HECTOR

D5.2

Data Management Plan (DMP)

Project number:	644052
Project acronym:	**HECTOR**
Project title:	Hardware Enabled Crypto and Randomness
Start date of the project:	1st March, 2015
Duration:	36 months
Programme:	H2020-ICT-2014-1

Deliverable type:	Report
Reference number:	ICT-644052/D5.2/1.0
Work package:	WP5
Due date:	August 2015 – M06
Actual submission date:	1st September 2015

Responsible organisation:	TUG
Editor:	Thomas Korak
Dissemination level:	Public
Revision:	1.0

Abstract:	The purpose of the DMP is to provide an analysis of the main elements of the data management policy that will be used by the applications with regard to all the datasets that will be generated by the project. The DMP should ensure that most important aspects regarding data management, like metadata generation, data preservation, and responsibilities, are identified in an early stage of the project. This ensures that data is well-managed during the project and also beyond the end of the project. Data which will be generated in the course of the project include output data of random number generators, PUF output data, measurement data, and source code. As the DMP is an incremental tool, it will be adapted in the course of the project.
Keywords:	data management policy; datasets; gathering process, accessibility, sharing, archiving

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 644052.

Editor

Thomas Korak (TUG)

Contributors (ordered according to beneficiary numbers)

- Martin Deutschmann, Corinna Kudler, Sandra Lattacher (TEC)
- Dave Singelee, Ingrid Verbauwhede (KUL)
- Victor Fischer (UJM)
- Alexandre Anzala-Yamayako (TCS)
- Bernard Kasser (STR)
- Guido Bertoni (STI)
- Michal Varcholda (MIC)
- Gerard Battum (BRT)

Executive Summary

This document represents the first version of the Data Management Plan (DMP) of the HECTOR project. It is a living document and will be updated on demand in the course of the project. The data for the DMP, which will be generated by the HECTOR project partners, have been identified. The data which have been identified so far involves output streams of TRNGs during normal operation and when applying active attacks, hardware signatures of PUFs during normal operation and when applying active attacks, leaked signal traces for side-channel analysis attacks, source code of hardware and software modules, and output data of statistical tests. Results of statistical tests and other relevant analysis and performance data will be published in the according deliverables for dissemination. Additional data that are identified in the course of the project will be considered in an updated version of the DMP accordingly.

The first version of the DMP also proposes several options for data sharing. Two parameters mainly influence the infrastructure, which will be used for data sharing. The first parameter defines if the data will be publicly available or only shared internally between project partners. A second parameter is the expected size of the generated data. Small amounts of data which will only be shared inside the project boundaries will be shared by using the already existing project SVN. This project SVN allows sharing and versioning of files between project partners. Larger amounts of data can be shared by using a public cloud infrastructure or a sharing infrastructure provided by the project partner generating the data. Here the decision, which sharing infrastructure to be used will be made on demand. This option can be applied for both, internal and publicly available data. All the required information for using the publicly available data will be provided on the HECTOR project homepage.

For the data provided to the public, also an appropriate licensing scheme is planned. The goal is to allow third parties to use, modify, and build on the data provided by the HECTOR project without or with small restrictions. Therefore, *Creative Commons* (see http://creativecommons.org/licenses/?lang=en) provides an online tool for selecting an appropriate license.

In summary, this first version of the DMP provides an early overview of the expected data that will be generated and disseminated during the HECTOR project. For dissemination, the DMP suggests sharing options for the data items in order to make them accessible for external partners like other research organisations or universities. Dissemination also includes publishing the results achieved during the HECTOR project in public deliverables and scientific publications. The HECTOR project homepage will serve as the main point of information on publicly available data.

Contents

Chapter 1 Introduction

In research projects it is common that several partners work together and produce a lot of data related to the project. Therefore, it is important to specify in an early stage of the project what data will be generated, how it will be shared between the project partners and if it will be publicly available. A data management plan (DMP) is a tool which should assist in managing the data created during the project.

In general, the DMP should specify what data will be generated, collected, and processed during the project. It should also provide information whether and how data will be exploited and open for public and re-use. The DMP should include information on what standards and methodologies will be used and how the data will be handled during and after the research project (how the data will be curated and preserved).

The DMP should result in a checklist for the future; it should serve as a reference for resource and budget allocation. Further, it should support and describe the data management lifecycle. The DMP is a living document, the first version is submitted in M06, and updated versions are planned for M18 and M36.

