



Investigation of the Effort Data Consistency in the ISBSG Repository



Déry & Abran



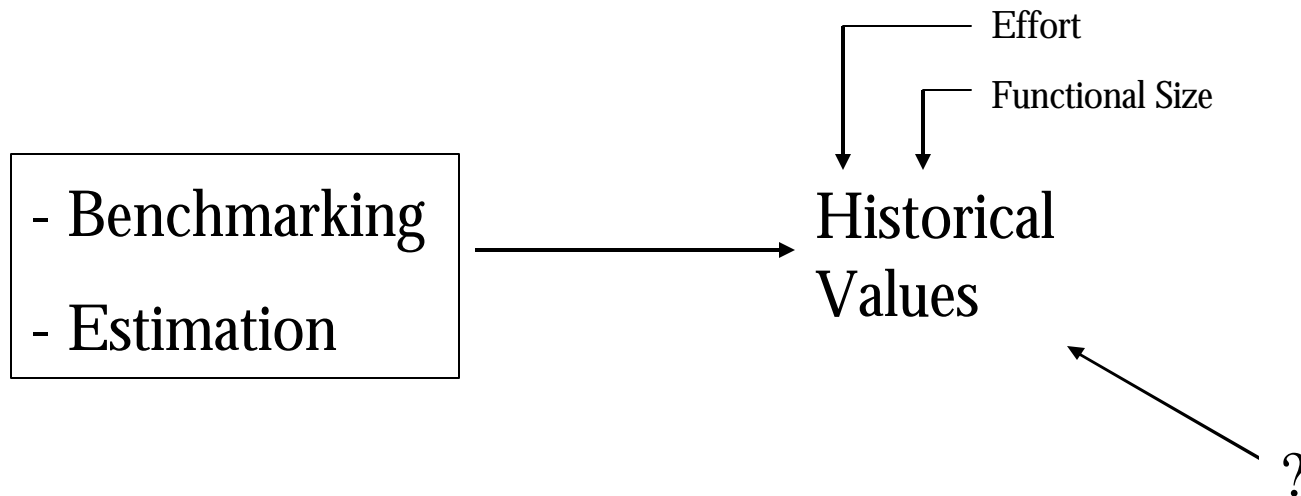
Agenda

- ***Context***
- ***The ISBSG Repository***
- ***Data Preparation***
- ***Observations***
- ***Conclusion***



Context

☞ To produce good estimates and perform benchmarks, historical data must be present





The ISBSG Repository of Project Data

- ☞ ISBSG contains information on +100 data fields for 3024 projects (release 9 - 2005)*
- ☞ ISBSG captures information about the project process*
 - The technology*
 - The people*
 - Effort*
 - And the product of the project*

