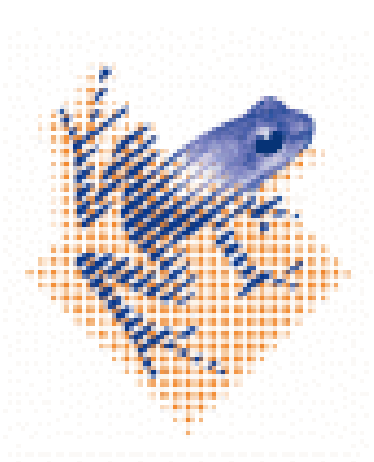




NEDERLANDSE FEDERATIE VAN  
UNIVERSITAIR MEDISCHE CENTRA



university of  
 groningen

# Data integrity

Ronald Stolk

*Professor of Clinical Epidemiology  
Director Research Data & Biobanking*

# Data integrity @ UMCG



umcg



university of  
groningen

- **Data stewardship**
  - FAIR data
  - National Dutch approach
- **Local UMCG solutions**
  - Workspace
- **Two nice projects**

# Data stewardship



umcg



university of  
groningen

- **Good research practice**
  - “patients deserve it, society desires it and funders demand it”
- **Scientific research gets more visible**
  - Journal editors
  - Funding bodies (H2020: 5%)
  - Health Care Inspectorate
  - Quality systems

