UNIVERSITEIT VAN AMSTERDAM

# onderwijsinnovatieprojecten FNWI

Dit formulier kan zowel gebruikt worden voor reguliere onderwijsinnovatieaanvragen als subsidieaanvragen in het kader van het facultair strategisch plan Blended Learning.

Voor ondersteuning bij het opstellen van de begroting kunt u contact opnemen met de bedrijfsvoerder Onderwijs bij het ESC, Ingrid van Loon, i.m.vanloon@uva.nl

## 1. ALGEMENE INFORMATIE

Aanvraag in het kader van facultair plan Blended Learning [X ] JA [ ] NEE

Titel van het project

| |
|---|
| Digital Examinations using Jupyter Notebooks |

Context project: opleiding(en)

| |
|---|
| Bachelor Informatiekunde, in particular the courses Netwerkanalyse (BsC Informatiekunde Year 1, period 2), Data Science (BsC Informatiekunde Year 3, period 4), Data Mining (BsC Informatiekunde Year 2, period 4), Collectieve Intelligentie (BsC Informatiekunde Year 1, period 5).<br><br>Afterwards, also courses in Bachelor AI, minor programming and probably even in other schools or even faculties (see below) |

Context project: curriculumonderdeel / leerlijn / nieuw vak*

| |
|---|
| Examination |

*in geval een nieuw vak, dan graag op deze plek ook de volgende gegevens vermelden: omvang (EC), type onderwijs (regulier, keuzevak, honours, minor)

Projectleider

| |
|---|
| Maarten Marx (FNWI, IvI, UD), |

Aangesteld bij instituut / capaciteitsgroep:

| |
|---|
| IvI |

Contactgegevens

| |
|---|
| maartenmarx@uva.nl, |

## 2. PROJECTVOORSTEL

Aanleiding / onderbouwing (bij nieuw vak: beoogde doelgroep, verwachte instroom)

| |
|---|
| For authentic assessment in courses which use computer software, in particular in the case of learning outcomes such as investigate, model and predict from real life datasets, the known commonly used digital assessment applications are not suitable to design exams which truly test these Intended Learning Outcomes (ILO) ELOs. In this project, we will develop a new alternative digital assessment method in which no classical digital assessment tool is needed but students use Jupyter Notebooks in which they directly solve exam questions/assignments.<br>All courses in the Bachelor Informatiekunde involving programming (and starting in |

2017-2018 also in the Bachelor AI) will use one programming language: Python. Python (but also languages like Matlab and R) can be programmed interactively in the form of Jupyter Notebooks. This notebook format allows a combination of interactive programming (with answers to code appearing in the notebook) and nicely marked up text (using a combination of MarkDown and LaTeX). These notebooks are ideal to test whether a student masters actively stated learning outcomes like investigate, model, predict from real life datasets. The advantages for the student are:

- Exam is similar to the practiced problem solving methods during the course and in the assignments.
- The notebook has lots of low level support helping the student to focus on deeper skills than rote-learning (e.g., autocompletion on variables and methods of objects, complete manual inside the notebook).
- Students can experiment with an answer and receive immediate feedback (about syntax errors, and, when they use good tests, about semantic errors; they do not receive feedback about the correctness of the answer). This is the natural way of solving problems.

The advantages for the teacher are:

- Notebook exam allows the teacher to really test the intended learning outcomes stated with active verbs.. Students work actively with the notebook during the exam, instead of producing prelearned facts. Exams can contain realistic problems about big authentic datasets. In an exam made by hand this is not feasible due to time constraints.
- Exam is fully aligned to assignments given during the course. The only difference is that the exam is individual, and it is not allowed to consult internet or others.
- The exam is "open" in the sense that the teacher must check and grade all questions. However, the notebook format makes that this can be done very efficiently. One loads the notebook, runs all cells, and immediately sees whether the provided code runs and yields the desired output. (This is an exact mirror of what the students have done during the exam.) Then one can inspect how the problem was solved. The cell by cell nature of the notebook ensures that it is easy to see which answer belongs to which question. Further speedup can be obtained using automatic grading through unit tests (this is supported by the nbgrader software [1] for these notebooks, but this aspect is not part of the current proposal). In a tryout with an exam Datascience made by 60 students, we could grade all exams with 4 people (the teacher and the 3 Tas) in 2 hours, calibrating scores on the spot. Three hours later the students knew their grade through Blackboard.

