Data Collection & Industry Standards
(Chapter 8 – Software Project Estimation)

Alain Abran

(Tutorial Contribution: Dr. Monica Villavicencio)

Topics covered

1. Introduction

2. The International Software Benchmarking Standards Group (ISBSG)

3. ISBSG Data Collection Procedures

4. ISBSG Individual Project Benchmarking Reports

5. Using the ISBSG Repository

8.1 Introduction

Data Collection Requirements

- Data from completed projects.

  - Projects with effort and functional size data, completed and with no uncertainty remaining.

- Consensus on data collection standards
  - Clearly identify, categorize, and measure using the same set of definitions, classification criteria, and rules for assigning quantitative values.

- Detailed data collection procedures
  - To ensure consistency in data collection, completeness and non ambiguity.

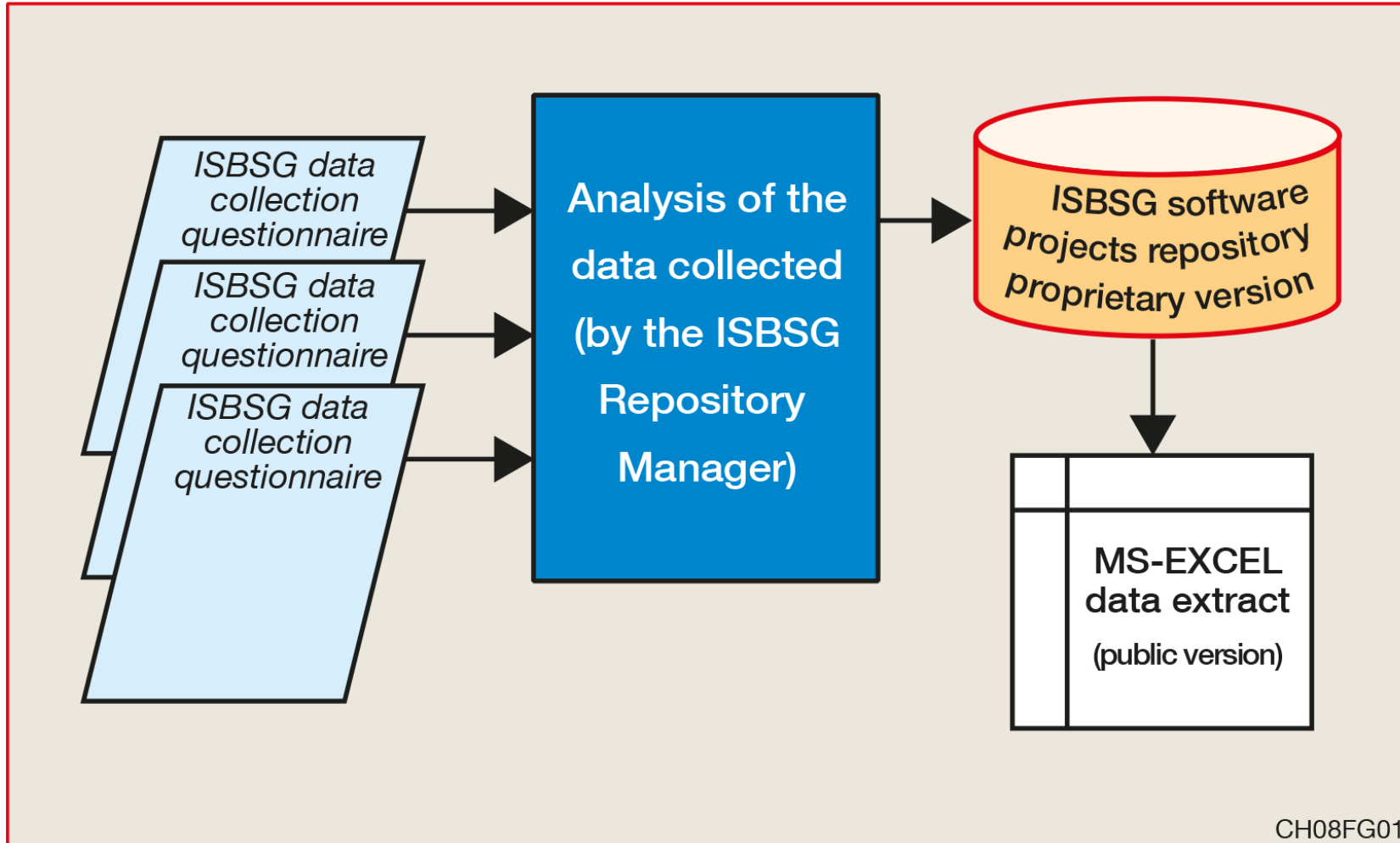# 8.2 The International Software Benchmarking Standards Group (ISBSG)

