H2020 – ICT-13-2018-2019

MUSKETEER

Machine Learning to Augment Shared Knowledge in Federated Privacy-Preserving Scenarios (MUSKETEER)

Grant No 824988

D8.4 Scientific dissemination activities

November 21

Imprint

Contractual Date of Delivery to the EC:	**31 Nov 2021**
Author(s):	**Angel Navia Vázquez, Jesús Cid Sueiro, Manuel Vázquez López (UC3M)**
Participant(s):	**Antoine Garnier (IDSA)**
Reviewer(s):	**Chiara Napione, Javier Gutierrez**
Project:	**Machine learning to augment shared knowledge in federated privacy-preserving scenarios (MUSKETEER)**
Work package:	**WP8**
Dissemination level:	Public
Version:	**1.1**
Contact:	**Angel Navia – angel.navia@uc3m.es**
Website:	**www.MUSKETEER.eu**

Legal disclaimer

The project Machine Learning to Augment Shared Knowledge in Federated Privacy-Preserving Scenarios (MUSKETEER) has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 824988. The sole responsibility for the content of this publication lies with the authors.

Copyright

Executive Summary

This document provides a detailed view on the MUSKETEER's project activities related to scientific dissemination. As a Research and Innovation Action, we expect to largely disseminate our results in the scientific community but also among various organizations interested or involved in the topics of the project. Therefore, our project combines external and internal means to disseminate and communicate the information among the partners and their networks, as well as among the scientific community.

Document History

Version	Date	Status	Author	Comment
1	01 Nov 2021	For internal review	Angel Navia	First draft
2	08 Nov 2021	Internal review	Javier Gutiérrez	Minor changes
3	08 Nov 2021	Internal review	Chiara Napione	Minor changes
4	11 Nov 2021	Review completed	Angel Navia	Minor corrections
5	19 Nov 2021	Final Review	Mark Purcell & Gal Weiss	Minor corrections

Table of Contents

List of Tables

List of Acronyms and Abbreviations

Abbreviation	Definition
AAAI	Association for the Advancement of Artificial Intelligence Conference
BDVA	Big Data Value Association
DCP	Dissemination and Communication Plan
ECCV	European Conference on Computer Vision
ECML	European Conference on Machine Learning
ESSAN	European Symposium on Artificial Neural Networks
ICAPAI	International Conference on Applied Artificial Intelligence
ICLR	International Conference on Learning Representations
ICML	International Conference on Machine Learning
ICMLA	International Conference on Machine Learning and Applications
IDP	Industrial Data Platform
IEEE	Institute of Electrical and Electronics Engineers
JN	Journal of Neurocomputing
KPI	Key Performance Indicator
MLC	Machine Learning Conference
NeurIPS	Neural Information Processing Systems
PR	Pattern Recognition
TIST	ACM Transactions on Intelligent Systems and Technology
TNNLS	Transactions on Neural Networks and Learning Systems
TPDS	IEEE Transactions on Parallel and Distributed Systems

1 Introduction

1.1 Purpose

The MUSKETEER main strategy is to raise attention on distributed model training without actually sharing the training data and widen the platform audience to new stakeholders in order to sustain the development of our industrial data space. The Dissemination and Communication Plan (DCP) described in D8.2 specifies the actions to disseminate and communicate the project results for the consortium in order to prepare the best exploitation phase possible. In this document we focus on the scientific dissemination activities carried out during the project, mainly contributions to Workshops, Conferences, Journals and whitepapers/book chapters with a clear technical orientation.

1.2 Related Documents

The Dissemination and Communication Plan drives WP8 activities, by defining the Project communication and engagement activities (D8.3), the Scientific dissemination activities (D8.4), the Community engagement and Technology Transfer activities (D8.5), providing inputs for the Evaluation and impact assessment (D8.6). The outcomes of the activities presented in deliverables D8.3, D8.4, D8.5 and D8.6 will provide inputs to adjust the business plans and prepare the exploitation plan (D8.7).

1.3 Document Structure

The document is divided in two main sections. The first one defines the dissemination and the communication strategy. The second section describes the tracking process of those activities.

