



**Ward Weistra**

Lead Translational Data Warehousing - The Hyve

[ward@thehyve.nl](mailto:ward@thehyve.nl)

# The latest in Glowing Bear

## *A modern open source interface for tranSMART, inspired by i2b2*

i2b2 2017 European Meeting, Paris, Oct 6<sup>th</sup> 2017



@WardWeistra  
@TheHyveNL  
#GlowingBear

# The Hyve - Open source solutions

Boston, USA



Utrecht, Netherlands

Singapore



Me

Kees (CEO and Founder)

Hyvie



# Towards one integrated infrastructure for personalised medicine & health research

2016



Health~Holland  
SHARED CHALLENGES. SMART SOLUTIONS



lygature



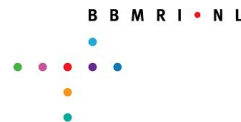
NETHERLANDS FEDERATION OF  
UNIVERSITY MEDICAL CENTRES



2014



2012



2009

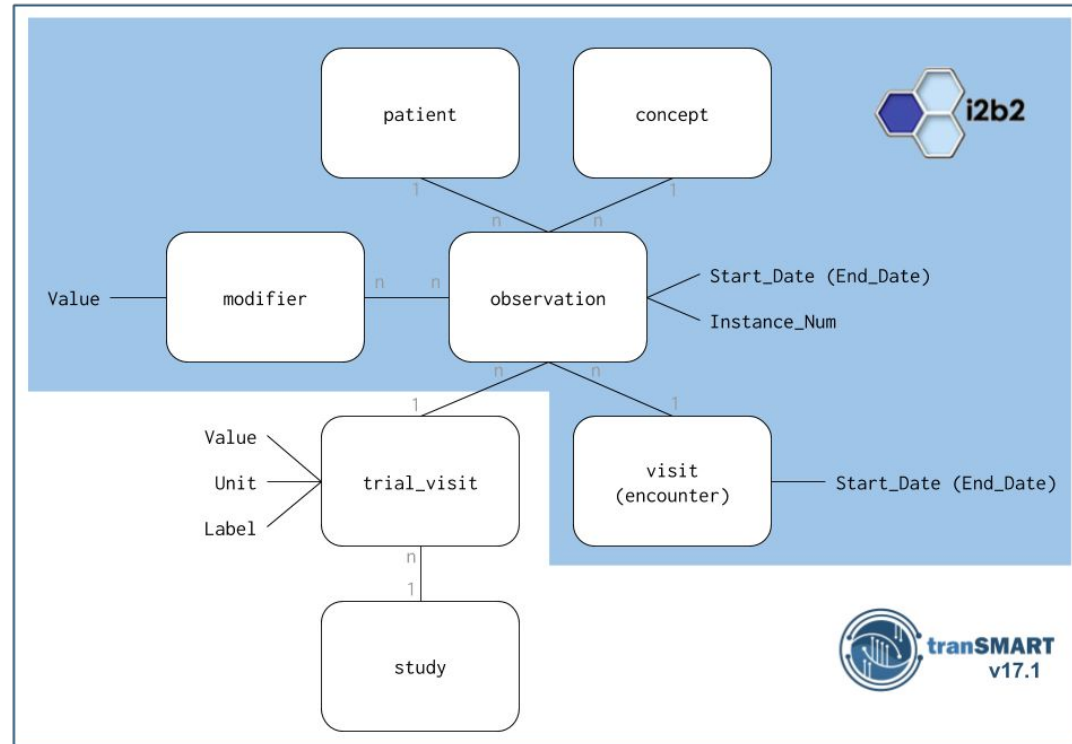
200 biobanks



**What made  
the Glowing Bear cohort selector  
possible and necessary?**

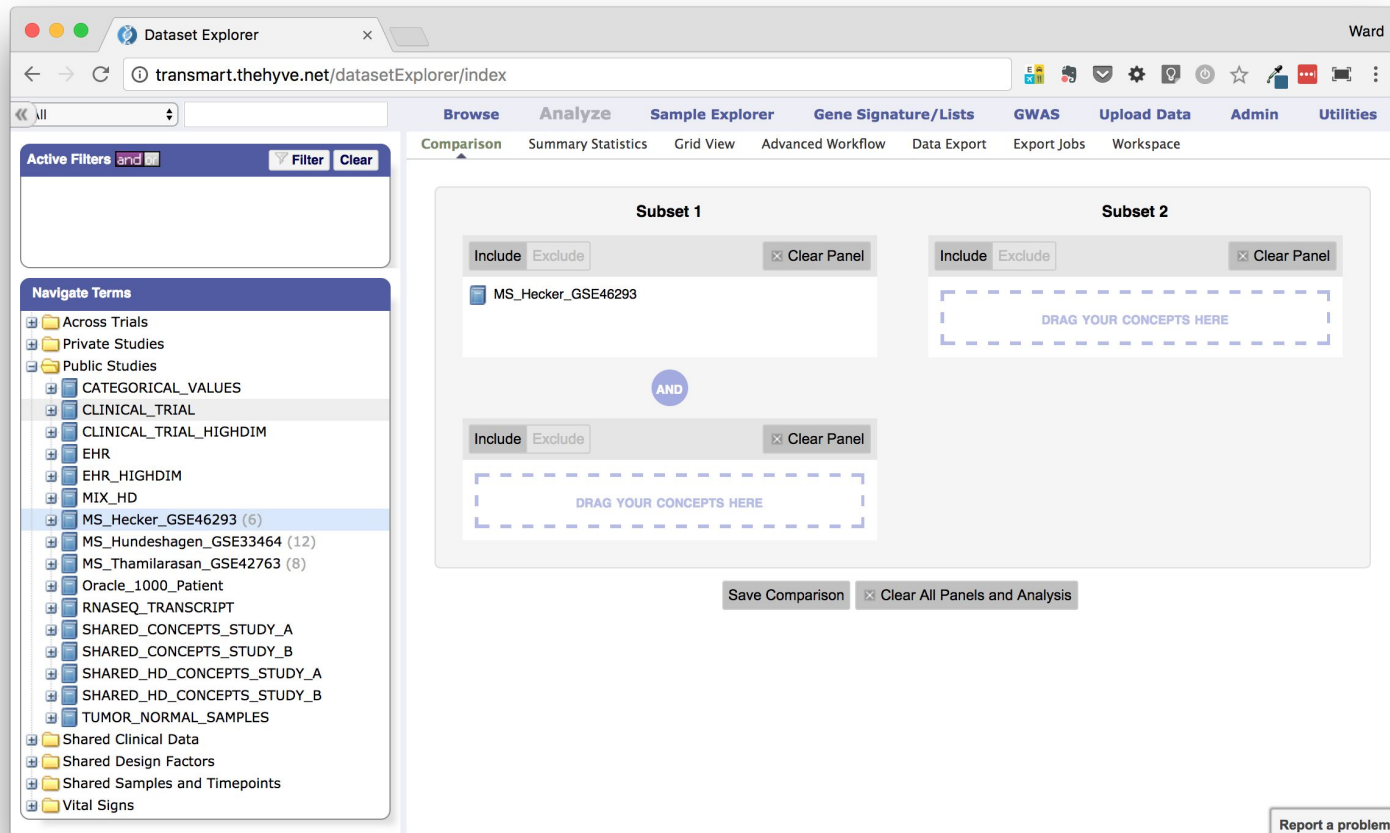
# tranSMART 17.1

- Supports:
  - absolute and relative time series
  - samples and replicates
  - cross-study concepts and ontologies
- Achieved by restoring the i2b2 database



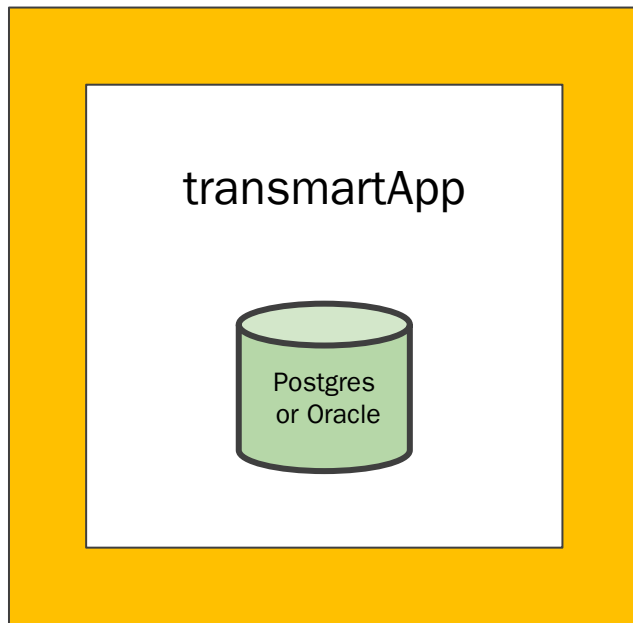
- Server version is released as beta on Github
  - <https://github.com/tranSMART-Foundation/transmart-core>
  - Production version expected in Oct 2017

