



Society of Actuaries in Ireland

INC Carlow - Building a Data Analytics framework

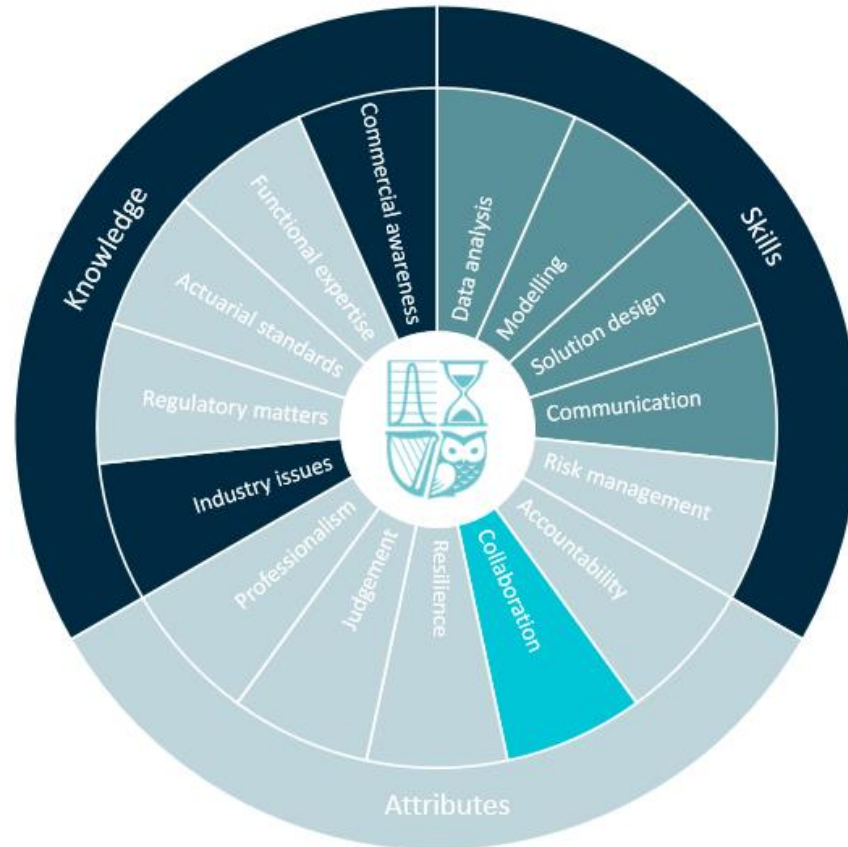
8th January 2021

© Society of Actuaries in Ireland

Disclaimer

The views expressed in this presentation are those of the presenter(s) and not necessarily those of the Society of Actuaries in Ireland or their employers.

Competency Framework Wheel

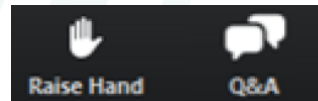


Q&A

Please click on the 'Raise Hand' icon
to ask a question
and
wait to be unmuted

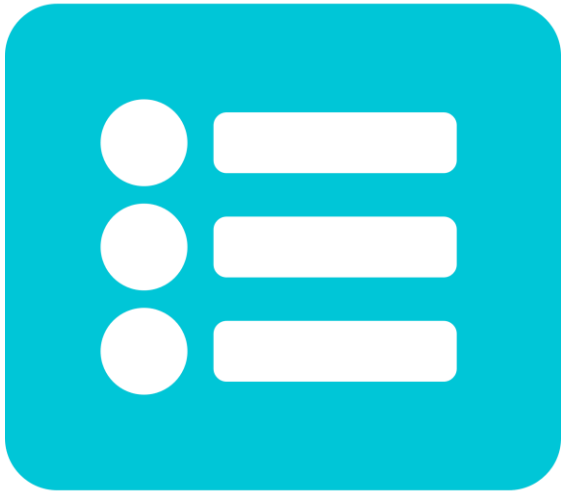
or

Use the Q&A function



Q&A is recorded!

Agenda



- **Dr. Greg Doyle** – How perform data analytics effectively and the components/methodology for a successful analytics project.
- **Pause for Q&A**
- **Aidan Mahon** – Insights into InsurTech scene and IT Carlow INC (InsurTech Network Centre).
- **Closing Q&A**

Speakers



A unique and exciting innovation

Insurtech Network Centre is a unique and exciting **Collaborative Insurance Innovation Platform** bringing established Insurance sector firms together with Insurtech Startups and the research, development and educational resources of Institute of Technology Carlow.

Member of:



Supported By:

Speakers



Dr. Greg Doyle

Dr. Greg Doyle is a lecturer in Computing Department at the Institute of Technology Carlow .

He is a programme director of the MSc in Data Science, and is a principal investigator and researcher in ITC's GameCORE (ICT centre of research, innovation and enterprise).

He is a Research Programme Board Director in Computing, Chair of the Research & Development committee of Academic Council, member of the Ethics Board and a long standing member of ITC's Academic Council as well as co-chair on a TUSEI (Technological University of the South East of Ireland) working group for Research & Innovation strategy development.

With over 20 years of experience in education, software engineering and data science research, innovation and their application, he is well placed to leverage the benefit of data science techniques across the insurance industry.

Speakers



Aidan Mahon

Mr. Aidan Mahon is the Manager of the Insurtech Network Centre (INC) and also works as an Associate Lecturer (Economics, Strategy, EBusiness and Research Supervision) with Institute of Technology Carlow.

He is also involved with the Research and Innovation (industry engagement) and Lifelong Learning working groups for the TUSEI (Technological University of the South East of Ireland) project.

Prior to the INC role Aidan had over 10 years' experience in the commercial insurance broking sector.

Aidan is keen to bridge the gap and foster collaboration between Insurtechs, Academia and Industry.



Society of Actuaries in Ireland

**Institute of Technology Carlow – Building a Data Analytics
framework**

8th January 2021

Data, data everywhere...

- My observations come from:
 - Personal experiences from teaching, research and professional consultancy work
 - Advisory engagements with various industries, organisations and SME's
 - Discussions with colleagues and company executives

Data, data everywhere...

“Drowning in data, but starving for insights” - Deloitte 2018

- What is the value of data in your business?
- Is the data related to business outcomes?
- Is the data producing business value?



Agenda

- Problem - where to start
- Framework/process model – need repeatability and reliability
- Learning/feedback loop – learning from failure/fail fast
- Real world examples
- Making a business case for analytics
- Technologies & tools to assist data (non data) scientists
- Industries doing data science well

Analytics - organisational challenges

- Model – centralised, decentralised, hybrid
- Legacy systems
- Data collection and integration
- ROI
- Business functions targeted
- Changing the culture to a truly data-driven culture

