

Course Proposal Form

Please see Page 2 for instructions on which parts of this form to complete, whom to consult with to avoid unnecessary effort, and where to send the completed form.

Proposer(s): Goldwater, Yorston, Gal

Date: 9/10/2019

Cover page: Basic permanent course information

Unless otherwise noted, items in this section are entered into EUCLID and **cannot** be changed without creating an entirely new course.

Course Name	Informatics 2 - Foundations of Data Science
Course Acronym (used by the School only, e.g., for the Sortable Course List)	INF2-FDS
Course Level If the course is only available to MSc students, then it must be classed as Postgraduate. All other courses, regardless of level, are Undergraduate.	<u>x</u> Undergraduate Postgraduate
Normal Year Taken	UG1 <u>x_</u> UG2UG3UG4UG5MSc
Also available in years [This can be changed later if need be.]	UG1UG2UG3UG4UG5MSc
SCQF Credit Level Level 8 should normally be used for pre-honours courses. Level 10 should normally be used for optional UG3 courses (so UG4 students may also take them) and for courses aimed mainly at UG4 students. Level 11 should be used for courses aimed mainly at MSc students, whether or not UG4 students can also take them.	<u>7 x</u> 8 <u>9</u> <u>10</u> <u>11</u>
SCQF Credit Points	10 <u>x</u> _20406080 Other:
Delivery Location	<u>x</u> CampusOn-line Distance Learning
Course Type	<u>x</u> Standard (default) <u>Dissertation</u> <u>Online Distance Learning</u> <u>Other (specify: Placement, Student Led Individually</u> Created Course, Year Abroad)
Marking Scheme By default, courses use a numerical marking scheme. If you wish to use a grade-only marking scheme, your course proposal below should justify this.	x_Standard (numerical) Letter grade only

Commented [s1]: Previously 'Data Analysis and Inferential Thinking'. We should decide which of these (or some other title).

Guidance for remaining sections:

For an initial course proposal, please complete the cover page and Section 1 (Case for Support), which asks you to describe the need for this course and to provide an overview of the course design, including the learning outcomes. Please discuss your plans as early as possible with the head of Curriculum Review to avoid unnecessary effort.

Send the form with these sections completed to the BoS Academic Secretary and head of Curriculum Review (listed on the BoS page) to obtain their comments before filling out the remainder of the form.

If a full proposal is invited, please complete the remaining sections and send to iss-bos@inf.ed.ac.uk.

2. Student-facing course description and additional feedback and assessment information. This section provides most of the information students see in the DRPS entry for this course, as well as related details for BoS consideration.

3. Further information for BoS consideration: sample materials.

4. Additional Course Details required for DRPS. [Administrative information such as delivery timing and prerequisites.]

5. Placement in degree programme tables. [Required for all level 9-11 courses; used to determine where the course will be added to existing degree programme tables.]

6. Comments from colleagues. [All course proposal should be sent to relevant colleagues in the area as well as to the appropriate year organizer and BoS Academic Secretary for comment in good time before the BoS meeting. Use this section to indicate what feedback has been solicited and received.]

Colour coding and item-by-item guidance:

Guidance is provided in italics for each item. Please also refer to the guidance for new course proposals at http://www.inf.ed.ac.uk/student-services/committees/board-of-studies/course-proposal-guidelines. Examples of previous course proposal submissions are available on the past meetings page http://web.inf.ed.ac.uk/student-services/committees/board-of-studies/course-proposal-guidelines. Examples of previous course proposal submissions are available on the past meetings page http://web.inf.ed.ac.uk/infweb/admin/committees/bos/meetings-directory but note that the proposal form was updated in Jan 2019.

Sections in gold are for student view and are required before a course can be entered into DRPS. You must complete these sections even if your course has already been approved based on other documentation.

Sections in orange are for School use but are still required for all courses (even those that have already been approved based on other documentation).

Section in gray are for consideration by the Board of Studies. They are normally required for all new course proposals but may be omitted in some circumstances (e.g., for invited course proposals) if you obtain permission in advance.

1. Case for support

This section is for consideration by the Board of Studies. The final two boxes (Learning Outcomes, Graduate Attributes) will also go into the student-facing course description.

Overall contribution to teaching portfolio and relation to existing curriculum

Please explain (a) what motivates the course proposal (e.g. a previous course having become outdated/inappropriate, an emergent or maturing research area or new research activity in the School, offerings of our competitors) and (b) how it relates to existing courses and degree programmes (including any prerequisite courses). Every new course should make an important contribution to the delivery of our <u>Degree Programmes</u>.