[1] Jess Hamrick, Brain Granger, (2015) http://nbgrader.readthedocs.io/

Doel van het project

UvA has excellent electronic test facilities at IWO and at Roeterseiland. Here students can make exams on UvA computers. These exams are created and taken using third party software (now Remindo, from next year on TestVision or SOWISO). It is possible to use external software to let the student compute things (e.g. Excel, SPSS).
We have asked twice to install Jupyter Notebooks, so we could run our desired

notebook exams. However, both times this was rejected by the UvA computer test dept, with the reason that the notebooks may affect the integrity of the system.

This is the reason that we now apply for funding to work out the idea ourselves.

Initial investigations and brainstorming with the Faculty of Science computer support makes us believe that it is possible to do exams using Jupyter Notebooks without additional risks of fraud. The main problem for security is that for logistic reasons students need to make the exam on a computer which is connected to the internet. The notebooks are rather powerful and give access to the operating system of the computer. These two combined give the student many opportunities for communication with others and thus fraud.

In this project we want to investigate how we can solve this problem by having the students work on virtual machines in the cloud (which exist only for the duration of the exam) which are, through firewall and/or parental control like mechanisms, disconnected from the internet and exclusively accessible for the space that we explicitly allow. This way the exams could in principle be taken in an UvA digital assessment room because the security of the UvA assessment system and network will not be put in any danger what so ever.

Luckily we are not the first, and an open source prototype for setting up virtual machines running notebooks exists [1], which has been tested for Computer Science exams at the University of Berkeley. We want to test two options: (1) using existing machines at the FNWI (the agile platform of FEIOG), which have also been used to create virtual machines for the Webprogrammeren course (1st year, about 260 students). Here our focus will be on security, scalability and fairness: how much is one student affected by heavy use of the computer by another student? (2) rent computers in the cloud, e.g., Amazon or Google. Scalability here is not an issue. We will focus on fairness again, on costs, and extra on security as it is outside UvA. The second option is scalable and can be of interest in case other courses or faculties want to use the developed system, or if we need to scale beyond the capacity of the FNWI machines.

We will work closely together with Natasa Brouwer, the FNWI expert on digital exams. She will assist us in evaluating the notebook exams. We plan to have the exams assessed by the FNWI testing-expert Stephanie Maijs, and plan to compare the notebook exams with traditional exams held in previous years.

[1] Jess Hamrick, JupyterHub, (2015) < https://developer.rackspace.com/blog/deploying-jupyterhub-for-education/>

In geval van een Blended Learning aanvraag: toelichting op de beoogde inzet van ICT

The courses in which this will be used are blended learning courses. The “blended” aspect lies in the fact that we use **interactive** ICT to boost digital exams to a level which could only be obtained in a traditional oral examination and assures a perfect blend.
The interactive nature of the notebooks enables the natural academic “think-try-out-receive feedback-evaluate-answer” cycle which we want our students to develop. This is also one of our learning outcomes that we could not adequatly **individually** test without the proposed project.

Beoogde uitkomsten

We intend to radically change how exams are being taken on computers. If the UvA wants to scale to the fast rising number of students, electronic exams are necessary. But automatically graded exams cannot test learning outcomes which ask for tasks as realistic problem solving, modeling, data processing and exploration, etc, especially when these activities are naturally done using computer applications. Such learning outcomes can better be tested in an interactive environment in which the student can try out things, receive natural feedback, has access to a manual, and need not make "dumb (often syntax) mistakes caused by stress". Recall that the students have been working in this environment during the whole course (in fact in Informatiekunde Bsc, during **all courses** in which programming is involved). This environment is provided by the Jupyter Notebooks.