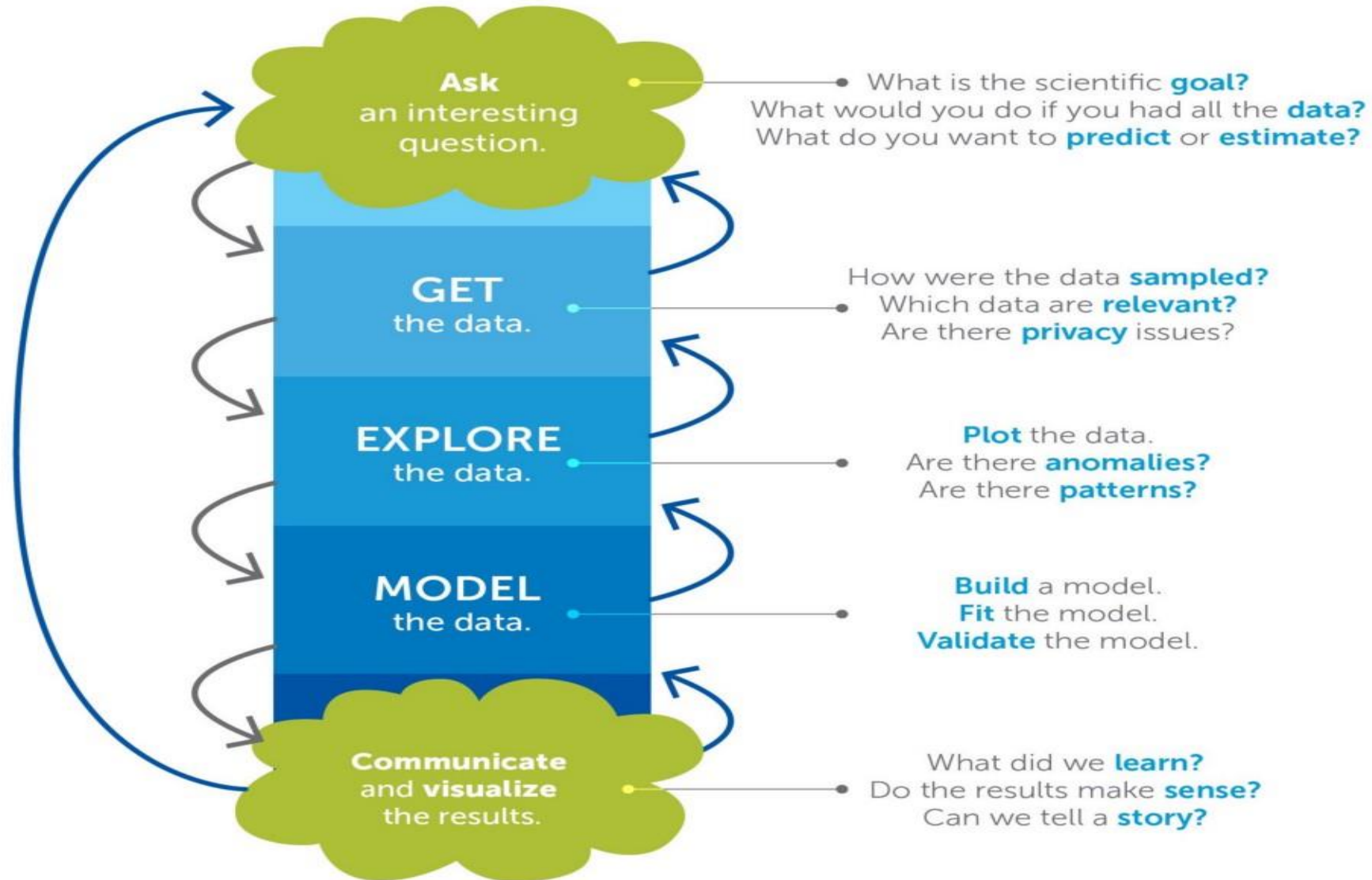
Analytics – where to start?

- Organisation-wide data science team
 - Takes time to see results, can be risky, hard to get buy in
- Organisation wide approach
 - Risky and potentially expensive
- Large-scale long-term investment
 - May be too long term
- Crack data science business approach – emerging as best practice
 - Cheaper high-value short term wins show the way

Analytics frameworks - process models

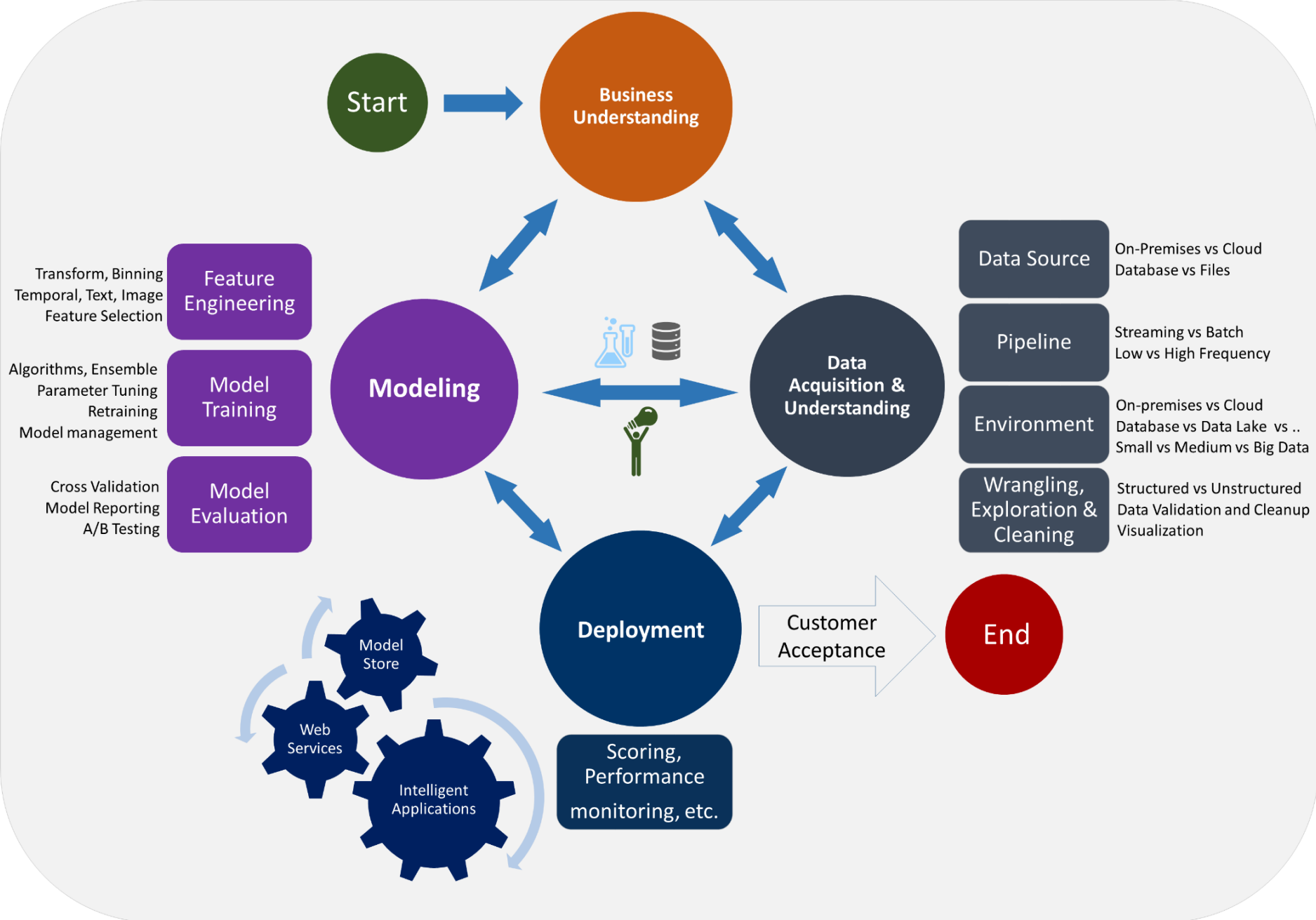
- Process reliable and repeatable by people no background
- Framework for recording experience
 - Allows projects to be replicated, scientific
- Aids project planning and management
- Comfort factor for new adopters
 - Demonstrates maturity of data mining
 - Reduces dependency on specific data mining experts

