DAIS ITA – BPP20

Distributed Analytics and Information Sciences

International Technology Alliance

Biennial Program Plan 2020

15th Jan 2020 – 14th Jan 2021

Summary of Accomplishments

Written on 15st March 2021
Agreement W911NF-16-3-0001

Table of Contents

Summary of Accomplishments

Executive Summary

In this report covering the BPP20 year one period we have chosen to highlight certain key publications for each of the four projects. Please refer to the Science Library (http://sl.dais-ita.org/science-library) for all publications during this period and the whole of the DAIS ITA program so far. At the time of writing there are a number of "in flight" publications that have not yet been published to science library since they are under review, or as yet unpublished in their external conference or journal format, however for the BPP20 year one period we have confirmed 17 journal papers and 56 external conference papers, all of which have been published to the science library. Our estimate is that up to an additional 7 journal papers and 36 external conference papers will eventually be published for BPP20 that were created and submitted in the year one period, but will be published in year 2. In addition, open source components continue to be maintained and will be added to, on github at https://github.com/dais-ita/.

The contributions and highlighted results of each project are summarised below.

Project 7: Policy-enabled Dynamic Infrastructure

In year one researchers from Imperial College, Purdue University, Yale University, Southampton University and University College London have made significant contributions on different aspects of policy-enabled dynamic infrastructure for efficient management of the networked computing system across coalition members, and (federated) policy learning methods that support transfer and domain adaptation between source and target domains of different coalition parties with limited access to data. Details are given below per task.

Task 7.1: Infrastructure Design & Distributed Control for Dynamic SDC

A team of researchers from Imperial College, Yale University and IBM U.S. have developed a new learning architecture for efficient reinforcement learning, based on separation of state and action spaces, to overcome the enormity of the state-action space in deep reinforcement learning (DRL) methods. They have shown that the convergence rate of this new approach is faster than DRL baselines (https://dais-ita.org/node/5659). In collaboration with DSTL U.K. and ARL U.S. the team has also investigated theoretical modelling and quantification of performance enhancement of distributed SDN architectures, exploring in particular how this is influenced by inter-domain synchronization levels and network structural properties (https://dais-ita.org/node/4663). Informed by these studies they have designed robust and scalable controller synchronization policies for distributed SDCs, by leveraging a combination of reinforcement learning and deep learning techniques, which have shown to render much superior performance when compared to existing schemes (https://dais-ita.org/node/4662). Researchers from Imperial College have also developed, in collaboration with IBM U.S., a novel embedding technique for efficient reinforcement learning, which jointly embeds states and actions combining aspects of model-free and model-based reinforcement learning (https://dais-ita.org/node/5393).

