Forensic Science Regulator
Overseeing Quality

Draft Guidance: Digital Forensics Method Validation

August 2014

This is a consultation draft and therefore should not be regarded or used as a standard. This draft is issued to allow comments from interested parties; all comments will be given consideration prior to publication. Comments should be sent to FSRConsultation1@homeoffice.gsi.gov.uk and should be submitted by 31 October 2014. This mailbox is not for general correspondence and is not routinely monitored so no acknowledgement will normally be sent.

THIS DRAFT IS NOT CURRENT BEYOND 31 OCTOBER 2014.

© Crown copyright 2014

The text in this document (excluding the Forensic Science Regulator's logo) may be reproduced in any format or medium providing it is reproduced accurately, is not otherwise attributed, is not used in a misleading context and is acknowledged as Crown copyright.

CONTENTS

1 EXECUTIVE SUMMARY *6*

2 INTRODUCTION *8*

2.1 Purpose 8

2.2 Scope 8

2.3 Reservation 8

2.4 Definition of a Method 8

2.5 Pre-Validation Requirements 10

2.6 Structure 11

2.7 Implementation 12

3 AN INTRODUCTION TO METHOD VALIDATION IN DIGITAL FORENSICS *12*

3.1 Purpose 12

3.2 Importance 13

3.3 Application 13

3.4 Challenges and Issues 14

3.5 Determining Methods Requiring Validation 14

3.6 Risk Assessment 14

3.7 Validation Requirements and Acceptance Criteria 14

3.8 Previously Validated or Adopted Methods 15

4 PLANNING VALIDATION *16*

4.1 Defining Requirements 16

4.2 Validation Strategy and Plan 16

4.3 Undertaking Validation 18

4.4 Further Guidance 18

4.5 Generation and Control of Test Data 18

4.6 Evaluation 19

5 CONCLUDING VALIDATION *20*

5.1 Validation Report 20

5.2 Statement or Certificate of Validation Completion 21

5.3 Implementation 22

6 POST-VALIDATION ACTIVITIES *22*

6.1 Maintenance of Documentation 22

6.2 Quality Assurance 22

6.3 Acceptance Testing of New Equipment 23

DRAFT

6.4 Review of Updates to Equipment or Software 23
6.5 Post-Project Review 23
7 ASSESSING UNCERTAINTY IN DIGITAL FORENSICS VALIDATIONS 24
8 COMPETENCY 26
8.1 Introduction 26
8.2 Technical Skills 26
8.3 Technical Interpretation 26
8.4 Evaluative Opinion 26
9 CHECKLIST 28
10 VALIDATION REQUIREMENTS FOR NEW METHODS FROM THE PERSPECTIVE OF THE COURT 30
11 VALIDATION AND CALIBRATION ASSESSMENTS FOR A LABORATORY 33
11.1 Starting the Validation Process 33
11.2 An Example of Determining the Validation Level 34
12 CONSEQUENCES OF FAILURE TO VALIDATE – COMPUTER ANALYSIS 37
12.1 Introduction 37
12.2 Sole Reliance on Case-by-Case Quality Assurance Procedures 37
12.3 Validating the Tool Rather Than the Method 37
12.4 Validating According to a Laboratory's Audit Schedule 38
13 CONSEQUENCES OF FAILURE TO VALIDATE – CELLSITE ANALYSIS 39
13.1 Introduction 39
13.2 Absence of Evidence Equals Evidence of Absence? 39
13.3 Provision of Opinion Without an Interpretation Framework 40
14 GLOSSARY 42
15 APPENDIX A: COMPUTER FORENSICS EXAMPLE – RECOVERY OF WEB BROWSING HISTORY RECORDS FROM A COMPUTER 48
15.1 Review of End-User Requirement and Specification 48
15.2 Risk Assessment 49
15.3 Defining the Method and Scope 49
15.4 Validation Strategy 51
15.5 Defining an Acceptance Criteria 51
15.6 Produce a Validation Plan 52
15.7 Competency Requirements of Validator 54
15.8 Comparison of Validation Results Against Acceptance Criteria 54

DRAFT

15.9 Statement of Validation Completion 56
15.10 Implementation Plan 56
15.11 Validation Library 57
16 APPENDIX B: MOBILE DEVICE FORENSICS EXAMPLE 58
16.1 Mobile Device Forensics Overview 58
16.2 Manual Verification 59
16.3 Dual-Method Verification 61
16.4 Mobile Device Forensics – Extraction of Call History Records from Nokia Series 40 Devices 62
17 APPENDIX C: CELLSITE ANALYSIS EXAMPLE – CALL DATA RECORD NORMALISATION TOOL 66
17.1 Introduction 66
17.2 Risks 66
17.3 Validation Requirements 66
17.4 Validation Strategy – Purpose 67
17.5 Validation Plan 67
17.6 Evaluation 67
17.7 Assessment of Uncertainty 68
17.8 Reporting 68
17.9 Other Activities 68
18 APPENDIX D: CELLSITE ANALYSIS EXAMPLE – SURVEY TOOL 69
18.1 Introduction 69
18.2 Risks 69
18.3 Validation Requirements 69
18.4 Validation Strategy – Purpose 70
18.5 Validation Strategy – Limitations 70
18.6 Validation Strategy – Approaches 71
18.7 Survey Methods 72
18.8 Validation Plan 73
18.9 Evaluation 74
18.10 Uncertainty in Reporting Serving Cell Results 75
18.11 Reporting Measurements in Standard Units 75
18.12 Reporting 76
18.13 Other Activities 76
19 APPENDIX E: CELLSITE ANALYSIS EXAMPLE – SURVEY METHOD 76
19.1 General 76

DRAFT

19.2 Risks 77
19.3 Validation Purpose and Requirements 77
19.4 Validation Strategy – Limitations 77
19.5 Validation Strategy – Approaches 78
19.6 Validation Plan 79
19.7 Evaluation 80
19.8 Uncertainty in Reporting Serving Cell Results 80
19.9 Reporting 81
19.10 Other Activities 81
20 VALIDATION GUIDANCE FOR FORENSIC AUDIO & SPEECH ANALYSIS *82*
20.1 General 82
20.2 Format conversion 82
20.3 Case-specific validation 90
20.4 Audio Enhancement 91
20.5 Speaker Comparison 98
20.6 Drawing of Conclusions 100
20.7 Qualifications 101
21 ACKNOWLEDGEMENTS *102*

DRAFT

1 EXECUTIVE SUMMARY

1.1.1 The validation exercise ensures that methods are fit for purpose prior to implementation in a 'live' forensic environment where the true answer is unknown. Validation involves an assessment of the risks associated with use of a method and usually includes an evaluation of the accuracy and precision of it in a controlled environment.

1.1.2 All methods have limitations; there are therefore no perfect methods, whether in digital evidence or in the wider field of forensic science. The purpose of validation is to enable informed choice of the "most appropriate"[1] method.

1.1.3 This document is intended to assist practitioners in the assessment of the limitations of their methods whether that be equipment, software or their own technical and / or interpretive expertise, so that the courts can have confidence in their competence to assess and present legitimate findings.

1.1.4 The onus is on the practitioner to demonstrate the method used is valid, not for others to show it is not; that which can be asserted without evidence can be dismissed without evidence.

1.1.5 Failure to perform validation exercises may result in the provision of incomplete or unsafe evidence, and for this reason validation is a key requirement of the Forensic Science Regulator's Codes of Practice and Conduct (the Codes).

1.1.6 The term validation in forensic science and as used this document is different to verification in software engineering, which is concerned with build quality in software development

1.1.7 Validation in the context of forensic science is focussed on the method not the tool and as such access to the source code is not a requirement; this applies to digital forensics just as much as in other areas of forensic science. For example, in the context of DNA profiling, the code used in the programs running the DNA sequencer is not the focus of the

[1] "most appropriate" could also be defined as "good practice", but no method can be perfect.

DRAFT

validation; not having access to the software code doesn't prevent validation in this field, nor does it in digital forensics. This type of testing is akin to 'black-box testing'.

1.1.8 Validation in forensic science is, wherever possible, through the assessment of known samples. Dual-method, or dual-tool verification is not a substitute for method validation. It may allow further confidence in evidence obtained via either method, but only if they are known to operate independently of one another. Method validation on both tools should be conducted if possible.

DRAFT

2 INTRODUCTION

2.1 Purpose

2.1.1 The validation exercise ensures that methods (see 2.4) are fit for purpose prior to implementation in a 'live' forensic environment where the true answer is unknown. Validation involves an assessment of the risks associated with the use of a method and usually includes an evaluation of the accuracy and precision of it in a controlled environment.

2.1.2 This document has been produced to provide guidance and advice on validation and how to perform it within the digital forensic sciences (digital forensics).

2.2 Scope

2.2.1 This document is intended to assist validation in the field of digital forensic science. Digital forensics as it is commonly termed, covers all scientific and systematic recovery and investigation of material stored digitally. This document may also assist the validation of methods used for analogue aspects that a digital forensic specialist is likely to be involved with. It covers the main disciplines that are deemed relevant. All methods are expected to be demonstrated to be valid, whether covered in this document or not.

2.3 Reservation

2.3.1 Every effort has been made to provide useful and accurate guidance of the requirements contained in the *Codes of Practice and Conduct for Forensic Science Providers and Practitioners in the Criminal Justice System* (the Codes). However, if the guidance supplied here inadvertently implies a lesser requirement than the Codes or ISO/IEC17025:2005 require, then the standard rather than the guidance will prevail.

2.4 Definition of a Method

2.4.1 A method is a logical sequence of operations, described generically for analysis (e.g. for the recovery of specific data from a hard drive, or

DRAFT

assessing the area over which a mobile phone cell may serve) or – more rarely in digital investigations - for evaluative comparison to establish origin or authenticity (e.g. assessing whether data are likely or not given a specific scenario).

2.4.2 For the purposes of validation, methods are classified into three types:

a. standard methods – methods validated by official bodies and recognised as standard;[2]
b. laboratory-developed[3] methods – methods conceived by the end-user requirements of a specific laboratory and validated by the laboratory for use;
c. non-standard methods – methods used by the laboratory once that are unique to a specific case requirement.

2.4.3 Many methods within digital forensics have been described as 'industry-standard' and 'best practice'. However, almost all methods employed by laboratories in this field cannot be considered standard methods as they have not been validated to the required level by an organisation authorised to do so.

2.4.4 If a method has been used many times by one or more practitioners, but it has never been validated, this is no indication that its output is accurate or that its limitations are known. It cannot therefore be 'best practice', or a 'standard method'. Without assessment, there is an unknown risk of incorrect outcomes where it has been used.

2.4.5 The methods used by laboratories are almost always laboratory-developed methods as they answer specific regularly requested needs by combining tools, techniques and expertise unique to the setup of the laboratory. Simplified examples of laboratory-developed methods in digital forensics include:

a. acquisition of a forensic image from a hard disk drive;
b. extraction of text messages from a mobile telephone;

[2] See note in ISO/IEC17025:2005, 5.4.1.
[3] 'Laboratory' is used here to mean the organisation providing the service, be it a police High-Tech Crime Unit or a forensic science provider.

DRAFT

c. normalisation of telephone network data for use in cellsite analysis;
d. removal of 'noise' at a certain frequency from audio recordings.

2.4.6 Non-standard methods are tasks requested of the laboratory that are unique or performed very rarely, differ in scope each time and are not repeatable outside of the exact requirements of the task. The method is still required to be validated under ISO/IEC17025:2005 (5.4.4) and although much of the advice provided here is applicable, this type of validation is not the specific subject of this document.

2.5 Pre-Validation Requirements

2.5.1 If no validation has been previously undertaken in a laboratory on any method, a logical starting point is to perform an analysis of the procedures, techniques and tools already in place and assess how these are used in provision of services to the laboratory's end-users.

2.5.2 The primary end-users of a laboratory's services are often determined by the environment within which the laboratory operates. Typically in digital forensics, laboratories operate within the following environments:

a. a department or unit within a law-enforcement organisation providing forensic services to internal customers within the organisation;
b. a public sector body providing forensic science services to law-enforcement organisations;
c. service providers, independent consultants or sub-contractors providing services to the prosecution, defence or both.

DRAFT

2.5.3 However, the body instructing or paying for the work will rarely be the sole end-user. For example, if the police request work to be performed by a laboratory the results will satisfy their demands, but any reports and evidence produced will be relied upon by other bodies within the criminal justice system. Examples include the prosecuting authorities, opposing counsel and the judiciary.

2.5.4 After identifying the laboratory's end-users, it is then important to determine the services that may be derived from their requirements. This could be performed a number of different ways, e.g.:

a. a review of current methods and processes employed within the laboratory;
b. a review of past requests for analysis, e.g. submissions or letters of instruction;
c. a review of legislation, case law and sentencing guidelines, which may determine what is required from the laboratory evidentially by the courts;
d. discussions with analysts within the laboratory to establish what is most often required of them.

2.5.5 Once the services are defined, the methods used within them can also be identified prior to formal validation.

2.6 Structure

2.6.1 Sections 3 to 13 detail the different steps involved with the validation process. They are based on material from the Regulator's Codes and other sources with particular emphasis (and with examples) on different aspects of digital forensics. These sections aim to provide a brief – and above all accessible – overview of the principles and processes involved.

2.6.2 The appendices provide in much greater detail worked examples of validations applied to specific aspects of digital evidence. Care has been taken to provide different types of validation in the different areas

DRAFT

to provide as complete a picture as possible, while ensuring that the most commonly encountered requirements are also provided.

2.7 Implementation

2.7.1 The Regulator already requires that validation is performed before a method is used in live casework, and that by October 2015, the validation of imaging of conventional hard drives is in the format required in the Codes.

2.7.2 This is a draft of a guidance document circulated for consultation. The requirements are set by the Codes.

3 AN INTRODUCTION TO METHOD VALIDATION IN DIGITAL FORENSICS

3.1 Purpose

3.1.1 The validation exercise ensures that methods are fit for purpose prior to implementation in a 'live' forensic environment where the true answer is unknown. Validation involves an assessment of the risks associated with the use of a method and usually includes an evaluation of the accuracy and precision of it in a controlled environment. Definitions of accuracy and precision may vary according to the discipline to which they are applied, but can broadly be defined as follows.

a. Accuracy is a measure of the closeness of a result to the accepted value, i.e. how close a result is to the true value (for a measurement) or whether all correct answers or matches are provided.

b. Precision is a measure of the repeatability or uncertainty of a test result. As such this is the spread of 'not-exactly-true' values returned and whether there are any answers or matches provided that are incorrect. For example, an assessment could include:

    i. whether a tool provides the same result on the same data when the same tests are run on multiple occasions; or
    ii. multiple tools provide the same result in the same environment.

DRAFT

3.2 Importance

3.2.1 Knowledge of the limitations of a method can:

a. enable the informed selection of the most appropriate technique;

b. mitigate limitations of a given method; and

c. improve efficiency of processes.

Failure to perform validation exercises may result in the provision of incomplete or unsafe evidence; validation is therefore a requirement under the Codes.

3.3 Application

3.3.1 The validation approach may vary according to what is being assessed. For example, whether the output is:

a. factual – absolutes (e.g. the following data were recovered);

b. technically interpreted – where the original output cannot readily be interpreted by a 'layperson'. The competence of the individual interpreting the data must also be included in the assessment; or

c. evaluative – use of a technique to enable an expert to give an opinion on a wider question. The competence of the expert must also be assessed not only in the use of techniques but on their ability to provide opinion (e.g. *"in my opinion, the data are of a type to be expected if …"*).

3.3.2 The validation method will therefore vary according to what is being assessed. For example:

a. for data recovery tests may be performed as to whether a search method recovers all legitimately matching files with no spurious data included;

b. if measurement values are presented, results against a calibrated sample (a sample with a known, externally assured, value) may be compared.

DRAFT

3.4 Challenges and Issues

3.4.1 It is not possible to perform infinite numbers of tests and thus guarantee the legitimacy of output of any method in any circumstances. There will therefore be limitations not just of a method but also of the validation process applied to it. Validation tests should therefore highlight the critical areas where accuracy and precision are required and any limitations, both of the method and the validation applied, must be clearly stated.

3.5 Determining Methods Requiring Validation

For determining the methods that require validation see the Codes, 20.4 to 20.11, and ISO/IEC17025:2005, 5.4.3 to 5.4.6.

3.6 Risk Assessment

3.6.1 An appropriate risk assessment is at the core of any validation requirement. The risks dictate the focus of the validation exercise. For example, the risks associated with a data recovery method for hard disks may be that it:

a. fails to recover all data present (including deleted data not yet overwritten);

b. appears to recover data that are not actually present (e.g. it may 'bleed through' data from a previous extraction in the results); and

c. changes data or corrupts the original evidence ('source') on the hard drive.

