Distributedlearning.ai: Towards a Distributed Learning Network for Cancer Registries

Gijs Geleijnse

Dilemma

- A research question that requires ***pooling*** sensitive data from multiple cancer registries:

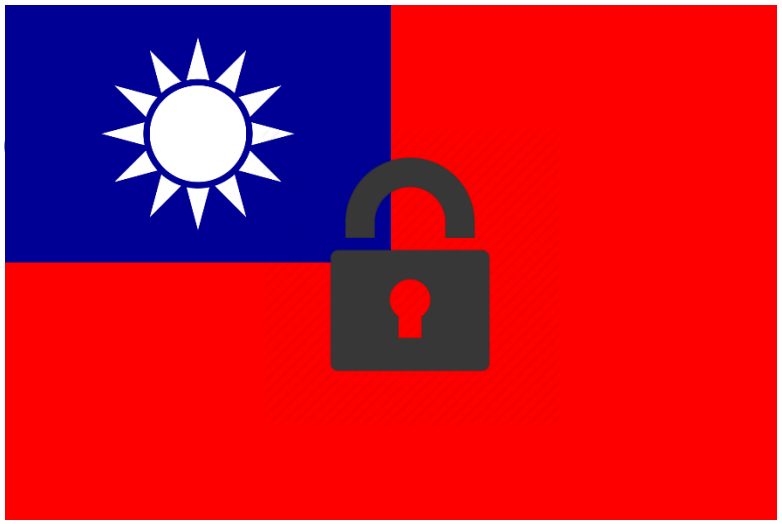
  - Because separate analyses do not give me all the insights
  - Because separate data sets are **too small**

Dilemma

- A research question that requires ***pooling*** sensitive data from multiple cancer registries:

ses

sets



- With current privacy protection acts in place, sharing patient data across borders may be **difficult**

A different approach

- If sharing is the problem: **don’t share the data!**
- If you can’t bring the data to your statistical software … you have to bring the statistics to the data

- Consequences:
  - The algorithm (logistic regression, Cox PH, ...) has to be distributed
  - Only aggregated data is shared and combined
  - The data formats have to be aligned

Te Data Matters

DataSHIELD: taking the analysis to the data, not the data to the analysis

Amadou Gaye,[1] Yannick Marcon,[2] Julia Isaeva,[3] Philippe LaFlamme,[2] Andrew Turner,[1] Elinor M Jones,[4] Joel Minion,[1] Andrew W Boyd,[1] Christopher J Newby,[5] Marja-Liisa Nuotio,[6,7] Rebecca Wilson,[1] Oliver Butters,[1] Barnaby Murtagh,[8] Ipek Demir,[9] Dany Doiron,[2] Lisette Giepmans,[10] Susan E Wallace,[8] Isabelle Budin-Ljøsne,[3]

Distrib
multi
conc

Arthu
Wim

[a] Departm
[c] Klinik fü
(CHU de l

H. Bre

Nu
Ed
Rc
Br

Communication-Efficient Learning of Deep Networks from Decentralized Data

---

WebDISCO: a web service for distributed cox model learning without patient-level data sharing