In particular, the data created by the HECTOR project will be in the form of:

- Bit streams generated by true random number generators (TRNGs)
- Hardware signature codes generated by physically unclonable functions (PUFs)
- Results of statistical testing methods for TRNGs using AIS 31, NIST SP 800-22 and NIST SP 800-90B methodologies
- Results of statistical testing methods for PUFs
- Measurement data of power consumption/electromagnetic emanation of the investigated devices observed during hardware side-channel attacks (leakage traces)
- Output data acquired during active attacks (e.g. fault attacks) targeting specific modules (e.g. TRNG, PUF)
- FPGA or ASIC specific HDL code describing modules for performing efficient cryptographic calculations or microcontroller-specific software for cryptographic computations

Parts of the created data will be made available for the public, e.g. the research community. Cloud storage (Dropbox, Google drive, …) or an IT service hosted by one of the project partners will be applied for that purpose. To make the data easily accessible there will be a direct link from the HECTOR homepage to the service where to download the data from incorporated with a detailed description of the data sets. Therefore, no specific data sharing infrastructure is required at the partner sites. This approach allows providing access to interested parties outside the project by e.g. simply sharing a URL.

It has to be considered that the size of some types of generated data (e.g. leakage traces) might exceed several Gigabytes (GBs) making it impossible to share using cloud storage or a comparable server-based solution. In such cases, only a subset of the data will be shared to limit the storage requirements. If external parties are interested in the whole dataset, an appropriate sharing solution can be set-up on demand. The required information therefore will be provided in the appropriate dataset description which can be found on the HECTOR homepage.

The data will typically be stored by the project partner generating it. E.g. the partner who performs side-channel measurements will store the corresponding leakage traces locally. Sharing the data between project partners will be done on demand. Depending on the size of the data to share, different approaches will be used: SVN, cloud storage, server-based approach, exchange USB sticks or hard drives.

For source code created during the project (e.g. HDL code of cryptographic modules, microcontroller code), only parts which do not include protection mechanisms against e.g., side-channel analysis attacks, will be made available for the public. The developer of the code will benefit from sharing the code in the way that other interested researchers can reuse the code. This reuse results in citations for the author. On the other hand the research community can benefit from the publicly available code in the way that implementing standard algorithms (e.g. authenticated encryption algorithms submitted to the CAESAR competition [1]) from scratch becomes unnecessary.

Results of side-channel analysis (SCA) attacks based on leakage traces, results of statistical tests for the TRNGs/PUFs, and implementation results of the cryptographic building blocks (area numbers, runtime) will be published in deliverables. Therefore these numbers are accessible for interested parties outside the project. Also scientific publications will ensure that the results are disseminated.

For publicly available data, an appropriate licensing scheme will be put in place. Interested third parties should be allowed to use, modify, and build on the provided data. One option to allow this is attaching a Creative Commons License (see http://creativecommons.org/licenses/?lang=en) to the data. Two examples are the CC0 license and the CC-BY license. While CC0 allows the author of data to waive the copyright completely, CC-BY allows the reuse of the data by a third party, but the original author has to be cited. Specific use-cases might require using a more-restrictive license (e.g. http://www.apache.org/licenses/LICENSE-2.0). If such cases are identified in the course of the project, decisions will be made on demand and the DMP will be adapted accordingly.

Currently there are no plans to use existing data. This might change if VHDL code for specific modules is already available by one project partner or if code from a third party can be used without license restrictions. This is a further adaption of the DMP which might become necessary during the lifecycle of the project.

Chapter 2 Data generation

Data Nr.	Responsible Beneficiary	Data set reference and name	Data set description			Research data identification				
			End user (e.g. university, research organization, SME's, scientific publication)	Existence of similar data (link, information)	Possibility for integration and reuse (Y/N) + information	D[1]	A[2]	AI[3]	U[4]	I[5]
1	UJM	Huge random bit streams and random data streams generated by proposed TRNGs in different technologies	University, research organisation, SMEs	No other similar data are available	Y; the data will be used within this project for the statistical evaluation and may be reused in other projects		x	x	x	x
2	TEC	Hardware signature codes generated by proposed PUFs in individual devices	University, research organisation, consortium	Data from PUFs that were developed in the course of the FP7 project UNIQUE, http://unique.technikon.com	Y; the data will be used within this project for the statistical evaluation and may be reused in other projects for advanced analysis		x	x	x	x
3	BRT	Results of TRNG statistical testing using AIS31, NIST SP 800-22 and NIST SP 800-90B methodologies	University, research organisation, SMEs	No other similar data available	Y; the data will be used within this project for the statistical evaluation and may be reused in other projects		x	x		x

