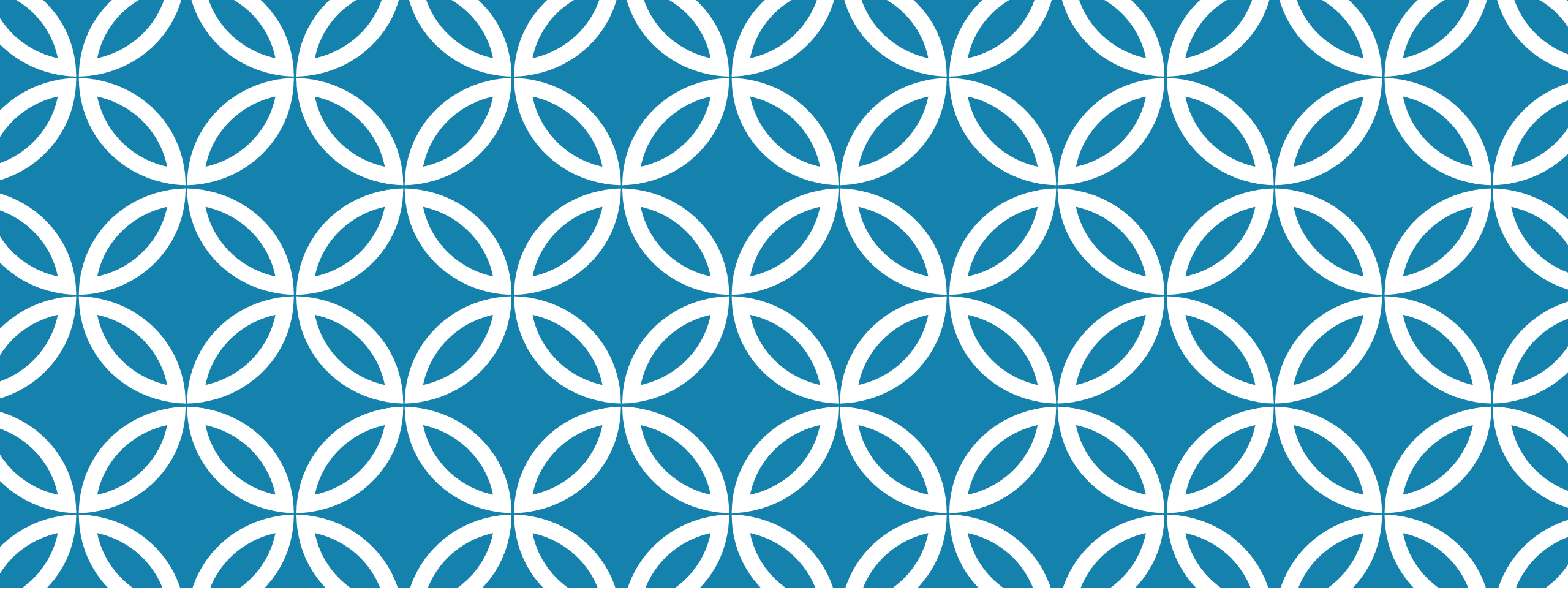


CAPSTONE PROJECT

DSI 3

Capstone Project

Part	Description	Deadline
1	Pitch	23rd March 2018, Friday
2	Dataset	13th April 2018, Friday
3	EDA	20th April 2018, Friday
4	Findings	11th May 2018, Friday
5	Presentation	18rd May 2018, Friday

SUMMARY OF SYLLABUS + ADDITIONAL DATA SCIENCE TOOLKIT

Algorithms:

Regression
- Linear
- Logistic

Classification and Regression Tree
- CART

Bayesian Algorithms
- Naïve Bayes

Clustering Algorithms
- K-means
- Hierachical

Neural Network
- Perceptron
- Back propagation

Deep Learning
- Convolutional Neural Networks

Ensemble
- Boosting
- Bagging
- Random Forests

Others..
- Natural Language Processing
- Recommender systems
- Reinforcement Learning

