

SNOW WHITE & THE 7 KEYS TO DATA LAKE HYDRATION

AGENDA INCLUDING INTROS

KATE McMAHAN



KATE McMAHAN

ANONYMIZATION
SPECIALIST

- Kate @ instem
- Royal Pains – The Problems that Need Solving
- Why Your Sleeping Data is the Fairest in the Land
- True Love's Kiss, how to maximize your data
- The Seven Keys that can help
- Happily Ever After

ROYAL PAINS – THE PROBLEMS THAT NEED SOLVING

- Drugs are expensive because each one must pay for 10 trials
- Clinical Trials are expensive
- Clinical Trials Fail
- Data + CSR submitted to FDA/EMA
- Often this data is never looked at again
- Due to GDPR, the data doesn't belong to you

We need to move beyond compliance,
so we can realize the value that we're holding on to!



YOUR DATA IS THE FAIREST IN THE LAND

- Cleanest Data
- Most Accurate

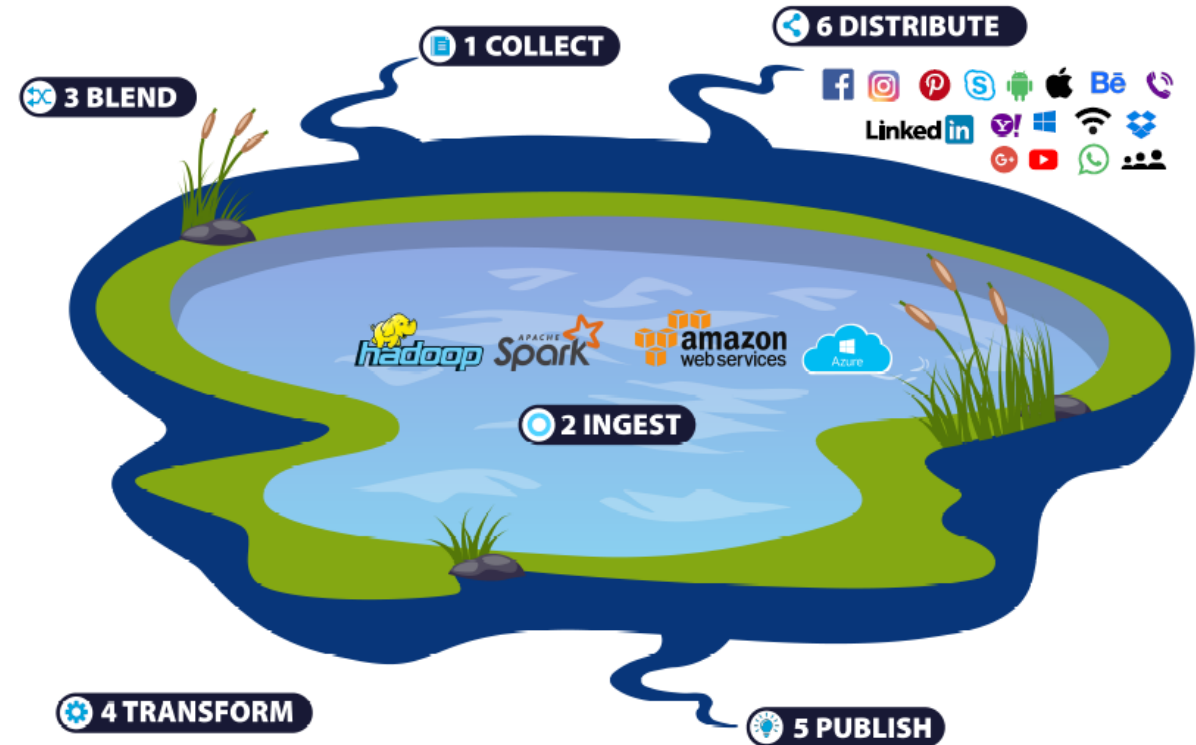
BUT IT'S ASLEEP!

- So much utility, but is just sitting on shelves (or in servers)
- As is, you cannot use it again



AWAKEN THE VALUE OF YOUR DATA WITH A CHARMING DATA LAKE

- Oil = Black Gold
- Data = ~~New Oil~~
- Data >>>> Oil





ANONYMISATION – SLEEPY

- You cannot use data until it is anonymized
- Anonymization breaks the link between the data and the patient

AUTOMATED ANONYMIZATION – HAPPY

- Doing this study by study would be slow and expensive
- This process can be automated using templates
- Quantitative risk assessments can be automated
- Rule selection can be automated
- Anonymization Software





DATA MAPPING - GRUMPY

- Map your data into a single format so it is easier to consume
- When do you map... before or after anonymization?