3.7 Validation Requirements and Acceptance Criteria

3.7.1 The validation requirements of a given method will depend on the risks and the output required of it. These should be defined at the outset of any validation, highlighting:

a. those features that must be tested as critical findings depend on them;

b. those features that have lesser importance but may be assessed as well; and

DRAFT

c. any issues expected or detected (including potential mitigation) for them.

3.7.2 For example, a computer forensics laboratory may wish to use a new method to detect, recover and produce e-mail messages from computers. The method comprises sub-methods depending on the type of e-mail message, the operating system of the computer and the software or script used to extract and present the message as evidence. The validation requirements should include the full range of activity required of the method and include the acceptance criteria required.

3.7.3 Requirements will vary according to the complexity or novelty of a method.

a. A new method will require comprehensive testing (to include the assessment of both the equipment or software and the approach taken when using it) to provide assurance that it is fit for purpose. It may be sufficiently novel to benefit from being published in a journal for dissemination.
b. New equipment, software or approaches applied to a pre-existing method may require testing targeted on specific aspects of it that inform critical findings.
c. An update to a method that has already undergone validation, where the method has not changed significantly, will not necessarily require full re-validation and a verification can be performed.

3.7.4 A validation will take the form of one or more tests of each of the requirements. A single test of a method in and of itself does not mean that a method is validated.

3.8 Previously Validated or Adopted Methods

3.8.1 If another organisation has validated a method, complete re-validation may not be necessary. To rely on the validation of others, some criteria need to be met.

a. Is the method described in sufficient depth to implement and is the end-user requirement the same as yours?

DRAFT

b. Is the validating organisation trustworthy for the validation or does it have a vested interest? (For example, is it the manufacturer of the equipment?[4] Is it a known independent and recognised authority or is it as yet an untested source? Is it known to lack credibility?)

c. Is there access to the validation report (including, if possible, the supporting data) for detailed review?

3.8.2 Where these criteria are met, all this documentation should be included in the validation records (see the Codes, section 20.16. Validation library. Verification will still be necessary.

4 PLANNING VALIDATION

(The Codes, 20.4 to 20.11, ISO/IEC17025:2005 5.4.3 to 5.4.6)

4.1 Defining Requirements

4.1.1 Prior to undertaking tests, an expectation as to how the method is expected to perform (potentially based upon advertised functionality or practitioner experience) should be defined. This definition should include acceptance criteria and whether specific capabilities are mandatory or desirable.

4.2 Validation Strategy and Plan

4.2.1 Once the requirements are defined they should be used to inform the approach taken for validation (i.e. the strategy). The **strategy** is an overview of the whole validation process and forms an outline of the **plan,** which is a series of discrete, achievable and measurable steps, each part of the process defining the specifics of the data used and the expected outcome. The strategy/plan should define the: following.

a. Equipment, software or process under review.

i. This should include all relevant details including the manufacturer and the versions of hardware, firmware and software.

[4] Validations by manufacturers are sometimes called developmental validations and can be very useful. However, the method, depth, rigour, relevance and level of peer review (e.g. whether it has been published in a journal relevant to the discipline) should be assessed before relying on the study.

b. Type of result being assessed:.

   i. Whether the method is factual, technically interpreted or opinion.

   ii. A technically interpreted method will probably also require an assessment of the validity of the factual output of equipment as well.

   iii. Likewise, when a method encompasses opinion, the technical interpretation and factual outputs that form parts of the overall process may also require assessment.

c. Source, quantity and reliability of data used for the tests.

   i. If data recovery assessments are being performed, a review of the source and type of data used should be undertaken; this should include whether the data are likely to provide problems for the system being assessed (i.e. whether the data enable a 'stress test'). For example, this could include non-standard character sets, formats, file locations or volumes of data.

   ii. If measurements involving standard units are being performed, the provenance and accuracy of the source (the traceable standard) should be established.

   iii. If technical interpretation or opinion assessments are being performed, blind trials may be used in addition to the other tests.

   iv. Blind trials should focus on non-obvious situations where a failure to assess correctly is a real prospect.

   v. If there is little or no control of the source data, this should be explicitly declared in the plan and the subsequent limitation declared.

d. The expected outcome for the tests performed, to include consequences or next steps if the expectations are not met. Expected outcomes should be wherever possible specific,

DRAFT

quantifiable and highlight the acceptable error margin (i.e. the defined accuracy and precision required of the method).

e. Limitations of the tests performed. For example, a limited data set has been used, or the data may potentially change with time.

4.3 Undertaking Validation

(The Codes, 20.4–20.11, ISO/IEC17025:2005, 5.4.3–5.4.6)

4.3.1 Once the requirements, strategy and plan have been defined the tests can be performed.

4.4 Further Guidance

4.4.1 Examples of specific validations are presented in the appendices. These provide examples of approaches to validations including problems encountered and how they were resolved.

4.5 Generation and Control of Test Data

4.5.1 The data relied on for validation are of critical importance. For example, a search or data recovery method may require bulk known data to access. These data should include the following.

a. Data or character types known to have caused problems with other tools, and should encompass wherever practical, all of the data types that the tool is envisaged to be required to work on.

b. A sufficient quantity of data to provide a real test of the process.

4.5.2 This is known as stress testing. It is not always possible to define the source data completely. For example, in cellsite analysis if a survey tool is used on a new technology (e.g. 4G) it is extremely unlikely that a validating organisation will have access to a controlled environment (e.g. a single known cellsite isolated from the wider radio environment). In such cases consistency of output between independent devices and blind trials may be the only assessment techniques practically available. If blind trials are undertaken, they should not all take place where the outcome is entirely predictable (e.g. in the cellsite example given, close to and directly in front of a particular cellsite).

DRAFT

4.5.3 Data created for and/or generated during the validation should be stored for later audit, if required.

4.6 Evaluation

4.6.1 Contemporaneous notes should be taken during evaluation exercises.

4.6.2 A note should be made for each test in the plan as to:

a. who undertook the test;

b. when the test took place;

c. what the test assessed;

d. what equipment was used;

e. the expected outcome;

f. what the results were; and

g. any other appropriate information (e.g. the raw results or a link to them and where the test was performed, if this may affect findings).

4.6.3 Each test in the plan should be carried out and the result compared with the expected outcome (i.e. the actual result versus the expected or acceptable outcome). An assessment as to whether the method has passed or failed each of the tests should be made. If a method fails an individual test, it may be possible to:

a. highlight methods by which to detect or mitigate the failure; or

b. re-assess whether the specific capability that failed the test is mandatory or desirable (i.e. whether the failure of the aspect tested should result in the entire method being discredited).

4.6.4 Consideration of uncertainty. Testing should not be limited to a single attempt. In assessment of a method, precision as well as accuracy should be taken into account. This can only be achieved by repeating tests, which can include:

a. different equipment run on the same data/in the same environment at the same time;

DRAFT

b. the same equipment on the same data/in the same environment at different times;

c. checks for bleed through of data from previous searches (perform search on large data set followed by search on smaller data set);

d. checks and assessment of a possible dynamic environment.

4.6.5 The range of results should be summarised and recorded. 'False positive' (when an answer known to be incorrect is output by the method) and 'false negative' (when an answer known to be correct is failed to be output by the method) should also be explicitly highlighted:

a. the observed precision of method versus the range of acceptable outcomes should be highlighted;

b. the competence of the practitioner planning, performing and assessing the validation should be defined.

4.6.6 Any deviation from the plan, along with the reason for this, should be noted. Within the contemporaneous notes, the findings should be summarised to include the following.

a. The original requirement for each test and a summary of the findings.

b. Whether the method meets the original requirement:

i. any areas in which the method fails to meet the requirement should be explicitly highlighted;

ii. any limitations of the validation approach and the method itself.

5 CONCLUDING VALIDATION

(The Codes, 20.12–20.17)

5.1 Validation Report

5.1.1 A report should be constructed that details the validation process performed. This should include the following.

a. The original requirement.

DRAFT

b. Reference to what is, and is not, validated.

c. A summary of the strategy, tests performed and the outcome of each test.

d. Reference to the data used and any limitations these may have on the tests performed.

e. Whether the method is fit for purpose:

   i. this should state whether the method is fully approved, partially accepted or not recommended for use.

f. Recommendations for use:

   i. to include any limitations of the method, the impact of these limitations and any additional steps required to detect and mitigate for them; and

   ii. define the required on-going quality regimen (e.g. quality assurance tests).

g. Effect of new approach/technique/equipment on existing methods:

   i. whether existing methods become obsolete and should be superseded or whether the method should be used as an alternative or in parallel.

h. Reliability of the validation process including any uncertainty in measurement encountered and the impact this may have.

5.2 Statement or Certificate of Validation Completion

5.2.1 The statement or certificate of validation completion should be a short (one or two page) summary of the validation report detailing what the method is and whether it is fully approved, partially accepted or not recommended for use. The certificate should highlight who is making the recommendation, their role (i.e. whether they are qualified to make such an assessment) and the date of implementation. The assessor should be independent from those undertaking the validation study.

DRAFT

5.3 Implementation

5.3.1 Once a method has passed validation and is approved for use, there will be further activities required before it can be used on live casework. These activities should include the following.

a. Training plan for users:
   i. competency requirements and testing.
b. Guidance for use:
   i. a technical handbook for the equipment;
   ii. inclusion of the method in quality systems;
   iii. on-going quality assurance should be defined.
c. Inclusion in existing systems (e.g. equipment logs, competency records, quality system).

6 POST-VALIDATION ACTIVITIES

(The Codes, 20.18., ISO/IEC17025:2005, 5.4.7)

6.1 Maintenance of Documentation

6.1.1 Reference to the validation may be included in quality documentation and the report should be included in the validation library held by the organisation performing it. There may also be links to other requirements that are not directly concerned with validation, e.g. equipment logs detailing changes in use. The documentation should be updated as new versions of equipment/software are tested and implemented.

6.2 Quality Assurance

6.2.1 On-going testing is recommended to ensure the continued correct operation of equipment. The test, expected result (with a range of acceptable results) and the frequency required should be defined and included in the training/equipment guidance documentation.

6.2.2 For example, in cellsite analysis if equipment is installed in vehicles, tests as to whether cables may have come loose may be advisable each time the vehicle is deployed. Alternatively, results involving measurements may be assessed against known traceable standards on

DRAFT

a calendar basis (e.g. daily/weekly/monthly). Repeating the same examination by different analysts is another approach that can be adopted, and assessing any differences may result in better working practices being identified.

6.3 Acceptance Testing of New Equipment

6.3.1 If new equipment of the same design (manufacturer, version) is purchased, acceptance testing may be required prior to the equipment being placed in service. This may be nothing more than running a quality assurance test and may form part of the equipment log.

6.4 Review of Updates to Equipment or Software

6.4.1 It is in the nature of digital forensics for updates of software or equipment to be fairly frequent. According to the nature of the update (e.g. whether it is a minor change in capability) additional assessments may be required.

a. Full validation may be required if there have been significant changes across the equipment/software.
b. Partial validation of the new functionality may be required if there is additional capability but the core capability remains unaltered (in addition, an acceptance test may still be required for the unaltered aspects as a safety check).
c. An acceptance test (or quality assurance test) may be required if the change is merely an update (efficiency saving, etc.).

6.5 Post-Project Review

6.5.1 A review of the validation process undertaken may be advisable to assess whether there are lessons to be learned for future validation exercises. For example, there may be data that can be used for other exercises (thus making the later exercises more efficient) or other resources may have been identified that may be of future use. These are not required for validations per se, but may be of benefit to organisations regularly undertaking such activity.

DRAFT

7 ASSESSING UNCERTAINTY IN DIGITAL FORENSICS VALIDATIONS

7.1.1 Forensic science is science applied in the service of the courts. Within digital forensics, there may be many fields employed including traditional sciences such as chemistry and physics but also areas such as computer science and statistics. There are different definitions of accuracy and precision according to each of these fields in which they are employed. This can result in the concepts being difficult to apply to digital forensics.

7.1.2 Uncertainty of measurement is a parameter associated with the result of a measurement that defines the acceptable tolerance bounds of the value relative to the error between the required and actual measured quantity. Its overall value is calculated by combining all relative uncertainty components and typically requires an associated confidence level in order to quantify the sampled data's potential error distribution.

7.1.3 For example, in cellsite analysis one approach for evaluating evidence is by applying the case assessment and interpretation model, as used in more traditional areas of forensic science; i.e. an the assessment as to whether the observed data are likely or unlikely given the alternative prosecution and defence propositions under consideration. In order to test these hypotheses, it is essential that survey measurements of the mobile phone cells are undertaken at several locations at and around the relevant locations, so that measurement uncertainties can be estimated and taken into account in the evaluation.

7.1.4 As part of a wider validation process, different types of equipment and a range of different methods have been assessed for consistency and against known expected outcomes, with particular focus on false exclusion measurements.[5]

7.1.5 Estimating the uncertainty of measurement can prove challenging in other fields of digital forensics. An assessment of digital evidence from computers and mobile phone devices often differs from that presented

[5] **Tart, M. Brodie, I. Gleed N. Matthews, J.** (2012) 'Historic cell site analysis – Overview of principles and survey methodologies'*, Digital Investigation* (8) 3–4, pp 185–193.

in other forensic disciplines as most often it is trying to establish an artefact's presence instead of a value or measurement. This presents difficulties in calculating a value of the uncertainty of this process as is the case in many pure scientific disciplines and the United Kingdom Accreditation Service has acknowledged this in granting ISO/IEC17025:2005 accreditation for services provided in this area.

7.1.6 An assessment of uncertainty is given in each of the appendices, but in broad terms 'accuracy' and 'precision' can be defined as follows.

Accuracy

7.1.7 The closeness of agreement between the mean of a set of results or an individual result and the value that is accepted as the true or correct value for the quantity measured:

a. Accuracy is a measure of whether or not a true answer is returned in the range of results from a method (and if not, how close the result is to the true answer).
b. For example, in an assessment of a search method in computing, this could be equated to whether all matching data are returned in a search (i.e. whether any matching data are not returned in a result).
c. In another example, in an assessment of a method for measuring voltage, this could be equated to how close the measurement is to the true value.

Precision

7.1.8 Precision is synonymous with reproducibility or repeatability, whereas accuracy is about obtaining the true or correct value for the quantity measured. An incorrectly calibrated device may be capable of giving reproducibly precise readings even though the data generated are not accurate.

a. Precision is a measure of the uncertainty of the result, the type or range of results provided that are not *exactly* the true answer.
b. In an assessment of a search method in computing, this could be equated to whether there are additional (spurious) data returned in a search that are *not* matching data.

DRAFT

c. In an assessment of a method for measuring voltage, this could be equated to the range of measurements returned for a single true value (i.e. the uncertainty in any individual reading).

8 COMPETENCY

8.1 Introduction

8.1.1 Assessment of a method involves both the validity of the technique and the competency of the practitioner (both initial and on-going). As such, the 'human factor' needs to be accommodated into any method validation as the practitioner is part of the method.

8.2 Technical Skills

8.2.1 If a method is to be deployed without any interpretation (i.e. is a set of reproducible steps, none of which require a wider competence) then competence assurance can be limited to an assessment of whether a method is correctly applied by a practitioner.

8.3 Technical Interpretation

8.3.1 If a method is to be deployed where the result is not obvious to a layperson, technical interpretation will be required. The competence of the individual must be assessed to:

a. select the method;
b. apply the method; and
c. correctly interpret the output of the method.

8.4 Evaluative Opinion

8.4.1 Competence in the use of technical methods does not in itself provide any assurance that the output can be correctly interpreted when applied to a wider scenario or question. In particular, opinion evidence (when a method is used to shed light on whether the evidence is expected given a specific activity) is prone to a range of additional concerns in addition to those concerning the validity of the method used. Competence in forensic interpretation (evaluative evidence) must be explicitly assessed if a practitioner is to produce opinion evidence. This would be in addition to validation exercises for a technical method.

DRAFT

8.4.2 One concern is that of defining and assessing inappropriate or misleading questions, a possible problem that would not be addressed in any technical validation exercise or through the purely technical competences of the practitioner.

8.4.3 For example, if comments are made on the likelihood of the scenario rather than the evidence, this is an example of a known and much-documented failing known as the 'prosecutor's fallacy'. Amongst the concerns with this approach is:

a. it is easy to make mistakes, as assessments that appear to be equivalent frequently are not;

b. it should be made clear what is 'expert' opinion and what is considered to be common sense; and

c. it may have an impact on the duty of the jury rather than that of the expert.