# 'The tranSMART UI' transmartApp

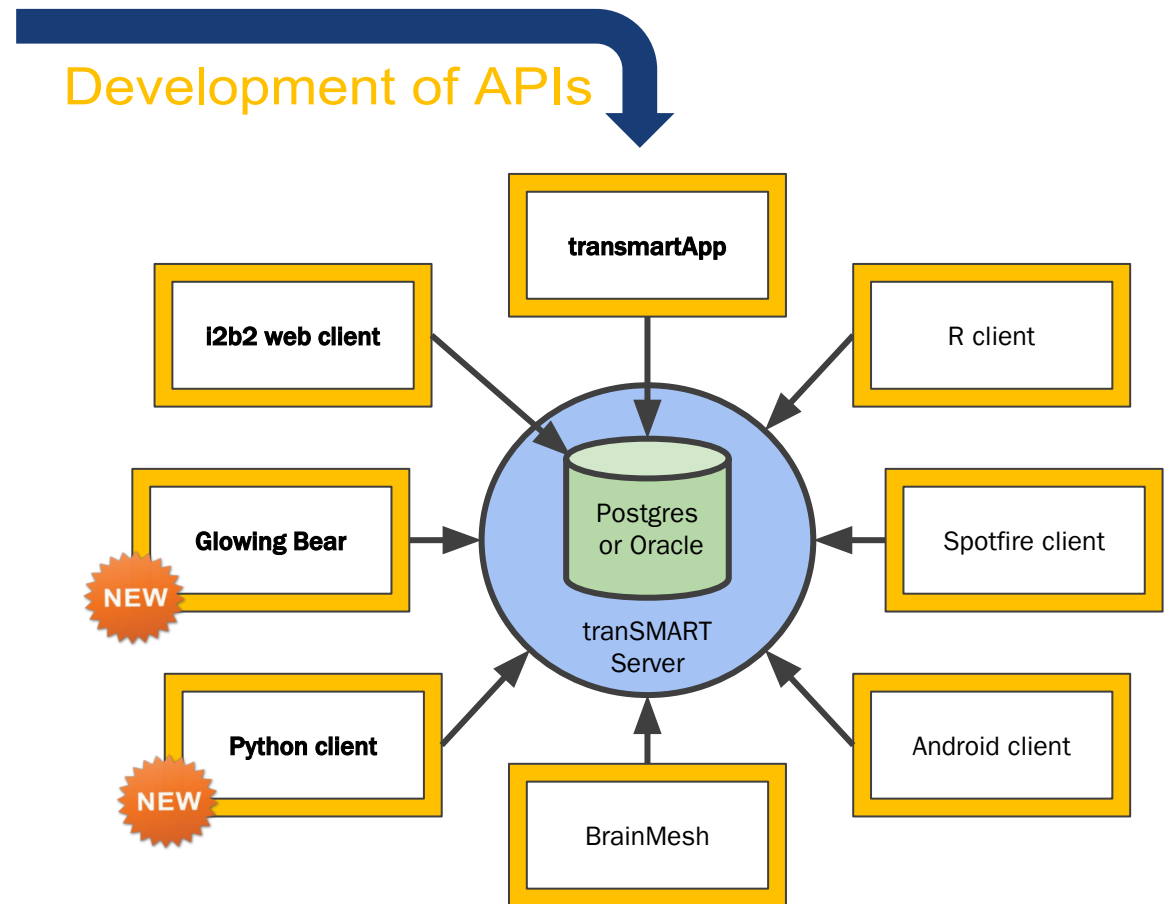


- Works with all data, without the new dimensions/features.
- No support for time series, samples and cross-study concepts

