Notes on Using National "Starter" Data Sets for LEAP
First Draft

Charles Heaps
LEAP Developer
June 2010

Table of Contents

Stockholm Environment Institute
Email: leap@sei-us.org
Web: www.energycommunity.org and www.sei-us.org

1 Introduction

SEI has developed a series of national level "starter" data sets for use in SEI's LEAP energy planning software.

These starter data sets are intended to be used by developing country energy planners as a starting point for their analyses. Each data set has been developed for a single country and one data set has been created for each country for which energy data is available from the International Energy Agency (IEA).

According to SEI's agreement with the IEA, these data sets are available only to qualified developing country organizations (governmental organizations, non profits and academic institutions). Each organization may access data sets only for their own country.

Please also note that it is a condition of downloading and using of these data sets that any improvements made to them should be documented and copies provided back to SEI (who in turn will make these changes available to the IEA).

It is important to note that these data sets are intended only as a **Starting Point** for analysis. The data they contain are based only on readily available and regularly maintained international sources of data. Users will need to check, refine and correct these starter data sets – typically by using their own locally available data, which in many cases will be superior in detail and quality to the data provided here. More importantly, the data in these data sets are fairly aggregate in nature. We expect that many users will wish to create more disaggregated data sets that, for example, including information on end-uses and technologies. In such cases, we hope that the more aggregate data in these data sets can serve as a jumping-off point for creating these more detailed data sets and as data that the more detailed data sets

Key Information

- National level starter data sets for use with LEAP.
- Available to qualified developing country energy analysts.
- Include aggregate historical data on energy consumption, production, energy sector emissions and non energy sector emissions.
- Based on a range of international data sources including data from the IEA, World Bank, IPCC, the UN, WEC and WRI.
- Includes simplified projections to 2030 assuming constant elasticities.
- Intended only as a starting point for analysis: NOT INTENDED AS COMPLETE PROJECTIONS.
- Users will need to check, refine and correct these starter data sets –typically by using their own superior locally available data.
- It is a condition of use of these data sets that any improvements made to them should be documented and copies provided back to SEI and the IEA.
- Each data set is provided as a single ".leap" data file and can be downloaded from the COMMEND web site.

can be calibrated against. Similarly, the forward-looking projections to 2030 in these data sets are intended as only a starting point for users to create their own scenarios. They are not intended as a finished forecast of what will, might or should happen in any country.

2 What Data is Included?

Each data set includes a comprehensive picture of historical energy demand and supply from 1990 to 2007 based on the IEA's database of World Energy Balances (2008)[1]. The energy data includes information on energy use by fuel in each major demand sector (households, industry, transport, services, agriculture, non energy use and non specified uses). On the supply side the data sets include information on distribution losses, own use, heat, electricity and CHP production as well as information on gas works, refineries and liquefaction. The data sets also include historical data describing statistical differences between demand and supply data, and annual stock changes in fuels.

Data on fuels is provided in terms of the simplified fuel categories used in the IEA's less detailed energy balances. As far as possible the data sets have been designed to match the data in the IEA's energy balances although some differences do exist and users should thoroughly review all data sets to ensure that the results coming from LEAP closely match the IEA's published balances.

Energy information is supplemented with the IPCC's standard Tier 1 emission factors to provide estimates of greenhouse gas emissions from the energy sector. The data sets also include simple estimates of non-energy sector GHG emissions taken from the CAIT database developed by the World Resources Institute (WRI, 2010). Energy Reserves data is also included, taken from the World Energy Council (WEC) Survey of Energy Resources, 2007.

Energy consumption and production data are supplemented by historical population data and the medium variant population projections of the United Nations (UN Population Prospects, 2008 revision).

Historical GDP data (in both MER and PPP terms) area taken from the World Bank's World Development Indicators database (World Bank 2010). The data sets also include various other useful historical indicators where available taken from the United Nations database (http://data.un.org) including urbanization rates, various transportation indicators (road, rail and air passengers carried; road, rail and air freight), development indicators (GINI coefficients, life expectancy, poverty, HIV prevalence, etc.) and data on electric sector capacity in Megawatts for various type of power plants.

