European Court of Auditors
Audit Methodology and Support **Unit**

GUIDELINE ON

DATA COLLECTION

CONTENTS

INTRODUCTION

Data collection from an auditee requires attention in the following areas:

- The security of the data, mainly during its transport to the European Court of Auditors (ECA), is subject to serious risks.
- Data must be obtained in a useable format, and must retain its integrity while it is being manipulated.

This guideline is addressed to auditors planning to collect electronic data during an audit.

Contact

IF YOU FEEL THAT THE INFORMATION PROVIDED IN THIS DOCUMENT COULD BE IMPROVED, PLEASE DO NOT HESITATE TO COMMUNICATE YOUR SUGGESTIONS: ECA-AMS.CONTACT@ECA.EUROPA.EU.

1: GENERAL FRAMEWORK

COMPLIANCE

Data collection must respect the **rules in force both at the Court of Auditors and at the auditee**, particularly those concerning data security and personal data protection.

Data belongs to the auditee

If the auditee has a data security classification and security procedures, these **must** be followed in your treatment of their data. Otherwise, treat the data as "ECA-LIMITED" [1].

Consider also **international guidelines and standards**, particularly those issued by ISACA[2].

RISKS & PREVENTIVE MEASURES

The following are the main risks in data collection, in decreasing order of impact:

RISK	Preventive measures
Data is lost or stolen	Apply security measures (encryption, custody) especially during transport to the Court of Auditors. The measures must comply with both the auditee's and the ECA's data security rules.
Data obtained is complicated to use	Ask for the data in a usable format. Test with a sample file.
Data loses integrity during treatment	Define integrity controls (number of rows, monetary totals, hash totals) and check them during data treatment.

[1] ECA data confidentiality levels are listed in Annex I.

[2] ISACA standards: Mainly document S6 "Audit rules on IS audit - executing audit work" and the guideline G3 "Use of computer assisted audit techniques."

2: PREPARATION

The need for data should be identified as soon as possible

The **need for data** from the auditee should be identified as soon as possible.

Time and effort in planning the data collection will help to reduce the risks shown above.

IDENTIFY THE GOALS AND THE DATA NEEDS

MUS ==> ID + amount

To perform a Monetary Unit Sample (MUS) you only need two "columns" for each transaction: an identifier and the monetary amount.

For advanced analysis you could need several tables that can be related (they share some fields).

No unnecessary data

Decide what analysis you want to perform, and ask only for the data you need.

Personal data: really needed?

If you absolutely need personal data (PD), be aware that:

- PD can imply a higher data classification level for the auditee and thus reinforced security procedures during collection and use.

- You have to inform the Data Protection Officers (DPO) of the ECA and of the auditee before you collect the data.

Tokenization: hiding sensitive data

One technique to use sensitive or personal data (PD) with reduced risks is to tokenize the data. This means replacing the data with a value (a token) calculated from it, for example an encryption of the data.

Tokenized data can be analysed in the same way as the original data because the same token always replaces the same data. For example, we could find several beneficiaries with the same address, without knowing what the address is.

Tokenization means work for the auditee

However, tokenization requires expertise and effort on the part of the auditee. There is a risk that they might refuse to do it.

STUDY THE IT SYSTEMS

You need to study the auditee's IT system to:

a) Recognize the data you want so that you can ask for it without ambiguity.

b) Have reasonable assurance that you are collecting reliable data (not incomplete, inconsistent or tampered with)[3].

BE PRECISE WHEN REQUESTING DATA

Use the enclosed form to specify the details

To ask for the data you need to specify:

1. **Data you want**: tables, columns, row selection criteria.
2. **Controls:** completeness and integrity.
3. **File format:** flat file or proprietary file (MS-Excel).
4. **Collection and Transport details**: security.

These are the contents of the **Data Request** template**,** which you can use to make a formal request for the data.

1) Data you want

Most probably the auditee uses a relational database. Therefore you can identify the required data by:

- **Table** (example: payments),
- **Columns** (example: postal code, beneficiary ID, date, amount in EUR).
- **Selection criteria** (example: where status = 'paid' AND Book_year = 2007).