The ISBSG organization

- **International Software Benchmarking Standards Group (ISBSG):**

  - Provides a worldwide repository of software projects.

    - Available standards for data collection (free of charge).

  - Contains descriptive variables and quantitative data from which a number of ratios can be derived for benchmarking purposes.

  - Is a not-for-profit organization.

    - Its members are the software measurement associations from Australia, India, Japan, the UK, USA, among others.

The ISBSG Repository

- Industry standardized data.

- Different types of projects (enhancement, new development and redevelopment).

- Categories:
  - Telecommunications
  - Banking
  - Insurance
  - Manufacturing
  - Engineering
  - Government, Public Administration, Regulation
 - Etc.

ISBSG repository process organization

ISBSG data collection questionnaire

ISBSG data collection questionnaire

ISBSG data collection questionnaire

Analysis of the data collected (by the ISBSG Repository Manager)

ISBSG software projects repository proprietary version

MS-EXCEL data extract (public version)

CH08FG01

8.3 ISBSG Data Collection Procedures

Data Collection Questionnaire

- It is available for free at www.isbsg.org

- It has 7 parts (structure).

- ISBSG provides a glossary of terms and measures
  - to assist in the collection of project data.
  - to standardize the way the collected data are analyzed and reported.

Structure of the ISBSG questionnaire

CH08FG02

11
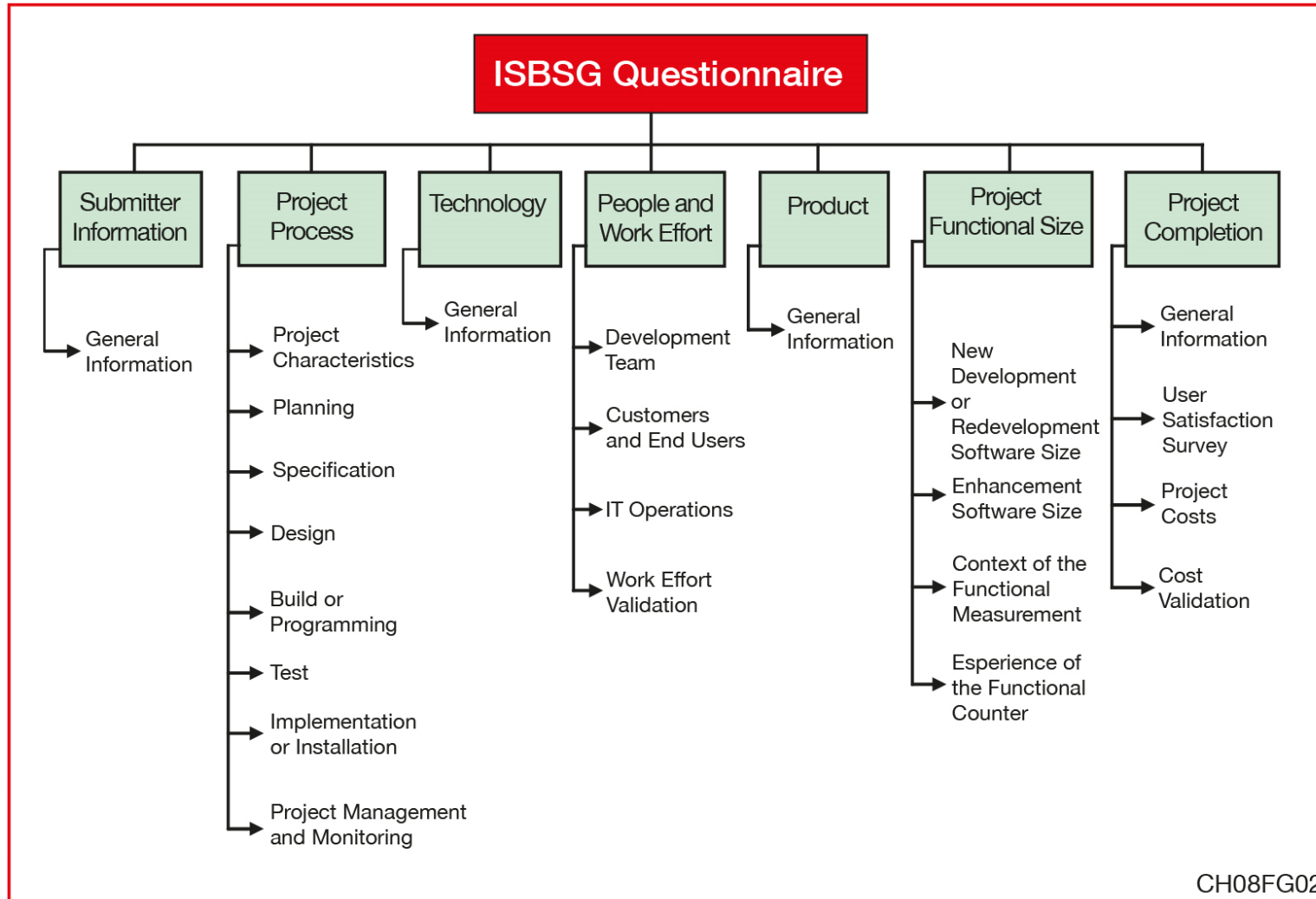
ISBSG Data Definition

- Data fields:

  - *Nominal* (project management tool name, requirements tool name, etc.)

  - *Categorical* (development platform: mainframe, mid-range, PC)

  - *Numerical* (effort in hours, size in Function Points, etc.).

Example 1 of detailed standards for the Effort data: the Resource Levels

Resource Level	Description