As for the efficient management of networked computing systems across coalition members, a team of researchers from Imperial College, University of Massachusetts, IBM U.S., BBM Technologies and U.S. ARL have explored resource sharing in mobile edge clouds (e.g., SDC), by modelling it as a multi-objective optimization problem and developed a game-theoretic framework, based on Cooperative Game Theory (CGT), showing that the proposed algorithms provide stable Pareto optimal allocations of all the services providers in a coalition. This work has been published in F. Zafari, K. Leung, D. Towsley, P. Basu, A. Swami, J. Li, "Let's Share: A Game-Theoretic Framework for Resource Sharing in Mobile Edge Clouds", IEEE Transactions on Network and Service Management, 2020 (https://dais-ita.org/node/5882). Efficient management is also enabled through robust learning of SDC control policies. Researchers from Yale, in collaboration with Imperial, have proposed a lightweight architecture of neural networks, called Binarized Neural Networks, for learning locally at the data plane level of switches. By means of these networks, switches can make real-time inference decisions even when fragmented from their controller, therefore enhancing the reliability of the SDC architecture. Binarized Neural Networks have also been combined with Federated Learning to enable coordination among switches and boost performance while preserving the privacy of the distributed data. Results have been published in Q. Qin, K. Poularakis, K. Leung, L. Tassiulas, "Line-Speed and Scalable Intrusion Detection at the Network Edge via Federated Learning, IFIP Networking 2020 (https://dais-ita.org/node/5892). Related to efficient management of networked systems is the problem of network fragmentation prediction. Researchers from Yale, in collaboration with IBM U.S. IBM U.K. and ARL, have proposed a deep neural network approach for predicting in advanced network fragmentations. The network fragmentation prediction is based on prediction of path loss value at a next time instance from a given (time window) past sequence of path loss values. Predicting fragmentation allows mobile nodes to self-organize and run a distributed mobile ad hoc network (MANET) protocol in order to preserve traffic routing even after controller fragmentation occurs (related demo https://dais-ita.org/node/5423). To improve even further the use of Deep Neural Networks in tactical ad hoc networks where dynamic changes in topologies are the norm rather than the exceptions, researchers from Yale University, in collaboration with IBM U.S., have explored the use of graph attention networks (GATs) to provide generalisable mechanisms for performing inference over different network topologies, including topologies not seen during the training of these networks. This has been applied in particular on the task of network congestion prediction and results are given in "Generalizable and Interpretable Deep Learning in Networking", submitted to IFI Networking 2021 (https://dais-ita.org/node/5890). At the core of the use Machine Learning for efficient resource utilisation is data collection. Researchers from Imperial College have developed an efficient method for collecting and forecasting resource utilisation data in distributed systems. This is a communication-efficient data collection algorithm that employs an adaptive data-reduction mechanism based on Fourier transforms and truncation in the frequency domain. This algorithm can be integrated with a deep-learning-based forecasting model to improve inference speed. Results have been published in P. Pritz, D. Perez, K. Leung, "Fast-Fourier-Forecasting Resource Utilisation in Distributed Systems", ICCCN 2020 (https://dais-ita.org/node/5467). Coalition operations often involve systems with multiple self-interested stakeholders (e.g. services or users belonging to different domains). Resource allocation can be optimised also with respect to these self-interests. Researchers from Southampton have developed novel auction-based algorithms that incentivise truthful reporting in edge cloud resource allocation settings with multiple self-interested service consumers or providers. This novel approach combines reinforcement learning with incentive compatible mechanism design, ensuring

that an allocation mechanism remains truthful despite adapting its behaviour based on historical data. Results have been published in S. Stein et al. "Strategy proof Reinforcement Learning for Online Resource Allocation", AAMAS 2020 (https://dais-ita.org/node/4969). Finally, RL has been proven to be versatile for system control, including resource and network management functions of SDC, but SDC may become fragmented due to natural component failures and man-made tactical reasons. Researchers from Imperial College, Purdue University, Yale, IBM US, IBM UK have investigated enhancement and performance improvement of RL in presence of SDC fragmentation by ensemble or meta-learning techniques to speed up the learning process following the reconnection of fragmented SDC components. Further discussion will be continued at the Bootcamp.

