

Project Report:

**Democratic People's Republic of Korea
Economic Statistics Project**
(April –December 2008)

March 2009

By

Mika Marumoto, Ph.D.
marumotomika@hotmail.com

Table of Contents

Table of Contents	2
List of Tables and Figures	3
List of Abbreviations.....	5
Executive Summary	6
Chapter I. Introduction	10
A. Project Background and Objectives	10
B. Expected Output and Beneficiaries	13
C. Methodology	13
Part One: Overall Assessment of North Korea Economic Statistics	
Chapter II. Availability and Characteristics of DPRK Statistics and Data.....	17
A. Geographical Coverage and Comprehensiveness	17
B. Language, Accessibility and References.....	18
C. Types of Datasets	21
Chapter III. Overall Assessment of DPRK Databases	24
A. Overall Assessment: Criteria and Limitations	24
B. Major Findings	25
Part Two: Comparative Assessment of North Korea Datasets - Case Studies	
Chapter IV. Analysis of Major Economic and Social Indicators	31
A. Analytical Focus: Selected Entities, Indicators and Rationale ...	31
B. Population and Health Indicators	32
C. Microeconomic Indicators - Prices and Exchange Rates	37
D. Macroeconomic Indicators - GDP and GDP Per capita	42
Chapter V. Analysis of Trade and Investment Statistics	57
A. Trade Statistics	57
B. Investment Statistics	81
Part Three: Conclusions and Recommendations	
Chapter VI. Toward Knowledge Sharing	84
A. Major Findings and Lessons Learned	84
B. Recommendations: Operationalization of the DPRK Statistics Databank	89
C. Concluding Remarks	93
Appendix * Project Implementation Schedule	94
*See separate excel files for Appendix: Data Comparison	
References	95

List of Tables and Figures

Tables:	Page
1: North Korea Statistics Data Availability and Comprehensiveness	18
2: North Korea Statistics Data Availability by Language	19
3: Key Characteristics of Four Categories of North Korea Data	21
4: Criteria for Data/Database Assessment	24
5: Distribution of Entities in terms of Assessment Points	26
6: North Korea's Crude Death and Birth Rates (per 1,000 people) 1995-2007	35
7: Population Dynamics: North Korea and the World	36
8: Share of Products with Prices Determined by Market (%)	40
9: Estimates of Private Sector's Contribution to GDP	41
10: Purchasing Power Parity (PPP)-based GNI	44
11: GDP Per capita Growth Rates: BOK vs. UN	46
12: PWT6.2 Indicators and North Korea Data (Selected Years)	54
13: PWT6.2 PPP-based GDP Per Capita: North Korea and Selected Countries 2003.....	54
14: Overall Evaluation of Macro Data	56
15: North Korea's Trade: Aggregate Level Comparison	59
16: Coverage of DPRK Trading Partners (Number of Countries or Regions)	59
17: Discrepancies in DPRK Trade Figures: KOTRA vs. GTIS (2007)	61
18: Discrepancies in DPRK Trade Figures: KOTRA vs. IMF (2006)	62
19: UNCTAD-DPRK's Trade with Developed and Developing Countries	64
20: IMF: DPRK's Export Trend by Trading Region (1998-2006)	65
21: IMF:DPRK's Import Trend by Trading Region (1998-2006)	65
22: GTIS Global Trade Atlas Output Example – Unit Cost Analysis	71
23: ITC Market Analyses Tool Output Example (1): Trade Performance Index - DPRK's Minerals Sector and Competitiveness in the World Market	73
24: ITC Market Analysis Tool Output Example (3): The Balassa Index – Revealed Comparative Advantage	75
25: Millennium Development Goal 8.6: Market Access Indicator (1) by UNCTAD-WTO-ITC	78
26: Millennium Development Goal 8.6: Market Access Indicator (2) by UNCTAD-WTO-ITC	89
27: Overall Assessment of Trade Data Usability	80
28: Comparison of FDI Inflow: China, DPRK and Vietnam	81
29: Expected Database Characteristics and Examples of Data Sources	90

Figures:

1: DPRK Statistics – Reverse Pyramid Structure of Data Sources	20
2: Overall Assessment of Databases Identified	25
3: DPRK Population Data by Entities	33
4: US Census Bureau-IDB Population Pyramids for North Korea: 1993-2007.....	35
5: ITC Market Analysis Tool Output Example (2): Trade Performance Index 2005.....	74
6: ITC Market Analysis Tool – Export Performance Dynamics Perspectives Charts 2004-2006.....	77
7: FDI Inflow Per Capita (flow and stock): DPRK, China and Vietnam...	82
8: FDI Inflow (flow and stock): DPRK vs. Vietnam.....	83

