Project Start Date: 1.9.2016

Project Duration: 30 months

GAP Data Management Plan

Deliverable details	
Deliverable number	1.1
Version number	1.0
Author(s)	Rachel Barrett & Anne Holohan
Due date	31st October 2016
Delivered date	31st October 2016
Reviewed by	Stephen Purcell (FAC)
Dissemination level	PU
Contact person EC	Carla Rocha-Gomes

History of Changes			
Date	Version	Person in Charge	Description
11/10/2016	v0.1	Rachel Barrett	Deliverable
24/10/2016	v0.2	Rachel Barrett	Added details of software supported in TARA. Added further reference to polces covered in GAP data and incidental findings policy. Add reference link to H2020 DMP template.
25/10/2016	v0.3	Mads Haahr	Added details of the software used to share game play data. Added details of standardized language. Added details about the size of the data.
27/10/2016	V0.4	Rachel Barrett	Added further detail about GAP ethics policies.
29/10/2016 30/10/2016	V0.5	Anne Holohan	Added further detail to Executive Summary, Introduction, Data Summary, overall content revision.

Executive Summary

The GAP data policy document sets out our principles and agreed practices in data management and sharing of data produced by the GAP project. Our goal is to ensure our research data is 'findable, accessible, interoperable and reusable (FAIR) to ensure it is soundly managed." We are guided by the principle that "good data management is not a goal in itself, but rather the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse."[1]

Our principles are described in relation to the different types of data collected for the GAP project, and our approach to data sharing, balanced with maintaining anonymity for data subjects, is presented. There are two rounds of interview data produced in GAP: in WP3 and WP5. Access to data collected in WP3 [interviews with personnel on their past experiences in deployment on CPPB missions] will be restricted, as "open access to data does not change the obligation to protect results in Article 27, the confidentiality obligations in Article 36, the security obligations in Article 37 or the obligations to protect personal data in Article 39, all of which still apply." [2]

Access to the interview data produced in WP5, during the evaluation of the game, will be open access, as will the data produced by the players within the game.

---

[1] H2020 Programme Guidelines on FAIR Data Management in Horizon 2020 Version 3.0 26 July 2016

[2] Article 29, H2020 AGA — Annotated Model Grant Agreement: V2.1.1 – 1 July 2016
http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/amga/h2020-amga_en.pdf

29.3   Open access to research data

[OPTION for actions participating in the open Research Data Pilot: Regarding the digital research data generated in the action ('data'), the beneficiaries must:

(a) deposit in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate — free of charge for any user — the following:

(i)   the data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible;

(ii)   other data, including associated metadata, as specified and within the deadlines laid down in the 'data management plan' (see Annex 1);

(b) provide information — via the repository — about tools and instruments at the disposal of the beneficiaries and necessary for validating the results (and — where possible — provide the tools and instruments themselves).

This does not change the obligation to protect results in Article 27, the confidentiality obligations in Article 36, the security obligations in Article 37 or the obligations to protect personal data in Article 39, all of which still apply.

As an exception, the beneficiaries do not have to ensure open access to specific parts of their research data if the achievement of the action's main objective, as described in Annex 1, would be jeopardised by making those specific parts of the research data openly accessible. In this case, the data management plan must contain the reasons for not giving access.]

The DMP will support the management life-cycle for all data that will be collected, processed or generated by GAP. It will include:

1. what data GAP will generate
2. whether and how it will be made accessible
3. how it will be maintained and preserved.

The DMP will be updated and completed (i.e. become more precise) as GAP evolves. New versions of the DMP will be created whenever important changes to GAP occurs (due to inclusion of new data sets, changes in consortium policies or external factors). This document, following the suggested Horizon 2020 FAIR Data Management Plan (DMP) template format given by the European Commission, outlines our approach to data sharing, within the limitation of ensuring privacy for GAP's research participants. Further information on our policy towards data collection, storage, protection, retention and destruction and compliance with national and EU legislation, and incidental findings may be found in the GAP Data Policy document and the GAP Incidental Findings document respectively.

At every stage, the GAP management and consortium will ensure the Data Management Plan is in line with the norms of the EU and Commission [as expressed in the General Data Protection Regulation (GDPR) (Regulation (EU) 2016/679)] and will promote best practice in Data Management.