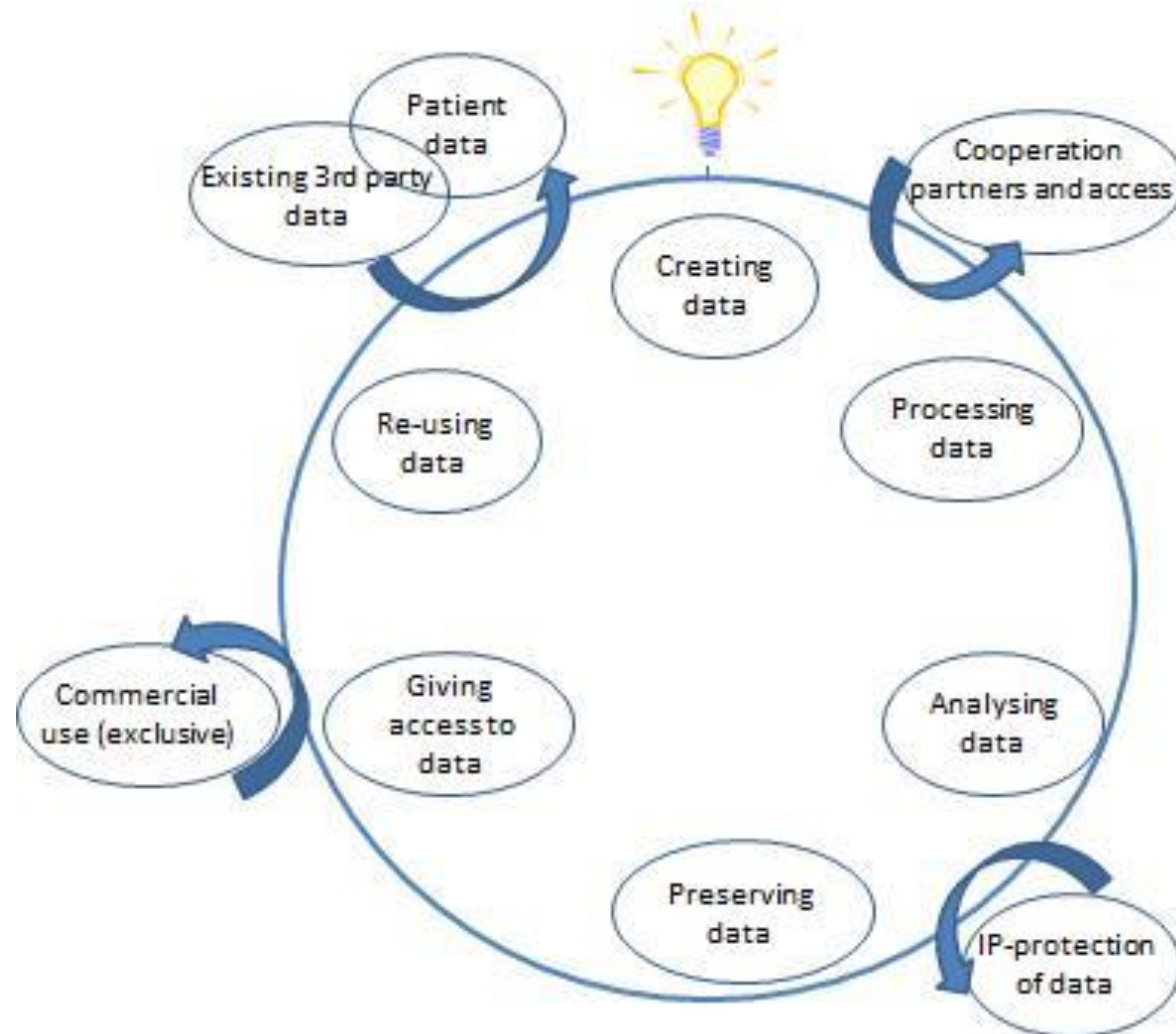
# Data life cycle



umcg



university of  
groningen



# FAIR data stewardship

- **Data Stewardship of life science information**
  - Findable – Easy to find, description of metadata
  - Accessible – Well-defined license and access conditions
  - Interoperable – To be combined with other datasets
  - Re-Usable – Future research, processed further



# Data stewardship

- Technological resources, e.g. storage space, support staff time
- Legal and ethical requirements, including privacy
- Able to share your finalised data publically
- Robust documentation
- Backed up and safe from sudden loss or corruption
- Remain accessible and comprehensible in the near, middle, and distant future
- Can be shared with others for scientific research, commercial development or validation



umcg



university of  
groningen

# Responsibilities

- **Researchers, clinicians**
  - Collect, analyse, or reuse data
- **Data managers, lab technicians, etc**
  - Procedures and technical systems for data stewardship
- **Institute managers**
  - Facilitate & train the professionals
- **Supervisory bodies**
  - Medical-ethical committees, privacy officers, etc

# Challenges for researchers



umcg



university of  
groningen

- **The safety & privacy of patients**
- **The integrity of research data**
  
- **Research complies with (inter)national laws, legislation and standards**
- **Research conduct is transparent, traceable and verifiable**





umcg



university of  
groningen

# Challenges for managers

- **IT facilities**
  - Storage, access
- **Regulations, policies, procedures**
  - Tailored to local situation
- **Training**
  - Continuous learning



