

Enabling Collaboration Across the DAIS ITA Research Program

Graham White
IBM Research, UK
gwhite@uk.ibm.com

Tien Pham
ARL

tien.pham1.civ@mail.mil

Simon Pierson
DSTL
sjpierson@mail.dstl.gov.uk

Maroun Touma
IBM Research, US
touma@us.ibm.com

Dave Braines
IBM Research, UK
dave_braines@uk.ibm.com

Abstract: The DAIS-ITA program is starting to increase in pace in terms of the level of experimentation being done to support the research. Demonstrations convey the work conducted and set it in a meaningful way against the context of the program. In this paper we seek to define “enablers” as a set of capabilities and resources to directly support collaboration, experimentation and demonstration across the DAIS ITA program. In order to reduce individual burden across projects and facilitate collaboration, some enablers are already available. This paper sets out these enablers and explores extensions to them. In addition, further enablers are possible that would extend the benefit into new areas. These are also set out and explored. Further work is expected in order to make these suggestions more concrete and to explore new areas beyond the scope of this paper.

1. INTRODUCTION

This paper sets out current enablers available to the researchers in the DAIS-ITA¹ program and identifies areas that could be developed during the course of the program. The intentions are to:

- facilitate working across projects as well as potentially outside of DAIS-ITA;
- reduce the risk of duplication of work across projects.

With the definition of current enablers and possible enabler areas, the areas to concentrate on initially are clearer and discussion can be promoted about which enablers to develop and what form these should take. Subsequent papers

may narrow down the options for which areas to develop.

2. DEFINING ENABLEMENT

The term enabler could be seen by many to have different definitions. In the context of this paper and the wider DAIS-ITA program, enabler is simply used to refer to any process, methodology or technology that aids research and can be made available for use across the program. This is typically for, but not limited to, experimentation and demonstration. It may consist of, but not be limited to, scenarios, data, software, infrastructure and working methods.

Enablers encompass a set of capabilities and resources that are shared by all participants in the program across multiple geographies and organizational affiliations. They support the creation and dissemination of new assets created by the DAIS-ITA researchers with the specific intents of accelerating the pace of our research and facilitating peer reviews.

3. CURRENT ENABLEMENTS

There is a set of enablers available for the DAIS-ITA program. These are not insignificant and have been built during the course of the research program or during the previous NIS-ITA² program. Use and re-use of these assets is encouraged and hence, we list them here to provide a foundation on which to build:

- Infrastructure and Software Defined Networking (SDN)
- DAIS-ITA Scenario

1 <https://dais-ita.org>

2 <http://nis-ita.org/>

- Repository for Papers and Collaboration

The following sub-sections define each of these in more detail.

3.1 INFRASTRUCTURE AND SDN

The most advanced and complete enablement is the area of infrastructure and SDN. This area covers the supply, virtualisation and connection of computer hardware, storage and networking including software defined networking. The SDN area is particularly advanced having incorporated the use of CORE³ and EMANE⁴ for the purposes of network emulation at the various layers in the network stack.

It is commonplace to find enablement facilities consisting of a set of computers upon which experiments can be run. These are linked with networking in order to create small clusters of machines. These clusters have been extended with SDN capabilities and include network emulation. This type of facility has been made available at DSTL, ARL and IBM. The facilities at DSTL [1] and ARL[2] are still in use. The facilities at IBM are currently shut down and not in use but could be resurrected if necessary.

The level of hardware virtualisation available is currently at the hypervisor level. This is either in the form of bare metal (such as VMWare ESXi) or operating system (such as KVM). There is currently no facility for the use of containers or for containerisation.

3.2 DAIS-ITA SCENARIO

A scenario has been produced by DSTL and ARL [3] that sets out a fictitious coalition engagement taking into account what is considered to be the future coalition battlespace around which the DAIS-ITA research should be based. The scenario has been produced to align the DAIS-ITA projects around a central theme that is considered realistic for experimentation and demonstration. The scenario has an overarching description of the broad phases of the mission and is designed to be extended via vignettes. These extensions are designed to be more detailed and specific to a particular situation or to enable the research focus of a specific project to be better accommodated.