List of Abbreviations

AMA	Analysis of Main Aggregate (database, by UN)
BOK	Bank of Korea
CIA	Central Intelligence Agency (of US)
CIC	Center for International Comparisons (University of Pennsylvania)
DPRK	Democratic People's Republic of Korea
ESDS	Economic and Social Data Services
ERINA	Economic Research Institute of Northeast Asia
EU	European Union
FAO	Food Agriculture Organization
GDP	Gross Domestic Products
GNI	Gross National Income
GTIS	Global Trade Information Service
IDB	International Data Base (US Census Bureau)
ITC	International Trade Center (UNCTAD/WTO)
KCS	Korea Customs Service
KDI School	Korea Development Institute School of Public Policy and Management
KIEP	Korea Institute for International Economic Policy
KINU	Korea Institute for National Unification
KITA	Korea International Trade Association
KOSIS	Korea Statistics Information Service (of NSO)
KOTRA	Korea Trade-Promotion Agency
IFI	International Financial Institution
IMF	International Monetary Fund
MDG	Millennium Development Goals (UN)
MOC	Ministry of Commerce (of China)
MOU	Ministry of Unification (of Korea)
NSO	National Statistical Office (of Korea)
OECD	Organization for Economic Cooperation and Development
PPP	Purchasing Power Parity
PWT	Penn World Table
ROK	Republic of Korea
SIPRI	Stockholm International Institute for Strategic Studies
TPI	Trade Performance Index
UNCTAD	United Nations Conference on Trade and Development
UNFPA	United Nations Population Fund
UNICP	United Nations International Comparisons Programme
UNICEF	United Nations Children's Fund
UNSD	United Nations Statistics Division
UNPD	United Nations Population Division
WDI	World Development Indicators (World Bank)
WHO	World Health Organization
WTO	World Trade Organization

PART ONE: OVERALL ASSESSMENT OF NORTH KOREA ECONOMIC STATISTICS

II. Availability and Characteristics of DPRK Statistics and Data

While identifying statistical sources and data, the Project studied four distinct and separate categories: (i) socio-economic indicators; (ii) microeconomic data such as prices; (iii) macroeconomic data such as GDP and GNI per capita; and (iv) trade data. As noted above, the first stage of the Project derived criteria for categorization of databases. Such criteria included geographical coverage and comprehensiveness (time series or single-year data), language (available in English or only available in Korean), and the types of datasets. To the extent possible, the frequency of data updating (regular or irregular) was also confirmed.

A. Geographical Coverage and Comprehensiveness

The Project covered and reviewed 162 entities drawing upon 221 data sources. Many entities have multiple databases or publications. The Project counted them as separate data sources in most cases.

Table 1 shows the geographical coverage of the databases covered and reviewed, and the availability of DPRK statistics and the degree of comprehensiveness. “Comprehensive” data is defined as any sub-category of data that has contains statistics covering a period of over ten years, thereby enabling analysts to conduct meaningful time-series analyses.

In terms of geographical coverage, the Project covered entities from the Republic of Korea (ROK), the United States (US), European Union (EU) member states, Japan (J), China (C), Hong-Kong (HK), other Asia Pacific Countries (APAC) such as Australia and New Zealand, as well as international organizations (Global). These entities were further categorized into sub-groups, namely, government entities (G), private institutions (P), or NGOs. Entities covered include a large number of ROK government institutions (27), followed in number by global entities such as UN organizations (21), ROK NGOs (17), US private research institutes (16), and ROK private institutions (15). Looking at individual databases rather than the producing entities, North Korean economic statistics are available in many databases compiled by international organizations (41), followed by databases issued by the ROK government (37), ROK NGOs (21) and the US government (21).

During the course of identification and collection of data, the first obstacle encountered was access to North Korean official statistics. Some ROK entities' websites containing North Korean data nominally in the public domain were only accessible by Korean nationals, or by the designated staff members of these entities.

Out of 221 data sources covered, about 64% (141 databases) have North Korea-related economic and social statistics data; of 141 databases, about 40% (56 databases) have comprehensive data, while 56% (79) have only partial data.

Table 1: North Korea Statistics Data Availability and Comprehensiveness

	Number		DPRK Data (Y/N)			Among Yes		
	Entities	Databases	Yes	No	Not known	Comprehensive	Not Comprehensive	Not known
APAC-G	2	3	3	0	0	0	3	0
APAC-P	1	1	1	0	0	0	1	0
EU-G	9	9	4	4	1	1	2	1
EU-P	7	15	10	5	0	5	4	1
EU-NGO	1	1	1	0	0	0	0	1
Global	21	41	32	8	1	14	18	0
HK-NGO	1	1	0	1	0	0	0	0
J-G	4	4	2	2	0	2	0	0
J-P	8	11	4	3	4	3	2	0
NK-G	5	5	2	1	2	1	0	1
NK-NGO	1	1	0	1	0	0	0	0
PRC-G	7	9	4	3	2	3	1	0
PRC-P	2	2	2	0	0	0	2	0
ROK-G	27	37	24	9	4	14	10	0
ROK-NGO	17	21	8	13	0	0	9	0
ROK-P	15	17	11	4	2	2	5	3
US-G	13	21	17	4	0	4	13	0
US-NGO	4	5	5	0	0	0	5	0
US-P	17	17	11	6	0	7	4	0
Total	162	221	141	64	16	56	79	7
Share (%)	-	100%	64%	29%	7%	40%	56%	5%