Task 7.2: Federated Policy Learning and Management

The second main objective of Project 7 is the development of approaches for federated learning of both local and global policies, and federated policy management that enable the composition of policies learned at local parties. Researchers from Purdue University, in collaboration with Imperial College and IBM U.S., have proposed and evaluated a new concept of hybrid policy-based ensembles, in which the DL models learned from different domains are be combined into an ensemble by means of policies expressed over properties of the training datasets in each of the domains. Experimental results, based on policy-driven federation of DL models, have been published in D. Verma, E. Bertino, A. Russo, S. Calo, A. Singla, "Policy-based ensembles for multi domain operations, SPIE 2020 (https://dais-ita.org/node/5066). Policy learned locally by one party of a coalition may be used by other parties. Policy transfer across coalitions, presents challenges due to the different parties' domains and local constraints. Researchers from Purdue University, in collaboration with Imperial College and IBM U.S., have advanced the start-of-the-art of policy management in coalition settings by proposing a policy transfer framework that considers minimal sharing of data and supports policy adaptation to address policy conflicts and privacy issues related to data sharing constraints. Results have been published in A. Jabal, E. Bertino, J. Lobo, D. Verma, S. Calo and A. Russo, "FLAP - A Federated Learning Framework for Attribute-based Access Control Policies", AAAI 2020 AAAI 2020 Fall Symposium (https://dais-ita.org/node/5381). At the core of policy transfer is the fundamental problem of transfer learning: how models trained to learn policies in one source domain can be transferred to a related by different target domain in particular when there is limited access to data in the target domain. Addressing this problem would allow local a local coalition party to leverage models learned by other source parties on source datasets (i.e., source models) when learning their own model in the presence of limited data. Researchers from Purdue University, in collaboration with IBM U.S. have addressed this problem by proposing a generative adversarial networks (GANs) approach for creating a domain-invariant mapping of the source and target datasets. The approach has been extensively evaluated on network intrusion datasets, considering the cases when source and target datasets have the same feature spaces, and when source and target datasets have different feature spaces. Results have been published in A. Singla, E. Bertino, D. Verma, "Preparing Network Intrusion Detection Deep Learning Models with Minimal Data Using Adversarial Domain Adaptation", ASIACCS 2020 (https://dais-ita.org/node/5753). Researchers from Imperial College and Purdue University have also tackled the problem of policy transfer across coalitions as an instance of policy migration. They have explored the challenges raised when policies that are enforced in a particular source computational device, need to be migrated and be enforced in a different target device. A formal framework has been proposed to evaluate the appropriateness of

a policy migration and results are reported in J. Lobo, E. Bertino. A. Russo, "On Security Policy Migrations" ACM SACMAT 2020 (https://dais-ita.org/node/5137). Researchers from Purdue University, Imperial College and IBM U.S. have also continued extending their BBP19 work on learning attribute-based access control policies, which has now been published in A. Jabal. E. Bertino. J. Lobo. M. Law, A. Russo. S. Calo, D. Verma, "Polisma - A Framework for Learning Attribute-Based Access Control Policies", ESORICS 2020 (https://dais-ita.org/node/3424).

Project 8: Generative Policy Models for Coalitions

A combined summary of all three project 8 tasks is included below.

Task 8.1: Distributed Online Learning with Multiple Learners

Task 8.2: Agile Analytics Enabled by Decentralized Continuous Learning in Coalitions

Task 8.3: Cognitive Workflows: Goal Directed Distributed Analytics Using Semantic Vector Spaces

In year one, researchers at UCL, UMass, and IBM studied the problem of decentralized, online, and multi-task learning to enable fast and adaptive decision making in a dynamic coalition environment. Several key results were devised with provable regret bounds or other optimality guarantees. The results were published in top machine learning venues such as NeurIPS and ALT (https://dais-ita.org/node/5823, https://dais-ita.org/node/5900, https://dais-ita.org/node/5903) and some further work on cooperative stochastic bandits with heterogeneous agents is under submission (https://dais-ita.org/node/5887).

Researchers at IBM, Imperial, PSU, and Yale studied the enablement of efficient machine learning in coalition environments. Two techniques were considered, namely, federated learning and coreset. For federated learning, novel algorithms were devised to perform adaptive model pruning for efficient training and inference. The effectiveness of these algorithms was verified on real low-power computation platforms involving Raspberry Pis. In addition, acceleration algorithms were studied that show insights between the efficiency and robustness of gradient-based learning techniques. On the coreset front, progress was made on the integration of coreset (cardinality reduction), dimensionality reduction, and quantization methods. Further, an initial comparison between coreset and federated learning techniques was done. These findings provide fundamental understanding and practical algorithms on efficient machine learning in coalitions with dynamically changing situations over time. The work was published in venues including ICDCS (https://dais-ita.org/node/4677), NeurIPS workshop (https://dais-ita.org/node/5825), and ICML workshop (https://dais-ita.org/node/5744). Moreover, the team also made initial progress in the areas of continual learning (https://dais-ita.org/node/5809 and decision making on non-Markovian processes (https://dais-ita.org/node/5878).