Further vignettes may be produced that provide a wider range of specific situations.

3.3 REPOSITORY FOR PAPERS AND COLLABORATION

Collaboration within the program is via the CENSE (Collaboration Environment for Scientific Exploration) platform. This provides many capabilities⁵ to researchers including the ability to publish materials and communicate with each other.

The CENSE environment is centred mostly around what may be considered traditional office style collaboration. However, this environment does not provide a central repository for code, collaboration around code and may also not be suitable for large volumes of data such as that required in machine learning based projects.

4. POSSIBLE NEW ENABLEMENT AREAS

There are a variety of enablement areas that could be added or improved. These should be focused towards helping all of the the DAIS-ITA projects across the program by providing centralised support. The purpose is to prevent repetition between projects and to provide overarching themes and consistency to the program.

The following list, which is unlikely to be exhaustive, are some of the options for creating new enablements:

- Data and pointers to open source repositories for training and testing AI models
- Code, shared libraries and commonly used AI models.
- programming tools and commonly used frameworks for building AI solutions
- training/testing environments
- user interface tools

These are in addition to those discussed throughout section 3:

- containerisation (section 3.1), container orchestration and container run-times
- scenario vignettes (section 3.2)
- data repository (section 3.3)
- code repository (section 3.3)

The following sub-sections explore the first list of possible enablement areas in more depth. The feasibility of each is not yet well understood. Hence, some more work needs to take place to identify which of these options, if any, should be taken forward. It is, however, important to recognise that any recommendation is unlikely to be suitable

3 <https://www.nrl.navy.mil/itd/ncs/products/core>

4 <https://www.nrl.navy.mil/itd/ncs/products/emane>

5 CENSE is built on Drupal <https://www.drupal.org/>

across the board so some amount of flexibility will be required when following them.

4.1 DATA

The analytics focus of DAIS-ITA and research into subject areas such as machine learning require data for both the fundamental research to take place but also experimentation and demonstration. The program currently offers no form of data (particularly data that might match the DAIS-ITA scenario [2]) that can be used for research e.g. as used for training and testing machine learning models. A centralised data set or collection of data sets would offer consistency throughout the program both from a technical point of view as well as a scenario point of view. If this area is explored further the data would need to be suitable across a wide range of projects and research areas within the DAIS-ITA program that will likely require multiple inter-related data sets.

4.2 CODE

There may be some commonality between the DAIS-ITA projects in terms of code re-use. This may be in many forms such as re-usable patterns of code or libraries that can be shared. In order for this to work best a common language or set of languages would need to be proposed, but not be restrictive or exhaustive so as not to stifle innovation.

Enablers in this area could range from organising and facilitating code re-use to offering assistance and help with writing code in a particular language to solve a particular problem along with links to documentation or testing approaches for common domains or techniques. Many of these resources could be generated from deep dive workshops into specific cross-project topics at the regular boot camp meetings, for example.

Further, shared and common code could also include specific machine learning libraries and models. These could be used in, for example, video, audio or text processing applications. Such models can be used to build more complex scenario vignettes for autonomous systems and specific experiments that support the DAIS-ITA research across multiple projects.

There is no central DAIS-ITA internal code repository since the purpose of the program is fundamental research, however a public open source organisation has been created on github⁶ to

facilitate sharing of code with both the DAIS-ITA community and the wider world when it is appropriate to do so.

4.3 PROGRAMMING TOOLS

To offer the best chance of portability and compatibility of any code arising from the DAIS-ITA research, it may be beneficial to standardise towards some specific tool kits. These tool kits are likely to provide a specialist feature or function within a wider piece of code associated with a project. There are, for example, a wide range of machine learning libraries and tools available that are offered as a framework for this task. Five common examples of these are Caffe, Scikit-Learn, Tensorflow, Theano and Torch.

