Course Approval Form

For instructions see:
http://registrar.gmu.edu/facultystaff/catalog-revisions/course/

Action Requested:

- [x] Create new course
- [ ] Inactivate existing course
- [ ] Modify existing course (check all that apply)
  - [ ] Title
  - [ ] Credits
  - [ ] Repeat Status
  - [ ] Grade Type
  - [ ] Prereq/coreq
  - [ ] Schedule Type
  - [ ] Restrictions
  - [ ] Other:

Course Level:

- [ ] Undergraduate
- [x] Graduate

College/School: College of Science
Department: CDS
Submitted by: D. Papaconstantopoulos
Ext: 3-3624
Email:
Subject Code: CDS
Number: 503
(Do not list multiple codes or numbers. Each course proposal must have a separate form.)

Effective Term:

- [x] Fall
- [ ] Spring
- [ ] Summer

Year 2016

Title:
Current:
Banner (30 characters max w/ spaces): Scientific Data Mining
New: Scientific Data Mining

Fulfills Mason Core Req? (undergrad only)

- [ ] Currently fulfills requirement
- [ ] Submission in progress

Credits: (check one)

- [3] Fixed — or
- [ ] Variable — to

Repeat Status: (check one)

- [x] Not Repeatable (NR)
- [ ] Repeatable within degree (RD) — Maximum credits allowed:
- [ ] Repeatable within term (RT)

Grade Mode: (check one)

- [x] Regular (A, B, C, etc.)
- [ ] Satisfactory/No Credit
- [ ] Special (A, B C, etc. +IP)

Schedule Type: (check one) LEC can include LAB or RCT

- [x] Lecture (LEC)
- [ ] Lab (LAB)
- [ ] Recitation (RCT)
- [ ] Internship (INT)
- [ ] Independent Study (IND)
- [ ] Seminar (SEM)
- [ ] Studio (STU)

Prerequisite(s):
Knowledge covered in CDS 130 or CDS 101; or permission of instructor

Corequisite(s):

Instructional Mode:

- [x] 100% face-to-face

Restrictions Enforced by System: Major, College, Degree, Program, etc. (include code)

Equivalencies: (check only as applicable)

- [ ] YES, course is 100% equivalent to:
- [ ] YES, course is being renumbered to/will replace the following:

Catalog Copy for NEW Courses Only (Consult University Catalog for models)

Description (No more than 60 words, use verb phrases and present tense)	**Notes** (List additional information for the course)
Data and databases used by scientists. Includes basics about database organization, queries, and distributed data systems. Student exercises will include queries of existing systems, along with basic design of database systems. Examples from different disciplines will be given.	

Indicate number of contact hours: Hours of Lecture or Seminar per week: 3 Hours of Lab or Studio:

When Offered: (check all that apply)

- [x] Fall
- [ ] Summer
- [x] Spring

Approval Signatures

Department Approval — Date

College/School Approval — Date

If this course includes subject matter currently dealt with by any other units, the originating department must circulate this proposal for review by those units and obtain the necessary signatures prior to submission. Failure to do so will delay action on this proposal.

Unit Name	Unit Approval Name	Unit Approver's Signature	Date

For Graduate Courses Only

Graduate Council Member — Provost Office — Graduate Council Approval Date

For Registrar Office's Use Only: Banner_____Catalog_____

revised 6/22/15

Course Proposal Submitted to the College of Science Curriculum Committee (COSCC)

The form above is processed by the Office of the University Registrar. This second page is for the COSCC's reference. Please complete the applicable portions of this page to clearly communicate what the form above is requesting.

FOR ALL COURSES

Course Number and Title: CDS 503- Scientific Data Mining

Date of Departmental Approval: 9/6/2015

FOR INACTIVATED/REINSTATED COURSES

FOR MODIFIED COURSES

FOR NEW COURSES

- Reason for the New Course: To provide these course topics at the graduate level. This new course will frequently be cross listed with CDS 303- Scientific Data Mining.
- Relationship to Existing Programs: A lower-level graduate course in scientific data mining that can be within reach of graduate students outside of the CDS department (the only other graduate courses that cover this subject are at the 700-level and above which are quite challenging for students outside of CDS's MS or PhD programs).
- Relationship to Existing Courses: Similar to CDS 303, but with graduate-level expectations.
- Semester of Initial Offering: Fall 2016
- Proposed Instructor: Dr. Edward Wegman

CDS 503
SCIENTIFIC DATA MINING

-- SYLLABUS --

Prerequisites: CDS 130 or CDS 101 or permission of instructor
Credits: 3
Instructor: Edward Wegman
Office Hours: TBD

Course Description:

This course provides a broad overview of the data mining component of the knowledge discovery process used by scientists. Scientific databases are growing at near exponential rates. As the amount of data has grown, so has the difficulty in analyzing these large databases. Data mining is the search for hidden, meaningful patterns in such databases. Identifying these patterns and rules can provide significant competitive advantage to scientific research projects and in other work settings. Data mining is motivated and analyzed as the killer application for large scientific databases. Data mining techniques, algorithms, and applications cover learning data types, data preparation, previwing, noise handling, feature selection, normalization, data transformation, similarity measures, and distance metrics. Algorithms and techniques will be analyzed specifically in terms of their application for solving particular problems. Several scientific case studies will be presented from the science research literature. The techniques that are presented will be analyzed specifically in terms of their application to solving particular problems. The techniques that are presented will include clustering, decision trees, regression, Bayes theorem, neural networks, and genetic algorithms. Topics will include informatics, semantic knowledge mining, and the integration of data mining with large scientific databases.

Lecture Content:

- Develop understanding of data mining and its scientific applications
    - Format Standards
    - Data Modalities: Image, Catalog, Time Series, Simulations
    - Curating data
    - Data reliability
- Data mining concepts, techniques, and algorithms
- Data mining for science problems
    - Significance of features
    - Feature selection
    - Feature reduction
    - Data training strategies
    - Learning curves
- Scientific Data Mining – KDD: Knowledge Discovery in Databases
    - Scientific Motivation: Killer App for Scientific Databases
    - Supervised Learning Methods
    - Unsupervised Learning Methods
    - Science Case Studies
        - Bioinformatics
        - Spatial-Temporal Data Mining
        - Outlier Detection: Discovery or Anomaly
    - Data Mining Tools and Software
- Neural networks and genetic algorithms

- Dynamic Data-Driven Discovery and Experiment Driving

Homework:

Students will use on-line and computational resources learn about the paradigms, languages, and methods in data mining. The emphasis will be on a hands-on approach using instructor supplied datasets.

Project:

There will be no class project for this class.

Exams:

Midterm and final exams will be given, based on the content of the lectures and the homework assignments. Short essays as well as analytic calculations about problem complexity, time, and simple examples from programs will be used.

Grades:

Homework (40), Projects (%), Midterm (30%), Final Exam (30%)

Required Text:

.Mathew North, Data Mining for the Masses, Global Text Project, 2012. ISBN 9780615684376.