CORK INSTITUTE OF TECHNOLOGY
INSTITIÚID TEICNEOLAÍOCHTA CHORCAÍ

PROGRAMMATIC REVIEW OF THE SCHOOL OF SCIENCE 2018

Phase 2: Programme Review

PROGRAMME PANEL REPORT

SCHOOL: Science
DEPARTMENT: Mathematics
DATE: 14th May 2018

PROGRAMMES SUBMITTED FOR REVIEW

Major Awards

Higher Diploma in Data Science and Analytics

Non-Major Awards

None

PROGRAMME REVIEW PANEL MEMBERSHIP

Dr Kieran Murphy (Chair), Lecturer, Department of Computing, WIT
Dr Katarina Domijan, Lecturer, Department of Mathematics & Statistics, NUIM
Dr Eoin Murphy, Senior Data Scientist – Research & Innovation, Allied Irish Banks, Dublin
Ms Eilish Bouse, Data Scientist, Arity, Belfast
Dr Noel Murray, Head of Department of Tourism & Hospitality, CIT

PROGRAMME REPRESENTATION

Programme Staff

Dr David Goulding, Head, Department of Mathematics
Dr Noreen Quinn, Lecturer, Department of Mathematics, Full-Time HDip and MSc Coordinator
Ms Sarah Murphy, Lecturer, Department of Mathematics, Part-Time HDip Coordinator
Dr Catherine Palmer, Lecturer, Department of Mathematics
Mr Aengus Daly, Lecturer, Department of Mathematics, Project coordinator

Dr Justin McGuinness, Lecturer, Department of Mathematics
Mr Adrian O'Connor, Lecturer, Department of Mathematics
Dr Robert Heffernan, Lecturer, Department of Mathematics
Dr Ted Scully, Lecturer, Department of Computer Science
Dr Ignacio Castiñeiras, Lecturer, Department of Computer Science
Mr Byron Treacy, Lecturer, Department of Computer Science
Dr Diarmuid Grimes, Lecturer, Department of Computer Science

Learner Representatives

Igea Accardo, Full-time student HDip in Data Science and Analytics
Stephen Murphy, Full-time student HDip in Data Science and Analytics
Ciaran O'Brien, Full-time student HDip in Data Science and Analytics
Nevan O'Keefe, Part-time student HDip in Data Science and Analytics

Graduates

Rose Bain, Nimbus, HDip in Data Science and Analytics Graduate
Niamh Minihane, McAfee, HDip in Data Science and Analytics Graduate
Christopher O'Donovan, Xanadu Consultancy, HDip in Data Science and Analytics Graduate
Kieran O'Driscoll, Heineken Ireland, HDip in Data Science and Analytics Graduate

External Stakeholders

Dr Sorcha Healy, Data Scientist, McAfee
(Member of MSc in Data Science and Analytics Validation panel)
Mr Mark Hodnett, Senior Data Scientist at AltViz
Dr Michael Phelan, Head of Supply Chain Data Science Group, DePuy Synthes Companies
(Lecturer on Part-time programme)
Dr David Hawe, Data Analysis Engineer, BioMarin Pharmaceutical Inc.
(Lecturer on Part-time programme)

A. PROGRAMME SUMMARY AND MAJOR CHANGES PROPOSED

1. HIGHER DIPLOMA IN DATA SCIENCE AND ANALYTICS

1.1. **Programme Summary**

The *HDip in Data Science & Analytics* runs over two (four) semesters for full-time (part-time) students. For the full-time students each semester accrues 30 credits, while in the part-time programme students will typically accrue 15 credits per semester. All 60 ECTS are at advanced level, in accordance with CIT's academic policy for taught Higher Diploma programmes.

The Semester 1 schedule (30 ECTS) consists of six 5-credit modules which provide the learner with the necessary foundation in Mathematics, Statistics and Computer Science. The Semester 2 schedule consists of four 5-credit modules and a capstone 10-credit project module.