1.Introduction

Gaming for Peace (GAP) is a project funded through the European Commission Horizon 2020 Secure Societies programme. GAP will collect deployment experiences from relevant civilian and military

informants in Europe (WP3). These will then be used, in conjunction with a review of state of the art current training, to develop a curriculum in soft skills (communication, cooperation, gender awareness and cultural competency) and this curriculum will be embedded in an immersive online role-playing game. Under WP5, there will be several rounds of evaluation of the curriculum and game, producing interview data with police, military and civil personnel and player metrics for those same individuals. The final version of the curriculum and game will be used by those being deployed in Conflict Prevention & Peace Building (CPPB) roles.

GAP supports the four key principles of FAIR - findability, Accessibility, Interoperability, and Reusability, and in its data management plan aims to go beyond proper collection, annotation, and archival to include "the notion of 'long-term care' of valuable digital assets, with the goal that they should be discovered and re-used for downstream investigations, either alone, or in combination with newly generated data."[3]

We adopt the definition of 'Digital research data' as information in digital form (in particular facts or numbers), collected to be examined and used as a basis for reasoning, discussion or calculation; this includes statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images.[4]

"As open as possible, as closed as necessary"

The GAP project was reviewed by the European Commission's Ethical reviewers in December 2015. In this review the issue of anonymity of interview and game participants was raised. As the pool of research participants has very defined parameters, the ethics reviewers expressed the opinion that ensuring anonymity of the participants would be a challenge. Following this assessment, it was agreed with the Commission Project Officer that GAP should not participate in the Open Research Data pilot. The ORD pilot aims to improve and maximise access to and re-use of research data generated by Horizon 2020 projects and takes into account the need to balance openness and protection of scientific information, commercialisation and Intellectual Property Rights (IPR), privacy concerns, security as well as data management and preservation questions. While open access to research data becomes applicable by default in Horizon 2020, the Commission also recognises that there are good reasons to keep some or even all research data generated in a project closed, as is the case for the first round (WP3) interviews in GAP.

However, the data generated in the second round of interviews in GAP (WP5) which comprises interviews with players evaluating the game will be open access as the likelihood of identifying information will be minimal, as the focus of the interviews is on the experience of playing the game.

---

[3] Mark Wilkinson et al *Scientific Data* 3, Article number: 160018 (2016) doi:10.1038/sdata.2016.18, accessed at http://www.nature.com/articles/sdata201618

[4] H2020 AGA — Annotated Model Grant Agreement: V2.1.1 – 1 July 2016,p. 219

2. Data Summary

2.1 Purpose of the data collection/generation and its relation to the objectives of the project & types and formats of data

GAP will gather interview and game play data primarily from the following project partners in two rounds of data gathering:

- Police Service of Northern Ireland (PSNI)
- Finnish Defence Forces (FINCENT)
- Armed Forces of the Republic of Poland (Via NDUW)
- Polish National Police (WSpol)
- Bulgarian Defence Force (BDI)
- Portuguese State Police (PSP)

Data will also be gathered from members of the project's 'End-user Advisory Board'. This is primarily comprised of NGO and voluntary groups working in the CPPB field. In the first round of data gathering (WP3), this is to be combined with state of the art knowledge in training personnel in CPPB missions to build realistic scenarios which capture the defined learning objectives of GAP. Personnel from the same organizations will be interviewed in the second round of data gathering (WP5), the purpose of which is to gather data on the experience of playing the game. Metrics from data generated by the players actually playing the game will also be collected and analyzed to ascertain the degree and kind of learning achieved in the game.

Table 1: Types of data to be collected throughout the project

1. Interview data:
• **Work Package 3 (WP3) interviews**: Interviews with participants' (drawn from end-user partner organizations and the end-user advisory board member organizations) on their experience of deployment in CPPB missions. The data from these interviews will contribute to the development of game play storylines, ensuring that the game is based on realistic scenarios. Interview data will also be analysed using sociological methods and theories to produce journal publications and will be the subject of analysis for PhD theses funded by GAP.
• **Work Package 5 (WP5) interviews**:   In the evaluation phase of GAP, there will be pre- and post-gameplay interviews. Participants will evaluate the game, both from a computer-user interaction perspective and from the perspective of impact on their organizational role and others organizational roles in CPPB missions. These interviews will focus on their experience of the game and their interpretation of and response to the various scenarios. This data will be used to modify the game play experience and the curriculum, and will be interrogated to contribute to sociological and psychological understandings of CPPB training and experiences.
2. Game play data:
The performance of test players will be recorded in the game and a data set will be