B. Language, Accessibility and References

Table 2 shows that out of 141 databases identified as containing North Korean statistical data, 85 data sources or about 60% are available only in English while about 18% is available only in Korean. About 3.5% is available in languages other than English or Korean. The accessibility of DPRK statistics available only in English is considerably higher than that of sources available only in the Korean language. Three observations are noteworthy. Given the lack of primary sources, the same DPRK statistical information from a relatively limited number of sources in English is circulated and recycled among these databases. Second, there are some English databases that are under-utilized by South Korean analysts. Third, although the databases available only in Korean represent a relatively small proportion of the total, some still offer high usability. Such databases

are under-utilized by non-Korean analysts. For example, it is highly likely that the ROK National Statistical Office's Korea Statistics Information Service (KOSIS) database has been under-utilized by non-Korean analysts despite its comprehensiveness and user-friendly functions. The KOSIS offers, under its domestic statistics section, a tab called South-North Korea Economic and Social Indicators Comparison, containing 13 files of 86 comprehensive data categories for the period of 1965/70 to 2006.¹¹

Table 2: North Korea Statistics Data Availability by Language

	Language						Accessibility	
	English only	English and Korean	English and Other language	Korean only	Others only	Sub-total	Open	Restricted (of which proprietary)
APAC-G	3	0	0	0	0	3	3	0
APAC-P	1	0	0	0	0	1	1	0
EU-G	1	0	2	0	1	4	4	0
EU-P	10	0	0	0	0	10	1	13 (1)
EU-NGO	0	0	1	0	0	1	0	0
Global	31	0	1	0	0	32	24	8 (2)
HK-NGO	0	0	0	0	0	0	0	0
J-G	0	0	2	0	0	2	2	0
J-P	0	0	2	0	3	5	3	2 (0)
NK-G	0	0	0	2	0	2	0	0
NK-NGO	0	0	0	0	0	0	0	0
PRC-G	0	0	4	0	0	4	2	2 (0)
PRC-P	1	0	0	0	1	2	0	0
ROK-G	6	3	1	15	0	25	0	5 (5)
ROK-NGO	1	6	0	2	0	9	0	0
ROK-P	0	3	0	6	0	9	0	1 (1)
US-G	17	0	0	0	0	17	0	0
US-NGO	5	0	0	0	0	5	0	0
US-P	9	0	2	0	0	10	2	4 (2)
Total	85	12	15	25	5	141	105	35 (11)
Share (%)	59.9%	8.5%	10.6%	17.7%	3.5%	100%	75%	26%

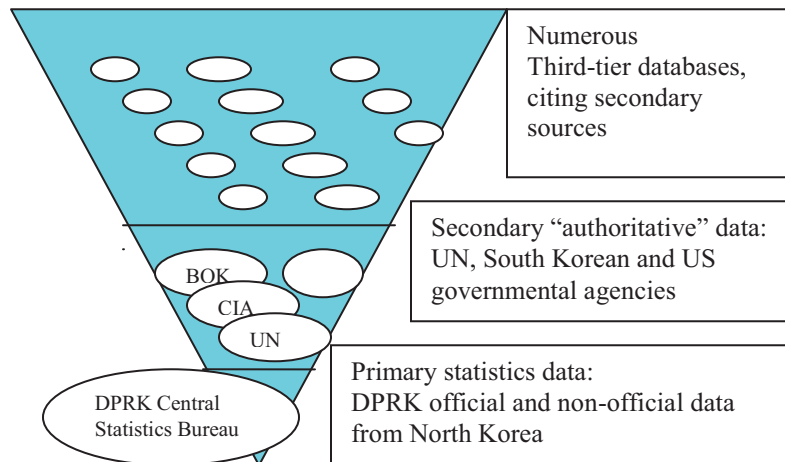
Aside from language barriers, other types of accessibility issues also became clear. So-called “proprietary” issues, defined here as obstacles to data access due to costs (fees), or other non-financial restrictions such as rules, regulations and procedures, are particularly relevant. For example, during the course of data identification and collection, it became clear that only South Korean nationals with national identification numbers can log in to become a member of certain ROK entities' websites. Also, even among South Korean

¹¹ National Statistical Office: Korea Statistics Information Service (KOSIS): <http://www.kosis.kr>.

nationals, there are those who can and cannot gain access to North Korean data and statistics made available for the public domain. For instance, the website of the Korea Institute for National Unification (KINU) discloses the list of North Korea statistics announced and released by the North Korean authorities that have been submitted to international organizations. During field visits and interviews, KINU advised that the only way to access such data on their website is to physically visit the KINU office in Seoul.¹²

