ANNEX B5 – 4 COMPARATIVE TABLE FOR RESEARCH SERVICES IN DENMARK AND THE NETHERLAND

SUBJECT	DENMARK	THE NEDERLAND
GENEREL The following institutions may be granted access to microdata	 Departments at Danish Universities Public Institutes for scientific research Danish Ministries (for policy advice or policy analysis) Danish research organizations as a part of a non-profit foundation Non-governmental organisations Consultancy firms Enterprises. However, single enterprises cannot have access to micro data with enterprise data 	 Dutch universities Institutes for scientific research. Organisations for policy advice or policy analysis. Statistical authorities in other EU countries. Other research institutions authorised to work with the microdata.
Definition/size of an institution	• Institutions authorized at Department level	• Institutions authorized at University level. Might be changed to department level in the future
Number of authorized institutions	• ~500 authorized institutions. Foreign researcher institutions cannot be authorized but can be granted access through an authorized Danish institution	• ~130 authorized institutions Include 30 foreign research institutions from 10 countries

Central unit for research services (RS)	 YES RS: 16 academics (generalists), 1 IT-specialists, 1 secretary Hours for consulting: 09:00-15:00. The Data Service Centre: 2 employees (Programmer) IT service desk: 2 employees Hours for technical support from IT service desk: 08:00-16:00. 	 YES RS: 16 people, 7 of them with an academic degree Hours for consulting: 08:00- 17:00 The Data Service Centre: 8 employees Hours for technical support: ?. 08:00- 17:00
Number of researchers and project	 2000 researchers 1200 research projects	530 researchers700 projects
Dataset available for research	 260 unique files names 4640 dataset >12.000 unique variables Each year around x files are placed in the secured environment 	 441 unique files names 3416 dataset x variables Each year around 400-600 files are placed in the secured environment
Mode of access	Remote access	 <i>Remote access</i> <i>On site access – mainly used for cause of death data and enterprise data. Might be closed down in the future</i>

Nork-flow in brief	Institutional authorisation	<u>Institutional authorisation online</u>
	Applications. No standard application	• <u>Project request</u>
	form. Does in most cases involve a meeting	• Intake meeting with RS and subject units -
	by phone or in person	mandatory
	• <u>Project request</u>	Contract
	• Intake meeting with RS and subject units if	• Confidentiality statement
	needed/on request	• Instalation of Fingerprint Box – New users
	• Selection of specific variables - Data	• Confidentiality test need to be passed
	extraction protocol	• Access granted to views of full datasets
	Contract	• Log-on to the research server – currently 3
	Confidentiality statement	factort authentication. Future authentication -
	• Access granted to costomized copies of	factor authentication
	dataset placed on a research server	Analysis
	• Log-on to the research server (3 factor	• <i>Getting output from analysis</i>
	authentication)	• Output received after max 3 days if
	Analysis	accepted
	• <i>Getting output from analysis</i>	• Output control by the RS by two
	• Output received instantly	independent controller per output
	• Sample output control by RS (one conroller per output)	

The need to know principle	• Need to know principle apply	• Need to know principle apply
	 Access to secure use files 	 Access to secure use files
	 Access customized populations after the need 	Access full population
	to know principles	Access full datasets – datasets contained
	• Access selected variables after the need to know principles	usually around 10 variables
Data base for extracting data	• Contain all data produced by DST	• Contain all data produced by CBS
producing data views	• Names removed and addresses replaced by numbers	• All direct identifier de-dentified
	• Data with direct identifier such as peronel ID (CPR) - might be changed to in the future to de-identified numbers	
De-identification of research	Secure use files	• Secure use files
data	• All direct identifier de-identified	• All direct identifier de-identified
	Project specifics keys	• Same key used for all projects and for the CBS database
Disclosure control before data delivery	• In most cases no disclosure control exept for prescription medicine where aggregation is required by a special law for prescription medicine	• In most cases no disclosure control
Data format	• Usually SAS datasets but data can be	• Data views (Multiple views).
	<i>delivered in any given format (Mulitiple datasets).</i>	• Tools to transpose microdata to SPSS by batch
	• Researcher can transfer data to any given format themself	
Confidentially contracts	• With research environments (click)	• With research environments (click)