Researchers at Cardiff, IBM, and Purdue showed how semantic vector representation of services can be used to perform efficient service discovery and workflow construction in a decentralized edge of network environment (https://dais-ita.org/node/5183) and demonstrated this in the context of the rapid dynamic reconfiguration of battlefield communications plans (https://dais-ita.org/node/5896). They were also the first to show how semantic vector representations of services generated by different coalition partners can be mapped between different semantic

vector spaces without the need to share the sensitive data needed to generate the semantic vector spaces. (https://dais-ita.org/node/5432)

Project 9: Agile Composition for Coalition Environments

The overall research context for Project 9 was on maintaining a reliable working relationship among partners in a coalition despite differences and attacks on cohesion and analytics by external and internal adversaries. We highlight below select activities and findings from each of the two tasks under this project.

Task 9.1: Interpretability of Neural Networks in Distributed & Contested Environments under Incomplete Trust

In BPP20 year 1, researchers from Cardiff University in collaboration with IBM US, IBM UK and ARL, advanced the state-of-the-art in explainable AI for systems that operate in real-time with multimodal data. This research addresses the problem of achieving coalition situational understanding (CSU) which involves the use of artificial intelligence (AI) based assets to assist human decision-makers, operating at or near the edge of the network, in responding to rapidly-changing events. CSU exploits data collected from sensors of multiple modalities, for example, visual and acoustic. A user will often need to rely on an AI asset created by a different coalition partner; therefore, the assets must be capable of generating explanations for their outputs to engender user trust. Moreover, the edge setting requires explanations to be as specific and targeted as possible. The DAIS team created a novel technique called Selective Audio Visual Relevance (SAVR) designed to support CSU in real-time, edge settings, by providing effective and resource-efficient explanations for AI-based activity detection from videos with audio soundtracks. The team demonstrated the applicability of SAVR as part of a real-time CSU 'dashboard' for monitoring events in an urban terrain setting. A version of SAVR is available as open source software in the form of a Jupyter notebook.

Additionally, in a paper that appears in the recently held NeurIPS 2020, researchers from the task empirically examined the efficacy of two primary forms of explainability that have emerged: explaining by annotating the relevant regions of the input data based on their relevance to the eventual prediction produced by the ML model, and explaining by referring to examples from the data that was used to train the model and thus affected the decision boundary. We performed a cross-analysis Amazon Mechanical Turk study comparing the popular state-of-the-art explanation methods to empirically determine which are better in explaining model decisions. The participants were asked to compare explanation methods across applications spanning image, text, audio, and sensory domains. Our findings indicate that explanation-by-example was preferred in all domains except text sentiment classification, where the LIME (local interpretable model-agnostic explanation) method of annotating input text was preferred. Our work highlighted qualitative aspects of employing the studied explainability methods and concluded with implications for researchers and engineers that seek to incorporate explanations into user-facing deployments.

Task 9.2: Network intelligence from negative ties

During Year 1 of BPP20, this task's research advanced an understanding of under-studied, negative, and conflictual, social networks (with 7 AFM2020 papers). New statistical, network techniques were developed to identify local network structure (i.e., network motifs) and applied to comparisons of multiple, negative and positive social networks. This new, statistical approach identifies unique "network signatures" in networks that can be used to distinguish potentially disruptive networks. One article was published in a Social Networks issue in Social Psychology Quarterly https://dais-ita.org/node/5472. A second was an AFM 2020 long paper and submitted for journal review https://dais-ita.org/node/5444 . A third was an AFM2020 short paper and was submitted to the INSNA conference https://dais-ita.org/node/5419 (Penn State and Cardiff).