[1] In some cases the data set includes time-series data going back further. For example, the IEA energy data goes back to 1970.

Additional high level economic indicators taken from the World Bank Database such as value added shares for services, agriculture and industry, are supplemented by calculations of the share of manufacturing value added coming from the various major industrial sectors for which the IEA maintains energy data (iron and steel, non ferrous metals, chemicals, non metallic minerals, paper and pulp, transport equipment, etc.). These calculations are derived from data in UNIDO's INDSTAT database. See below for an explanation of how these data are put together in the LEAP data set and used in LEAP's calculations.

Finally, the data sets also include a skeleton structure suggesting a format for a simple mitigation assessment in which users specify the energy and emissions reductions possible in various different sectors of the economy. For information on this, please refer to section 4.

In the next section we introduce the basic methodology used in the starter data sets and describe how it makes use of the standard data sources outlined above.

3 The Structure and Methodology of the Starter Data Sets

Each starter data set is structured using LEAP's standard tree branch structure. The top level branches in the tree are Key Assumptions, Demand, Statistical Differences, Transformation, Stock Changes, Resources and Non Energy Sector Effects. Each of these is described in turn.

3.1 Key Assumptions

The key assumptions branches in the starter data sets store a variety of demographic, economic, and development indicators extracted from a variety of data sources. Some of these data are referenced in LEAP's energy calculations (see descriptions below) while other data is included for information purposes only or for potential later use by users of the data sets.

The **Key Assumptions** branch contains the following sub branches:

- **GDP**

  Current Accounts: Under the GDP branch you will find historical data on GDP in both market exchange rate (MER) and purchasing power parity (PPP) terms in units of billions of year 2000 constant dollars. This data comes from the World Bank World Development Indicators (WDI 2010) database and typically covers the years 1971-2006.

  Baseline Scenario: In the baseline scenario GDP is calculated as the product of population and per capita income.

- **Value Added**

Current Accounts: Under the Value Added branch you will find historical data on the percentage share of GDP derived from services agriculture and industry, as well as the percentage share of GDP coming from manufacturing. This data comes from the World Bank World Development Indicators (WDI 2010) database. Industry consists of Manufacturing and Construction. In other words manufacturing is a subset of industry. Thus we include a second variable labeled "Manufacturing2" that calculates manufacturing as a percentage share of industry.

Baseline Scenario: The baseline scenario assumes no future changes in value added shares.

- **Population**

Current Accounts: Under the Population branch you will find historical data on the population of the country in units of millions of people. This data comes from the UN Population Prospects , medium variant, 2008 Revision and typically covers the years 1950-2050.

Baseline Scenario: The baseline includes the UN's medium variant population projections to the year 2050.

- **Income**

Current Accounts: In Current Accounts the Income variable is calculated by dividing GDP by the population of the country. Income is calculated both in MER and PPP terms in units of $/person.

Baseline Scenario: The baseline includes simple growth forecasts for per capita income to 2030 based on the regional growth assumptions of the US EIA's International Outlook (EIA, 2009). The various regional growth assumptions in that report are applied to all of the countries in the corresponding region.

- **Transportation**

Current Accounts: Under the Transportation branches you will find various historical indicators for both passenger and freight transport. This data comes from the World Bank World Development Indicators (WDI 2010) database and is available for various different years depending on the country.

Baseline Scenario: The baseline has no projections of these variables.

These variables are not used in LEAP's energy calculations in the starter data set.

- **Development**

  Current Accounts: Under the Development branches you will find various historical development indicators. This data comes from the World Bank World Development Indicators (WDI 2010) database and is available for various different years depending on the country. The indicators are GINI index (a measure of income inequality[2]), life expectancy (in years), poverty (% of population), HIV Prevalence (% of population), undernourishment (% of population).