[1] Discoverable
[2] Accessible
[3] Assessable and intelligible
[4] Usable beyond the original purpose of which it was collected
[5] Interoperable to specific quality standards

Data Nr.	Responsible Beneficiary	Data set reference and name	Data set description			Research data identification				
			End user (e.g. university, research organization, SME's, scientific publication)	Existence of similar data (link, information)	Possibility for integration and reuse (Y/N) + information	D[1]	A[2]	AI[3]	U[4]	I[5]
4	TEC	Results of PUF statistical testing using new proposed methodology	University, research organisation, consortium	http://unique.technikon.com	Y; advanced analysis may be based on this data		x	x		x
5	BRT	Leaked signal traces observed during hardware SCA	University, research organization	Power measurements for the DPA contest, http://www.dpacontest.org/v4/rsm_traces.php	Y; Might be reused in other projects to evaluate e.g. novel attack methods		x	x		
6	UJM	Test output data acquired during active attacks on proposed modules and demonstrators	University, research organisation, SMEs		Y; may be used in other projects too			x		
7	TUG	VHDL code of building blocks for demonstration and evaluation in WP4	University, research organization	ASCON hardware implementations at github, https://github.com/ascon/ascon_collection	Y; The building blocks might be reused for other projects and scientific research.	x	x	x		x

Table 1: Data generation

Explanation of Table 1:

Data set reference and name:

Identifier for the data set to be produced

Data set description:

Description of the data that will be generated or collected, its origin (in case it is collected), nature and scale to whom it could be useful, and whether it underpins a scientific publication. Information on the existence (or not) of similar data and the possibilities for integration of reuse.

Research Data Identification

The boxes (D, A, AI, U and I) symbolize a set of questions that should be clarified for all datasets produced in this project.

Discoverable:

Are the data and associated software produced and/or used in the project discoverable (and readily located), identifiable by means of a standard identification mechanism (e.g. Digital Object Identifier)

Accessible:

Are the data and associated software produced and/or used in the project accessible and in what modalities, scope, licenses (e.g. licencing framework for research and education, embargo periods, commercial exploitation, etc.)

Assessable and intelligible:

Are the data and associated software produced and/or used in the project assessable for and intelligible to third parties in contexts such as scientific scrutiny and peer review (e.g. are the minimal datasets handled together with scientific papers for the purpose of peer review, are data provided in a way that judgements can be made about reliability and the competence of those who created them)?

Useable beyond the original purpose for which it was collected

Are the data and associated software produced and/or used in the project usable by third parties even long time after the collection of the data (e.g. is the data safely stored in certified repositories for long term preservation and curation; is it stored together with the minimum software, metadata and documentation to make it useful; is the data useful for the wider public needs and usable for the likely purposes of non-specialists)?

Interoperable to specific quality standards

Are the data and associated software produced and/or used in the project interoperable allowing data exchange between researchers, institutions, organisations, countries, etc. (e.g. adhering to standards for data annotation, data exchange, compliant with available software applications, and allowing re-combinations with different datasets from different origins?)

It is recommended to make an “x” to each applicable box and explain it literally in more detail afterwards.

Chapter 3 Processing and explanation of generated data

The following sections provide some additional information to the listed data introduced in Chapter 2. This information includes the entity which is responsible for the data, how the data is collected, an identification of the end-users of the data, and research data identification.

3.1 Huge random bit streams generated by proposed TRNGs in different technologies

3.1.1 Responsible Beneficiary

Random data will be generated and recorded by the parties performing evaluations of random number generators. This task will be mainly performed by UJM, so they take the main responsibility of the data. It is probable that similar data will also be produced by other parties, e.g. KUL, BRT, STM, TCS or MIC, as these parties also have expertise in random number generation.

3.1.2 Gathering Process

Random data will be essentially generated using HECTOR evaluation boards and demonstrator in various conditions including border and corner operating conditions. Two types of data will be generated: the raw random data streams and the post-processed random data streams. Random data can be bits, bytes, 16- or 32-bit words. Two data formats will be available: the binary stream and the stream of random words (bytes, 16- or 32-bit words). Stream of random words can be useful for example when the raw random data is the output of a counter of random events.