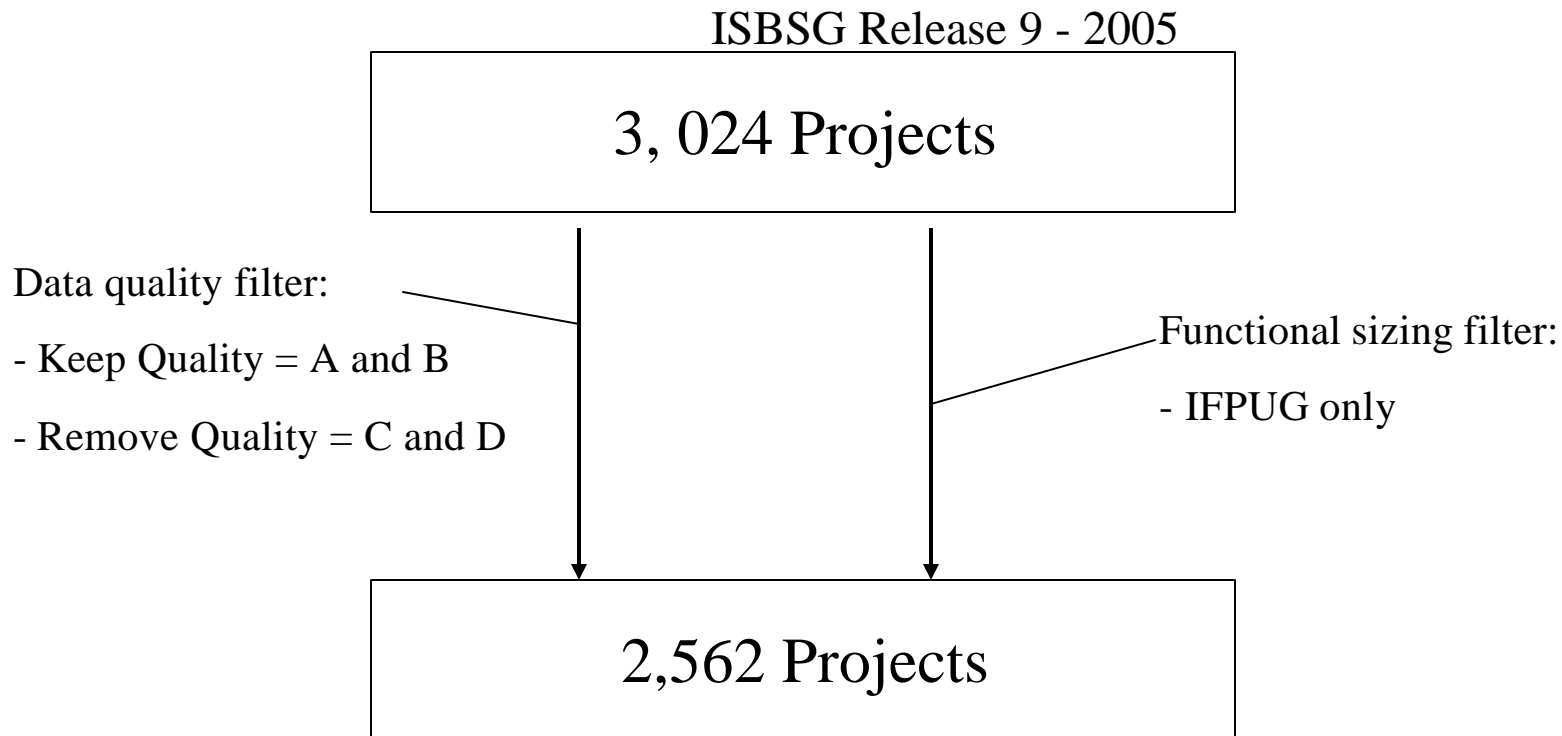


Consistency of Data in ISBSG

- Not all of 100 data fields have complete data for all for 3024 projects.*
- Not all projects have the same number of phases.*