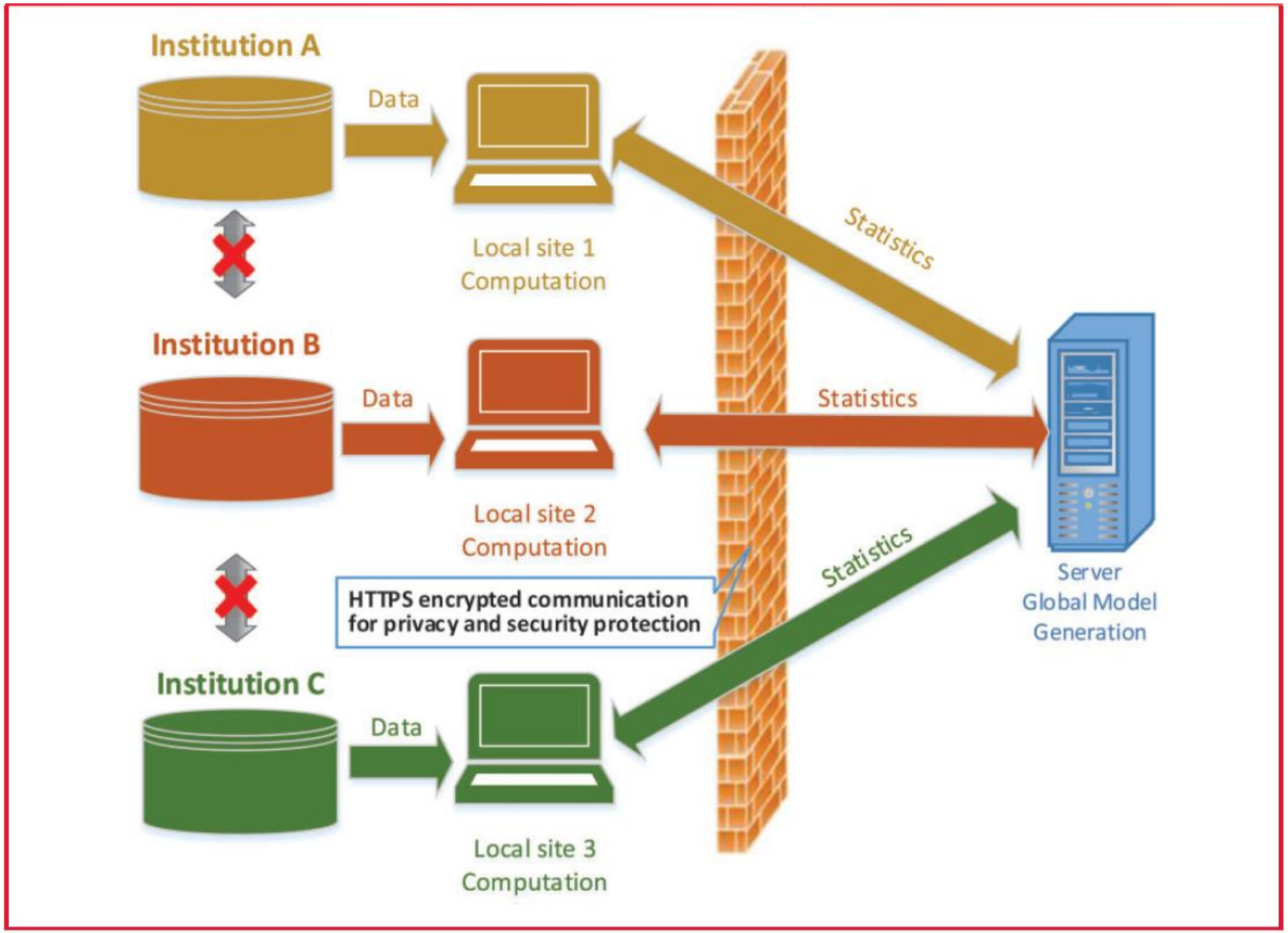
Chia-Lun Lu[1], Shuang Wang[1,*], Zhanglong Ji[1], Yuan Wu[2], Li Xiong[3,4], Xiaoqian Jiang[1,*], Lucila Ohno-Machado[1]

Vincent Ferretti[2,25†] and Paul R Burton[2,24†*]

RECEIVED 24 November 2014
REVISED 16 May 2015
ACCEPTED 26 May 2015
PUBLISHED ONLINE FIRST 9 July 2015

AMIA INFORMATICS PROFESSIONALS. LEADING THE WAY.

OXFORD UNIVERSITY PRESS



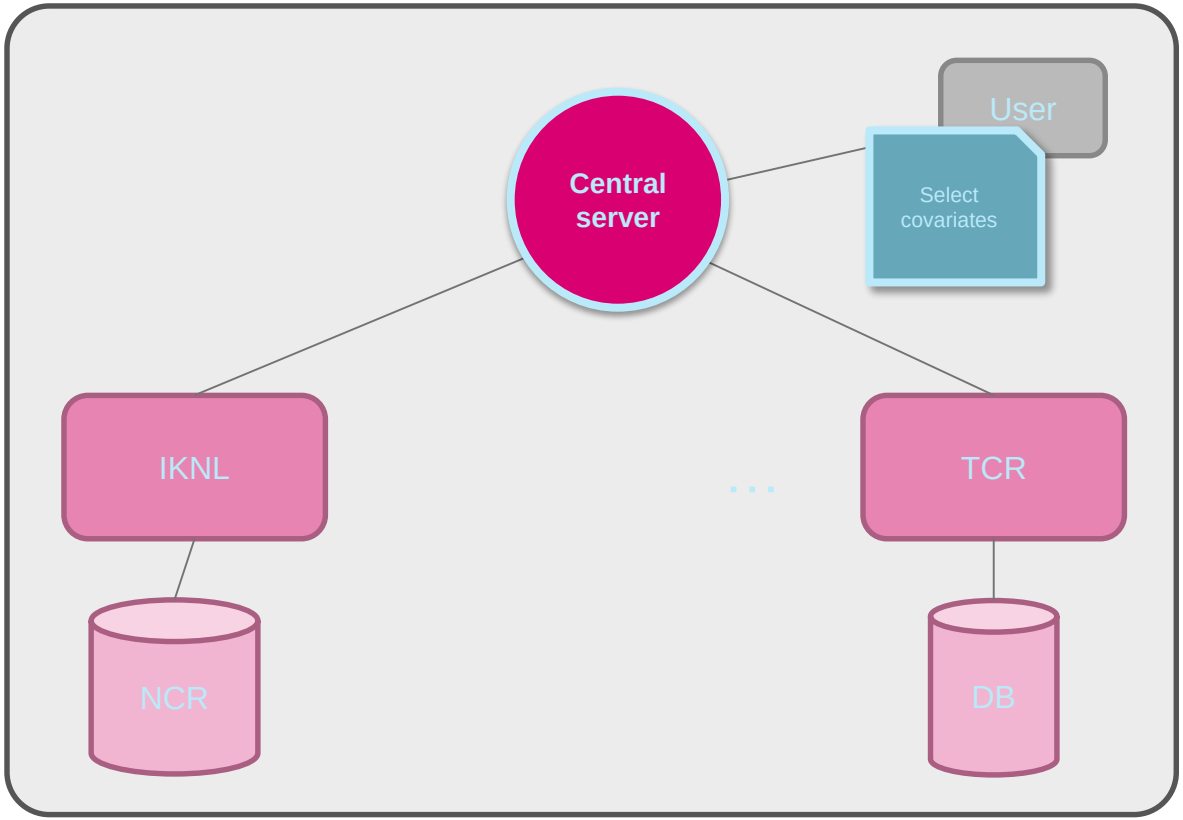
From: https://www.ncbi.nlm.nih.gov/pubmed/26159465