  Baseline Scenario: The baseline has no projections of these variables.

  These variables are not used in LEAP's energy calculations in the starter data set.

- **Urbanization**

  Current Accounts: Under this branch you will find data on the percentage of urban and rural households in the data set country where available. This data comes from the World Bank World Development Indicators (WDI 2010) database and is available for various different years depending on the country. The rural percentage is calculated as 100-Urban.

  Baseline Scenario: The baseline has no projections of these variables.

  These variables are not used in LEAP's energy calculations in the starter data set.

- **Manufacturing Value Added**

  Current Accounts: Under this branch you will find data on value added in major manufacturing sector in current US$. This data is derived from the Value added data in the UNIDO INDSTAT database. That data is mapped to the major manufacturing sectors in which the IEA reports its energy statistics. The mapping of 3 digit Rev3 ISIC Codes is displayed in Table 1. For China, the mappings are based on 3 digit Rev2 ISIC Codes.

  Baseline Scenario: The baseline has no projections of these variables.

[2] See: http://en.wikipedia.org/wiki/Gini_coefficient

IEA Sector	For China 3 digit Rev2 ISIC Codes	All Other Countries 3 digit Rev3 ISIC Codes
Iron and Steel	371	2710, 2731
Chemicals & Petrochemicals	351, 352, 353, 354	241, 242, 2430
Non Ferrous Metals	372	2720, 2732
Non Metallic Minerals	355, 356, 361, 362, 369	2610, 269
Transport Equipment	384	3410, 3420, 3430, 351, 359
Machinery	381, 382, 383, 385	3110, 3120, 3130, 3140, 3150, 3190, 3210, 3220, 3230
Food and Tobacco	311, 313, 314	151, 1520, 153, 154, 155, 1600
Paper Pulp and Print	341, 342	210, 221, 222, 2230.
Wood & Wood Products	331, 332	2010, 202
Textiles & Leather	321, 322, 323, 324	171, 172, 1730, 1810, 1820, 191, 1920
Other	The balance of all manufacturing VA - the sum of the sectors listed above.	
All	All manufacturing VA.	

Table 1: ISIC Codes used for manufacturing in the LEAP Starter Data Sets

- **Manufacturing Value Added (Shares)**

  Under this branch you will find calculated the percentage shares of value added in major manufacturing sectors. This data is calculated based on the dollar values under the Key Assumptions\ Manufacturing Value Added branches.

- **Electric Generating Capacity:**

  Current Accounts: Under this branch you will find data on installed capacity of different types of power plants (thermal, nuclear, hydro, wind, solar, geothermal, and wave/tidal) reported in megawatts (MW). The data is divided into public and self producer power plants. The data is from the UN data web site (http://data.un.org).

  Baseline Scenario: The baseline has no projections of these variables.

  These variables are not used in LEAP's energy calculations in the starter data set.

3.2 Demand

The Demand branches in the starter data are the place where LEAP calculates all final energy consumption of energy as well as direct demand-side emissions of pollutants. Total demand is broken down into seven sectors following the approach used in the IEA's energy balances: households, industry, services, transport, agriculture, non energy use and non-specified.

3.2.1 Current Accounts

Since the IEA data records total energy consumption, not energy intensity data, the LEAP data set calculates historical energy intensities in each sector for its Current Accounts data. In the household, services, agriculture, non energy use and non-specified sectors these intensities are calculated for the sector as a whole, while in the transport sector intensities are calculated by mode for road, rail, and domestic marine transport, international aviation, domestic aviation, pipelines and non specified transportation. For industry, intensities are calculated for the major industrial subsectors listed in Table 1.

Energy intensities are calculated using the following formula in each year:
$EI_y = E_y / A_y$

Where:
EI_y = Energy Intensity in year y
E_y = Total Energy in year y
A_y = Activity level in year y.