The raw random bit stream files have extension *.rbs, the raw random data stream files have extension *.r08, *.r16 or *.r32 for data streams with bytes, 16- and 32-bit words, respectively. The post-processed bit stream files have extension *.pbs and the post-processed data stream files have extension *.p08, *.p16 or *.p32.

Random bit stream files with extension *.rxx or *.pxx (raw bit streams or post-processed bit streams) represent the most common file format, since this format is required by most general-purpose statistical tests (e.g. AIS 31, NIST SP 800-90B or NIST SP 800-22).

In generation and evaluation of random numbers, the order of bits, bytes and words is important, since it can change the existing pattern (if there is some). Random bytes are written into the files in the same order as they arrive. Bits are placed into the bytes in the following manner: the first arrived bit is placed to the least significant bit and the last arrived bit to the most significant bit, i.e. byte=bit8|bit7|bit6|bit5|bit4|bit3|bit2|bit1. The 16-bit words have the following format: word16=byte2|byte1 and the 32-bit words are as follows: word32= byte4|byte3|byte2|byte1.

3.1.3 End-User of the Data

The end-users of this type of data will mainly be the producers of data and other partners of the HECTOR project. It can happen that the generated data would need to be shared with another institution. The data file sizes of at least 2 MB will be needed for applying the AIS31 tests, sizes of 1 MB for applying the NIST SP 800-90B test suite and thousands of files of 125000 bytes for applying the NIST SP 800-22 test suite. The technique to share the data depends on the amount of data. A small amount (<100MB) can be shared using the existing SVN. For medium amounts (<1GB) some cloud storage infrastructure might be applied. Huge amounts (>10GB) might require to share USB sticks or external hard disks.

To allow parties outside the project to evaluate the generated data, the data files will be made publicly available. Depending on the size of the measurement data, only subsets of the data files might be publicly shared. All information concerning data acquisition will be available at the same place where the generated data can be downloaded. If one interested party requires the full set of measurement data, a custom sharing method can be set up.

3.1.4 Research Data Identification

The TRNG output data will not be discoverable in public search engines or in a global registry of research data repositories, but within the consortium internally. It will be accessible by means of an existing project subversion repository or if necessary, exchanged via data storage media. The quality and reliability of the data can be evaluated by statistical evaluation. The data may be useable in upcoming projects as well, but the purpose of the data will not change from a present-day perspective. TRNG data are used within frameworks that allow the interoperability between the existing components based on the conformity to the same standards.

3.2 Hardware signature codes generated by proposed PUFs in individual devices

3.2.1 Responsible Beneficiary

The data of Physically Unclonable Functions (PUFs) will be generated and recorded by parties performing evaluations on the data, mainly TEC and KUL. The driving partner will be UJM in this context. It will be decided at a later stage which PUF type will be used.

3.2.2 Gathering Process

There are several different sources that may be used for PUF data generation. At the current stage, responses can be derived from 65nm PUF ASICs including SRAM, Latch, D Flip-flop, Buskeeper, Arbiter and Ring Oscillator PUFs. This PUFs were developed in the course of the FP7 project UNIQUE. Another possible source is an FPGA structural and behavioural emulation of an SRAM-like PUF implemented in VHDL by TEC (realized during the FP7 project HINT). There are also ring oscillator PUF implementations ongoing that may be used in this context.

The raw PUF data have extension *.bin, for data streams in binary files and deliver sequences of bytes either in hexadecimal or binary format. For existing ASICs and for correlation analysis the physical proximity plays an important role, since "0xA1" may correspond either to "10100001" or "01011000".

3.2.3 End-User of the Data

The end-user of this type of data will be mainly the partners within the HECTOR project or organisations that perform analysis on PUF data. The generated data can be shared with other institutions. The size of the *.bin files are different for the PUF types since they show different response length, but have a maximum size of 16kB per response.

When performing statistical tests on PUF data, a lot of data is required and needs to be shared. This might lead to an exchange of the data via USB sticks or external hard disks. If PUF data is only used for a low number of reconstructions within a framework, a small amount of responses can easily be shared via the existing SVN or a cloud storage infrastructure.

3.2.4 Research Data Identification

The PUF data will not be discoverable in public search engines or in a global registry of research data repositories, but within the consortium internally. It will be accessible by means of an existing project subversion repository or if necessary, exchanged via data storage media. The quality and reliability of the data can be evaluated by statistical evaluation. The data may be useable in upcoming projects as well, but the purpose of the data will not change from a present-day perspective. PUF data are used within frameworks that allow the interoperability between the existing components based on the conformity to the same standards.