This is discussed further in Section 13 'Consequences of Failure to Validate'.

DRAFT

9 CHECKLIST

The following checklist highlights the main steps required for a validation exercise.

Documented Evidence Required (Validation Library)	Task	Sub-task	Reference
User Requirement	Define user requirement		4.1
	Risk assessment	Negative outcomes if method provides incorrect output	3.6
	Novel technique	Full validation	3.7, 3.8,10
	New version of existing technique	Acceptance testing ('verification') of new version of previously validated method	3.7
	Previously validated technique	Review evidence of validation	3.5, 3.7
Validation Strategy	Produce validation strategy	Define validation approach taken	4.2
	Define acceptance criteria	Quantity, variety and types of tests employed	3.7, 4.1
		Definition of accuracy and precision used or other applicable technical quality standards	8
		Highlight limitations of validation approach taken	4.2, 5.2
	Assessment of uncertainty	Define test undertaken	4.2
		Define data set used	5.2
		Define 'true' answer and acceptable deviation from it	4.2, 5.2

DRAFT

Documented Evidence Required (Validation Library)	Task	Sub-task	Reference
	Undertake tests	Make contemporaneous notes	4.5
		Define tester, equipment used with version, expected outcome, actual outcome	4.2
Validation Report	Produce validation report		5.1
	Validation certificate	Independent internal review of validation material	5.2
	Publication of findings?	Independent external peer review if method is novel	
Implementation Plan	Produce implementation plan		5.3
	Training plan		5.3
	Competency assessment		8
	Guidance for use	Produce document	6.2
	Inclusion into quality systems	Produce SOPs (standard operating procedures)	7.1
	Post-implementation review	(optional, but good practice)	7.5
On-going use	Quality assurance testing regimen	Include in SOPs	7.2
	On-going competency requirements	Include in SOPs	7.2

DRAFT

10 VALIDATION REQUIREMENTS FOR NEW METHODS FROM THE PERSPECTIVE OF THE COURT

(The Codes, 20.15)

10.1.1 The ultimate end-user for forensic science is often the court, and for innovative science to be used for the first time in a prosecution, the Crown prosecutors must be able to answer positively the following three questions, using documentary evidence included in the validation library.

a. Can the evidence be used in court?
b. Is the evidence reliable?
c. Is the evidence relevant?

10.1.2 The Criminal Procedure Rules (CrimPR) 2014 requires that the expert's statement explicitly provides information to assist the court in determining whether the evidence should be admissible (33.4.h).

10.1.3 To support this change, the Lord Chief Justice of England and Wales has amended the Criminal Practice Directions, providing the following factors which the court may take into account in determining the reliability:

a. the extent and quality of the data on which the expert's opinion is based, and the validity of the methods by which they were obtained;

b. if the expert's opinion relies on an inference from any findings, whether the opinion properly explains how safe or unsafe the inference is (whether by reference to statistical significance or in other appropriate terms);

c. if the expert's opinion relies on the results of the use of any method (for instance, a test, measurement or survey), whether the opinion takes proper account of matters, such as the degree of precision or margin of uncertainty, affecting the accuracy or reliability of those results

DRAFT

d. the extent to which any material upon which the expert's opinion is based has been reviewed by others with relevant expertise (for instance, in peer-reviewed publications), and the views of those others on that material;

e. the extent to which the expert's opinion is based on material falling outside the expert's own field of expertise;

f. the completeness of the information which was available to the expert, and whether the expert took account of all relevant information in arriving at the opinion (including information as to the context of any facts to which the opinion relates);

g. if there is a range of expert opinion on the matter in question, where in the range the expert's own opinion lies and whether the expert's preference has been properly explained; and

h. whether the expert's methods followed established practice in the field and, if they did not, whether the reason for the divergence has been properly explained.

10.1.4 The Criminal Practice Directions could be considered to clarify the end-user's requirement (i.e. the courts in England and Wales) for the method to be valid, as well as certain features to be teased out in the validation. Time will tell how courts will interpret the directions and whether the other jurisdictions in the United Kingdom will adopt or give cognisance to them.

10.1.5 In addition, the courts need a clear explanation of the novel science and any surrounding issues and limitations in its use. This should be provided as a two-page 'Strengths and Weaknesses' summary document, written in plain English that can be readily understood by lawyers, judges, jurors and other non-scientists within the criminal justice system (CJS).[6] Points addressed should include the following.

[6] **Tully, G., Sullivan, K., Vidaki, A. and Anjomshoaa, A.** (2013) *Taking Forensic Science R&D to Market*, Forensic Science Special Interest Group. Available at: www.tinyurl.com/FoSciSIG [Accessed 12/09/14].

DRAFT

a. Has the science been validated, and if so, what is the evidence of the validation?
b. Is the service provider accredited for carrying out the test, and if so, what is the evidence of accreditation?
c. Does the forensic service provider comply with the Codes, and if so, how is compliance demonstrated?
d. Is the individual who carried out the test competent to do so, and how is this competence evidenced?
e. Caveats about the use of the method.
f. Approved uses of the method, which could be by case type or exhibit type.
g. Circumstances in which the use of the method would be inadvisable.
h. Additional work that should be undertaken in combination with the result.

DRAFT

11 VALIDATION AND CALIBRATION ASSESSMENTS FOR A LABORATORY

11.1 Starting the Validation Process

11.1.1 Attempting to look at an entire set of processes as a single object will, unless they are very simple, make it nearly impossible to develop an effective method validation policy.

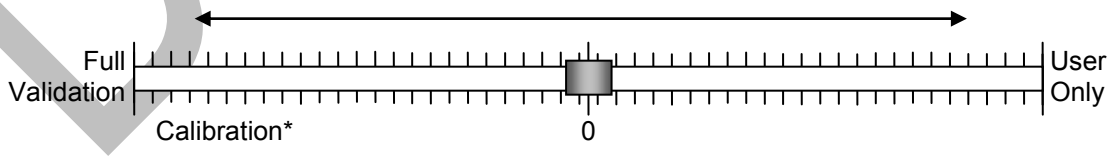
11.1.2 A good rule to observe is to subdivide [atomise] processes down to a level that will enable more accurate determination of the requirements of each part.

11.1.3 The subdivision process itself should be an iterative-based method that, when completed, must be assessed and then repeated at least once.

11.1.4 The reason for doing this is that the first run will be based on existing assumptions. It may very well result in the re-evaluation of some of the original assumptions if they do not appear exactly as expected.[7]

11.1.5 A good starting point is to see if the seemingly different blocks of processes can be grouped into unique sets.

11.1.6 Once this has been done they can be associated along a sliding scale as to whether it is believed they are fully automated processes, or processes that are wholly dependent on the person carrying out the procedure.



* It may be thought that calibration and validation are two separate processes, but in reality a calibration is a simplified subset of validation. A number of items may require calibration in order that a validation requirement can be satisfied. In addition practitioners will still require competence to use the calibrated equipment (which in itself is part of the validation requirement set).

[7] The assumptions that did pass scrutiny have probably already previously been through a similar local process of evaluation, but the process has been so automatic that the user may not have realised that they had done so.

DRAFT

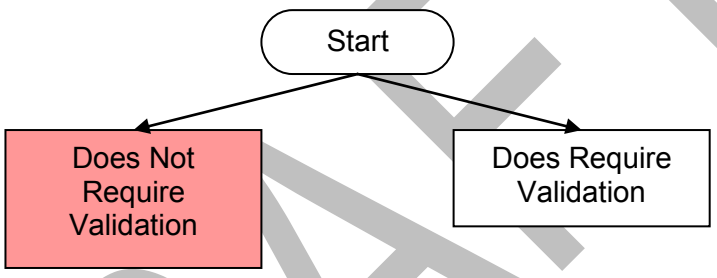
11.1.7 The primary concept to keep in mind is that it is highly unlikely that any process can be associated with either extreme of the sliding scale.

11.1.8 A fully validated process still requires human interaction to interpret the results, and equally, all users will normally have to rely on at least one generated result for at least one process.

11.2 An Example of Determining the Validation Level

11.2.1 If it is unclear how to start the process then a useful approach may be as follows.

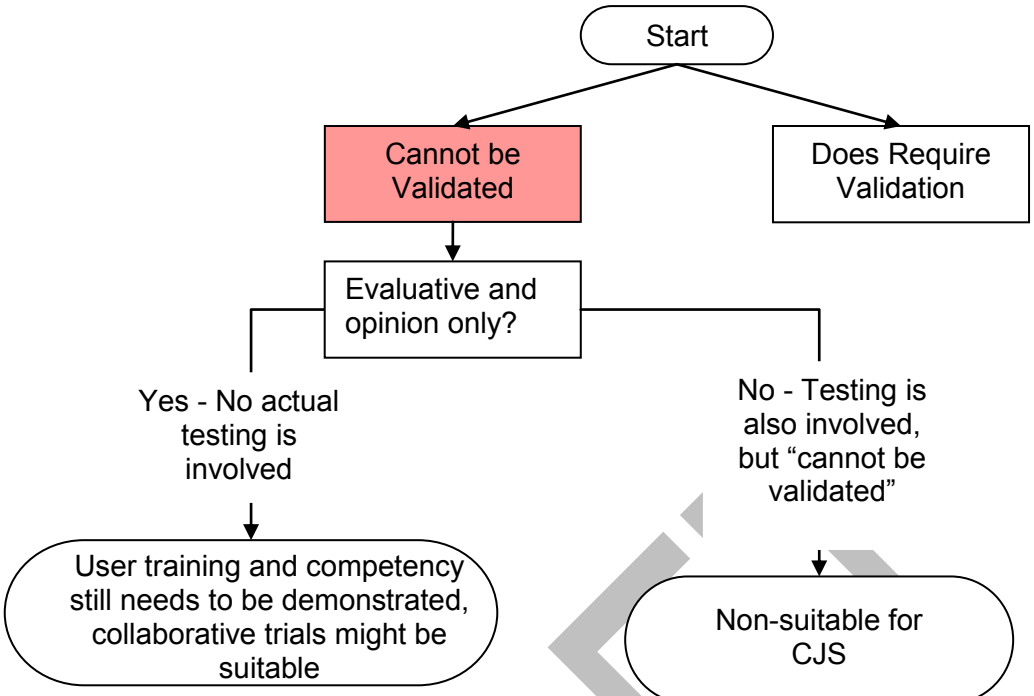
a. Subdivide all processes into the following two groups.



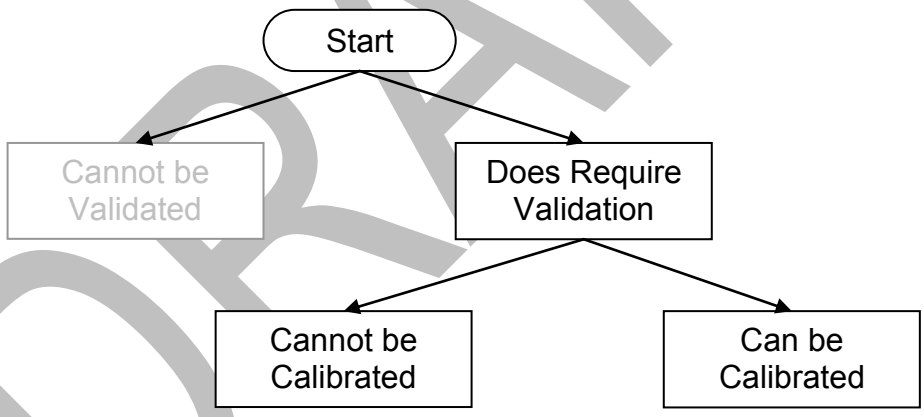
b. If all processes fall under the group '*Does Not Require Validation*' then the work conducted is either not suitable for court use or the assessment has not been competently undertaken and the process must be repeated.

c. For instance, if it is believed that nothing requires validation as all processes are wholly dependent on the expertise of the analyst then it has been forgotten that the user training and competency assessment process will require a very detailed validation design and implementation plan that must also be demonstrated. A more thorough model may be as follows.

DRAFT



d. Once there is a list of processes that require validation then they may be split into one of the following subgroups.



e. If processes can be applied to '***Calibration***' then this will generally simplify validation processes as the calibration can be carried out by a United Kingdom Accreditation Service (UKAS) accredited body, which although costing money saves the production and maintenance of a detailed validation plan.

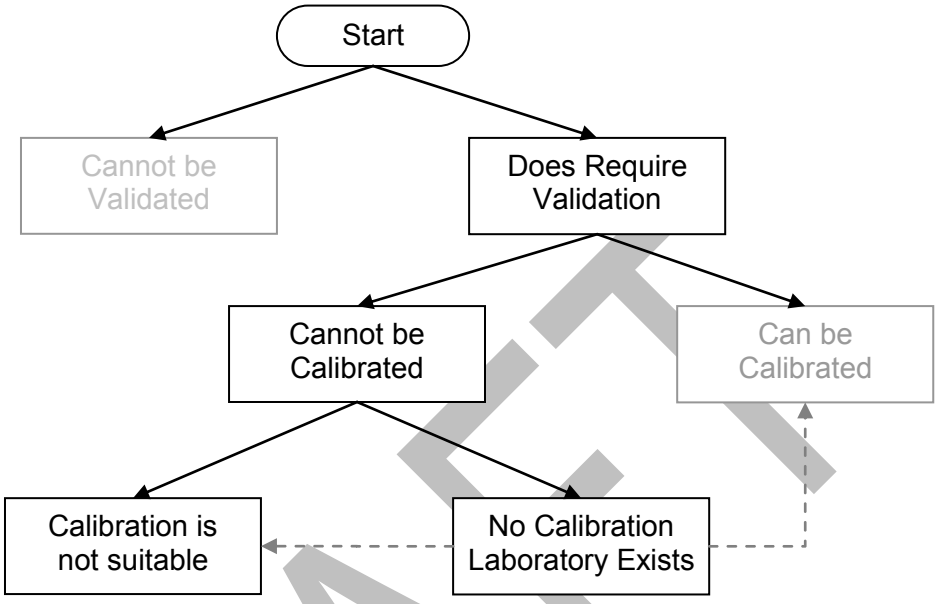
f. All that needs to be specified is effectively '***Send item for calibration n times a year***'.[8]

[8] The calibration interval may be derived in part from the manufacturer, but the end-user would need to assess how their implementation might influence performance drift and therefore calibration interval.

DRAFT

g. If it cannot be calibrated then a detailed validation methodology must be devised to show that the method is fit-for-purpose.

h. The next useful stage of subdivision is as follows (again only two subgroups).



i. At the end of the review process the table should be populated, see below.

No Validation Required	Validation Required		
	No Calibration	Internal Calibration	External Calibration
Word processor used to write reports and statements. General computer monitors, keyboards, etc.	Staff training, software analysis, indication only. Only previously calibrated kit will need to be labelled with '***Indication only***' stickers	Bespoke calibrations	Multimeters, oscilloscopes, bench power supplies, electronic filters, analysis machines, electromagnetic compatibility cages. All kit that was purchased with a calibration certificate.

DRAFT

12 CONSEQUENCES OF FAILURE TO VALIDATE – COMPUTER ANALYSIS

12.1 Introduction

12.1.1 The examples provided may be focused on specific areas of digital evidence, but the principles provided apply to all areas.

12.2 Sole Reliance on Case-by-Case Quality Assurance Procedures

12.2.1 It may be tempting to suggest that quality procedures implemented during the provision of casework (such as dual-tool verification and peer review) are adequate to demonstrate that the methods used are legitimate. The 'true' answer is unknown during active casework.

a. Dual-tool verification is a process that checks that one tool is producing the same results as another from the same exhibit on a case-by-case basis. However, both tools may share some of the same source code or libraries and could therefore produce the same erroneous results (i.e. they may be essentially the same tool with a different user interface). Unless the tools can be demonstrated to be truly independent there is no assurance that any correlation between outputs means that the results are legitimate. Validation of one or both tools can be undertaken to show this.

b. Peer review is an important tool for checking analyst competence, consistency of usage of methods, and error trapping on a case-by-case basis. However, peer review cannot assess whether the method used is producing reliable, repeatable results. The only assurance given for the methods used is that obvious errors or omissions from a method may be detected in the check.