In the project module, the student applies the knowledge, skills and competences acquired in the taught modules to the research and development of a Data Science problem, and successfully complete the project in accordance with a project plan. Many of the projects undertaken are linked to industry and the Department actively seek collaboration with industrial partners for projects. Giving the student the opportunity to work on an authentic data science problem will maximise employment potential on graduation.

1.2 Major Changes Now Proposed

Based on the feedback from graduates, current students and industry, the key change to the programme is that the Department proposes introducing a dedicated module in R Statistical programming in Semester 1 (*Intro to R for Data Science*). The introduction of this module will allow for the consolidation of the instruction of R within the programme. This will however necessitate the removal of one of the current modules within the programme and the Department propose removing *DATA8003 – Unstructured Data and Visualisation* and amalgamating some of the learning outcomes of that module with the new R module and a revamped *Data Visualisation & Analytics* module.

This will further allow for the removal of potentially overlapping content between *DATA8003* and *DATA8007* as prescribed in the current version of the programme. Another key change is the proposed name change for *COMP8042 – Analytical and Scientific Programming* to *Scientific Prog in Python*, with the proposed name change better reflecting the content delivery of the module. In an effort to respond to recommendations from the *MSc in Data Science and Analytics* validation panel, it is proposed that an introduction to command line Linux is included in the updated module while based on feedback from industry source code control is now also included in the module.

Current Programme

	Stage 1	Semester 1
	Mandatory	
Order	**Module Code**	**Module Title**
1	DATA8001	Data Science and Analytics
2	STAT8006	Applied Stats & Probability
3	MATH8009	Maths Methods and Modelling
4	COMP8042	Analytical and Scientific Prog
5	DATA8002	Data Management Systems
6	DATA8003	Unstructured Data & Visualis'n
	Stage 1	Semester 2
	Mandatory	
Order	**Module Code**	**Module Title**
7	STAT8007	Statistical Meth for Big Data
10	DATA8005	Distributed Data Management
11	DATA8006	Data Science Analytics Project
	Group Elective 1	
Order	**Module Code**	**Module Title**
8	DATA8007	Data Visualisation & Analytics
8	DATA8004	Data Mining & Knowledge Discovery
	Elective	
Order	**Module Code**	**Module Title**
9	COMP8043	Machine Learning
9	STAT8008	Time Series & M-V Analysis

Proposed Programme

	Stage 1	Semester 1
	Mandatory	
Order	**Module Code**	**Module Title**
1	STAT8006	Applied Stats & Probability
2	No Code	Intro to R for Data Science
3	COMP8042	Scientific Prog in Python
4	MATH8009	Maths Methods and Modelling
5	DATA8002	Data Management Systems
6	DATA8001	Data Science and Analytics
	Stage 1	Semester 2
	Mandatory	
Order	**Module Code**	**Module Title**
7	STAT8007	Regression Analysis
10	DATA8005	Distributed Data Management
8	No Code	Data Visualisation & Analytics
11	DATA8006	Data Science Analytics Project
	Elective	
Order	**Module Code**	**Module Title**
9	COMP8043	Machine Learning
9	STAT8008	Time Series & PCA

Panel Findings and Recommendations

1. Overall Recommendation to Academic Council on Revalidation

Contingent upon confirmation of the fulfilment of any Panel conditions and the successful completion of the internal programme and module moderation process, the Panel **recommends to Academic Council that the listed programmes be revalidated** for five years or until the next Programmatic Review, whichever is sooner, with effect from September 2018.

As a condition of revalidation, the following Panel requirement(s) must be met:

1.1. **Requirement:** The Panel requires that GDPR is formally incorporated in the indicative content of the programme, and addressed accordingly.

2. General

2.1 Commendation: Both students and industry were largely unanimous in their positive feedback on the course. Students and graduates had positive feedback on the course, as well the professionalism of staff and the quality of teaching.

