PROBLEM STATEMENT

- TO BUILD A MODEL WHICH DEFINES THE PHYSICAL AND LOGICAL STRU TABLES
- THE MODEL IS USED TO DETECT TABLES AND TO ANALYZE AND DECOM DETECTED TABLES

EXISTING SOLUTIONS

- CURRENT SOLUTIONS ARE BUILT AROUND PDFs
- THEY STRUGGLE WITH IMAGES AND ARE ALSO NOT EFFIC
  UNDERSTANDING THE UNDERLYING TABLE STRUCTURE

CHALLENGE

BUILDING A MODEL WHICH DEALS WITH IMAGES AS EFFIC

WITH PDFs AND TO UNDERSTAND AND EXTRACT WIDE VAR

TABLE STRUCTURES

COMPLEXITY

NO BOUNDARIES

Your organisation's country of establishment (indicate your country of residence if answering as an individual person):

	replies	%
AT - Austria	79	(4.6%)
BE - Belgium	142	(8.2%)
DE - Germany	258	(14.9%)
DK - Denmark	28	(1.6%)
EL - Greece	80	(4.6%)
ES - Spain	123	(7.1%)
FI - Finland	22	(1.3%)
FR - France	132	(7.6%)
IE - Ireland	33	(1.9%)
IT - Italy	147	(8.5%)
LU - Luxembourg	1	(0.1%)
NL - Netherlands	89	(5.1%)
PT - Portugal	81	(4.7%)
SV - Sweden	68	(3.9%)
UK - United Kingdom	157	(9.1%)
CY - Cyprus	12	(0.7%)
CZ - Czech Republic	10	(0.6%)

PARTIAL BOUN

Table A Annual growth rates of debt securities issued by euro a

(percentages; end of period)

	2006	2007	2008
Total general government	2.5	2.8	8.1
Long-term	*3.4*	*2.3*	*3.7*
Fixed rate	3.4	2.0	3.5
Floating rate	3.4	5.4	5.1
Short-term	*-8.8*	*9.6*	*61.9*

Source: ECB.

Table B Structure of amounts outstanding of debt securities iss

(percentages of total debt securities issued by general government; end of period)

	2006	2007	2008
Long-term	*92.9*	*92.5*	*88.8*
Fixed rate	83.9	83.2	79.7
Floating rate	8.0	8.2	8.1
Short-term	*7.1*	*7.5*	*11.2*
Total general government in EUR billions	4,710.7	4,841.8	5,266.2

Source: ECB.

