University of Glasgow | School of Computing Science

Session types in the wild

A. Laura Voinea

Primary Supervisor: Dr. Simon Gay
Secondary Supervisor: Dr. Wim Wanderbaude

1st Year PhD Report

June 12, 2016

Contents

1 Introduction

Nowadays, distributed systems are ubiquitous and communication is an important feature and reason for its success. Communication-centred programming has proven to be one of the most successful attempts to replace shared memory for building concurrent, distributed systems. Communication is easier to reason about and scales well as opposed to shared memory, making it a more suitable approach for systems where scalability is a must, as in the case of multi-core programming, service-oriented applications or cloud computing[1].

Communication is usually standardised via protocols that specify the possible interactions between the communicating parties in a specific order. Mainstream programming languages fail to adequately support the development of communication-centred software. Thus, implementations of communication behaviours are based on informal protocol specification and thus informal verification. As a result they are prone to errors such as communication mismatch, when the message sent by one party is not expected by the other party, or deadlock, when two parties are waiting for a message from each other causing the system to block[1]. To allow formal protocol specification within the programming language session types have been devised. Session types describe communication by specifying the type and direction of messages exchanged between two parties[6]. Programmers can express a protocol specification as a session type, which can guarantee, within the scope where the session type applies, that communications will always match and the system will never deadlock. The main goal of the ABCD project[1] is to improve the practice of software development for concurrent and distributed systems through the use of session types. This is to be accomplished through built-in language support for protocol codification in existing languages such as Java or Python, in new languages such as Links[1], inter-language interoperability via session types, and through adapting interactive development environments and modelling techniques to support session types. Logical and automata foundations of session types will be further developed to express a wider class of behaviour and, as need arises, to support the former. Empirical studies to assess methodologies and tools are to be carried out, with results being used to improve language and tool design and implementation.

1.1 Thesis Outline

Eaten by apple

1.2 Thesis Statement

Session types are a useful addition to the syntax and semantics of modern languages. Programming with session types and the constraints that this comes with i.e. linearity can be understood and used by real world programmers. Moreover session types can help programmers understand the problem they are tying to solve with more ease and structure their code better. Session typed languages provide useful additional safeguards and diagnostic information that lead to a system with the expected behaviour with less effort.

2 Litrerature review

2.1 Session types, a short history

Session types have been introduced by Honda[12], and further extended by Takeuchi et al.[19], Honda et al.[13], and others.

A session type describes the types of individual messages and the state transitions of the protocol i.e. the allowed sequences of messages. A session type is associated with a communication channel upon creation. Type-checking may be used to verify processes compliance to the session type system and establish properties such as deadlock and race freedom, or session fidelity between two parties.

```
% global protocol Bookseller(role Buyer1, role Buyer2, role Seller) {
%       price_request(String) from Buyer1 to Seller;
%       choice at Seller {
%               price_response(int) from Seller to Buyer1;
%               quote(int) from Buyer1 to Buyer2;
%               choice at Buyer2 {
%                       agree(String) from Buyer2 to Buyer1, Seller;
%               } or {
%                       quit(String) from Buyer2 to Buyer1, Seller;
%               }
%       } or {
%               outOfStock() from Seller to Buyer1, Buyer2;
%       }
% }
%
```

Session types have thus far been the subject of a wide body of work aimed at extending its theoretical foundations as well as developing new tools and techniques. For example, session types have been augmented with subtyping polymorphism to enable protocols to describe richer behaviours[10]. They have been extended to ensure the progress property and deadlock freedom [7]. Session type theory has also been extended from binary session types to multiparty session types to support communication instances with more than two participants while still guaranteeing the absence of deadlock[14]. In other work [5], global session types have been used to detect choreographies that can be realized in the context of web services. Session types have been successfully applied in functional programming languages [20], [4], object-oriented languages like Java [15], [11], low-level programming languages like C in [18], dynamically-typed languages like Python[17] or Erlang [16], or in the operating systems context with the Sing# language [9].

