

Documents without Borders

Dear International Organizations: Please Don't Delete Your Data

Jim Church

In the US government-information community, domestic issues dominate the conversation, as they arguably should. Yet at times I feel discouraged at how little traction international issues receive—the situation is just as serious, if not more so, than with the US Federal Depository Library Program. International organizations, including the United Nations, have been effectively ending depository programs and paywalling their publications, in spite of the UN Sustainable Development Goal 16, Target 10, to provide “access to information.”¹ International government data is likewise under duress. While the proliferation of online international data has resulted in tremendous research gains, unless the data is deposited in trusted repositories and subjected to best practices, international organizations may alter or delete the data for a multitude of reasons. This is in fact what has happened.

The Human Development Index

The United Nations Development Program's (UNDP) *Human Development Report* is one of the UN's great success stories. In 1990 the inaugural issue was published in collaboration with Oxford University Press, and immediately attracted attention for its critical assessment of conventional development models and their reliance on GDP growth. Its most famous metric, the *Human Development Index* (HDI), was developed by Economist Mahbub ul Haq, a colleague of Amartya Sen who won the 1998 Nobel Prize in the Economics. Sen wrote a paper on the HDI methodology which remains on the UNDP website. The premise of the index, which takes into account factors such as life expectancy, knowledge, and a decent standard of living, is one of human choice or “development as freedom.” As Sen says, “It is the lives people lead that is of intrinsic importance, not the commodities or income that they happen to possess.”²

I use this data often, and in 2015 I was disturbed to notice that some of the annual data had vanished. From 1990 to 2010 the online tabular data was only available in ten and five-year intervals, and finding the five-year data was difficult.³ Concerned, I asked some colleagues: no-one knew. At a meeting at the Academic Council of the United Nations System in 2016, I attended a panel with representatives from the UN Statistics

Division and the UNDP and asked about this. The UNDP representative replied the annual data had been removed because of changes to the methodology. As it turns out, changes to the HDI have been numerous and well-documented by academics,⁴ with some stating the index is not comparable over time as a result.⁵ In the 2016 Human Development “Reader's Guide” the UNDP admits as much and states, “the values and ranks presented in this Report are not comparable to those published in earlier editions” and refers users to the five-year tables.⁶ For a time, the only access to the annual data was through the statistical tables in the print and online *Human Development Reports*.

These methodological changes may have been important innovations. But what should concern us is the UNDP's removal and revision of globally cited data because of new methodologies. Official government data should not simply be removed or overwritten: when revisions are necessary the obsolete data should be archived as discrete data files. Any methodological changes should be clearly specified in documentation that can be easily found.⁷ Ideally for each change there should be a specific dated version, with documentation, on a single webpage or directory.

The annual HDI data has since resurfaced on the UNDP web site, it is not clear as to when the current annual data was revised, if users should consult the five-year intervals, or if the entire index is unreliable for chronicling historical development trends. The only apparent way to construct the HDI over time is to consult the data in the annual yearbooks or to search the Internet Archive for prior data files. Because the UNDP often published their data via dynamically generated databases (which cannot yet be web archived) this can be a daunting task.

The UNCTAD *World Investment Directory* and *Country Profiles*

In the 1990's the United Nations Conference on Trade and Development (UNCTAD) *World Investment Report* was the hottest international document around: it had data on Foreign Direct Investment (FDI) at a time when the growth of international capital flows were taking off. Lesser known was the UNCTAD *World Investment Directory*, a series of regional volumes with more detailed bilateral FDI data: flows of direct investment data between two countries, at times by economic sector. Altogether there are ten *World Investment Directories*, but out of these only three are now available on the UNCTAD website. I have searched for the other seven editions and cannot locate them on any archive. This is unfortunate because two of these volumes were about Asia, which was attracting the most FDI at the time.

Several years later, UNCTAD released a series of *Investment Country Profiles* with similar data. For these publications there is solid evidence of take-downs: the current site lists 24 profiles published between 2011 and 2013,⁸ while on the Internet Archive there are 124 of them.⁹ UNCTAD has apparently removed 100 out of 124 of these publications. The distinguishing feature of the remaining ones seems to be their attractive tables and color covers. To make matters more confusing, UNCTAD now publishes a series of “General” and “Maritime” country profiles. These are not the same.