COMPLEXITY

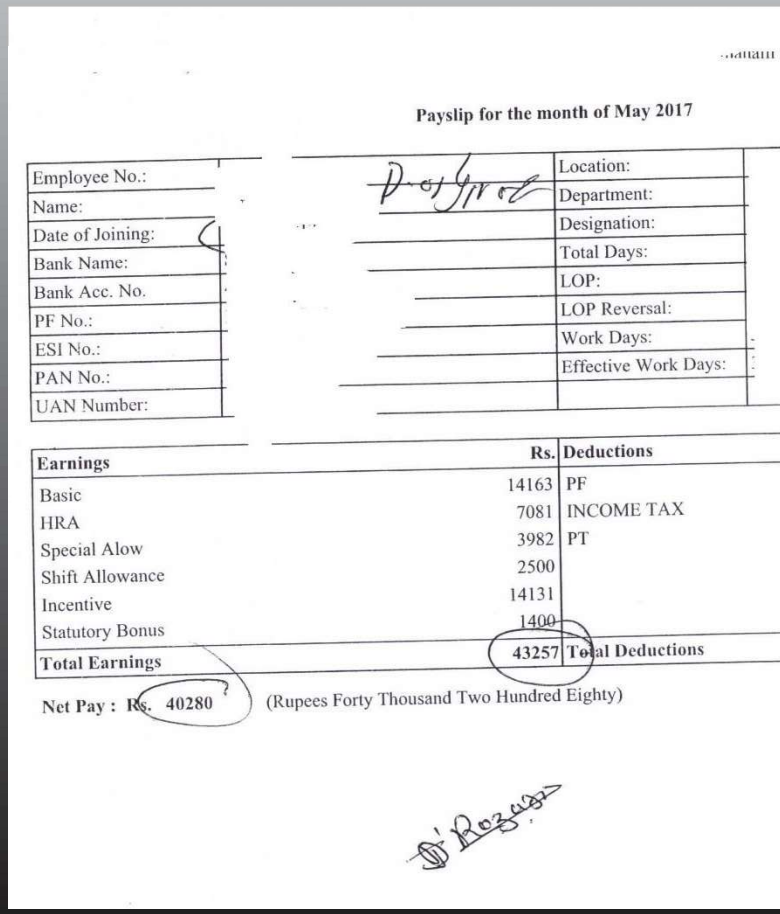
NESTED ROWS AND COLUMNS

Table 6
Correlations between the difficulties encounters and the perceived impacts on pupils, teachers and the school as a whole (Pearsons correlation coefficient*)

	Difficulties encountered with respect to		
Impacts on participating pupils	Lack of interest/ acceptance from colleagues	Lack of interest of pupils	Lack of interest of parents
Knowledge and awareness of different cultures	-0,1651	-0,2742	-0,1618
Foreign language competence	-0,0857	-0,1804	-0,1337
Social skills and abilities	-0,1237	-0,2328	-0,1473
Acquaintance of special knowledge	-0,1355	-0,2008	-0,1234
Self competence	-0,1291	-0,2466	-0,1636
Impacts on participating teachers			
Knowledge/appreciation of school system and education in the partner countries	-0,1505	-0,1636	-0,1349
Foreign language competence	-0,0545	-0,0997	-0,0519
Social skills and personal commitment	-0,2558	-0,2235	-0,1302
Professional knowledge and abilities	-0,2145	-0,2319	-0,1003
Impacts on the school as a whole			
European/International dimension of the school	-0,2438	-0,1945	-0,1030
School climate	-0,2976	-0,1810	-0,1012
Innovation in teaching and school management	-0,2586	-0,2557	-0,0928
Training of teachers	-0,1839	-0,1703	-0,0518
Involvement of external actors in the every day school-life	-0,2346	-0,2343	-0,2237
International mobility of pupils	**	**	-0,0583

* Significance p = 0,000
** No significant correlation

NOISY IMAGES



Payslip for the month of May 2017

Employee No.:		Location:	
Name:		Department:	
Date of Joining:		Designation:	
Bank Name:		Total Days:	
Bank Acc. No.		LOP:	
PF No.:		LOP Reversal:	
ESI No.:		Work Days:	
PAN No.:		Effective Work Days:	
UAN Number:			

Earnings	Rs.	Deductions
Basic	14163	PF
HRA	7081	INCOME TAX
Special Alow	3982	PT
Shift Allowance	2500	
Incentive	14131	
Statutory Bonus	1400	
Total Earnings	**43257**	**Total Deductions**

Net Pay : Rs. 40280 (Rupees Forty Thousand Two Hundred Eighty)

METHODOLOGY

PRE-PROCESSING

- SKEW-CORRECTION
- LINE EXTRACTION
- TEXT INVERSION

TABLE DETECTION

- BOUNDED FIGURES & TABLE DETECTION
- TEXT BLOCK FORMATION
- HORIZONTAL & VERTICAL ALIGNMENT VERIFICATION

TABLE EXTRACTION

- ROWS & COLUMNS IDENTIFICATION
- HEADERS' IDENTIFICATION
- MAPPING CELL'S ROW & COLUMN HEADER(S)

EXAMPLE

INPUT

- SI2.5: Country Net Yearly Salary Average in EURO

Country	Net Yearly Salary average
Austria	31.552
Belgium	25.408
Bulgaria	3.361
Croatia	12.477
Cyprus	33.865
Czech Republic	11.514
Denmark	35.874
Estonia	7.865
Finland	26.268
France	30.030
Germany	28.351
Greece	20.105
Hungary	9.423
Iceland	33.936
Ireland	34.930
Israel	26.703