From the literature surveyed, two main areas of interest will be presented bellow, namely typestate and language usability and evaluation.

2.2 Typestate

What typestate is

Why is it useful to session types

Plural, Plaid, Fugue, Mungo

Eaten by apple

2.3 Language usability and evaluation

Why is evaluation important in the context of programming languages?

Workshop on Evaluation and Usability of Programming Languages and Tools (PLATEAU) at SPLASH

3 Research Activities

3.1 Mungo

Mungo[1] is a Java front-end tool, developed by Dr. Dimitrios Kouzapas at the University of Glasgow used to statically check the order of method calls i.e. its typestate. It is implemented using the JastAdd framework[2]. A protocol or session type is represented as a separate typestate file, associated with a Java class. The protocol definition is described as a sequence of method calls, the order of which determines the validity of the protocol.

Mungo checks that the object instantiating the class performs method calls as defined by its typestate. If the protocol is not violated, standard .java files are produced for every .mungo file in the package. The resulting code can then be ran as any standard Java code.

Typestate is a refinement of the concept of a type in programming [7, 8]. Typestate is the constraining of a sequence of calls. In programming certain methods can sometimes only be called in certain states. Typestate checks what method is available to be called in which state and what the subsequence state of each method is. Furthermore Typestate can check at compile time that specifications are followed. Through Typestate analysis, it is possible to track the degree of initialisation of variables, guaranteeing that operations would never be applied on improperly initialised data.

The following work was undertaken on the Mungo tool:

- the tool has been moved from the Java 1.4 compiler to the Java 1.8 compiler and adapted to work with the new framework.(joint work with Dr. Dimitrios Kouzapas)
- moving the tool to a new compiler framework was a good opportunity for some refactoring(such as getting rid of dead code or method extraction).
- the tool has been extended to support Java enumerations.(joint work with Dr. Dimitrios Kouzapas)
- updated repositories
- work has been undertaken to support generic types

[1]http://www.dcs.gla.ac.uk/research/mungo/
[2]http://jastadd.org/web/extendj/

- work has been undertaken to typecheck exceptions, in a simple form at the moment

Areas of future work:

- aliasing
- typecheck generic types
- support annotations

3.2 StMungo

StMungo (Scribble to Mungo)[3] is a Java-based tool, developed by Dr. Ornela Dardha at the University of Glasgow, that translates Scribble[4] local protocol specifications into Mungo specifications.

Scribble is a language that can describe application-level protocols among communicating systems. A protocol represents an agreement on how participating entities interact with each other.

After the Scribble protocol is translated to a Mungo specification, Mungo**??** is used to generate a Java implementation for the protocol. This tool allows an easy transition from a Scribble global protocol definition to working Java implementation. We start by specifying distributed multiparty protocol in Scribble. We can then use the Scribble tool chain to validate and project the global protocol into a local one describing the interactions from the point of view of a specific participant. For every Scribble local protocol, StMungo will produce .mungo files containing: a typestate specification describing the local protocol as a sequence of method calls, an API for the participant implementing the typestate methods and a main class skeleton calling the methods in the typestate.

To improve this tool various extensions, refactoring has been undertaken:

- extended the tool to translate messages with no payload i.e.
  ```
  message_operator ()
  ```
- extended the tool to translate messages with multiple payload i.e.
  ```
  message_operator ( payload_type1 , ... , payload_typen )
  ```
- extended the tool to translate messages without a message signature i.e.
  ```
  ( payload_type1 , ... , payload_typen )
  ```
- extended the tool to translate messages with annotated payloads i.e.
  ```
  message_operator ( annotation:payload_type)
  ```
- various small improvements to allow most translations to run without having to be edited by a human