Oral Cavity: Taiwan & The Netherlands

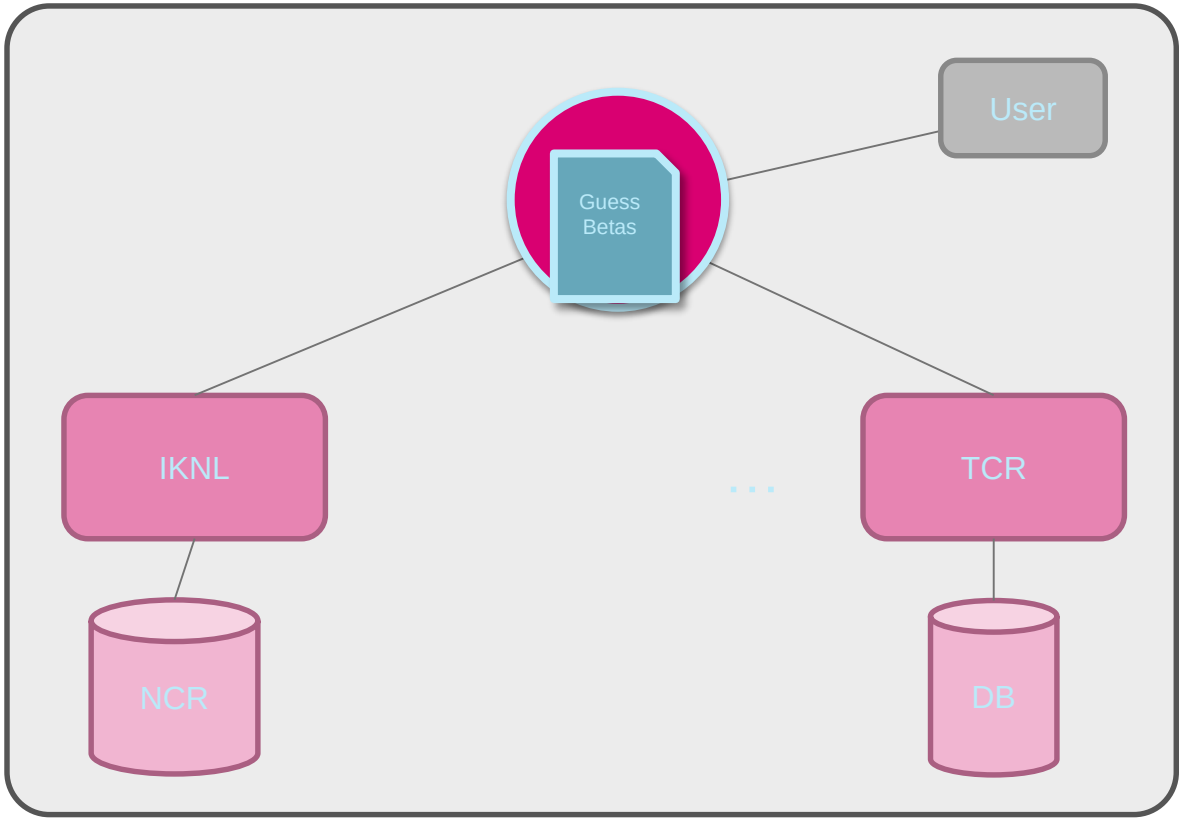
- Can we implement an open source, distributed, privacy preserving Cox PH analysis?

- Does survival of oral cavity carcinoma differ between Taiwan and the Netherlands? And if so, which factors contribute to the variation?