The Data Science Process

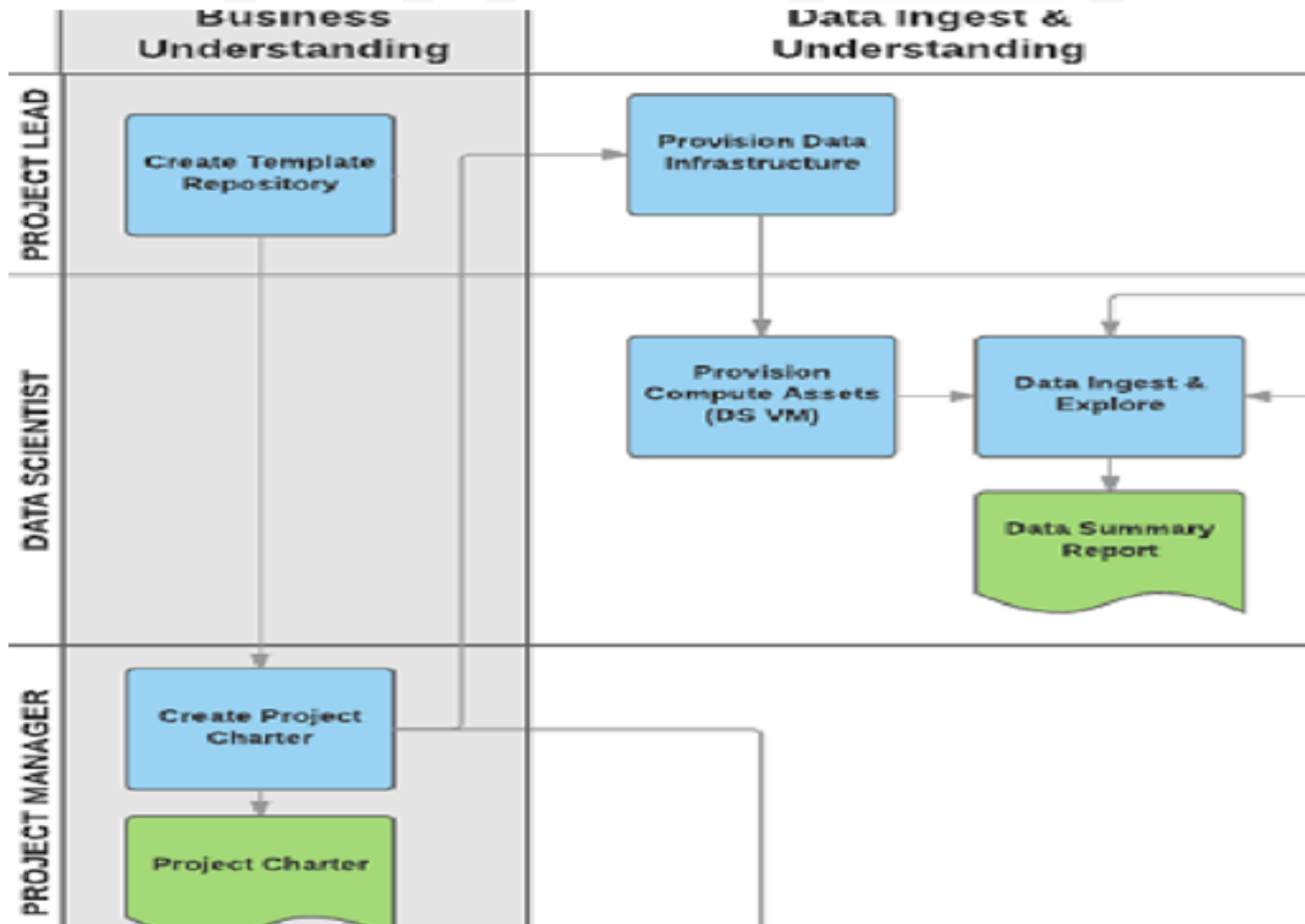


Derived from the work of Joe Blitzstein and Hanspeter Pfister, originally created for the Harvard data science course <http://cs109.org/>.

Data Science Lifecycle



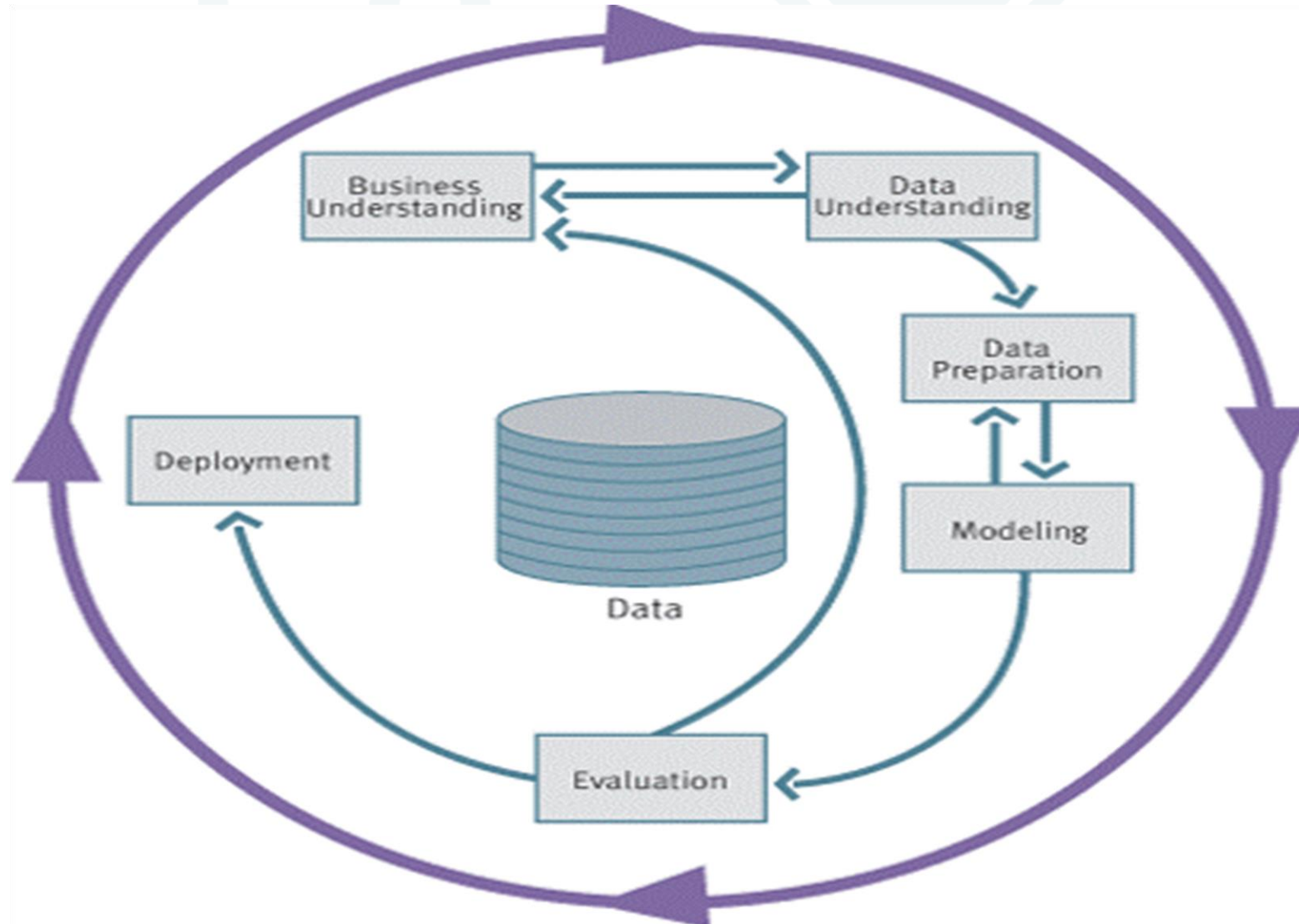
Microsoft TDSP - Tasks & artefacts



Microsoft TDSP - Key components

- **A data science agile, iterative lifecycle** definition
- **A standardized collaborative (team) project structure**
- **Infrastructure and resources** for data projects – on-site/cloud datasets/DB, big data (SQL or spark) clusters, ML services (Azure Machine Learning)
- **Tools and utilities** recommended for project execution
- Source: <https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/overview>