Sector	Activity Level
Households	Total Population
Industry	Subsector Value Added if available, otherwise Industry Value Added
Transport	GDP
Services	Services Value Added
Agriculture	Agriculture Value Added
Non Specied	GDP
Non Energy Use	GDP

Table 2: Activity Level Drivers Used in the LEAP Starter Data Sets

Different activity levels are used in each sector to calculated energy intensities as shown in Table 2. In the services sector the activity level is specified as the product of GDP (in PPP terms) and the percentage value added from the services sector, both of which are stored as data in the Key Assumptions branches. The equation in LEAP is written as:

Key\GDP\GDP PPP[Billion us$] * (Key\ValueAdded\Services[%]/100)lo

A similar equation is used to calculate the Agricultural value added as a function of GDP (PPP) and agricultural value added.

In each industrial subsector, GDP is first allocated to all industries.

In countries where subsectoral value added data is available, this amount is then allocated among manufacturing and other industrial value added. Finally, manufacturing value added is allocated among the various manufacturing subsectors using equations like this:

For total industry:
Key\GDP\GDP PPP[Billion us$] * (Key\ValueAdded\Industry[%]/100)

For subsectors (e.g. Iron and Steel) the value calculated above is further multiplied by the following:

Key\ValueAdded\Manufacturing2[%] * (Key\Manufacturing ValueAdded Shares\Iron and Steel[%]/100)
For countries where subsectoral value added shares are not available, the activity level is simply total industrial value added and hence only the first of the above two equations is used.

Historical energy consumption data is stored in a user-defined variable (initially defined in the General: User Variables screen) labeled **TotalEnergy** with units of Thousands of Tonnes of Oil Equivalent (kTOE). These data are extracted from the IEA's World Energy Balance database 2008. This data is stored as the total for each sector for households, services, agriculture, non energy use and non specified energy use, and by mode for transport and by subsector for industries. In addition to total energy consumption, LEAP also stores data on the fuel shares (% shares of TOE) for each fuel consumed within these sectors and subsectors. Fuel share data is also derived directly from the IEA's World Energy Balance database 2008.

Energy intensities are simply calculated as the total energy consumption divided by total activity level for each sector/subsector using the following type of equation in LEAP:

TotalEnergy[TOE]/Total Activity

3.2.2 Baseline Scenario

In the baseline scenario, the TotalEnergy Variable is not used. Instead users can make separate projections of the activity level, energy intensity and fuel share variables and LEAP will then automatically use these to calculate the energy consumption for each fuel in each sector in each year of the scenario.

The starter data sets make no explicit projections of future fuel shares, future energy intensities or future value added shares. These are all set to stay constant over time. Users of the data sets will need to make their own projections of these variables.

However, at the highest level, we do provide default projections of the main demographic variables. Population projections to 2050 are provided from the United Nations (UN Population Projections, 2008 revision). GDP projections (in PPP terms) are based on the regional growth assumptions of the US EIA's International Outlook (EIA, 2009). The various regional growth assumptions in that report are applied to all of the countries in the corresponding region. Income is then multiplied by population to give overall future GDP (PPP) projections. Note that these projections should be viewed as a starting point only. We expect that many if not most users will wish to use their own assumptions for this key variable.

3.3 Transformation

The starter data sets include specific representation of the major Transformation sectors (modules) listed in Table 3.

Transformation Sectors in LEAP Starter Data sets.
Distribution Losses
Own Use
Coal Transformation
Gas Works
Heat Production
CHP Production
Electric Generation
Oil Refining
Liquefaction

Table 3: Transformation Sectors

Efficiency and historical production data for each sector are derived from the IEA's World Energy Balance database 2008. Other supplementary variables required to allow LEAP's Transformation calculations to proceed are all set to SEI assumed values. These include: lifetime, interest rate, maximum availability, and capacity credit values. Values for the exogenous capacity variable (in MW) are rough values calculated internally based on default assumptions about availability and the IEA's data on historical production – they do not reflect actual data on installed capacity. Users may wish to make use of the capacity data stored under the **Key Assumptions\Electric Generating Capacity** branches or better still they may wish to make use of their own local data sources.