Country	Net Yearly Salary average
Italy	21.821
Latvia	9.150
Lithuania	6.340
Luxembourg	45.847
Malta	19.688
Netherlands	36.791
Norway	39.407
Poland	8.059
Portugal	18.111
Romania	5.766
Slovakia	6.055
Slovenia	14.279
Spain	22.930
Sweden	28.075
Switzerland	64.123
Turkey	14.565
United Kingdom	37.001

Table 48 - Country Net Yearly Salary Average (2006, N=6.934, all currencies in EURO)

Note: A different analysis has been carried out for the net yearly salary costs obtained in the survey in order to detect unusual observations. The final sample had 7.018 correct replies, including 84 answers from Marie Curie fellowships. Marie Curie answers has not been considered for the calculation of the country net yearly salary average, as a result the sample had N=6.934.

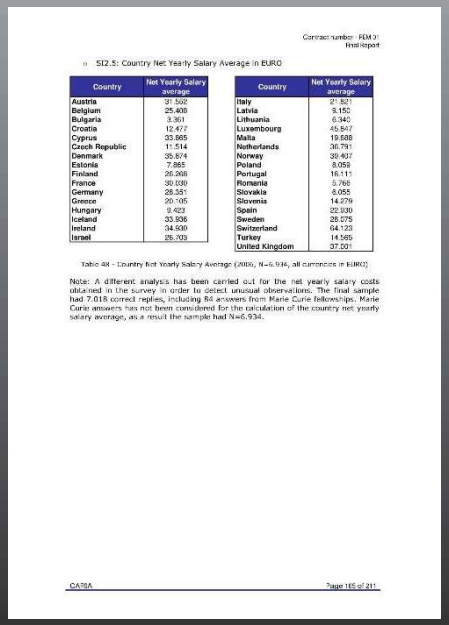
OUTPUT

[MAIN COLUMN HEADER/ROW HEADER/CO

	A	B	C	
1	1			
2	Country	Austria		Net Yearly S
3	Country	Belgium		Net Yearly S
4	Country	Bulgaria		Net Yearly S
5	Country	Croatia		Net Yearly S
6	Country	Cyprus		Net Yearly S
7	Country	Czech Republic		Net Yearly S
8	Country	Denmark		Net Yearly S
9	Country	Estonia		Net Yearly S
10	Country	Finland		Net Yearly S
11	Country	France		Net Yearly S
12	Country	Germany		Net Yearly S
13	Country	Greece		Net Yearly S
14	Country	Hungary		Net Yearly S
15	Country	Iceland		Net Yearly S
16	Country	Ireland		Net Yearly S
17	Country	Israel		Net Yearly S
18	2			
19	Country	Italy		Net Yearly S
20	Country	Latvia		Net Yearly S
21	Country	Lithuania		Net Yearly S
22	Country	Luxembourg		Net Yearly S
23	Country	Malta		Net Yearly S
24	Country	Netherlands		Net Yearly S
25	Country	Norway		Net Yearly S
26	Country	Poland		Net Yearly S
27	Country	Portugal		Net Yearly S
28	Country	Romania		Net Yearly S
29	Country	Slovakia		Net Yearly S
30	Country	Slovenia		Net Yearly S
31	Country	Spain		Net Yearly S
32	Country	Sweden		Net Yearly S
33	Country	Switzerland		Net Yearly S
34	Country	Turkey		Net Yearly S
35	Country	United Kingdom		Net Yearly S

PRE-PROCESSING

SKEW-CORRECTION → LINE DETECTION → LINE REMOVAL → NOISE REDUCTION →



Contract number - REM 01
Final Report

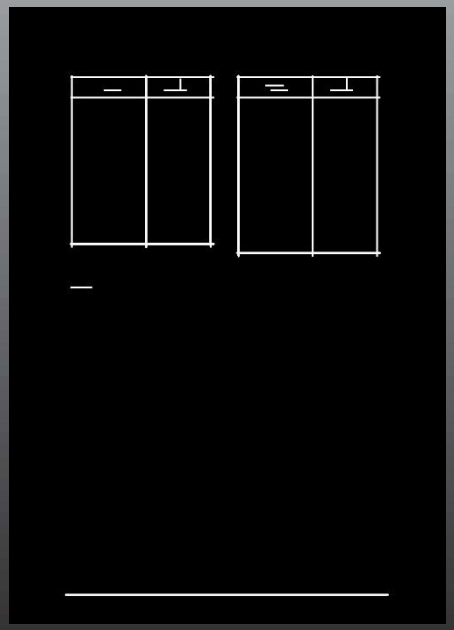
SI2.5: Country Net Yearly Salary Average in EURO