[3]http://www.dcs.gla.ac.uk/research/mungo/
[4]http://www.scribble.org

- various improvements allowing the tool to crash gracefully
- adapted the tool to work with multiple versions of scribble specification
- improved the tool by implementing support for special cases of recursions nested in choice structures A simple example of a problematic scribble specification is:

  ```
  global protocol Example(role S, role C) {
  choice at C{
          rcpt(String) from C to S;
      } or {
          msg(String) from C to S;
          rec loop {
                  subject(String) from C to S;
                  continue loop;
              }
          }
      }
  }
  ```

- improved the tool by implementing support for special cases of nested choice inside a recursion

  A simple example of a problematic scribble specification is:

  ```
  global protocol ProtocolName(role S, role C) {
    command(String) from C to S;
    rec overall {
    choice at S
        {
        ok(String) from S to C;
                        choice at S {end(String) from S to C;}
                        or {
                        sum(String) from S to C;
                        }
                        message(String) from S to C;
        } or {    error(String) from S to C;      }
        continue overall;
  }
  }
  ```

- extending the tool to translate to support inlined protocols and sub-protocols(ongoing)
- kept it up to date with changes in Mungo
- refactoring(such as method extraction or getting rid of dead code) to keep everything simple
- regression testing to find any new bugs introduced

Areas of future work:

- test harness
- extension to translate more complex constructs such as interruptible or parallel
- keeping it up to date with changes in Scribble and Mungo

3.3 Usecases

To better understand the expressive power of current session type technology together with any limitations that may need to be addressed, the current use case repository[5] was surveyed as a first step. As a second step, new real-world examples were sought. From the various protocols looked after, representations were attempted for two, Paxos and the File Transfer Protocol(FTP).

The first protocol chosen as a usecase was Paxos. Paxos is a protocol for solving consensus in a network of unreliable processes. It ensures that a single value among the proposed values can be chosen. It assumes an asynchronous, non-Byzantine model.[**?**]

Two major advantages of Paxos are that it is provably correct in asynchronous networks that eventually become synchronous and it does not block if a majority of participants are available.Furthermore it has provably minimal message delays in the best case. Despite it's reputation of being difficult to understand & implement it is widely, a couple of examples would be Google in Chubby, Yahoo use something based on it in ZooKeeper. The protocol comes with three roles and a two-phase approach. A proposer responsible for initiating the protocol, that handles client requests and proposes values to be chosen. An acceptor that responds to messages from proposers by either rejecting them or agreeing in principle and making a promise about the proposals it will accept in the future. An a listener or learner, who wants to know which value was chosen. Each Paxos server can act as any or all 3 roles.

Some representations of the algorithm have been attempted using Scribble, combined with StMungo and Mungo to give a working Java code. However in trying to represent it some shortcomings of the toolset became apparent:

- representing broadcasting
- representing quorum/a majority
- representing express the dynamic aspects such as processes failing, restarting
- express multiple instances of the protocol

The work on Paxos has been presented during the last ABCD meeting in January 2016. Work on a better Paxos representation has been paused for the moment, other tasks taking priority such as improving the tools.

The second protocol chosen was the File Transfer Protocol described by Request for Comments(RFC): 959[6]. FTP is a standard network protocol for transferring files between a client and server on a network. FTP is an unusual protocol in that it utilizes two ports, a data port and a command(control) port. FTP may run in active or passive mode, which determines how the data connection is established. In both cases, the client creates a TCP control connection from a random, usually an unprivileged, port number to the FTP server command port 21. Some representations of the algorithm have been attempted using Scribble, combined with StMungo and Mungo to give a working Java code. However in trying to represent it some shortcomings of StMungo became apparent. Hence, work on an FTP representation has been paused to improve StMungo and Mungo, to allow a better representation.

[5]https://github.com/epsrc-abcd/session-types-use-cases

[6]https://tools.ietf.org/html/rfc959

Work on both usecases is planned to be restarted in the autumn, and make the base for a paper later on in the year.