# TranSMART REST API (V2)

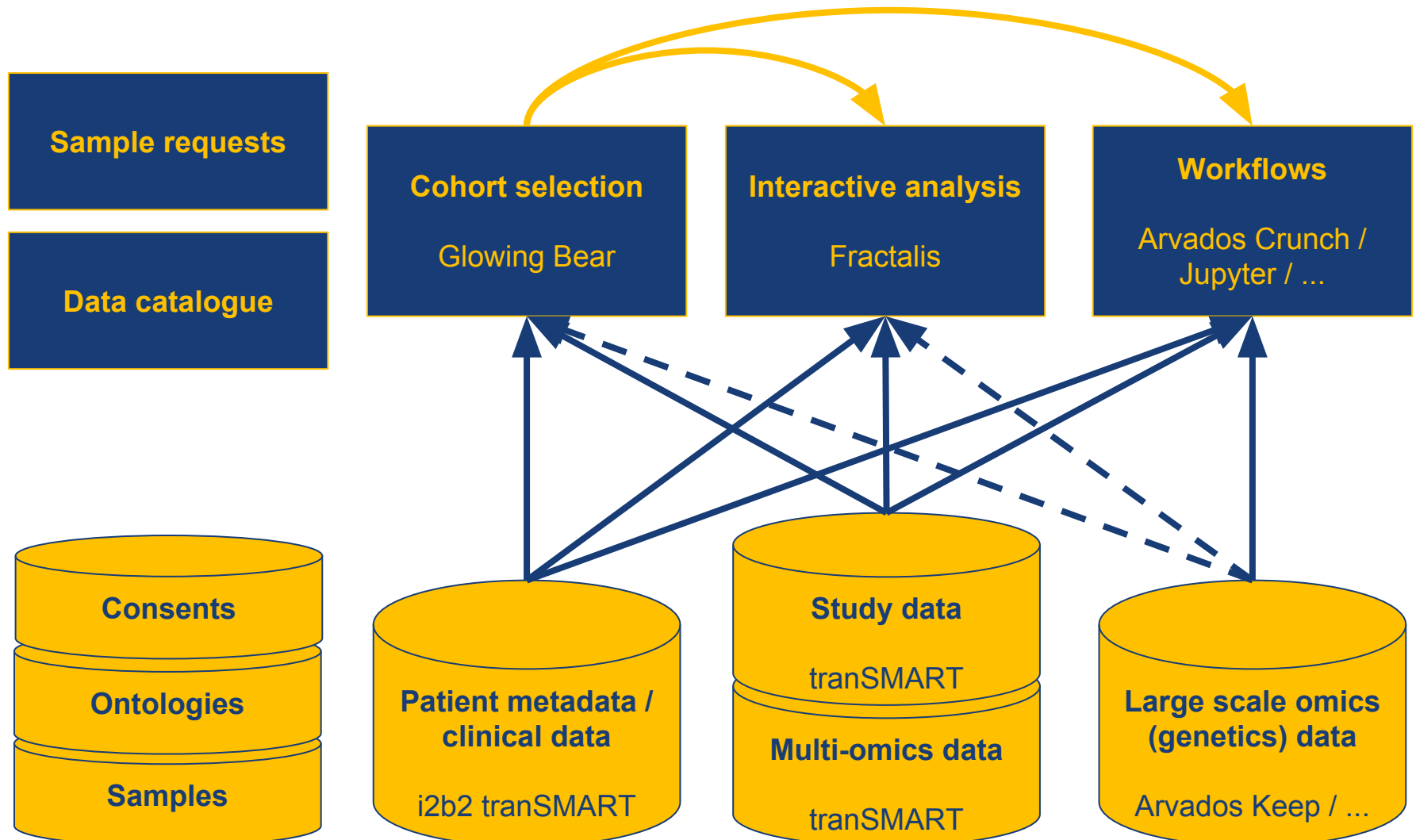


TranSMART 1.0



TranSMART 17.1

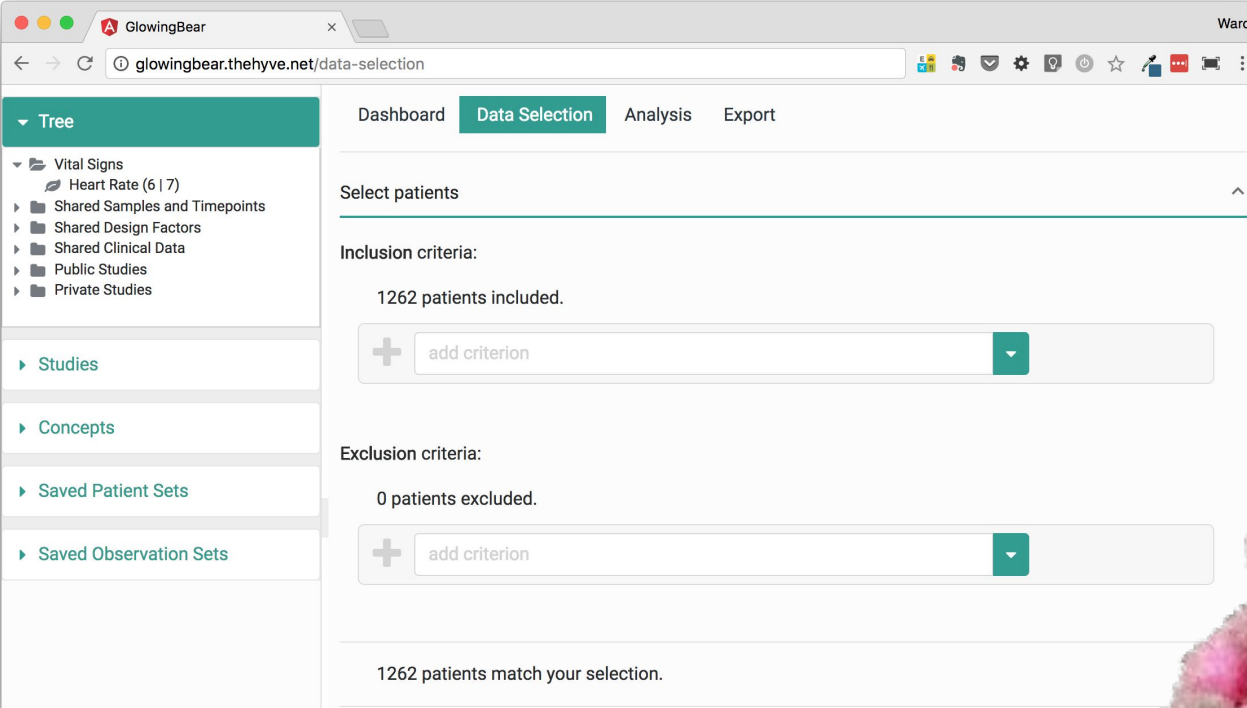
# Microservices architecture





# Glowing Bear cohort selector

# The **Glowing Bear** cohort selector



The screenshot shows a web browser window with the URL `glowingbear.thehyve.net/data-selection`. The interface includes a navigation menu on the left with sections for 'Tree', 'Studies', 'Concepts', 'Saved Patient Sets', and 'Saved Observation Sets'. The main content