3.3 Results of TRNG statistical testing using AIS 31, NIST SP 800-22 and NIST SP 800-90B methodologies

3.3.1 Responsible Beneficiary

The results of the statistical testing are mainly produced by those who perform evaluations on TRNG data (e.g. UJM, KUL, BRT, STM or MIC) described in Section 3.1.

3.3.2 Gathering Process

Test outputs (test results) are produced from TRNG output data described in Section 3.1. Results of statistical testing using AIS 31, NIST SP 800-90B and NIST SP 800-22 methodology are generated by corresponding standard tests as log (text) files. It is important to maintain the link between tested data and test output using convenient file naming. The filename before extension must therefore be the same for input and output data of each test.

Output of tests of the raw data will have file extension:

*.r31 – for the AIS31 test suite output,

*.r22 – for the NIST SP 800-22 test suite output,

*.r9i – for the NIST SP 800-90B test suite for iid data,

*.r9n – for the NIST SP 800-90B test suite for non-iid data.

Correspondingly, output of tests of the post-processed data will have file extension:

*.p31 – for the AIS31 test suite output,

*.p22 – for the NIST SP 800-22 test suite output,

*.p9i – for the NIST SP 800-90B test suite for iid data,

*.p9n – for the NIST SP 800-90B test suite for non-iid data.

Since the NIST SP 800-90B test suite needs different input data format: one random sample per output byte (or two-byte word) must be saved. Some conversion program to convert formats described in Section 3.1 to this specific format will be needed.

3.3.3 End-User of the Data

Most of the resulting data including detailed explanations will be incorporated within (public) deliverables. So the actual main end-user of the results will be project partners and/or universities or research organisations that may use this data to build additional statistical analysis on the given results, or use them for comparisons. External end-user may also build up new analysis on already existing results or use the raw data for their own evaluations.

3.3.4 Research Data Identification

The results of the statistical evaluations will not be discoverable in public search engines or in a global registry of research data repositories, but within the consortium internally. Because of the

small size of the output data, the data can be easily made accessible by means of an existing project subversion repository. The realization and the results of the statistical tests will be published together with scientific papers and/or deliverables within the project. Therefore, the produced data can be assessed. There is no additional purpose conceivable. The interoperability is given with the exchange of the statistical evaluation between researchers.

3.4 Results of PUF statistical testing using new proposed methodology

3.4.1 Responsible Beneficiary

The results of the statistical testing are mainly produced by those who perform evaluations on PUF data (e.g. TEC, KUL) described in Section 3.2.

3.4.2 Gathering Process

PUF raw data are *.bin files, which may be read in by a MATLAB script, to subsequently perform statistical analysis. The sequence of bytes needs to be converted from a hexadecimal or decimal form to binary bit strings. When performing statistical analysis, the output parameters will be stored within a structure array that can be saved within *.mat file. A *.mat file with 1440 different output parameters (evaluation of 12 different chips) makes up about 16KB.

3.4.3 End-User of the Data

The *.mat file with the resulting parameters of the statistical analysis needs to be combined with a read-me file that will describe the structure of the stored variables. Most of the resulting data including detailed explanations will be incorporated within (public) deliverables. So the actual main end-user of the results will be project partners and/or universities or research organisations that may use this data to build additional statistical analysis on the given results, or use them for comparisons. External end-user may also build up new analysis on already existing results or use the raw data for their own evaluations.

3.4.4 Research Data Identification

The results of the statistical evaluations will not be discoverable in public search engines or in a global registry of research data repositories, but within the consortium internally. Because of the small size of the output data, the data can be easily made accessible by means of an existing project subversion repository. The realization and the results of the statistical tests will be published together with scientific papers and/or deliverables within the project. Therefore, the produced data can be assessed. There is no additional purpose conceivable. The interoperability is given with the exchange of the statistical evaluation between researchers.

3.5 Leaked signal traces observed during hardware side channel attacks

3.5.1 Responsible Beneficiary

Leakage signal traces will be recorded by the parties performing evaluations of the side-channel resistance of specific cryptographic building blocks. This task will be mainly performed by BRT, so they take the main responsibility of the data. It is very likely that similar data will also be produced by other parties, e.g. TUG or KUL, as these parties also have expertise in side-channel measurements.