This course is part of the new pre-hons curriculum and will run as a long thin course throughout UG2. Its goal is to introduce foundational technical concepts from data science (e.g., tabular data, exploratory data analysis, basic inferential statistics and machine learning) using an approach that is more integrated with practical and real-world issues (common tools, visualisation, ethical implications) than our current courses where some similar technical content is taught (e.g., Inf1-DA, PwA, Inf2B-Learning).

The coverage of machine learning techniques will be different, and somewhat less, than in Inf2B-Learning. This is not a problem because most students also take IAML, and nearly all topics in Inf2B are also covered in IAML. We aim to reduce this redundancy.

Inferential statistics are covered briefly in Inf1-DA but more time will be spent here, with a much broader variety of examples, using simulations to emphasize the importance of randomness and sampling. Inf1-DA also discusses relational databases and relational algebra. This course will cover some material from these topics, but with a less formal and more practical focus: in particular, we plan to use Python packages such as Pandas that support dataframes, and discuss dataframe operations such as selection and join, as well as how to effectively organize data (e.g., Tidy Data).

The course will have a co-requisite of either Discrete Mathematics and Probability (the updated version of DMMR), or Probability (10 credit Maths course). Both are taught in semester one. See below for how this is accommodated in the course structure.

Target audience and expected demand

Describe the type of student the course would appeal to in terms of background, level of ability, and interests, and the expected class size for the course based on anticipated demand. A good justification would include some evidence, e.g. by referring to projects in an area, class sizes in similar courses, employer demand for the skills taught in the course, etc

Required UG2 course, likely enrolment 300ish (depending on UG1 intake)

Anticipated Resource Requirements

Estimate how much lecturing, tutoring, exam preparation and marking effort will be needed in steady state, and any additional resources needed to set the course up initially. Provide estimates relative to class size where applicable and discuss how support staff will be recruited and supervised, if the class is likely to be very large. Please mention any scaling limits due to equipment or space. If equipment is required, say how it will be procured and maintained.]

Quotas, special arrangements or unusual characteristics

Please specify if this course requires any special arrangements such as quotas or other registration arrangements; is a collaboration with another school or institution, or has other atypical characteristics that may affect finances or student registration. Further justification/information may be requested for such courses.

none

Narrative description of the course aims and structure

Please describe the main goals of the course and how the course design will allow students to achieve those goals. This section should be consistent with the student-facing information provided below, but should provide additional information to help colleagues at BoS understand the vision and structure of the course. This description may refer to the learning outcomes and graduate attributes (next two boxes) and should explain how activities such as tutorials, labs, or in-lecture activities will support them, and how the proposed assessments will assess them.

For courses that are important pre-requisites for other courses, this section may also provide content/syllabus information which is too detailed for the student-facing description, such as a lecture-by-lecture syllabus.

The course will have a co-requisite of either Discrete Mathematics and Probability (the updated version of DMMR), or Probability (10 credit Maths course). Both are taught in Sem1. Therefore during Sem1 this course will focus on topics that do not require probability theory, and Sem2 will build on the concepts with additional methods that do rely on probability theory. The idea of introducing statistical concepts before introducing probability is not new: a similar syllabus is described in Gelman and Nolan's 'Teaching Statistics' book and is also used by Berkeley's Data 8 course, with an open-source textbook and materials that we plan to draw on.

The course will use Python-based tools (e.g., Pandas, Seaborn, Jupyter notebooks). Case studies will introduce some of the major themes of the course, including sampling bias and other potential ethical issues. Students will also be guided in critical thinking and writing skills, for example through reading and discussion of data science articles and peer feedback on short writing exercises, to help prepare them for a mini-project with report.

We plan to hold 1-2 hours of lecture per week (no more than 3 in every 2 weeks), plus workshopstyle weekly tutorials for discussion and/or paper exercises. A lot of the work for the course will be lab-based but we are not sure yet whether labs will be scheduled or just have drop-in help available.

1. Data wrangling and exploratory data analysis

- Working with tabular data (selecting rows and columns, joining tables, Tidy Data)
- Visualization of numerical and categorical data (histograms, scatterplots, line and bar charts, transforms; how to use color/style/markers, clutter vs clarity)
- Descriptive statistics (mean, median, mode, percentiles, standard deviation)
- Linear regression and correlation (descriptively) (MSE and optimization, difference between correlation and causation)

- High dimensional data (PCA for dimensionality reduction and visualization)
- k-means clustering (similarity measures, feature transforms)

2. Supervised machine learning

- Classification: k-nearest neighbors
- Regression: linear regression (now for prediction, also show how feature transforms apply here; warning about interpreting coefficients)
- Generalization and regularization (train/dev/test, motivated by choices of similarity measures, feature transforms, etc. Also discuss potential differences between scientific practice, competitions, and real-world use.)

