

## DATA ACCESS AND ARCHIVING POLICY AND PRINCIPLES FOR THE INTERNATIONAL COMPARISON PROGRAM

Trewin, Dennis  
Dennis Trewin Statistical Consulting  
38 Mirning Crescent  
Aranda, ACT, 2614, Australia

*(dennistrewin@grapevine.net.au)*

### **1. Introduction**

The International Comparison Program (ICP) is a worldwide statistical initiative to collect comparative price data and estimate purchasing power parities (PPPs) for individual countries or economies. Using PPPs instead of market exchange rates to convert currencies makes it possible to compare the output of economies and the welfare of their inhabitants in real terms - that is, controlling for differences in price levels across countries.

The ICP entails an extensive price data collection and detailed national accounts expenditure data compilation. The resulting ICP databases at the participating countries, the regional coordinating agencies, and the ICP Global Office contain valuable price and national accounts information, which could support the empirical research of many international, regional, and national agencies, as well as universities and research centres. However, some countries are concerned about the confidentiality of their price data. They are also concerned about the quality of the ICP data at that detailed level and criticisms they may receive if it was made available for research analysis, noting that the ICP was not designed to provide reliable data at the detailed level.

In the 2005 round, the Data Archive and Access Policy strongly limited access to data below the basic heading level data (there are nearly 200 basic heading expenditure categories). Also, in some regions, the memoranda of agreement between the regional coordinating agencies and the countries restricted researchers' access to detailed price data. However, at the end of the 2005 round, it was clear that some of the more important researchers were mostly interested in more detailed data than was available from the 2005 round notwithstanding the quality limitations.

Based on the experience of the 2005 round, the 2011 round will attempt to achieve two fundamental goals in the area of data access: (1) securing the confidentiality and security of data, and (2) ensuring effective dissemination and use of data. A trade-off arises between achieving those two goals. The 2011 round will try to strike a balance between users' needs for access to detailed micro data and participating countries concerns about this level of access including both confidentiality and quality concerns. However it should be recognized that country's laws on confidentiality cannot be breached under any circumstances.

Also, this new ICP round will provide greater focus on collecting, archiving, and providing access to metadata, to increase the quality and utility of the data collected. The limitations in the available meta data for the 2005 round made the ICP data seem like a 'black box' to the users of detailed ICP data.

One of the innovations of this round will be the development of a new methodology to link the regional results. This methodology involves putting in place a global list of products and services that will be priced by all participating countries, including the countries of the OECD/Eurostat PPP program. This will be a unique exercise that will result in a database of comparative price data of up to 170 countries for around 350 products and services. This database of average prices should be extremely valuable to researchers around the world.

Given the proposed new direction of the 2011 round in terms of responding to user's needs by providing wider access to detailed data and increasing the focus on metadata, and given other innovations of the round that will result in valuable price and national accounts information, there needs to be a policy on data access and archiving to guide practice in these areas.

## **2. Policy Objective**

The core objective of the ICP Data Access and Archiving Policy is that: *Data derived from the ICP should be utilized to the maximum extent possible for statistical and analytical purposes.* Specifically, the proposed objective is to provide users with access to detailed data beyond what was accessible through the 2005 ICP as follows:

1. PPPs, Price Level Indexes (PLIs) and Expenditure data for all countries will be published at the analytical level (ie combinations of basic headings), with supporting metadata.
2. PPPs, PLIs and Expenditure data at various levels of detail below the published level for all countries will be available to data users, with supporting metadata. The most detailed data would be at the basic heading level.
3. National Average Price data at the product level for items on the global core list for all countries - with appropriate disaggregation, measures of quality and supporting metadata – will be available to data users.
4. National Average Price data at the product level for regional items not on the global core list for all countries - with appropriate disaggregation, measures of quality and supporting metadata – will be available to data users.
5. Individual Price Observations with supporting metadata are available where permitted by the laws of individual countries as long as confidentiality is protected.

At the time of writing, there was agreement to objectives 1-3 and 5 but objective 4 was subject to discussion with participating countries.

## **3. The Core Elements of the 2005 Policy**

It might be useful to revisit the 2005 policy to provide a context. The following extract from 2005 ICP Data Access Policy is relevant.

