

Complex landscape



Statistics Sweden:
40 national registers
and surveys of interest
to health related
research




Health
registers



12 central
authorities
with registers



100
national
health care
quality
registers



BBMRI and
the National
Biobank
Register with
some 600
research and
clinical
biobanks

The SIMSAM-INFRA application 2011 listed the coordination needs

Project management	Central registers	Healthcare quality registers	Biobanks	Researcher generated data
Databases/computing				
Metadata/documentation				
Legal and ethical aspects				
Researcher support				
International collaboration				



**Karolinska
Institutet**

Survey of data management and analysis within SIMSAM

Marie Reilly

National SIMSAM steering group meeting 17 Sept 2012

SIMSAM "sharing survey"

- Marie Reilly, Anna Holmström, Luwis Diya
- The first step after critique from evaluation 2012
- To learn what is being done already
- Identify potential for sharing resources
- Get suggestions for how to promote sharing
- Fuel discussion concerning how to proceed
- Received some 80 responses across all the SIMSAM1 nodes

Role

PI/co-PI:	31%
Postdoc	19%
PhD	50%

Main work in SIMSAM

Medical Science 50%

Social Science 47%

Epidemiologist 44%

Social Scientist 24%

Demographer 24%

Statistician 19%

Geographer 7%

Data Manager 7%

Database Designer

/Administrator 4%

Other (student, economist, clinician) 15%

Data Preparation work

58% Yes

Stata	60%
SAS	28%
SPSS	23%
R	13%
ArcGIs	8%
Access	2.5%
dBase	2.5%

Code for data preparation

Found available code **31%**

class coding/origin, case/control, cohorts, families,
standardised incidence.....

Wrote code **87%**

case-control, cohorts, connecting kin, spouses/partners,
siblings, identifying duplicates, ICD, class coding, person-
time, time-dependent exposure

Data Analysis work

87% Yes

Stata	53%
SPSS	27%
SAS	23%
R	18%
ArcGIs	6%
Other	16%

Code for data analysis

Found available code	20/68	(29%)
Would recommend to others	10	(50%)
Needed to write code	23/62	(37%)
(relatively) easy for others		~ 50%

Willing to share code

Within group	52/57 (91%)
Within SIMSAM	42/52 (81%)
Anyone	27/47 (57%)

Interested in code from others: 55/72 (**76%**)

Obstacles to sharing code

Lack of documentation/commenting	68%
Individual logic/style	65%
Finding resources	39%
Changing versions of software	36%

Useful resources listed

Books, manuals, web pages, R package, SAS help,
Lexis macro, stataлист,

Conclusion

From survey, evidence of inefficiency and need for sharing, but also clear **opportunity**:

Interest

Willingness

Expertise

Information

Next steps?

- SIMSAM INFRA
 - We asked nodes to list data management teams/competencies
 - We tried to set up a SIMSAM DM working group
 - We opened a discussion list on registerforskning.se
- We failed .. Why?
- How can we succeed? We need your help...