3.5.2 Gathering Process

Leaked signal traces are typically recorded using an oscilloscope, independent whether power measurements or EM measurements are performed. Modern digital oscilloscopes allow storing the captured traces in different file formats. Such file formats can e.g. be CSV (comma separated file), MAT (MATLAB data file), or a proprietary format. Due to the fact that most of the formats can be easily converted into other formats it is not necessary for the different parties to agree on a common format. In case of proprietary format (BRT), a conversion tool will be provided to the partners of the consortium.

3.5.3 End-User of the Data

The end-users of this type of data will mainly be the producers itself. It is common that the institution measuring the side-channel information also evaluates the amount of leakage which can be extracted out of the measurements by applying methods like differential power analysis (DPA), template attacks (TA) and others. Of course also cases might arise where the measurement data has to be shared with another institution having more computing power for the evaluations or want to test and apply novel analysis methods. Here, the technique to share the data highly depends on the amount of data. A small amount (<100MB) can be shared using the existing SVN. For medium amounts (<1GB) some cloud storage infrastructure might be applied. Huge amounts (>10GB) will require to share USB sticks or external hard disks.

To allow parties outside the project to reproduce the side-channel analysis results or to apply new methods, the leakage traces will be made publicly available. Depending on the size of the measurement data, only subsets of the measurements might be publicly shared. All information required to use the measurements (e.g. corresponding plain text and cipher text to each leakage trace, oscilloscope model which has been used for capturing the data, measurement parameters) will be available at the same place where the measurement data can be downloaded. If one interested party requires the full set of measurement data, a custom sharing method can be set up. One existing example for sharing measurement data are the power measurements for the DPA contest available at http://www.dpacontest.org/v4/rsm_traces.php.

The results of the side-channel analyses will be reported in (public) deliverables. So additional end-user of the results will be project partners and/or universities or research organisations that may use this data to perform additional side-channel analysis with the given measurements, or use them for comparisons.

3.5.4 Research Data Identification

The leaked signal traces will not be discoverable in public search engines or in a global registry of research data repositories, but within the consortium internally. Because of the expected large size of the data, sharing it using the existing project subversion will not be applicable. Sharing options like cloud storage or a sharing infrastructure provided by the responsible project partner will be applied. If the size even exceeds several gigabytes, exchange of physical data storage devices like USB sticks or hard disks can be arranged. Results of side-channel analysis based on specific leakage traces will be published in scientific papers and/or deliverables within the project. Therefore, the achieved results based on the measurement data can be assessed.

3.6 Test output data acquired during active attacks on proposed modules and demonstrators

Investigations of the influence of active attacks on the hardware signature codes of PUFs and the random bit streams generated by the TRNGs will be performed in the course of the HECTOR project. The goal is to evaluate to what extent the investigated PUF/TRNG modules are vulnerable to active attacks in order to include appropriate countermeasures. Format and gathering process of the output data do not change when applying active attacks so for a detailed description to the corresponding data formats we refer to Section 3.1 for the TRNG case and to Section 3.2 for the PUF case, respectively.

Type and parameters of the active attacks are important information for further analyses and also for the countermeasure development. This additional information will be incorporated to the dataset description and poses the main difference to the data sets recorded without active attacks.

3.7 VHDL code of building blocks for demonstration and evaluation in WP4

3.7.1 Responsible Beneficiary

VHDL code for cryptographic building blocks will be mainly developed by the parties KUL, STI and TUG. Although the focus of TUG is more on evaluating countermeasures they will also contribute to the hardware design and act as the responsible beneficiary for this type of data.

3.7.2 Gathering Process

Hardware building blocks are typically modelled using a hardware description language (HDL) such as VHDL or Verilog. For more complex building blocks, the source code can be divided into several files which then form a project. Each project will be accompanied by a short readme file explaining the file structure providing a quick overview of the project.

Some building blocks might also be developed as software modules running on microcontrollers. Here the software is typically developed in C or a comparable high-level programming language. Projects typically consist of vendor-specific files including e.g. standard configuration routines of the target microcontroller and user-specific files including the actual program for the microcontroller.