- Project effort can not be compared amongst project that are composed of a different set of phases.*



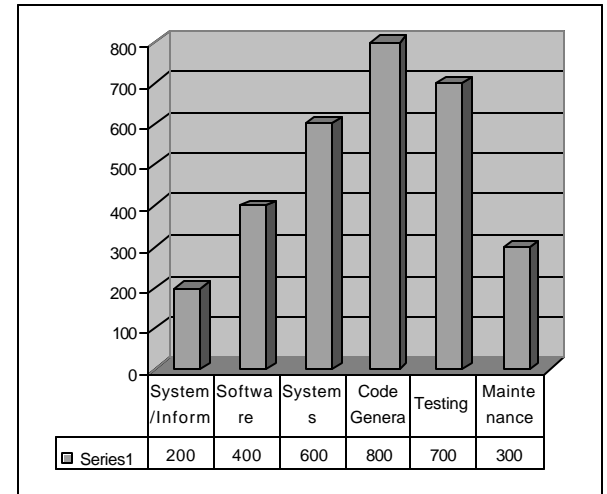
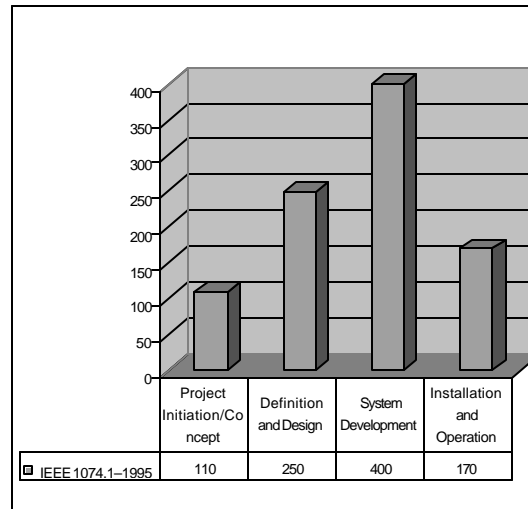
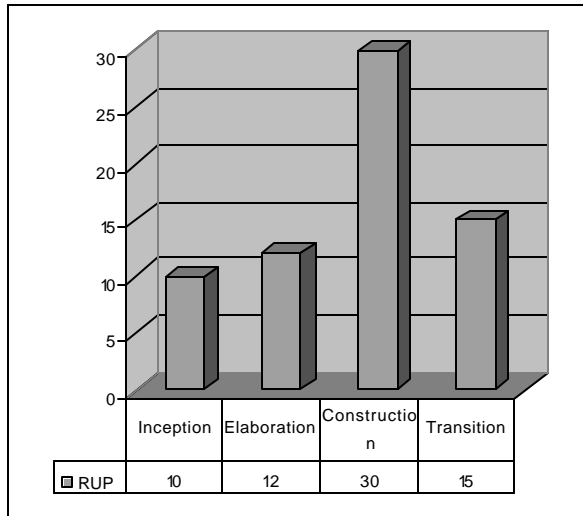
Initial ISBSG Data Preparation