area has tabs for 'Dashboard', 'Data Selection', 'Analysis', and 'Export'. Under the 'Data Selection' tab, there is a 'Select patients' section with 'Inclusion criteria' showing '1262 patients included.' and 'Exclusion criteria' showing '0 patients excluded.' Each criteria section has an 'add criterion' button. At the bottom, it states '1262 patients match your selection.'

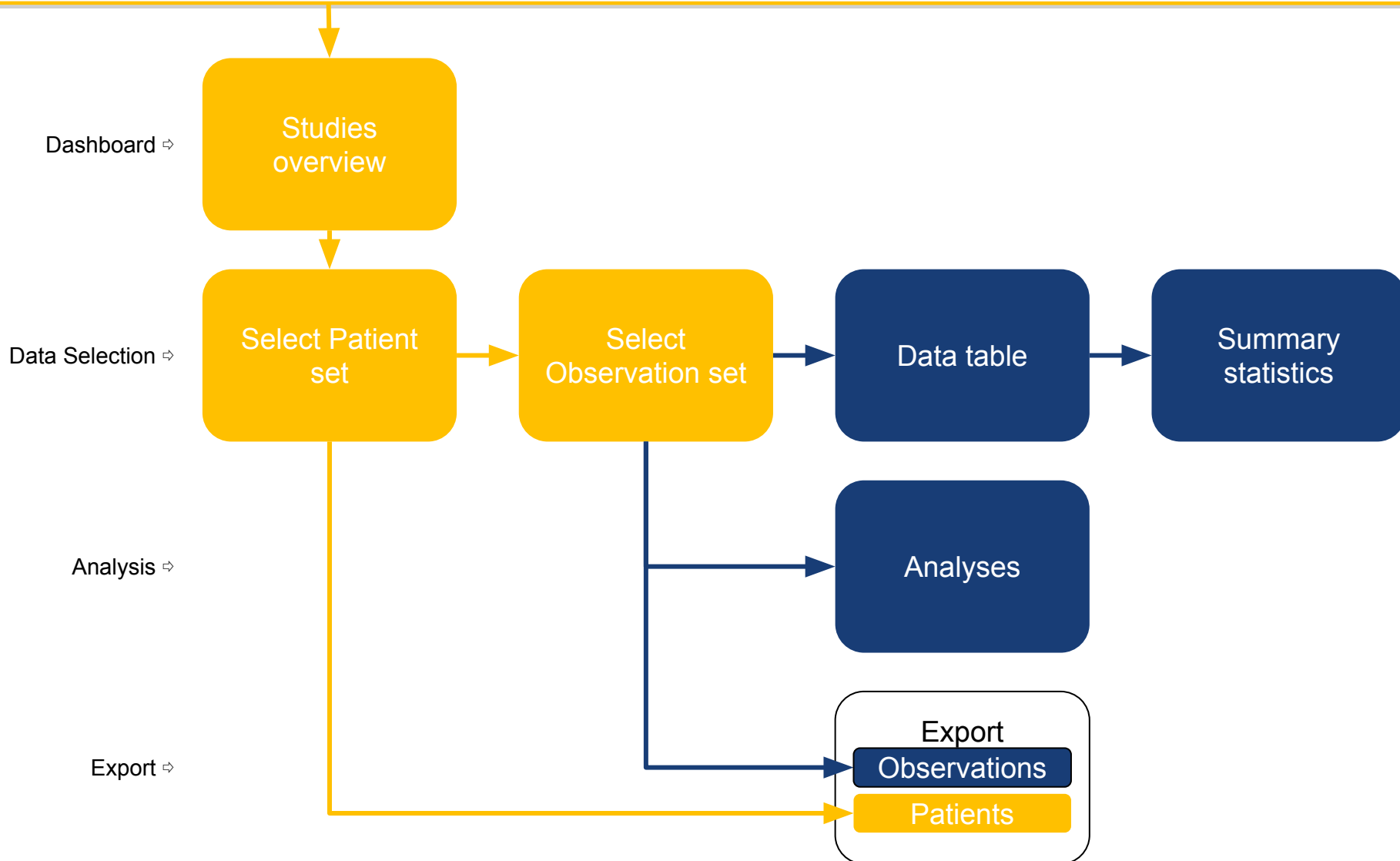


A modern user interface supporting the full tranSMART 17.1,  
via the tranSMART REST API V2