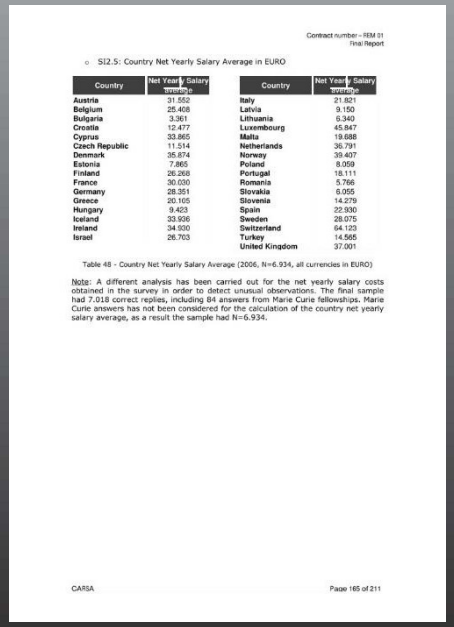
Country	Net Yearly Salary average	Country	Net Yearly Salary average
Austria	31.552	Italy	21.821
Belgium	25.408	Latvia	9.150
Bulgaria	3.361	Lithuania	6.340
Croatia	12.477	Luxembourg	45.847
Cyprus	33.865	Malta	19.688
Czech Republic	11.514	Netherlands	36.791
Denmark	35.874	Norway	39.407
Estonia	7.865	Poland	8.059
Finland	26.268	Portugal	18.111
France	30.030	Romania	5.766
Germany	28.351	Slovakia	6.055
Greece	20.105	Slovenia	14.279
Hungary	9.423	Spain	22.930
Iceland	33.936	Sweden	28.075
Ireland	34.930	Switzerland	64.123
Israel	26.703	Turkey	14.565
		United Kingdom	37.001

Table 48 - Country Net Yearly Salary Average (2006, N=6.934, all currencies in EURO)

Note: A different analysis has been carried out for the net yearly salary costs obtained in the survey in order to detect unusual observations. The final sample had 7.018 correct replies, including 84 answers from Marie Curie fellowships. Marie Curie answers has not been considered for the calculation of the country net yearly salary average, as a result the sample had N=6.934.

CARSA Page 165 of 211





Contract number - REM 01
Final Report

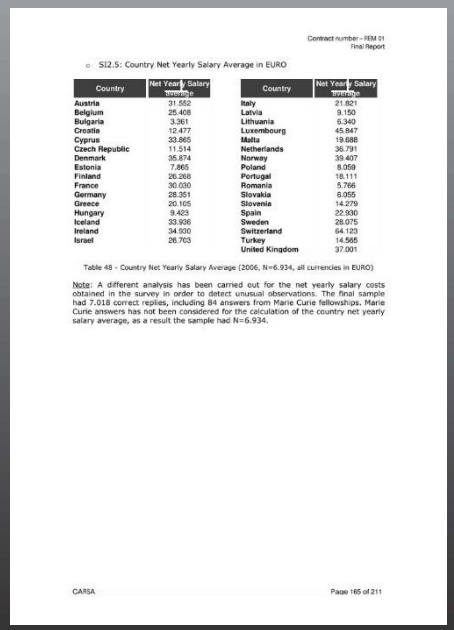
SI2.5: Country Net Yearly Salary Average in EURO