Level 1: Development Team	• Those responsible for the delivery of the application under development. The team or organization that specifies, designs, and/or builds the software also typically performs testing and implementation activities. This level comprises the following: Project Team Project Management Project Administration Any member of IT Operations specifically allocated to the project
Level 2: Development Team Support/IT Operations	• Those who operate the IT systems that support the end-users and are responsible for providing specialist services to the development team (but not allocated to the team). This level comprises the following: Data Base Administration Data Administration Quality Assurance Data Security Standards Support Audit & Control Technical Support Software Support Hardware Support Information Centre Support
Level 3: Customers/End-Users	• Those responsible for defining the requirements of the applications and sponsoring/championing the development of the application (including the software's end-users). The relationship between the project customer and the software's end-users can vary, as can their involvement in a software project. This level comprises the following: Application Clients Application Users User Liaison User Training

Example 2: Time Recording Methods

Time recording method	Description
Method A: Staff Hours (recorded)	The daily recording of the entire WORK EFFORT expended by each person on project-related tasks. For example, a person working on a specific project from 8 am until 5 pm with a 1-hour lunch break will record 8 hours of WORK EFFORT.
Method B: Staff Hours (derived)	It is possible to derive the WORK EFFORT when it has not been recorded on a daily basis as in Method A. It may only have been recorded in weeks, months, or even years.
Method C: Productive Time Only (recorded)	The daily recording of only the productive effort (including overtime) expended by a person on project-related tasks. Using the same example as in Method A above, when the non productive tasks have been removed, (coffee, liasing with other teams, administration, reading a magazine, etc.), only 5.5 hours can be recorded.

