

# Integrating large datasets for the Movember Global Action Plan on Active Surveillance for low risk Prostate Cancer



Tim Hulsen<sup>1</sup>, Henk Obbink<sup>1</sup>, Wim van der Linden<sup>1</sup>, Cees de Jonge<sup>2</sup>, Daan Nieboer<sup>3</sup>, Sophie Bruinsma<sup>4</sup>, Monique Roobol<sup>4</sup>, Chris Bangma<sup>4</sup>

<sup>1</sup>Philips Research, department of Professional Health Solutions & Services, Eindhoven, the Netherlands

<sup>2</sup>Philips Research, department of eScience, Eindhoven, the Netherlands

<sup>3</sup>Erasmus MC, department of Public Health, Rotterdam, the Netherlands

<sup>4</sup>Erasmus MC, department of Urology, Rotterdam, the Netherlands

## Introduction & Objectives

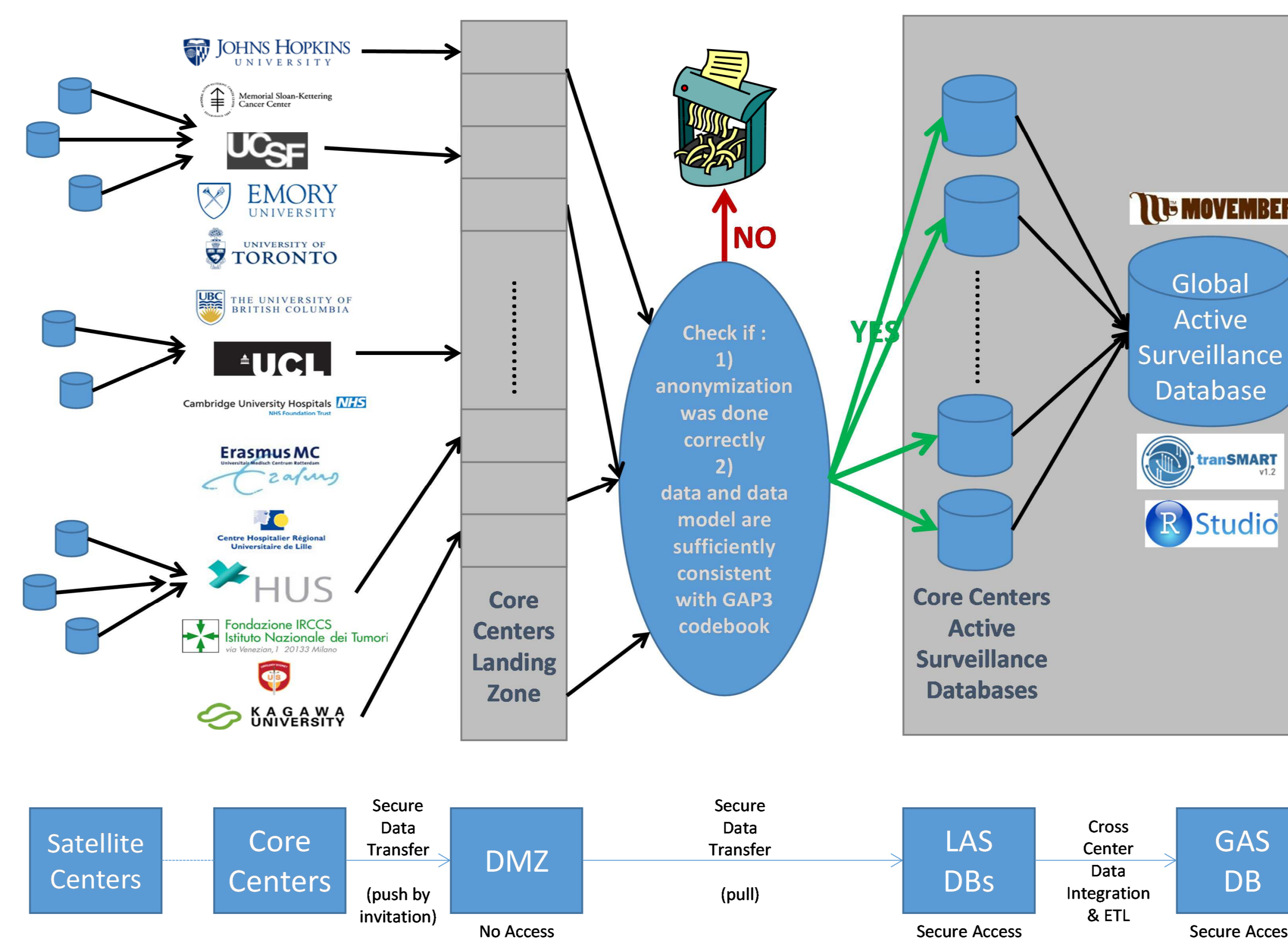
The Movember Global Action Plan (GAP) on active surveillance for low risk prostate cancer includes the integrated 30 months activity of 19 institutions in 14 countries in the 5 Movember regions (Australasia, Europe, UK, Canada, and USA). The initiative is also open to other eligible centres. Milestones of the project include a global Active Surveillance (AS) database for clinical, biospecimen, imaging and biomarker data (including a virtual biobank), as well as worldwide tailor-made guidelines on AS and a web-based platform on AS. The database needs to be accessible for integrated analysis on all datasets from all participating institutes.