The *Investment Country Profiles* were not exactly best-sellers. Typically between twenty to forty pages in length, they were mostly tables. But some of these little booklets included “FDI flows in the host economy, by geographical origin” for small developing economies. That is hard to find and of great interest to researchers working on country investment policies. Interestingly, UNCTAD now publishes a very useful series of “Bilateral FDI Statistics,”¹⁰ but users must download the data one country at a time on separate excel sheets. This data is not available on the main UNCTAD statistics portal, UNCTAD-Stat, where most users will look.

I cannot understand why UNCTAD keeps doing this. Their most interesting data disappears, only to pop up elsewhere in formats that are difficult to find and use. This especially pains me because UNCTAD presents data on topics few governments acknowledge—“creative services” for example. UNCTAD also offers us an international economic counter-culture distinguished from its more neoliberal brethren. In 2014, the IMF and other governments implemented new guidelines for FDI data based on the sixth edition of the IMF *Balance of Payments and International Investment Position Manual*. The revised guidelines categorize FDI as assets and liabilities, not inward and outward investments, showing FDI flowing in and out of countries. This may make sense for budgetary analysis, but for policy-making the “directional principle” of investment is much more interesting and useful. Thankfully, UNCTAD still uses the prior methodology. More about this later.

The International Labour Organization: Laborsta and ILOSTAT

I was recently helping a student locate gender wage data for countries around the world.¹¹ Surprisingly this is not easy to find: many countries distribute periodic labor surveys to determine wage/earnings levels, but it’s a tall order to compile these into one database facilitating cross-country comparisons. The International Labour Organization (ILO) has done an admirable job of this. But in December 2013, the ILO implemented a new statistical data system, *ILOSTAT*, replacing the historic

LABORSTA database. *ILOSTAT* is much better organized and documented, but I was puzzled because there were significant gaps in the data. Searching diligently for other sources, I finally looked at the Labor section of the UN Statistics Portal (*UN Data*). There we found the UN Statistics Division had archived much of the historic ILO data going back to the 1970s, and were able to find additional data.

This concerned me, so I wrote to the ILO to ask why there were gaps. The first thing they said was, “The data from *LABORSTA* are completely obsolete and should not be used” and they would write to the UN to check (the data is still there). They also noted, “a massive cleaning exercise was done when moving data from *LABORSTA* to *ILOSTAT* and this is why some data can be missing in *ILOSTAT* compared to the previous system.”¹²

I am sure the older data had problems. But this made me shudder. It first of all shows a serious lack of coordination between intergovernmental organization (IGO) statistical offices. Why did it take an academic librarian to notice this? Why is the current ILO data not on the UN web portal while the old data remains—evidently against the ILO’s wishes? Did the UN Statistics Division intentionally archive the data, or was this just inertia? None of us should feel good about either scenario, but if it is the latter, here’s to inertia: we could never have found the older data without it. By all means fellow librarians, when our governments decide that data is “obsolete” archive it—or urge your institutions to do so. And going forward, IGO statisticians, please don’t undertake any massive cleanings of your data without archiving, publishing and documenting the prior versions.

The IMF Balance of Payments Manual

In the entry for “Balance of Payments” in the first edition of the *Concise Library of Economics and Liberty*, economist Herbert Stein quipped “few subjects in economics have caused so much confusion—and so much groundless fear—in the past four hundred years as the thought that a country might have a deficit in its balance of payments.”¹³ This is amusing and still true. The Balance of Payments (BoP) is the record of all transactions between residents of one country with another, including direct investment abroad and international trade. Changes made to the way the BoP is calculated can dramatically alter its usefulness. In 2014, as noted above, some countries and International Organizations adopted the revised IMF guidelines for the compilation of FDI data. The IMF now calculates on an asset/liability basis instead of the directional principle (inward or outward). As UNCTAD notes in the FDI information note on UNCTADStat:

While the presentation on an asset/liability basis is appropriate for macroeconomic analysis (i.e. the impact on the balance of payments), the presentation on directional principle is more appropriate to assist policymakers and government officials to formulate investment policies. *This is because the presentation of the FDI data on directional basis reflects the direction of influence by the foreign direct investor underlying the direct investment.*¹⁴ (author's emphasis).

UNCTAD goes on to say that “the absence of information on FDI on the directional basis may even hamper policymakers from making appropriate decisions and formulating investment policies for development.”¹⁵ I am very glad, as I am sure others are as well, that UNCTAD continues to use the directional method.

