Bank Procedure

Development Dataset Acquisition, Archiving and Dissemination

Bank Access to Information Policy Designation
Public

Catalogue Number
4.05-PROC.125

Issued
January 22, 2020

Effective
January 22, 2020

Content
Procedural instructions to Bank Staff on the archiving and dissemination of development datasets and guidance on acquisition.

Applicable to
IBRD, IDA

Issuer
Managing Director, Operations, MDOPS

Sponsor
Director, DECDG
SECTION I – PURPOSE AND APPLICATION

1. This Procedure provides instructions for archiving and dissemination of Development Datasets, as long as those Datasets do not contain Personal Data.

2. This Procedure applies to the Bank.

SECTION II – DEFINITIONS

As used in this Procedure, the capitalized terms and acronyms have the meanings set out below.

1. Corporate Data: means information about the business and business functions of the Bank, including data on the Bank’s budget, human resources, corporate finance as well as lending and trust-funded operations and advisory services and analytics.

2. Data: means the physical representation of information in a manner suitable for communication, interpretation, or processing by human beings or by automatic means. It provides characteristics or information, usually numerical, that are often collected through observation.

3. Databank: means an analysis and visualization tool that contains collections of time series Data on a variety of topics.

4. Data License Agreement or Agreement: means a written document legally binding agreement, between two or more parties, which specifies the rights and obligations of each party governing the access, acquisition, archiving, delivery, dissemination, sharing, and use of Development Data, without any exchange of payment.

5. Data Provider or a Dataset Provider: means an organization that provides Data and/or Metadata. This includes: government entities, international organizations, non-governmental organizations, universities, commercial vendors, etc. It does not include individuals directly providing their own personal Data or respondents in a Data collection exercise.

6. Dataset: means any organized collection of Data that does not contain Personal Data. In this Procedure, the terms ‘Dataset’ and ‘Development Dataset’ are used interchangeably.

7. Dataset Acquisition: means a Dataset is delivered by electronic or other means (hard copy or CD, USB drive) and Staff are in possession and/or have access to the Dataset, in paper form, or on their desktops/network drive or Cloud storage. Datasets can be acquired from a Data Provider for a fee or for no fee.

8. Derivative Works: means works based on or derived from one or more already existing works. Derivative works include derived data and analytical products, including but not limited to: research papers, analytical studies, data visualizations, derived Indicators, aggregated and/or derived databases, and other outputs (e.g. publications, CDs, mobile device apps, blogs, online data products, etc.) created using the Dataset(s) and metadata in question.
9. **Development Data**: means data about countries that does not contain Personal Data and that can be used for reference or analysis in the process of development, typically in sectors such as the economy and finance, poverty, education, health, public administration, private sector development, agriculture, land use, gender, climate change, environment, infrastructure, trade and others.

10. **Development Dataset**: means any organized collection of Development Data that does not contain Personal Data, but excludes Bank’s Corporate Data, Datasets acquired outside the provisions of this Procedure, e.g. produced or collected during collaborative ‘hackathon’ events, individual observations, or small-scale Data that do not form a Dataset (e.g. Data received by Staff from Data Providers as single or multiple Data points).

11. The **Development Data Council** provides governance and coordination for the World Bank’s vision, priorities and activities related to development data. The Council has three broad strategies: first, increase country capacity to produce, disseminate, and use data of all kinds; second, improve the Bank’s internal capacity to access, manage, and disseminate data; and third, produce high quality statistical indicators and data products to help improve development outcomes.

12. **Development Data Hub (DDH)**: means the centralized dataset and metadata repository mandated by the Development Data Council of the World Bank. DDH comprises a suite of Dataset repositories – Microdata Library, Databank – that help Staff discover, manage and use Development Datasets held across the Bank, and also Datasets that are not in the physical possession of the Bank. It provides policies, services, a searchable central Dataset catalog, and other software tools for Dataset curation, storage, access and use, and dissemination.

13. **License or License Agreement**: means a legally binding agreement between two parties, typically granting rights from a licensor to a licensee and which, in exchange, the licensee usually submits to a series of conditions regarding the use of the licensor's property, which may include payment.

14. **Metadata**: means ‘data about data’. Metadata describes data. It helps to understand the meaning of data, or provides useful information about its provenance or licensing status.

15. **Microdata**: means unit-level data obtained from sample surveys, censuses, and administrative systems. It provides information about characteristics of individual people or entities such as households, business enterprises, facilities, farms or even geographical areas such as villages or towns.