3.7.3 End-User of the Data

Several end-users can be identified for the cryptographic hardware building blocks. First, the evaluators will use these building blocks in order to evaluate their resistance against implementation attacks such as differential power analysis (DPA) attacks or fault attacks. By evaluating designs without and with countermeasures, evaluators can rate the efficiency of the integrated countermeasures. Second, some of the building blocks will be integrated into the demonstrator platform by MIC. Finally, some of the building blocks will also be made publicly available. The decision whether the building blocks will be publicly shared will be discussed on demand but at the moment it is planned to apply the following rule:

If the implementation does not include countermeasures and is likely to be reused by other parties for comparison reasons or as foundation for integrating improvements, it will be made publicly available. In order to share the code and distribute it across the community, a web-based hosting provider for software projects like github (https://github.com/) will be used. A link to the github repository will be provided on the HECTOR homepage. This approach is already in use by TUG to make hardware and software implementations of their CAESAR submission named ASCON publicly available (https://github.com/ascon/ascon_collection).

Implementations including specific countermeasures against implementation attacks will not be made publicly available, they can be shared using the internal project SVN service.

3.7.4 Research Data Identification

The publicly available source code will be discoverable by public search engines using the name of the implemented algorithm. Links to the source code repository will also be provided from the HECTOR homepage. Project-intern source code will be shared by applying the project SVN where only the project partners have access. Implementation results (area numbers, cycle count, …) will be published in scientific publications and (public) deliverables and will therefore be publicly accessible. The interoperability is given with the exchange of the implementation results between researchers.

Chapter 4 Accessibility - Data sharing, archiving and preservation

Access to and sharing of data helps to advance science and to maximize the research investment. A whitepaper[6] by the University of Michigan reported that when data is shared through an archive, research productivity and often the number of publications increases. Protecting research participants and guarding against disclosure of identities are essential norms in scientific research. Data producers should take efforts to provide effective informed consent statements to respondents, to identify data before deposit when necessary, and to communicate to the archive any additional concerns about confidentiality. With respect to timeliness of data deposit, archival experience has demonstrated that the durability of the data increases and the cost of processing and preservation decreases when data deposits are timely. It is important that data is deposited while the producers are still familiar with the dataset and able to fully transfer their knowledge to the archive.

In particular potential users can find out about generated and existing data most likely through the project's dissemination activities (scientific publications and papers), deliverables, presentations and technical events (conferences, trade shows) etc. During the project lifetime these documents and data will be published on our official project website (www.hector-project.eu) where a broad community has access to the project information. Besides the HECTOR public websites also marketing flyers or the internal project subversion repository will be used as a tool to provide and exchange the requested data.

In principle, the data will be shared within the HECTOR consortium according to our Consortium Agreement (with respect to any IPR issues) via a secured data repository as soon as the data is available. To the public community, data will be shared according to the dissemination level of the data via the public project website. Besides the data repository and the website, the consortium is also willing to handle requests directly. Public deliverables will be made available as soon as they have been approved by the European Commission.

Generally, the consortium's opinion is that it will not be necessary to destroy any data for contractual, legal, or regulatory purposes. However, as described before, there will be the case that the confidential deliverables will be restricted. The data generated will serve as basis for future scientific research work and reports on device performance as well as for benchmarking.

With regards to the retention and preservation of the data, HECTOR will retain and/or preserve the produced data at least for three years after the project end. Due to the broad range of data generated during the HECTOR project, there will not be a single solution for data sharing. Small amounts of data (e.g. source code of hardware modules or microcontroller code, example measurement data) up to 100MB will be shared by applying the already existing project SVN repository https://hector.technikon.com. This allows easy synchronization as well as data versioning. It has to be noted that only project partners have access to the project SVN. Therefore, publicly available data needs to be shared in another way. For publicly sharing software projects (source code), the file-hosting service github (https://github.com/) has been established within the research community during the last years. The ASCON designers at TUG use the file-hosting service github to promote their software and hardware implementations of the ASCON authenticated encryption algorithm. The github repository is accessible via a link on the ASCON homepage (http://ascon.iaik.tugraz.at/links.html). For software and hardware implementations created within the HECTOR project, which will be publicly shared, a similar approach is planned. The source code of cryptographic hardware and software modules which are secured by means of countermeasures will

[6] http://deepblue.lib.umich.edu/handle/2027.42/78307

not be made publicly available. This is on the one hand due to the protection of the intellectual property of the project partners and on the other hand due to security-related considerations. Here, the internal SVN will be applied.