----- Topics below here rely on probability theory ------

• Logistic regression (probabilistic modelling, optimization)

3. Statistical inference

- Randomness, simulation and sampling
- Confidence intervals, law of large numbers (error bars: difference between standard deviation and confidence interval)
- Randomized studies, hypothesis testing (focusing on the concept rather than on lots of specific tests)

4. Ethics, context, and meta-skills (Interspersed, discussed using case studies during lecture time and/or workshops)

- Where does data come from? (Sample bias, licensing and/or privacy issues)
- Visualization: misleading plots, color and accessibility
- Classification: algorithmic bias, discrimination
- Evaluation and use: differences between development and deployment; does using the system change the data distribution?
- Advantages and disadvantages of Notebooks; when to move to a real program
- Scientific communication; structure of a lab report
- Reading and critique of data science findings/articles; role of baselines and controls, claims and evidence (can we conclude X?)

Summary of Intended Learning Outcomes (MAXIMUM OF 5)

List the learning outcomes of the course. These must be assessable (i.e., observable), so must specify what the student should be able to do concretely, not simply what they should "understand". Use concrete verbs that indicate (a) what type of assessment would be appropriate, and (b) what level of knowledge/thinking is expected (from recall to analysis to

novel creation). Example verbs: define, explain, implement, compare, justify. Assessments (described later) should be tied to the learning outcomes.

Outcomes should typically focus more on the types of thinking/skills developed than on the detailed course content, and the level of thinking should be appropriate to the level of the course: outcomes for a Level 11 course should include more higher-level thinking skills than for a Level 8 course. Further guidance on writing learning outcomes can be found at https://www.ncl.ac.uk/ltds/assets/documents/res-writinglearningoutcomes.pdf

On completion of this course, the student will be able to

1. Describe and apply good practices for storing, manipulating, summarising, and visualising data.

2. Use standard packages and tools for data analysis and describing this analysis, such as Python and LaTeX.

3. Apply basic techniques from descriptive and inferential statistics and machine learning; interpret and describe the output from such analyses.

4. Critically evaluate data-driven methods and claims from case studies, in order to identify and discuss a) potential ethical issues and b) the extent to which stated conclusions are warranted given evidence provided.

5. Complete a data science project and write a report describing the question, methods, and results.

Graduate Attributes, Personal & Professional Skills

List the personal attributes and generic transferrable skills this course will help develop. Examples include

Cognitive skills: problem-solving, critical/analytical thinking, handling ambiguity

Responsibility, autonomy, effectiveness: independent learning, self-awareness and reflection, creativity, decisionmaking, leadership, organization and time management, flexibility and change management, ethical/social/professional awareness and responsibility, entrepreneurship

Communication: interpersonal/teamwork skills, verbal and/or written communication, cross-cultural or cross-disciplinary communication

2. Student-facing course description and additional feedback and assessment information

Except where noted, all fields are required and will go into the DRPS entry for the course (for use by students). Important: any text in DRPS is effectively a contract with students, so should not include details that are likely to change from year to year.

Summary Description Provide a brief official description of the course, around 100 words. This should be worded in a student-friendly way, it is the part of the descriptor a student is most likely to read. If this course replaces another course, please say so in this summary. Keywords	
Give a list of searchable keywords.	
Course Description A more detailed student-facing description of the course, which should normally include (a) a more in-depth academic description of the learning aims, nature and context of the course, (b) a rough outline of the content or syllabus, often as bullet points, and (c) a description of how the course will be taught, how students are expected to engage with their learning and how they will be expected to evidence and demonstrate their achievement of the intended learning outcomes.] Assessment Weightings:	
These should correspond approximately to the proportion of learning outcomes that each component assesses. More than 30% coursework requires specific justification. The expectation for a 10pt course is 20% coursework with the equivalent of one 15-20hr assessed assignment (but possibly split into smaller pieces). See 'components of assessment' below.	Written Exam _60_% Practical Exam% (<i>for courses with programming exams</i>) Coursework _40%
Further Assessment Information Provide any further information that should go on DRPS for students. E.g., if the assessment includes required group work or if students must pass some individual component of assessment as well as the course overall.	Course also includes Mini project (end of semester 1) that is peer graded and formative.
Components of assessment and time spent on assignments (for BoS only) If not already included in the course narrative description, please describe the type of assessments (oral presentation, report, programming, etc) and how each component of assessment will assess the intended learning outcomes. Where coursework involves group work, it is important to remember that every	Likely two assessed courseworks: (1) in-class quiz at the end of semester 1, worth 10% - this will focus on basic technical skills including Python, to make sure students are prepared for the formative and later summative projects. (2) final project in semester 2, worth 30%.

