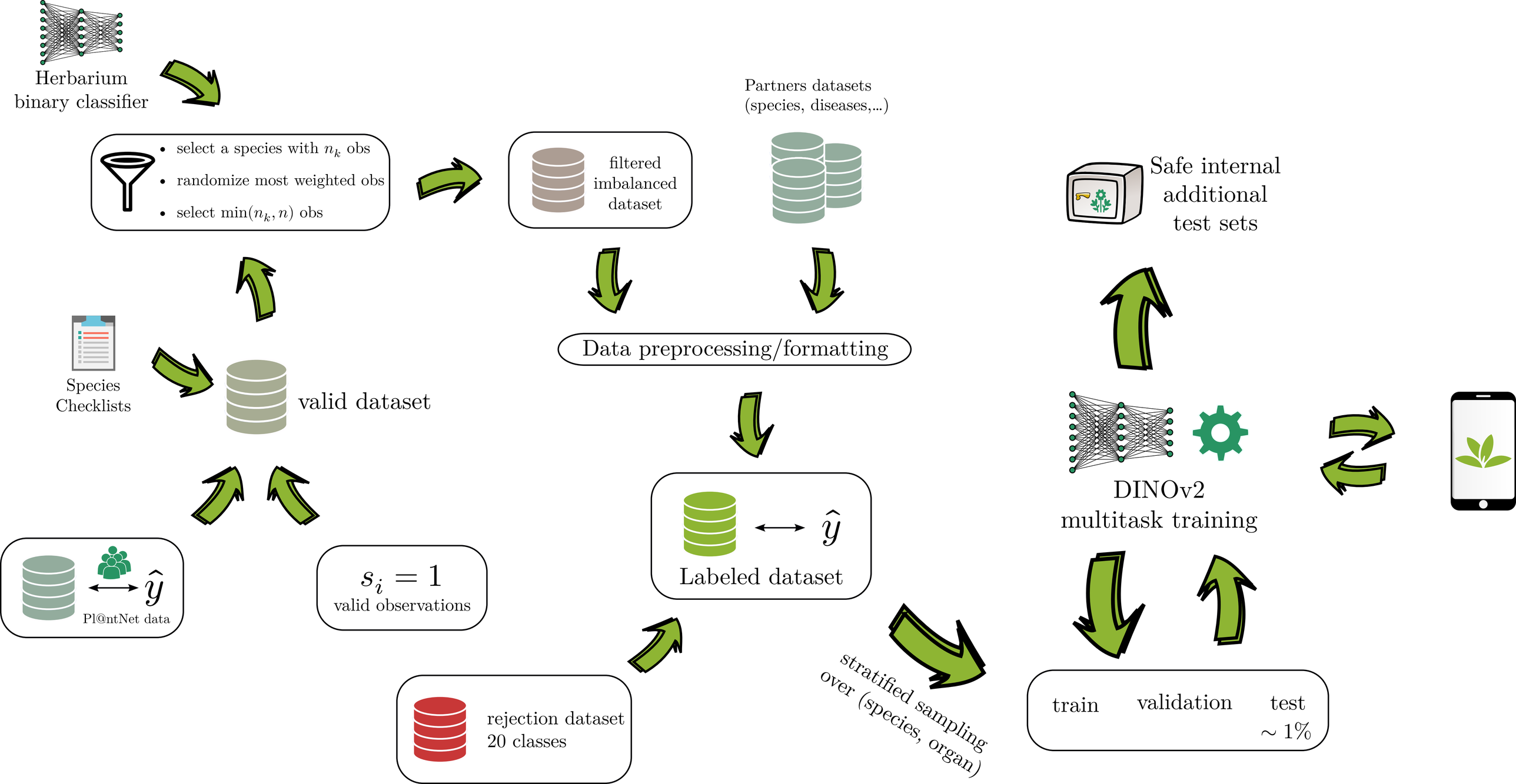
Herbarium
binary classifier
select a species with n_k obs
randomize most weighted obs
select $\min(n_k, n)$ obs
filtered
imbalanced
dataset
Partners datasets
(species, diseases,...)
Safe internal
additional
test sets
Species
Checklists
valid dataset
Data preprocessing/formatting
DINOv2
multitask training
$\hat{y}$
Pl@ntNet data
$s_i = 1$
valid observations
Labeled dataset
rejection dataset
20 classes
stratified sampling
over (species, organ)
train
validation
test
$\sim 1\%$