Progress was made in addressing negative, problematic content on the social media site Reddit. A paper titled "Assessing temporal and spatial features in detecting disruptive users on Reddit" was accepted for presentation at The 10th Workshop on Social Network Analysis in Applications (SNAA 2020) co-located with the 2020 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2020) https://dais-ita.org/node/4866 . Another paper on Reddit COVID-19 misinformation communities and graphlets was submitted to a special issue of Online Social Networks and Media (Cardiff and Penn State).

Research advanced our knowledge regarding the ability to predict the spread of negative ties (e.g., curse words) on social media with the use of LSTM. One paper (presented at the NASN network conference) demonstrated the utility of Graph LSTM over other approaches to predict the diffusion of curse words on Twitter https://dais-ita.org/node/5877. Work also advanced in producing a Demo that showcases the ability to predict the spread of negative content on Reddit with the use of statistical and Graph LSTM (joint with IBM US, Penn State, Cardiff, and ARL).

Progress was made on the application of AI/ML and statistical network modelling to the examination of the negative ties of terrorist networks. An article was published in the Journal of Quantitative Criminology (Penn State and Cardiff) https://dais-ita.org/node/1501. Another paper (joint effort between IBM UK, IBM US, and Penn State) was published in the Proceedings of the SPIE-DCS 2020 https://dais-ita.org/node/5133.

Finally, work progressed regarding refining the concept of negative network ties in a paper accepted for SPIE 2021 https://dais-ita.org/node/5813 (IBM UK, Cardiff, Penn State, ARL). New developments proceeded on the implementation of the Human-Agent Knowledge Fusion (HAKF) concept in the Cogni-Sketch environment (Cardiff)-> https://dais-ita.org/cogni-sketch.

Project 10: Instinctive Analytics in a Coalition Environment

In this project we pursue the organization, integration and autonomy of both human and machine agents to fulfill coalition objectives pertaining to multi-domain scenarios with context and situational awareness.

Task 10.1: Coherence in Coalitions: understanding internal group behavior and dynamics in complex multi-domain environments

In task 1 we have focused on understanding the function and operation of coalition-based groups in terms of their coherence and ability to make effective decisions.

- Work on cultural influence in organisations (https://dais-ita.org/node/5406) has resulted on insights into how organisational structure can influence the spread of culture (i.e., beliefs that can be contagious) as a function of breadth and depth. This is a collaboration between Cardiff and ARL.
- Southampton, in collaboration with IBM UK and ARL, have further advanced work on competitive influence maximisation. This has involved a non-linear model for budget settings (https://dais-ita.org/node/5355), the effect of negative ties/influence in the diffusion process on signed network https://dais-ita.org/node/5356, and determination of how human strategies compare to theoretical optimal solutions https://dais-ita.org/node/5349.
- Cardiff in collaboration with Yale has developed computational insights into the evolution of identity fusion, which represents a personal alignment with a group. This concept is important because it offers the current best explanation as to why individuals can become empowered to act selflessly for a group (e.g., the devoted actor concept). A comprehensive manuscript has been submitted to Nature Scientific Reports https://dais-ita.org/node/5468.
- ARL in collaboration with IBM UK and Cardiff developed an agent-based model of collaborative problem solving and investigated the impact of team structure on the accuracy and speed of finding solutions to complex problems. A paper on this has been accepted at SPIE 2021. Simultaneously, researchers from ARL, Southampton and IBM UK investigated the same problem adopting the reinforcement learning framework in an attempt at developing a real-time platform guiding C2 structures in coalition operations.