2.2 Commendation: The explicit renaming of modules (and their importance) which include R & Python is to be commended

2.3 Commendation: The quality of the programme documentation as well as the rationale provided for changes in the programme is also to be commended

2.4 Commendation: In the module Scientific Prog in Python, using Linux is deemed really beneficial to the programme. Also the introduction of Git and Version Control practices is a positive addition to the programme which will allow for greater opportunities for graduates.

2.5 Commendation: Staff are very approachable from a student's perspective, and are very knowledgeable and passionate about the programme.

2.6 Commendation: The applied approach with multiple real life projects is extremely beneficial to the development of the student's skill set, which will provide them with significant career opportunities.

2.7 Recommendation: Fundamental linear algebra concepts are needed across the programme, however linear algebra does not explicitly appear in the indicative content of any module. Given its importance linear algebra should be inserted in the indicative content within some modules

2.8 Recommendation: All book resources to be checked and updated in all modules.

2.9 Recommendation: Where module recommendations (in prerequisite learning) are specified in the module descriptor, please ensure that these are all updated to the new module codes and titles.

2.10 Recommendation: Students need to be exposed to more 'dirty data' sets in their assignments, as this will better reflect what they will encounter in industry.

2.11 Recommendation: As a medium term/longer term view – the programme board should consider reducing the data management (DATA8002 & DATA8005) modules into one module. The motivation for the reduction of these two modules is based on both student and industry feedback. In a similar vein, the programme board should consider splitting the content of COMP8043 – Machine Learning into two modules, with possibly the second of these modules being an elective module. Again the motivation for this is student and industry feedback.

2.12 Recommendation: Use the seminar series to show real life data sets, to allow students to see real-life issues. Include cloud tools as part of the seminar series.

2.13 Recommendation: Have an accelerator course for programming (as is available already in maths in learning support). The panel would recommend the provision of dedicated out of hours learning support for part-time students during the first few weeks of the programme, where possible.

3. Entrant and Graduate Profile, Award and Professional Environment

3.1 **Commendation:** Based on the feedback from industry, this programme appears to be well designed and meets industry needs.

4. Programme Operation and Performance

4.1 **Commendation**: Both the report and student feedback suggests that there are a lot of opportunities for graduates. Job prospects are extremely positive for both current students and past graduates and reflects very well on the programme.

4.2 Recommendation: Review the range of marks in all modules, as student feedback suggests that some may be struggling in some cognate areas, although this does not seem to be evident in the grades presented in the programme documentation.

4.3 Recommendation: Encourage Part Time students to develop peer-to-peer support networks.

5. Proposed Programme Specification (incl. Delivery and Assessment)

5.1 **Commendation:** The re-sequencing of modules, and the reshaping of module content has resulted in a more coherent programme. The re-sequencing of the statistics content to have a specific module on regression is to be commended. The introduction of the Intro to R module, is a very positive change to the programme.

5.2 Recommendation: Insure that the sequence of topics covered in Intro to R for Data Science is aligned to the topics covered in Scientific Prog in Python (COMP8042).

6. MODULES

This section presents the findings and recommendations from an indicative review of modules carried out by the members of the Peer Review Panel. The Panel notes that a comprehensive survey of module specifications could not be carried out in the context of this review.

Therefore, a recommendation of the Panel to revalidate the programme(s) under review is contingent on the successful completion of the subsequent internal programme and module moderation process carried out by, or on behalf of, the CIT Registrar's Office.

6.1 **Scientific Prog in Python**

Recommendation: May need to introduce a practical on the installation of software as part of indicative content and assessment.
Recommendation: Mention Pandas & NumPy as part of indicative content.

6.2 Data Science and Analytics

Recommendation: Keep some excel in the module content.
Recommendation: Make the 'Tutorial' a 'Lab' in the workload of the FT programme.
Recommendation: Spelling mistake on 'Theory' in PT workload.

6.3 Data Management Systems:

No change

6.4 Intro to R for Data Science –

Recommendation: LO1 verb needs to be updated.
Recommendation: A 5th LO to be added focused on Document Generation.
Recommendation: Introduce Git as part of this module content.
Recommendation: Remove the explicit mention of Bayesian statistics in indicative content.
Recommendation: Add explicit reference to control statements in R in indicative content.