CRISP-DM – Process model



CRISP-DM – Hierarchical model

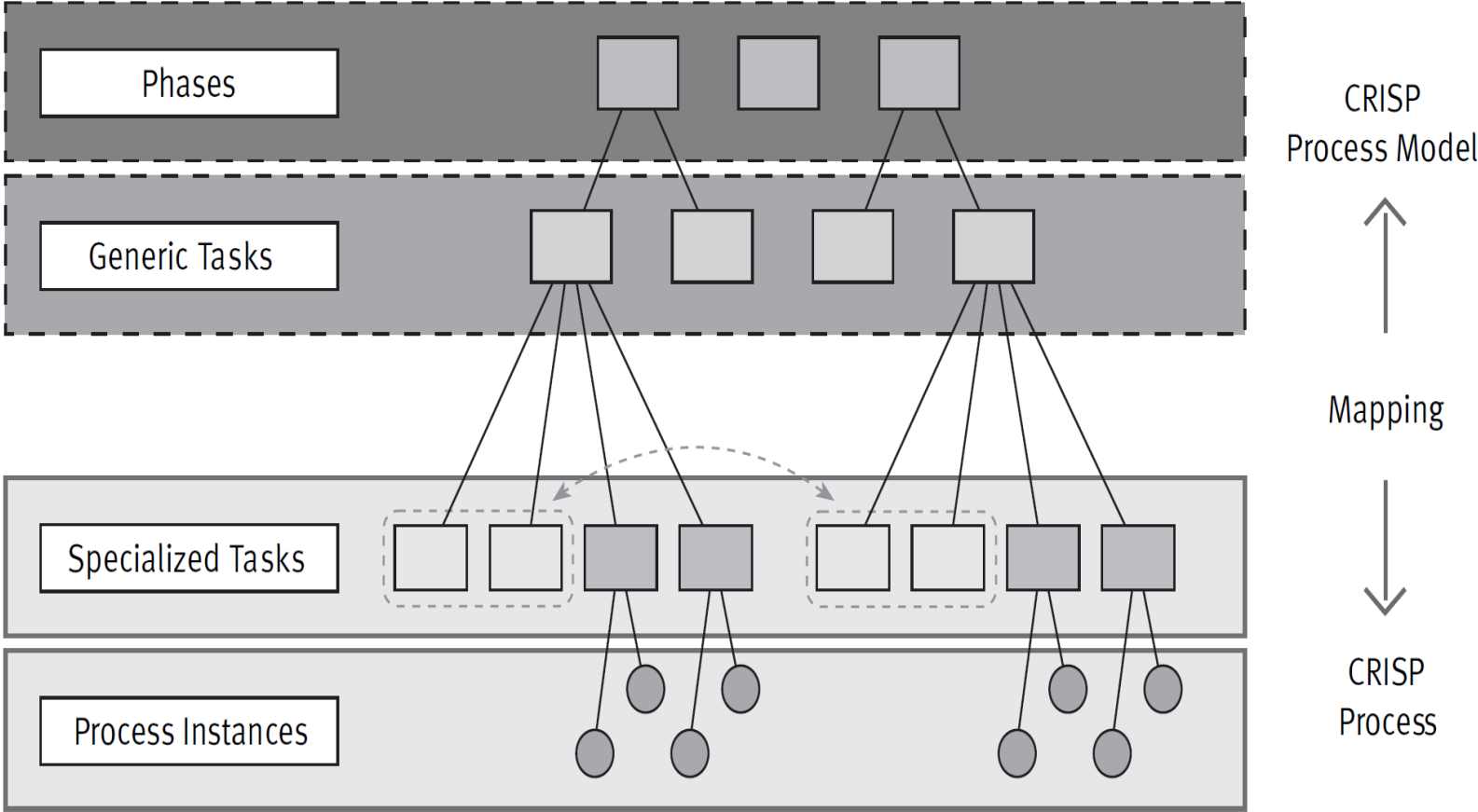
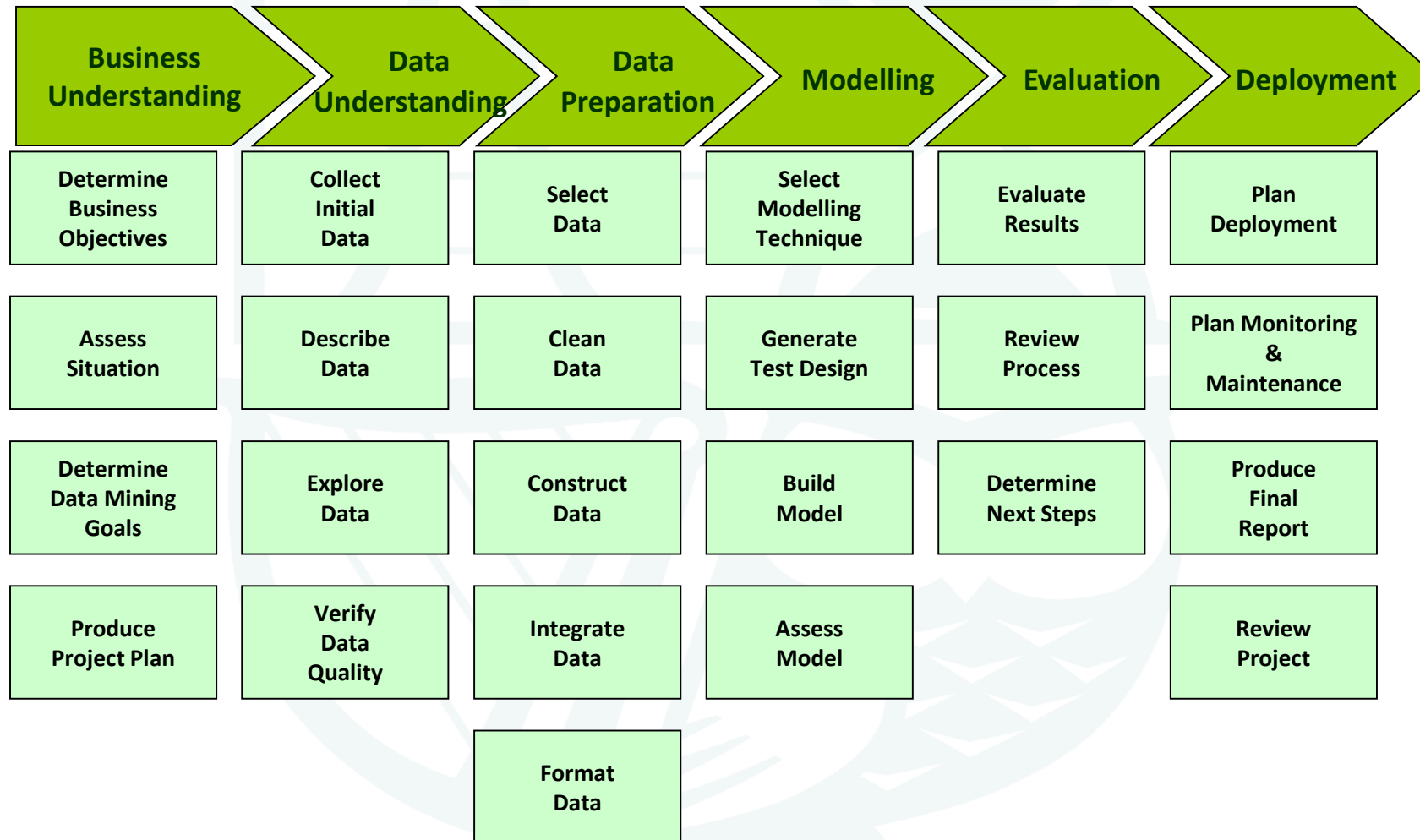