12.3 Validating the Tool Rather Than the Method

12.3.1 It is a method that produces the results, a tool is only part of a method. For example, a write blocker is a device that allows a storage device from an exhibit to be connected to a forensic examiner's computer, preserving evidential integrity during preview or forensic imaging. It is

DRAFT

prudent to validate that the write blocker is not malfunctioning, e.g. allowing data to be written back to the storage device or corrupting data as they are read through it. However, if this is the only part of the forensic imaging method that is checked or validated, it cannot be known whether consistent and full results are produced on each occasion. It is therefore important to validate the entire forensic imaging method, from the continuity and handling of the original exhibit through to the production of a verified set of forensic images for analysis, and including all intermediate steps.

12.4 Validating According to a Laboratory's Audit Schedule

12.4.1 Due to the reactive nature of casework it is often difficult to find time to review validation requirements. If a laboratory's requirements are not reviewed on a regular basis and only approached when there are impending deadlines to meet (e.g. the visit of an auditor) this could impact on the provision of up-to-date, fully validated services that a laboratory can offer. Examples of when requirements could change are as a result of a:

a. new release or major update of a mainstream operating system or software application, changing the way artefacts are represented on computer exhibits submitted to the laboratory;

b. new release or major update of a forensic tool, with the ability to identify artefacts from new or updated software.

DRAFT

13 CONSEQUENCES OF FAILURE TO VALIDATE – CELLSITE ANALYSIS

13.1 Introduction

13.1.1 The following are intended as examples of the risks associated with incorrect, or absence of, appropriate validation for a whole method (i.e. a method including both the technique used and the competence of a practitioner in the interpretation of the output).

13.2 Absence of Evidence Equals Evidence of Absence?

13.2.1 If a cell has not been detected as serving at a location, this may indicate that it is unlikely (or impossible) that a phone may have been there at the time of phone activity. However, if no assessment has been made as to the likelihood of false exclusions (false negative results, where a legitimately serving cell is not detected by a method) through the validation of the method used, the failure to detect a cell at a location may not, in itself, indicate that the phone could not have been there. It is unknown if the negative result provided by the method is itself legitimate.

13.2.2 In summary, failure to perform validation to assess false negative reporting can, and has, caused issues at court.

13.2.3 If asked what evidence there is to support an opinion, comments such as: “It matches because I say it matches” or “The evidence is **my** opinion” are both unhelpful and may indicate the method used (as applied by the practitioner) is not robust.

13.2.4 There is uncertainty in all areas of forensic science. The purpose of validation is to provide a level of assurance that the limitations of a technique are known and have been assessed prior to use. If an expert is unable or unwilling to explain how a result or conclusion is reached and what the limitations of the result/their opinion are, this may indicate that the expert is actually ignorant of the limitations of both the method and their own competence. As such the approach taken and their resultant findings may be unsafe.

DRAFT

13.2.5 In summary, the limitations of the equipment, process and competence of the practitioner are required to be defined and explicable to a court. This is best achieved with reference to validation. If this assessment has not been completed then this limitation, or caveat must be must be communicated to the investigating officer and ultimately the court.

13.3 Provision of Opinion Without an Interpretation Framework

13.3.1 Cellsite analysis can be defined as the process of inferring an area where a mobile phone may have been at the time of activity. This is predominantly a technically interpretive exercise, i.e. a number of methods may be used to reach a technical judgement of whether data would be expected if a phone were at a specific location when used.

13.3.2 These technical methods can and should be validated. However, a practitioner using the output of validated methods to provide evaluative (opinion) evidence does not in itself provide any assurance that the practitioner's opinion is correct. Examples of evaluative exercises in cellsite analysis include assessments of whether, given a call data record, that record would be expected if a specific sequence of events took place, or whether a person may have been the user of a phone that is contested. Without awareness of the difference between technical and evaluative opinion, a cellsite analysis practitioner may stray into areas beyond their expertise without knowing it.

13.3.3 Comments such as *"It is the most likely location for them"* and *"It is more likely someone else used this phone"* without framing the range of alternative scenarios considered may indicate a lack of competence in providing evaluative opinion (i.e. awareness of evidential issues beyond the use of technical methods) and findings may be unsafe as they are assessments of the scenario rather than of the evidence given the scenario.

13.3.4 Comments in expert evidence should be limited to whether the data are expected (i.e. the evidence in the call data records involving cells used at the times of activity) given the scenario presented as, by the definition as given above, this is the expertise of the cellsite analysis

DRAFT

practitioner. Comments on whether the scenario would be expected given the evidence may appear to be the same but they are not. For example, if a cell demonstrably provides service over an area that includes a location of interest, comments such as: *"The data are of a type to be expected if the phone were at the location of interest"* are valid and are not the same as: *"The phone is expected (or likely) to have been at the location of interest"*. The phone could be *anywhere* in the service area of the cell and, for a single call, there is no way of knowing specifically where within the service area of that cell the phone was. The phone could be at the location of interest, next door to it or significantly distant from it (although still in the service area of the cell used) and the evidence (the cell in the call data record) would be the same in each case, so no assessment as to which of these scenarios is more likely can be made.

13.3.5 In addition, competence in inferring where a phone may have been at the time of activity does not automatically translate into wider expertise in everything else to do with phone usage (e.g. social behaviour). As such a cellsite analysis practitioner as defined above cannot offer any greater expertise than a juror in matters such as (but not restricted to):

a. whether movement of a phone or a person is likely or unlikely;
b. how many other randomly selected people may have moved from one area to another in the same period as the phone under consideration.

13.3.6 While comments can be made to highlight relevant wider information so that others – e.g. the jury – can take a view, assessments that could be considered 'common sense' should be separated from assessments that are expert opinion.

13.3.7 In summary, the competence of the practitioner in forensic assessment and interpretation requires validation in addition to the technical methods used by them.

DRAFT

14 GLOSSARY

Accreditation

Third-party attestation related to a conformity assessment body conveying formal demonstration of the forensic science provider's competence to carry out specific conformity assessment tasks.

Accuracy

The closeness of agreement between the mean of a set of results or an individual result and the value that is accepted as the true or correct value for the quantity measured (see also **precision**).

Blind Trial

A blind trial is when the outcome of the test is known by someone other than the person performing it.

Calibration

The set of operations that establish, under specified conditions, the relationship between values indicated by a measuring instrument or measuring system, or values represented by a material measure, and the corresponding known values of a **measurand**.

[The] Codes

The *Codes of Practice and Conduct for Forensic Science Providers and Practitioners in the Criminal Justice System*, published by the Forensic Science Regulator in 2011. Available from: https://www.gov.uk/government/publications/forensic-science-providers-codes-of-practice-and-conduct.

Competence

The skills, knowledge and understanding required to carry out a role, evidenced consistently over time through performance in the workplace. The ability to apply knowledge and skills to achieve intended results.

DRAFT

Contamination

The undesirable introduction of substances or trace materials.

Criminal Justice System

The criminal justice system (CJS) is the collective term used in England and Wales for the police, the Crown Prosecution Service, the courts, prisons and probation, which work together to deliver criminal justice.

Customer

Whether internal or external, it is the organisation or a person who receives a product or service (e.g. the consumer, **end-user**, retailer, beneficiary or purchaser).

Databases

Collections of information designed to provide information rather than for archive, which are stored systematically in hard copy or electronic format and are, e.g. used for:

a. providing information on the possible origin of objects or substances found in casework; and/or

b. providing statistical information.

End-user

The end-user of forensic science is the **criminal justice system**, essentially the courts. A **method** or tool may not be directly used by the courts, but it is assumed that the results will be.

Evidence

Anything that may prove or disprove an assumption to be true, e.g. an exhibit or the lack of expected findings.

Evidential

The Crown Prosecution Service applies an evidential test to decide whether there is enough **evidence** to prosecute and importantly whether the evidence is reliable and can be used in court.

DRAFT

Exculpatory

Exculpatory **evidence** is broadly favourable to the defendant.

Expert (Witness)

An appropriately qualified and/or experienced person familiar with the testing, evaluation and interpretation of test or examination results, and recognised by the court to provide live testimony to the court in the form of admissible hearsay evidence.

False Positive/False Negative

A False Positive is the inclusion of a result in an output that is incorrect. A False Negative is the exclusion of a correct result from an output.

Five by Five by Five (5x5x5)

The five by five by five refers to an intelligence report/product, and is part of the **National Intelligence Model**. Each five refers to a grading of the evaluation of the source, **intelligence** and a handling code.

Inculpatory

Inculpatory **evidence** is broadly favourable to the prosecution case.

Intelligence

Intelligence is information transformed through an analytical process.

Investigating Body

A relevant law-enforcement body as defined in s63A(1A) and (1B) of the Police and Criminal Evidence Act 1984, as amended.

Measurand

A physical quantity, property, or condition quantity that is being determined by measurement.

Method

A logical sequence of operations, described generically for analysis (e.g. for the identification and/or quantification of drugs or explosives, or the determination of a DNA profile) or for comparison of items to

DRAFT

establish their origin or authenticity (e.g. fingerprint/shoemark/toolmark examination; microscopic identifications).

Method Validation

The process of verifying that a **method** is fit for purpose (i.e. for use for solving a particular problem).

Organisation

A group of people and facilities with an arrangement of responsibilities, authorities and relationships (e.g. a company, corporation, firm, enterprise, institution, charity, sole trader, association, or parts or combination thereof).

Precision

Precision is synonymous with reproducibility or repeatability, whereas **accuracy** is about obtaining the true or correct value for the quantity measured. An incorrectly calibrated device may be capable of giving reproducibly precise readings even though data generated are not accurate.

Presumptive Test

The first test carried out on a specimen for the purpose of determining a presumption of a positive or negative identification or assay. Such tests include the Kastle-Meyer test for blood; it can show that a sample is unlikely to be blood (i.e. a low false negative) or that the sample is probably blood (a high false positive) but other substances are known to cross-react and give a false positive result. Usually positives are followed by a confirmatory test.

Provider

The term ‘provider’ is used to include all providers of forensic science, whether commercial, public sector or internal to the police service (e.g. scenes of crime, fingerprint bureau).

Qualitative

DRAFT

Results or requirements based on some quality rather than on some quantity i.e. the identity of the compound rather than concentration.

Quality

The totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs.

Quantitative

A measurement or requirement based on some quantity or number.

Risk

The probability that something might happen and its effect(s) on the achievement of objectives.

Robustness

The capacity of an analytical procedure to remain unaffected by small, but deliberate, variations in method parameters.

Root-Cause Analysis

Is a problem solving process for investigating an identified incident, error, problem, unexpected result or non-conformity.

Standard Methods

A 'standard **method**' is published by certain prescribed **organisations** and has the following characteristics:

a. contains concise information on how to perform the tests;
b. does not need to be supplemented or rewritten as internal procedures; and
c. can be used as published by the operating staff in a laboratory.

Based on the full definition ISO/IEC17025:2005 under Section 5.4.1, at the time of writing (2013) there appears to be no 'standard methods' in the traditional forensic sciences in the UK.

Stress Testing

A data set used in **validation** specifically designed to expose expected or reasonable deficiencies of the **method** under test.

DRAFT

Uncertainty of Measurement

The estimation of the uncertainty of measurement is a BS EN ISO/IEC17025:2005 requirement and is based on the principle that all measurements are subject to uncertainty and that a value is incomplete without a statement of **accuracy**. Sources of uncertainty can include unrepresentative samples, rounding errors, approximations and inadequate knowledge of the effect of external factors.

Validation

The process of providing objective **evidence** that a **method**, process or device is fit for the specific purpose intended.

Verification

Confirmation, through the assessment of existing objective **evidence** or through experiment, that a **method**, process or device is fit (or remains fit) for the specific purpose intended. This is an overriding requirement that there is evidence that the **provider's** own competent staff can perform the method at the given location. Some forms of verification exercise may also take the form of an acceptance or **quality** assurance test.

DRAFT

15 APPENDIX A: COMPUTER FORENSICS EXAMPLE – RECOVERY OF WEB BROWSING HISTORY RECORDS FROM A COMPUTER

15.1 Review of End-User Requirement and Specification

Identification of end-user requirement

15.1.1 It has been determined that in a large number of cases received, the laboratory is asked to recover and produce history artefacts generated by web browsers from computers as evidence. The aim of this is to assess whether the computer may have been used to visit a web page or download content using these applications.

Current provision of requirement

15.1.2 Laboratory's forensic examiners have tended to use a software tool (e.g. Tool 'X') to recover web browsing history records from forensic images of a computer. The tool can be used in various different ways, and the laboratory has not defined a procedure for its use.

15.1.3 The user manual states that Tool 'X' can recover artefacts from Browser 'A' versions 1–4, Browser 'B' versions 1–3 and Browser 'C' version 1. The manual also states that the tool can recover deleted history records, but does not state which artefacts or the state of these deleted records.

15.1.4 The laboratory's examiners have reported that they have identified that version 2 of Browser 'C' has been released and are increasingly noting its use on computers they are examining during the course of everyday casework. Examiners have noted that Tool 'X' appears to recover history artefacts from this browser, although no testing has been performed to ascertain the effectiveness of this 'feature'.

Current validation of requirement

15.1.5 Each time web browsing history records are recovered in the laboratory, manual verification and a quality check in the form of a peer review is performed. The examiner is expected to verify the results to ensure that they are accurate prior to producing them as evidence,

DRAFT

relying on the competence and experience of the examiner to identify any spurious or erroneous results.

15.1.6 There are no accredited methods produced by recognised standards bodies for the recovery of such artefacts using Tool 'X', and the laboratory itself has not performed any prior validation of the tool or the process in which it is employed.

15.1.7 The laboratory has therefore identified this technique as a novel, laboratory-defined method that will require full validation for its continued use in the laboratory. The laboratory has identified that improvements could be made to its overall efficiency in validating this tool, as the amount of manual checking performed each time results are produced could be reduced if a validated method is employed.

15.2 Risk Assessment

15.2.1 A risk assessment has identified the following risks that may arise from the laboratory continuing to produce web browsing history records as evidence without further validation of this method:

a. recovery of duplicate history records from a computer;
b. failure to recover a history record present on a computer;
c. recovery of records generated by another device 'synced' with the computer;
d. recovery of records from other exhibits or cases examined on the same laboratory computer.

15.2.2 These factors could all potentially cause a miscarriage of justice with significant reputational damage or financial loss to the laboratory and/or its examiners.

15.3 Defining the Method and Scope

15.3.1 The scope of the method intended for the recovery of web browsing history artefacts, which will be implemented subject to passing validation, may be defined as follows.

a. The preparation of source data prior to analysis.

DRAFT

b. Loading into and processing the source data with Tool 'X' version 3, recovering history artefacts from the latest versions of the Browser 'A' version 4, Browser 'B' version 3 and Browser 'C' version 1 (the latest supported versions).
c. The output from Tool 'X' and any manipulation and interpretation required to produce evidence.

N.B. The method description has been simplified for the purpose of this example. This section would need to include an exact definition of the method and how it is to be performed. This could reference another document such as instruction manuals or standard operating procedures.

15.3.2 Not included within the scope of this method, and therefore requiring a separate validation exercise, are the following.

a. Previous versions of Browser 'A' and Browser 'B', as examiners have determined through a sample of recent casework that prior versions are no longer found on computers submitted for examination.
b. Browser 'C' version 2, although being the latest version, should not be validated as the method does not document that this version is supported. Other methods should be found for the recovery of these artefacts.
c. The software user manual states that Tool 'X' supports the recovery of other artefacts from these browsers, e.g. cookies and cached web pages. These features are not included in this method.
d. This method will not encompass the in-built search feature of the method that allows users to search and filter the data following data recovery.
e. The method will not encompass recovery of any other artefacts that can be recovered by Tool 'X'.
f. The method will not encompass recovery of artefacts from any other web browsers, as these are not supported by Tool 'X'. Validation of methods that recover these artefacts will be performed separately.

DRAFT

g. The method will not cover the use of other versions of the software prior to or succeeding version 3.

N.B. Any of the above 'features' could be included in the method as long as they are appropriately validated. However, if a method becomes too complex to assess during validation, it may be more effective to split or 'atomise' the method into more than one sub-method, e.g. the cited 'in-built search feature' could be validated in another separate, simpler process.

15.3.3 From this, the laboratory has developed an instruction guide as to how this method should be used for both validation purposes and, if validation is successful, on-going use by the laboratory.

15.4 Validation Strategy

15.4.1 Working with their forensic examiners, the laboratory has developed a strategy for the validation of this method. The web browsers will be installed onto a computer and a predetermined set of websites visited in each respective browser. The computer will then be examined and, using Tool 'X', any artefacts will be recovered and produced evidentially.

