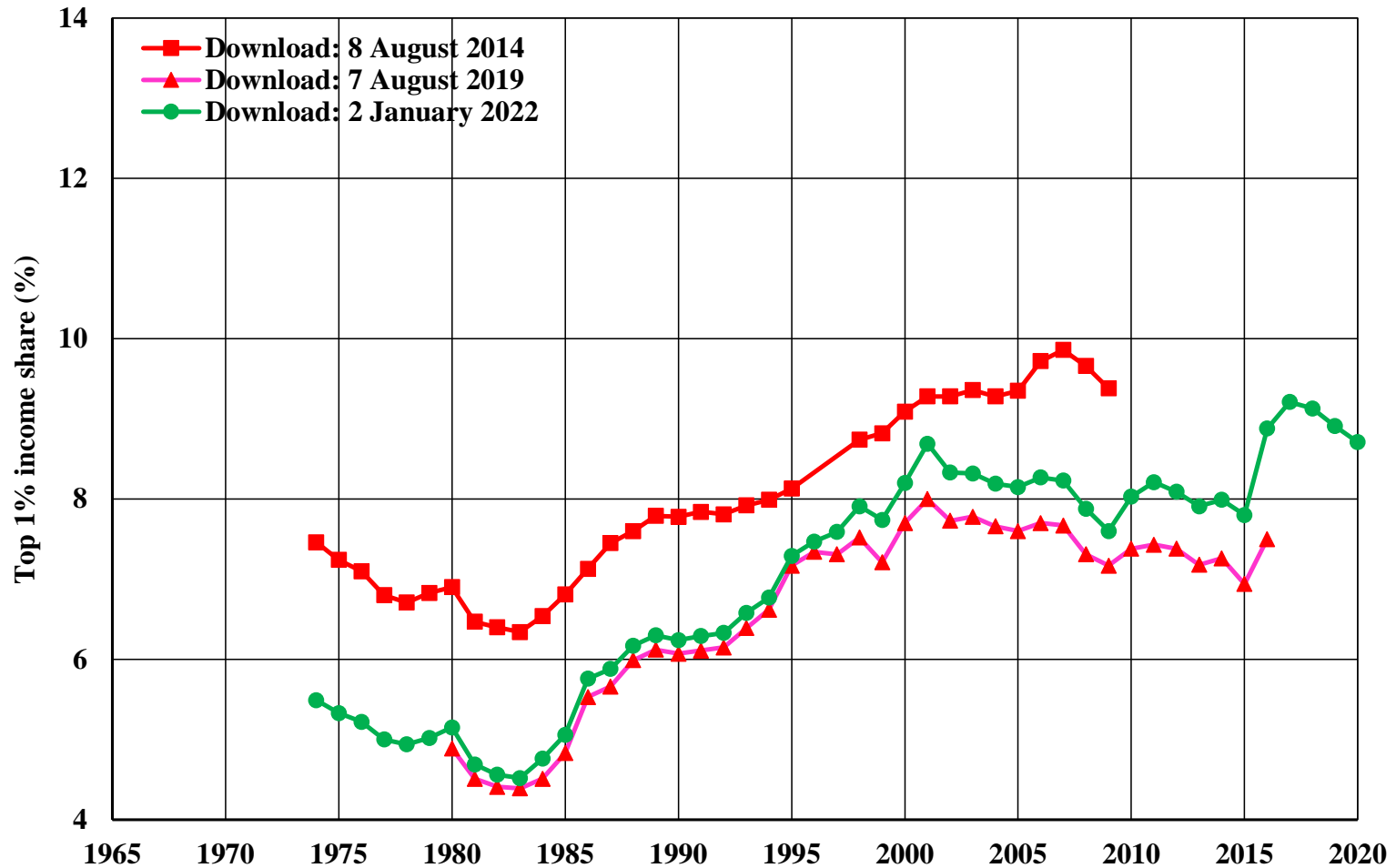
Different releases of WTID/WID



Source: websites of WTID/WID at different dates.

