

Access to Microdata: On Site and Remote Access

Leo Engberts, Consultation and Information Services
Department



Datasets – Catalogue files I

- Uniformity
- SPSS datafiles
- Full labeling
- File per subject / period
- All files linkable (same key)
- No direct identifiers
- Access to entire data file
- With highly sensitive data
- 3700 databases available, incl CBS surveys and Government registers



Datasets – Customer datafiles

- Customer brings own datafile
(at the moment 12.000 files are imported)
- CvB prepares datafile
 - Adding the key for linking
 - Removal of identifiers
 - Labelling and documentation (“light”)

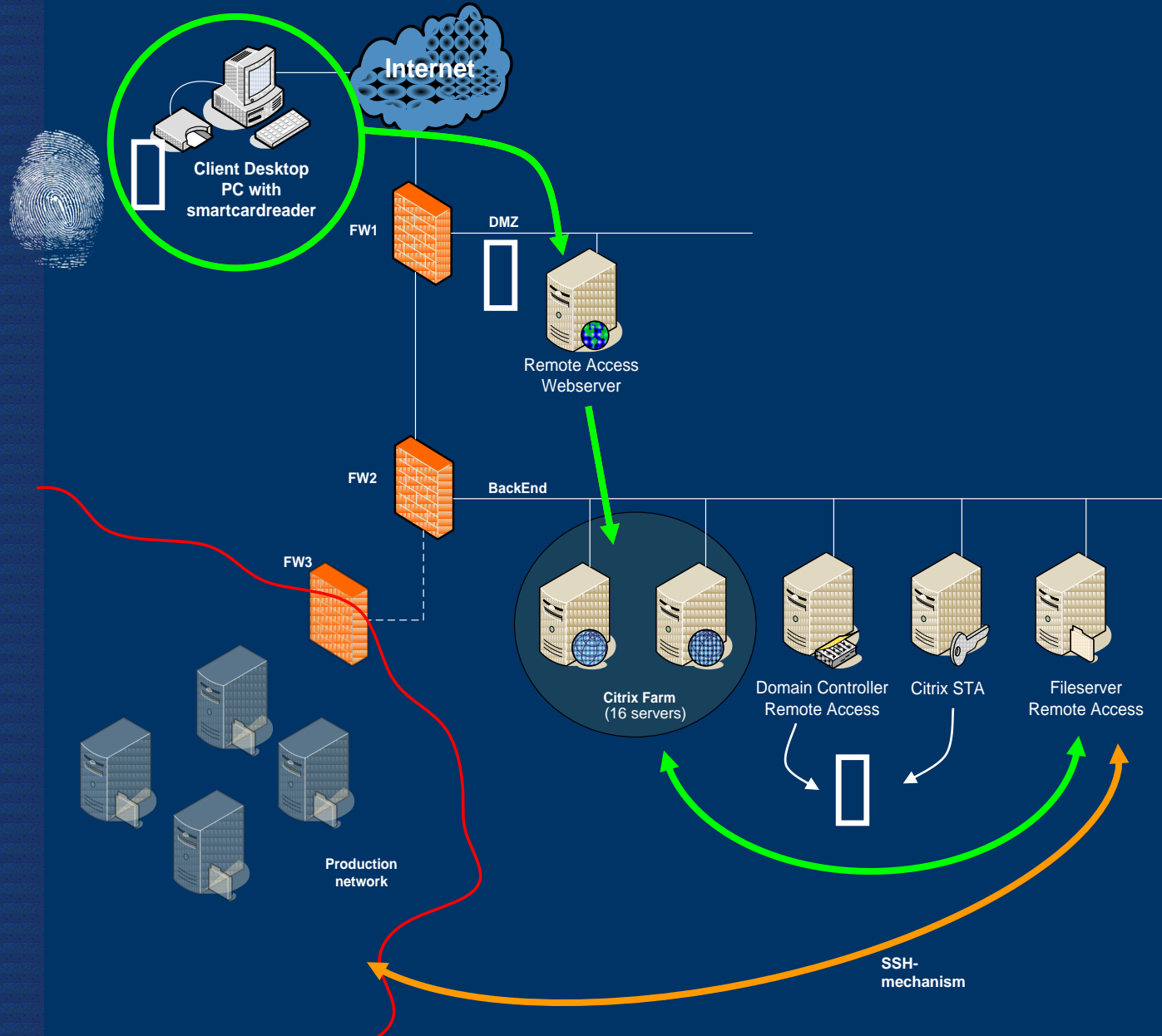


Microdata access facility

- Seperate, secure IT-network
 - Use of CITRIX
- Two entry modes: On Site/Remote Access
- On Site
 - Available for more than 11 years already
 - Safe room at CBS offices (8:30-17:00)
 - Renting of computer per half a day
- Remote Access
 - Available since september 2006 (after initial pilot)
 - 24/7 access from own workspace (internet)
 - Fingerprintreader/smartcard



Technical aspects RA



Remote Access PC

- Customer provides computer
- Dedicated Remote Access PC
 - “Light” PC
 - Internet connection
 - USB-port
 - No laptop !
- Separate office that can be locked
- Visit by CvB
 - Inspect location
 - Install fingerprintreader



New project – contracts

- Intake: Researcher and accountmanager, permission to use data, research aim, necessary datafiles, timeframe, users, additional wishes
- Basic contract Remote Access
- Contract agreement researchproject
- Confidentiality statement researcher
- Standardised documents (templates)



New project – Useraccount

- Personal account
- Different accounts for each project
- 20 - 80 GB storage space
- Access to necessary datasets
- Access to additional software
- Shared workspace for the project
- Introduction: by CvB employee, smartcard (one per person, enrollment fingerprint etc.



Output checking

- All output is checked
 - OS output within 1 day
 - RA output within 5 days
- Only for disclosure control
- Rules for presenting output
- Manually
 - 35-40 people who check output
- CvB AND statistical department
- Researchers remain responsible
- Mailing of output to emailaddress at office



Publications

- Researcher responsible for publications
- CBS not responsible for conclusions or quality
- CvB receives copy (as information, not control)



Customers

- Remote Access:
 - 80 computers
 - 200 current research projects
 - approx. 350 active researchers
- On Site:
 - 10 computers
 - > 50 current research projects
 - approx. 100 active researchers OS



Personnel (OS and RA)

- 3 accountmanagers
- 7 people to make documentation
- 6 people to support RA (and OS)
 - contracts
 - accounts
 - introduction
 - support
- 35-40 people who check output
- IT maintance and support (1 fte?)



Resources-diskspace

- Central repository : 2,5 TB of datafiles
- On OS/RA: 24 TB of userdata



Another microdata service

Scientific use files

- Distribution of 'safe files' on CD-ROM
- Data have been distorted to make it safe:
 - Recoding, regrouping, dropping variables etc
- Limited number of files (10 subjects)
- Very limited number of users (mainly universities)