student has to be assessed individually for their contribution to any jointly produced piece of work. Also estimate how many hours students will spend on assignments. Please see the <u>School</u> policy on Workload and Assessment, which states that students should not be expected to spend more than 6-7 hrs/wk per 10 credits, including contact hours. Note that it often desirable to include formative assignments which are not formally assessed but submitted for feedback, often in combination with peer assessment.	Since there's no S1 exam, we also plan to have a mini- project due near the end of the S1 exam period which will be formatively assessed and probably done in groups of 2-3 students. This would allow students to present some exploratory data analysis and perhaps very simple machine learning in a report style similar to the final project. Students will also be expected sometimes to bring work to the workshop groups for peer review/discussion.
Feedback Information Provide a high-level description of how and what type of feedback will be provided to students, for inclusion in DRPS.	
Additional Feedback Information (for BoS use only) If not already included in the course narrative, provide further details on planned feedback arrangements. This includes how course feedback is solicited from the class and responded to, as well as what feedback students will get (either on work that contributes to their final mark, or not). The University is committed to a <u>baseline of</u> <u>principles</u> regarding feedback that we have to implement at every level, and the School encourages submission of at least one piece of written work for formative feedback. In general, formative feedback: • Should say how students can improve. • Need not be on individual work (e.g., consider a lecture or document summarizing common issues.) • Can include oral feedback during labs/tutorials • Can include feedback from peers • Clickers/TopHat/equivalents can provide in- class feedback for both students and lecturer. • Is returned in time for other forms of assessment to which it relates, to allow feedforward.	

Breakdown of Learning and Teaching Activities

State how many hours students spend on each part of the course. The total should be 10 x course credits, but please also see the <u>School</u> <u>policy on Workload and Assessment</u>,which states that students should not be expected to spend more than 6-7 hrs/wk per 10 credits, including contact hours.

Assume 10 weeks of lectures slots and 10 weeks of tutorials, but these need not all be used. As a guideline, a 10-pt course typically has 18-20 lecture hours, but should have only around 15 lectures of examinable material; the rest should be used for guest lectures, revision sessions, introductions to assignments, etc.

Reading List/Learning Resources You are encouraged to create resource lists using LEGANTO

Contact hours	
Hours	Туре
	Lecture Hours
	Seminar/Tutorial Hours
	Dissertation Project Supervision Hours
	Supervised practical/Workshop/Studio hours
	Feedback/Feedforward hours
	Summative assessment hours
	Revision Session Hours
Non-contact hours	
Hours	Туре

21 S	iype
	Directed Learning & Independent Learning
	hours

Total hours:

3. Further information for BoS consideration: sample materials

A full proposal for a new course must include examples of exercises and assessment. Please provide these below, along with publicity information if the course is to be advertised outwith the School.

Course information and publicity The course web page (typically the Learn landing page) will be linked from the Sortable Course List, and information such as timetables and assignment deadlines must be made available prior to the start of the academic year. Please specify here if any additional info/publicity is needed for your course: typically only if it is aimed largely at non-Sol students.	
Sample tutorial/lab sheet questions Provide a list of tutorial questions and answers and/or samples of lab sheets. These need not be fully fleshed out but should indicate what sort of exercises will be provided to help students learn the material.	
Sample assessment materials If the course is primarily assessed by exam, provide a sample exam question with model answers. Any non- standard exam format must be justified. The online list of past exam papers gives an idea of typical and alternative exam formats: http://www.inf.ed.ac.uk/teaching/exam_papers/. If the course is largely or primarily assessed by coursework, provide a sketch of a possible assignment with an estimate of effort against each sub-task and a description of marking criteria.	
Any other relevant materials Include anything else that is relevant, possibly in the form of links. If you do not want to specify a set of concrete readings for the official course descriptor, please list examples here.	

4. Additional Course Details for DRPS

Except where otherwise noted, these fields are required for entry into EUCLID and will be visible to students in the DRPS entry.