produced from this. Test players will again be sought from the participating armies and police services, with additional testers sought from organisations in the End-user Advisory Board. The GAP game will require players to take on different roles within the game than they hold in real life (e.g. a male member of the army could play the role of a female NGO worker). Data produced will be used to study the degree and kind of learning, including the achievement of various soft-skills (communication, cooperation, gender awareness, cultural competency) by the players. The data will be used to modify the game play experience and the curriculum as necessary, and also will provide the basis for a 'skills passport' that will provide a means to standardize results and thus enable game players to document their achievement of targeted soft-skills.

2.2 Re-use of existing data

GAP will not reuse existing data.

2.3 Origin of data

GAP will produce original research data from two rounds of interviews carried out in six countries, and play metrics within the subsequently developed game [see 2.1 above].

2.4 Expected size of the data

WP3 interviews: Interviews with 15-30 individuals will be carried out in each country lasting 45 minutes – one hour. This will produce approximately 30 typed pages of transcript per interview. This will produce approximately 180 individual interviews x 30 pages of interviews= 2,700-5,400 pages of transcript.

WP5 interviews: 15-30 interviews will be carried out in each country lasting 15-30 minutes pre-game play and 30/45 minutes post-game play. This will produce approximately 30 typed pages of transcript per interview. This will produce approximately 180 individual interviews x 30 pages of interviews= 2,700-5,400 pages of transcript.

Game data: The exact size of the gameplay database will depend on the specific assessment metrics produced by WP4.2, but it is not expected that it will consist of more than 1 GB of data.

2.5 To whom is this data useful?

This data will be useful to academics, practitioners (end users) and policymakers. It will be useful to researchers who are interested in the development of serious games, specifically games as a tool for learning; and researchers who aim to understand the acquisition of specific soft skills, specifically but not limited to those skills as used by CPPB personnel. It will be useful to end users (organizations such as the end user partners in the Consortium of GAP) who can use the data to expand and attain a greater understanding of how to provide and deliver a curriculum of appropriate soft skills to personnel being deployed in CPPB.

3. FAIR Data

3.1 Making data findable

Our policy is guided by the 3 steps outlined in Annex 1.

 Open access to digital research data involves 3 steps:

Procedure for open access (research data):

Step 1 — Deposit the digital research data, preferably in a research data repository.

Step 2 — Provide open access by taking measures to enable users to access, mine, exploit, reproduce and disseminate the data free of charge (e.g. by attaching a 'creative commons licence' (CC-BY or CC0 tool) to the data). Open access must not be given immediately; for data needed to validate the results presented in scientific publications, as soon as possible; for other data, beneficiaries are free to specify embargo periods for their data in the data management plan (as appropriate in their scientific area).

Step 3 — Provide information, via the repository, about tools and instruments for validating the results.

3.1.1 Naming conventions
GAP interview data will follow the naming convention (Interview number_Initials of interviewer_date of interview_GAP name and project).   Thus an example of a file name would be "Int1_RB_10.1.2017_GAP 700670"

The game play dataset will be labelled as such, with metadata providing more information on background to the data.

3.1.2 Search keywords
Metadata data will be included with each item to be deposited in the open access repository (see section 3.2.2 for details on the repository). This will allow for keyword searching.

3.1.3 Metadata
Each item to be included in the open access repository will include metadata giving background explanation of the GAP project, and specifics on how the data was created and the data format. We will follow the guidelines set out in the TARA repository (see section 3.2.2)

3.2 Making data openly accessible

3.2.1 Which data produced will be made openly available
Anonymised WP5 interview transcripts and game play data will available for open access sharing. WP5 interviews concern the participants' experience of the game, and are likely to contain less sensitive information that WP3 interviews (although identifying information or vignettes may be present). WP5 interview transcripts will be checked to ensure anonymity by the responsible team member at each organisation (see table one of the GAP Data Policy documents for further details) before wider release to the Consortium. Any identifying details or vignettes (after anonymization) will be redacted before release to the wider Consortium.