Perhaps most significantly, the Project also confirmed a “reverse pyramid” structure of data sources as determined by the references noted in the various databases. Some 22% of databases refer to primary sources, while 25% are considered to be authoritative entities reporting DPRK-related statistics. The remaining 50% are entities which merely cite and repeat data provided by the secondary authoritative sources. In other words, there are only a handful of institutions which rely on primary data sources provided by North Korean authorities (i.e., the United Nations), creating a fragile foundation for DPRK data and statistics sources. There is also a relatively limited number of frequently cited “authoritative sources” including the Bank of Korea (BOK), Korea Trade-Promotion Agency (KOTRA), and the United States Central Intelligence Agency (CIA). The majority of institutes that possess and disclose DPRK statistics on their websites or publications simply cite these authoritative sources, in most cases without adding their own analysis. It should also be noted that this “reverse pyramid” structure for DPRK

Figure 1: DPRK Statistics – Reverse Pyramid Structure of Data Sources



data sources does take into account those individuals and institutions which are known to possess DPRK statistics but use them only internally, or on a strictly proprietary basis. Most of these entities seem to be reluctant to share such data with others in part to protect exclusive relationships they may have established with data providers in North Korea.

¹² Author’s interviews with a ROK government official and a former KINU official. Seoul. August 2008. <http://www.kinu.or.kr>.

C. Types of Datasets

Table 3 shows major findings regarding the characteristics of North Korea-related data, explained for each of the four data categories. There are distinct characteristics or features of the available North Korea data which become clearer when considered for each of the four types, namely, macroeconomic indicators, microeconomic indicators, social indicators and trade data. These differing characteristics imply that different approaches are necessary to conduct in-depth analysis on entities and their datasets in Part II. Also, there are implications for the operationalization of a data bank beyond the current phase of the Project. Each category of data could require a different method for data bank operationalization.

Table 3: Key Characteristics of Four Categories of North Korea Data

	Findings	Implications for needed follow-up
Macro-Data	<ul style="list-style-type: none"> * Many entities refer to a few sources and analyses (i.e., BOK, CIA, KOSIS) * Methodologies not transparently explained in the public domain 	<ul style="list-style-type: none"> * In-depth comparative analyses among entities and methodologies * Interviews * Recommendations to data-producers
Micro-Data	<ul style="list-style-type: none"> * Lack of data and reliability issues * Major original sources are either from surveys conducted by international organizations and individuals (connected or contracted by outside organizations), or observations by entities/individuals in NK (i.e. foreign embassies in PY) 	<ul style="list-style-type: none"> * Collection/compilation of micro data will continue to be a challenge * A data depository system as a model for compiling micro-data in the future
Trade Data	<ul style="list-style-type: none"> * Comprehensive data is available from multiple sources * Specialized entities providing data on a subscription basis. * Technical notes/methodologies, value-added analyses are more available, compared to other categories 	<ul style="list-style-type: none"> * Comparative analyses among different sources (i.e., IMF DOT vs. KOTRA/KITA) * Recommendations to data-producers * Technical analysis (TA) can be a “user guide” for DPRK trade data users
Socio-econ data	<ul style="list-style-type: none"> * Specialized entities/agencies are cited as original sources (i.e., health indicators, TB rate) * Existing data is under-utilized 	<ul style="list-style-type: none"> * Compilation of various data, categorization, and analytical focus need to be determined for the Project. * Link to macro-data analyses

Analyses of the databases covered indicate that for **macroeconomic data** such as GDP and GNI many institutes rely on only a few entities, namely the BOK and the CIA. This finding is not new; it confirms the already-known fact that those are the most frequently quoted sources. Why are they the most frequently cited? Aside from the BOK and CIA,

are there any other institutions that announce North Korea's macro data on a regular basis?

Regarding the first question, the frequent citation of the BOK and the CIA statistics can be attributed to the fact that both are governmental institutions which are believed to have direct access to DPRK data. In addition, their data is easily accessible through websites and publications, and the data has been regularly updated over the past decade or so. As a result, these two institutions are viewed as "authoritative" sources of North Korean economic statistics, especially in South Korea and the United States. The BOK and the CIA are almost exclusively relied upon as the original sources of North Korea GDP data. But until their underlying purposes, rationale and methodologies are better understood, gaps in the GDP data they publish will continue to be a source of contention. The problem becomes even more complicated when third-tier entities cite GDP figures from the CIA and the BOK without much consideration to their comparability with other sources of GDP data for other different countries.