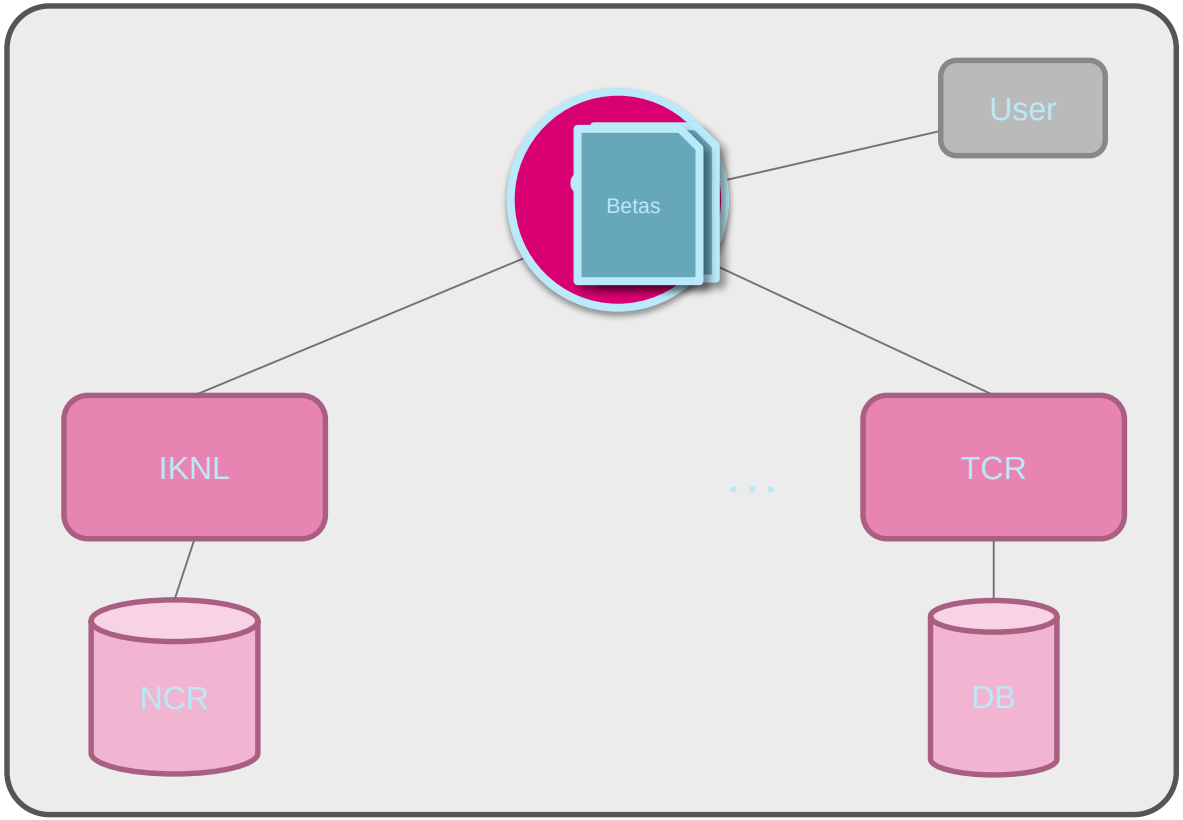
Process flow – Cox PH



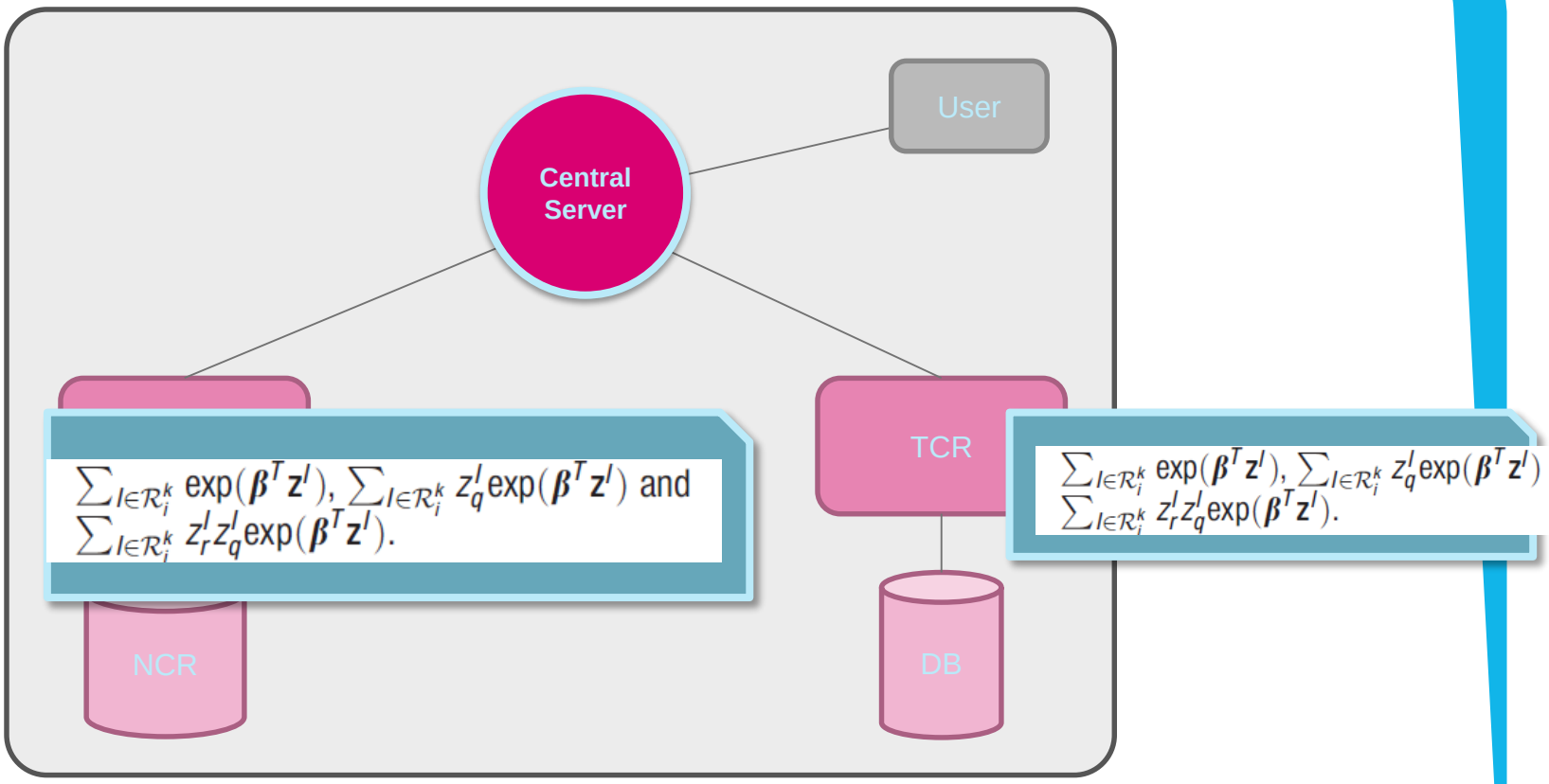
Process flow – Cox PH



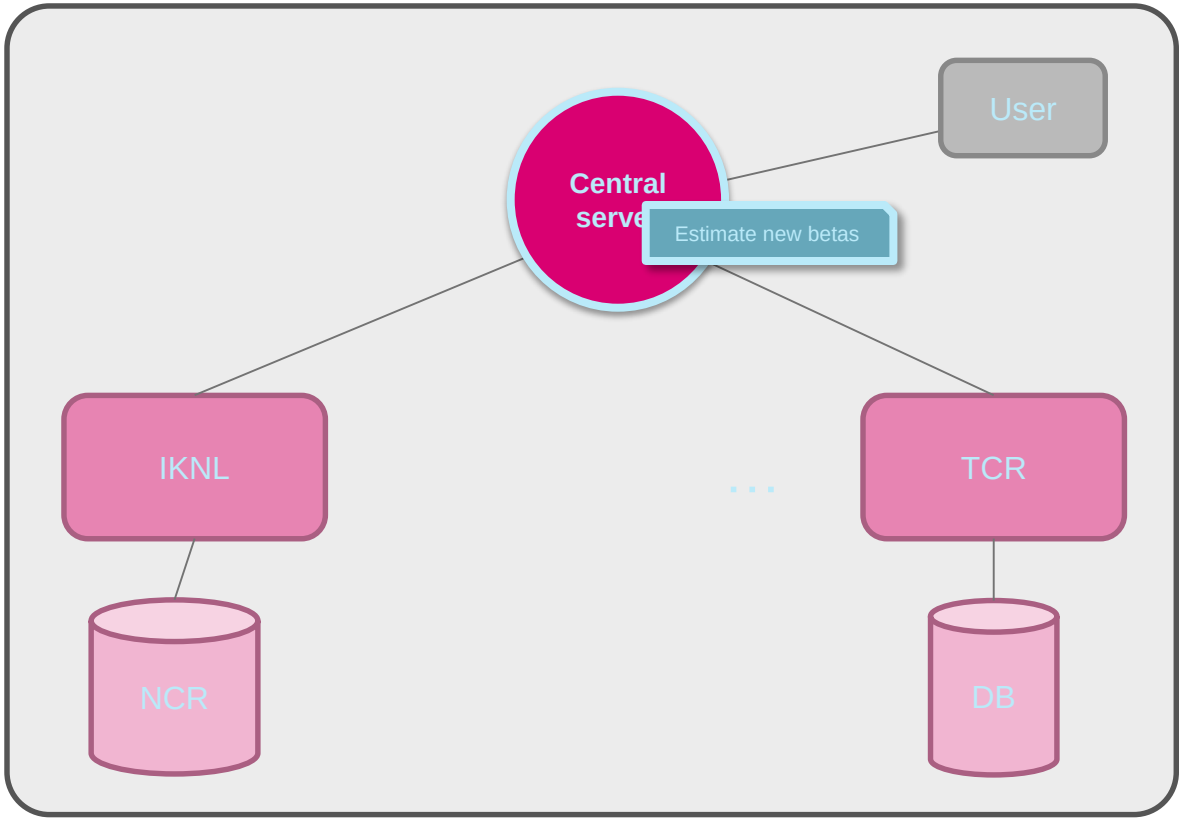
Process flow – Cox PH



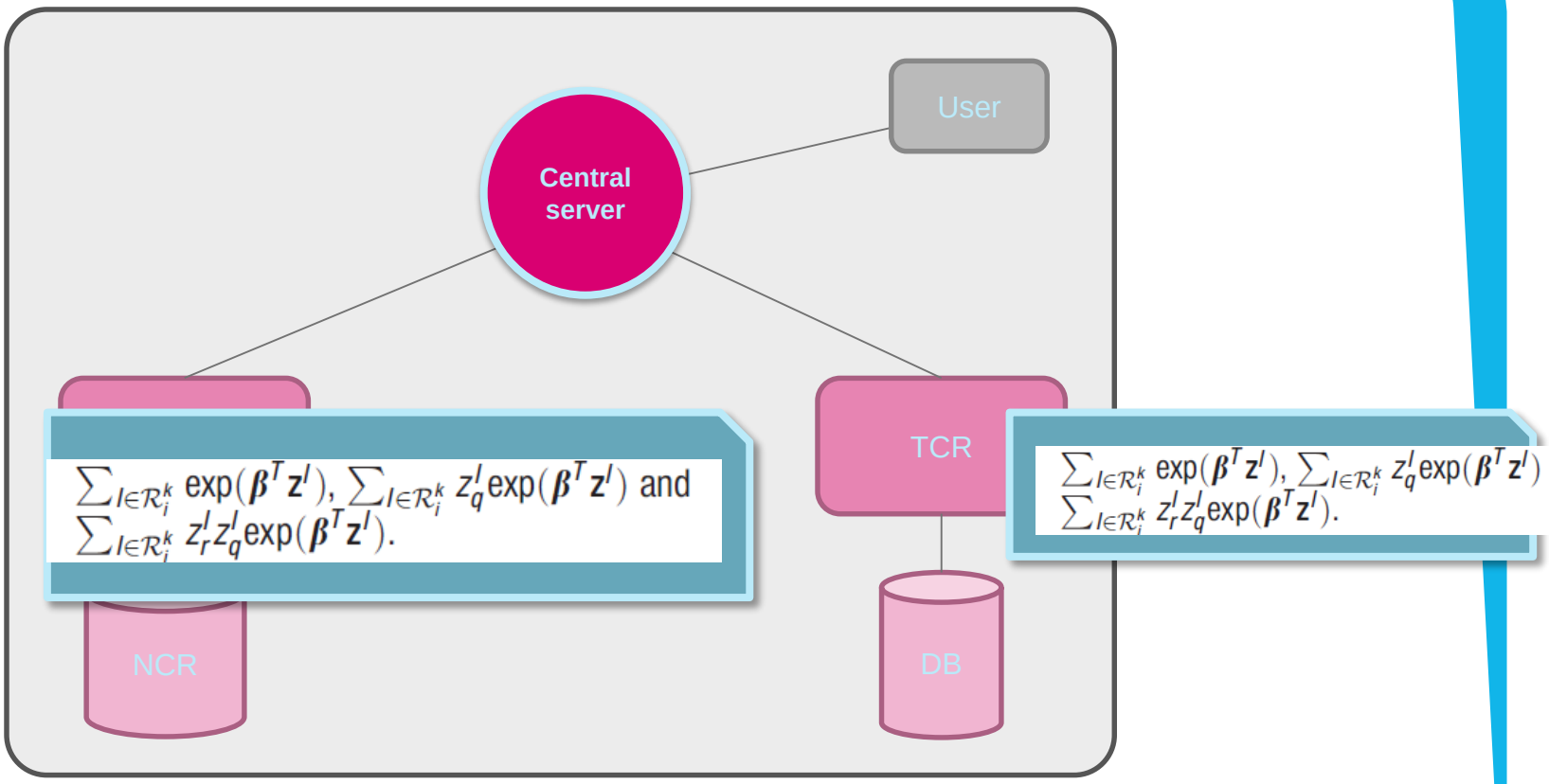
Process flow – Cox PH



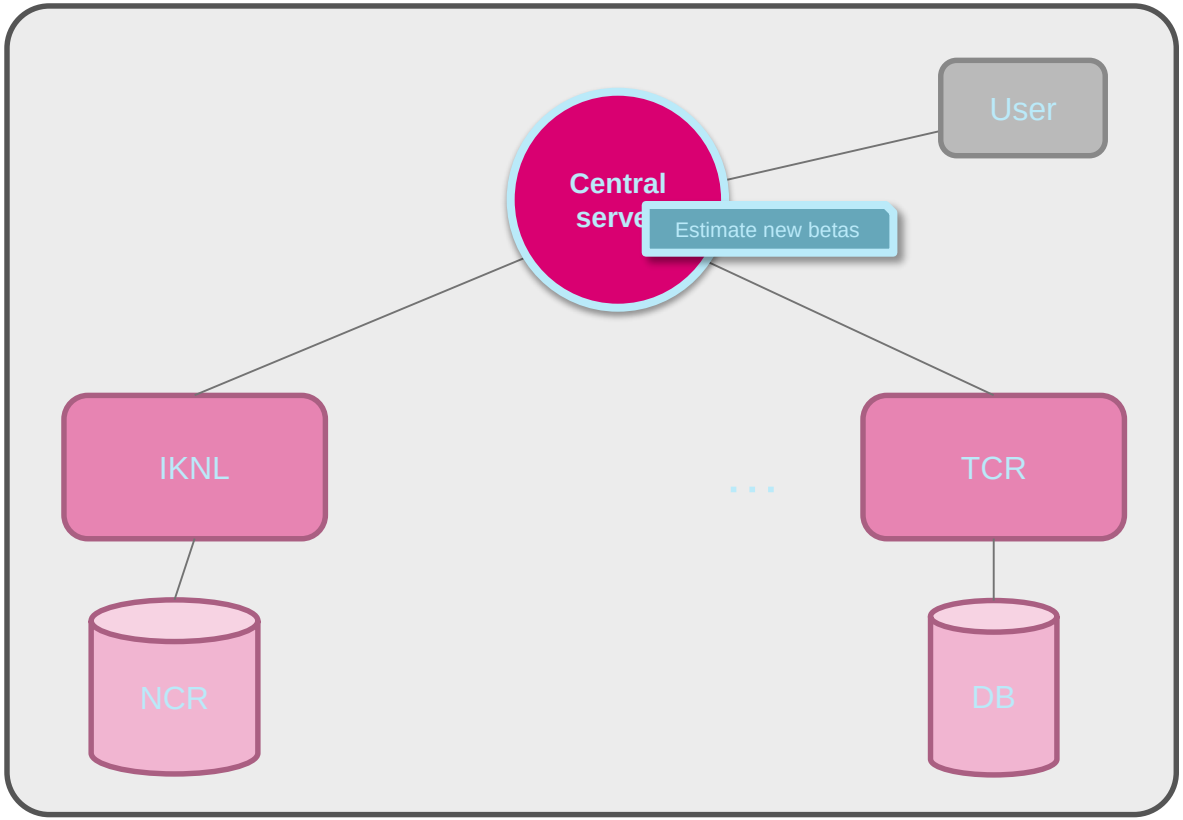
Process flow – Cox PH



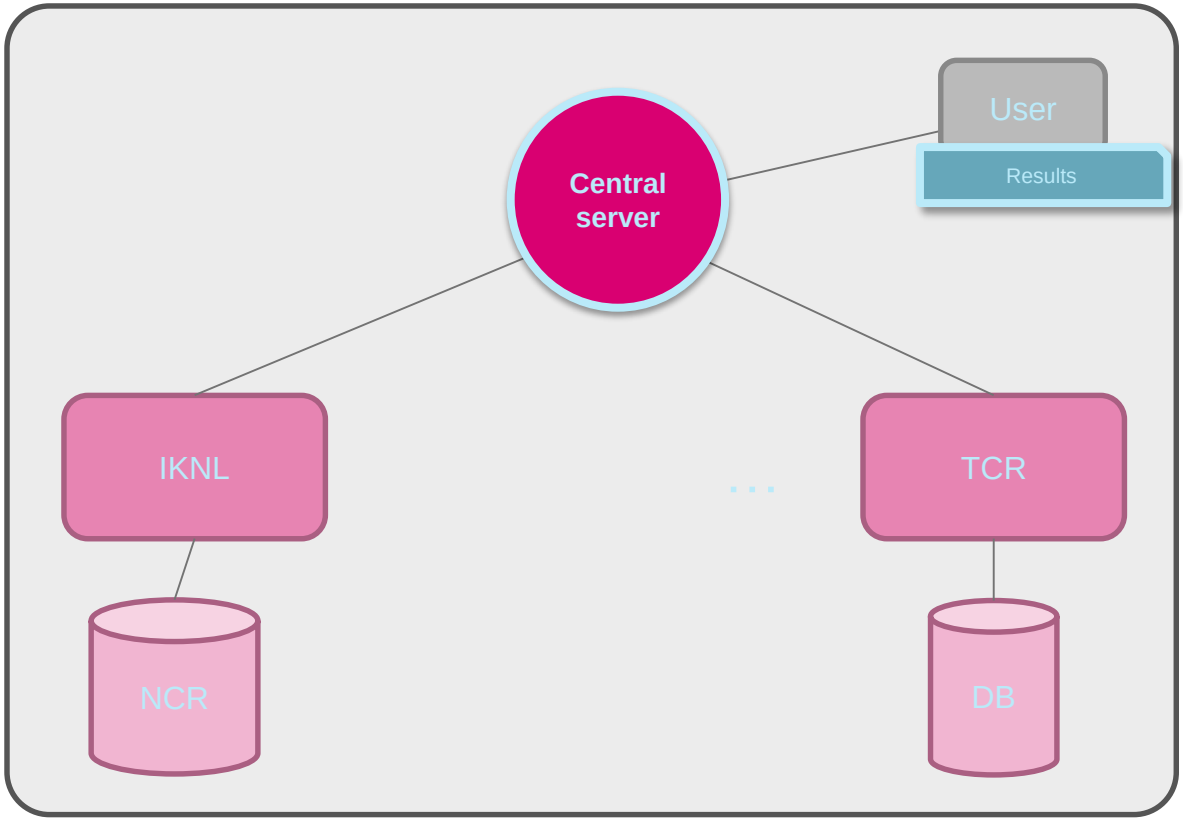
Process flow – Cox PH



Process flow – Cox PH



Process flow – Cox PH