Pre-tax income inequality in Italy

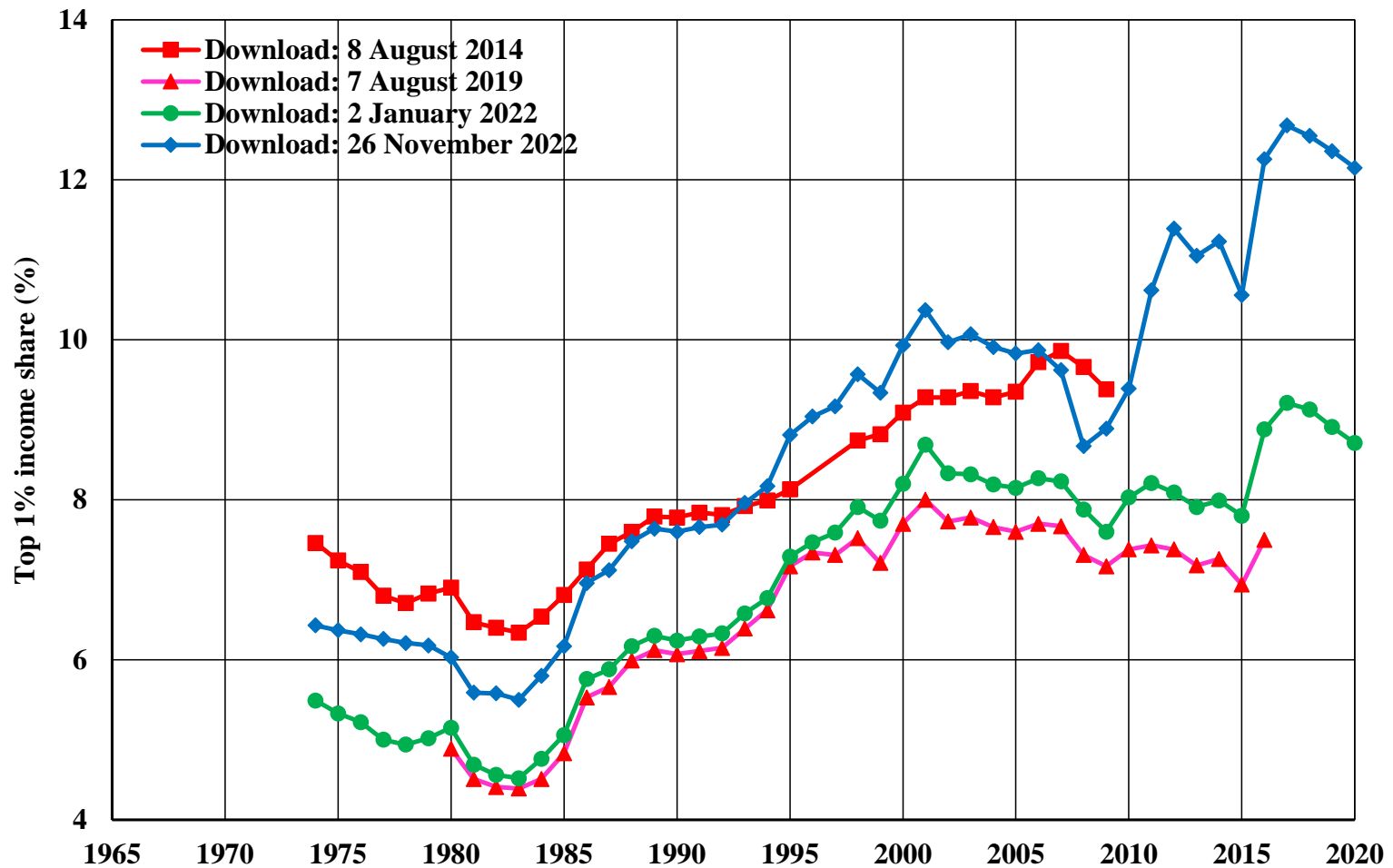
Different releases of WTID/WID



Source: websites of WTID/WID at different dates.

Pre-tax income inequality in Italy

Different releases of WTID/WID



Source: websites of WTID/WID at different dates.

Embarrassment of riches? A test

How has inequality varied in Italy since World War II?

- Can we draw a coherent story?

The answer is no. There may be good reasons for the differences. But if I had to tell how inequality has changed in Italy over the last half a century, I would be in trouble. Unless I choose one source and decide it is “the best—really the only—available” (Berg et al. 2018)

- Atkinson-Brandolini (2001) recommendation to users still holds

We are not convinced that at present it is possible to use secondary data-sets safely without some knowledge of the underlying sources, and we caution strongly against mechanical use of such data-sets

From data producers to data stewards?

- Paul Schreyer again

*The challenge for NSOs and the OECD alike is to reap the benefits of the data deluge while maintaining the level of quality that underpins trust in statistics. In this sense this new wealth of data is both an opportunity and a challenge. Where international organisations and national administrations previously acted as the principle (and often only) producer of statistics, we are now becoming **stewards** or gatekeepers at the centre of a diverse data ecosystem.*

One important responsibility of a data steward is ... to co-ordinate access to administrative data sources for statistical and research purposes, while fully respecting confidentiality standards. Such co-ordination through common classifications, inter-operable data systems and the right institutional framework also reduces the response burden on people and businesses from traditional surveys.

- A bigger challenge: guide “pure” users through multiple sources

Thanks for your attention!