Country	Net Yearly Salary average	Country	Net Yearly Salary average
Austria	31.552	Italy	21.821
Belgium	25.408	Latvia	9.150
Bulgaria	3.361	Lithuania	6.340
Croatia	12.477	Luxembourg	45.847
Cyprus	33.865	Malta	19.688
Czech Republic	11.514	Netherlands	36.791
Denmark	35.874	Norway	39.407
Estonia	7.865	Poland	8.059
Finland	26.268	Portugal	18.111
France	30.030	Romania	5.766
Germany	28.351	Slovakia	6.055
Greece	20.105	Slovenia	14.279
Hungary	9.423	Spain	22.930
Iceland	33.936	Sweden	28.075
Ireland	34.930	Switzerland	64.123
Israel	26.703	Turkey	14.565
		United Kingdom	37.001

Table 48 - Country Net Yearly Salary Average (2006, N=6.934, all currencies in EURO)

Note: A different analysis has been carried out for the net yearly salary costs obtained in the survey in order to detect unusual observations. The final sample had 7.018 correct replies, including 84 answers from Marie Curie fellowships. Marie Curie answers has not been considered for the calculation of the country net yearly salary average, as a result the sample had N=6.934.

CARSA Page 165 of 211



Contract number - REM 01
Final Report

SI2.5: Country Net Yearly Salary Average in EURO

Country	Net Yearly Salary average	Country	Net Yearly Salary average
Austria	31.552	Italy	21.821
Belgium	25.408	Latvia	9.150
Bulgaria	3.361	Lithuania	6.340
Croatia	12.477	Luxembourg	45.847
Cyprus	33.865	Malta	19.688
Czech Republic	11.514	Netherlands	36.791
Denmark	35.874	Norway	39.407
Estonia	7.865	Poland	8.059
Finland	26.268	Portugal	18.111
France	30.030	Romania	5.766
Germany	28.351	Slovakia	6.055
Greece	20.105	Slovenia	14.279
Hungary	9.423	Spain	22.930
Iceland	33.936	Sweden	28.075
Ireland	34.930	Switzerland	64.123
Israel	26.703	Turkey	14.565
		United Kingdom	37.001

Table 48 - Country Net Yearly Salary Average (2006, N=6.934, all currencies in EURO)

Note: A different analysis has been carried out for the net yearly salary costs obtained in the survey in order to detect unusual observations. The final sample had 7.018 correct replies, including 84 answers from Marie Curie fellowships. Marie Curie answers has not been considered for the calculation of the country net yearly salary average, as a result the sample had N=6.934.

CARSA Page 165 of 211

TABLE DETECTION

BOUNDED FIGURES & TABLE DETECTION

BLOCK FORMATION

HORIZONTAL & VERTICAL ALIGNMENT VERIFICATION

Country	Net Yearly Salary average
Austria	31.552
Belgium	25.408
Bulgaria	3.361
Croatia	12.477
Cyprus	33.865
Czech Republic	11.514
Denmark	35.874
Estonia	7.865
Finland	26.268
France	30.030
Germany	28.351
Greece	20.105
Hungary	9.423
Iceland	33.936
Ireland	34.930
Israel	26.703

Country	Net Yearly Salary average
Italy	21.821
Latvia	9.150
Lithuania	6.340
Luxembourg	45.847
Malta	19.688
Netherlands	36.791
Norway	39.407
Poland	8.059
Portugal	18.111
Romania	5.766
Slovakia	6.055
Slovenia	14.279
Spain	22.930
Sweden	28.075
Switzerland	64.123
Turkey	14.565
United Kingdom	37.001

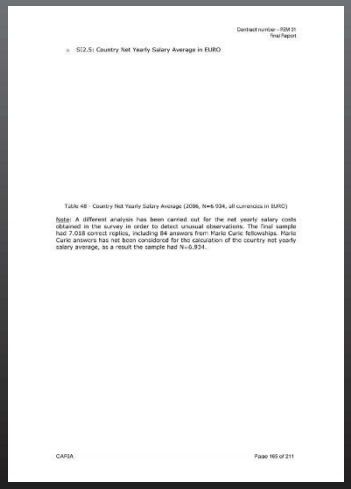
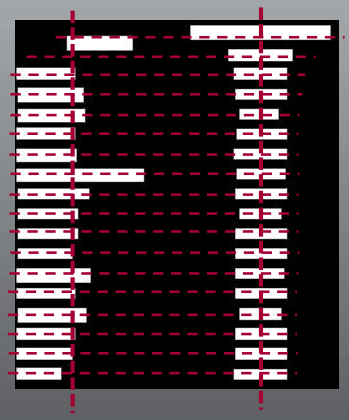