6.5 Applied Stats & Probability

Recommendation: Years of books to be updated, and latest editions added. Need to focus on R as opposed to SPSS.

6.6 Maths Methods & Modelling

Recommendation: Rename the module – suggested title to be changed to more adequately reflect the content.
Recommendation: Remove 1st recommended book. Need to update all books for this module based on changes to the module.

6.7 Regression Analysis

Recommendation: Change Part Time workload to Independent Learning for 4 hours.

6.8 Data Science Analytics Project

Recommendation: Update module recommendations in prerequisite learning.

6.9 Distributed Data Management

No change (other than recommendation in section 2.11).

6.10 Data Visualisation & Analytics:

Recommendation: State clustering in the indicative content if it is being covered in this module.
Recommendation: Data communication needs to be added in as indicative content and as a learning outcome (may just rephrase LO 4 to reinforce the importance of data communication).

6.11 Machine Learning

Recommendation: Final project assessment timing needs to change to 'semester end' from 'week 12', and scale the indicative content down (as there are too many elements covered), to highlight specifically what is being covered in the module, e.g. remove decision trees, and show where this is being covered in other modules.
Recommendation: Re-examine Learning Outcomes 2, 3, & 4 to provide a short list of examples of what the student will know on completion of the module.

6.12 Time Series & PCA

Recommendation: Review allocation time for PT & FT students, when the module is being shared.
Recommendation: Update Learning Outcomes as follows:
Rewrite LO1 to read: Perform PCA to reduce dimensionality of datasets.
LO2: Delete and replace with: Describe the assumptions underlying PCA & time series models.
LO3: no change
Rewrite LO4 to read: Apply regression and time series models for prediction, and give an account of the paradigm under which the forecasts are being made, along with their reliability.
Rewrite LO5 to read: Apply statistical software to generate time series models and preform diagnostic analysis and forecasts for these models.

7. Other Findings and Recommendations

7.1 **Commendation:** There was a wide variety of graduates and industry partners engaged in the review process, and their feedback was generally very positive, which is to be highly commended.

7.2 **Commendation:** The programme documentation was exceptional, especially in reaching out to the industry in Ireland in relation to their needs.

8. Derogations Sought

The panel seeks a continuation of derogation from free-choice electives. This rationale is based on similar programmes running having only cognate electives and the programme would seek a derogation from free-choice and prescribe a cognate elective.

B. PROGRAMME FINALISATION

1. IMPLEMENTATION OF PANEL REQUIREMENTS

Requirement(s)	Department Response	Registrar's Office
1. Overall Recommendation to Academic Council on Revalidation		
1.1 Requirement: The Panel requires that GDPR is formally incorporated in the indicative content of the programme, and addressed accordingly.	We note the requirement of the panel in relation to the incorporation of GDPR within the programme content. Data protection guidelines including GDPR has previously been taught within DATA8001 Data Science and Analytics but not reflected in the module content. In order to address this shortcoming, the programme board has reviewed the content of **DATA8001 Data Science and Analytics** and propose the following changes: LO5 is replaced with the following: "LO5: Develop a deep understanding of data protection, data privacy and other ethical issues." A section explicitly dealing with ethics and data privacy has been added to the indicative content: "**Ethics, data privacy and security** Investigate ethics, data privacy, security, data protection legislation, including GDPR and related topics in data governance."	Complete
2. General		
2.7 Recommendation: Fundamental linear algebra concepts are needed across the programme, however linear algebra does not explicitly appear in the indicative content of any module. Given its importance linear algebra should be inserted in the indicative content within some modules	We note the recommendation of the panel in relation to the inclusion of Linear Algebra within the indicative content of some of the modules. We propose including a section on Linear Algebra in **MATH8009 – Maths Methods and Modelling** LO1 and LO2 from the previous descriptor have been combined and a new LO2 has been included: "LO2: Examine linear systems, matrix operations and Gaussian elimination."	Complete