<i>Responsible for confidentiality output from RS</i>	 With researchers (<u>click</u>) The Research institution The researcher RS provide guidance 	 With researchers (click) Shared responsibility of CBS and the researcher RS provide guidance
Output control	 Output received instantly Automated warning system checking for specified patterns Sample output controlled <u>after</u> sent home Sample output controlled by the RS – linspector per sample output The option/right to send output home can be restricted to researchers approved by the Institution <u>Guidelines</u> Minimum 3 units per cell. Special rules apply for buisness data (Dominance rule) No requirement for description of output Approx. number of output per day: 1000 Costs: Free of charge Limitation of file type: Yes 	 Option of output light and normal output Output received - same day for output light and after max 3 daysfor normal output. Output controlled <u>before</u> sent home. CBS is currently testing the options of checking after for experienced users Output control led by the RS – 2 inspectors per output (Changing pairs of checkers) <u>Guidelines</u> Each output need to be accompanied by a description of the output (Standard form). For standard form please consult the output <u>Guidelines</u> Appendix 2 Minimum 10 units per cell. For large table t-Argus is used to for confidientiality risk Approx. number of output per day: 3-4 Costs: € 200 for a normal output Limitation of file type: Yes

Sanction in case of breach	 Sanctions against both the entire institution (Department level) as well for the researcher Exclusion for at least a month for all researchers associated with the institutions. Written explanation as well as plan for future policy requested In worst cases – legal penalty (Never occurred) 	 In case a researchers output in not in alignment with the confidentiality rules the researchers is contacted in order to make sure they get a full understanding of the confidentiality rules In worst cases – legal penalty (Never occurred)
Training in confidientiality rules	• On demand	• Mandatory –About five courses of 2-3 days offered per year
Tests	• No	 Initial on-line testwith multiple questions Random popup tests while working on the server - appr. every 30 min. In case of wrong aswers new questions will pop up. CBS has a pool of about 300 questions for tests

IT TECHNICAL AND			
SECURITY SET UP			
Technical setup	Physical research servers	•	Virtual research servers
	Encrypt SSL/VPN connection	•	28 servers
	Windows environment	•	1 SAS server
	• 2 SAS servers	•	12 POWER servers
	• 2 Stata servers	•	Each server is 4 CPU
	• 2 SPSS server	•	Two technologies: virtual technology is
	• 1 GAUSS server		VMWARE and CITRIX for connection
	• 2 RDP frontend servers		technology- CITRIX was chosen for security and
	• 1 WPS server		operational reasons. The other CITRIX they use
	• 2 server for other applications like R		is in order for the research staff to access their
	• NAS for data 180 TB		remote desktop at home.
	• Researchers connection through TS		
	• Operational system is based on Microsoft		
Security	• When establishing the system- hired external risk	٠	Separated from the production network
	analytics that checked the system	•	Blocked from the internet
	• Separated from the production network	•	Current authentication - 3 factort authentication
	• Blocked from the internet		Fingerprint (Box) Personnal password and SMS
	• 2 factort authentication by Username, Personnal four		Token/physical token
	digit code, password and SMS Token/physical token	•	Future authentication - 2 factor authentication:
	• Multiple firewalls		Token, SMS-code, VPN
	• The researchers can't print the environment	•	Multiple firewalls
	• SSL/VPN to encrypt traffic over the Internet	•	The researchers can't print the environment
	• Keep the ISO27001 but only partially. They aren't	•	Data security- 3 rd generation (דור שלישי). 1 st
	obligated to the full ISO27001 standard- want to get		generation- until 2008; 2 nd generation- 2016-
	certified in specific areas		2017; 3 rd generation- in pre-test stage, should be
	• Data Security - light data security level: segmentation,		implemented by the end of 2017.

	 anti-virus, SIEM (logpoint. DST don't have FW DB) All outputs are logged and stored on a server 	• CBS check applications that are installed in the environment- for malicious content.
	 The researchers can choose the email they want to send the outputs to- including different emails every time 	• The researchers can't print from the environment
	• The monitoring is done after the output was sent.	
	• The system can check if there is the same type of output sent a few times and give a warning.	
	• There is no limit of number of outputs	
	• There is a list of types of outputs that can be sent home.	
	• Log management- an external company looking into the log files and notifying in case of suspicious activity(during working hours only)	
Directory structure and	Rawdata- read only	• Views
projects at the researcher	• Workdata- new folders can be made per the researcher	Personel folders
environment	themself	• x
	• Metadata	