As such notebooks exist for several languages (e.g. Python, Matlab, R) which are used in many courses from different facilities, we expect that the proposed way of examining will be picked up broadly at the UvA.

Our intention is to integrate the "notebooks exams" as much as possible in the currently existing computer exam environment of the UvA and HvA.

This development has a high potential and is a natural continuation in a digital assessment process in a curriculum when approaching high cognitive goals and discipline specific learning processes.

**Dissemination**
The development and the results will be communicated at the the level of FNWI and UvA.
We aim for a publication about the project in a professional journal directed to higher education.

Evt. neveneffecten

We reiteratie: as such notebooks exist for several languages (e.g. Python, Matlab, R) which are used in many courses from different facilities, we expect that the proposed way of examining will be picked up broadly at the UvA and also at other universities.

SOWISO, one of the digital assessment applications at the UvA with a license, extensively used at the Faculty of Science, has computer algebra Maxima as engine which makes it possible to design randomized question sets, automatic check of open questions and instant feedback in formative assessments. SOWISO is currently developing an extension to other other engines, like R and Matlab. These applications will improve the possibilities for scaffolding in practicing and self assessment of problem solving in closed problems. For assessment of realistic problem solving, real modeling, data processing and exploration assessment the notebooks exams is the best suitable solution. However, the two technologies could be integrated in a logical way and could follow each other in a curriculum in different science disciplines.

Plan van aanpak & fasering

We first intend to study similar solutions and best practices. Based on that we will set up a test environment, perform stress testing for testing fairness and scalability and do experiments with mock exams (period 2, course Netwerkanalyse, 100 students).

We also intend to have the system "hacked" by master students System and Network Engineering in period 2. If all goes well, we use the notebooks for real exams in the Data Science course (period 4, 60 students). If that test works too, we apply the notebook exams also in the Data Mining and Collective Intelligence courses (period 5). We note that in 2017, we have done the Data Science exam and resit using notebooks on students own laptops together with intense surveillance and online Proctoring.

We intend to talk to FNWI SOWISO team about the possible collaboration in this project.

Looptijd

1 year

Wijze van evaluatie / effectmeting (is beoogde doelstelling behaald?)

Measurable outcomes: As we better test what we have taught our students, we expectbetter constructive alignment of the courser and thus

- higher grades on the exam
- higher appreciation in student evaluations about the exams

Second, we hope to convince others to follow our proposed approach.

**3. PROJECTORGANISATIE**

Andere leden projectteam (indien van toepassing): naam, instituut/afdeling, rol binnen project

- Robin de Vries (FNWI, Datanose) Role: technical development
- Natasa Brouwer (FNWI Digital Exam coordinator): Role: liaison UvA digital exams, digital exam expert
- Student Assistant. Role: quality control, testing, grading, auto-grading testing

Andere incidenteel betrokkenen (klankbord, advies, verantwoording)

Help from Stephanie Maijs for checking whether we indeed better test the desired learning outcomes.

Wensen t.a.v. ondersteuning vanuit ESC of extern

## 4. FINANCIEN / BEGROTING

- kosten inzet wp
- kosten inzet student-assistenten (op basis van aantal uren & looptijd)
- kosten externe inhuur
- materiële / systemische kosten
- overige uitgaven

Begroting totaal (excel met specificaties toevoegen als bijlage)

| |
|---|
| See attached Excel sheet . Totaal 10.307 Euro |

Opmerkingen / bijzonderheden

| |
|---|
| Hours needed: Best practice investigation (20H), technical setup on cloud platform (40H), testing + adjusting (60H), manual, example exams, publicity (30H), automatic and semi-automatic grading (30H), PI: 20H.<br><br>Resources needed: Cloud platform (either our own, or we rent at Amazon/Google). We have booked 2K Euro to experiment with different external cloud providers and to start quickly. |

## 5. ONDERTEKENING

Aanvraag besproken en geaccordeerd door opleidingsdirecteur:

| |
|---|
| Naam: Jacobijn Sandberg<br>Datum: 2017-07-07 |