Figure 1: Four level breakdown of the CRISP-DM methodology

Framework – CRISP-DM phases & tasks

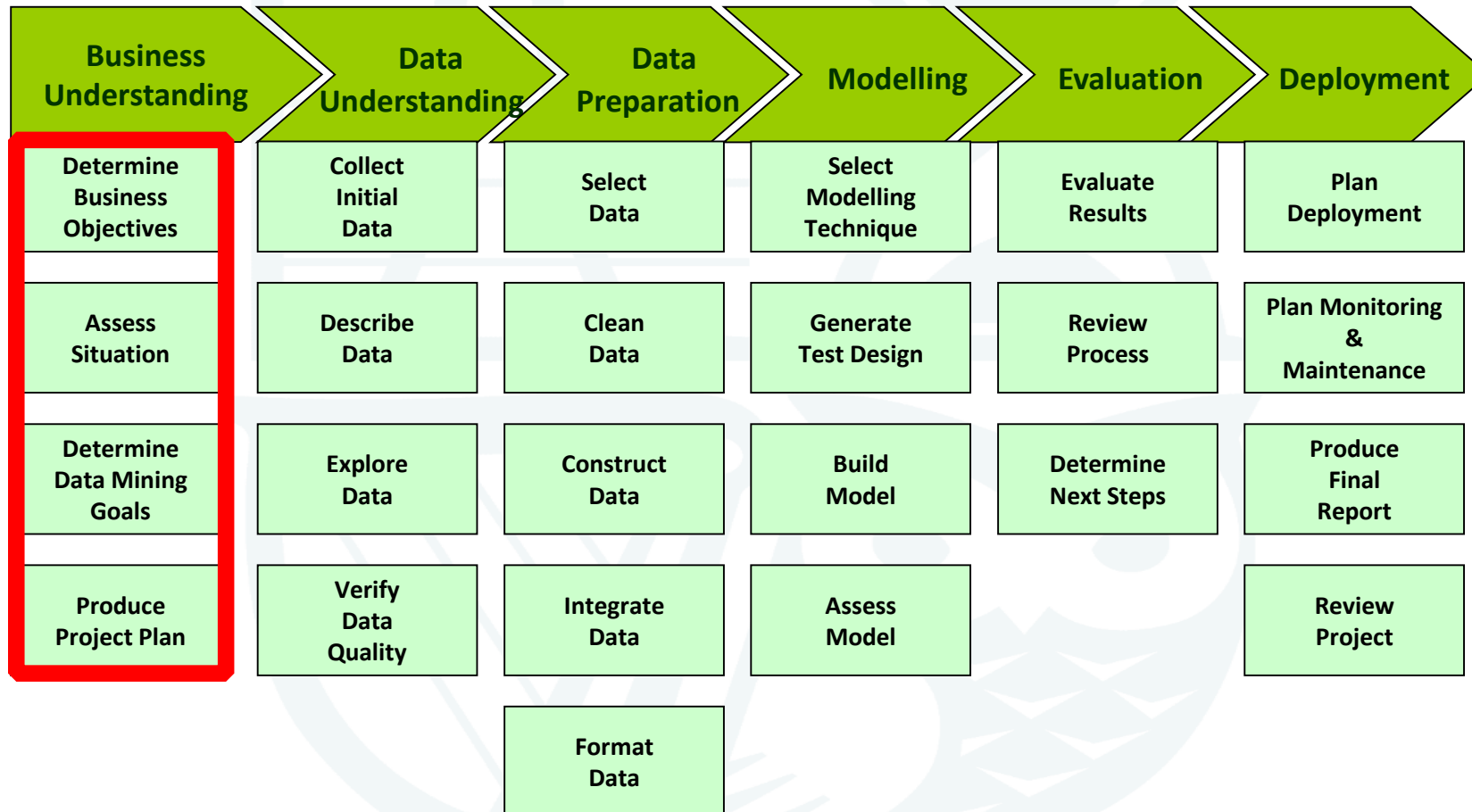


Framework – CRISP-DM tasks & outputs

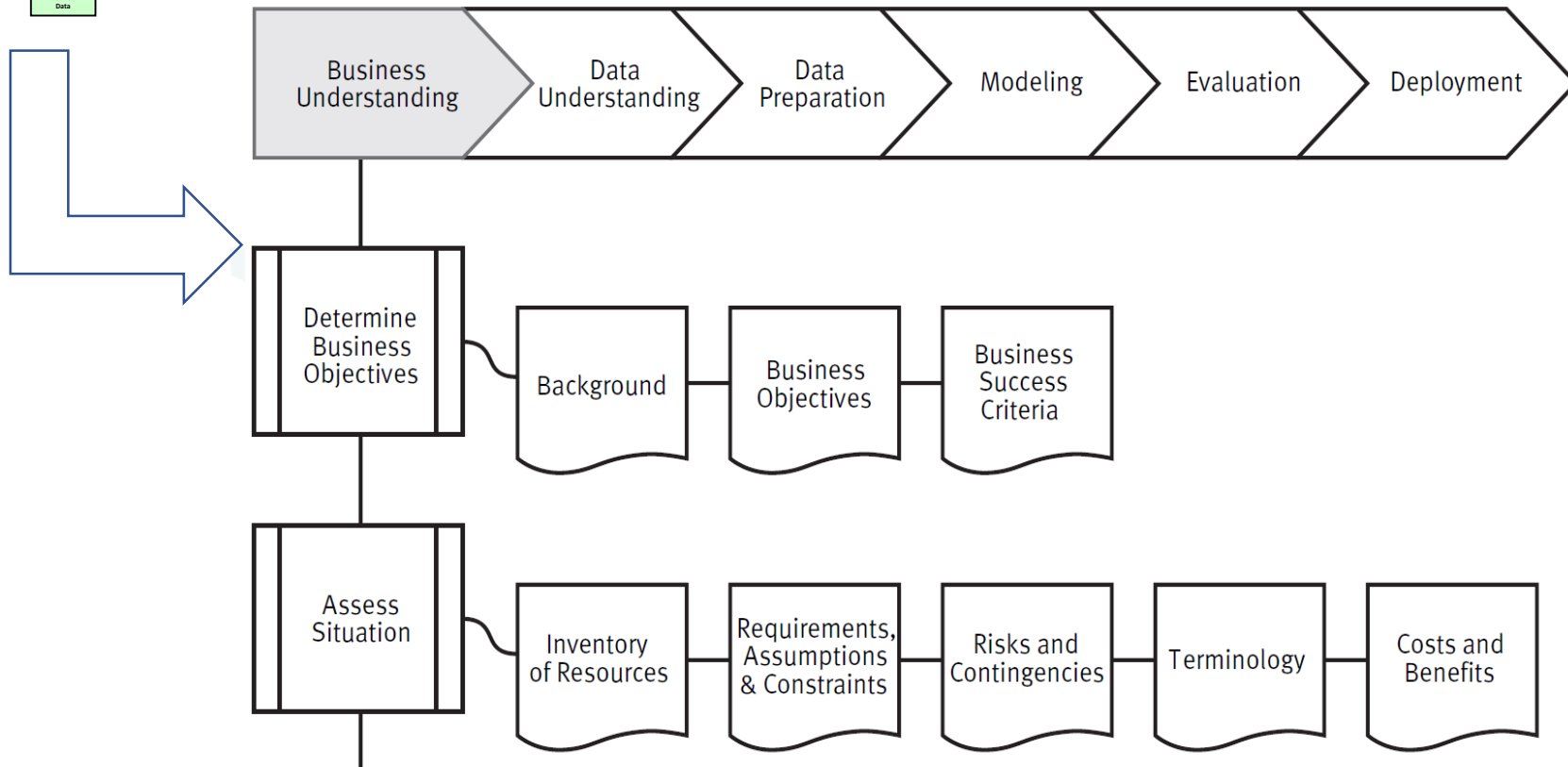
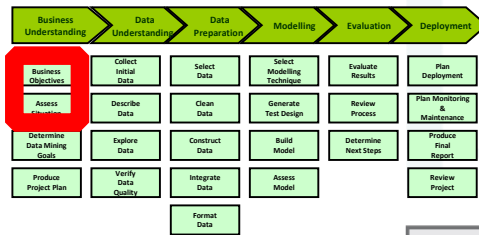
Business Understanding	Data Understanding	Data Preparation	Modeling	Evaluation	Deployment
<p>Determine Business Objectives <i>Background Business Objectives Business Success Criteria</i></p> <p>Assess Situation <i>Inventory of Resources Requirements, Assumptions, and Constraints Risks and Contingencies Terminology Costs and Benefits</i></p> <p>Determine Data Mining Goals <i>Data Mining Goals Data Mining Success Criteria</i></p> <p>Produce Project Plan <i>Project Plan Initial Assessment of Tools and Techniques</i></p>	<p>Collect Initial Data <i>Initial Data Collection Report</i></p> <p>Describe Data <i>Data Description Report</i></p> <p>Explore Data <i>Data Exploration Report</i></p> <p>Verify Data Quality <i>Data Quality Report</i></p>	<p>Select Data <i>Rationale for Inclusion/ Exclusion</i></p> <p>Clean Data <i>Data Cleaning Report</i></p> <p>Construct Data <i>Derived Attributes Generated Records</i></p> <p>Integrate Data <i>Merged Data</i></p> <p>Format Data <i>Reformatted Data</i></p> <p><i>Dataset Dataset Description</i></p>	<p>Select Modeling Techniques <i>Modeling Technique Modeling Assumptions</i></p> <p>Generate Test Design <i>Test Design</i></p> <p>Build Model <i>Parameter Settings Models Model Descriptions</i></p> <p>Assess Model <i>Model Assessment Revised Parameter Settings</i></p>	<p>Evaluate Results <i>Assessment of Data Mining Results w.r.t. Business Success Criteria Approved Models</i></p> <p>Review Process <i>Review of Process</i></p> <p>Determine Next Steps <i>List of Possible Actions Decision</i></p>	<p>Plan Deployment <i>Deployment Plan</i></p> <p>Plan Monitoring and Maintenance <i>Monitoring and Maintenance Plan</i></p> <p>Produce Final Report <i>Final Report Final Presentation</i></p> <p>Review Project <i>Experience Documentation</i></p>

Figure 3: Generic tasks (bold) and outputs (italic) of the CRISP-DM reference model

CRISP-DM – BU tasks



CRISP-DM – BU tasks & outputs



CRISP-DM – BU task details

1.1 *Determine business objectives*

Task

Determine business objectives

The first objective of the analyst is to thoroughly understand, from a business perspective, what the customer really wants to accomplish. Often the customer has many competing objectives and constraints that must be properly balanced. The analyst's goal is to uncover important factors at the beginning of the project that can influence the final outcome. A likely consequence of neglecting this step would be to expend a great deal of effort producing the correct answers to the wrong questions.

Output

Background

Collate the information that is known about the organization's business situation at the start of the project. These details not only serve to more closely identify the business goals to be achieved but also serve to identify resources, both human and material, that may be used or needed during the course of the project.