2 Scientific dissemination

As a Research and Innovation Action, we have a continuous scientific activity along the project. Our aim is to highlight the potential of technological innovation in order to drive adoption and facilitate realising the value of data. This should also be a way to draw more researchers towards addressing our challenges. Our strategy for scientific dissemination relies on several types of actions:

- **Journals**: this is the preferred channel for scientific dissemination, since it implies a well implemented and widely recognized peer-review process. At the same time, it is the most challenging for achieving results, since the journal acceptance rates have significantly decreased because of the huge amount of research activity in the Machine

Learning field. Also, we have observed that the review times are getting longer, in some cases, more than 12 months to get a first reply about a submitted manuscript, which hampers the completion of the initiated research activities, and also slows down possible continuity papers on initiated research tracks.

- **Technical whitepapers/book chapters:** given the high relationship of the MUSKETEER project with the Big Data Value Association (BDVA), part of the scientific dissemination has been redirected towards whitepapers or book chapters linked to BDVA. Since in most cases there is a relevant technical and scientific contribution, we will also include here those contributions.
- **Conferences:** this dissemination channel is also very important, especially in the case of well-known international events such as Neural Information Processing Systems (NeurIPS), Machine Learning Conference (MLC), European Conference on Machine Learning (ECML), International Conference on Machine Learning (ICML), International Conference on Learning Representations (ICLR), European Symposium on Artificial Neural Networks (ESSAN), International Conference on Applied Artificial Intelligence (ICAPAI 2021), IEEE International Conference on Machine Learning and Applications (ICMLA), Association for the Advancement of Artificial Intelligence Conference (AAAI), etc. In this case, the increased research activity in the machine learning field has also affected the acceptation rates. Additionally, the covid restrictions have reduced the amount of events for several months, as well strongly limiting the kind of possible interactions among the participants (many have been reduced to online events).
- **Workshops:** we include here the scientific dissemination activities carried out in the context of workshops and other technical forums, where the MUSKETEER project (as a whole or any of its components) has been presented and discussed with a technical audience as a target. Again, given the high relationship of the MUSKETEER project with the Big Data Value Association (BDVA), several of those presentations took place in BDVA context events.

Therefore, among all the dissemination activities carried out along the project and listed in Table 1, we will focus in this deliverable on the scientific activities related to items No. 4, 6, 9, 10, and 11. The rest of events or dissemination activities will be described in the context of D8.5.

Table 1 Different types of initiatives related to dissemination

01. **Online:** Newsletter, email
02. **Online:** Website
03. **Online:** Social Media
04. **Event:** Workshops
05. **Event:** Trade show, exhibitions
06. **Event:** Conference
07. **Event:** Presentation / Lecture
08. **Event:** Hackathon
09. **Publications:** Conference papers
10. **Publications:** White papers
11. **Publications:** Scientific paper
12. **Publications:** Article / Interview (press, other media)
13. **Liaisons** with Innovation Actions
14. **Liaisons** with 3rd Parties
15. **Liaisons** with National Initiatives

3 Scientific dissemination: timeline and results

We will detail here the scientific dissemination activities carried out during the project, indicating the date and status of every activity, and grouping them according to the KPI objectives defined in the DOW and further specified in D8.3, as follows:

Table 2 Scientific publication activities (sorted by date)

Scientific publications (JCR journals, conference publications, book chapters)

(KPI target is 12 activities)

1. L. Muñoz-González, K. T. Co, E.C. Lupu. **Byzantine-Robust Federated Machine Learning through Adaptive Model Averaging**. Association for the Advancement of Artificial Intelligence Conference AAAI'19. Febr. 2019. Rejected.

2. G. Collinge, E. Lupu, L. Muñoz-González. **Defending against Poisoning Attacks in Online Learning Settings**. Proceedings of the 27th European Symposium on

Artificial Neural Networks, Computational Intelligence and Machine Learning. ESANN. April 2019. Bruges, Belgium.