Planned Academic Year of Delivery (The first year you anticipate the course running, e.g. AY 2019-20)	
Course Organiser (By default, the course proposer)	
Intended Delivery Period	Semester1 Semester 2 Full Year Summer Other (please specify):
Timetable considerations/conflicts For School use. Please specify any constraints to be considered (e.g. overlap of popular combinations, other specialism courses, external courses etc). Include whether the semester delivery is constrained or could be flexible.	
Is this course available to visiting students?	Yes (default) No If no, please provide a justification here:
Required pre-requisite courses Use sparingly: these are enforced in PATH and can only be waived by approval from the School's Curriculum Approval Officer. Note that cross-year required pre- requisites may prevent MSc students from registering; consider using recommended pre-requisites or "other requirements" instead.	No Yes (please specify full course name(s) and code(s)):
Recommended pre-requisite courses	No Yes (please specify full course name(s) and code(s)):
Required co-requisite courses Specify any courses that must be taken in parallel with the existing course. Note that this leads to a timetabling constraint that should be mentioned elsewhere in the groups of	No Yes (please specify full course name(s) and code(s)):

Prohibited Combinations Specify any courses that may not be taken in combination with the proposed course].	No Yes (please specify full course name(s) and code(s)):
Other Requirements/Additional Information This information is often used by MSc students and students from other Schools to see if they have appropriate background without having done our School's courses. So please avoid course titles, instead list specific knowledge and skills (such as mathematical concepts, programming ability or specific languages, etc). Also list any other constraints on registration, for example: "Only available to 4th Year Informatics students including those on joint degrees." or "This course is open to all Informatics students including those on joint degrees, and to students in the School of Mathematics. Other external students whose DPT does not list this course should seek permission from the course organiser."	No Yes (please specify):
Visiting Student Pre-requisites	Same as "other requirements" Different than "other requirements" (please specify):

5. Placement in degree programme tables: for level 9-11 courses only

This section is for consideration by the Board of Studies and will be used later by ITO to determine where the course will be added to existing degree programme tables.

Is this course restricted to students on a specific degree? E.g., some courses are only available to students on a specific CDT or MSc.	No Yes (please specify and provide justification):
Is this course compulsory for students on any degree(s)?	No Yes (please specify and provide justification):
Any issues for part-time students? Normally, part-time students have access to the same courses as full-time students on the equivalent degree. If you anticipate any problems with this, please specify here.	

For optional courses:

If this course is available but non-compulsory for students on various degrees (most courses), please fill in this section. The choices here determine where the course appears in degree programme tables (DPTs) and the 2-3 character tags are displayed in the Informatics sortable course list.

Should this course be tagged as 'ML' (machine learning foundations and methods)? Courses with the ML tag are typically very high-demand and most degrees limit the number of ML credits. If your course might appeal to a similar audience but draw off students from these large courses, please select 'no' and choose one of the tags below.	No Yes
It you chose 'no', please choose at least one of the following tags Ideally, select exactly one, unless there is a good argument for more than one. These three are used in various combinations for many of our degrees.	 FSS (CS foundations, systems, and software) AIA (artificial intelligence applications and paradigms) COG (cognitive science: including HCI and NLP courses, but not most other AI courses. Please restrict to courses most relevant to natural cognition.)
and also tick if any of the following tags or categories apply. <i>Do not tick any of these if you selected</i> <i>'ML' already.</i>	 NS (natural systems: e.g., computation by or about biological or social systems. Many COG courses are also NS. This tag is mainly relevant for MSc in Informatics.) SE (software engineering: including courses that are highly relevant to SE degrees. All SE courses should also be FSS. This tag is mainly relevant for UG SE degrees.)

	 Databases and data management systems (used for Data Science MSc and MSc(R)) Unstructured data and applications (used for Data Science MSc and MSc(R)) Level 11 Security courses (used for Security MSc)
If you are not sure which tags are most appropriate or have other questions about this section, please note any comments/issues here.	

6. <u>Comments from colleagues</u>

All course proposal should be sent to relevant colleagues in the area as well as to the appropriate year organizer and BoS Academic Secretary for comment in good time before the BoS meeting. Please indicate here what feedback has been solicited and received.

Additional Comments Summarise any comments received from relevant individuals prior to proposing the course. If you have not discussed this proposal with others please note this.	
Year Organiser Comments Year Organisers are responsible for maintaining the official Year Guides for every year of study, which, among other things, provide guidance on available course choices and specialist areas. The Year Organisers of all years for which the course will be offered should be consulted on the appropriateness and relevance on the course. Issues to consider here include balance of course offerings across semesters, subject areas, and credit levels, timetabling implications, fit into	
the administrative structures used in delivering that year.]	
BoS Academic Secretary Comments Proposals must be checked by the Secretary of the Board of Studies prior to discussion at the actual Board meeting. This is a placeholder for their comments, mainly on the formal quality of the content provided above.	