# The **Glowing Bear** cohort selector

- Funded by
  - Health-RI (BBMRI)
  - Netherlands Twin Registry
- Being installed in production now
- Working on integrations with partners
  - Fractalis (SmartR successor) interactive analysis
  - HiDome: Cohort selection on omics data
  - Jupyter Hub / Notebook
- Supports time series, samples and cross-study queries
- Built on tranSMART REST API Version 2 with Angular 4
- Open source: <https://github.com/thehyve/glowing-bear>

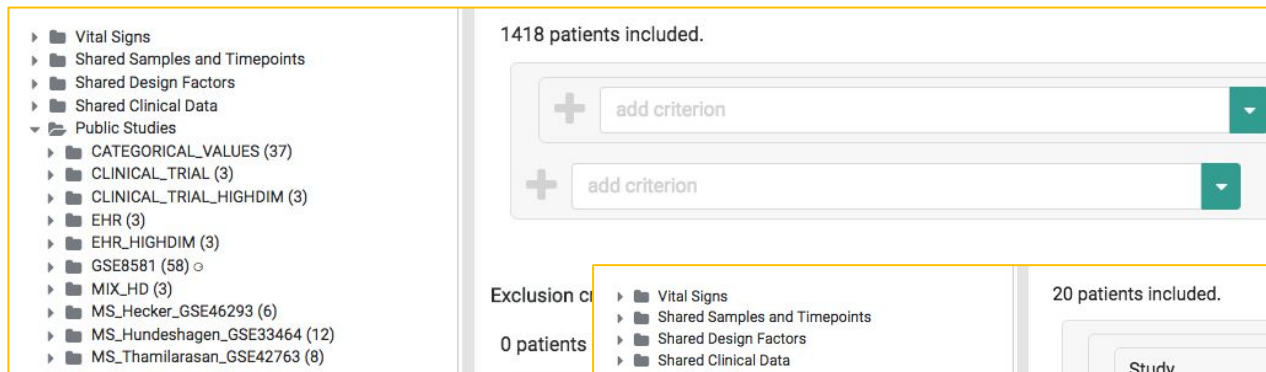
# The **Glowing Bear** cohort selector



# Cross-study

- Glowing Bear is cross-study by default.
- One or several studies can be dropped into filter box to limit subset to selected studies.

1418 subjects in total over all studies



1418 patients included.

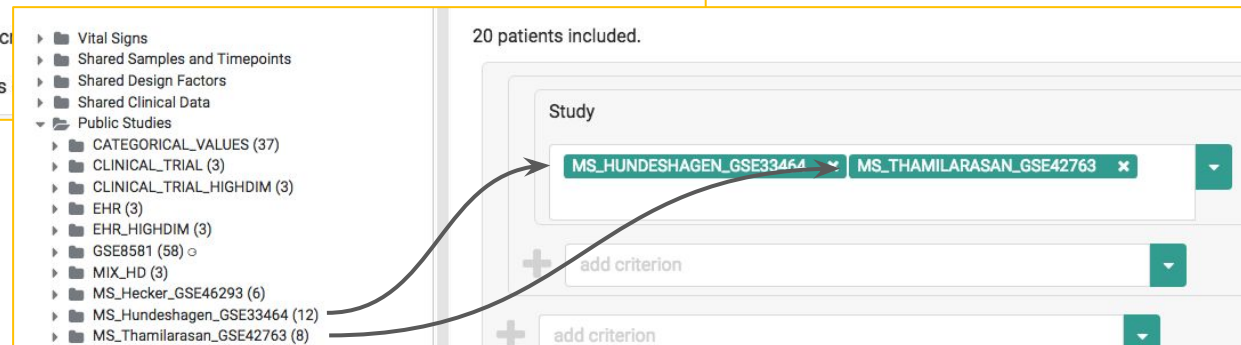
0 patients

Exclusion criteria

Public Studies

- ▶ CATEGORICAL\_VALUES (37)
- ▶ CLINICAL\_TRIAL (3)
- ▶ CLINICAL\_TRIAL\_HIGHDIM (3)
- ▶ EHR (3)
- ▶ EHR\_HIGHDIM (3)
- ▶ GSE8581 (58) ◦
- ▶ MIX\_HD (3)
- ▶ MS\_Hecker\_GSE46293 (6)
- ▶ MS\_Hundeshagen\_GSE33464 (12)
- ▶ MS\_Thamilarasan\_GSE42763 (8)

20 subjects in  
2 MS studies selected



20 patients included.