Activities

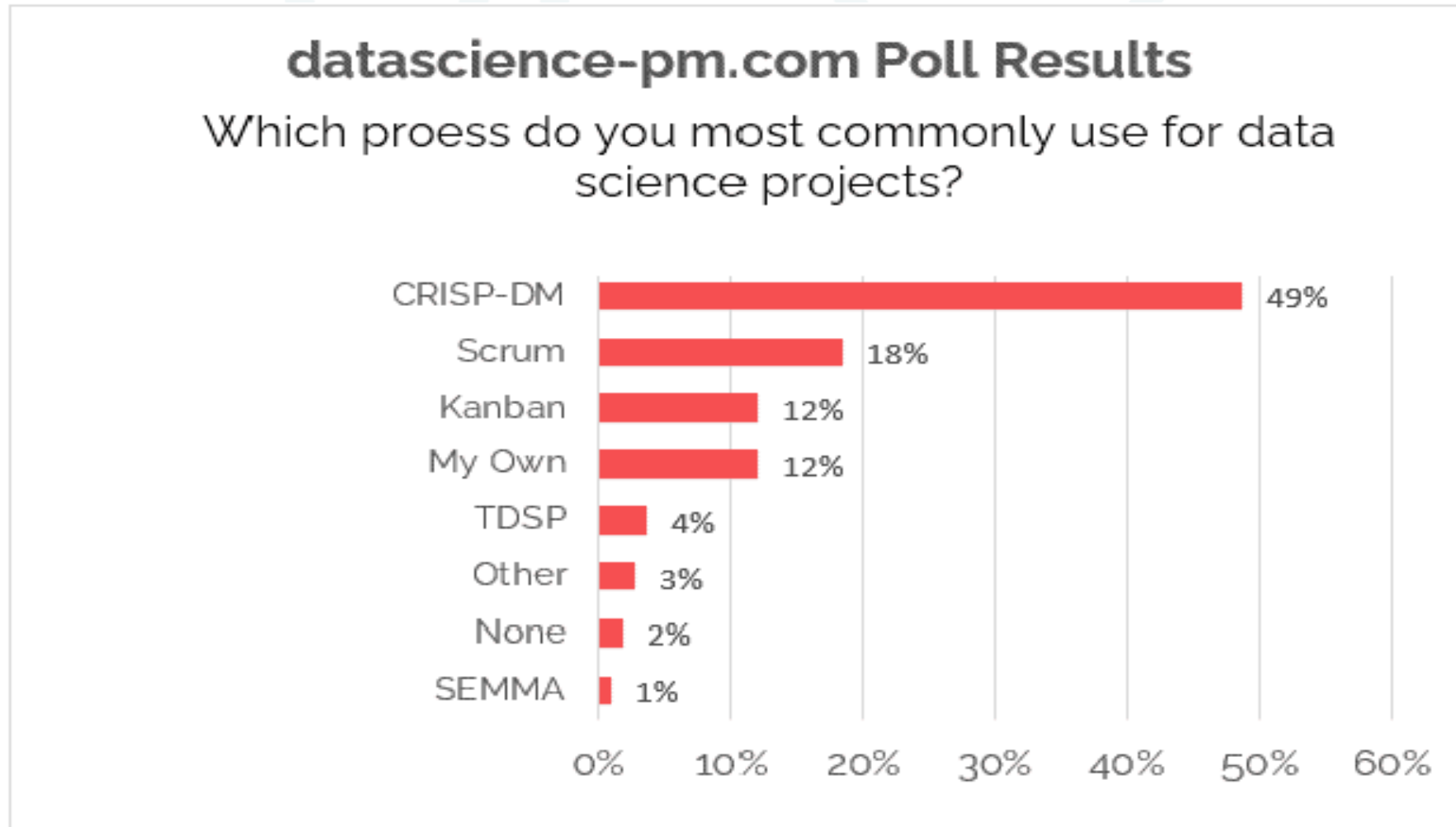
Organization

- Develop organizational charts identifying divisions, departments, and project groups. The chart should also identify managers' names and responsibilities
- Identify key persons in the business and their roles
- Identify an internal sponsor (financial sponsor and primary user/domain expert)
- Indicate if there is a steering committee and list members
- Identify the business units which are affected by the data mining project (e.g., Marketing, Sales, Finance)

Problem area

- Identify the problem area (e.g., marketing, customer care, business development, etc.)
- Describe the problem in general terms
- Check the current status of the project (e.g., Check if it is already clear within the business unit that a data mining project is to be performed, or whether data mining needs to be promoted as a key technology in the business)

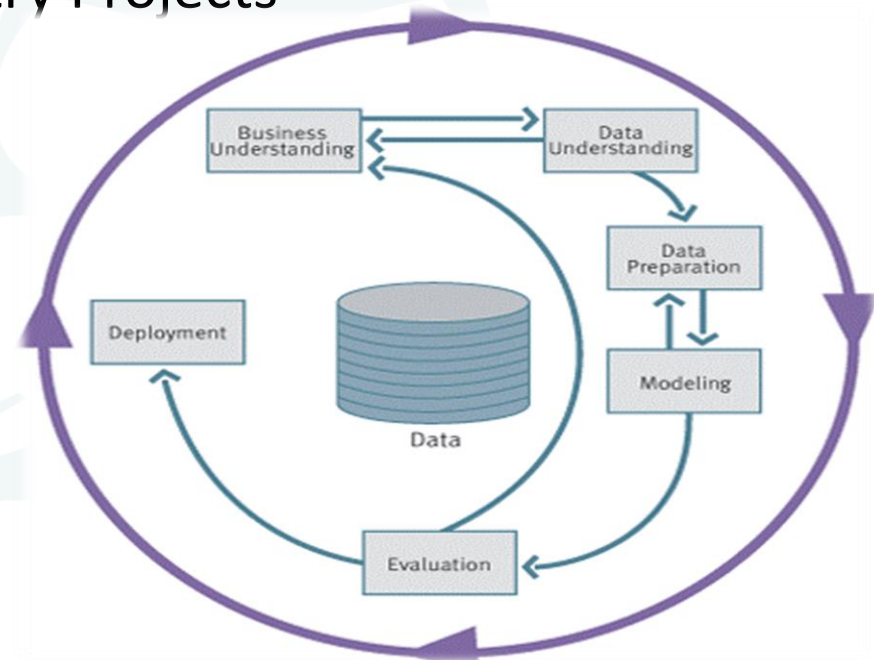
Frameworks – most common models



Sources <https://www.datascience-pm.com/crisp-dm-still-most-popular/> and <https://www.kdnuggets.com/>

CRISP-DM - On our projects

- Enterprise Ireland Innovation Vouchers (EI IV) & privately funded projects
- Enterprise Ireland Innovation Partnership Project (EI IPP)
- M.Sc. In Data Science(DS) Industry Projects
- Research projects



EI Innovation Vouchers (6-8 weeks)

- BU – easy to get from domain experts as projects smaller in scale
- DU – subject to readily available data (often MS Excel, csv), hard to assess due to lack of company knowledge
- DP – we prepare what is provided, we often need/ask for more
- Modeling – straightforward/often more descriptive or visual in nature
- Evaluation – dependent on models required, descriptive statistics, regression
- Deployment – variable, assessment of where company currently is/recommendations