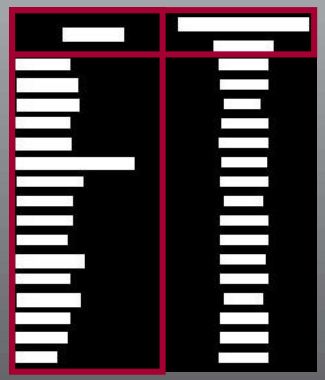
TABLE EXTRACTION

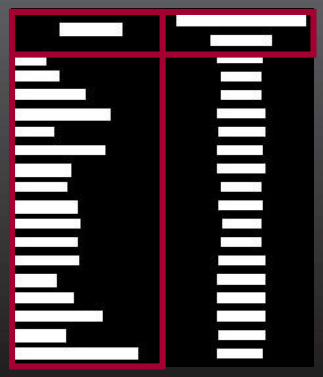
ROWS & COLUMNS IDENTIFICATION

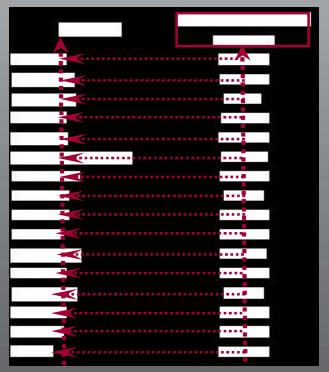
HEADER IDENTIFICATION

MAPPING CELL'S ROW & COLUMN HEADER(S)









CHALLENGES

NESTED HEADER MAPPING

Spain					
Scientific domain	Total Yearly Salary Costs of Researchers (data from the study)		Equivalent(s) profession(s) defined by ISCO classification	Category with available data in Eurostat	Total Year similar p fro
	Male	Female			Male
Social and Human Sciences	27.301 €	16.806 €	2.4.4	2.4	46.657 €
Economics	42.978 €	33.669 €	2.4.1	2.4	46.657 €
Chemistry	39.091 €	20.296 €	2.1.1	2.1	42.138 €
Physics	27.197 €	16.067 €	2.1.1	2.1	42.138 €
Life Sciences	36.523 €	22.139 €	2.2.1	2.2	37.111 €
Mathematics	45.389 €	41.107 €	2.1.2	2.1	42.138 €
Information Sciences	26.059 €	25.893 €	2.1.3	2.1	42.138 €
Engineering Sciences	34.316 €	25.435 €	2.1.4	2.1	42.138 €
Environment and Geosciences	29.210 €	14.847 €	2.1.4	2.1	42.138 €