As regards the second question of whether other entities also announce or estimate North Korean GDP, information based on North Korean authorities' periodic but rather inconsistent announcements of data, particularly data from the 1990s, are available from some sources including the Ministry of Unification's North Korea Information Center.¹³ The publication called *UN Data* also makes available North Korean GDP data through multiple channels, including the UN's Analysis of Main Aggregates (AMA) database. On the other hand, neither the World Bank nor the International Monetary Fund (IMF) includes North Korean macroeconomic data in most of their databases such as the World Development Indicators (WDI) online or the International Financial Statistics. This is mainly because North Korea is not yet a member country of these International Financial Institutions (IFIs). Most recently, the Center for International Comparisons at the University of Pennsylvania, in its latest Penn World Table (PWT6.2), has started to estimate North Korea's PPP-based GDP data starting in 1970 using various methods.

In sum, the circular nature of DPRK data is particularly noticeable in the case of macroeconomic data -- the "reverse pyramid" structure of economic data sources is especially prominent in the macroeconomic data category. Part II of this report directly addresses these issues, by conducting in-depth comparative analyses of some of the above-mentioned entities.

As expected, the top problem identified for **microeconomic data** is its overall absence. In addition, North Korean microeconomic data has serious problems with reliability even when it is available. Price data is fundamental to estimation of national accounts. Such data is especially important given that North Korea is in *de facto* transition – slowly, but clearly nonetheless – from a planned economy to a market economy. Within the limited microeconomic datasets that are available, a majority of data sources are surveys conducted by individual researchers, NGOs or foreign embassies in Pyongyang. These institutions often rely on individual (untrained) North Korean citizens to provide the

¹³ <http://unibook.unikorea.go.kr/new2/>

information. As a result, systematic collection and compilation of microeconomic data will continue to be a tremendous challenge.

Regarding **socio-economic data**, there are a number of specialized agencies producing unique and useful information related to health and social sectors. Such unique databases vary from ones produced by global entities such as the United Nations, the World Health Organization (WHO) and UNICEF (i.e., UN-Millennium Indicators Database offering DPRK historical data in the limited categories of socio-economic indicators) to specialized South Korean entities including the Korean National Tuberculosis Association and the Inter-Korean Summit Secretariat. The Inter-Korea Summit Secretariat provides statistics on inter-Korean dialogue sub-meetings by theme, such as politics, economy, culture and military. Due to the fragmented nature of socio-economic data, compilation of statistics from various sources and classification is a challenge that requires prioritization. Part II of this report will take up population and some health-related datasets for in-depth analysis.

The nature of issues revolving North Korea **trade data** and available trade databases is quite different from those of other data categories. Comprehensive databases of North Korea's external trade (for periods over 10 years) are available from multiple sources. The challenges faced by DPRK analysts in using and assessing trade data are therefore different. For example, why do the available trade databases show different aggregate figures that differ beyond what could be expected conceivable errors in data derived from mirror statistics? Which specific data sources should we use for which specific analytical purposes? What are major differences among the trade databases in terms of methodologies and classifications, as well as the entities' decisions in modifying reported official data when compiling mirror statistics?

Before addressing these specific technical issues in Part II of this report, the next section will make a more detailed overall assessment of the databases identified, in order to explain the logic behind choosing specific entities and datasets for more in-depth analysis.

III. Overall Assessment of DPRK Databases

A. Overall Assessment: Criteria and Limitations

The second stage of the Project focused on general assessment and classification of identified or collected databases containing North Korean economic statistics. Assessment criteria were presented and discussed with the Steering Committee members as well as at an interim report seminar in Seoul in August 2008. In the end, the assessment criteria used in this study included both supply-side and demand-side criteria, as summarized in Table 4. Next, points were assigned depending on the degree to which each database performs according on each criterion, as shown in the table.

Table 4: Criteria for Data/Database Assessment

Criteria	3 Points	2 Points	1 Point
Data update frequency	Regular updating for a period over 5 years	Irregular updating or track record of updating over the past 3 years or so.	One time publication
Institutionalization	Specialized institutions, or divisions in charge of collection/analysis of DPRK data	Specialized personnel of organizations/institutions	Individual researchers
Sources	Primary sources of datasets (incl. rare, unique data, official trade customs data); secondary authoritative sources	Citing primary or secondary “authoritative institutions” w/ their own analyses/notes	Citing other institutions only, or no references indicated
Methodology	Solid methodology indicated/shared	Citing institutions, using solid methodology	No indication of methodology or unknown sources
Comprehensiveness of data	Over 10-year time series data	Data covering 2~9 years	Single-year figures
Data accessibility	Open to public through internet/publications	Purchase/subscription required, but available at university libraries	Proprietary data or difficult-to-obtain information (i.e., cost consideration)
Language	Available in English and other languages (i.e., Korean)	Available practically in one language (English).	Available only one non-universal language
Data presentation format/functions	Excellent: User-friendly formats/functions for data analyses (i.e., downloadable)	Good: Table formats	Not easy: Data and figures in text or figures only.