# Re-using existing data

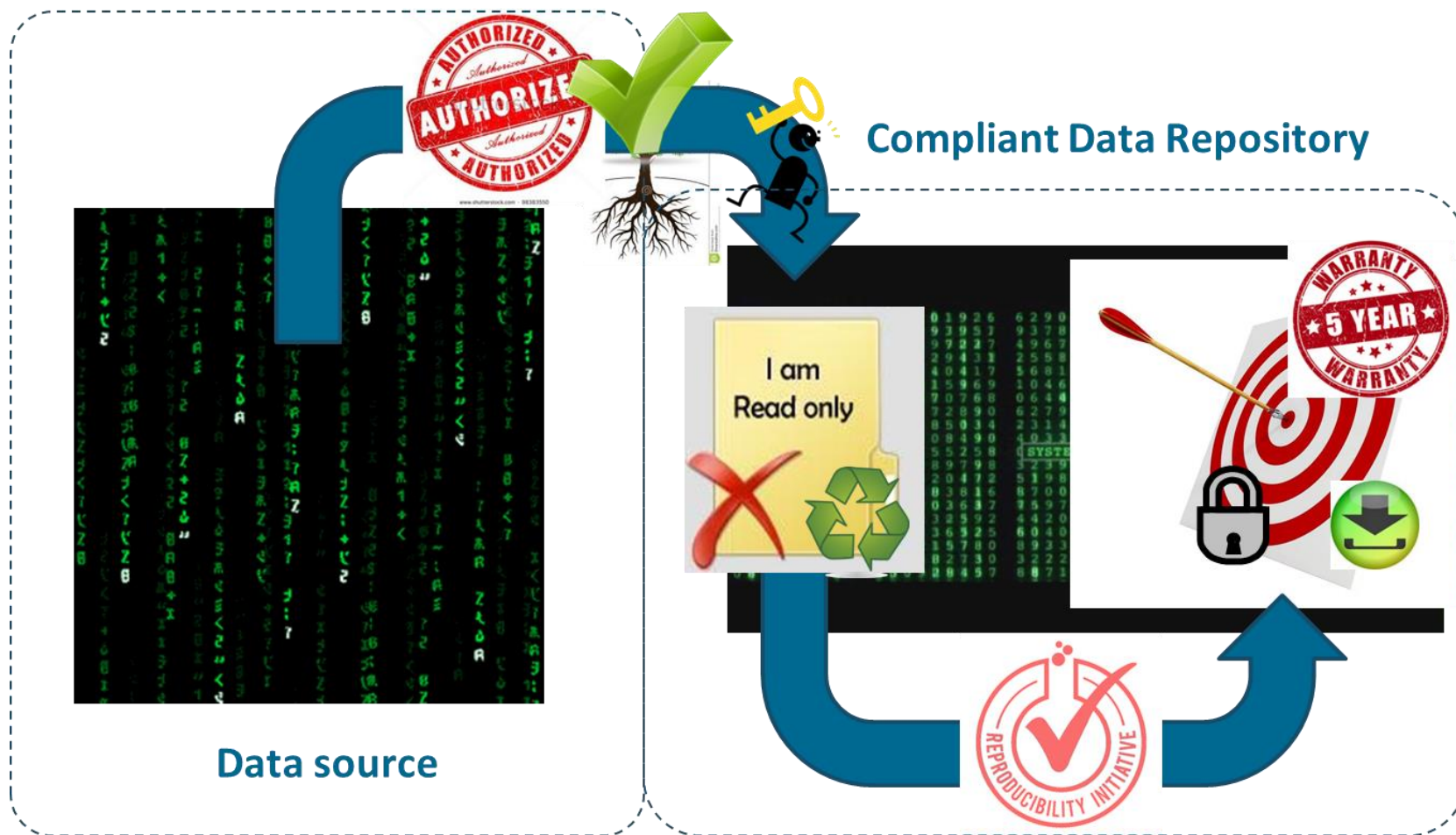
- **Existing biobank/cohort studies**
  - Numerous smaller studies
  - National Cohort
- **International biobank/cohort studies**
  - [bbmri-eric.eu](http://bbmri-eric.eu)
  - Genetic consortiums, European collaborations
- **Clinical files: EPR**
  - Informed consent/no objection

# Practical aspects data storage

- **Pseudonimized**
  - Also for joining data from multiple sources
- **Standardized coding**
  - HL7, SNOWMED, ICD10, ...
- **Rigorous file formats**
  - xml, DICOM, VCF, ...
- **Meta data**
  - Catalogue



# FAIR storing data



# Dutch program research data

- Towards a shared data infrastructure for biomedical research
- Access, linkage, sharing, training, ...



NETHERLANDS FEDERATION OF UMCS  
DATA4LIFESCIENCES

## HANDS

Handbook for Adequate Natural  
Data Stewardship

What is data stewardship?

You are a data steward

FAIR principle

Accountability and responsibility

Preparing your study

Creating and processing your  
data

Analysing your data

Preserving your data (archiving)

Giving access to your data

About this Handbook

Glossary

Appendices

## Handbook for Adequate Natural Data Stewardship



The Federation of Dutch UMCs (NFU) has developed this handbook to promote good data stewardship at the eight Dutch university medical centres (UMCs). HANDS was written by experts with knowledge of the current regulations and