3.4 User study

New programming language constructs are more often than not introduced without first exploring how well suited their are for their purpose or how they would be used in the real world. While proving they solve the problem is a good thing, checking how well they solve it would be nice. To explore how well programmers would interact with session types an initial study has been devised. This can be found in appendix **??**. It is intended to explore whether programmers can learn what session types are, identify how they should be used, and reason about session typed code correctly. At this stage there are little constraints on the type of participant. Since one of the claims behind session types is that they make reasoning about communication easier, participants with less programming experience are welcome. It would be interesting to see if there are any differences and what differences there are between participants of different backgrounds. The only requirement in this case being a working knowledge of Java.

Difficulties: choosing meaningful for the user to carry out, deciding what constitutes a good understanding of session types.

4 Other Activities

As part of the first year of my PhD, various additional activities have been undertaken, such as training courses(detailed in appendix**??**), ABCD group meetings of various sizes, seminars and talks(e.g. Tht Scottish programming language seminar series, FATA seminars) which improved my knowledge of the field and gave me some insight of the exciting research ongoing, as well as my own ignorance. Some notable events attended were the BETTY (Behavioural Types for Reliable Large-Scale Software Systems)[7] meeting in March, and Wadlerfest[8], held in celebration of Philip Wadler's 60th birthday, or LFCS30[9], to celebrate the 30th anniversary of the founding of the LFCS.

5 Future Work

Some of the future work planned has been highlighted throughout this report. The activities planed for next year fall into several interconnected themes:

- Language usability and evaluation. As mentioned above in 3.4 a user study is in plan for the near future, its results to be analysed and written up and submitted to PLATEAU 2016[10]. A second more focused study is also in plan for the later part of the year.

[7] http://www.behavioural-types.eu/meetings/wg-mc-meetings-17th-18th-march-2016-in-malta