16. **Microdata Library**: means a searchable Bank catalog and repository of survey Microdata and documentation. It facilitates access to Microdata that provides information about people living in developing countries, their institutions, their environment, their communities and the operation of their economies. It includes Datasets from other international organizations, statistical agencies and other agencies in low and middle-income countries.

17. Mixed Dataset is a Dataset which includes Development Data and Personal Data.

18. **Open Data**: means any Data that can be freely accessed, used, modified and shared by anyone for any purpose, which is both legally open (available under a public open data
license) and technically open (available in machine-readable and bulk form). If data is to be released as Open Data, it does not contain Personal Data.

19. Open Data License: means any legally binding instrument, offered to all potential users, which grants permission to access, re-use, and redistribute a dataset with few or no restrictions, and provides consistent and broadly recognized terms of use. Examples include ‘Creative Commons’ Licenses, as well as the Open Data Commons Open Database License. See Public Licenses Procedure.

20. Open File Format: means a format that is platform independent, machine readable, and made available to the public without any restrictions that would impede the re-use of that dataset.

21. Personal Data as defined in the World Bank Group Policy Personal Data Privacy, Section II, paragraph 3.

22. Processing as defined in the World Bank Group Policy Personal Data Privacy, Section II, paragraph 5.

23. Sensitive Personal Data is a subcategory of Personal Data relating to the core of a data subject’s private sphere such as Personal Data concerning health, sexual orientation, political orientation and fiscal identity, which, in case of unauthorized Processing or accidental loss, destruction or damage, results in a high risk to the data subject’s privacy.

24. Staff: means persons in the Bank holding an appointment under Staff Rule 4.01, “Appointment”.

25. Terms-of-Use, also known as terms of service or terms and conditions, means rules or conditions to which a user agrees in order to use a service or information.

26. TTL: Task Team Leader.


28. World Bank Group: means the Bank, together with the International Finance Corporation (IFC), Multilateral Investment Guarantee Agency (MIGA), and International Centre for Settlement of Investment Disputes (ICSID).

SECTION III – SCOPE

1. This Procedure applies to all Development Datasets to which the Bank has access, from the effective date of the Procedure unless such datasets are subject to other pre-existing arrangements. This Procedure applies to Development Datasets acquired by the Bank including those from countries under any Bank funded loan/credit/grant, including Implementation Status and Results (ISR) Data or any Bank-executed ASA or Bank-executed Trust Fund, where consent is obtained at the point of data collection to provide the Data to the Bank. If the dataset includes Personal Data, please contact the legalhelpdesk and the Data Privacy Office.
2. This Procedure does not apply to the Bank’s Corporate Data, Datasets acquired outside the provisions of this Procedure, e.g. produced or collected during collaborative ‘hackathon’ events, individual observations, or small-scale Data that do not form a Dataset (e.g. Data received by Staff from Data Providers as single or multiple Data points).

3. Except as otherwise prescribed herein, this Procedure covers two processes: 1) Archiving of the Dataset, and (2) Dissemination of the Dataset - making the Dataset available to internal and external users.

**Dataset Acquisition**

4. This Procedure does not cover the manner of acquisition of Datasets. [See section XI: Related Documents, for guidance on acquisition.]

**Dataset Archiving**

5. Staff deposit Development Datasets and their Metadata into the DDH, following the steps provided on the [DDH Data Deposit Site](#) and the following provisions:

a. During the deposit process, Staff determine whether the content of the Development Dataset is public or restricted from public access pursuant to the Bank Policy: Access to Information. When the Development Dataset is restricted from public access, Staff assign a security classification pursuant to [WBG Procedure: Information Classification and Control Policy (AMS 6.21A)](#) : Official Use Only, Confidential, or Strictly Confidential.

b. A Development Dataset is deposited no later than six months after it is acquired, with the provision to also later deposit any revisions or updates. Once deposited, at the discretion of the depositing Staff, the TTL decides if the Dataset is embargoed from dissemination pending the publication or release of any Derivative Work(s). Benefits of early deposit include preventing accidental data loss and inadvertent/inappropriate data release in the case of restricted data.

c. A Development Dataset which is part of a database system (one or more databases) or a web portal that is owned, managed or commissioned by the Bank does not need to be deposited in the DDH. In such cases, it is sufficient to provide to the DDH a link to access the database system/s, a web or data service that serves to virtualize the data in the database systems, or the portal. However, the associated Metadata for such a Dataset is still deposited in the DDH.

d. Where feasible, a Development Dataset being archived in the DDH is provided in an Open File Format to facilitate its re-use.