15.5 Defining an Acceptance Criteria

15.5.1 Upon completion of the validation process, this method should fulfil the following requirements in order to be validated unconditionally for use in the laboratory.

a. All live and deleted history records should be recovered from the three browsers, accurately and correctly representing the following information:
   i. browser;
   ii. uniform resource locator (URL) of the page visited;
   iii. title of the page visited (where page visited has one);
   iv. date and time of the visit;
   v. device used to visit site (where the browser supports cross-device syncing);

DRAFT

vi. physical location of history record on disk (path, row ID, file/sector offset).

b. There shall be no false positives recovered (i.e. no irrelevant data).

c. There shall be no duplicate records recovered.

d. There shall be no cross-exhibit contamination.

N.B. This is the ideal. Should the acceptance criteria not be met this is not necessarily an outright validation failure. However, these constraints should be made clear in the validation report and in the implementation plan.

15.6 Produce a Validation Plan

N.B. For the purposes of this example, suggested section headings and examples of things to consider are below.

Validation scope

15.6.1 This should cover the following:

a. a detailed explanation of the method to be tested including all of its steps;

b. the browsers (and versions) that will be tested, and the operating system(s) that the browsers will use;

c. details of the tool including advertised functionality in relation to the browsers being tested;

d. detailing the laboratory computer(s) (architecture, operating systems, etc.) used to perform the testing and also the analysis would also be beneficial.

Testing approach

15.6.2 Elaborating on the strategy, this section defines the specific test data that will be input into the browsers including sample size/variety, details of exactly how the tests will be performed and how notes of the process are recorded (the times that specific URLs are visited in the browsers, etc.).

DRAFT

Testing steps

a. The chosen web browsers will be installed on a 'clean' test computer.
b. Each web browser will be tested to an agreed script that emulates the various user interactions that result in the creation of web browsing history on the computer.
c. A forensic image will be taken of the test computer's hard disk drive and verified as a complete copy of the original data.
d. The method will be performed, strictly according to the instruction guide, over the forensic image.
e. On the test computer, the web browsing history will then be cleared/deleted using the browser's built-in features.
f. Another forensic image will be taken of the test computer's hard disk drive.
g. The method will be performed again over the new forensic image, to test the method's efficacy in recovering deleted records.

Defining the test sample

15.6.3 The sample of test data would be defined in this section. In this example, testing consists of visiting websites/pages in the web browsers on the test computer(s). Therefore the following factors could be considered when generating a sample of web pages to visit.

a. Ensuring that the sample is representative of realistic user activity – in this example this could mean including typed URLs, visited links, and search engine activity in the test data.
b. Variety of sample – e.g. consider whether the pages to be tested include symbols in languages that are realistically expected to be encountered, or whether both short and long URLs or page titles are tested for.
c. Ensuring that the sample is of a sufficient size – consider whether the test script will put the tool through its paces and mimic realistic usage of the tool.

DRAFT

d. Frequency of the tests – defining how many times the tests will be run. Once may suffice, but consider the efficacy of the tool.

N.B. This is not an exhaustive list of things that should be considered. Many of these factors will be determined from experience of the artefact or forensic tools/techniques; please refer to the Codes and the appendix to the Codes, Digital Forensic Services FSR-C-107.

15.7 Competency Requirements of Validator

15.7.1 Consider a minimum or expected level of qualification or experience that may be required to perform, interpret and make recommendations on the results of the validation process. For example, it may be determined that the validator must be an experienced computer forensic examiner or analyst to perform these tests as the method requires such technical ability. In addition, this individual should have some training or experience in quality assurance or validation.

Validation Report

N.B. There are several requirements set out in the Codes for this document, including defining a document title, describing the method validated, the validation process, the individual performing the validation, etc. that will not be repeated here. Some key points relevant to this example have been explored below.

15.8 Comparison of Validation Results Against Acceptance Criteria

15.8.1 All live and deleted history records should be recovered from all browsers, accurately and correctly representing the following information:

a. browser;

b. URL of the page visited;

c. all live history records were recovered from all three browsers tested, with complete accuracy and provenance;

DRAFT

d. deleted history records were recovered from the Browser ‘A’ and Browser ‘C’;

e. no deleted history records were recovered from Browser ‘B’.

15.8.2 There shall be no false positives recovered.

a. Ten additional live Browser ‘A’ history records were recovered that were not input during testing. These were dated six months prior to the date that the computer was set up and all point to pages on the Browser ‘A’ website. Further testing proved that these records are default and present when this version of the Browser ‘A’ software is installed on any computer.

15.8.3 There shall be no duplicate records recovered.

a. No duplicate records were recovered.

15.8.4 There shall be no cross-exhibit contamination.

a. No artefacts from other exhibits were recovered.

Method limitations

15.8.5 This method is not capable of recovering deleted history records from Browser ‘B’ version 3. Therefore this method must not be relied upon for this purpose.

15.8.6 Default records are recovered from Browser ‘A’ version 4. Provisions must be made to account for this.

Method implementation recommendations

15.8.7 This method has shown to be effective in performing certain tasks, and is therefore recommended for use for the following.

a. Recovery of live (present, not deleted) history records from Browser ‘A’ version 4, Browser ‘B’ version 3 and Browser ‘C’ version 1.

b. Recovery of deleted history records from the Browser ‘A’ version 4 and Browser ‘C’ version 1.

DRAFT

c. Prior to implementation of this method a list of default Browser 'A' version 4 records must be compiled and upon use of this method this list must be checked against any final results. Matching records must be then excluded to avoid producing these records as evidence that these websites have been visited.

Validation limitations

15.8.8 The data set chosen during testing is the major limitation to this particular validation exercise. For example, if the variety or size of the sample of websites visited during testing is not representative of 'real' web browsing activity, then this should be identified as a limitation of the validation.

15.9 Statement of Validation Completion

15.9.1 This is a short, non-technical summary of:

a. how the method was validated;

b. limitations; and

c. recommendations for the implementation of the method, as explored above.

The audience of this document is effectively the end-user, to help them to evaluate the method and understand the weight that can and cannot be given to the evidence produced.

15.10 Implementation Plan

15.10.1 Building on the results of the validation and recommendations made, an implementation plan could include the following.

a. Assessment of user training and competency – details the minimum competence required for the use of this method and any training required for the method. In the web browsing history example above, a laboratory member with demonstrable experience to identify spurious results would be required.

DRAFT

b. Guidance for use – a thorough user guide should be developed showing the laboratory member how to perform the method.

c. Inclusion into quality systems – generation of standard operating procedures for the use of the method, where applicable.

d. Review interval – define the terms of when this method should be reviewed. This could be a timed six-month interval, or in this example it could be when major new versions of the web browsers are released.

e. Quality assurance procedure – explore how the laboratory can check the quality of results before they are produced as evidence, e.g. by peer review, manual verification or dual-tool/method validation.

15.11 Validation Library

15.11.1 Documentation created during this process should be kept and maintained in a validation library. Appropriate version control and backups of this documentation should also be in place as part of a wider quality system.

DRAFT

16 APPENDIX B: MOBILE DEVICE FORENSICS EXAMPLE

16.1 Mobile Device[9] Forensics Overview

16.1.1 Due to the fast, ever-changing nature of mobile devices, method validation in mobile device forensics is vitally important. New platforms are released to consumers regularly, which may mean that specific, validated methods may not be fit for purpose when it comes to analysing new platforms, or new versions of known platforms. This example is prevalent in smart devices, where a small change in the structure of storage files may mean the difference between getting all of the data, or none of it. In addition, challenges faced by a mobile device forensic analyst may also include restricted access to the raw data present on the device. This is significant as the type of verification of the evidence recovered may be dependent on the type of device being analysed. Issues with evidence obtained via forensic tools may include:

a. data extracted by the tool of choice may be incomplete, e.g. Short Message Service (SMS) messages extracted from a handset with timestamps missing;

b. data extracted may be incorrect, e.g. the tool has not recovered a particular section of a concatenated SMS message;

c. data may not have been extracted at all by the tool.

16.1.2 In the examples given above, the forensic tool used will often give no clear indication as to whether the extraction of specific data types has been successful, or has failed. Many of these issues are common when analysing basic mobile devices that do not allow the end-user to access the file system where such data are stored. The tools of choice request data from the target mobile device, the device will respond to these requests, and the forensic tool will parse the replies into the data seen by the analyst. However, it is vitally important that the tools of

[9] A mobile device is defined as a small, often handheld device usually consisting of a display with some form of input feature, such as a touch screen or a keyboard. Examples of such devices include mobile phones, tablets or satellite navigation devices.

DRAFT

choice have been validated within reason to identify limitations that the tool may possess in relation to the platform of the device being analysed.

16.1.3 As stated previously in this document, it can be impracticable, due to the rapid advances of mobile technology, as well as the increasing abundance of mobile platforms, to validate a method for every situation, for every platform that the method supports; many popular mobile forensic tools support hundreds of mobile platforms. This is where the first stage of the validation process, 'user requirements', is very important. Critical requirements of the method should be identified, with analysis tasks that are most common having a higher priority than activities that are rarely needed, or not at all. In addition, the specific requirements of the tool and/or method for data extraction from a specific platform should be assessed for the requirements of the examiner, and for the remit of the investigation. However, even then, with so many mobile device platforms, operating systems, and the sheer diversity of devices, it is important to implement active verification of data extracted, in partnership with method validation.

16.2 Manual Verification

16.2.1 Manual verification is the practice of actively comparing data extracted from the forensic tool of choice, with data that are displayed to the user of the device. In many circumstances, in parallel with method validation, this allows the analyst to be confident of the accuracy and precision of the records extracted. This practice is critically important when analysing mobile platforms where the data are not readily available in raw form (i.e. the files that store these data are not accessible). This provides assurance that the correct number of records has been obtained, and that the data are precise.

16.2.2 If the forensic tool has undergone no validation regarding the platform in analysis, then 100 per cent verification of the data extracted should be carried out to determine the accuracy of the method/forensic tool, as the limitations of the method regarding that particular platform are

DRAFT

unknown. However, if the method has undergone prior validation to the scope of the investigation in relation to the platform in question, then less checking may be carried out so long as the noted limitations are taken into consideration. For example, if the tool is known to misinterpret dates and times for call history records, then the analyst should check the dates and times of all call history records. However, if another forensic tool at the analyst's disposal has been tested and obtains the call records correctly, then the analyst should refer to that tool, with a smaller manual verification set.

16.2.3 While manual verification is essential for handsets where the analyst has no access to the raw data set, it is also important for the analysis of devices where the raw data are available and can be queried. 'Smart devices' typically store data in SQLite databases and other 'human readable' file formats, which allows the analyst with the correct tools to view the raw data in the native format. Upon doing so, and with the right skills and knowledge of these particular file formats, the analyst can verify data that the forensic tool has parsed. However, due to the ever-changing nature of the structure of these file formats, normally associated with firmware updates (which happen more regularly on these type of platforms) the forensic tools rarely keep up to date.

a. Advantages

    i. Allows active verification of extracted data, providing confidence in the precision and accuracy of the data extracted.

b. Limitations and guidance

    i. Not suitable for handsets where all extracted data may not be readily presented to the analyst.

    ii. Should not be used solely to provide confidence in the data extracted. It is therefore advised that a log is kept as to which records have been manually verified, and the analyst should make a sensible decision on how many records are to be sampled.

DRAFT

16.3 Dual-Method Verification

16.3.1 Dual-method, or dual-tool, verification is the practice of using more than one method to verify data extracted. In doing so, a comparison is made between the two data sets to conclude the accuracy and precision of the data. Dual-method or dual-tool verification can be considered as one whole method in its own right. While this practice may seem advantageous to determine the quality of evidence, there are also a few main limitations, as discussed in the following points.

a. Advantage

    i. Allows further confidence in the evidence obtained via the first method, but only if prior method validation is conducted on one or both of the tools, and they are known to operate in an independent manner.

b. Limitations and guidance

    i. Many tools operate using the same protocols, e.g. the method used by one forensic tool may be a standard protocol that other forensic tools also use. If the analyst is in a situation where manual verification cannot be carried out, and no prior validation of the method has been undertaken, then dual-method verification cannot be relied upon.

    ii. Validation of one or both methods prior to dual-method verification is a must to determine its limitations, otherwise, how can the analyst conclude if any limitations exist? It may not be obvious whether or not different tools are truly independent.

16.3.2 It may be suitable, when coming across new data artefacts such as smartphone/tablet apps, to develop new methods to target the specific data required. However, this increases the need for prior validation to be carried out so that the method can be tailored.

16.3.3 It can also be said that if a particular analyst is competent in the analysis of the file type requiring analysis, then prior validation of the

DRAFT

bespoke method is not necessarily required. However, detailed notes on the method approach should be kept, verification of the data extracted should be implemented, and a standard methodology for the analysis of the file type should be adhered to.

16.4 Mobile Device Forensics – Extraction of Call History Records from Nokia Series 40 Devices

Defining the user requirement

16.4.1 The extraction of call history records from a Nokia device that belongs to the 'Series 40' platform. The test is necessary to determine whether all call history records are extracted from the device, and to measure the precision of the extracted data.

16.4.2 For this particular handset platform, validation is required as the device only displays one call history record per contact, per call type. For example, if there were five contacts, each with five dialled calls on the device, the device would only display five dialled calls to the user.

16.4.3 Novel technique: The novel technique of 2014 involves extracting the data from the device using a forensic tool that is available that supports the extraction of call records from this particular Nokia Series 40 device. This is then followed by manual verification of the extracted data. However, as the device does not present all data records stored on the device to the user, validation of the tool is required to determine whether all records are extracted.

16.4.4 New version of existing technique: The new version of the existing technique will depend very much on the outcome of the validation test.

16.4.5 Previously validated technique: Examine the device using the tool of choice. Old Series 40 devices allowed the user to view all records to allow for accurate manual verification. As this behaviour is new to the platform, the method must be tested using a device exhibiting this updated behaviour.

Risk assessment

DRAFT

16.4.6 Risks: The incorrect number of records will be extracted from the handset, and as the user cannot view all records present, it may be assumed that the tool is correct.

Validation strategy

16.4.7 A Nokia Series 40 device that exhibits the same behaviour of obscuring the call data records is populated with a known data set, which includes the following parameters:

a. a mix of dialled, missed, and received calls;

b. the device must be populated with more than one call record per contact per call type;

16.4.8 the device will then be analysed using the selected forensic tool, and the results will be compared with the known data set.

Acceptance criteria

16.4.9 The method will only be regarded as successful if all records on the handset have been extracted accurately. The precision of such records is also coherent.

Produce validation plan

16.4.10 Measurement-based versus interpretive-based: The test to be carried out is measurement-based. No interpretation is required; the output can be assessed by a layperson with no technical competencies in the field.

Assessment of uncertainty

16.4.11 At this stage the test set used within documentation and the 'true answer' should be defined. For this type of test, along with the data set, the only 'true' answer will be the tool extracting all the test set records correctly. No deviation from this outcome will be accepted as the new method.

DRAFT

Undertake tests

16.4.12 Define the test's components, including:

a. the make and model of the test platform (including firmware version where appropriate);

b. the outlines of the method, i.e. the tools and actions contained within this new method that will be used;

c. the order in which the tools will be used; and

d. any sub-methods that may be included.

16.4.13 The tests will then be carried out at this stage, as defined in the validation strategy.

Assessment of actual outcomes versus acceptance criteria

16.4.14 Compare the actual outcomes of the tests with the acceptance criteria defined.

Produce validation report

16.4.15 Produce a validation report with the outcome of the findings, whether the method has been accepted after comparison between the results and the acceptance criteria detailed earlier. In this specific example, all call records from the known data set were extracted correctly, as expected. However, in accordance with the 'active verification' of data methodology previously detailed, this test may have to be carried out using a separate tool to ensure that the data can be actively verified during analysis using the separate tool, as in this situation, the number of call records present on the handset will be unknown to the analyst.

Produce implementation plan

16.4.16 An implementation plan for the newly validated method must be created, detailing:

a. the method in full and how other analysts may implement it, including the outcome of the test;

DRAFT

b. the situations where it is to be used; and

c. its limitations (if applicable).

In this example, the method will state the following.

a. The test was successful and the forensic tool 'ABC' can be used to extract call records from a Nokia mobile phone of the Series 40 platform.

b. The test was conducted using a known data set and was successful. However, as only one Nokia Series 40 was examined and the analyst does not have access to the raw data set to determine manually how many records are present, a second tool should be used to verify the precision and accuracy of the call records.

16.4.17 The implementation plan should also include:

a. details of training and competency (if applicable);

b. possible tests for new versions of the software that are released (if required); and

c. inclusion into quality systems, such as standard operating procedures (SOPs) or other internal systems (such as handling guides).