3. T. Timan, Z. Mann (Eds.), R. Araujo, A. Crespo-García, A. Farkash, A. Garnier, A. Vivian-Kiousi, P. Koster, A. Kung, G. Livraga, R. Díaz-Morales, M. Önen, A. Palomares, A. Navia-Vázquez, A. Metzger (contributors). **Data protection in the era of artificial intelligence. Trends, existing solutions and recommendations for privacy-preserving technologies**, October 2019. BDVA position paper. https://www.bdva.eu/node/1384

4. Gusmeroli S., Dalle Carbonare D. (eds). **Big Data challenges in Smart Manufacturing Industry** (ed. 2020). BDVA Whitepaper. https://www.bdva.eu/sites/default/files/BDVA_SMI_Whitepaper_2020.pdf

5. L. Muñoz-González, B. Pfitzner, M. Russo, J. Carnerero-Cano, E.C. Lupu. **Poisoning Attacks with Generative Adversarial Nets**. International Conference on Machine Learning (ICML 2020). Febr 2020. M14. Rejected.

6. A. Navia-Vázquez, M.A. Vázquez, and J. Cid-Sueiro. **Robust On-line Data Value Estimation in Federated Machine Learning scenarios**. Submitted to Journal of Neurocomputing. March 2020. Rejected.

7. A. Navia-Vázquez, M.A. Vázquez, and J. Cid-Sueiro**. First vs Second Order Double Confidential Federated Machine Learning for Logistic Regression**. Submitted to Journal of Neurocomputing. March 2020. Rejected.

8. L. Muñoz-González, B. Pfitzner, M. Russo, J. Carnerero-Cano, E. C. Lupu. **Poisoning Attacks with Generative Adversarial Nets**. International Conference on Learning Representations (ICLR 2020). May 2020. Rejected.

9. A. Navia Vázquez, M.A. Vázquez-López and J. Cid-Sueiro. **Double Confidential Federated Machine Learning Logistic Regression for Industrial Data Platforms.** International Workshop on Federated Learning for User Privacy and Data Confidentiality in Conjunction with ICML 2020 (FL-ICML'20) July, 2020.

10. K. Co, L. Muñoz-González, L. Kanthan, B. Glocker, E. Lupu. **Universal Adversarial Perturbations to Understand Robustness of Texture vs. Shape-biased Training**. European Conference on Computer Vision. ECCV'20.Aug. 2020. Rejected

11. A. Navia-Vázquez, M.A. Vázquez, and J. Cid-Sueiro. **First vs Second Order Doubly Confidential Distributed Learning for Logistic Regression**. IEEE Transactions on Parallel and Distributed Systems. Submitted Jan 2021.

12. A. Navia-Vázquez, R. Díaz-Morales, M. Fernández-Díaz, **Budget Distributed Support Vector Machine for Non-ID Federated Learning Scenarios.** ACM Transactions on Intelligent Systems and Technology. Special Issue on

Federated Learning: Algorithms, Systems, and Applications. Submitted March 2021.

13. S. Rossello, S., L. Muñoz-González, and R. Díaz Morales. "**Data protection by design in AI. The case of federated learning**." Computerrecht: Tijdschrift voor Informatica, Telecommunicatie en Recht 3 (May 2021): 273-279.

14. S. Bonura, D. Dalle Carbonare, R. Díaz-Morales, A. Navia-Vázquez, M. Purcell and S. Rossello**. Increasing Trust within a Data Space with Federated learning**, in **Data Spaces: Design, Deployments, and Future Directions**, BDVA Book Chapter. 2021.

15. S. Bonura, D. Dalle Carbonare, R. Díaz-Morales, M. Fernández-Díaz, L. Morabito, L. Muñoz-González, C. Napione, A. Navia-Vázquez, M. Purcell. **Privacy Preserving Technologies for Trusted Data Spaces**. BDVA Book Chapter. 2021.

16. A. Rawat, G. Zizzo, M. Zaid Hameed, and L. Muñoz-González, **Security and Robustness in Federated Machine Learning**. Book chapter in "Federated Learning: A Comprehensive Overview of Methods and Applications". Springer, 2021.