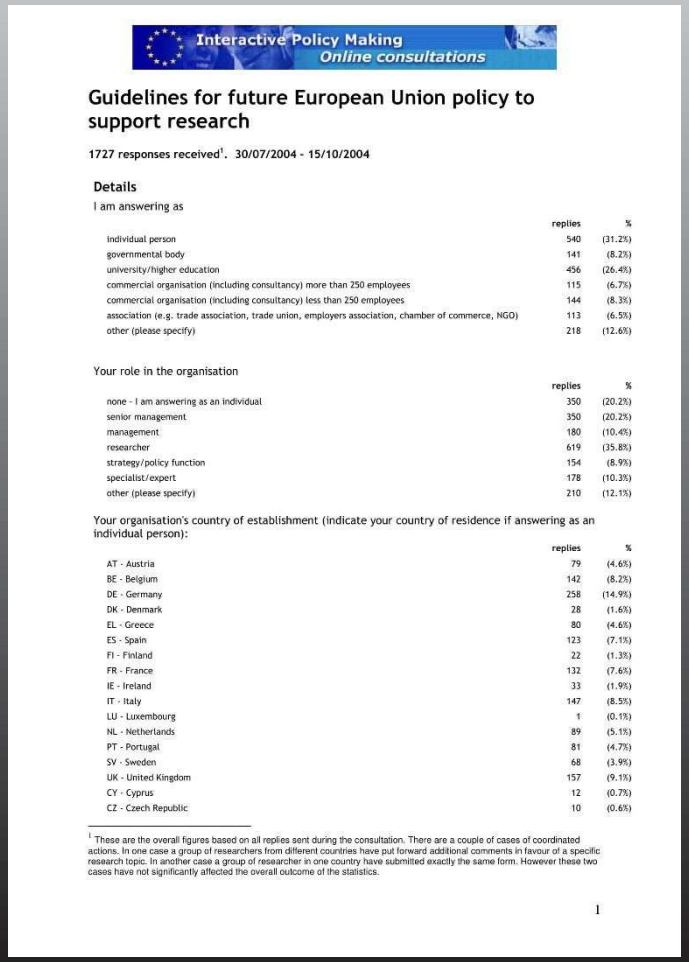
CHALLENGES

REPETITIVE COLUMN/ROW HEADERS

Country	Number of responses	% of total responses	Country	Number of responses	% of total responses
DE - Germany	258	14.9%	CH - Switzerland	25	1.4%
UK - United Kingdom	157	9.1%	FI - Finland	22	1.3%
IT - Italy	147	8.5%	LT - Lithuania	22	1.3%
BE - Belgium	142	8.2%	IL - Israel	17	1.0%
FR - France	132	7.6%	PL - Poland	16	0.9%
ES - Spain	123	7.1%	CY - Cyprus	12	0.7%
NL - Netherlands	89	5.2%	CZ - Czech Republic	10	0.6%
PT - Portugal	81	4.7%	EE - Estonia	10	0.6%
EL - Greece	80	4.6%	HU - Hungary	10	0.6%
AT - Austria	79	4.6%	BG - Bulgaria	8	0.5%
SV - Sweden	68	3.9%	SK - Slovak Republic	7	0.4%
TR - Turkey	44	2.5%	SL - Slovenia	6	0.3%
IE - Ireland	33	1.9%	MT - Malta	5	0.3%
NO - Norway	31	1.8%	LV - Latvia	4	0.2%
Other country	30	1.7%	LU - Luxembourg	1	0.1%
RO - Romania	29	1.7%	IS - Iceland	1	0.1%
DK - Denmark	28	1.6%	*TOTAL*	*1727*	*100%*

CHALLENGES

TABLE SPLITTING OVER MULTIPLE PAGES



Interactive Policy Making
Online consultations

Guidelines for future European Union policy to support research

1727 responses received[1]. 30/07/2004 - 15/10/2004

Details

I am answering as

	replies	%
individual person	540	(31.2%)
governmental body	141	(8.2%)
university/higher education	456	(26.4%)
commercial organisation (including consultancy) more than 250 employees	115	(6.7%)
commercial organisation (including consultancy) less than 250 employees	144	(8.3%)
association (e.g. trade association, trade union, employers association, chamber of commerce, NGO)	113	(6.5%)
other (please specify)	218	(12.6%)