[8] http://events.inf.ed.ac.uk/wf2016/

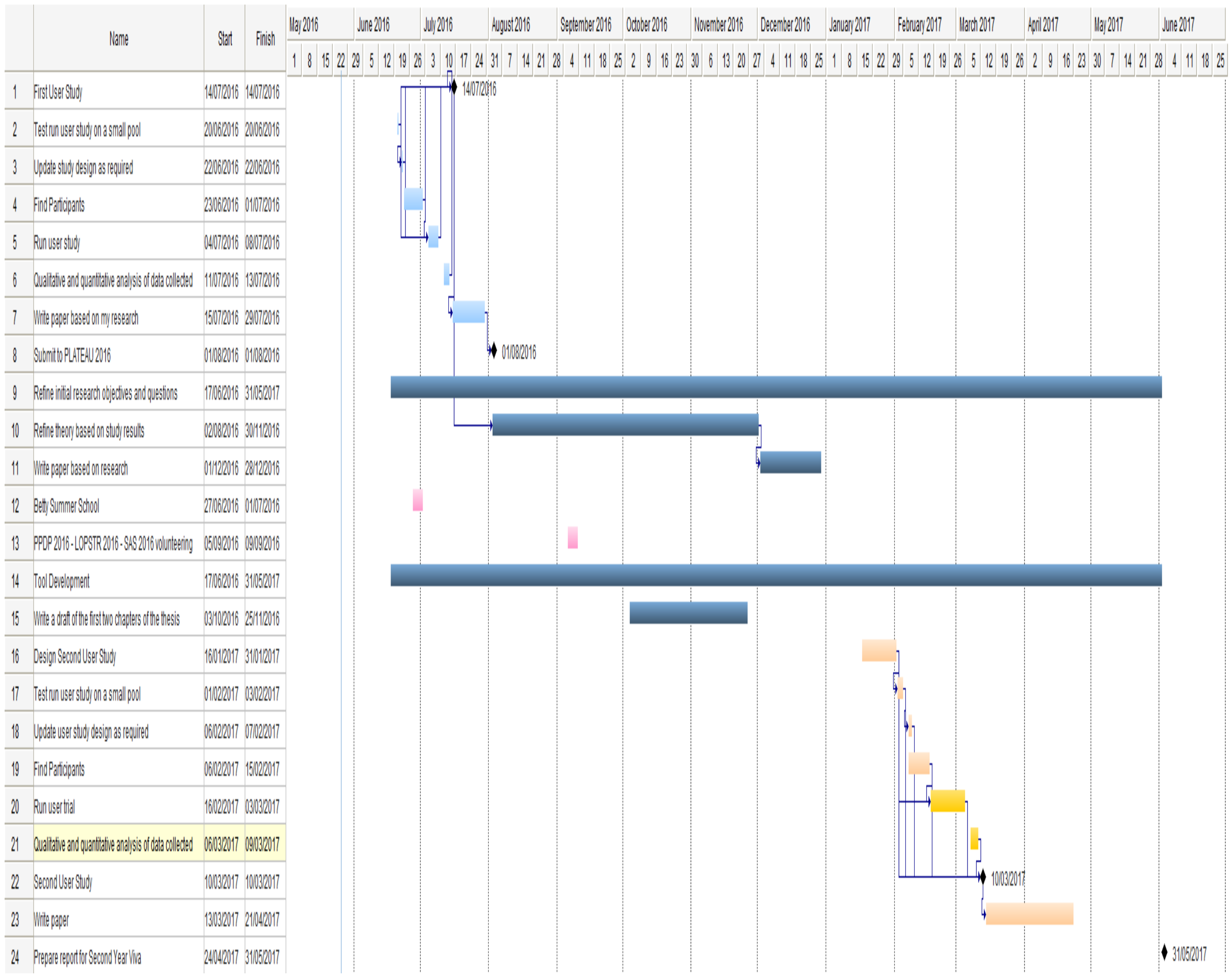
[9] http://events.inf.ed.ac.uk/lfcs30/

[10] http://2016.splashcon.org/track/plateau2016

- Tool development. Some of the work that is to be done has already been discussed earlier, however it is expected that the need for various different improvements will arise. Thus this will be an ongoing activity through the year.
- Training and development. This category contains any training courses required by the school as well as any academic events to be attended. Over the summer, I will attend BETTY (Behavioural Types for Reliable Large-Scale Software Systems) in Limassol, Cyprus. The summer school will cover closely related areas of interest such as multiparty session types, linear logic, Practical programming with session types. In September I will help with the PPDP, LOPSTR and SAS conferences in Edinburgh.
- Refining initial research objectives and questions. This is really the miscellaneous category, containing work to be continued such as the usecases**??**, and any new ideas that will arise from the user study.

For a detailed plan, a Ganntt chart is available in appendix A.

A Second Year Plan



	Name	Start	Finish
1	First User Study	14/07/2016	14/07/2016
2	Test run user study on a small pool	20/06/2016	20/06/2016
3	Update study design as required	22/06/2016	22/06/2016
4	Find Participants	23/06/2016	01/07/2016
5	Run user study	04/07/2016	08/07/2016
6	Qualitative and quantitative analysis of data collected	11/07/2016	13/07/2016
7	Write paper based on my research	15/07/2016	29/07/2016
8	Submit to PLATEAU 2016	01/08/2016	01/08/2016
9	Refine initial research objectives and questions	17/06/2016	31/05/2017
10	Refine theory based on study results	02/08/2016	30/11/2016
11	Write paper based on research	01/12/2016	28/12/2016
12	Betty Summer School	27/06/2016	01/07/2016
13	PPDP 2016 - LOPSTR 2016 - SAS 2016 volunteering	05/09/2016	09/09/2016
14	Tool Development	17/06/2016	31/05/2017
15	Write a draft of the first two chapters of the thesis	03/10/2016	25/11/2016
16	Design Second User Study	16/01/2017	31/01/2017
17	Test run user study on a small pool	01/02/2017	03/02/2017
18	Update user study design as required	06/02/2017	07/02/2017
19	Find Participants	06/02/2017	15/02/2017
20	Run user trial	16/02/2017	03/03/2017
21	Qualitative and quantitative analysis of data collected	06/03/2017	09/03/2017
22	Second User Study	10/03/2017	10/03/2017
23	Write paper	13/03/2017	21/04/2017
24	Prepare report for Second Year Viva	24/04/2017	31/05/2017