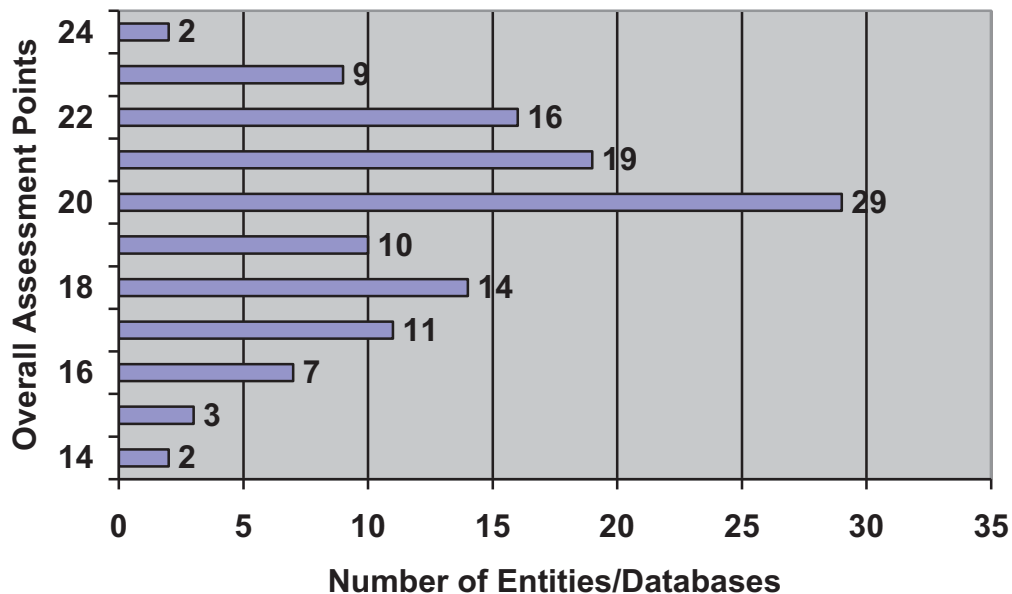
The supply-side factors included: (1) data update frequency; (2) institutionalization of staff members who collect and analyze North Korea economic statistics; and (3) sources of data and technical notes/explanations. The demand-side criteria included: (1) reliability of methodology; (2) comprehensiveness of data; (3) data accessibility; (4) language of availability; and (5) data presentation format and functions (Table 4).

Each database was evaluated based on the above criteria, using a 3-point system. The maximum possible point score that each database can earn is 24 points. This system of evaluation was not intended to rank the identified databases. Rather, the system was used to ensure that a logical methodology was in place for selecting entities or databases for more in-depth technical analyses as case studies. These criteria and point scores can also serve as decision-making guidelines, although not as definitive criteria, for the inclusion of specific entities or databases for inclusion in a larger-scale databank that might be constructed beyond the current phase of this Project.

B. Major Findings

Figure 2 shows the overall assessment of databases identified. In total, 121 entities were evaluated, of which 11 databases earned 24 points (a full score) or 23 points, followed by 16 databases with 22 points, 19 databases with 21 points, and 29 entities with 20 points. The weighted average point score is 19.6 points and the median point score is 20 points.

Figure 2: Overall Assessment of Databases Identified



The entities producing the highest-rated databases have certain characteristics in common. Most of them are global entities, providing comprehensive trade data or economic and social indicators in their databases in the public domain, including UN Comtrade, OECD

databases, and UNCTAD trade databases. But national organizations such as the South Korean Ministry of Unification's website for inter-Korean trade and Japan's Ministry of Finance/Customs database also earned high overall points.

Table 5 shows a more detailed distribution of entities in terms of points assigned based on each assessment criterion. From this analysis, we can confirm many of the characteristics and issues of DPRK statistics that we have intuitively perceived.

Table 5: Distribution of Entities in terms of Assessment Points

Criteria	3 Points	2 Points	1 Point
Data update frequency	70 (57%)	38 (31%)	14 (11%)
Institutionalization	92 (75%)	29(24%)	1 (1%)
Sources	72 (59%)	48 (39%)	2 (2%)
Methodology	72 (59%)	47 (39%)	3 (2%)
Comprehensiveness of data	57 (47%)	49 (40%)	16 (13%)
Data accessibility	93 (76%)	24 (20%)	5 (4%)
Language	24 (20%)	76 (62%)	22 (18%)
Data presentation format/functions	36 (30%)	81 (66%)	5 (4%)

Update Frequency: The majority of databases assessed, accounting for 57%, are updated periodically, on a yearly basis or more frequently. However, the distribution skewed toward regular updating should be interpreted carefully. Most entities adopt necessary reviews and procedures as a matter of practice for all the countries concerned. But this does not necessarily mean that figures are updated for North Korea based on new or fresh data. Many of the UN databases covering population statistics, for instance, the UN Statistics Division (UNSD) or UN Population Division (UNPD), review North Korea's data and statistics on an annual basis, but not always based on receipt of new or necessary data. Many databases including UNSD and UNPD are assigned 3 points for conducting regular updating, but the reality is merely procedural, rather than reflecting the updating of actual data in a true sense. These issues are of course tied to the relative absence of primary source data.