8.4 ISBSG Individual Project Benchmarking Reports
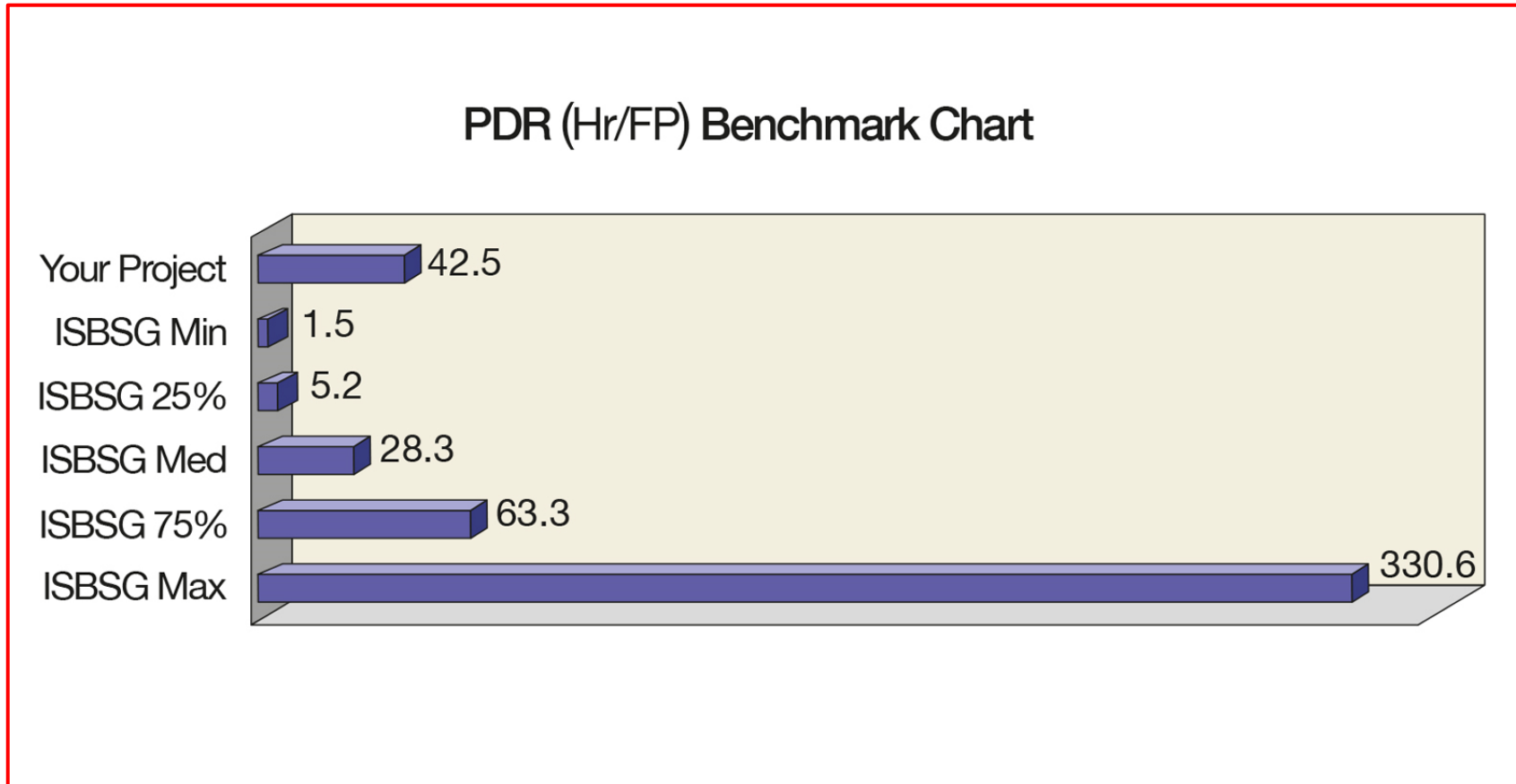
Benchmarking report: Sections

1. Productivity benchmarking

2. Quality benchmarking

3. Assessment of the project data submitted

4. Normalization of the effort reported

Project Delivery Rate (PDR)

- Comparison of the productivity of the project submitted with those in the repository.