Task 10.2: Learning and Inferencing in Neuro-Symbolic Hybrids for UncertaintyAware Human-Machine Situational Understanding

In task 2 we address the need to rapidly integrate machine analytic components in a way which (1) is aware of uncertainties; (2) exploits synergies; and (3) supports human decision makers. Our objective is to achieve a step change in free-flowing composition of uncertainty-aware human-agent and agent-agent information analytics. The first year of BPP 2020 has seen significant success in our neuro-symbolic artificial intelligence (AI) research.
Cardiff, UCLA and ARL developed AI architectures that can learn and understand complex events, enhancing the trust and coordination between human and machine needed to successfully complete battlefield missions. This work addresses the challenge of sharing relevant knowledge between coalition partners about complex events, i.e., compositions of primitive activities connected by known spatial and temporal relationships. For such events, he said, the training data available for machine learning is typically sparse. Two different approaches were developed by the team to enable learning at the neural layers by propagating gradients through the logic layer. The first, Neuroplex – published at SynSys 2020 (https://dais-ita.org/node/5382) – uses a neural surrogate for the symbolic layer. The second, DeepProbCEP – published at ICLP 2020 –

uses DeepProbLog to propagate the gradients (https://dais-ita.org/node/5340). Neuroplex was evaluated against pure deep learning methods over three types of complex events formed by a sequence of images, a sequence of sound clips and a nursing activity data set collected from motion capture, meditag and accelerometer sensors. The experiments and evaluation showed that Neuroplex is capable of learning to efficiently and effectively detect complex events, which cannot be handled by state-of-the-art neural network models. During the training, Neuroplex not only reduced data annotation requirements by one hundred times, but also significantly sped up the learning process for complex event detection by four times. Similarly, experiments on urban sound clips demonstrated over a two times improvement in complex event accuracy for DeepProbCEP against a two-stage neural network architecture.

Task 10.3: NSPL – A Neural-Symbolic Learning of Generative Policies in Coalition Environments

In task 3 we explore how to enable coalition systems and devices to operate with minimal human intervention in highly heterogeneous, and dynamic contexts whilst maintaining a level of security, to guarantee robust distributed analytics. Four topics have been addressed.

- Imperial in collaboration with IBM US, IBM UK and ARL, have advanced Hybrid Interpretable Learning From Noisy Raw Data (https://arxiv.org/abs/2012.05023). The underlying idea is to enable different deep learning models of parties of a coalition operation to extract features from the local surrounding environment and use their classifications as contextual information for a symbolic machine learning system capable of computing interpretable and transferable models (e.g, policy models) for classification.
- Scalable Symbolic Machine Learning in Uncertain Domains has been addressed by researchers from Imperial, in collaboration with Purdue. This has extended the symbolic machine learning system FastLAS developed in BBP18. FastLAS2 has features that make it scalable and effective in learning interpretable models in uncertain domains. This work is under review at IJCAI2021.
- Imperial have developed a formal framework and theoretical foundations for exact computation of non-monotonic semantics of logic programming, typically used in common-sense reasoning, in continuous vector space. This is the first ever method for approximating non monotonic inference in continuous space. This is very significant as it provides a natural way for injecting semantics and prior knowledge in deep learning architectures to enable their training from unstructured data in the absence of labels.
- Purdue have developed a framework to constraint reinforcement learning (RL) systems by symbolic policies (https://dais-ita.org/node/5408). In this work, we have designed Jarvis, a novel constrained RL framework for simulated RL environment in order to find the optimal device actions according to the user's goals, such as saving energy and minimizing cost. By observing the specific IoT environment, Jarvis dynamically builds a simulated environment in terms of device states and actions. An agent, constrained by security policies, can traverse the simulated environment in multiple episodes of specific time periods and find the optimal safe actions in terms of functionality requirements provided by the user. Our analysis shows that Jarvis provides significant advantages over

normal device behavior in terms of functionality and over general unconstrained RL frameworks in terms of safety and security.

