MSc. in Computing
Practicum Approval Form

Section 1: Student Details

Project Title:	An Enhanced Data Catalog to Support Trustworthy Federated Learning
Student ID:	21264466, 20210611
Student name:	Stefano Marzo, Royston Pinto
Student email:	stefano.marzo2@mail.dcu.ie royston.pinto2@mail.dcu.ie
Chosen major:	Artificial Intelligence, Data Analytics
Supervisor:	Dr Rob Brennan
Date of Submission:	27/01/2022

Section 2: About your Practicum

Please answer all questions below. Please pay special attention to the word counts in all cases.

What is the topic of your proposed practicum? (100 words)

Development of a data analysis tool integrated into a data governance platform that produces metadata about the fairness of a federated dataset. The metadata created can be used to classify data more accurately by reducing bias related to the presence of minority groups within the dataset. The project, motivated by recent scientific results, aims to implement unfairness quantification as part of the fundamental principles described in the Ethics guidelines for trustworthy AI authored by the High-Level Expert Group on Artificial Intelligence (AI HLEG) on behalf of the European Commission.

Please provide details of the papers you have read on this topic (details of 5 papers expected).

1. Chen, Z., Tian, P., Liao, W., & Yu, W. (2021). Zero knowledge clustering based adversarial mitigation in heterogeneous federated learning. IEEE Transactions on Network Science and Engineering, 8(2), 1070–1083. https://doi.org/10.1109/TNSE.2020.3002796
2. Deng, Y., Kamani, M. M., & Mahdavi, M. (2021). Distributionally robust federated averaging.
3. Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., D'Oliveira, R. G. L., Eichner, H., Rouayheb, S. E., Evans, D., Gardner, J., Garrett, Z., Gascón, A., Ghazi, B.,

Gibbons, P. B., . . . Zhao, S. (2021). Advances and open problems in federated learning.
4. Taskesen, B., Nguyen, V. A., Kuhn, D., & Blanchet, J. (2020). A distributionally robust approach to fair classification.
5. Xie, M., Long, G., Shen, T., Zhou, T., Wang, X., Jiang, J., & Zhang, C. (2021). Multi-center federated learning

How does your proposal relate to existing work on this topic described in these papers? (200 words)

As [3] reports, fairness is a major concern in the context of federated learning. To enhance fairness in a federated environment, distributionally robust optimization (DRO) [2, 4], together with multi-center model aggregation [5] are proposed. One of the main problems due to federated learning (FL) is considering the processed data as heterogeneous and representative of the whole population, which can lead to bias when processing data belonging to minority groups. As reported in [2,4], population characteristics can have different distributions among subgroups. The information extracted from the data may be strongly based on the distributions of the majority groups, introducing unfairness. As suggested by [5], it is paramount to recognise possible subgroups within the population *a priori*, so that the influence of minorities can be balanced. To achieve these results without violating the privacy principles underlying FL, Zero Knowledge Proof (ZKP) models can be employed to record metadata regarding the proportions of population clusters [1]. The collected metadata can make a difference in predictions and decision making processes as they can help to identify population groups and calibrate analyses more accurately.

What are the research questions that you will attempt to answer? (200 words)

- *To what extent can we mitigate federated data bias by using data governance systems according to EU guidelines for data ethics and trustworthy AI?*

- *To what extent can zero knowledge proof metadata about the proportions of population clusters generated in a federated learning environment can be used to enhance trustworthiness in AI?*

How will you explore these questions? (Please address the following points. Note that three or four sentences on each will suffice.)

- What software and programming environment will you use? *Using the CKAN data governance platform, we will develop the proposed framework for fair data analysis using mainly python.*

- What coding/development will you do? *We want to develop the proposed data analysis framework in the form of a plug-in for the CKAN data governance platform.*

- What data will be used for your investigations? *In order to conduct experiments on data fairness detection we will use the* ***UTKFace*** *dataset available here:* *https://susanqq.github.io/UTKFace/*.
The aforementioned is a collection of more than 20,000 face images labeled with age, gender, ethnicity.

- Is this data currently available, if not, where will it come from? *dataset available*

- What experiments do you expect to run? *In a simulated federated environment:*

We want to build a statistical tool for measuring unfairness in a federated environment, we will show that for low unfairness values, predictions on the data are more accurate for minority groups.

We want to apply multi-center aggregation to make more accurate predictions for data regarding a minority group. We will measure the improvement in predicting one or more features with respect to the cluster of which data belong. We want to show that with this method is it possible to reduce bias.

- What output do you expect to gather? *The expected output will be in the form of statistics that we can easily display in the form of "before - after" comparison*

- How will the results be evaluated? *Result evaluation will be made through the use of proper statistical tests to emphasize the significance of results.*