The fear that powerful countries may exert a sinister political influence on their direct investment recipients is a longstanding one, at times leading to accusations of neocolonialism: I leave that debate to the pundits and professors. But what concerns me is the online wiping of historical government data due to changes to statistical methodology. You cannot go to the IMF BoP online tabular data now and find a historical table for “Brazil—Direct Investment Abroad” or “Direct Investment in China.” It is now an asset or a liability. In order to document when this happened, I consulted the print yearbooks: the change seems to have taken place in 2013, but the online data has been recalculated as far back as I can tell. If users want to access the historic data in tabular format (as opposed to PDFs) they need to use the historic IMF CDs or DVDs or a commercial service such as IHS *Global Insight*. I hope I am wrong here and would love to be so proven, but I don't think so.

Conclusion

Perhaps all this should not bother me, but it does: I hate it when online data just vanishes, or reappears with new names. It needs to stop. The best practice would be for IGOs to document and explain changes to methodologies where users are likely to first encounter the data. IGOs should never delete renowned data cited by researchers the world over: in the interest of reproducibility and transparency, the historical versions should be archived as discrete files, with dates and documentation for each version.¹⁶ A perusal of the practices undertaken by the Data verse Network, ICPSR, and other data archives, and spelled out in the Data Seal of Approval could serve as a first step to ensure that data created by international organizations¹⁷—not to mention national governments—remains both accessible and usable.

Acknowledgment

With special thanks to Bobray Bordelon, economics, finance and data services librarian at Princeton University, for his helpful edits and suggestions.

References

1. See “Sustainable Development Goal 16,” Sustainable Development Knowledge Platform, 2017, <https://sustainabledevelopment.un.org/sdg16>, target 10.
2. Amartya Sen and Sushil Annand, “Human Development Index: Methodology and Measurement” (paper written in preparation for the 1994 Human Development Report): <http://hdr.undp.org/en/content/human-development-index-methodology-and-measurement>, p. 1.
3. See “Human Development Data (1980–2015),” United Nations Human Development Programme, Internet Archive Wayback Machine, March 1, 2017, <https://web.archive.org/web/20170301010747/http://www.hdr.undp.org/en/data>. This only has five-year intervals starting from 1990, and annual after 2010.
4. Stephen Morse, “Stirring the Pot. Influence of Changes in Methodology of the Human Development Index on Reporting by the Press,” *Ecological Indicators* 45 (October 2014): 245–54, <https://doi.org/10.1016/j.ecolind.2014.04.023>.
5. Jean Claude Saha, “Comparability over Time of the Human Development Index: A Shapley Value Approach,” *Canadian Journal of Development Studies* 28, no. 3/4 (2009): 475–87, <https://www.tandfonline.com/doi/pdf/10.1080/02255189.2009.9669225>.
6. United Nations Development Programme, “HDR 2016 Reader's Guide,” <http://hdr.undp.org/en/content/hdr-2016-readers-guide>.
7. The UNDP documents methodological changes in their Human Development Reports. But these are not obvious for data users and it took an entire research article to summarize them all.
8. See “Investment Country Profiles,” United Nations Conference on Trade and Development, accessed September 25, 2017, <http://unctad.org/en/Pages/Publications/Investment-country-profiles.aspx>.
9. See “Search by Country/Economy,” United Nations Conference on Trade and Development, Internet Archive Wayback Machine, May 14, 2008, <https://web.archive.org/web/20080514004529/http://www.unctad.org/Templates/Page.asp?intItemID=3198&clang=1>.

10. See “Bilateral FDI Statistics,” United Nations Conference on Trade and Development, April 2014, <http://unctad.org/en/Pages/DIAE/FDI%20Statistics/FDI-Statistics-Bilateral.aspx>.
11. Technically “earnings” data. Earnings can include income from sources other than employment.
12. Personal email communication from the ILO Statistics office dated April 11, 2018.
13. Herbert Stein, “Balance of Payments,” in *Concise Encyclopedia of Economics*, ed. David R. Henderson, <http://www.econlib.org/library/Enc/BalanceofPayments.html>.
14. See “Foreign Direct Investment: Inward and Outward Flows and Stock, Annual,” United Nations Conference on Trade and Development, accessed September 25, 2018, <http://unctadstat.unctad.org/wds/TableView/tableView.aspx?ReportId=96740>. Click the “i” icon next to the word “annual” in the table title.
15. See “Methodological Note,” World Investment Report 2016, http://unctad.org/en/PublicationChapters/wir2016chMethodNote_en.pdf, p. 5.
16. The World Bank has done a commendable job: see “World Development Indicators,” World Bank, 2018, <https://data.worldbank.org/products/wdi>. However, not all the data files have readme.txt files.
17. See “Assessment,” Data Seal of Approval, accessed September 25, 2018, <https://www.datasealofapproval.org/en/assessment/>.