		univar HR (95% CI)	multivar HR (95% CI)
	Age	1,03 (1,03 - 1,03)	1,03 (1,03 - 1,03)
Gender	Male	1 (ref)	1 (ref)
	Female	0,95 (0,90 - 1,00)	0,81 (0,76 - 0,85)
Geography	Taiwan	1 (ref)	1 (ref)
	Netherlands	1,20 (1,14 - 1,26)	0,97 (0,91 - 1,03)
Period of diagnosis	2004 - 2010	1 (ref)	1 (ref)
	2008 - 2011	0,93 (0,88 - 0,99)	0,96 (0,91 - 1,02)
	2012 - 2016	0,87 (0,82 - 0,93)	0,80 (0,75 - 0,86)
Topography	lip	1 (ref)	1 (ref)
	floor of mouth	1,96 (1,55 - 2,48)	1,68 (1,33 - 2,13)
	gum	2,15 (1,70 - 2,72)	1,30 (1,03 - 1,66)
	other/unsp parts of mouth	1,69 (1,34 - 2,13)	1,39 (1,10 - 1,75)
	other/unsp parts of tongue	1,54 (1,22 - 1,94)	1,67 (1,32 - 2,11)
	palate	2,42 (1,85 - 3,16)	1,26 (0,96 - 1,64)
Stage	I	1 (ref)	1 (ref)
	II	1,51 (1,40 - 1,65)	1,34 (1,23 - 1,46)
	III	2,04 (1,87 - 2,23)	1,74 (1,58 - 1,91)
	IVA	3,39 (3,27 -3,73)	2,92 (2,69 - 3,17)
	IVB	5,80 (5,18 - 6,48)	4,37 (3,84 - 4,98)
	IVC	15,74 (13,16 - 18,82)	5,91 (4,89 - 7,15)