# Data integrity @ UMCG



umcg

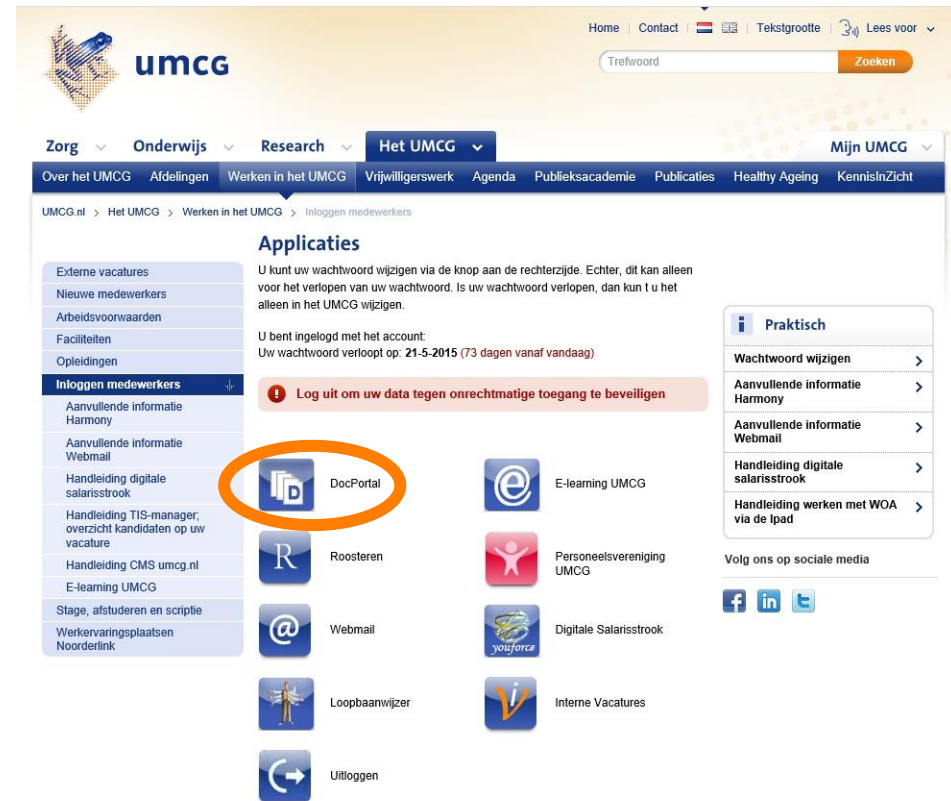


university of  
groningen

- **Data stewardship**
  - FAIR data
  - National Dutch approach
- **Local UMCG solutions**
  - Workspace
- **Two nice projects**

# ISO 9001 (27001)

- Care, research, teaching
- Quality management system
  - PDCA cycle
  - Quality managers
- DocPortal
- Audits
  - Internal
  - External



The screenshot shows the UMCG website interface. The main navigation bar includes 'Zorg', 'Onderwijs', 'Research', and 'Het UMCG'. The 'Het UMCG' dropdown menu is open, showing options like 'Over het UMCG', 'Afdelingen', 'Werken in het UMCG', 'Vrijwilligerswerk', 'Agenda', 'Publieksacademie', 'Publicaties', 'Healthy Ageing', and 'KennisinZicht'. The 'Werken in het UMCG' section is active, displaying a list of services. The 'DocPortal' icon, which is a blue square with a white 'D' and a document icon, is circled in orange. Other icons include 'Roosteren', 'Webmail', 'Loopbaanwijzer', 'Uitloggen', 'E-learning UMCG', 'Personeelsvereniging UMCG', 'Digitale Salarisstrook', and 'Interne Vacatures'. A search bar is visible at the top right, and a 'Log uit om uw data tegen onrechtmatige toegang te beveiligen' warning is present.





umcg



university of  
groningen

# Research toolbox

- **Toolbox**
  - Decision support flow charts
  - Templates
  - References to HANDS and other sources
- **Research registry**
  - Including type of consent
- **Clinical Research Office**
  - Support

# Workspace

- **Research workspace environment for researchers in health care research**
- **Remote access to high performance storage and compute**
  - No local data storage