https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/

CHOOSING A CAPSTONE PROJECT

1. Problem statement approach
2. Dataset approach

DATASET LINKS

https://www.kaggle.com/datasets

https://archive.ics.uci.edu/ml/datasets.html

https://en.wikipedia.org/wiki/List_of_datasets_for_machine_learning_research

http://dataportals.org

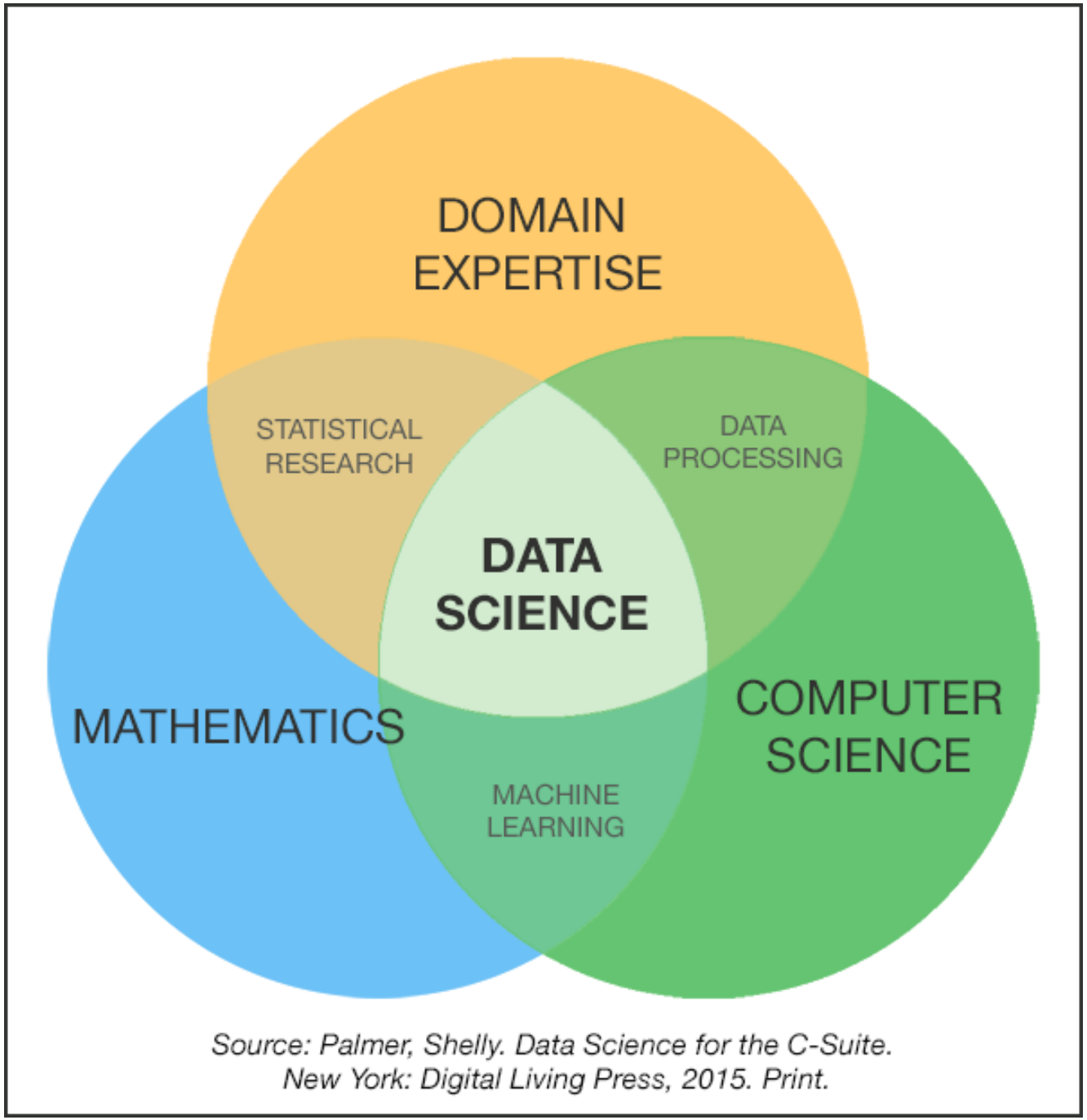
http://mldata.org/repository/data/

https://data.gov.sg

https://aws.amazon.com/public-datasets/

https://github.com/awesomedata/awesome-public-datasets

https://www.kdnuggets.com/datasets/index.html



Source: Palmer, Shelly. Data Science for the C-Suite. New York: Digital Living Press, 2015. Print.