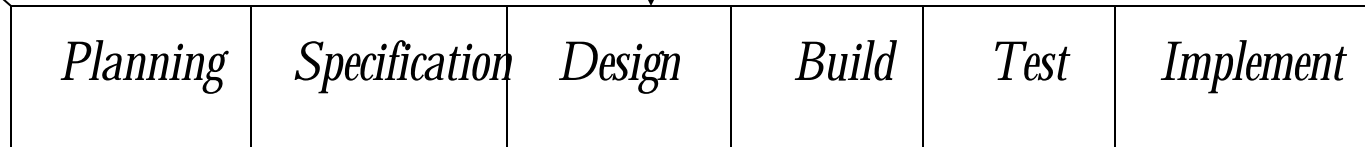
The ISBSG Project Phases

Industry



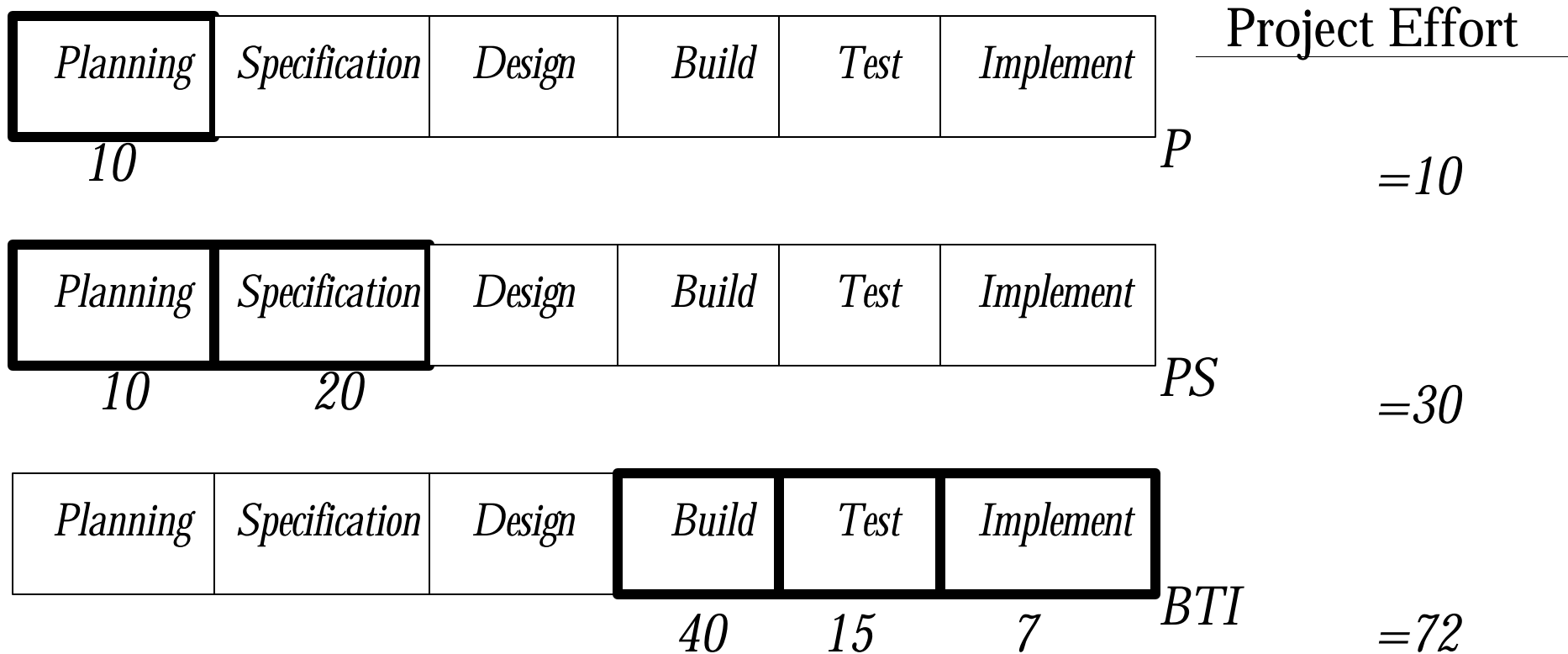
Mapping

ISBSG





The ISBSG Project Phases



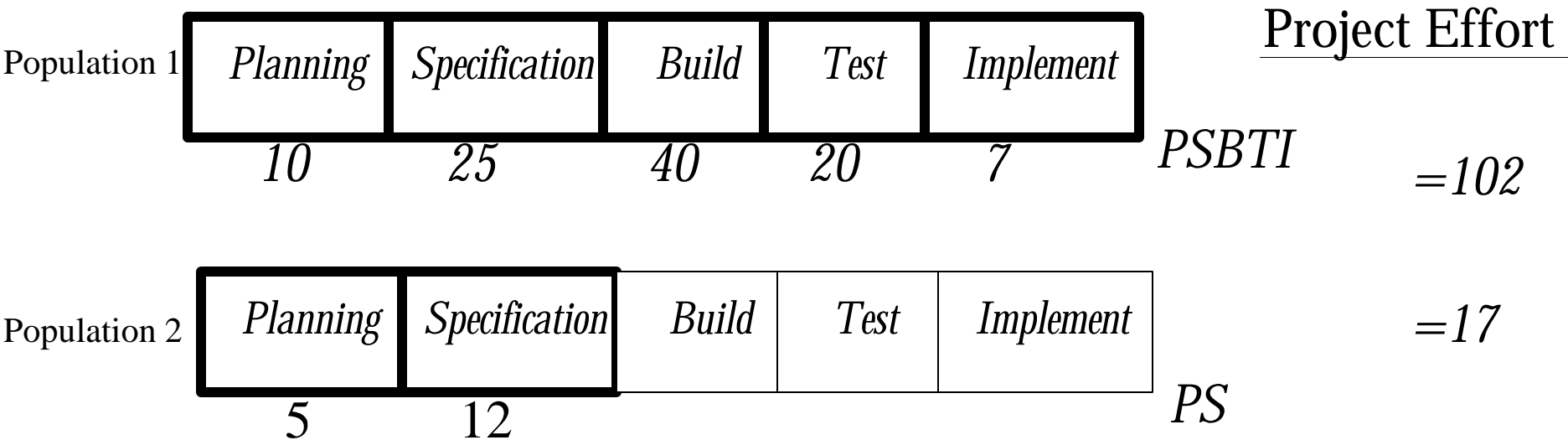


Observations

- ☞ *Comparing projects that have not the same project span will provide misleading conclusions*
- ☞ *A conversion process has to be found to compare projects with different project spans*
- ☞ *ISBSG has defined a predefined set of phases*
- ☞ *All projects in ISBSG have been fitted to those categories*



Extrapolation of Phases



☞ *What would be the total project effort of population 2 if phases BTI were part of the project?*