The baseline scenario has no data about changes in conversion efficiencies and thus is set to assume that all efficiencies stay constant in future years. It does include default rates of growth for generating capacity, which are derived from the regional trends for generating capacity seen in the US EIA's International Outlook (EIA, 2009).

3.4 Statistical Differences

The starter data sets include historical data on statistical differences (the difference between final consumption values and energy supplies). For Current Accounts this data is taken directly from the IEA's World Energy Balance database 2008. The baseline scenario has no explicit projections of these values, so that all future statistical differences are set to zero.

3.5 Stock Changes

The starter data sets include historical data on stock changes (the supply of primary energy from in-country stocks). The convention used in LEAP is that negative stock changes indicate an increase in stocks. For Current Accounts this data is taken directly from the IEA's World Energy Balance database 2008. The baseline scenario has no explicit projections of these values, so that all future stock changes are set to zero.

3.6 Resources

The starter data sets include historical data on reserves of primary resources. For Current Accounts this data is taken from the World Energy Council's 2007 Survey of Energy Resources.

For most fuels, the reserves are specified at the end of 2005. The baseline scenario has no explicit projections of these values, so that reserves are assumed to decline as they are consumed in future years in the scenario.

3.7 Energy Sector Emissions

The starter data sets include the IPCC's standard Tier 1 emissions factors taken from LEAP's TED database. These factors include default values for CO_2, CO, CH_4, VOCs, NO_x, N_20 and SO2 for the combustion of different fuels in each different demand and transformation sector. Most factors are specified in units of Tonnes per TJ of fuel combusted. All emission factors are assumed to be constant in the starter data set.

3.8 Non Energy Sector Effects

The starter data sets include historical data on non-energy sector GHG emissions taken from the CAIT database developed by the World Resources Institute (WRI). The baseline scenario has no explicit projections of these values, so that all non energy sector emissions are assumed to decline as they are consumed in future years in the scenario.

4 Template for Mitigation Assessment

The starter data sets include suggested branch structures but no data for carrying out mitigation assessments. Users should feel free to adapt or change these branches in developing their own mitigation assessments, but we hope they will serve as a useful default for organizing analyses.

5 References

European Commission. 2010. European Economic Forecast - Spring 2010. European Commission. http://ec.europa.eu/economy_finance/eu/forecasts/2010_spring_forecast_en.htm

Heaps, C. 2010. *LEAP User Guide.* Online document. Stockholm Environment Institute, Stockholm, Sweden. http://www.energycommunity.org

International Energy Agency, 2008. *World Energy Balances 2008 Edition.* CD ROM OECD/IEA, Paris, France.

Intergovernmental Panel on Climate Change. 1996. Revised 1996 IPCC Guidelines for National Greenhouse Gas Inventories. Volume 3: Reference Manual. http://www.iea.org/ipcc/invs1.html

United Nations. 2009. United Nations Data. http://data.un.org

United States Energy Information Administration. 2009. International Energy Outlook 2009. Washington, DC. http://www.eia.doe.gov/oiaf/ieo/

United Nations. 2009. *Word Population Prospects 2008 Revision.* United Nations. New York, USA. http://esa.un.org/unpp/

United Nations Industrial Development Organization (UNIDO). 2009. INDSTAT 4 2008 ISIC Rev. 3. CD ROM. UNIDO, Vienna, Austria.

World Energy Council. 2007. Survey of Energy Resources 2007. WEC. London. http://www.worldenergy.org/documents/ser2007_final_online_version_1.pdf

World Resources Institute. 2010. Climate Analysis Indicators Tool (CAIT) version 7.0. Web Site. World Resources Institute, Washington, DC. http://cait.wri.org

World Bank World Development Indicators 2010. World Bank. , Washington DC, USA. http://data.worldbank.org/data-catalog