Game players will be given an identifying number, with their corresponding identifying details encrypted and kept in a secure location by project partner Haunted Planet Studios (HPS). Prior to depositing email addresses used for creating game accounts, hashed passwords, all IP addresses logged during play and any chat logs that maybe be created will be removed. As game play data can be fully anonymised, and players will be taking on a role different from their role in real-life, the ability to identify game players will be severely circumscribed, thus this data will be suitable for depositing in the TARA repository.

WP5 Interview data to be deposited will include:

- Anonymised interview transcripts from pre and post-game play interviews.

Game data to be deposited will be:

- Description of the scenarios and the soft-skills learning goals associated with these
- Anonymised quantitative data, on actions taken within the game, broken down by scenario

WP3 interview concern the real experience of CPPB by members of the participating police and armies (listed in section 1.1) plus participating NGO and civil society members. Due to the sensitive nature of the data collected via interview, and the risk of interviewee identification, sharing of interview data on the interviewee's experience of CPPB activities will not be included in the open access repository. Rather, WP3 interviews will be made available for sharing on a case-by-case basis following a request to the GAP PI. Anonymised WP3 interview transcripts will be available to other researchers on request. Requests will be reviewed by the GAP PI and, if the request is deemed appropriate, anonymised transcripts with identifiable vignettes redacted, will be supplied.

As per Article 38 of the Grant Agreement, Visibility of EU funding Any dissemination of results (in any form), even when combined with other data, must include the reference to EU funding set out in the GA.

3.2.2 Method for data accessibility

Open access means taking measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate data — via a research data repository. A 'research data repository' means an online archive for research data; this can be subjectbased/thematic, institutional or centralised. The GAP project will use Trinity's Access to Research Archive (TARA) to make the above outlined data openly accessible. TARA is designed to store, index, distribute, and preserve the digital materials of Trinity College Dublin. Content, deposited directly by Trinity faculty and staff, may include research papers, photographs, videos, theses, conference papers, or other intellectual property in digital form. The content is then distributed through a searchable Web interface. TARA uses DSpace open source software, developed by MIT and Hewlett Packard.

Data uploaded to TARA is requires metadata attached to each item. This metadata entails descriptive information about an item that allows it to be found via keyword searching. Thus each item of GAP data will include metadata.

3.2.3 Methods or software tools needed to access the data
TARA is accessible from a standard web browser.

3.3 Making data interoperable

3.3.1 Interoperability of data produced

TARA has a predefined list of software packages that it supports. These software types are 'recognised' by TARA and will be maintained as such by the system on an ongoing basis and if/when the content is exported or moved, or the server is changed. Qualitative interview data will be provided on Microsoft word, which is a package 'recognised' in TARA. All transcripts will be provided in English and will follow a standard transcript template developed by the research team.

Game play data will be provided in machine-readable eXtensible Markup Language (XML) format, which is also supported in TARA. Supporting data on the scenarios that the data relates to will also be supplied in XML format.

3.3.2 Standard vocabularies

In the conduct of qualitative interviews it is likely that technical terms, specific to conflict prevention and peace building, may be used. In this case, we will provide a mapping of these terms to accompany the transcripts.

The gameplay data will be expressed in an XML application designed specifically to reflect the metrics produced in WP4.2. A Document Type Definition (DTD) will be provided, which will help with making the XML data set fully machine-readable, searchable and interoperable.

3.4 Increase data re-use

By depositing WP5 interview and game play data in the TARA repository, this data will be made freely available to anyone via the internet. TARA provides a persistent web link that remains constant and allows anyone worldwide access over the internet. No licencing would be required for accessing of the deposited data.

WP3 data, shared if appropriate, will be provided in a standardised PDF format. The PI will liaise with those accessing the data about format usability – while ensuring the security of the data.

3.4.1 When will the data be made available for re-use?

Data collected during the project will be used for two purposes 1) creation of the training game and 2) academic publication. We will embargo release of the data until both of these objectives have been achieved. The anticipated time period for depositing the data into the TARA repository will be late-2019.

3.4.2 How long is it intended that the data remains re-usable?

The TARA repository provides long-term stewarded preservation of deposited materials; no specific time limit will put on the use of data deposited in TARA.

4. Allocation of resources

By using the TARA repository, we are able to avail of a service provided, at no additional charge, to TCD academics. Thus, additional funding for making GAP data open access is not required.