The data

Profile#	#pages	Planning	Specification	Design	Build	Test	Hybrid	Integration	Rollé	Cycle	Build	Total
1	11											
2	2											
3	19											
4	13											
5	1											
6	29											
7	5											
8	1			H/L								
9	3											
10	5											
11	20											
12	7											
13	64											
14	37											
15	3											
16	9											
17	68											
18	39											
19	405											
20	350											
21	1											
22	4											
23	5											
24	1											
25	1											
26	1											
27	0											
28	8											
29	3											
30	349											
31	92											
32	1006											
Total	2562											



Observations

- *Rows 1, 9, 10 and 27: a single phase effort*
- *Rows 2, 3, 11, 15, 16, 25 and 28 :a two phases effort*
- *Rows 7 (full life cycle) and 22: cover all 6 phases*



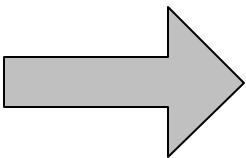
Observations

- ➡ Row 32 indicates that 1,006 projects do not have any indication of project phases*
- ➡ Rows that include “Design” have too few projects for statistical analysis (row 5 (1p), row 8 (1p), row 21 (1p), row 22 (4p))*



Prerequisite for analysis

☞ *In order to use the data for statistical analysis, at least two requirement must be met:*



- There must be enough historical data in order to have statistical value*
- The data must be homogeneous enough to provide meaningful interpretations*

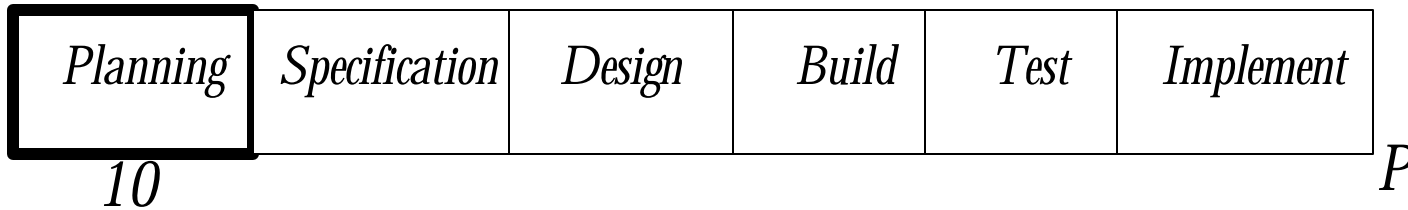


Some neighboring profile > 30 projects

- *Profile 19 has 405 projects and 4 phases (PSBT)*
- *Profile 20 has 350 projects and 5 phases (PSBTI)*
- *Profile 30 has 349 projects and 4 phases (SBTI)*