- Project delivery rate (PDR): the number of hours to deliver a functional size unit.

- Example of factors impacting PDR:
    - Development platform
    - Language type
    - Maximum team size
    - How the development methodology was acquired.

Example of an ISBSG benchmark report on: Project Delivery Rate (PDR) in hours per FP

PDR (Hr/FP) Benchmark Chart

Category	Value
Your Project	42.5
ISBSG Min	1.5
ISBSG 25%	5.2
ISBSG Med	28.3
ISBSG 75%	63.3
ISBSG Max	330.6

Example of an ISBSG Quality benchmarking report: Project Defect Density

Defect Type	Project	ISBSG		
		Min	**Mean**	**Max**
Total no. of defects	*0/1000 FP	*0	17	91

The benchmarking report compares the project's defect density (defects per 1000 Function Points) to projects in the ISBSG repository.

ISBSG Assessment of the data collected

The ISBSG data administrator assesses the quality of the data submitted and assign the following rating to the data collected.
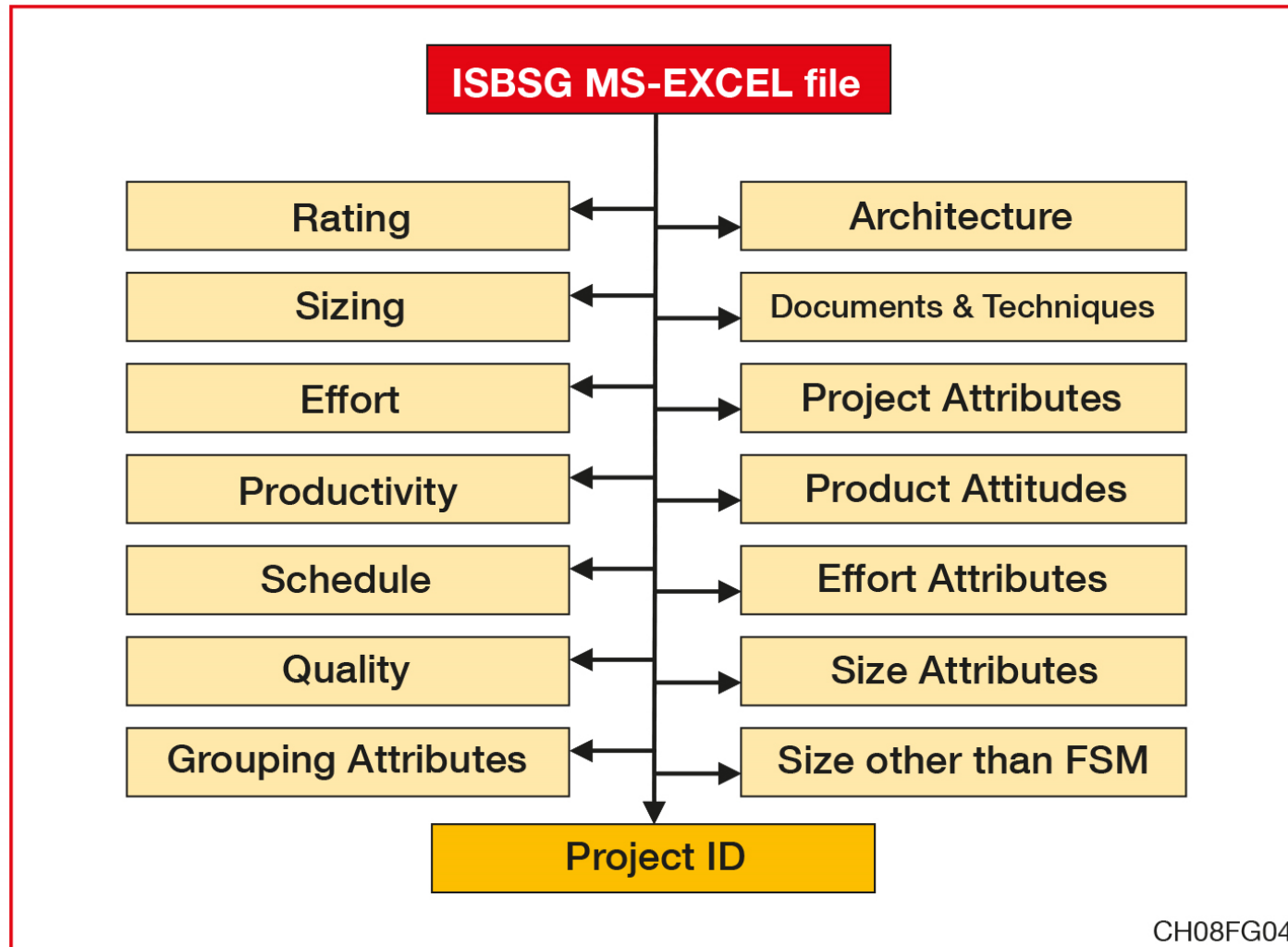
Rating Code	Rating Description
A	The data submitted were assessed as sound, with nothing identified that might affect their integrity.
B	The submission appears fundamentally sound, but there are some factors that could affect the integrity of the submitted data.
C	Owing to significant data not being provided, it was not possible to assess the integrity of the submitted data.
D	Owing to one factor or a combination of factors, little credibility should be given to the data submitted.