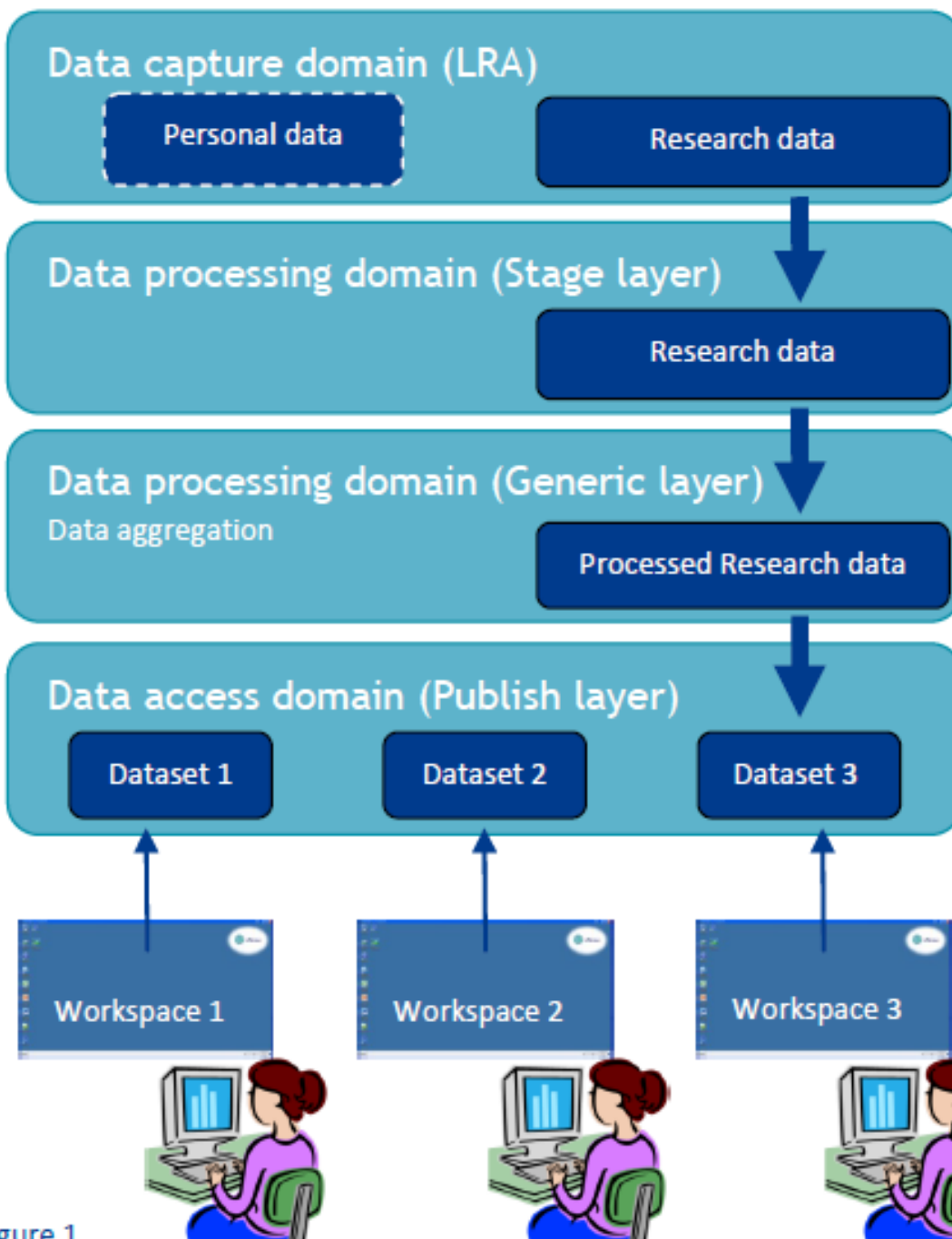


Figure 1

# Functional requirements

## Workspace



umcg



university of  
groningen

- **Secured access to limited number of applications**
  - LifeLines: fully locked
  - UMCG: flexible locking
- **Global accessible, 24/7**
- **Scalable – linkage with Linux cluster**
- **Easy to use**
- **Fast**



# Deployment

- Universal access to desktop, apps and data
- Deployable on PC, laptop, mobile device (any device)
- Based on virtualization technology



Windows  
PCs & laptops



Tablets &  
smartphones



Macintosh  
computers



Thin clients

# Look and feel



Lifelines Logon

Username

Password



Please enter your SMS token.