4.1 Responsible persons

The GAP Principle Principal Investigator (Anne Holohan) will be responsible for ensuring that interview transcripts and game data are deposited into TARA. Once deposited, the TCD library team, who manage the TARA repository, will be responsible for the long term storage of the GAP data.

5. Data security

All data deposited into TARA is backed up on the TCD servers. The GAP Principal Investigator will also ensure that GAP data is stored on departmental servers.

6. Ethical Aspects

As discussed at the start of this document, ethics concerns mean that we cannot have full open-access sharing of all data to be produced by the GAP project.   GAP ethics documents outline: 1) our policy towards in data collection, storage, protection, retention and destruction and compliance with national and EU legislation, and 2) our Incidental Findings policy. In brief, our policies on the above are as follows:

- Data Collection: Data will be collected in WP3 and WP5 via interview and game testing. All participating individuals will be provided with information sheets on the project and asked to sign consent forms. Via these we will ensure that that our policies on handling, storing and retaining their data, their right to withdraw and our policy on incidental findings are understood by the study participants. Participants in interviews and game play may involve the same individuals, or may be conducted with separate people. In either scenario, separate consent will be sought before interviews and game evaluation.
- Data storage: Interview data will be immediately transferred off recording devices and encrypted. Access to un-anonymised data will be strictly limited and only anonymised data will be transferred amongst consortium members. Consent forms will be kept in a secure and separate location from transcripts and game play data.
- Data protection: The content of data produced by the GAP project will be specified, and we will provide copies of appropriate authorizations according to the legal requirement of the area where the research is planned to take place. This includes all partner institutional/college ethical approval committees, operating under the auspices of EU regulations on Data Protection and Privacy, notably the Data Protection Directive (Directive 95/46/EC) and the General Data Protection Regulation (Regulation (EU) 2016/679).
- Original recordings will be stored for the duration of the project (2.5 years); after this time they will be irreversibly destroyed by overwriting the file with other sound. The consent forms will be stored at a different location to the transcripts and recordings for the duration of the project. Then they will all be destroyed by shredding of hard copies (informed consent) and irreversible overwriting of soft copies. Those who collected original recordings in each jurisdiction will be responsible for destruction of the aforementioned material in this manner at the end of the project.
- Incidental findings: Should interview data result in an incidental finding (such as reports of illegal/prohibited behaviour or reports of PTSD) the research team will follow the

escalation/reporting protocol (set up in the GAP Incidental findings policy document). In the case of illegal prohibited behaviour, this will be reported in the first instance to the appropriate person in their organisation, who will then escalate this to an outside agency if appropriate. In the case of mental health issues, the research team will direct the participant to support services both inside and outside their organisation.

Full details of these policies are available in the GAP Data Policy Document and the GAP Incidental Findings document.

6.1 Informed consent

GAP will comply with Article 39.2 of the grant agreement:

39.2 Processing of personal data by the beneficiaries

The beneficiaries must process personal data under the Agreement in compliance with applicable EU and national law on data protection (including authorisations or notification requirements).

The beneficiaries may grant their personnel access only to data that is strictly necessary for implementing, managing and monitoring the Agreement.

The beneficiaries must inform the personnel whose personal data are collected and processed by the *Agency or the* Commission. For this purpose, they must provide them with the service specific privacy statement (SSPS) (see above), before transmitting their data to the *Agency or the* Commission.

GAP's open data policy will be outlined in information sheets and consent forms for participants in pre and post-game interviews. Game players will also receive information and consent forms that inform them how the data will be used.

7. Timetable for Updates.

The DMP will be updated minimally at Month 12 and Month 28, after the scheduled Ethics Committee meetings but also whenever important changes to GAP occur (due, for example, to inclusion of new data sets, changes in consortium policies or external factors).

8. References

European Commission Participant Portal H2020 Online Manual
http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

H2020 Programme Guidelines on FAIR Data Management in Horizon 2020 Version 3.0 26 July 2016
http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

H2020 AGA — Annotated Model Grant Agreement: V2.1.1 – 1 July
2016http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/amga/h2020-
amga_en.pdf

Wilkinson, Mark et al *Scientific Data* 3, Article number: 160018 (2016) doi:10.1038/sdata.2016.18,
accessed at http://www.nature.com/articles/sdata201618