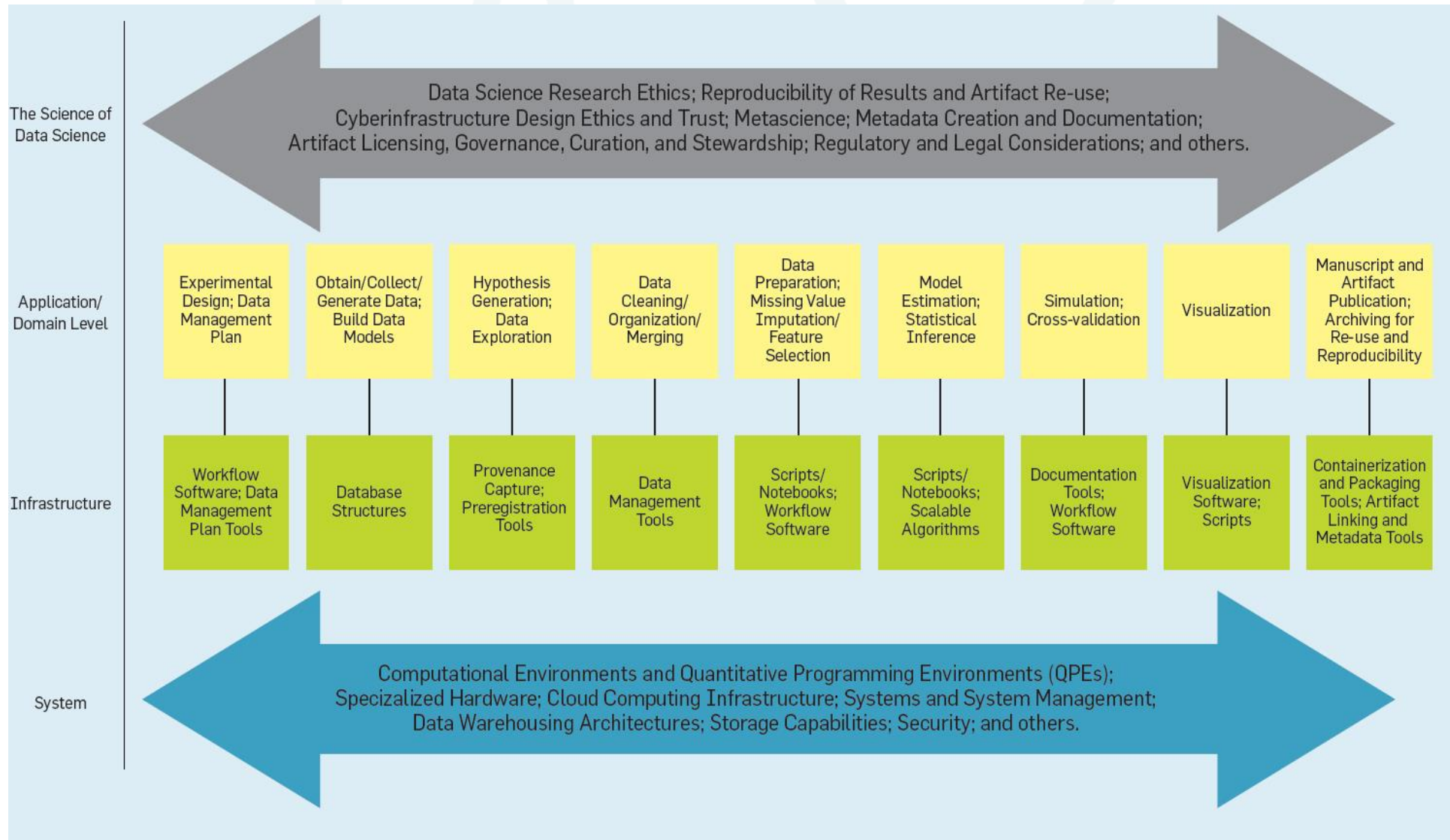
M.Sc. In DS industry projects (6-8 months)

- BU –from domain experts, projects medium size, longer term
- DU – much more time for EDA, additional data gathering
- DP – much fuller undertaking based on company requirements
- Modeling – more advanced predictive modelling, usually ML, nice to solve business problem
- Evaluation – more time to assess the models, iteration, testing, validation stronger
- Deployment – variable, but may serve as a basis for further work/POC

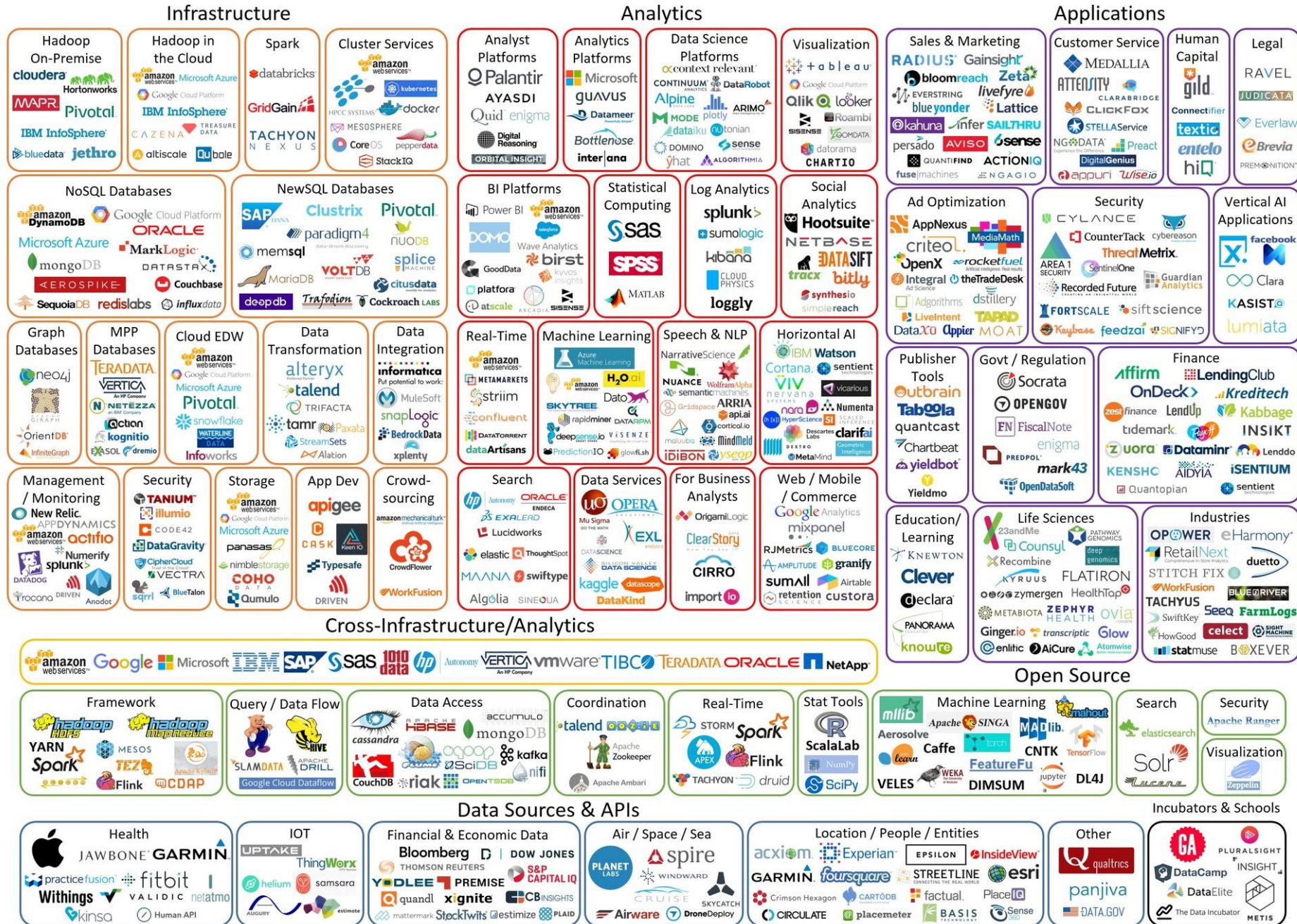
EI IPP (18 months)