Study

MS\_HUNDESHAGEN\_GSE33464 MS\_THAMILARASAN\_GSE42763

Public Studies

- ▶ CATEGORICAL\_VALUES (37)
- ▶ CLINICAL\_TRIAL (3)
- ▶ CLINICAL\_TRIAL\_HIGHDIM (3)
- ▶ EHR (3)
- ▶ EHR\_HIGHDIM (3)
- ▶ GSE8581 (58) ◦
- ▶ MIX\_HD (3)
- ▶ MS\_Hecker\_GSE46293 (6)
- ▶ MS\_Hundeshagen\_GSE33464 (12)
- ▶ MS\_Thamilarasan\_GSE42763 (8)



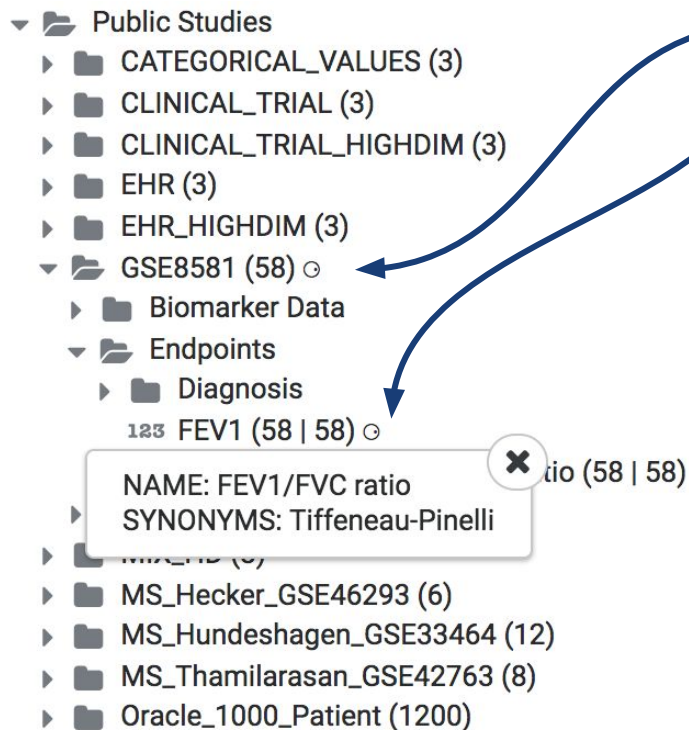
# Ontologies

- ▶ Vital Signs
- ▶ Shared Samples and Timepoints
- ▶ Shared Design Factors
- ▶ Shared Clinical Data
  - abc Duration from Diagnosis to Therapy Initiation (months) (26 | 26)
  - 123 Number of Relapses (26 | 95)
  - abc Treatment Type (26 | 26)
- ▶ Public Studies
  - ▶ CATEGORICAL\_VALUES (3)
  - ▶ CLINICAL\_TRIAL (3)
  - ▶ CLINICAL\_TRIAL\_HIGHDIM (3)
  - ▶ EHR (3)
  - ▶ EHR\_HIGHDIM (3)
  - ▶ MIX\_HD (3)
  - ▶ MS\_Hecker\_GSE46293 (6)
    - ▶ Biomarker Data
    - ▶ Clinical Data
      - 123 Duration from Diagnosis to Therapy Initiation (months) (6 | 6)
      - ▶ Duration of Treatment (months)
      - 123 EDSS Score (6 | 18)
      - 123 Number of Relapses (6 | 18)
      - abc Treatment Type (6 | 6)
    - ▶ Design Factors
    - ▶ Samples and Timepoints
    - ▶ Subjects
  - ▶ MS\_Hundeshagen\_GSE33464 (12)
    - ▶ Biomarker Data
    - ▶ Clinical Data
      - abc Disease (12 | 12)
      - ▶ Duration from Therapy Initiation to the First Relapse (months)
      - 123 Expanded Disability Status Scale (EDSS) (12 | 47)
      - 123 Number of Relapses (12 | 47)
      - ▶ Treatment
    - ▶ Subjects
  - ▶ MS\_Thamilarasan\_GSE42763 (8)

- Data can loaded at the same time as:
  - Shared concepts, mapped to standard ontologies
  - Study specific concepts, with study based hierarchy