For bigger amounts of data in the range of gigabytes, which needs to be shared, it is foreseen to utilize commodity clouds with usage of internal infrastructure and data bases from the partners or external platforms. Costs for data storage and archiving will occur, in particular for server provision (infrastructure) and maintenance (security updates). The coordinator, Technikon, has foreseen appropriate costs in the project budget for the active project time. At a later stage of the project it can be better assessed, if further costs for data storage will occur. These costs will then be covered by the partners with their own resources.

Another potential solution for quickly sharing huge amounts of data it the direct use of public cloud solutions. The cloud storage provider dropbox figured out to offer a well-fitting solution with Dropbox Pro (https://www.dropbox.com/upgrade). It offers 1TB of data storage, 30 days versioning of files, folders can be shared by using links and shared links can be protected by passwords. Furthermore, the duration of the sharing can be limited and file permissions for different users can be set (e.g. read, modify, …). The cost for this service is 99€ per year. The cloud storage can be used for sharing data between project partners on the one hand and also to offer publicly available data.

At the current stage of the project, no data which requires some kind of embargo period has been identified. Of course this can change during the lifecycle of the project and will then be reported in an updated version of the DMP.

In order to allow third parties to access, mine, exploit, reproduce, and disseminate the publicly available data, an adequate license scheme has to be put in place. For publicly available data provided at the github repository or via another sharing infrastructure from the HECTOR homepage we plan to attach an appropriate *Creative Commons* License (http://creativecommons.org/licenses/?lang=en). Different types of licenses are provided by that service, differing in the restrictions. These restrictions include the right for modification, commercial usage, naming the original author, and passing on under the same conditions. The license with the lowest restrictions is CC0, which allows authors to waive the copyright protection on their work (“No Rights Reserved”). As a consequence, a third party can freely build upon, enhance, and reuse CC0-licensed data. The *Creative Commons* website provides a tool which allows adding several of the previously listed restrictions to enhance the CC0 license.

Chapter 5 Summary and conclusion

This data management plan outlines the handling of data generated within the HECTOR project, during and after the project lifetime. As the deliverable will be kept as a living document it will be regularly updated by the consortium. The partners put into write their plans and guarded expectations regarding valuable and publishable data.

The generated data such as leaked signal traces will not only be of interest for the project partners but also for the scientific community outside of the HECTOR project. These signal traces serve as foundation for practically verifying new methods for e.g. security evaluations. The same is true for the random bit streams generated by the TRNG designs applied in the HECTOR project. Not all institutions have the facility to generate this data on their own. This institutions benefit from the data provided by the HECTOR project. As another advantage, the public data sharing enables comparing TRNG designs across the HECTOR project borders. This will further result in citations of HECTOR project results in external scientific publications. The scientific community will also benefit from publicly available source code created during the HECTOR project. It enables e.g. comparisons of metrics like runtime, or resource consumption of algorithms created in the HECTOR project with algorithms created by external researchers. This will again lead to citations of HECTOR-related results.

The HECTOR consortium is aware of proper data documentation requirements and will rely on each partners' competence in appropriate citation etc. The Consortium Agreement (CA) forms the legal basis in dealing with IPR issues and covers clear rules for dissemination or exploitation of project data. Besides the HECTOR public website, which targets a broad interest group, also marketing flyers or the SVN repository will be used as a tool to provide data. With regards to the retention and preservation of the data, HECTOR partners will retain and/or preserve the produced data for several years, three years after the project end at least.

The HECTOR consortium is convinced that this data management plan ensures that project data will be provided for further use timely, available and in adequate form, taking into account the IPR restrictions of the project.

Chapter 6 List of Abbreviations

Abbreviation	Explanation
ASIC	Application-Specific Integrated Circuit
DMP	Data Management Plan
DPA	Differential Power Analysis
FPGA	Field-Programmable Gate Array
HDL	Hardware Description Language
IID	Independent and Identically Distributed
IPR	Intellectual Property Rights
NIST	National Institute of Standards and Technology
PUF	Physically Unclonable Function
SCA	Side-Channel Analysis
SRAM	Static Random-Access Memory
SVN	Subversion
TA	Template Attack
TRNG	True Random Number Generator
URL	Uniform Resource Locator
VHDL	Very High-Speed Integrated Circuit Hardware Description Language

Chapter 7 Bibliography

[1] D. J. Bernstein, "CAESAR: Competition for Authenticated Encryption: Security, Applicability, and Robustness," [Online]. Available: http://competitions.cr.yp.to/index.html. [Accessed 19 06 2015].