APPROPRIATE PLATFORM/TOOLS - DOPEY

- Software that helps engineer more efficient trails
- Visualization Tools (Spotfire, SAS Visual Analytics, etc.)
- Machine learning to use anonymized data to predict challenges
- Python
- R
- SAS





PEOPLE - BASHFUL

- Do you really need data scientists to get started?
- SAS programmers!

```
2=proc surveyselect data=dm_ae_ser
3   method=srs n=500 out=train;
4 run;
5
6=data AE_Predict;
7 set dm_ae_ser;
8 drop aesern;
9 run;
10
11=proc discrim data=train
12             test=AE_Predict
13             testout=Predict_out
14             method=npair k=10
15             testlist;
16   class aesern;
17   var sexn age countryn racen ethnicn actarmcdn siteidn;
18 run;
19
```

https://phuse.s3.eu-central-1.amazonaws.com/Archive/2022/Connect/EU/Belfast/POS_PP10.pdf

APPROPRIATE ACCESS – SNEEZY

- Summary Dashboards
- Summary Tables
- Visualization Tools (Spotfire, SAS Visual Analytics, etc.)
- Individual Patient Level Data (anonymized)
- Automated Trial Design Documents (creating trial design without having access to patient level data)
- Ability to include/exclude data from algorithms
- Limit Data Access (summary level vs. individual level)
- Problem (how/what to restrict and to who)



COMBINATION OF DATA - DOC

- Pull in other sources
- Public data e.g., chemical make up of compounds and how they relate to safety/efficacy
- Pre-clinical data
- Genetic Data
- Real world evidence
- Actual Data Lake: eTRANSAFE
- clinicaltrials.gov
- dailymed.nlm.nih.gov
- eTOXsys
- FAERS
- Biomarkers Databases
- Sponsor Data clinical/pre-clinical





OUTCOMES- HAPPILY EVER AFTER

- Higher trial success
- Lower trial cost
- Better recruitment
- Fraud detection
- Safer Trials
- Cheaper drugs
- Personalized Medicine
- Better lives

WHAT DID YOU LEARN

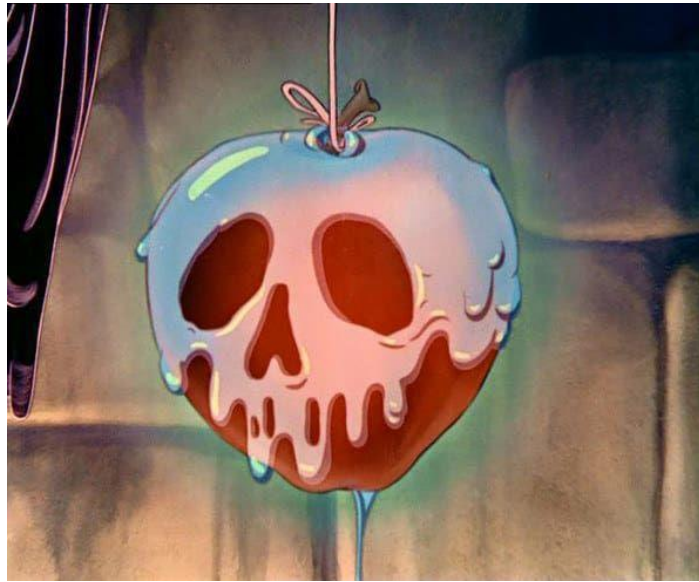
- Anonymization can start to awaken your data!
- Automated Anonymization will make you Happy by making the process better!
- Think about the benefits and challenges of Data Mapping. But do it!
- Use the tools you already have!
- Use excellent people/resources to get you started!
- Consider data lake access and potential!
- Think about data sources beyond that of clinical!



QUESTIONS & CONTACT US

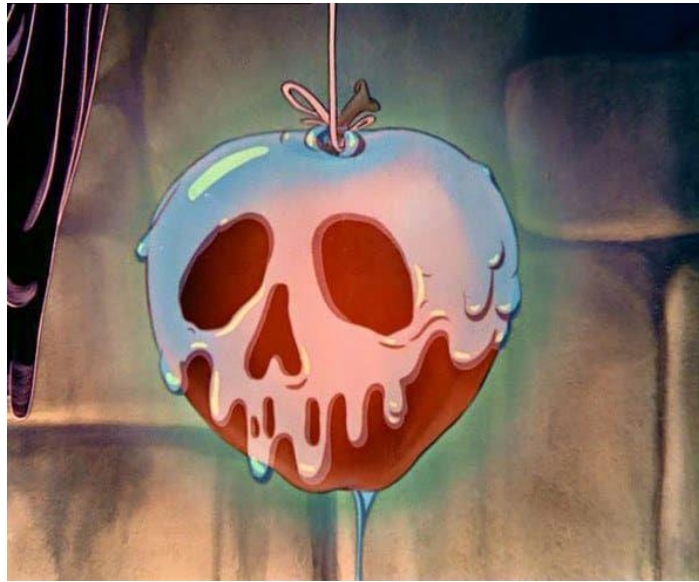


POP QUIZ!



POP QUIZ!

WHAT IS MY FAVORITE DWARF?





POP QUIZ!

What is MY favorite dwarf?

HAPPY!

POP QUIZ!

What Grumpy challenge do we need to be innovating to overcome?



POP QUIZ!

What Grumpy challenge do we need to be innovating to overcome?

Data Mapping!





POP QUIZ!

Give me three ways Data Lake Hydration can lead to a happily ever after:

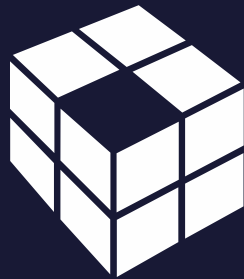


POP QUIZ!

Give me three ways Data Lake Hydration can lead to a happily ever after:

- Higher trial success
- Lower trial cost
- Better recruitment
- Fraud detection
- Safer Trials
- Cheaper drugs
- Personalized Medicine
- Better lives

QUESTIONS & CONTACT US



INSTEM

instem.com | info@instem.com