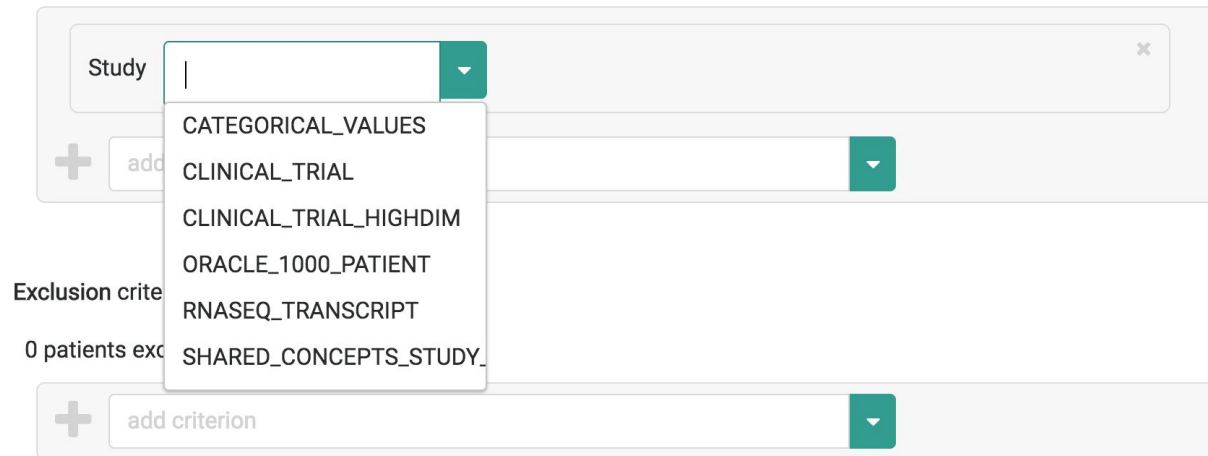
# Metadata



Icons indicate for which tree nodes metadata is available

- Supported on both studies and variables

- On Studies



Study |

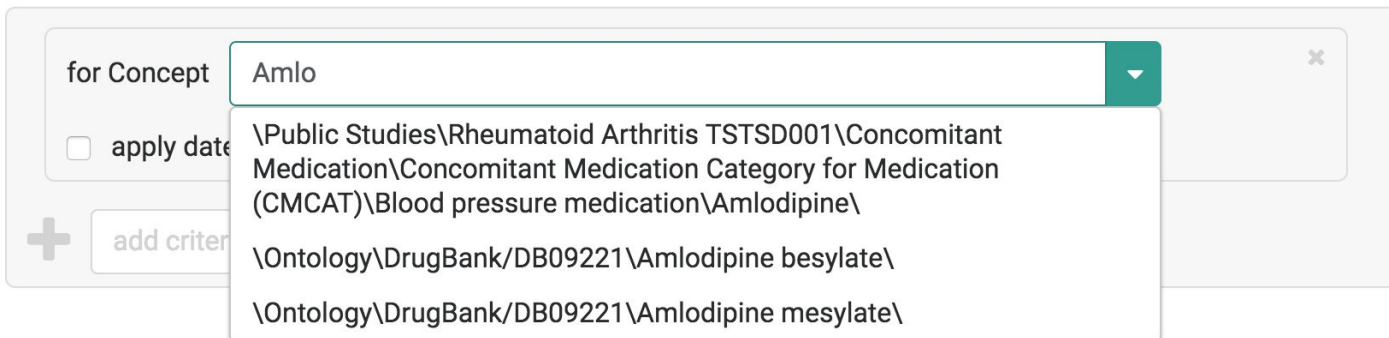
- CATEGORICAL\_VALUES
- CLINICAL\_TRIAL
- CLINICAL\_TRIAL\_HIGHDIM
- ORACLE\_1000\_PATIENT
- RNASEQ\_TRANSCRIPT
- SHARED\_CONCEPTS\_STUDY\_

Exclusion criteria

0 patients excluded

+ add criterion

- On Concepts



for Concept Amlo

- \Public Studies\Rheumatoid Arthritis TSTD001\Concomitant Medication\Concomitant Medication Category for Medication (CMCAT)\Blood pressure medication\Amlodipine\
- \Ontology\DrugBank/DB09221\Amlodipine besylate\
- \Ontology\DrugBank/DB09221\Amlodipine mesylate\

apply date

+ add criterion

# Filtering on time series

01	NSG001	Moderate	7/2/15	9/23/15	Headache
01	NSG001	Moderate	8/2/15	10/23/15	Headache
01	NSG001	Moderate	9/2/15	11/23/15	Headache
01	NSG404	Mild	1/31/16	2/7/16	Insomnia
01	NSG405	Mild	2/1/16	2/8/16	Insomnia
01	NSG410	Mild	2/2/16	2/9/16	Insomnia
01	NSG411	Mild	2/3/16	2/6/16	Insomnia
01	NSG012	Severe	6/12/16	6/19/16	Tuberculosis
01	NSG013	Mild	6/13/16	6/20/16	Diarrhea
01	NSG016	Mild	6/14/16	6/21/16	Diarrhea
01	NSG018	Mild	6/15/16	6/22/16	Diarrhea
01	NSG021	Moderate	6/16/16	6/23/16	Diarrhea
01	NSG023	Moderate	6/17/16	6/24/16	Diarrhea
01	NSG025	Mild	8/18/16	8/25/16	Diarrhea
01	NSG027	Mild	8/19/16	8/26/16	Diarrhea
01	NSG029	Mild	8/20/16	8/27/16	Diarrhea
01	NSG200	Mild	8/21/16	8/28/16	Diarrhea
01	NSG201	Mild	8/22/16	8/29/16	Diarrhea
01	NSG202	Mild	8/23/16	8/30/16	Diarrhea
01	NSG300	Mild	8/24/16	8/31/16	Insomnia
01	NSG301	Moderate	8/25/16	9/1/16	Insomnia
01	NSG303	Moderate	8/26/16	9/2/16	Insomnia
01	NSG305	Mild	8/27/16	9/3/16	Insomnia
01	NSG307	Mild	8/28/16	9/4/16	Insomnia
01	NSG400	Mild	8/29/16	9/5/16	Insomnia
01	NSG401	Severe	8/30/16	9/6/16	Squamous ce

6 patients included.

for Concept

with values

and start date *between*  and

+ add criterion

2 patients included.

for Concept

with values

and start date *between*  and

and

for Concept

with values

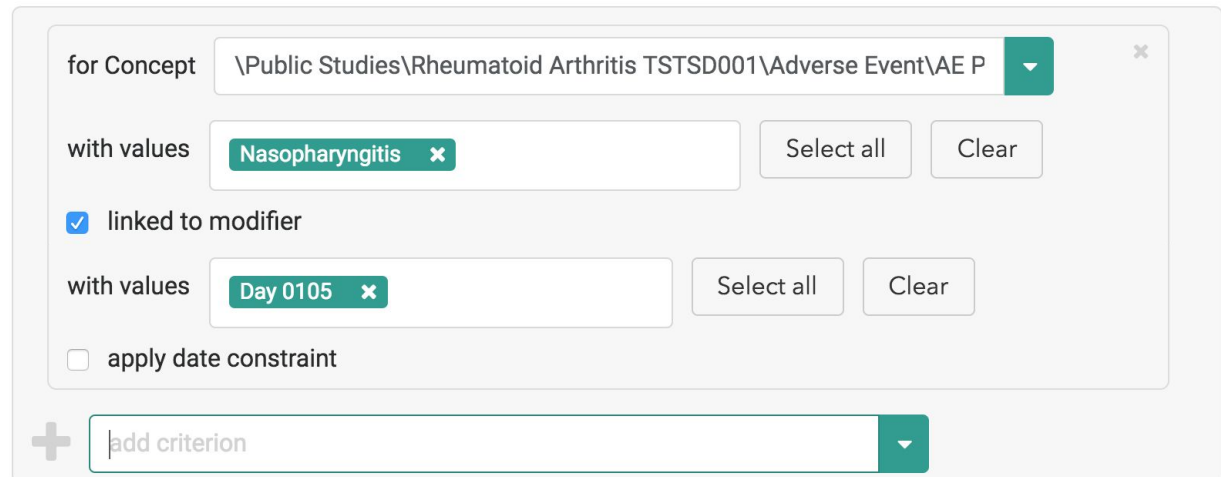
apply date constraint

+ add criterion

# Filtering on modifiers (coming soon)

STUDY_ID	SUBJ_ID	AE Preferred Term (AE Study Day of End of Event)	
TSTSD001	NSG001	Headache	Day 0105
TSTSD001	NSG001	Nasopharyngitis	Day 0105
TSTSD001	NSG003	Headache	Day 0105
TSTSD001	NSG003	Nasopharyngitis	Day 0105
TSTSD001	NSG004	Tuberculosis	Day 0096
TSTSD001	NSG006	Nasopharyngitis	Day 0096
TSTSD001	NSG010	Headache	Day 0096
TSTSD001	NSG027	Diarrhoea	Day 0105

2 patients included.