**Supporting Documentation**

e. Staff deposits either an electronic or a scanned copy of: the email, fax, hard-copy, the applicable Terms of Use, License, signed Data License Agreement, or URL links to those documents into the DDH along with the Development Dataset being deposited. These documents are deposited no later than six months after the Dataset’s Acquisition. Signed Agreements are filed with LEG in WBG official records management system, following instructions provided on the [DDH website](#).
f. Other documentation or URL links to them, may include: terms-of-reference for a Development Dataset; Licenses and/or subscription contracts for a Dataset that is subscribed to or purchased. Such other documentation or URL links are deposited no later than six months after the Dataset’s acquisition.

Dataset Dissemination

6. Subject to the security classification set forth in the WBG Procedure: Information Classification and Control Policy (AMS 6.21A) assigned and the terms and conditions of any License applicable to a Development Dataset, the TTL may be able to share the dataset with Staff and external parties as follows:

a. Public Datasets: A Development Dataset classified as “Public” is shared with Staff and the public provided:
   - there are no restrictions by the Dataset Provider on re-use of the Dataset, under the terms of a public License, consistent with the requirements of the Bank Procedure: Public Licenses; or
   - dissemination is consistent with the Terms-of-Use/License/Data License Agreement with the Dataset Provider.

b. Restricted Datasets: A Development Dataset classified by the Bank as “Official Use Only”, “Confidential”, or “Strictly Confidential” pursuant to the WBG Procedure: Information Classification and Control Policy (AMS 6.21A) is shared with Staff and/or External Parties as defined under AMS 6.21A, only as permitted in the Terms of Use/License/Data License Agreement with the Dataset Provider, and according to all applicable Bank policies, directives, procedures, including, but not limited to WBG Procedure: Information Classification and Control Policy (AMS 6.21A).

Key Roles and Responsibilities

7. DECDG is responsible for providing guidance to Staff on:
   a. The process of depositing Development Datasets, Metadata and other documentation in the DDH;
   b. Templates for Terms-of-Reference for hiring Consulting Firms to collect Development Datasets.

8. The Legal Vice Presidency is responsible for all legal agreements including the Data License Agreement template and any deviations. Staff shall use the standard Data License Agreement template prepared by the Legal Vice Presidency and available online through the proper internal electronic platforms. The Legal Vice Presidency clears any deviations made to the standard Data License Agreement template and other forms of agreements. The legal agreements are signed consistent with the Procedure: Signature Authority. For legal advice, staff may contact the Legal Help Desk at legalhelpdesk@worldbank.org.

9. The Development Data Council is responsible for resolving issues over deposit and dissemination of Development Datasets.

Staff Responsibilities
10. Staff are responsible for:

a. Depositing the Development Dataset, its Metadata and other documentation into the DDH in accordance with this Procedure;

b. Profiling the data according to the Bank’s Access to Information Classification and Control Policy (AMS 6.21A) throughout all stages of the data life-cycle irrespective of where the data is managed and disseminated. Staff refer to the Data License Agreement together with the Annex: Data Sharing and Use Protocol, Public Information, and WBG Procedure: Information Classification and Control Policy (AMS 6.21A) for details on how to profile the data.

c. Complying with the terms of the Data License Agreement/ License/ Terms-of-Use entered into with the Dataset Provider; and

d. Any other storage and document management requirements governing Data.

SECTION IV – EXCEPTION

None

SECTION V – WAIVER

The Issuer of this Procedure, with the advice of the Director, Development Data Group (DECDG), Development Economics, may waive any provision of this Procedure.

SECTION VI – OTHER PROVISIONS

None

SECTION VII – TEMPORARY PROVISIONS

None

SECTION VIII – EFFECTIVE DATE

This Procedure is effective as of the date on its cover page.

SECTION IX – ISSUER

The Issuer of this Procedure is the World Bank Managing Director, Operations.

SECTION X – SPONSOR

The Sponsor of this Procedure is the Director, Development Data Group (DECDG), Development Economics.
SECTION XI – RELATED DOCUMENTS

WBG Procedure: Information Classification and Control Policy


Bank Guidance: Development Dataset Acquisition and Archiving, May 18, 2018. Catalogue Number CEO4.05-GUID.108


Data License Agreement Template

Summary of Open Data Terms of Use

Terms of Reference Template for a Contracted Firm for Development Data Collection


World Bank Group Policy - Personal Data Privacy, effective May 25, 2018, Catalog No. SEC4.05-POL.101

Questions regarding this Procedure should be addressed to the Sponsor.