*“Two kinds of data access will be allowed. First, the World Bank, regional, and national coordinating organizations may commission research to evaluate the methodology and suggest where improvements can be made for the next round of the ICP. If this work is done by consultants under contract to these organizations or the World Bank, they*

*should be given access to the full set of regional and global prices and quantity data as well as basic heading PPPs and expenditures providing they sign a declaration of confidentiality.*

*The second level of access comes from individuals or organizations outside the ICP framework requesting access for research purposes. This type of access should be allowed since it increases the knowledge base on the uses of PPPs and may provide valuable input for future improvements in data quality.”*

This policy required some interpretation and, with the benefit of hindsight, could have been more clearly specified. The way it was interpreted was that:

- National average price data at the product level was only available for methodology research undertaken by the World Bank, regional, and national coordinating organizations or commissioned by them;
- Other approved researchers would only have access to data at the Basic Heading level.
- Access to individual price observations was only possible with the approval of the country concerned.

#### **4. User Needs**

The two most important research needs for ICP data are to (1) support poverty analysis and (2) ICP methodological research. The latter was reasonably well supported by the 2005 policy whereas the former was not so well supported.

The basic requirement for poverty analysis is national average price data at the product level preferably with some geographic disaggregation (especially a capital city/urban/rural breakdown) and store type to better understand the prices paid by the poor and how they differ from national average prices. This implies the sample is designed to provide this level of detail which was not always the case although attempts are being made to improve the geographic distribution of the sample for the 2011 round.

Data was not available at this level of detail for the 2005 round. The users also require expenditure data at as a detailed level as possible although realistically this will not be available at a more detailed level than the Basic Heading level and even then there will be a degree of estimation for most countries.

Another important use is investigating how the ICP works from an index compilation perspective with the objective of improving understanding of results as well as improving methods. Research into alternative methods, including aggregation methods for index compilation purposes, requires access to more detailed data.

There are genuine concerns by regions and countries about the lack of quality with the detailed ICP data but the more sophisticated data users accept that the ICP was not designed to provide high quality data at this detailed level. However, they say it was still useable if there was a quality indicator (eg co-efficient of variation) and sufficient meta data to enable them interpret and understand the detailed data. Among other things, this would allow users to adjust for extreme observations.

## 5. Issues from a Country Perspective

The previous section highlight that the most important requirement from a data user perspective is for access to national average price data at the product level together with a quality indicator and supporting meta data. This requirement is also high priority for the World Bank and the Regional Co-ordinating Authorities. Access to individual price observations would also be useful for some purposes.

From a country perspective, what are the issues with providing national average price data at the product level? The stated concern from national statistical offices about providing access to more detailed data is often around confidentiality. However, the real concerns are more often around data quality confirmed by a survey conducted of countries. There are two aspects to this. First, there are concerns that the quality of data at the detailed level is not good enough to support research uses. Second, there are concerns that the countries and Regional Co-ordinating authorities might be publicly criticised because of these shortcomings. There might also be concerns that the ICP itself might be criticised because of the greater knowledge of quality weaknesses. These concerns about quality have to be recognised in the ICP policy on data access.

From a country perspective, there are also issues about providing individual price observations in about 50% of countries. Even though prices may be largely observable (although not in all cases), the statistical laws of some countries still prevent the release of this data. Other countries can only release micro data under very specific conditions which may restrict ICP uses.

## 6. Analysis of the Key Issues

As outlined in the previous Section, quality is a major issue that has to be managed in providing broader access to national average price data. Criticisms of quality are something that many official statistical agencies, that provide access to detailed data including micro data, need to currently manage. It is not a new problem. With respect to the ICP, the most important messages to be provided to the users of detailed data are:

- The ICP is of good quality for the purposes for which it was primarily designed.
- It was not designed to provide data at a very detailed level.
- However, it is recognised that data at that level might be useful for some specialist users particularly if supported by quality indicators and appropriate meta data.

Establishing realistic expectations on the quality of the data is also very important. This implies that access should be restricted to those data users who understand and accept the limitations and whose proposed uses of the data are not dependent on data of a higher quality.

With respect to individual price observations, given the variety of legal arrangements for releasing micro data, the only practical solution is for the data users to liaise direct with those countries for which they would like access to individual price observations.

## 7. Guiding Principles for ICP Data Access and Archiving

The proposed guiding principles for ICP Data Access and Archiving Principles are set out below. At the time of writing, they had not been finalised so there may be some modifications. Operational procedures will be based on those principles.