16.4.18 In this example the new method will be added to the internal system. Analysts within the laboratory will be made aware of the new method's existence, and that it should be used for future examinations.

On-going use

16.4.19 Quality assurance testing regimen: Checks should be made to determine whether current SOPs cover the quality assurance checks of the new method. If not, an allowance for the new method should be made during the quality-checking phase.

16.4.20 On-going competency requirements: Details relating to on-going competency requirements, if applicable.

DRAFT

17 APPENDIX C: CELLSITE ANALYSIS EXAMPLE – CALL DATA RECORD NORMALISATION TOOL

17.1 Introduction

17.1.1 This example is for a proposed data normalisation tool. Call data records (CDRs) may be provided from a variety of networks in a variety of formats. A tool may therefore be used to standardise ('normalise') that data.

17.2 Risks

17.2.1 The tool excludes legitimate information held in the CDRs from the output (e.g. does not process all of the data, or falsely exclude legitimate information).

17.2.2 The tool includes illegitimate information in the output (e.g. data from a previously normalised CDR).

17.2.3 The tool incorrectly converts data.

17.3 Validation Requirements

17.3.1 The specification of the tool needs to be highlighted so that the validation requirements can be defined. For example, that the data normalisation tool needs to be relied on:

a. to normalise data formats (e.g. times/dates, location information for cellsites);

b. to normalise the terms used for common events (e.g. 'outgoing call' replacing all the other terms used to represent such events);

c. to exclude data that are potentially misleading (e.g. cellsite information related to phones other than the one for which the data were requested).

DRAFT

17.4 Validation Strategy – Purpose

17.4.1 The requirements are all 'technical'; interpretation of output is a separate method.

17.4.2 Validation Strategy – Limitations

17.4.3 An issue surrounding validation of software replacing manual activity is the quantity and variety of data likely to be encountered. Known data in the formats expected to be encountered can be input to the software and the output can be compared with an entirely definable and predictable expectation.

17.5 Validation Plan

17.5.1 Now that the strategy has been defined, a detailed plan can be drafted. The method is essentially an efficiency saving, and there is no risk of contamination or negatively affecting a 'live' analysis, rather than delivering additional analytical capability to be verified. There is therefore no issue with testing it on live casework in parallel with existing methods (i.e. the output of the tool should not be relied upon in itself, but can be compared with the output already relied upon).

17.5.2 A variety of validation approaches can be defined, adopted and documented.

a. In-code error trapping and pre-implementation testing on known data sets.

b. Verbatim check of the output against:

   i. the original file;

   ii. traditional re-formatting location data via plot.

17.6 Evaluation

17.6.1 A log of records used in the tests should be kept.

a. Periodic assessment of the records used can take place.

DRAFT

b. When all combinations of networks formats and sufficient data quantities have been converted with no issues, the tool could be deemed fit for use.

c. These requirements could be defined in advance (i.e. the acceptance criteria). For example, (**accuracy)** 0 errors for (**precision)** 5 CDRs in every format known to exist, to include at least 1 record with over 1,000 entries.

17.7 Assessment of Uncertainty

17.7.1 The tool should, if working properly, provide a discrete and defined output not subject to uncertainty. Dip checking of output should still be undertaken (as noted below under 'other activities') to assure continued valid operation.

17.8 Reporting

17.8.1 Once a sufficiently extensive data set has been assessed the software may be deployed for operational use. A validation certificate highlighting the tests performed and the locations of the detailed assessments can be issued.

17.9 Other Activities

17.9.1 Successful validation does not mean that the software can be used in casework with no caveats, and the guidelines may include the requirement for on-going dip checking (verification tests) taking place. These tests potentially review that the correct number of records have been normalised, misleading data have been removed, etc.

DRAFT

18 APPENDIX D: CELLSITE ANALYSIS EXAMPLE – SURVEY TOOL

18.1 Introduction

18.1.1 This example is for a method involving a proposed survey tool. The survey tool may have been purchased from an organisation other than that deploying it (e.g. a network tool used by telecomm providers) or may be an application developed in-house (in which case there may be additional code-level assessments also possible).

18.2 Risks

18.2.1 If the method does not accurately reflect the operation of an actual mobile phone when compared with call data, the results from it may provide:

a. false negatives – failure to detect a legitimately serving cell may appear to exclude use of it from a location at which it was actually present;

b. false positives (provide a result that indicates a phone may have been at a location even though it could not have actually been there).

18.3 Validation Requirements

18.3.1 The specification of the tool needs to be highlighted so that the validation requirements can be defined. For example, the survey tool may need to:

a. reliably replicate the operation of an actual mobile phone (highlighted in 'risks' above);

b. detect and record a serving cell ID;

c. record a location (potentially also defining the co-ordinate system to be used);

d. provide other data, e.g. received signal strength, frequency, neighbour cell data. Some of the requirement may be to report absolute measurements (e.g. detected signal strength).

DRAFT

18.4 Validation Strategy – Purpose

18.4.1 The method may be:

a. technical (when a survey is conducted but the output not assessed);

b. technically interpreted (when the output is assessed); or

c. fully evaluative (when the output is assessed in the context of a wider question and an opinion may be given).

18.4.2 The validation requirements should reflect which of these outputs are required of the method and may include wider validations to encompass survey strategy or interpretation. For the benefit of this example, it is assumed that the method is restricted to exclude any interpretation of the output (i.e. it is a technical validation).

18.5 Validation Strategy – Limitations

18.5.1 The issues surrounding validation of survey tools include the source data (i.e. the air interface radio environment) being outside the control of the validation exercise. This is unusual for most validation areas, as the easiest way to assess the accuracy and precision of a method is to test it on a defined data set where an explicit comparison with a known, completely true, answer can be achieved.

18.5.2 In the absence of an ideal and entirely predictable mobile phone network controlled by the person performing the validation, the complete 'true' answer will be unknown. Thus, if there is a range of possible answers, these may be difficult or impossible to define these accurately, although it may be possible to define a subset of correct answers.

18.5.3 There is therefore a limitation to the validation from the outset.

a. The complete range of 'true' answers is unlikely to be definable. The true accuracy and precision of the equipment cannot be easily tested.

DRAFT

b. The assessment of the validity of the tool will be affected by the validity of the manner in which it is used, which is likely to require separate validation.

18.5.4 The validation strategy may include more than one approach and becomes more robust if combinations of them are adopted.

18.6 Validation Strategy – Approaches

Consistency assessment

18.6.1 Tests as to whether the tool is consistent with its own output at a different time or with other identical devices at the same time can be performed. Ideally, two or more devices would be available for simultaneous deployment enabling direct comparison of their output. In addition/alternatively, if other tools have already been through a full validation, and are accepted as legitimate devices for comparison, they can be simultaneously deployed and the outputs compared. The method of deployment should also be varied so as to 'stress test' the tool (i.e. expose it to a variety of conditions and therefore increase the likelihood of detecting shortcomings).

18.6.2 This approach has a number of virtues:

a. the 'true' answer does not need to be known as it is a straightforward comparison of output from different tools that is being performed;

b. While the 'accuracy' cannot be assessed (as the true answer is unknown), the differences in output can be assessed and hence a comparison of the uncertainty of measurements can be made. In this example, this is related to the 'precision' of the device.

18.6.3 Examples of types of deployment are given below.

Blind trials

18.6.4 An individual makes calls and makes a record of where they are at the time of the calls. The call data records are then requested from the relevant telecomm service provider. If the trial also forms part of a

DRAFT

competency assessment, the location of the caller should not be shared with the person performing the analysis.

18.6.5 This approach has a number of virtues:

a. the approach tests the equipment in the same situation that it is likely to be deployed in live casework;

b. at least one 'true' answer is known, i.e. if the cell that was used at the time of the call is detected using the equipment under test as showing where the call took place, this is clearly a valid result.

18.6.6 There is a known issue in that only one cell can be recorded as a serving cell in a call record at a given time. There are likely to be other legitimately serving cells at a location and these could not also be in the original call data record. Selection of any of these other, legitimate, additional cells by the test equipment is not an incorrect answer, but as these other 'correct' answers cannot be specified in advance it may not be clear if the equipment is actually functioning as hoped.

18.6.7 If a legitimately serving cell (i.e. that in the call data records) is not detected at the location where it is known to have served, additional assessment may therefore be required. For example, the approach adopted for deployment of the equipment may be at fault (some methods are known to be more prone to false exclusions than others) or there may be some other reason (e.g. the cell in question may have been off air at the time of the test survey). As such, even blind trials cannot be definitive and should not be used in isolation from other validation approaches.

18.6.8 Ultimately, this approach is unlikely to highlight false positives (including an invalid cell in the test output) but may spot false negatives (artificially excluding a legitimate cell).

18.7 Survey Methods

Location surveys

18.7.1 The equipment is deployed to survey a specific location and the cell ID(s) detected serving there can be compared with either:

DRAFT

a. the same or a similar device at a different time;

b. a similar device at the same time; or

c. a known, expected, result.

18.7.2 :There are a number of survey approaches that could be adopted (e.g. static surveys, limited movement surveys or targeted area surveys). These survey approaches may be separately validated prior to the equipment being tested so their effects can be predicted. This type of survey has the virtue that it can be easily linked to a blind trial to assess accuracy (but does not need to be) and can also enable assessment of the variability of results at a location. A disadvantage would be that the environment in which the survey takes place may not stress test the device.

Route survey

18.7.3 The equipment is deployed to survey a specific route and the cell ID(s) detected along it can be compared with either:

a. the same (or a similar) device at a different time; or

b. a similar device at the same time.

18.7.4 This type of survey has the virtue that, if the route is carefully selected, it can stress test the device by moving through a number of types of environment (e.g. rural, suburban, urban) and can move through different Location Area Codes (LACs).

18.8 Validation Plan

18.8.1 Now that the strategy has been defined, a detailed plan involving combinations of the options above can be drafted. This may include detailed planning and documentation of the following.

a. Blind trials at known locations, testing both the primary risk (that the method does not replicate a ‘real’ phone) and the **accuracy** of the method (i.e. whether a ‘true’ answer generated by a ‘real’ phone is reported). This also tests both the equipment and the operator.

DRAFT

b. Consistency trials, testing the **precision** of the method (i.e. whether the range of results returned is replicable). This may include a comparison of output in different environments (e.g. at a location, along a route, in a rural area, in an urban area) for:

   i. the test method in parallel with a previously validated method;

   ii. multiple test devices deployed simultaneously;

   iii. the same device deployed in the same environment at different times.

18.8.2 Expected test results for a 'valid' method can be defined (e.g. that a known serving cell is detected in a blind trial).

18.9 Evaluation

18.9.1 Reliably replicate the operation of an actual mobile phone:

a. test via blind trials.

18.9.2 Detect and record a serving cell ID:

a. test via blind trials, consistency tests using the same device at different times, other devices of the same type or which have successfully undergone independent validation.

18.9.3 Record a location (potentially also defining the co-ordinate system to be used):

a. test via plotting survey data on maps and compare these with where the survey is known to have been undertaken.

18.9.4 Provide other data, e.g. received signal strength, frequency, neighbour cell data. Part of the requirement may be to report absolute measurements (e.g. detected signal strength):

a. test via blind trials, consistency tests using the same device at different times, other devices of the same type or which have successfully undergone independent validation;

DRAFT

b. if absolute measurements are to be reported (values with units), measurements of standard signals can be assessed.

18.10 Uncertainty in Reporting Serving Cell Results

18.10.1 For each of the deployment methods, the variation of results should be defined. For example, if analysing a blind trial the location survey results should be defined.

Accuracy

18.10.2 Was the cell that was known to serve the location within the results specified by the tool? If not, is there a reasonable explanation as to why not (e.g. cell off air)?

Precision

18.10.3 If more than one cell was detected were the same cells also detected if the survey was repeated?

a. Were the same cells detected by other units simultaneously deployed?

b. Were the same cells detected by other validated systems?

18.10.4 As the complete 'true' answer is unknown (i.e. the full list of serving cells is unknown, only those cells selected in the blind trials are known), a quantitative assessment of accuracy and precision is not reasonable, but this does not mean that they cannot be assessed at all.

18.11 Reporting Measurements in Standard Units

18.11.1 This may be achieved by comparison of measurements against a known, externally assured, standard signal, preferably in a radio isolated environment (e.g. a Faraday cage). The expected received power at a specific distance from the signal generator can be calculated using established methods, and the output of the method compared with the known true value. Comparisons of the measured and true values can then be made to establish the closeness of each result (and the mean of all results) to the known correct value (the **accuracy**) and the range of values (the **precision**).

DRAFT

18.11.2 Once this has been established, the effect of the actual value measured on the question to be addressed (e.g. how the absolute signal strength affects selection of the serving cell) would also need to be assessed for it to have any meaning.

18.11.3 This example is based on reporting standard radio frequency power measurements (e.g. dBm), but could just as easily be audio frequencies or any other method that produces output results in standard units (e.g. Hz, nm). It is difficult to see how any method reporting measurements in standard units could be validated without reference to an externally assured standard unit.

18.12 Reporting

18.12.1 Once the tests have been conducted and evaluated, a report and associated documentation can be drafted.

18.13 Other Activities

18.13.1 Standard operating procedures should be drafted covering how to use the device (set up, deployment, 'in field' checks, recovering data from the device). This could be a guidance manual and is to enable technical operation of the unit by a trained operator. Any practical issues should be highlighted.

18.13.2 If the output is to be interpreted in any manner, this interpretation needs to be tested.

18.13.3 Just because a tool is assessed as valid for reporting legitimate cell information, this does not mean that anyone using it is automatically competent to interpret the output or give an opinion on the meaning of the results. Competence for these activities must be explicitly assessed in addition to the tool itself.

19 APPENDIX E: CELLSITE ANALYSIS EXAMPLE – SURVEY METHOD

19.1 General

19.1.1 This example is for a proposed survey method. A survey will rely on a survey tool and so this example shares many characteristics with the validation of the tool described above in Appendix D, but with a

DRAFT

different focus. Many of the details will remain the same, however, so this example should be read in conjunction with that for the survey tool above.

19.2 Risks

19.2.1 If the method does not accurately reflect the actual radio environment, when compared with call data the results from it may provide:

a. false negatives – failure to detect a legitimately serving cell may appear to exclude use of it from a location at which it was actually present;

b. false positives (provide a result that indicates a phone may have been at a location even though it could not have actually been there).

19.3 Validation Purpose and Requirements

19.3.1 The specification of the method needs to be highlighted so that the validation requirements can be defined. For example, the method may need to detect cells serving at a location, or demonstrate the area over which a cell provides service. A separate validation would be required for each activity; both are discussed below.

19.4 Validation Strategy – Limitations

19.4.1 Issues include the source data (i.e. the air interface radio environment) being outside the control of the validation exercise. This is unusual for most validation areas, as the easiest way to assess the accuracy and precision of a method is to test it on a defined data set where explicit comparison against a known, completely true, answer can be achieved.

19.4.2 In the absence of an ideal and entirely predictable mobile phone network controlled by the person performing the validation, the complete 'true' answer will be unknown. Thus, if there are a range of possible answers, these may be difficult or impossible to define accurately, although it may be possible to define a subset of correct answers.

19.4.3 There is therefore a limitation to the validation from the outset.

DRAFT

a. The complete range of 'true' answers is unlikely to be definable. The accuracy of the method may be tested but the precision cannot be so easily assessed.

b. The assessment of the validity of the method will be affected by the validity of the tool used, which is likely to require separate validation.

19.4.4 The validation strategy may include more than one approach and becomes more robust if combinations of them are adopted.

19.4.5 There may also be additional environmental factors to consider in that the environment to be measured may also be prone to change (e.g. over time).

19.5 Validation Strategy – Approaches

Example 1 – location surveys

19.5.1 There are a number of survey approaches that could be adopted (e.g. static surveys, limited movement surveys or targeted area surveys). Each survey approach should be validated separately by comparison of results.

19.5.2 This comparison can include an assessment of whether the approach produces the same results each time it was deployed using the same method in the same environment. If an entirely different list of cells are presented, this indicates that the method under test is both imprecise and inaccurate (i.e. the approach is entirely inconsistent and the 'true' result – whatever that may be – was not detected on at least one occasion).

19.5.3 A blind trial would also be strongly recommended, so that at least one 'true' answer is known to enable assessment of **accuracy**. This comparison can explicitly address consistency, false positive and false negative results, between methods (an assessment of **precision**).

DRAFT

Example 2 – service area survey

19.5.4 The equipment is deployed to survey a specific cell ID when the mast location (and preferably other data, such as antenna point direction and height) is known.