17. A. Navia-Vázquez, M.A. Vázquez, and J. Cid-Sueiro. **"A Priori" Shapley Data Value Estimation for Risk-Balanced Data Monetization in Federated Learning**. Pattern Recognition, Elsevier. Submitted Nov. 2021.

Table 3 Scientific dissemination activities (sorted by date)

Conference presentations

(KPI target is 12 activities)

1. **Conference Big Data and AI Tech World 2019**, 12/03/2019. https://www.bigdataworld.com/conference-programme

2. **ECR 2019, European Society of Radiology**. 28/02 -03/03/2019. Austria, https://www.myesr.org/past-congresses/ecr-2019

3. **AI Law & Ethics Conference**, KU LEUVEN, 28/02/2019, Belgium, https://www.law.kuleuven.be/citip/en/news/item/citip-conference-through-the-looking-glass-of-ai-platforms-between-global-governance-and-techno-regulation

4. **9th Annual Data Protection and Privacy Conference**, 20/03/2019, Belgium, https://eu-ems.com/summary.asp?event_id=4382&page_id=9802

5. **Towards Value Centric Big Data Workshop (E-Sides)**, 02/04/2019, Belgium, https://www.evensi.be/centric-big-data-connect-people-processes-technology-imec-smit-vub/295256964

6. **Spotlight on Ethics and Bias in AI for Healthcare**. London Clinical and Health Data Science Meetup, 16/05/2019, https://www.meetup.com/LonClinDatSci/events/259386351/

7. **Biotronics3D Workshop: welcome to the future of Radiology**. 24/05/2019 www.biotronics3d.com

8. **Workshop on Artificial Intelligence for Manufacturing, EFFRA**, BDVA, euRobotics, 02/07/2019. Belgium, https://ec.europa.eu/digital-single-market/en/news/workshop-artificial-intelligence-manufacturing

9. **Liaisons with National Initiatives. The London Medical Imaging & Artificial Intelligence Centre for Value Based Healthcare**. 8/09/2019, UK. https://www.kcl.ac.uk/bmeis/research-impact/london-medical-imaging-and-ai-centre-for-value-based-healthcare

10. **Delivering Data Protection in Real Time Workshop**, Oxford University / OASIS, 09/09/2019, UK, https://privacyworkshop19.oasis-open.org

11. **Common European data spaces for Smart Manufacturing Conference**, 16/09/2019, Belgium, https://ec.europa.eu/digital-single-market/en/news/common-european-data-spaces-smart-manufacturing

12. **Big Things 2019**, 01/11/2019, https://www.bigthingsconference.com/, https://www.youtube.com/watch?v=Pjjd53MwLGA

13. **Digitalization of Knowledge and Industrial Technologies Conference**. University of Bologna, 7/11/19, Italy, https://eventi.unibo.it/dokit-2019/agenda

14. **Theory and Practice of Differential Privacy (TPDP) at 26th ACM Conference on Computer and Communications Security (CCS 2019)**, 11/11/2019, https://www.sigsac.org/ccs/CCS2019/

15. **Conference on Big Data LDN**. London, UK, 13-14/11/2019, https://bigdataldn.com/

16. **Liaisons with National Initiatives. The London Medical Imaging & Artificial Intelligence Centre for Value Based Healthcare**. 21/11/2019, UK. https://www.kcl.ac.uk/bmeis/research-impact/london-medical-imaging-and-

ai-centre-for-value-based-healthcare

17. **Privacy in Machine Learning (PriML) at NeurIPS 2019**, 14/12/2019, Canada, https://priml-workshop.github.io/priml2019/

18. **AI Law & Ethics Conference**, 18/02/2020, Belgium, https://www.law.kuleuven.be/citip/en/citip-conferences/lailec/lailec-2020/programme

19. **Conference for Participation on drafting of the Best Success Story for BDV**. BDVA Activity Group Conference, BDVA, 30/04/2020.