	X	2,99 (2,57 - 3,49)	1,23 (1,04 - 1,47)
Grade	Well differentiated	1 (ref)	1 (ref)
	Badly/undifferentiated	2,10 (1,92 - 2,29)	1,59 (1,45 - 1,75)
	Moderately differtiated	1,51 (1,41 - 1,61)	1,38 (1,29 - 1,48)
	Unknown	1,84 (1,70 - 1,99)	1,32 (1,22 - 1,43)
Treatment	Primary surgery	1 (ref)	1 (ref)
	No treatment	8,40 (7,72 - 9.13)	5,38 (4,89 - 5,93)
	Other treatment	6,85 (5,84 - 8.03)	4,21 (3,55 - 4,99)
	Primary radiotherapy	4,15 (3,87 - 4,46)	1,98 (1,82 - 2,16)
	Surgery and chemo	1,63 (1,35 - 1,97)	1,40 (1,15 - 1,70)
	Surgery and radiotherapy	1,81 (1,71 - 1,92)	1,00 (0,93 - 1,07)
	Unknown	3,89 (3,19 - 4,74)	4,19 (3,37 - 5,22)

Conclusions

- We presented a working, open-source system to analyse combined privacy-sensitive data – without sharing data

- System can support cross-registry research by addressing data governance issues

integraal
kankercentrum
Nederland

integraal
kankercentrum
Nederland