Software provided on	• Operating system	• Operating system
ervers	• Windows Server 2008 R2 Enterprise, Microsoft® Windows	• x
	Version 6.1 (7601, Service pack 1), x64-based (64-bit)	• Statistical package
	Statistical package	• SPSS
	• SAS 9.4	• STATA ?
	• STATA/MP 14.2 for Windows	• <i>R</i> ?
	• IBM® SPSS® Statistics Version 22	• Gauss
	• WPS Workbench	• MLWin
	• Gauss	• Ox GiveWin
	• <i>R</i>	• SAS Base
	• <i>aML</i> , version 2.09	• Micro-soft and word processing programs
	• GAMS	• Microsoft ®
	LatentGOLD®	• x
	• Mplus Version 6.12 Demo	• Other program
	• SCD/DIGRAM	
	SPSSToDigram	
	• PLINK! v1.07	
	• Micro-soft and word processing programs	
	• Microsoft ®	
	• GNU Emacs, version 24.3.1	
	• gVim - VIM - Vi IMproved, version 7.1	
	• Tinn-R, version 4.0.3.5	
	• MiKTeX TeX	
	• Adobe ® Reader, version 9.1.0	
	• Other program	
	• Stat/Transfer	
	• WinZip®	

Supporting tools on the servers	 Acess to documentation and metadata Codelist (Format library) 	• ?
General IT Information	 Researchers don't need to download any software on their PC in order to use the system 2000 users on 1200 projects. There can be up to 300 simulations users on the remote access SAS license is unlimited Researchers can convert files to other file formats The researchers have no limit in storage space for data but pay if their storage exceed 5 Gigabyte (GB) There is a common internal databank of all registers from which the research unit creates files for researchers IT security group that consists of: head of IT, operations and IT security coordinator Every project has its own environement 	 900 users but only about 600 active researchers on 700 projects. There can be up to 150-200 simultaneous users on the remote access There can be 4 users using the same virtual server simultaneously. Researchers cannot convert files Each researcher receives 50GB of space for data they pay a set fee for space (see booklet) They have a license of 500 concurrent sessions Every project has its own environement Microdata is stored in flat files and they can be accessed in flat files in SPSS format.
Management of the IT system	 Switching hardware every 3-4 years Each researcher has at least one AD Cluster system that enables dealing with problems without shutting down the whole system Disaster Recovery is partial- not for all systems 70 servers bougth by research institutions and hosted by DST Troubleshooting- eg deleting or finding folder through the log. Logs are protected from deleting and changes 	• Disaster Recovery is full – for all systems and is placed in a secure remote environment.

Monitoring of the IT system	 Monthly performance reports Proactive intervention if reports show any problems SLA on performance counters CPU, 85% of all Available memory Available diskspace on workspaces Disk queue length on SAS workspace Runtime for predefined standard job Knowledge of how many researchers are working on the system. There is no monitoring system that records key-strokes of the researchers 	 Import and export folders- data sent to researchers. There is an automated process that makes sure that if a researcher places something in the export, the CBS will get and alert and they will d a disclosure check. Same goes for the research import folders.
	 Oracle- Within the databases, there is a log of all the inquiries that go 6 months back SAS- there is a log and an alarm if someone looks for a specific ID number 	
Administrative system for supporting research services	 SLA- service level agreement between the research department and the IT department. The IT provide the RS with quarterly reports on activity 	• ?
Services provided to researchers by the IT Department	 2 full time employees and 1 part time employee Hours for technical support- help desk: 8:00- 16:00. After office hours support is provided for servers. Support is given through teamviewer. The IT team can log in from remote access If researchers request specific applications to be installed the IT will look into the software carefully before approving 	 1 full time employee as a manager for IT support assisted by other IT employees for technical support Hours for technical support: 8:00-17:00 If researchers request specific applications to be installed the IT will look into the software carefully before approving