1. **Appropriate use:** data should only be made available for analytical and statistical purposes.
2. **Preservation of microdata confidentiality:** provision of data should be consistent with legal and other necessary arrangements that ensure the confidentiality of country microdata.
3. **Transparency:** the principles and procedures for access to detailed ICP data, as well as the uses of this data, should be transparent and publicly available.
4. **Consistency:** the principles and procedures for data access should strive to be consistent across all regions and countries, to promote equality in the treatment of all countries and reciprocity between countries to the maximum extent possible.
5. **Reliability:** releases of ICP data should be accompanied by appropriate metadata, including metadata that describes the quality limitations of the data.
6. **Quality limitations:** access to ICP data should be limited to users who are informed of the quality limitations and agree that the data is still useful for their purposes.
7. **Serviceability:** the ICP data should be archived to ensure it can be used to service future approved requests for access to data, it is available for possible use in future ICP rounds, and it is available as back-up in case this data is lost through disaster or other reasons by a region or a country.
8. **Disclosure limitations:** researchers accessing the detailed ICP data should ensure that their published results do not reveal directly or indirectly more detailed data than what the ICP has published without permission.
9. **Ease of access:** data access procedures should ensure a simple and expedited process for access to ICP data.

## 8. Operational Procedures – Data Users

The operational procedures have not been finalised. This section, and the following two sections, describes the procedures I proposed as part of a consultancy I undertook for the World Bank on ICP Data Access.

One operational requirement is to determine the approved data users. The following conditions are to be satisfied.

- (i) There must be a written application seeking approval which, among other things, outlines the uses to which the ICP will be put.

- (ii) The senior member of the data user group must sign an undertaking agreeing to the conditions of access.
- (iii) The conditions of access include:
  - they confirm that they understand the quality limitations and that the data is satisfactory for their use,
  - they will not publish data at a more detailed level than the analytical categories which will be the most detailed level at which ICP data is published,
  - they will provide feedback to the ICP Global Office (eg a copy of a research report) of the results from their the use of the ICP data,
  - they will not provide the ICP data to third parties, and
  - the data will only be used for analytical and statistical purposes.
- (iv) The data user is able to demonstrate that the value of the proposed use of the ICP data warrants the effort involved in supporting the data request.

### **9. Operational Procedures – Release of Data**

The analytical categories to be published cannot be prescribed at this stage. They will depend on other issues particularly data quality, both with respect to the PPPs and the National Accounts weights. Also, there should be some commonality between the Basic Headings that are combined so that the analytical category can be interpreted in a sensible way. The analytical categories published by Eurostat should provide some guidance and may provide a useful starting point for deciding what should be published for the ICP.

Data at the Basic Heading Level will only be provided to approved data users. This is similar to the 2005 policy except that it is proposed that this data be provided through the ICP web site. However, by using user name/password protection, it would only be accessible by approved users. Releasing through the web should reduce the effort involved in these type of requests. Data on the standard deviation of PPPs, available from the Dikhanov tables, should also be provided.

National Average Price data at the product level is the core data set for satisfying those important data users working on poverty analysis. There will needs to be some adjustments to provide a special treatment for confidential items (eg branded items) or items that cannot be released at this detail for other reasons. There may also be a need to exclude some countries because of quality or other considerations. For these reasons, an 'externally releasable' version of national average prices should be created to save effort and ensure consistency in releases. This will provide the data base from which approved requests might be serviced.

There is likely to be special interest in the global product list because it will facilitate analysis across the regions. If so, a version of national average prices for global list products only should be created.

To support poverty analysis, it is necessary to identify the prices paid by the poor. This will be assisted greatly if national average prices can be produced separately for certain areas (eg rural) and for selected outlet types. This would not be necessary for countries with relatively low levels of poverty or where it can be demonstrated that prices paid by

the poor do not differ greatly from those paid by others. The World Bank and the Regional Co-ordinating Authorities should agree on the countries for which the dissection is important. The precise arrangements may need to be determined on a one-on-one basis.

A quality measure, such as the standard deviation or co-efficient of variation, should also be provided to assist interpretation and allow Users to make some adjustment for extreme observations. Data on the number of observations would be useful but may have to be adjusted so that those national average prices with one or two observations cannot be identified (eg it could be said that three or less observations are involved in the relevant cases).