Normalization

- **The ISBSG calculate the normalized PDR when the project reported effort is only a *portion* of the project life cycle.**

- **Note: it takes a number of other variables into consideration to derive this 'Normalized Effort'**

  (including the Resource level, etc.)

8.5 Using the ISBSG Repository

ISBSG data extract structure

ISBSG MS-EXCEL file

| Rating | Architecture |
Sizing	Documents & Techniques
Effort	Project Attributes
Productivity	Product Attitudes
Schedule	Effort Attributes
Quality	Size Attributes
Grouping Attributes	Size other than FSM

Project ID

CH08FG04

Data preparation: Quality of the data collected

- Data quality verification
  - The DQR Rating: from Very Good (A) to Unreliable (D)

- Data completeness verification.
  - Look at other data fields that may provide important clues about the quality of the collected variables (effort and functional size).

DATA PREPARATION

http://www.point2.com/listingSyndication/howItWorks.html
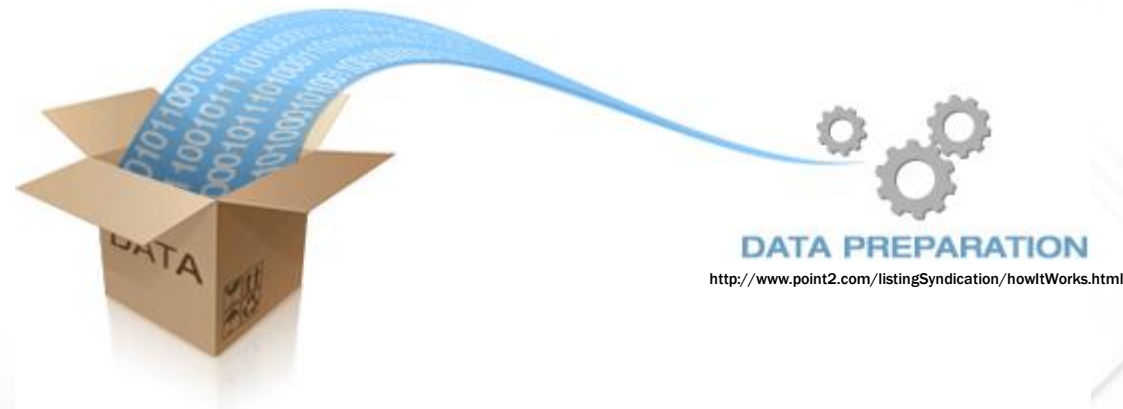
Missing data