19.5.5 The area over which the cell ID is detected serving can be compared with that detected by:

a. either the same or a similar device at a different time; or

b. a similar device at the same time.

The intra- and inter-device uncertainty in measurements can therefore be assessed (related to the precision of the technique).

19.5.6 If a blind trial is also conducted, the presence (or absence) of the cell ID within the service area can be reviewed (**accuracy**).

19.5.7 The data can also be compared with the reasonable expectations of the service area (e.g. that there is more service in the azimuth direction than behind it, that it is constrained in this direction by known terrain). This will only reasonably highlight those data that are obviously erroneous (**precision**).

19.5.8 The cell could also be mapped at different times of the day or year (via season) to greater assess uncertainty inherent to the method.

19.6 Validation Plan

19.6.1 Now that the strategy has been defined, a detailed plan involving combinations of the options above can be drafted. This may include detailed planning and documentation of the following.

a. Blind trials at known locations, testing both the primary risk (that the method does not replicate the actual radio environment) and the **accuracy** (i.e. whether a 'true' answer generated is reported) of the method. This also tests both the equipment and the operator.

b. Consistency trials, testing the **precision** of the method (i.e. whether the range of results returned is replicable). This may include a

DRAFT

comparison of output in different environments (e.g. at a location, along a route, in a rural area, in an urban area) for:

i. the test method in parallel with a previously validated method;

ii. multiple test devices deployed simultaneously;

iii. the same device deployed in the same environment at different times.

19.6.2 Expected test results for a 'valid' method can be defined (e.g. that a known serving cell is detected in a blind trial).

19.7 Evaluation

19.7.1 To test whether a method reliably replicates the radio environment:

a. test via blind trials;

b. consistency tests using validated devices both simultaneously deployed and at different times.

19.8 Uncertainty in Reporting Serving Cell Results

19.8.1 For each of the deployment methods, the variation of results should be defined. For example, if analysing a blind trial and the location survey results the following should be defined.

Accuracy

19.8.2 Was the cell that was known to serve the location within the results specified by the tool?

a. If not, is there a reasonable explanation as to why not (e.g. cell off air)?

Precision

19.8.3 If more than one cell was detected were the same cells also detected if the survey was repeated?

a. Were the same cells detected by other units simultaneously deployed?

b. Were the same cells detected by other validated systems?

DRAFT

19.8.4 As the complete 'true' answer is unknown (i.e. the full list of serving cells is unknown, only those cells selected in the blind trials are known), a quantitative assessment of accuracy and precision is not reasonable, but this does not mean that they cannot be assessed at all.

19.9 Reporting

19.9.1 Once the tests have been conducted and evaluated, a report and associated documentation can be drafted.

19.10 Other Activities

19.10.1 Standard operating procedures should be drafted covering:

a. what the limitations are for each method;
b. when the usage of a method is appropriate;
c. when the usage of a method is inappropriate.

19.10.2 If the output is to be interpreted in any manner, this interpretation needs to be tested.

19.10.3 Just because a tool is assessed as valid for reporting legitimate cell information, this does not mean that anyone using it is automatically competent to interpret the output or give an opinion on the meaning of the results. Competence for these activities must be explicitly assessed in addition to the tool itself.

DRAFT

20 VALIDATION GUIDANCE FOR FORENSIC AUDIO & SPEECH ANALYSIS

20.1 General

20.1.1 The areas covered by this guidance are format conversion, audio enhancement and speaker comparison using auditory-phonetic *cum* acoustic analysis. The format conversion guidance is likely to be relevant to all speech and audio practitioners. The audio enhancement and speaker comparison guidance is likely only to be relevant to certain groups of practitioners depending on their activities.

20.1.2 There are other areas of work within forensic speech and audio that are not addressed in this guidance – for example, authenticity examinations of recordings, sound source analysis, sound propagation testing at crime scenes, speaker profiling, disputed utterance analysis and transcription. Also within the areas that are covered, there are methods and approaches that are not addressed here – for example, no guidance is offered with regard to the use of automatic speaker recognition systems with speaker comparison. The areas and methods have been selected on the basis that they represent the majority of forensic speech and audio casework currently being undertaken in the UK. Other areas and methods may be addressed in future publications.

20.2 Format conversion

20.2.1 Format conversion may be the sole purpose of an examination, or an activity carried out as part of a more complex task, e.g. converting a recording to a standard format prior to enhancement or speaker comparison. In almost all cases some format conversion or copying is required, and therefore it is important to ensure that conversions are carried out using reliable, tested methods in order to ensure the integrity of the recording.

<u>What types of conversion need to be validated?</u>

20.2.2 Providers should determine which recording formats they encounter most often in casework and develop validation strategies for

DRAFT

procedures for converting them to a standard uncompressed digital format. The formats that are commonly encountered are likely to include a range of digital audio and video file formats. Additionally, at the time of writing, CD-DA (audio CD) and DVD-Video formats are likely to be commonly encountered, as well as analogue formats including compact cassette.

20.2.3 The design of validation strategies and selection of test materials should acknowledge that some digital audio and video formats do not relate to a single standard format, but may refer to a group of standards (e.g. mp3) which may be coded with different implementations of the standards by different manufacturers, and may be coded with a range of bit rates and sampling rates. Also, formats such as wav and avi are container formats that can contain materials encoded by a variety of codecs.

20.2.4 It is not realistic or practical to expect providers to validate methods in advance for all audio formats, as there is a significant number of formats and recording devices, many of which may never be encountered in case work. As rarely encountered, new or proprietary formats appear in casework, case-specific validation will need to be performed (see Section 20.3).

20.2.5 As well as procedures for converting the format of submitted materials to a standard digital format, it is necessary to validate procedures for the production of materials by the provider. The laboratory should have defined output formats and technical procedures for producing them.

20.2.6 Copying to analogue formats is not recommended and therefore will not be covered here. Should it be necessary to convert to analogue formats in a particular situation, then this conversion will need to be validated. Under normal circumstances, the only digital to analogue conversion that should take place in a forensic audio and speech laboratory is for the purposes of listening to recordings. Analytical listening is an integral part of many tasks, and problems with the equipment or its configuration may influence the outcome of an

DRAFT

analysis or enhancement; therefore the laboratory's methods and equipment for listening to audio should be validated to ensure that the audio signal is reliably reproduced.

20.2.7 Generally, laboratories should avoid producing material in compressed formats. An exception to this is DVD-Video, on which audio is usually compressed. If the laboratory routinely produces material on DVD-Video, this conversion must be validated with respect to the audio quality and whether this is fit for the intended purpose. For example, if the purpose of the conversion to DVD is only for listening, then the validation can be carried out by listening to and comparing the output of test material before and after conversion to DVD.

20.2.8 Methods for sample rate and bit rate conversion also require validation. A validation of sample rate conversion could address, for example, whether the required sample rate is actually achieved, whether there is any change in speed (pitch/file length) as a result of the sample rate conversion, whether appropriate anti-aliasing filters have been employed by the method when downsampling, and whether the whole spectrum and bandwidth is adversely affected when upsampling. For example, a particular method of upsampling from 8 kHz to 44.1 kHz was tested with a white noise sample and it was found that where the long term average spectrum (LTAS) was approximately flat before conversion (up to 4 kHz), after conversion roll off occurred from around 3 kHz. This made the method unfit for purpose as important parts of the speech spectrum were modified.

DRAFT

<u>What should be taken into consideration when validating conversion methods?</u>

20.2.9 The output of audio format conversion depends on the equipment and method used and, as long as the operator is adequately trained, should be the same for any operator. The Forensic Regulator's Codes of Practice and Conduct divide methods into measurement-based methods and interpretive methods (FSR-Codes 20.7.4) but format conversion does not fall neatly into either category. However, while

format conversion does not result in measurements or identifications, accuracy and precision are still relevant concepts. The accuracy of a conversion relates to how well the output represents the input, and the precision relates to how similar the results are on multiple occasions, with different equipment or different operators. The accuracy and precision of the conversion is particularly important when the output is used for subsequent measurements and analyses in casework.

20.2.10 For format conversion, the criteria under consideration in the validation tests may include the following:

a. All audio in the original recording should be present and intact, i.e. nothing should be missing from the start or end of the recording and there should be no additional silences or drop outs.

b. No audio should be added to the recording.

c. No audible distortions or artefacts should be introduced, and there should be no audible loss of quality.

d. The peak and RMS level of the recording should be unchanged.

e. Frequencies of tones in a test recording should be unchanged when measured on a spectrum analyser.

f. The sampling rate of the recording should be the same or higher (unless the process involves downsampling from sample rates higher than 44.1 kHz, in which case appropriate anti-aliasing should be employed).

g. The bit rate of the recording should be the same as the original or higher.

h. Repeating the method on a given recording gives auditorily indistinguishable results (except for analogue to digital conversions where slight differences in level are unavoidable).

20.2.11 Depending on the conversion being validated, it may not be possible, or relevant, to test for all of the above due to the issues described in Sections 15.4.1.3 and 15.4.1.4.

20.2.12 In accordance with the Forensic Regulator's Codes of Practice, and Section 4.1 of this document, the validation plan should specify whether

DRAFT

each requirement is *mandatory* or *desirable*. It may be practically sensible to set most if not all of the above requirements as *desirable*, because in practice it may be that for a given conversion no method can be found for which all the above criteria are satisfied. In this situation the method which carries *least risk* to the accuracy of the converted recording should be selected as the most suitable and any problems with it investigated and documented so the issues are known and controlled so as to mitigate any risks. For example, in comparing two methods for converting *format x* to *format y*, Method A may be found to consistently increase the overall level of the output relative to the original recording by 2 dB which may cause clipping, while Method B is found to add 2 seconds of silence to the end of a recording. In assessing the results of the validation tests, neither meets the ideal criteria. If no other methods are available then it would be most sensible to adopt Method B as the laboratory's standard method and to document in standard operating procedures that this method is known to add 2 seconds of silence to each recording. Additionally the criteria themselves and the extent to which they are mandatory or desirable, depends on what the recording is to be used for after conversion. For example a small change in the spectral characteristics of a recording may be acceptable if the purpose of the conversion is only to enable the recording to be listened to and played in court, but may be unacceptable if the recording is to be analysed in a forensic speaker comparison. In setting the pass criteria in the validation plan, the provider should therefore take into consideration the purpose(s) or potential purpose(s) of the conversion.

DRAFT

Difficulties with validating audio file format conversion methods

20.2.13 There are some fundamental problems with validation of format conversion when dealing with codecs, which must be acknowledged. At first, it may seem a simple task to compare the recording before conversion with the recording after conversion and see what has changed. To do this we would need to be able to open the original recording and the converted recording in analysis software in order to compare the audio, its level, spectrum etc. before and after conversion. However, this is often impossible because for many formats there is no

software with adequate analysis tools that allows a file to be opened directly. Even when software does allow a format to be opened, it often does this by performing some kind of conversion as it opens the file (i.e. using a codec). This means we may never be able to directly listen to or analyse the source file but only a converted or decoded version of it. We are therefore comparing the converted file with another converted file, not with the original 'source' recording.

20.2.14 A second problem is that for a validation exercise we ideally need to start with a test file of known content (e.g. some speech and noise with known spectral content, levels, duration, signal to noise ratio etc.), but to produce this we need to start by converting the test signals *to* the compressed format in question. So the attributes of the test signal we start with are subject to the effects of the conversion to the compressed format and it is this conversion which is likely to have the biggest impact on the integrity of the signal. So the problem is that we have no 'known' version of the signal that existed prior to performing the conversion being tested (i.e. from the compressed to the uncompressed format).

20.2.15 For these two reasons it may be impossible for some audio formats to directly assess the effect of the conversion on the audio signal as part of a validation exercise. This means that different strategies must be employed. One such strategy may be:

a. to produce a test recording of known content in an uncompressed digital format;

b. to convert it to the compressed format in question using any single available technique;

c. to convert it back to the uncompressed format using a number of different methods under test;

d. to compare the resulting files with each other, and with the original signal, to assess any differences and determine which of the conversion methods produces results closest to the original signal.

DRAFT

20.2.16 So using amr as an example of a compressed file format which cannot be directly analysed:

Test.wav -> compression -> Test.amr

Test.amr -> conversion method A -> Test_output1.wav

-> conversion method B -> Test_output2.wav

-> conversion method C -> Test_output3.wav

20.2.17 It should be acknowledged that different software may be implementing exactly the same algorithm or codec to do the conversion, meaning that in effect the same method is being tested twice.

20.2.18 Where it is possible to listen to or analyse the source file directly, the audio signal before and after conversion should be directly compared.

20.2.19 For some file types it is not possible to convert to the format in question to create a test file because, for example, the format may be a proprietary one originating from an item of equipment or software which is not available. In this situation the guidance given in Section 15.4.1.5 for case-specific validation may be followed.

Issues with conversion from analogue formats

20.2.20 Analogue audio and video formats are now almost obsolete outside of the forensic field, and within it their use is in rapid decline. At the time of writing, compact cassette tapes are still regularly dealt with in some forensic audio / speech laboratories due to the fact that, in some police forces, PACE interviews are still recorded on this format.

20.2.21 Laboratory procedures for conversion from analogue formats (or digital formats with no digital output) may be split into two parts: the first being the output from the replay equipment and the second being the analogue to digital conversion.

20.2.22 Analogue replay procedures may consider equipment maintenance (such as head cleaning and demagnetisation), setting

DRAFT

output levels appropriately (to avoid dynamic compression or distortion on the output or clipping at the input to the ADC), adjusting azimuth for maximum high frequency output, and ensuring all audio is copied.

20.2.23 These aspects of the procedures can be validated by carrying out them out on test recordings and ensuring that the output is fit for purpose. Test tapes, or commercially recorded tapes, may be used to check that the equipment is capable of playing a recording, that the signal chain is properly set up, and that there are no obvious quality problems or other anomalies. Any test tapes made at the laboratory should be made on a different recording device than the one being tested for replay. If multiple replay units are available, the output can be compared between units to determine whether there are inconsistencies in quality across them.

20.2.24 To optimise playback of analogue tape-based media, speed, wow and flutter and frequency response should ideally be measured for each replay device using calibrated tests tapes to ensure these properties are within acceptable tolerances. However, it has become very difficult, if not impossible, to obtain calibrated test tapes needed to accurately test these criteria. Furthermore, it is now very difficult to buy professional grade equipment. It is therefore not expected that all providers will be able to accurately determine these characteristics for their analogue replay equipment. Providers must therefore consider what kind of effect speed errors, poor frequency response or poor wow and flutter performance at the replay stage may have on the resulting output and acknowledge the limitations that these may impose on any subsequent analyses, or inferences drawn from digitised versions of these recordings.

20.2.25 Speed: On some analogue recordings, signals or timing information on the recording can be used to correct the speed of the digitised copy. For example, PACE interviews are recorded with a time track on the right channel, where time announcements and a beep are recorded at 10 second intervals. For recordings where no time or frequency reference is available, the speed accuracy of the original

DRAFT

recording equipment cannot be known, and therefore even if the replay equipment is correctly calibrated the practitioner cannot know whether a submitted analogue recording is being replayed at the correct speed. For these reasons, accurate calibration of replay speed may be considered non-essential. However, laboratory produced test tapes or commercially recorded tapes can be used to assess significant speed errors and equipment should be repaired or replaced if necessary.

20.2.26 Wow and flutter and frequency response: Significant wow and flutter and frequency response errors may affect the intelligibility of speech and may have implications in speaker comparison examinations. It is advised that compact cassette machines are, where possible, shown to be working adequately in these respects. This may be achieved through servicing and testing of equipment, or where servicing and testing is not possible, playing laboratory produced test tapes of known material or commercially recorded tapes.

20.2.27 Procedures for the second stage of the conversion, the analogue to digital conversion at the computer, will include selection of appropriate sampling rate, bit depths, channel configuration and format, and identification of the sound cards or interfaces and recording software that may be used. Drivers for the recording interface may need to be specified as well as the operating system in use on the computer. Validation of these procedures should aim to show that the specified recording characteristics are fit for the intended purpose, and that the equipment is capable of producing recordings to this specification without introducing unacceptable levels of distortion, noise or other artefacts, or resulting in signal drop outs.

20.3 Case-specific validation

20.3.1 There will be formats where the validation guidelines suggested above are not feasible, for example when it is not possible to produce a test recording in the format under investigation because no available software or hardware allows the user to record in that format or convert

DRAFT

to it, and the equipment used to make the recording is not available. This is likely to happen when a recording system produces a proprietary format. In these situations it is likely that the recording can only be played and/or converted in one piece of software (or hardware) which may have to be procured specifically for a particular case. Clearly, in cases where the laboratory does not have the device or software to create test recordings, it is not possible to validate the method using test material.