RESPONSIBILITY AND COOPERATION		
Production of data	• Subject unit	• Subject unit
Production of metadata	 Subject unit Subject unit, researchers and RS for High quality quality documentation 	• Subject unit
Management of metadata	• Subject unit	• Data service centre (DSC) – cooperative repository for storing microdata and metadata
Management of data catalog	Research Service Unit	• Data service centre (DSC) – cooperative repository for storing ad sharing statistical microdata and metadata
Management of data library for the researchers	Research Service Unit	• Data service centre (DSC) – cooperative repository for storing ad sharing statistical microdata and metadata
Advise/guidance	 Research Service Unit Subject unit if needed - before, during and and after the research is performed IT help desk on technical matters 	 Research Service Unit Subject unit at intake meeting Subject unit if needed during and after the research is performed Data service centre on technical matters

GUIDELINES AND TRAINING		
Step-by step procedure	• Can be found on-line	• Can be found on-line
Application forms	 Guidance and forms can be found on-line. Please click <u>here</u> for further details 	 Guidance and forms can be found on-line Please click <u>here</u> for further details
Data catalog	 On the internet Please click <u>here</u> for further details Interactive catalog with link to documentation 	 On the internet Please click <u>here</u> for further details Pdf data catalog
Documentation and metadata	 On the internet Please click <u>here</u> for further details Interactive metadata with linkage between datacatalog, register documentation, varable documentation and code lists Language: For registers: Danish and English. For variables: Dansih Special documentation for researchers: Yes – with special focus on time series 	 Metadata as Pdf (Contain datacatalog, register documentation, varable documentation and code lists) Please click <u>here</u> for further details Language: Dutch Special documentation for researchers:?
<i>Guidance for output</i>	 On the internet. Please click <u>here</u> for further details Courses on demands 	 On the internet. Please click <u>here</u> for further details Mandatory on-line course for all new users
Courses	• Occationally thematic courses for researchers (e.g. in labour market statistics, health statistics, statistics methodology, secirurity etc)	• Five courses for researchers per year in disclosure control (2-3 days)

BUISNESS PLAN AND PRICE STRUCTURE

SIRUCIURE	
All services are based on hourly rates (about $\in 150$ for public Inst and	All sevices are paid by the researchers
220 for private inst). Parts is finaced by a yearly grant from Danish	• Services prior to the project start-up
Agency for Science and Higher Education the rest is paid by the	◦ Feasibility study of available data. (Hourly rate: $€ 100$)
researhcers	• Services during the project start-up
• Services prior and during the project start-up	 Basic start fee (€ 1600)
 Consulting and meetings 	 Costs per dataset topic (€ 160)
o Authorization of new researcher institutions and	• Authorization of new researcher (\notin 375)
researchers	 Call out fee (€ 530)
 Cost for construction of population(s) 	○ Fingerprint card ($€ 150$)
 Costs for customizing dataset 	o Importing own micro-data (Between € 210-1300
 Importing external data 	depending on complexibility)
\circ Setting up new users, creation of password,	\circ Use of software (SAS € 560 per months, other software
administration	€ 40-60 per months)
Services during an ongoing research	• Services during an ongoing research
 Consulting 	◦ Consulting (€ 80 per month per researcher and € 16 per
• Output checking	month per dataset, for extensive questions € 100 per
• Update and adding datasets to a research project	hour)
Remote Access facility services	• Output checking (€ 200 per output)
• Storage (First 5 GB free paid by the yearly grant then	• Adding datasets to a research project (€ 160 per
€ 2 per GB per quarter).	dataset)
• Use of the remote access system	• Extra storage capacity (\notin 25 per month per 50 GB)
Public Inst. – through the yearly grant	Remote Access facility services
Private Inst. € 25 per log-on per day ("log-on fee")	○ Installation of Remote Access computer (€ 1500)
 Options of buying their own servers 	◦ Service contract RA-PC (€165 per month)
• Use of software - financed trough the yearly grant for	Administrations, maitanance and development
public inst. and part of the log-on fee for private Inst.	• Paid by the fees obtained from the researchers
• Creation of new users, resetting passwords, IT related	
consulting, administration (Paid by the yearly grant)	
Administrations, maitanance and development	
• Paid by the yearly grant, DST and from earnings from	
the researchers fees	