Token



UWP Logon

Username

Password


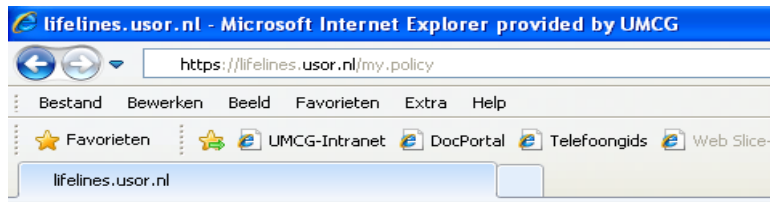


UMCG Logon

Username

Password

# Login


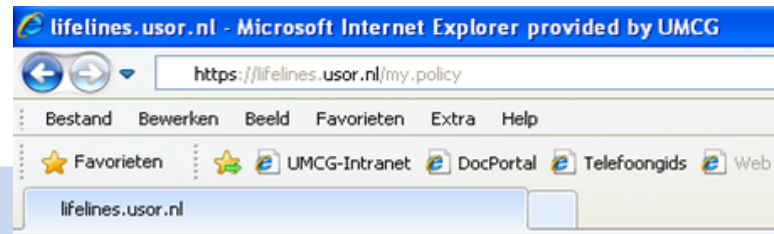


umcg

UMCG Logon

Username

Password

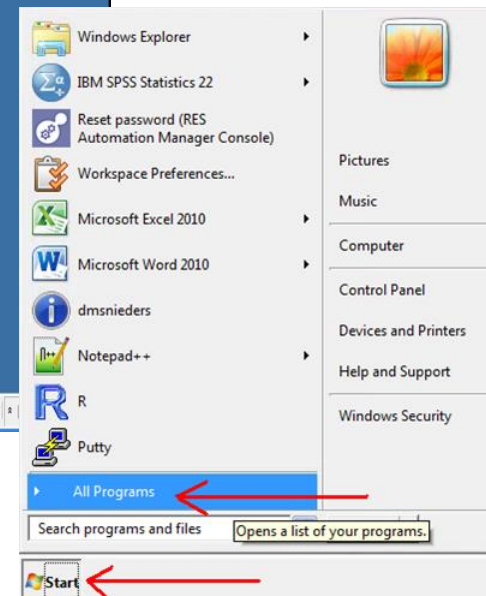


umcg

Please enter your SMS token.

Token

# Easy to use





# User roles

- **LifeLines researcher**
  - NO up- and download functionality
  - NO Internet access, no printer and USB devices supported
  - MS-Office, SPSS, Plink, R, Putty, WinSCP, Adobe, Notepad, ...
  - On request: SAS, STATA, ...
  - Read and write access private share
  - Read access global share
- **UMCG researcher**
  - Limited up- and download functionality
  - ...