20.3.2 Instead, the recommended course of action is to determine first whether the proposed method, i.e. using the proprietary software, allows the user to export to an uncompressed format or to the target format directly. If so, the exported file should be compared auditorily with the original file as it is heard on replay using the software. If there is an unacceptable audible loss of quality through the export function, the best course of action may be to play and digitally re-record the original file in real time. This process should be documented in the case notes.

20.4 Audio Enhancement

General

20.4.1 The aim of audio enhancement is generally to improve the intelligibility of speech on a recording or the aesthetic 'listenability' of a recording (i.e. to make the recording easier, or more pleasant, to listen to).

Issues affecting validation of audio enhancement

20.4.2 The effectiveness of audio enhancement is dependent on the equipment used, and on the practitioner's skill and judgement. The balance between the effects of these two factors varies depending on the range of equipment available and on the recording itself. Some recordings are simple to enhance using basic techniques, while for others it may not be possible to make *any* improvement to the intelligibility or listenability.

20.4.3 There is no single correct enhancement strategy for a given audio recording. Many different strategies may be employed depending on

DRAFT

the available tools and the practitioner's preferences, judgement and experience, and different tools or even classes of tools can be applied to the same problem with comparable results. It is not generally possible to determine *objectively* which is the best strategy, whether any strategy may be degrading the speech intelligibility or whether a particular strategy could be improved upon given a defined set of tools.

20.4.4 Because of the wide variety of recording problems and tools available and the subjective nature of the output, it may not be appropriate, or indeed useful, to attempt to validate specific methods for dealing with particular types of recording problems. Practitioners must develop appropriate enhancement strategies for each task they are faced with, using a range of tools in various combinations and orders and with appropriate settings. While there may be general recommended approaches to various types of enhancement situations, attempts to prescribe fixed strategies for dealing with particular problems may prevent practitioners from producing the optimum results.

20.4.5 As practitioner competence plays such a vital part in determining the effectiveness of enhancement, the practitioner's role should be recognised in validation exercises concerning whether the processes used are capable of making subjective improvements to the intelligibility or listenability of recordings. Practitioners carrying out audio enhancement work should be trained in elements of audio signal processing and audio engineering.

DRAFT

What should be taken into consideration when validating audio enhancement?

20.4.6 While the equipment can be tested to show that it is performing as expected, the absolute accuracy of the audio processes themselves is not always critical. What matters is what the practitioner chooses to do with the available tools and the effect of the chosen strategy, which will often combine various different processes, on the *speech* and the *noise*. For example whether or not a filter set to a particular cut-off frequency actually cuts off at *exactly* that frequency is not usually relevant providing the practitioner uses their ears (and spectral analysis

tools where necessary) and sets the filter appropriately. Incorrect functioning of a filter may hinder work in some circumstances; for example, when attempting to reduce a tone or set of tones that have been measured using a spectrum analyser, or if a filter introduces unexpected distortion. Therefore validation should address whether the processing tools function adequately, and generally behave as expected.

20.4.7 In validating the use of adaptive filters in audio enhancement practitioners should consider how their method (i.e. the way that they *use* their selected adaptive filters) deals with the effects of adaptation time and adaptation rates. Most adaptive filters may be trained on a selected part of the recording where there is only noise, and then 'fixed' to avoid rapidly changing noise profiles during the first few seconds of the recording. For filters which do not allow pre-training, methods should be established which counteract this problem.

20.4.8 In validating audio enhancement, the equipment can be looked at in isolation but this will not tell us much about what will happen to the audio when used by a practitioner on a particular recording. Therefore, in addition to evaluating the accuracy and repeatability of the tools, and ensuring adequate practitioner training and experience, validation of enhancement methodology may be best approached by considering the effectiveness of the available range of tools at a laboratory when used on a range of different test recordings by practitioners. This may be done using a subjective assessment of the effectiveness of various enhancement strategies selected by practitioners for a range of recordings using a defined range of tools.

Example of a validation plan for audio enhancement

20.4.9 The aim of this guidance is to give forensic providers some ideas about how they may go about validating audio enhancement. It does not prescribe any specific validation strategy and other approaches may be found that are more appropriate depending on the circumstances within each laboratory.

DRAFT

20.4.10 *Method under evaluation:* Filtering of audio recordings to improve listenability and/or intelligibility using any combination of processes/filters available in "Software X/Y/Z" (enhancement software) when used by competent practitioners within the laboratory.

20.4.11 *Risk assessment:* One risk associated with audio processing is that recordings are over-processed. This may potentially cause speech sounds to become more similar to other speech sounds, decreasing intelligibility and giving rise to the possibility of the content being misinterpreted. The impact of this problem depends on the extent to which the recording is over-processed and the nature of the processing. The risk can be controlled by ensuring practitioner competence in avoiding over-processing by selecting appropriate tools and settings, as well as by peer checking and by ensuring that the original recording is always left unprocessed so it can be referred to in case of doubt.

20.4.12 Other risks are that the recording may not be processed as effectively as it could be given a different approach to using the available tools, or that material is missed or extra material inserted when the signal is processed.

20.4.13 *User requirements:* The end-user is usually the court or the investigating officer, but there may be intermediate users whose requirements must be taken into consideration; for example, the person who will be transcribing the processed recording. There are essentially two different user requirements for audio enhancement. The first is a requirement to improve the intelligibility of the speech on a recording and the second is a requirement to improve the listenability. These are not necessarily mutually exclusive: sometimes both are required. A fundamental requirement for all enhancement work is that the intelligibility is not reduced by the processing. There is also a requirement that processing is repeatable and auditable.

20.4.14 *Validation acceptance criteria:*

a. When carried out by a competent practitioner, the processing should not decrease the intelligibility of speech. (This may need to be assessed subjectively).

DRAFT

b. Using the range of tools available in the laboratory it should be possible to make subjective improvements to intelligibility or listenability in recordings that are degraded by a range of commonly encountered types of noise/distortion problems.

c. The processing is auditable and repeatable, i.e. the processing settings can be saved (or otherwise recorded) and recovered/recreated in sufficient detail that auditorily indistinguishable results are achieved on repeated processing of a given audio file with the recovered settings.

d. Processes should operate as expected according to the settings selected by the operator. For example, the cut-off frequency of a low-pass filter should be approximately correct and there should be a fairly flat response in the pass-band and a suitable amount of rejection in the stop-band[10]. An adaptive filter should be observed to be adapting and if there is an option for freezing the adaptation for example, this should be tested to ensure it freezes. The pass criteria may not need to be very strict in many cases, as it is the overall effect on the speech and the noise that is important. For example, the accuracy of the adaptation rate of an adaptive filter is probably less important than whether the methodology being assessed ensures that the speech is not adversely affected by rapidly changing noise profiles during adaptation transition periods.

e. When no processes are active, the system used should be transparent, i.e. audio files opened in the software and saved as new files should be identical or equivalent to the input files. Some small changes may be inevitable (for example introduction of a short delay), but providing they are reproducible and do not compromise the integrity of the information contained within the recording these may be considered acceptable.

[10] The person producing the validation plan may set specific pass criteria for each filter, e.g. +/-20 Hz, +/-3 dB ripple in pass-band, at least 30 dB attenuation, or may choose not to set specific quantifiable pass criteria but instead to determine what the characteristics are and then determine whether these are acceptable and what needs to be done to counter any limitations.

DRAFT

f. No audio material is missed or extra audio material inserted when the recording is processed.

Suggested validation strategy

20.4.15 The strategies suggested here relate to the acceptance criteria numbered a to f, set out above.

20.4.16 For a and b, a set of test recordings may be constructed from a selection of specifically generated test material or other available recordings, chosen to represent the range of types of challenges commonly encountered in casework. The set of recordings could include broadband noise, car/traffic noise, tonal stationary noise, tonal varying noise, music, noise/distortion/interference caused by defective equipment, distortion due to clipping, reverberation, GSM interference etc. The recordings should be selected to represent a range of levels of difficulty of enhancement such as may be encountered in casework. If the recordings are too easy or too difficult to enhance, the tests will not provide any useful information.

20.4.17 The practitioner(s) are given the test recordings and asked to process them to aid intelligibility or listenability (or both) as specified by the person setting the test. The practitioners then process the recordings using an agreed range of tools[11]. Detailed auditable notes are kept by the practitioners and the settings are, where possible, saved to enable repeated processing.

20.4.18 A designated assessor or panel of assessors evaluates the recordings subjectively in terms of whether intelligibility has in their opinion been improved, stayed the same or decreased, and whether listenability has been improved, stayed the same or decreased. The results are then used to provide information about the validity of the specified range of tools as a whole in achieving the goal of effective enhancement without loss of intelligibility. The practitioner or practitioners performing the validation tests should be experienced in performing audio enhancement. The same test can be used as part of

[11] This may be all the tools available in the laboratory, or a subset under test.

practitioners' competency assessments with the aim of the competency assessment being to show that the practitioners are capable of making appropriate decisions regarding effective enhancement strategies without over-processing recordings. It may be that not all available tools are used in these tests, but the point is to show whether an appropriate range of tools is available in the laboratory to enable effective enhancement.

20.4.19 The assessors should look at the enhancement strategies employed by the practitioners and determine which worked well and which, if any, were not as successful, and give feedback to the practitioners.

20.4.20 In an ideal world, intelligibility would be assessed objectively; for example, by using transcription before and after processing (with predefined speech material), or other objective intelligibility measures, but in reality this may be impractical owing to the time it would take. Signal to noise ratio is generally not a good indicator of listenability/intelligibility.

Using the range of tools available in the laboratory it should be possible to make subjective improvements to intelligibility or listenability in recordings that are degraded by a range of commonly encountered types of noise/distortion problems.

For c, using some of the recordings from the tests for requirements i and ii, the filter settings may be retrieved or otherwise reconstructed and the same audio passed through the arrangement of filters used previously. The output can then be compared to determine whether the process is repeatable.

20.4.21 For assessing d, whether the filters are functioning correctly different types of test recordings will be needed for different types of filters. For testing a band-pass filter, for example, white noise would be a suitable source with its spectrum being averaged over several minutes using an audio spectrum analyser. For testing certain adaptive filters, recordings with speech and varying tonal noise may be suitable,

DRAFT

and for testing parametric filters white noise with stationary tones may be suitable. For each type of filter or process being tested, suitable test recordings will need to be determined and produced by the provider.

20.4.22 With e, recordings should be compared before and after resaving with no processes in place to determine whether any changes have occurred to the signal.

20.4.23 With f, recordings should be compared before and after processing with each tool to determine whether any audio is removed from or added to the recording.

20.5 Speaker Comparison

20.5.1 Speaker comparison is a complex method involving a combination of a) non-analytic technical procedures, b) analytic technical procedures and c) human-based interpretation of speech features (i.e. auditory phonetic analysis). The conclusion arrived at by the method is based on an interpretation of the findings from b) and c).

20.5.2 The validation requirements for a), b), and c) and for the drawing of conclusions are different. For b), c) and the drawing of conclusions the degree of analyst-dependency is such that the method cannot be validated independently of the practitioner; it is inextricably linked to individually-held subject knowledge, skills and competencies.

20.5.3 The features of voice and speech most relevant to a comparison, or set of comparisons, will vary somewhat from case to case and cannot be stipulated in advance. Practitioners' abilities to select relevant parameters and features for material under examination should form part of the validation process.

<u>Non-Analytic Technical Procedures</u>

20.5.4 Examples of non-analytic technical procedures include transfer of audio from CDs, DVDs and other storage media to computer, and format conversion prior to analysis. Guidance on these steps is provided under 20.2.

DRAFT

Analytic Technical Procedures

20.5.5 Examples of analytic technical procedures are the editing and preparation of recordings and measurement of various parameters of the speech signal.

20.5.6 Editing may be considered analytic insofar as it involves exercising judgement in respect of, for example, the selection of representative sections of a recording and the location of comparable material in the recording it is to be compared with. In addition to editing, preparatory work may involve the filtering of recordings. This may be necessary, for example, in cases where there are aliasing artefacts. Also, if the frequency bandwidth of one recording is significantly different to another, filtering may be used for the purposes of channel equalisation prior to auditory analysis. These technical procedures may be considered analytic in that they involve practitioners exercising judgment over the selection of filters and settings based upon analysis of the signal. Guidance concerning the validation of digital filters is provided under 20.4. Practitioner competence in editing and filtering may be demonstrated via proficiency testing.

20.5.7 In respect of measuring parameters of the speech signal, e.g., fundamental frequency, formant frequencies, voice onset times and articulation rate, validation is required of the method used to make the measurements. Minimally, this would involve the testing of the software against reference materials (e.g. tones, synthetic speech, real speech) to ensure its accuracy. Consideration must be given to the influence that recording format, bandwidth limitation, poor quality and distortion can have on measurement accuracy. Scripts and spreadsheets that perform logging or calculations must also be validated to ensure their correct operation. The validation process should be repeated when software is updated to newer versions.

20.5.8 Practitioners, as part of their proficiency testing, must be able to demonstrate competence in the extraction of appropriate values using computer software including the logging of the extraction point and the

DRAFT

settings used when performing the extraction. While the actual values extracted are likely to vary to a certain degree across individuals, one would nevertheless expect them to fall within a relatively narrow range of variation. The validation process must address the issue of consistency of measurements, both within and across practitioners. Collaborative exercises involving practitioners within the same laboratory and/or different laboratories are an appropriate testing ground for such consistency checks. The findings of the exercises can be incorporated into practitioner training and standard operating procedures to improve the consistency of measurements.

Auditory Phonetic Analysis

20.5.9 This includes the assignment of speech and voice features to conventional phonetic categories at the segmental and supra-segmental levels.

20.5.10 At the segmental level, practitioners might be expected to competently identify where in a file a particular target occurs, apply IPA symbols and diacritics to consonant and vowel sounds and to explain the sounds in terms of articulatory processes.

20.5.11 At the supra-segmental level practitioners might be expected to competently apply voice quality, rhythmical and intonational descriptors.

20.5.12 As with the making of measurements, consistency within and across practitioners may be ensured through repeated testing of personnel and intra and/or inter-laboratory collaboration. Again, the outcomes of such exercises can be incorporated into practitioner training and standard operating procedures to improve the consistency of analyses.

20.6 Drawing of Conclusions

20.6.1 In drawing conclusions from findings, practitioners address two main questions. The first concerns assessing the degree of similarity between samples, and the second concerns evaluating the distinctiveness, or otherwise, of features found.

DRAFT

20.6.2 The assessment of similarity requires that practitioners are aware of, and able to competently take account of factors that may affect intra-speaker variation (situational, psychological, physical) as well as technical factors including channel differences and recording quality.

20.6.3 The assessment of distinctiveness requires that practitioners are able to bring to bear knowledge of the canonical patterns for each parameter examined, in order to identify deviations from the norm. Norms – social, regional, ethnic - are indexed to the language varieties under examination, and knowledge of them may have been gained through education in sociophonetics (see 20.7) and previous casework supplemented, as necessary, by reference to research literature and/or databases.

20.6.4 Competence of practitioners in respect of assessing similarity and distinctiveness may be established via proficiency testing based around recordings with an accompanying set of analytic findings. The test recordings should reflect the realities of casework in terms of technical quality, duration, etc., and should include a mixture of same speaker and different speaker comparisons.

20.6.5 As with other parts of the speaker comparison method, one would expect some variation across practitioners with respect to the conclusions they draw from findings relating to the same material. However, one would also expect the degree of such variation to be reasonably constrained. Participation in intra- and inter-laboratory collaborative exercises may serve to identify and reduce inconsistencies in performance both within and across individuals.

20.7 Qualifications

20.7.1 Given the heavily analyst-dependent nature of the method overall, audio practitioners would be expected to hold a postgraduate level qualification involving substantial components of phonetics, socio-phonetics and speech acoustics.

DRAFT

21 ACKNOWLEDGEMENTS

21.1.1 This draft appendix was produced following a competitive tender. The following were involved in preparing the text:

Anna Bartle and Gregory Webb (Metropolitan Police Service)

Kevin Sullivan (Principal Forensic Services Ltd); and

Matthew Tart[12], Joshua Tedd and Peter Williams (CCL Solutions Group Ltd).

21.1.2 Further assistance and review was supplied by:

Iain Brodie, Rosanna Hume, Richard Olejnczak, Nicholas Patrick-Gleed, Karen Sabin, Stephen Swabey (CCL Solutions Group Ltd).

DRAFT

[12] Lead author, compiler and editor for this appendix.