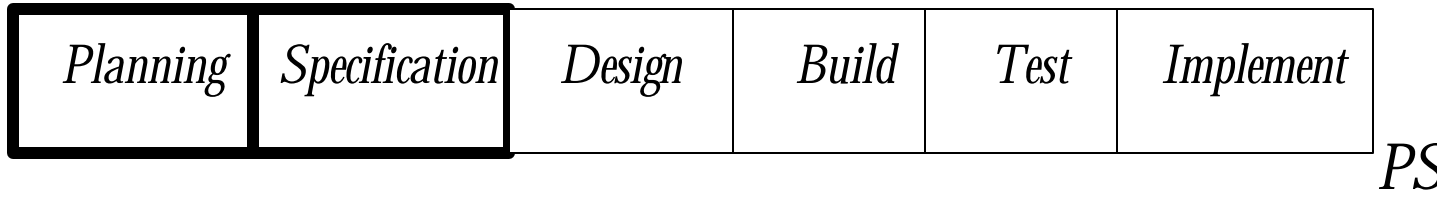


The ISBSG Project Phases



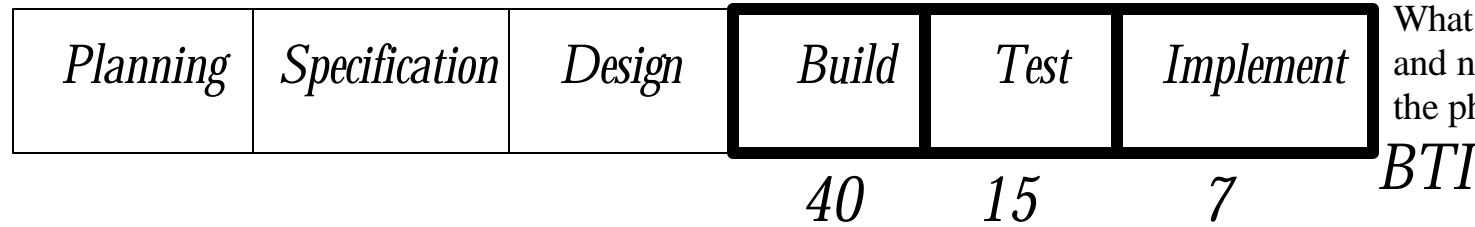
Project Effort

=20??



What if totals do not match with efforts per phase?

=30??



What if there is only a total and not efforts attributed to the phases of the profile?

=90?

?



Second Preparation

Table 1: PSBT-related profiles: with phase tags and effort data by phase

Project Profile	Number of projects		
	With phase tags	With detailed effort by phase	Phase effort consistent with Summary Effort
20: PSBTI	350	113	76
19: PSBT	405	200	100
31: SBTI	92	12	3



Prerequisite for analysis

- ☞ In order to use the data for statistical analysis, at least two requirements must be met:*
 - There must be enough historical data in order to have statistical value*
 - ➔ – The data must be homogeneous enough to provide meaningful interpretations*



Out of the 76 projects

Third level of data preparation

Average effort distribution for 34 projects for with very high Specification Effort

Profile 20: PSBTI

Phase	P	S	B	T	I
34 Projects	0,1%	98,5%	0,7%	0,5%	0,2%

Average effort distribution for the PSBTI (excluding outliers and unusual distribution)

Profile	No. of Projects	Project Phases - % Effort				
		P	S	B	T	I
PSBTI	41	9,1	24,7	39,1	19,7	7,3
PSBT	62	11,2	18,3	34,6	35,9	0
SBTI	3	0	27,6	49,0	15,3	8,1

Note: an additional outlier was removed: one with the greatest amount of effort but with no FP



Normalization of Effort

Profile percentage from ISBSG internal correspondence (referring to 2001 data)

	P	S	B	T	I	# Projects
PSBTI	9.3	19.4	47.4	16.5	7.3	47
PSBT	11.3	22.0	39.1	20.3	(7.3)	11
SBTI	(9.3)	15.1	46.6	23.2	5.7	9

Compared to the numbers obtained in the current paper with release 9 (2005) and 41 projects:

Profile	P	S	B	T	I
PSBTI	9,1	24,7	39,1	19,7	7,3



Normalization of Effort

: Statistical characteristics of profile samples

#	Profile	No. projects	Avg (Effort)	σ (effort)	Avg (size)	σ (size)
1	PSBTI	350	6379	12969	473	841
2	PSBTI	76	4417	5889	435	623
3	PSBT	405	4802	7557	544	981
4	PSBT	100	5311	6613	553	573
5	SBTI	92	7132	13793	352	539
6	SBTI	3	35042	50903	1121	191
7	Cont2-4-6	179	5705	10189	512	604

Building a new sample with the concatenation of neighbouring profiles.



Extensivity and range of ISBSG normalization

- ☞ ISBSG has normalized only 43% of the dataset looked at in this paper:*
 - 1,465 projects out of 2,562 projects initially qualified (57%) had the same value for the normalized effort and the summary work effort*
 - The normalization ratio used by ISBSG may vary within the same profile*



Concluding remarks

- ☞ The 100 data fields is a mixed blessing:*
 - It help group projects that have similar traits*
 - But not all fields are mandatory which leads to several projects with incomplete data: hence often too few projects to use statistical techniques*



Concluding remarks

- ☞ *ISBSG has attempted to tackle the issue of the variability of phases included in a project:*
 - *It has included a normalization process.*
 - *But not all fields are mandatory : when a discrepancy was observed, the data was discarded. However, the ISBSG data repository manager may keep these data based on additional information not made available to the general public.*
 - *It is surprising to see that the design phase is excluded in so many profiles.*



Thank You !



aabran@ele.etsmtl.ca, david.dery.1@ens.etsmtl.ca