The screenshot shows a search filter configuration window. It includes a dropdown menu for the concept, two input fields for values with 'Select all' and 'Clear' buttons, and checkboxes for 'linked to modifier' and 'apply date constraint'. A plus sign and an 'add criterion' input field are at the bottom.

for Concept

with values

linked to modifier

with values

apply date constraint

+



# Glowing Bear

## Modern cohort selector on the i2b2 tranSMART database

- New since the June Harvard i2b2 tranSMART meeting:
  - Data tree: Concept search, dynamic loading and interactive
  - Observation filtering on concepts
  - Data export
  - Query management
- Adopted by Health-RI, child cancer center Princess Maxima Center, Dutch twin registry (NTR), US Pharma
- Coming soon
  - Integration with Fractalis for interactive analytics
  - Integration of HiDome for cohort selection on Omics
  - Export to SPSS, filtering on family relations, strong test suite, ...

[More information  
and demo server  
HERE](#)



We empower scientists by building on open source software

### Our collaborators:

Sascha Herzinger - Université du Luxembourg  
Denny Verbeeck - Johnson & Johnson  
Matteo Gabetta, Mauro Bucalo - Biomeris

### Our team, in particular:

Bo Gao, Wibo Pipping, Gijs Kant  
Jochem Bijlard, Ruslan Forostianov



**Kees van Bochove**  
kees@thehyve.nl



**Ward Weistra**  
ward@thehyve.nl

# We're also making data loading easier

## “The Arborist and *tmtk*: A researcher friendly data loading toolkit for tranSMART”

- The Arborist: [Demo](#) / [Github](#)
- *tmtk*: [Docs](#) / [PyPi](#) / [Github](#)

**Details: Categorical**  
File: OBS336-201\_pk.txt (7)

Label: Treatment  
DLS:  
CVCD: SNOMEDCT/74040009

Search ontology  
Label: Protein measurement  
Code: SNOMEDCT/74040009

- Observable entity
  - Social / personal history observable
  - Clinical history/examination observable
- Procedure
  - Procedure by method
    - Evaluation procedure
      - Measurement
        - Measurement of substance
          - Lipids measurement
            - Lipoproteins measurement
              - Low density lipoprotein measurement
              - High density lipoprotein measurement
            - Protein measurement

Apply changes Templates Exit

### The Arborist and *tmtk*: A researcher friendly data loading toolkit for tranSMART

Jochem Bijlard<sup>1</sup>, Ward Weistra<sup>2</sup>, Stefan Payralbe<sup>3</sup>, Natalia Boukharov<sup>4</sup>, Marijka Biekens<sup>5</sup>, Annemieke Hiemstra<sup>1</sup>, Remond Fijneman<sup>2</sup>, Jan-Willem Bolten<sup>6</sup>, Gerrit Meijer<sup>3</sup>

Correspondence: jochem@thehyve.nl, ward@thehyve.nl

**From Excel to tranSMART in five simple steps**

- Import:** start the import wizard to create a study based on your study data.
- Validate:** let the toolkit check the tranSMART-specific requirements.
- Edit:** make changes to your tree with the visual Arborist editor.
- Save:** store the study on disk as tranSMART-ready staging files.
- Load:** use tranSMART-batch to load your data to tranSMART.

**The Arborist Visual editor**  
Collaborate on data modelling with non-technical data experts in the secure Arborist web application.

- Restructure the tranSMART tree with drag and drop
- Rename variables and values
- Add and edit metadata for any tree node
- Work with both low and high dimensional data

Try it at <http://arborist-test-trait.thehyve.net/demo>  
Code at <https://github.com/thehyve/arborist> under GPL v3 license.

**tmtk Python library**  
Library that allows users to create and load studies without the need for tranSMART specific knowledge.

- Quickstart studies from tabular files (eg. XLS, TSV, CSV)
- Extensive dataset validation
- Use The Arborist directly embedded into Jupyter Notebook
- Load studies to The Arborist web application for collaboration
- Many functions to work with low and high dimensional data
- Minimal technical and tranSMART specific knowledge required

Install for Python3: `pip install tmtk`  
Documentation: <https://tmtk.readthedocs.io>  
Code at <https://github.com/thehyve/tmtk> under GPL v3 license.

**tmtk notable python commands**

<b>create_study</b> Create a study from a tabular file. Quickstart your treatment study.	<b>validate_all</b> Many of the objects in the data have validating methods. These methods can easily be extended by adding more.	<b>tranSMART abstract</b> Visual view and drag editor for the tranSMART concept tree. Change word mappings, use metadata, and map concepts to ontologies.
<b>visual_create_study</b> Create a study from a tabular file. Quickstart your treatment study.	<b>load_data</b> Loading the data into tranSMART is a wrapper for tranSMART-batch. For easy use and better progress bars!	<b>search</b> Search the Arborist embedded into Jupyter.
<b>visual_test</b> Generate fully randomized study object, great for testing stuff!	<b>save_to_file</b> Load your study to tranSMART from Jupyter or the command line.	<b>publish_to_batch</b> Send data tree to Arborist web application for easy collaboration.