20. G. Zizzo, A. Rawat, M. Sinn, B. Buesser. FAT: Federated Adversarial Training. **NeurIPS 2020 Workshop on Scalability, Privacy, and Security in Federated Learning (SpicyFL)** December 5th-12th, 2020 - Virtual Only. https://nips.cc/virtual/2020/public/workshop_16123.html

21. **Data protection impact assessment: the MUSKETEER's project as a use-case**. 12/10/2021. Workshop with data protection law experts of the KUL's Center for IT and IP Law.

22. Citip's Safe-Deed closing event "**Addressing legal, technical and ethical challenges in the data market context**". 02/12/2021. https://www.law.kuleuven.be/citip/en/news/item/safe-deed-closing-event-addressing-legal-technical-and-ethical-challenges-in-the-data-market-context

In Table 4 below, we summarize the activity corresponding to scientific dissemination along the project, in relation to the KPI described in D8.3.

Table 4 Achievements in regard to the DoA KPIs, interpreted as in D8.3

Activity	Targeted communities	KPIs (at least)	Responsible	M36
Scientific publications (JCR journals, conference publications, book chapters)	Scientific/research community, industrial companies, SMEs	**12** activities	All; led by scientific partners.	**17** submitted, **8** accepted. **3** under review, **6** rejected,
Conference presentations	Scientific/research community, industrial companies, SMEs	**12** activities	All; led by the project coordination	**22** participations

4 Conclusions

The main objective of the MUSKETEER project was to implement an open source Industrial Data Platform (IDP) such that everyone could obtain improved Machine Learning models by training them on a distributed setup, without the participants directly exposing their training data. A great deal of effort went into the first part of the project, for the implementation of the needed libraries and software components, so that a working version of MUSKETEER was available to the consortium as soon as possible. This approach has served the objective that the KPIs associated with the deployment and assessment of libraries and end user cases were met on time, but at the cost of somehow delaying the scientific dissemination activities. The main dissemination activity concerning general aspects of the project has mainly concentrated during the first year of activity (2019), to present the MUSKETEER initiative to the largest possible scientific community (see scientific dissemination events, Table 3). The more specific research activity was concentrated in the last part of the project, since to run some experiments we needed an operative version of the platform and preliminary implementations of the proposed methods and algorithms. This scientific dissemination activity is best observed in Table 2, where most results are concentrated in the last part of the project (years 2020 and 2021), with several works still under the review process. We have observed some publishing difficulties due to the huge amount of research activity in the Machine Learning area, and also excessively prolonged response times from some journals (we are still waiting for response about some papers submitted almost one year ago). Anyhow, we think we have completed a full scientific dissemination strategy, to be continued during the upcoming months, since many of the research lines are of long-term interest for the involved research groups. We summarize in Table 4 the achievements in relationship with the corresponding KPIs specified in D8.3. We consider that the KPI objectives have been fulfilled, especially taking into account that some additional research works product of the MUSKETEER activities are still pending for completion/submission.

5 References

AAAI Association for the Advancement of Artificial Intelligence Conference https://www.aaai.org/

BDVA Big Data Value Association https://www.bdva.eu/

ECCV European Conference on Computer Vision https://eccv2020.eu/

ECML European Conference on Machine Learning https://2021.ecmlpkdd.org/

ESSAN European Symposium on Artificial Neural Networks https://www.esann.org/

ICAPAI International Conference on Applied Artificial Intelligence http://www.icapai.org

ICLR International Conference on Learning Representations https://iclr.cc/

ICML International Conference on Machine Learning https://icml.cc/

ICMLA International Conference on Machine Learning and Applications https://www.icmla-conference.org/

JN Journal of Neurocomputing https://www.journals.elsevier.com/neurocomputing

MLC Machine Learning Conference https://mlconf.com/

MUSKETEER Website http://musketeer.eu/

NeurIPS Neural Information Processing Systems https://nips.cc/

PR Pattern Recognition https://www.journals.elsevier.com/pattern-recognition

TIST ACM Transactions on Intelligent Systems and Technology https://dl.acm.org/journal/tist

TNNLS Transactions on Neural Networks and Learning Systems https://cis.ieee.org/publications/t-neural-networks-and-learning-systems

TPDS IEEE Transactions on Parallel and Distributed Systems https://www.computer.org/csdl/journal/td