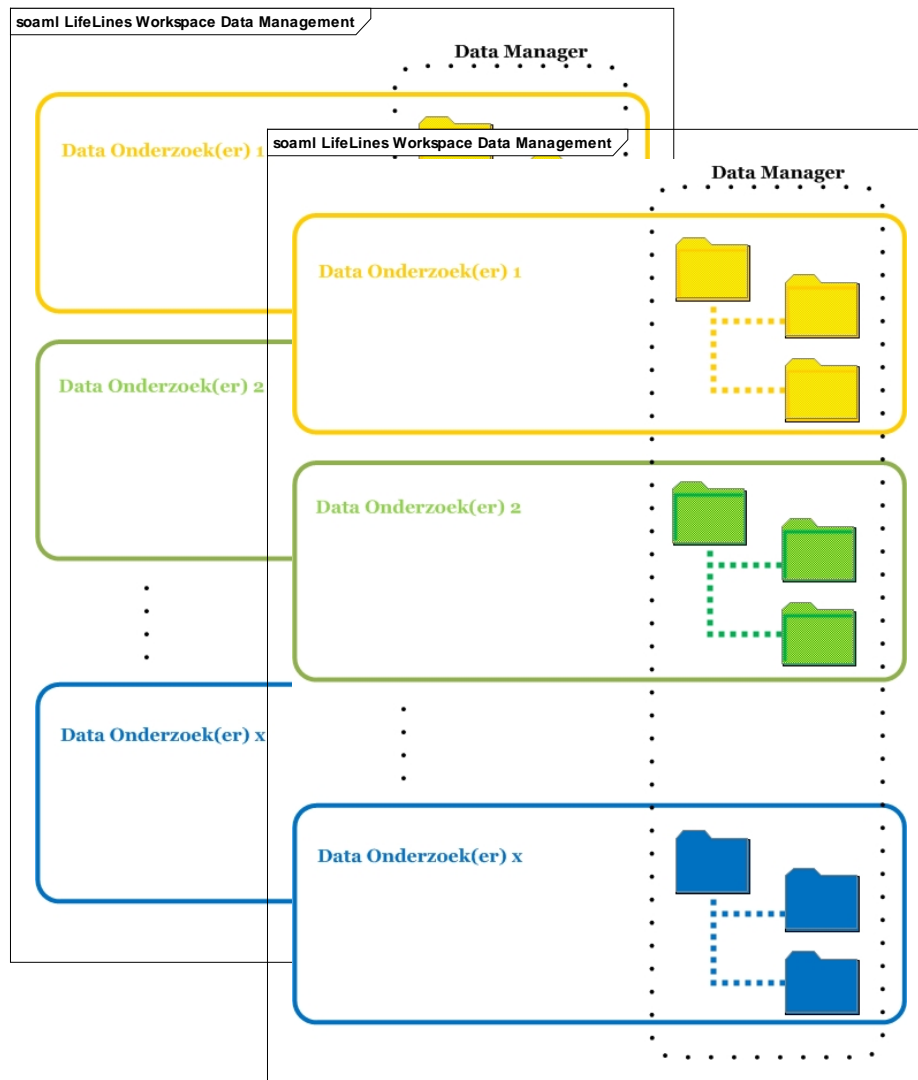
# Work together in groups



umcg



university of  
groningen



# Towards IaaS

## infrastructure as a service



umcg



university of  
groningen

### Clients

- 1 SFTP file server
- 1 SSH commandline access
- 2 Windows workspace 2 cores 2Gb ram 1TB storage
- 0 Linux workspace 8 cores 16Gb ram 6TB storage

### Policies

- Enable data download\*
- Enable internet access

\* Always enabled for SSH clients

### High performance compute

- 128 cores
- 4Gb RAM per core (total: 512Gb)

Advanced options

### Bulk storage

- 5TB High available storage (incl nightly backup)
- 40TB High performance storage (temp, no backup)
- 60TB Tape archiving

### Applications

- SPSS
- R
- OpenClinica
- MOLGENIS
- Galaxy
- Imputation pipeline
- NGS pipeline
- Etc

### Users

- swertzma
- imhamf

+

Request configure research infrastructure for: 4,680 euro/month

Cancel

OK

# Data integrity @ UMCG



umcg



university of  
groningen

- **Data stewardship**
  - FAIR data
  - National Dutch approach
- **Local UMCG solutions**
  - Workspace
- **Two nice projects**



- **Dedicated knowledge resource for the local and international research community**
  - Big data
  - Analysis of biobank/registries
  - Biomaterial
- **Dutch export product**
  - Oldenburg EMS cohort initiative



umcg



university of  
groningen

# Registration at the source

- **Single registration, multiple use**
  - Uniform care documentation
  - International standards