Institutionalization: Institutionalization is also a difficult criterion that requires elaboration and careful interpretation. Many of the institutions operating large-scale multi-country databases are equipped with staff members who are in charge of not only North Korea but also other countries. Highest points (3) were assigned to those organizations even if their databases contain only limited datasets for North Korea, as long as they are likely to swiftly mobilize the necessary staff members to work on North Korean data when and if needed or available. For instance, North Korea is not yet a member country of IFIs such as the IMF and the World Bank. The Bank's World Development Indicators Online database, for instance, has not filled many of the data categories for North Korea due to the unavailability and questionable reliability of North

Korea statistics; out of over 800 data sub-categories contained in the World Development Online database, less than 200 datasets are present for North Korea. But the Bank is institutionally equipped to start providing necessary data once more reliable primary data, which passes the Bank's certain criteria and reliability tests, becomes available.

On the other hand, lower points were assigned to some data sources reviewed, if the North Korea data can be viewed as an individual rather than institutional asset, or if there is no institutional setup for data sharing. Field interviews revealed that some entities had staff members or North Korea experts create websites presenting rare and unique primary data. But the sites have not been updated after these particular staff members left the organizations.

Sources: A majority of entities/databases earned at least 2 points or above for this criterion, given that very few entities display North Korea data without citations or sources of data. In this overall evaluation, databases were categorized slightly differently from during the preliminary evaluation: those listing both primary sources and authoritative secondary sources earned 3 points. For instance, Good Friends earned 3 points for almost all of their survey reports disclosed on their website, as they use primary data sources or interviews, with value-added analyses or observations concerning the food situation or nutritional surveys. Many trade databases earned 3 points, as their source data such as partner countries' official trade/customs statistics can be interpreted as "primary" data even though they are mirror statistics. Examples include customs offices or trade authorities of partner countries, such as the China Customs and Japanese Ministry of Finance/Customs databases, global entities' databases such as UN Comtrade and IMF Direction of Trade, and private ventures such as the Global Trade Information Services (GTIS) that compiles trade data from foreign customs authorities.

If some research publications utilized primary data sources and conducted value-added reliability analyses, such works earned highest points. For instance, KOTRA, while using GTIS's database as well as overseas official customs data as its original sources, makes adjustments if there are questionable and inconsistent data. Similarly, reports compiled by the US Congressional Research Service received 3 points for the same reasons, identifying and verifying some major errors or inconsistencies in official trade statistics and adjusting figures accordingly. Such examples will be discussed in detail as part of this report's technical assessment of trade data (Box 4).

Methodology: About two-thirds of the databases assessed provided some technical notes and methodologies concerning how they obtained and estimated North Korea statistics. Most of them were global entities which disclosed general methods, methodological steps and procedures taken for data collection, verification or estimations. However, the most critically-needed information are the specific assumptions and methods applied to DPRK data, as it often does not fall into the same general categories as data from other countries. Some databases provide such crucial information (for example, the Center for International Comparisons - Penn World Table).

In most cases, however, North Korea-specific methodological information cannot be obtained in the public domain, requiring interviews and discussions with those organizations. It was a major challenge to ascertain the detailed methods or assumptions specifically applied to North Korea by most institutions contacted during the course of Project implementation. Specific questions were directed to relevant entities, but very few responded to such queries.

For instance, the UNCTAD *World Investment Report*, an annual publication as well as an interactive database, contains detailed information on its sources and methods for compiling and/or estimating FDI figures. But North Korea does not fall into any of the groups of countries for which conventional methods and procedures were applied. Further information was necessary for in-depth analysis. Despite these caveats, however, such databases earned 3 points due to their generally-solid technical methodologies and procedures disclosed, while revealing information gaps in their methodological steps specific to North Korea.

Comprehensiveness of data: More than one-third of the databases with DPRK statistics offer comprehensive datasets, defined as covering over ten years, as discussed in the previous section. Examples include the UN National Accounts Main Aggregate database, which contains extrapolated estimates to make datasets available from 1970, the IMF Direction of Trade database, and CIC's latest Penn World Table.

Data Accessibility: The Project consciously focused on investigating publicly available open sources as well as databases available at the SAIS online library network. As a result, databases that earned the highest score of 3 points accounted for three-fourths of databases evaluated.

There are some subscription-based databases that were not available at SAIS and earned low points due to cost factors, but were worth noting as valuable data sources nonetheless. For example, a US private venture, GTIS World Trade Atlas (WTA), was awarded only one point due to cost factors (subscription-based databases), but its utility as a trade database seems to be very high. GTIS has filled in a market niche by successfully serving not only US government organizations but also major entities in South Korea including KOTRA. GTIS's WTA will be included for further technical assessments in Part II of this report.