## Materials & Methods

The global database IT infrastructure (**left figure**) is based on the tranSMART prostate cancer instance of the TraIT IT infrastructure developed by Philips within the Dutch CTMM-TraIT and CTMM-PCMM projects ([www.ctmm-traits.nl](http://www.ctmm-traits.nl), [www.pcmmpoject.org](http://www.pcmmpoject.org)). This infrastructure offers support for collecting and combining the various large, longitudinal datasets from the participating institutes. Transfer of data takes place using the Secure Data Transfer tool provided by Philips.

## Results

The Movember Active Surveillance database currently contains datasets of more than 10,000 patients. The clinical data has been gathered using a common data model (**right figure**), specifically designed to answer the research questions defined by the principal investigators at the start of the project. This data can be browsed through the web-based tranSMART platform, which is only accessible for a selected group of users. tranSMART supports a number of statistical analyses, such as correlation analysis, logistic regression and survival analysis. Genomics data can be analysed in tranSMART through built-in analysis methods such as group tests and heatmaps. Imaging data will be analysed through imaging analysis pipelines, and the resulting data will be stored in tranSMART, with links to the original images. The Movember Active Surveillance tranSMART instance is connected to R-studio to enable the statisticians to execute their own R scripts on the database.



Codebook GAP-3 global database Version: 22-01-2016		
<b>Inclusion</b>		
Year of birth	YYYY	Integer
Height at diagnosis	cm	Numeric
Weight at diagnosis	kg	Numeric
BMI at diagnosis	kg/cm <sup>2</sup>	Numeric
Race	Asian Black White Native American Native Hawaiian or other pacific islander Mixed Other	Categorical
Patient refused	Yes No	Categorical
Ethnicity (Hispanic or Latino)	Yes No	Categorical
Country of origin		String
Marital/relationship status at diagnosis (With/Without partner)	Yes No	Categorical
Highest Educational level	Less than high school High school University	Categorical
Family history of prostate cancer (Death of PCa in 1 <sup>st</sup> or 2 <sup>nd</sup> degree relative)	Yes No	Categorical
Biomarkers available for this patient	Text	String
Year of diagnosis	YYYY	Integer
Diagnosis done by TURP	Yes No	Categorical
PSA at inclusion	ng/ml	Numeric
free-PSA	Percentage	Numeric
T-Stage at DRE	Tx T1 T1a T1b T1c T2 T2a T2b	Categorical
TNM-staging system used	1992 1997	Categorical

## Conclusions

We show here that it is feasible to construct a global Active Surveillance database, which enables researchers to browse and analyse the clinical, biobank, imaging and genomics data in a secure and efficient manner. The infrastructure, that includes tranSMART and R-studio, will be used in future projects that handle large clinical datasets.