Additional details

Additional details about the BPP20 period plans and goals can be found at https://dais-ita.org/BPP20 A summary of accomplishments and publications for the BPP20 and the whole of the program can be found at https://dais-ita.org/project_qprs and also https://dais-ita.org/qpr_archives

In the above summary of accomplishments, we have chosen to highlight certain key publications for each of the four projects. The next section of this report is a summary of publications occurring during the BPP20 year one period. Please refer to the Science Library (http://sl.dais-ita.org/science-library) for all publications during this period and the whole of the DAIS ITA program so far.

Summary of Publications

Introduction

The data and statistics in this document are summarised from the publicly available DAIS ITA Science Library [http://sl.dais-ita.org/science-library]. In some cases there are publications listed in this document that are not yet published to Science Library, e.g. because they have not yet been published in their external conference or journal venue (See Appendix 1). Eventually, all successfully accepted papers will be published to Science Library, even after the BPP20 period has concluded. The papers counted in this summary document will tally closely to those reported throughout the BPP20 period in the Quarterly Progress Reports (QPRs) but they are reported separately here as a single succinct summary as the repeating periodic nature of the QPR document means that papers are regularly reported in numerous quarters as they progress from submitted to accepted and published, plus each QPR only deals with the papers relevant to that quarter. The purpose of this stand-alone summary document is therefore to provide a short and simple overall summary for the period, backed up by detailed statistics on science library and more detailed contextual reporting in each of the QPRs.

Overall Summary

The BPP20 year one period (12 months) saw 17 journal papers published at external peer reviewed journals, and 56 conference papers published at peer reviewed external conferences or workshops. This compares to 6 journal and 103 conference/workshop papers in the IPP period (16 months), and 36 journal papers along with 182 conference/workshop papers in the BPP18 period (24 months). It should be noted that papers are "counted" in the period that they are published, and this may not be the same period as when the effort was spent in writing the paper, especially for journal papers which may require more than 6 months for publication. With this important caveat in mind, the publication data is broken down by project in the table below, with citation data[1] for each project and period. The IPP and BPP18 data is included purely for comparative purposes.

[1] Citation data was taken manually from Google Scholar for each DAIS paper on 11th Mar 2021.

Period		External Journal	External Conference	Citations
BPP20 (Year 1) Jan-20 to Jan-21	BPP20 P7	8	8	14[2]
	BPP20 P8	5	16	39
	BPP20 P9	3	3	7
	BPP20 P10	1	12	9
	Cross-project			
	Total			
BPP18 Jan-18 to Jan-20	BPP18 P1	13	28	459
	BPP18 P2	4	31	105
	BPP18 P3	4	25	267
	BPP18 P4	9	25	100
	BPP18 P5	3	47	353
	BPP18 P6	3	16	47
	Cross-project	0	10	70
	Total	**36**	**182**	**1491**
IPP Sep 2016 - Jan-2018	IPP P1	1	10	80
	IPP P2	0	23	148
	IPP P3	3	8	299
	IPP P4	0	13	31
	IPP P5	0	7	42
	IPP P6	2	18	297
	Cross-project	0	11	44
	Total	**6**	**90**	**941**

Table 1: Overall confirmed external publication summary (by project)

[2] It should be noted that the citation data for the BPP20 period is substantially lower than BPP18 because the snapshot is after 1 year maximum (in many cases only a few months) and it takes time for citations to accrue. The BPP18 snapshot was taken after a 2 year period, hence allowing at least double the time for citations when compared to BPP20.

Period		Journal[3]	External Conference
BPP20 Year 1 Jan-20 to Jan-21	BPP20 P7	2	15
	BPP20 P8	2	10
	BPP20 P9	2	3
	BPP20 P10	1	8
	Total	**7**	**36**

Table 2: Possible external publication summary (by project), including papers not yet confirmed as published, and found in Google Scholar but not reported in CENSE. These are likely maximum values, i.e. if all submitted papers are accepted and published.

[3] We estimate the type of paper, but sometimes it is not specified, and in other cases the venue may change if the paper is rejected from the first choice venue.