Aside from cost factors, other accessibility issues such as individual and institutional level constraints were taken into consideration in assessing data accessibility. For instance, price data which is often gathered by individual researchers is neither easily nor fully accessible to a wider audience. Institutional level accessibility issues include some discriminatory rules and procedures applied to certain groups of people (i.e., Korean nationals vs. non-Koreans).

The Project's data accessibility assessment revealed another important issue. Due to the relative absence of DPRK primary data and constraints on accessibility to such data, data users tend to cite more easily accessible "authoritative" entities without paying

appropriate attention to their technical notes and methodologies. For instance, the CIA's *World Factbook* is widely used due to its openness. But sometimes the CIA's PPP-based GDP estimates have been misinterpreted or compared with data from non-comparable sources.

Language: Among the databases reviewed, about one out of five databases containing North Korea statistics is available in both Korean and English. Another one out of five is available only in non-universal languages, including Korean. The majority of databases are accessible only in English.

Language factors are critical in terms of usability. Highest points were awarded to bilingual databases taking into consideration both Korean and non-Korean speaking analysts' perspectives. However, this does not necessarily mean that there is lower usability or reliability for databases available only in English or in non-universal languages. To the contrary, we identified some databases that are sadly under-utilized due to language barriers (i.e., the South Korean KOSIS database), as briefly discussed in the previous section. The Project aimed to identify such databases as one of the Project's key objectives.

It is interesting to note that "bilingual" websites or publications do not necessarily present identical contents. Often, only parts of datasets or information are made available in English, if the producing entities are from non-English speaking countries. In other words, more datasets and user-friendly formats/functions tend to be made available in the producing entity's own language. The Chinese Ministry of Commerce, for instance, presents trade datasets on its websites in both Chinese and English. But the Chinese version is much more user-friendly in terms of datasets included and presentation. Deutsche Bundesbank's publication, although available only in German, offers DPRK exchange rates both in US dollar and Euro terms. While higher points were awarded to bilingual data sources, the Project attempted to select carefully databases for technical analyses in Part II.

Data Presentation and format: The majority of data sources covered earned 2 points, demonstrating good presentation of data in table formats with actual data points. One-third of the databases offered user-friendly database functions, including downloadable options and/or embedded value-added analyses in databases. The US Census Bureau's International Data Base (IDB), for example, has functions enabling analysts to derive population pyramid structures instantly. GTIS also has simple and powerful analytical tools embedded in its database, such as deriving unit costs of imports and exports. UNCTAD's *Handbooks of Statistics* offers pages indicating in a simple table format North Korea's historical trade structure by county groups or region, such as developed and developing countries.

Low points were assigned if relevant information such as technical notes or methodologies cannot be located easily, for example if methodologies are separately presented from dataset tables without any indication of where they can be accessed, causing confusion (the Bank of Korea database has this problem). Also, if some research

works, while relevant and significant, have only graphs or figures without offering specific figures, only one point was provided as it poses constraints on data gathering and analysis efforts, or integration of such data into a databank.

The Project set a general cut-off point at 20 points (the roughly weighted average or mean point score): databases with 20 points or more were in principle considered for further analysis. But the Project has taken into consideration not only overall points but also specific criteria for understanding the situation of available DPRK statistics and data. For example, some databases are extremely user-friendly, but may not earn highest points due to their proprietary nature. Conversely, some databases that earned high points (trade databases) while not necessarily satisfying a specific analytical purpose for this study were not included (for example, the official bilateral trade databases of New Zealand and Australia).

Summary: Preliminary and general assessments of available databases during the first and second stages of Project implementation provided the information needed to proceed with more in-depth technical assessment of selected datasets.

- For **macroeconomic data**, Part Two will focus on a few of the most frequently cited entities and their databases, including the BOK, the CIA and the UN, as well as as possibly under-utilized databases such as CIC's Penn World Table.
- Selected **microeconomic data** (prices) and **socio-economic indicators** (population and health indicators) will be analyzed in the context of deriving macro data such as GNI per capita. One of the most frequently cited data sources for price data, Good Friends, will be introduced as a case study. Good Friends is generally open about sharing data compared to other institutions or individuals.
- Technical analysis for **trade data**, on the other hand, will focus on the introduction of different types of data sources including the IMF's *Direction of Trade* database, UN Comtrade, and data from the South Korean Ministry of Unification (MOU), Korea Trade-Investment Promotion Agency (KOTRA) and Korea International Trade Association (KITA), as well as the trade and customs authorities from Japan and China. Also, there are specialized trade data sources, both proprietary and non-proprietary data, with value-added technical analysis already embedded in such databases. These include GTIS's World Trade Atlas and the UNCTAD/International Trade Center. One entire section of the report is devoted to trade data, which can also serve as a "user guide" of trade databases for specific analytical purposes.¹⁴

¹⁴ The author is indebted to William Newcomb for the concept of a "user guide" approach to conducting the Project's in-depth analysis of trade databases.