Enablers in this area could range from: creating simple recommendations towards certain tools; or guidelines for configuring and using particular tools; to: creating code samples and boilerplates; or offering assistance and help with a particular toolkit. At a higher level, common patterns of practices for rapidly building entry level capabilities, for example, CNNs (Convolutional Neural Networks), or how to use grid search techniques for hyperparameter tuning would also be potentially valuable and can often be found in the public domain already.

At a minimum, these enablers should include a set of tutorials on common tools as well as examples on how they are being used in the program. The primary goal would be to create awareness of the available tools and capabilities that are used by any one of the DAIS-ITA projects and encourage replication across other projects.

4.4 TRAINING/TESTING ENVIRONMENTS

Traditional code can be tested using small amounts of data and compute. However, machine learning models generally require a large amount of data and compute resource during training and testing. Specialised GPU hardware is also very beneficial for speeding up this process. At present there is no advice or resource offered to help with this aspect of DAIS-ITA research.

Enablers in this area could range from recommendations for approaches to take and resources to use through to providing a centralised or cloud infrastructure on which models can be

⁶ See <https://github.com/dais-ita/>

trained that would incorporate any required specialised hardware.

The DAIS-ITA requirement in this area may also be quite specific. The distributed nature of the analytics being performed might require a particular configuration of heterogeneous network of machines in order to truly train and test models (potentially distributed models) in the real-world scenarios being researched.

4.5 USER INTERFACE TOOLS

Experiments do not require the level of presentability needed for a demonstration. Demonstrations, by nature, are designed to be shown to an audience and should be presented in an attractive and understandable format in order to convey the potential impact of the research and to start more detailed conversations, especially with military or government stakeholders. Today, many demonstrations are given using web browser based technologies since these offer an easy path to production and a well understood level of effort and maintenance. However, not all demonstrations can or should be given in this manner.

Enablements in this area could range from the recommendation of methodologies or tooling for different platforms. Examples might include particular web frameworks for browser based demonstrations or be broader in their scope such as multi-modal, haptic or conversational interfaces.

5. CONCLUSIONS

This paper has created a definition for enablement and offered a review of currently available enablements alongside proposing possible areas for developing new enablements within the DAIS-ITA program.

Some enablements are already in place to support experimentation such as: infrastructure; software defined networking; the DAIS-ITA scenario; and the CENSE platform. These have been explored and could be extended to include: containerisation; scenario vignettes; and repositories for code and/or data. These existing enablements may also partially support demonstrations.

In addition to extending current enablements, possible new areas of enablement have been explored including but not limited to: creating training/testing/development data; creating re-usable code; recommending or creating templates for particular coding frameworks and tools; maintaining training/testing/development environment(s); and facilitating the creation of user interfaces for demonstration.

We do not, at this stage, seek to make recommendations for which enablements should be adopted but simply set out the positioning and explore the options available. Further work will be required within the DAIS-ITA program to research which are the most beneficial such that recommendations can be made to follow up on those that have the most utility.

6. ACKNOWLEDGEMENTS

This research was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defence under Agreement Number W911NF-16-3-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

7. REFERENCES

- [1] Christopher Gibson, Rosie Lickorish, "Experimentation Environment Study", TIN052 Research Report, 2018.
- [2] K. Marcus, M. Touma and F. Bergamaschi, "Network Science collaborative experimentation methods and tools to accelerate network science innovation within the ITA and NS-CTA," *MILCOM 2015 - 2015 IEEE Military Communications Conference*, Tampa, FL, 2015, pp. 1385-1390.
- [3] "DAIS ITA Scenario" presented by Maj S J Pierson R SIGNALS at the DAIS-ITA Bootcamp, June 5th 2018.