	A section explicitly dealing with Linear Algebra has been added to the indicative content: "**Linear Algebra** Introduce linear systems, matrices and Gaussian elimination, in addition to covering topics such as matrix operations and inverse of a matrix and linear independence."	
2.8 Recommendation: All book resources to be checked and updated in all modules.	All book resources have been checked and updated and the programme board plan to maintain an up to date central repository of online books for students in the Data Science and Analytics programmes.	Complete
2.9 Recommendation: Where module recommendations (in prerequisite learning) are specified in the module descriptor, please ensure that these are all updated to the new module codes and titles.	Module recommendations have been updated to reflect the new module codes and titles.	Complete
2.10 Recommendation: Students need to be exposed to more 'dirty data' sets in their assignments, as this will better reflect what they will encounter in industry.	Suitable example 'dirty' data sets have been sourced and the programme board will work with our colleagues in industry to ensure that the students are exposed to realistic datasets throughout their studies.	Complete
2.12 Recommendation: Use the seminar series to show real life data sets, to allow students to see real-life issues. Include cloud tools as part of the seminar series.	Discussions between the programme board and industry representatives has begun in an effort to ensure that a full programme of industrial seminars and events are in place for the coming semester.	Complete
4. Programme Operation and Performance		
4.2 Recommendation: Review the range of marks in all modules, as student feedback suggests that some may be struggling in some cognate areas, although this does not seem to be evident in the grades presented in the programme documentation.	All finalised module marks from 2012-2017 were obtained from Student Records and examined by the programme board. Table 1 below shows the breakdown of module marks for the duration of the programme while Figure 1 presents a boxplot of the same results. It is worth noting that the student performance is high across all modules with extremely high pass rates in all modules (both cognate and non-cognate). The results shown below are directly in-line with the results presented by the programme team during the review process. Notes: (1) No Grade includes withdrawal, deferrals and non-presenting students. (2) DATA8001 is an elective module for Business Information Systems students and this accounts for the increased numbers for this module.	Complete

5. Proposed Programme Specification (incl. Delivery and Assessment)		
5.2 Recommendation: Insure that the sequence of topics covered in Intro to R for Data Science is aligned to the topics covered in Scientific Prog in Python (COMP8042).	The programme team will work closely together to ensure that the sequencing of topics in both modules are fully aligned.	Complete
6. Modules		
6.1 **Scientific Prog in Python** **Recommendation**: Mention Pandas & NumPy as part of indicative content.	This module has been updated to now include a section in the indicative content listing some of the Python packages used throughout the module "**Python Packages** NumPy, SciPy, Matplotlib, Seaborn, Pandas"	Complete
6.2 Data Science and Analytics **Recommendation:** Keep some excel in the module content. **Recommendation:** Make the 'Tutorial' a 'Lab' in the workload of the FT programme. **Recommendation:** Spelling mistake on 'Theory' in PT workload.	Excel is mentioned within the indicative content of the module and the programme board propose maintaining some Excel within both this module and the programme as a whole. Other recommended changes have been completed.	Complete
6.4 Intro to R for Data Science **Recommendation:** LO1 verb needs to be updated. **Recommendation:** A 5th LO to be added focused on Document Generation. **Recommendation:** Introduce Git as part of this module content. **Recommendation:** Remove the explicit mention of Bayesian statistics in indicative content. **Recommendation:** Add explicit reference to control statements in R in indicative content.	LO1 verb changed from 'Understand' to 'Evaluate'. Additional LO added 'Develop best practice in terms of reproducible documentation and version control.' – this LO is added to account for both this and the following recommendation. Changed 'Reproducible Documents' in indicative content section to 'Reproducible Documentation and Version Control' and added line 'Implement version control through the integration of Git in R.' Removed and updated to correctly reflect content as follows: 'Investigate how R can be used in statistical modelling techniques (e.g. naive Bayes classifers).' Added the line 'Examine the implementation of control structures (loops and functions) in R.' to the indicative content of the module	Complete