- Effort data can vary considerably from one organization to another:
  - Effort from initial planning to full deployment.
  - Only effort for the programming and testing activities.

- In ISBSG:
  - Data collectors have to map their own life cycle to a standardized ISBSG life cycle (Planning, Specification, Design, Build, Test, and Implement).
  - Total project effort is a mandatory.
  - The detailed effort per phase is optional (often missing).
  - An additional field (calculated data) is provided, which includes the derivation of normalized work-hours to represent a full life cycle.

Exercises

1. Why would an organization spend effort and money on a benchmarking exercise?

2. Why is it important to adopt standards in data collection?

3. At what organizational process maturity level(s) is benchmarking worthwhile? Would the benefits be the same at each level? Explain.

4. Identify key success factors in benchmarking.

5. Identify key failure factors in benchmarking.

6. How is the ISBSG organized?

7. How is ISBSG data collection organized?

8. What sections of the ISBSG data collection questionnaire are devoted to software projects?

Exercises

9. Which ISO standards have been adopted by the ISBSG for the collection of the functional size of the software delivered by a project?

10. What are the key differences between the three levels of human resources?

11. In the ISBSG benchmarking report, what is the Project Delivery Rate – PDR?

12. In the ISBSG benchmarking report, how does the PDR of your project compare to that of other projects in the ISBSG repository?

13. In section 4 of the ISBSG benchmarking report, what is involved in effort normalization?

14. How does the ISBSG assess the quality of the data submitted to its data administrator?

15. Why is the normalization of effort data necessary for benchmarking purposes and for building productivity models?

Term Assignments

1. Look at your time reporting system and map it to the ISBSG definitions for the time recording methods and work effort breakdown in Box 1.

2. Of the data fields in the ISBSG questionnaire, what is the ratio of quantitative data to nominal data?

3. When you have to carry out a data analysis on the ISBSG dataset (or any other dataset), what preparation steps are important?

4. What is the significance (and impact) of bundling projects sized with different functional size measurement methods?

5. What is the significance (and impact) of bundling projects for which effort was collected with different time recording definitions?

6. If your organization does not have a repository of completed projects, how can you test the relevance for your organization of using an external dataset (such as the ISBSG, or any other similar repository) in your organization? What would you recommend to an organization in such a context?

Term Assignments

7. Collect the data from a team project carried out in another software engineering course, and compare its performance with that of similar projects in the ISBSG repository. Discuss and comment on the interpretation context.

8. Take the documentation for a completed project, including the total effort that was needed to complete the project. Access the project estimation information based on which the project budget was initially allocated. Compare the figures, comment on them, and prepare recommendations for your management.

9. Identify five organizations you would like to propose to your CIO for a benchmarking exercise. Provide a detailed rationale for your recommendation. What do you know specifically about the software organizations you recommended to your CIO? Why would these five organizations be ready to share data with your organization?

10. The ISBSG publishes how many data are from which country on its website. How many projects are from your country? Explain this number, relative to countries of similar size. Why would countries of similar size have more projects? What factors would lead to such a situation?

11. Download a Data Collection Questionnaire from the ISBSG website. Use it to collect data from your project. For how many fields could you obtain data? What is the ratio of those fields to the total number of fields? Why were you not able to collect all the data in your organization? For the data you could not collect, were they meaningless for project monitoring and project management?