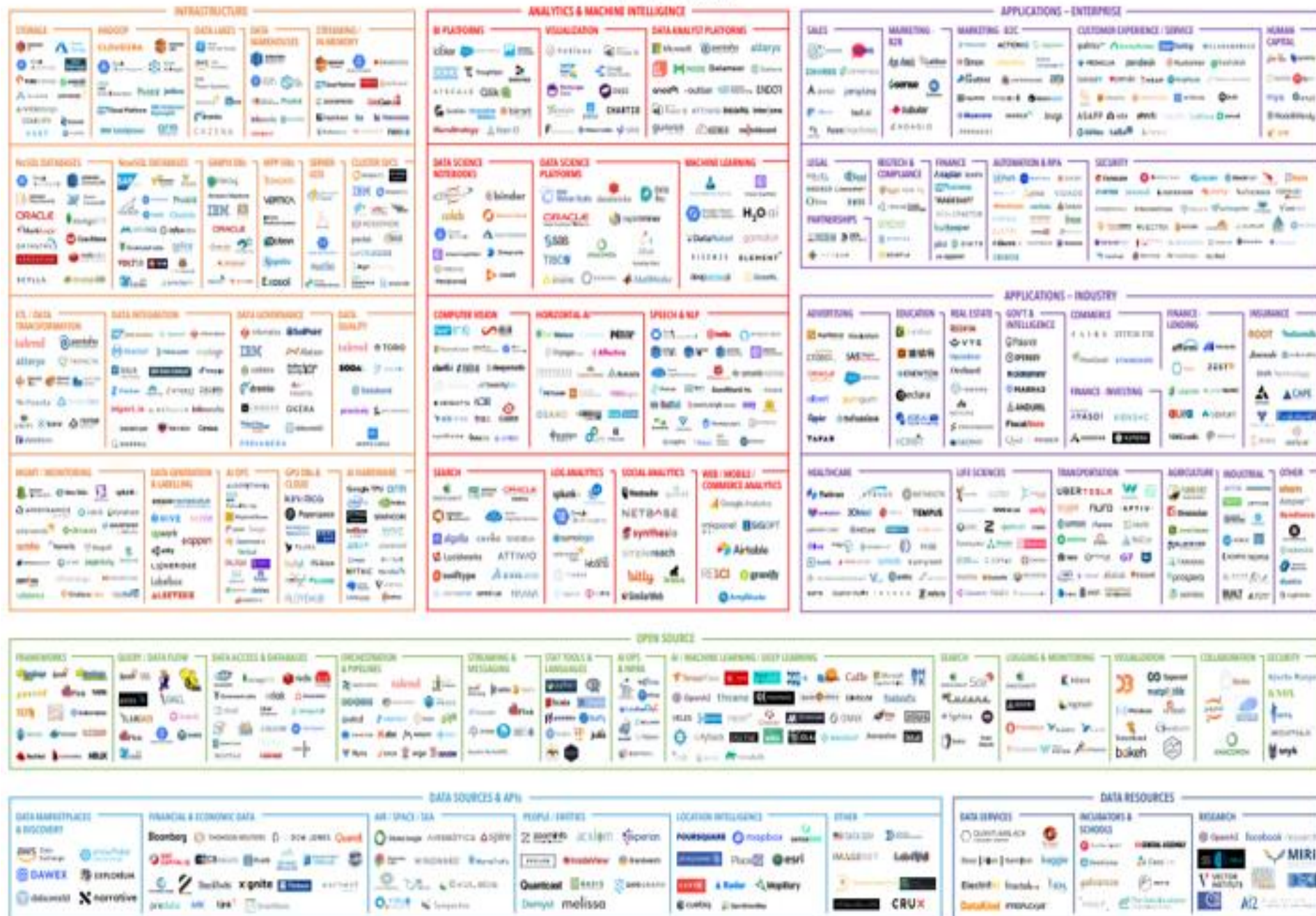
- BU – from domain experts, projects much large in scale
- DU – subject to & limited by the equipment/need
- DP – we gather what we need and iterate early & often
- Modeling – much more advanced and exploratory, fail fast and iterate
- Evaluation – dependent on client needs and accuracy and precision required
- Deployment – variable but may serve as a basis for further work/POC

Technologies on projects



Big Data Landscape 2016 (Version 3.0)





Technologies on our/other projects

EI IV, EI IPP, M.Sc. In DS

- Infrastructure – Microsoft Azure, Amazon Web Services AWS, Hadoop cluster, local machine
- Data manipulation – SQL, NoSQL, NewSQL
- Databases/datasets – MySQL, MS Excel, SQL Server, flat files, csv
- Programming – Python, R, Julia

Technologies on our/other projects

EI IV, EI IPP, M.Sc. In DS

- Visualisation –Matplotlib, Dash & Plotly, ggplot, R Shiny apps
- Data science/machine learning platforms– RapidMiner & Weka, KNIME, Azure ML, MATLAB, SPSS
<https://www.gartner.com/reviews/home>
- Other – Jupyter notebooks <https://jupyter.org/>, GitHub

Business case for analytics

- Optimise people
- Optimise processes
- Optimise material management
- Fraud reduction
- Data based decision making
- Etc.

Business case for analytics

- Short term easy wins
- Address business drivers for the company/leaders
- Analytics to help decision makers (perf. v obj.)
- Connect data & analytics governance to business outcomes/objectives
- **Data quality is key**

Embracing Data Science well?

According to IBM the 10 industries with the greatest demand for data science:

- 1. Finance and Insurance: 19%**
- 2. Professional, Scientific, and Technical Services (incl. medicine): 18%**
- 3. Information Technology: 17%**
4. Management of Companies and Enterprises: 13%
5. Manufacturing: 12%
6. Utilities: 10%
7. Wholesale Trade: 9%
8. Mining, Quarrying, and Oil and Gas Extraction: 9%
9. Public Administration: 7%
10. Agriculture, Forestry, Fishing, and Hunting: 6%

Conclusion & recommendations

- CRISP-DM - the de facto industry leader, traditional or agile
- Makes data analytics process more reliable & repeatable
- Learn more about CRISP (TDSP or another process model/framework)
- Learn Python and/or R
- Learn how to use Jupyter notebooks
- Undertake/review some of Andrew Ng ML courses

Conclusion

The full CRISP-DM model documentation is available here:

<https://www.the-modeling-agency.com/crisp-dm.pdf>

ASUM DM

http://gforge.icesi.edu.co/ASUM-DM_External/index.htm#cognos.external.asum-DM_Teaser/deliveryprocesses/ASUM-DM_8A5C87D5.html

and

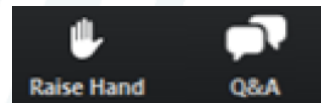
https://www.Researchgate.Net/publication/321944704_combining_process_guidance_and_industrial_feedback_for_successfully_deploying_big_data_projects

Questions

Please click on the 'Raise Hand' icon
to ask a question
and
wait to be unmuted

or

Use the Q&A function



Any questions?





INSTITUTE of
TECHNOLOGY
CARLOW

Institiúid Teicneolaíochta Cheatharlach

Insurtech Network Centre (INC) Carlow

Aidan Mahon

Supported by



Relevance of the Actuary – The hunt for the purple unicorn!

- Data science and actuarial science: necessary bedfellows for the future
- Actuaries can focus data science efforts on the problems worth tackling
- What can the data science and data engineering community learn from data intensive adjacent industries
- With more sophisticated technology comes the need for a more advanced skillset to exploit the tech to its full capability.
- The highly sought-after hybrid data scientist-actuary is so rare they are referred to in insurance recruitment as ‘purple unicorns’
- The best solution can often be to put the data science tools into the hands of actuaries.

Future of the Actuary

- Technology entrepreneurs and software engineers are driving the “tech” part of InsurTech, actuaries can be the engineers driving the “insur” part—because actuaries understand how the many pieces of the business fundamentally work together.
- Actuaries are perfectly positioned to serve as advisors to insurtech and innovation efforts
- Actuaries with learning mindsets, experimental attitudes and personal drives to be problem-solvers are well-positioned to thrive in the InsurTech space. (Deloitte)

Insurtech Network Centre (INC)



Think of the INC as an Exchange or Market Place.

Do you have ideas for a project you want to do in one of these areas, but you can't get the resources or budget? We can help!

- **Start** - Insurtech Accelerator.
- **Expand** - Innovation by accessing the Research & Development capabilities of IT Carlow. (via EI supports)
- **Engage** - Educational resources of IT Carlow
- **CIRDAS** (Centre for Insurance Risk and Data Analytics Studies)
HCI funded opportunity – heavily / fully subsidised targeted education programmes for industry

Sample Challenges / Solutions

- Low /No Digitisation
- Low Retention
- Low Value Repetitive Administrative Tasks
- Commoditisation
- Disintermediation
- Poor Customer Journey / Touchpoints
- Poor Targeting & Low share of wallet
- Inconsistent service
- Regulatory and Compliance
- Lack Resources

A snapshot of the Irish Insurtech landscape

Ireland has a vibrant Insurtech environment (many Fintech, Regtech, Medtech and Agtech solutions can have an insurance application too!). Some examples of the diverse offering include:

- **Reitigh** – Automating some of the most complex administration processes
- **Traxsit** - Remote Asset Tracking Platform
- **Aerial Agritech** - Satellite Multi Spectral & High Res 3D GIS Drone Imaging
- **CameraMatics** – Advanced Telematics / Tracking Solution
- **Cybrisc** - Smart Real Time Cyber Risk Monitor Assessment / Validation
- **GECKO Governance** - Revolutionary Regtech system

INC / IT Carlow education offering

The INC strategy on education has been industry led.

Two streams:

1. Equipping Data Scientists to work in the area of insurance.
2. Upskilling current insurance company employees in the discipline of Data Science / Data Analytics.

The education and upskilling offering will be continuously reviewed and will combine topics such as:

- Data Management for the Insurance Industry
- Behavioural Analytics
- Claims Analytics – Fraud Detection
- Design thinking
- EDI in the workplace and in insurance practice

Thank you for your time



Greg.Doyle@itcarlow.ie

Aidan.Mahon@itcarlow.ie

The Insurtech Network Centre is now online at:
www.insurtechcarlow.ie

On Twitter at: @InsurtechInc and personally on LinkedIn

To join INC mailing list please email
Aidan.Mahon@itcarlow.ie

My mobile number is 087 7618574

Any questions?