B User Study Plan

1 Experimental Description

The next part of this document is concerned with what will be evaluated and how. The following sections will define the methodology and approach that will be taken.

1.1 Tasks to complete

The participants will be given three main tasks to complete with various subtasks.

1. Iterator example in Mungo. Code provided. Spot and correct the error. Add the remove operation.
2. SMTP[1] example in Scribble and Mungo. Code provided. StMungo introduced as well. Spot and correct the error one in scribble specification, second one in Mungo specification.
3. Implement a small example based on an informal specification. Buyer seller would be a good candidate here.

1.2 Variable/Experimental conditions

Independent variables:

- Programming expertise

  Dependent variables to be measured:

- time to complete each task
- how many times the code is compiled
- how many bugs/errors
- is the end result correct

1.3 Data Collection/Measurements/Observations

The purpose of this study is exploratory. Dependent variables to be measured:

- time to complete each task
- how many times the code is compiled
- how many bugs/errors are encountered in the process
- is the end result correct
- participants' opinion about the tools/session types

1.4 Design

Considering that the aim of this experiment is exploratory and the learning effect (performance improving with experience, even over very short periods of time) is of little concern in this case a within subjects design, every participant perform every task under every condition, will be used.

[1] http://www.ietf.org/rfc/rfc2821.txt

1.5 Participants/Subjects/Users (Controlled Variable)

The only requirement for participants is that they have some Java experience. Factors like level of programming proficiency, background, or upfront knowledge of behavioural types or session types will be taken into consideration.

1.6 Experimental location

The experiment will be carried out in Sir Alwyn Williams building, room F112. This office has been chosen as it allows a higher degree of control over the environment, the programming one in particular.

1.7 Experimental schedule

This section outlines all of the parts of the experiment and a rough estimate how long each will take. After completing all tasks, anticipated to take roughly 45 minutes, the participants will be asked to answer a short survey rating their experience using session types. The time taken to complete the experiment may be longer or shorter depending on the level of programming expertise.

1. 5 minutes to explain the experiment
2. 15 minute practical tutorial on session types, typestate, the tools to be used
3. approximately 10 minutes for task 1
4. approximately 15 minutes for task 2
5. approximately 20 minutes for task 3
6. 5 minute short informal interview
7. 5 minute to complete survey
8. 5 minutes for any questions the participant might have

C Training Needs Analysis Form

University of Glasgow | College of Science & Engineering

Training Needs Assessment and Record 2015

Name	Adriana Laura Voinea	Year of Study	First	School	Computing Science		
Date Discussed with supervisor	03/06/2016	Date Submitted	16/06/2016	Funding Source	EPSRC	Total Credits	11

The areas below correspond to those detailed in the Researcher Development Framework . You should select the areas you would like to develop and discuss these with your supervisors. You may be able to obtain the skills you require through training and/or practical experience. The completed TNA/Record should be submitted to the Graduate School when you submit your paperwork for your annual progress review.