Your role in the organisation

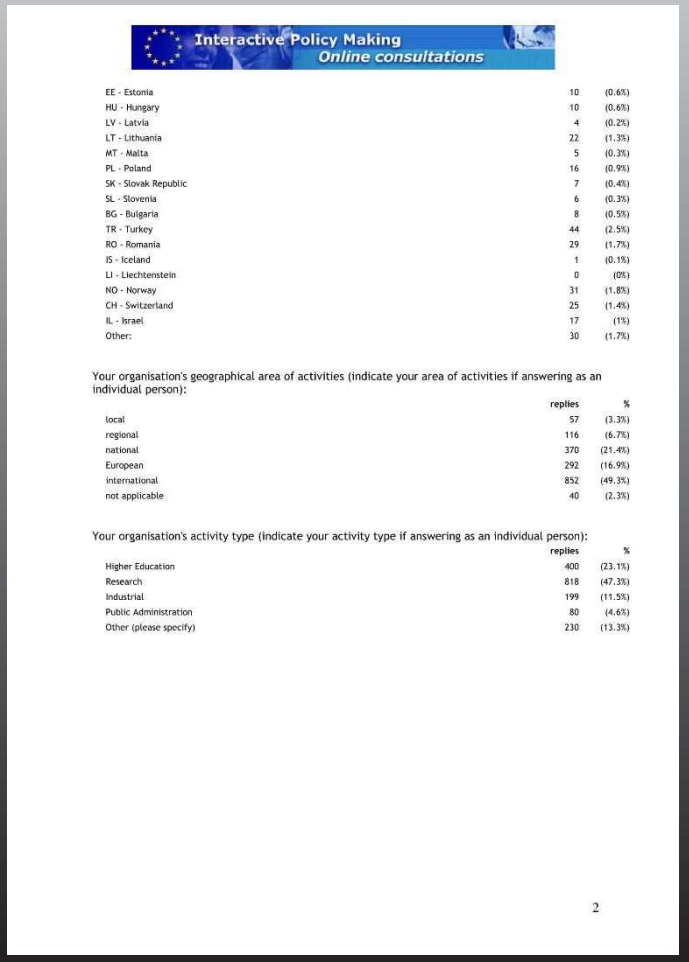
	replies	%
none - I am answering as an individual	350	(20.2%)
senior management	350	(20.2%)
management	180	(10.4%)
researcher	619	(35.8%)
strategy/policy function	154	(8.9%)
specialist/expert	178	(10.3%)
other (please specify)	210	(12.1%)

Your organisation's country of establishment (indicate your country of residence if answering as an individual person):

	replies	%
AT - Austria	79	(4.6%)
BE - Belgium	142	(8.2%)
DE - Germany	258	(14.9%)
DK - Denmark	28	(1.6%)
EL - Greece	80	(4.6%)
ES - Spain	123	(7.1%)
FI - Finland	22	(1.3%)
FR - France	132	(7.6%)
IE - Ireland	33	(1.9%)
IT - Italy	147	(8.5%)
LU - Luxembourg	1	(0.1%)
NL - Netherlands	89	(5.1%)
PT - Portugal	81	(4.7%)
SV - Sweden	68	(3.9%)
UK - United Kingdom	157	(9.1%)
CY - Cyprus	12	(0.7%)
CZ - Czech Republic	10	(0.6%)

[1] These are the overall figures based on all replies sent during the consultation. There are a couple of cases of coordinated actions. In one case a group of researchers from different countries have put forward additional comments in favour of a specific research topic. In another case a group of researcher in one country have submitted exactly the same form. However these two cases have not significantly affected the overall outcome of the statistics.

1



Interactive Policy Making
Online consultations

EE - Estonia	10	(0.6%)
HU - Hungary	10	(0.6%)
LV - Latvia	4	(0.2%)
LT - Lithuania	22	(1.3%)
MT - Malta	5	(0.3%)
PL - Poland	16	(0.9%)
SK - Slovak Republic	7	(0.4%)
SL - Slovenia	6	(0.3%)
BG - Bulgaria	8	(0.5%)
TR - Turkey	44	(2.5%)
RO - Romania	29	(1.7%)
IS - Iceland	1	(0.1%)
LI - Liechtenstein	0	(0%)
NO - Norway	31	(1.8%)
CH - Switzerland	25	(1.4%)
IL - Israel	17	(1%)
Other:	30	(1.7%)

Your organisation's geographical area of activities (indicate your area of activities if answering as an individual person):

	replies	%
local	57	(3.3%)
regional	116	(6.7%)
national	370	(21.4%)
European	292	(16.9%)
international	852	(49.3%)
not applicable	40	(2.3%)

Your organisation's activity type (indicate your activity type if answering as an individual person):

	replies	%
Higher Education	400	(23.1%)
Research	818	(47.3%)
Industrial	199	(11.5%)
Public Administration	80	(4.6%)
Other (please specify)	230	(13.3%)

2

CHALLENGES

NOISY SCANNED DATA