Although some data users would like access to individual price observations, this is prevented by law in more than 50% of countries or only available under prescribed conditions. It is not possible to develop a general policy that allows access. In reality, data users only require individual price observations for specific studies in specific countries. The number of justifiable requests is likely to be quite small. It is most sensible if the data user negotiate direct with the country concerned. The ICP Global Office or the Regional Co-ordinating authority might assist by providing appropriate contacts.

As with all uses of statistics, the meta data requirements are important. Meta data is vital for understanding the data and its quality. For example, it is important that data users understand where reference PPPs have been used rather than the PPPs being compiled on the basis of actual data. In order to keep service costs to a minimum, a single meta data document will be produced for all users. An international meta data standard for statistics should be used. The Quality Assurance Framework that is being put in place for the ICP should be the overarching framework for providing the meta data required to support these data users.

## **10. Operational Procedures - Treatment of Breaches by Data Users**

Breaches of the conditions of access should be treated seriously to discourage inappropriate practice. The consequences of breaches should be spelt out in both the Request for Data Access and the approval letter in response to requests.

If breaches do occur,

- (i) The Data User group should be asked to return the data and forbidden from future access to ICP data.
- (ii) The institution(s), to which the User Group belongs, should be advised of the breach.
- (iii) The institution itself may be forbidden from accessing non-published ICP data for a period of time (eg until a satisfactory response to the breach is obtained).

For some Data Users (eg internal World Bank users), it may be possible to take legal action if the breach is sufficiently serious in nature.

## 11. Feedback to Countries

The following quote, made in the survey of countries, captures the feeling of many countries.

*“A mechanism should be in place to keep National Statistical Offices (NSOs) informed of details of all data access requests and .....inform NSO once the results are published. In particular, the following information regarding the approved access should be provided: identity of the researcher and the institutions he/she belongs to; research proposal outlining the objectives of the research, how the ICP data will be used, research methodology, expected results, intended use of results, what data will be publishes and dissemination plan.”*

There is a need to find a way to efficiently provide feedback to countries whose data has been made available through approved ICP data access requests if their support is to be maintained.

## 12. Archival Requirements

There needs to be agreement on the archival arrangements. At the time of writing this had not happened. It has been agreed that the data should be archived but the precise arrangements are still subject to consultation. It is most important that the relevant ICP data be archived so:

- it can be used to service future approved requests for access to data,
- it is available for possible use in future ICP rounds, and
- it is available as back-up in case this data is lost for one reason or another by a region or country.

The World Bank is best placed to provide this service although Regional co-ordinating authorities may also want to retain copies of this data for their own purposes.

The data that should be archived includes:

- Basic Heading Level data (PPPs, PLIs and expenditure weights)
- National Average Prices at the product level, disaggregated where appropriate
- Quality measures derived from the Quaranta and Dikhanov tables
- Supporting meta data.

## 13. Conclusion

The ICP provides a potentially rich source of data for a number of important purposes. In the 2005 round, the emphasis was on aggregate data particularly PPPs at the national level for broad commodity groupings. This data is the main reason the ICP is conducted but the detailed data underpinning it also invaluable for a number of important uses especially poverty analysis even though the quality at this detailed level is not good. In the 2005 round these latter uses were not sufficiently well supported in the 2005 round. Access was restricted. As discussed in this paper, this will be rectified in the 2011 round. However, it is important that users understand the quality limitations in the data if they are to use it effectively.

**Abstract:** The International Comparison Program (ICP) is a worldwide statistical initiative to collect comparative price data and estimate purchasing power parities (PPPs) of the world's economies. Using PPPs instead of market exchange rates to convert currencies makes it possible to compare the output of economies and the welfare of their inhabitants in real terms - that is, controlling for differences in price levels.

The 2011 ICP is underway and will be the largest ICP round to date, covering up to 170 economies. The ICP entails large price data collection and detailed national accounts expenditure data compilation. The resulting ICP databases contain valuable price and national accounts information, which could support the empirical research of many international, regional, and national agencies, as well as universities and research centres.

In the 2005 round, the Data Archive and Access Policy strongly limited the researchers' access to detailed data mostly as a result of country's confidentiality concerns, although in reality the main concern was data quality. Based on the experience of the 2005 round, the 2011 round will attempt to provide access to more detailed data whilst striking an appropriate balance between participating countries' confidentiality requirements and the users' needs for access to detailed data. A strategy policy on data access for the 2011 round has been agreed and the key features of this policy are outlined in this paper.