6.5 Applied Stats & Probability **Recommendation:** Years of books to be updated, and latest editions added. Need to focus on R as opposed to SPSS.	All books have been updated and the module is now solely focused on the use or R rather than SPSS.	Complete
6.6 Maths Methods & Modelling **Recommendation:** Rename the module – suggested title to be changed to more adequately reflect the content. **Recommendation:** Remove 1st recommended book. Need to update all books for this module based on changes to the module.	The programme board thank the panel for the suggestion but given the changes to the module in response to 2.7 Recommendation (i.e. the inclusion of Linear Algebra) – we now believe that the title is now more relevant following this change. First book has been removed and other books sources and resources updated to include material on Linear Algebra as well.	Complete
6.7 Regression Analysis **Recommendation:** Change Part Time workload to Independent Learning for 4 hours.	Part time workload changed.	Complete
6.8 Data Science Analytics Project **Recommendation:** Update module recommendations in prerequisite learning.	Module recommendations in prerequisites updated as per 2.9 Recommendation.	Complete
6.10 Data Visualisation & Analytics: **Recommendation:** State clustering in the indicative content if it is being covered in this module. **Recommendation:** Data communication needs to be added in as indicative content and as a learning outcome (may just rephrase LO 4 to reinforce the importance of data communication).	Indicative content expanded to now include clustering within the module description "Investigate clustering techniques e.g. partitioning methods, hierarchical clustering and advanced methods - fuzzy clustering, density based and model based clustering." LO3 expanded to include reference to data communication "Design and implement appropriate data visualisation techniques to solve data analytical problems; use a variety of data communication methods to present the corresponding results." Second assessment also changed to indicate key role of communication skillset "Evaluate and implement a visualisation technique to solve a problem; research and communicate a data analytics topic."	Complete
6.11 Machine Learning **Recommendation:** Final project assessment timing needs to change to 'semester end' from 'week 12'	Timing of final project assessment changed to Semester End.	Complete

6.12 Time Series & PCA **Recommendation:** Review allocation time for PT & FT students, when the module is being shared. **Recommendation:** Update Learning Outcomes as follows: Rewrite LO1 to read: Perform PCA to reduce dimensionality of datasets. LO2: Delete and replace with: Describe the assumptions underlying PCA & time series models. LO3: no change Rewrite LO4 to read: Apply regression and time series models for prediction, and give an account of the paradigm under which the forecasts are being made, along with their reliability. Rewrite LO5 which now reads: Apply statistical software to generate time series models and preform diagnostic analysis and forecasts for these models.	Allocation to be adjusted if module is shared between PT and FT students. LO1 changed. LO2 replaced as suggested. LO4 updated to reflect suggestion of the review panel. LO5 rewritten as "Perform diagnostic analysis and forecasts for both PCA and time series models, using statistical software."	Complete

3. Module and Programme Moderation

C.2.1 Completion of Programme and Module Moderation

Complete

C. APPENDIX – TIMETABLE OF PHASE 2 MEETINGS

Programmatic Review of the School of Science & Informatics
Department of Mathematics

14th May 2018 - Meeting Room, Tourism & Hospitality Building (Room T201)

Sunday May 13th 2018	
7.30 pm	**Panel Dinner (Optional)**
Monday May 14th 2018	
9.00 am	Private Panel Meeting including Presentation by Registrars Office
9.30 am	**Departmental Presentation**
10.00 am	**Meeting with Dept. Teams re Programme Operation and Performance**
11.00 am	Private Panel Meeting (Tea/Coffee)
11.30 am	**Meet with Dept. Teams re Proposed Changes to Programme Structures**
12.45 pm	Private Panel Lunch
1.45 pm	**Meet with Dept. Teams re General Review of Modules**
2.45 pm	**Meet with Students**
3.15 pm	**Meet with Graduates/ Employers**
3.45 pm	Private Panel Meeting (Tea/Coffee)
4.00pm	Private Panel Meeting to draft outline report
4.30pm	Feedback to School and Department Management