DOMAIN A – KNOWLEDGE AND INTELLECTUAL ABILITIES				
	Development Required	Courses identified from the Doctoral Research training programme and elsewhere and any practical experiences you intend to undertake to develop the skills required	Courses attended and dates	Credits
A1 – Knowledge Base Includes subject knowledge, research methods, information search skills and management, languages	Knowledge of Statistics	RSDA 6107 Introduction to Statistical Inference RSDA 6108 Introduction to Statistical Modelling RSDA 6109 Introduction to Design and Analysis of Experiments	Attended on 27/05/16 Attended on 28/05/16 Attended on 29/05/16	1 1 1
	Programming language theory	Scottish programming language seminars Betty meeting	Attended on various dates throughout the year Attended on 17-18 March 2016	N/A N/A
A2 – Cognitive abilities	Literature review	RSDA 6083 Literature critique/ review SC	Not attended as it was cancelled due to illness	N/A

Includes analysing, synthesising, critical thinking, evaluation and problem-solving skills				
A3 – Creativity Includes developing inquiry skills, intellectual insight, innovation, constructing an argument, intellectual risk	Inquiry skills, constructing an argument	Project meetings Scottish programming language seminars FATA seminars	Attended on various dates throughout the year Attended on various dates throughout the year Attended on various dates throughout the year	N/A N/A 1

DOMAIN B – PERSONAL EFFECTIVENESS				
	Development Required	Selected Courses from the Doctoral Researcher Development Training Programme	Date Course(s) Attended	Credits
B1 – Personal qualities Includes enthusiasm, perseverance, integrity, self-confidence, responsibility	Self-confidence, responsibility	Mini Project	10/15-01/16	N/A
B2 – Self management Includes preparation and prioritisation, commitment, time-management, work-life balance, responsiveness to change	Prioritisation, time-management, project management	RSDC 6001 Project management – An Introduction	Attended on 01/03/2016	1
B3 – Professional and career development Includes career management, continuing professional development,	Networking, continuing professional development	Attended Seminars (SPLS) Attended Betty Meeting WadlerFest/LFCS30	Attended on various dates throughout the year Attended on 17-18 March 2016 Attended on 11-13 April 2016	N/A N/A 1

responsiveness to opportunities, networking				

DOMAIN C – RESEARCH GOVERNANCE AND ORGANISATION				
	Development Required	Selected Courses from the Doctoral Researcher Development Training Programme	Date Course(s) Attended	Credits
C1 – Professional conduct Includes health and Safety, ethics, principles and sustainability, IPR/copyright, respect and confidentiality, attribution and co-authorship, appropriate practice	Improve knowledge of IPR/copyright, respect and confidentiality, attribution and co-authorship	RSDC 6023 Research Integrity Equality and Diversity Online Training	Attended on 21/01/2016 Completed on 25/05/2016	1 1
C2 – Research management Research strategy Project planning and delivery Risk management	Improve knowledge of data management	RSDC 6025: Research Data Management	Attended on 17/02/2016	1
C3 – Finance, funding and resources Includes incomes and funding generation, financial management, infrastructure and resources	Funding generation, financial management, infrastructure and resources	FATA seminars	Throughout the year	N/A

DOMAIN D – ENGAGEMENT, INFLUENCE AND IMPACT				
	Development Required	Selected Courses from the Doctoral Researcher Development Training Programme	Date Course(s) Attended	Credits
D1 – Working with others	Collaboration skills	Attended Project Meetings	Multiple dates: September 2015 – now	N/A

Includes collegiality, team-working, people management, supervision, mentoring, influence, leadership, collaboration, equality and diversity				
D2 – Communication and dissemination Includes communication methods, communication media, publication	Improve presentation skills	RSDD6002: Presenting with Impact	Attended on 12/05/2016	1
D3 – Engagement and impact Includes teaching, public engagement, enterprise, policy, society and culture and global citizenship	Teaching Teaching training	Tutored 2nd Year Course: JOOSE2 Tutored 2nd Year Course: WAD2 Graduate Teaching Assistant Statutory Training	10/2015 – 03/2016 01/2016 – 03/2016 Attended on 26/20/2015	N/A 1

References

[1] A basis for concurrency and distribution. http://groups.inf.ed.ac.uk/abcd/.