## Registratie aan de bron

Visie op documentatie  
en gebruik van  
zorgegevens 2013-2020



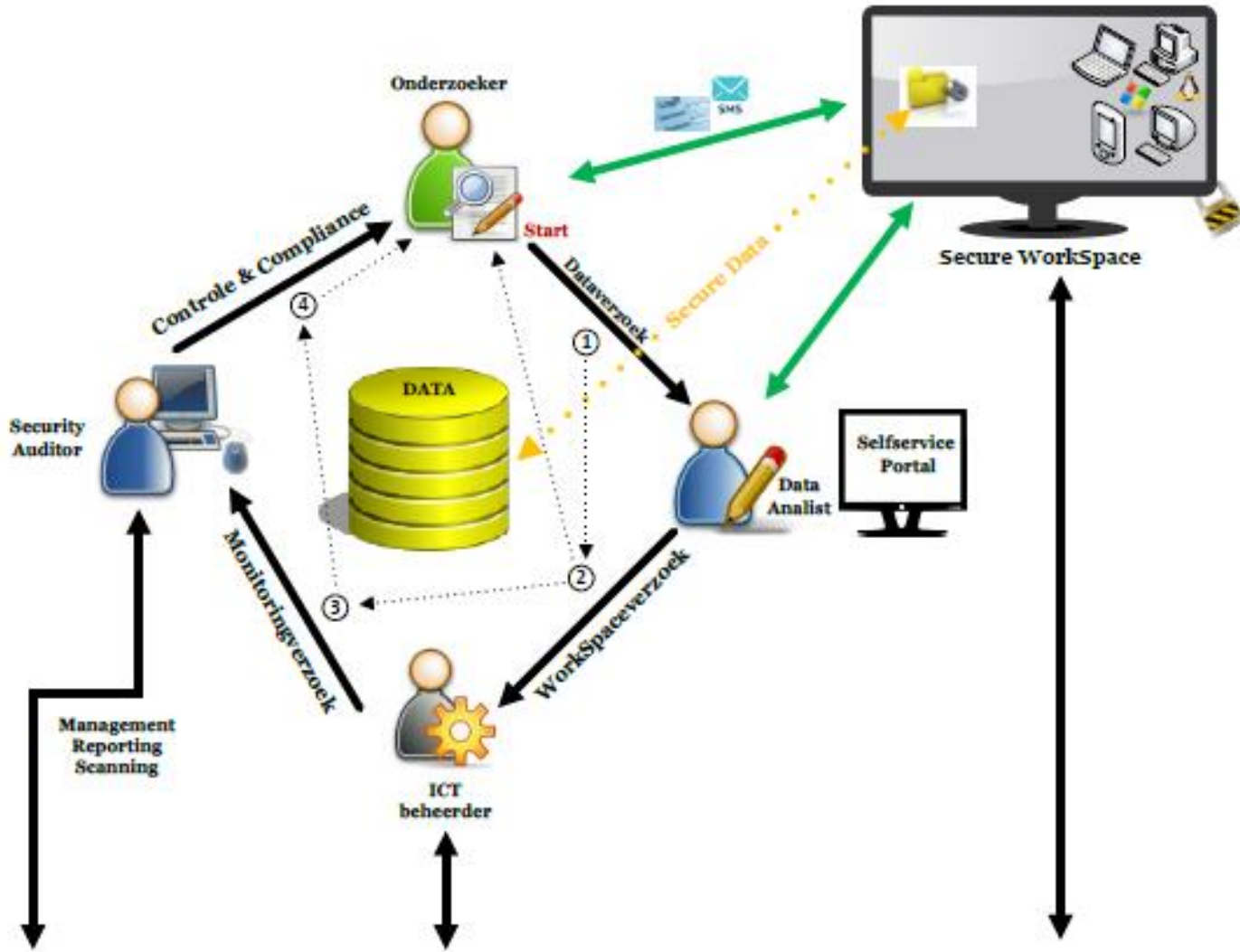
# Team for data integrity



umcg



university of  
groningen



# Questions - Answers



[rdb.umcg.nl](http://rdb.umcg.nl)  
[www.biobank.bike](http://www.biobank.bike)  
[www.lifelines.net](http://www.lifelines.net)  
[www.dtls.nl](http://www.dtls.nl)  
[www.bbmri-eric.eu](http://www.bbmri-eric.eu)

Summer School Biobanking  
20-24 June, 2016

[R.P.Stolk@umcg.nl](mailto:R.P.Stolk@umcg.nl)

Linked 





# Technology