2) Controls

Completeness control

You should ask for some sort of **completeness control** on the data. For example: the sum of all the payments could be equal to a total in some official report.

[3] The INTOSAI Auditing standards state, in explanatory paragraph 144, that "Where accounting or other information systems are computerized, the auditor should determine whether internal controls are functioning properly to ensure the integrity, reliability and completeness of the data." Explanatory paragraph 153 states that "When computer-based systems data are an important part of the audit and the data reliability is crucial to accomplishing the audit objective, auditors need to satisfy themselves that the data is reliable and relevant".

Integrity controls:

You should ask for **integrity controls** which help detect if data is "damaged" while being processed: e.g. rows are not lost, values are not changed.

- Number of rows
- Monetary totals
- Hash totals

The minimum items to check are: the number of records and the total of the monetary columns. You can also check the total of some other numerically coded columns (i.e. postcode, budget line, id-codes). This is known as a hash total and detects changes to values in that column.

3) File format

Tool formats are fine, but beware of versions

If you know which tool you will use you can ask for the data in that format. **Beware**: check that the auditee's version of the tool is the **same or older** than yours to ensure compatibility. The ECA currently uses MS-Office 2010.

Our advice is to ask for data in **text files** that can be read by Excel, ACL, Access, Oracle.

Text files with delimited columns are likely to be compatible

A fairly compatible file specification is the following: flat text files encoded in ASCII or UTF8; one file per table; one row per record; columns separated by TAB characters; decimal point is point or comma (no mix); no thousands separator (not even space); text fields not enclosed in quotation marks; no omission of last fields in row (some tools omit them if null).

Early testing helps

Hint: ask for a file with some sample data (text, dates and numbers) to test compatibility. This is both for proprietary formats and for text files.

4) Collection and transport details

Based on your request, the auditee will determine the **data security classification** and **estimated size** of the data. Then agree with the auditee:

- the transport method appropriate for the data classification and size.
- if you need to go to the auditee's premises to supervise the extraction or transport the data.

Avoid accessing the auditee's systems

The auditee should extract the data. The auditors should only exceptionally access the auditee applications and data, and then only after formal authorisation and under supervision.

Transporting the data is the riskiest operation

Security of the data transfer to the Court of Auditors is of paramount importance. The goal is to protect the data from disclosure. Make sure that you comply with the auditee's and the Court's security rules.

A Court's laptop with personal custody is very safe

The most secure way to transport data to the Court is by strongly-encrypted media under custody. A way to do this is to carry the data on the hard disk of an ECA laptop.

"PostFiles" instead of e-mail annexes

Use the "**PostFiles**" tool to receive (or send) files in a secure way. Avoid using e-mail, which is limited in size and unsafe (e.g. non encrypted transmission, multiple copies in mail servers).

Removable media sent over public channels require specific transmission protocols

If data will be sent by post/courier on a removable media (CD, memory stick), you must agree an appropriate protocol, with the approval of the security officers of the auditee and the ECA.

For sensitive data this could involve encrypting the data with a single use key plus sending the key only when the media has reached the ECA.

3: COLLECTION & TRANSPORT

Carry out the data collection and transport as planned.

Data Reception

The moment the data is received by the auditor, a few checks called "Data Reception" must be done on each file:

- Open the file and check the overall format (Excel, readable text).
- Check that the file contains the requested fields with adequate format (e.g. decimal points, date formats).
- Check the data controls defined: number of rows, monetary totals and hash totals.

4: STORAGE & TREATMENT

Storage

While in the Court of Auditors, the data obtained should be subject to the auditee's and the ECA's security rules.

For data needing stronger security than “ECA-LIMITED”, consult your Information Security Officer (ISO).

Treatment

Carry out your analysis or sampling as planned.

While manipulating the data remember to:

- keep track of the manipulations done to the data so that they can be checked or reproduced. ACL does this tracking automatically.
- keep checking the data integrity controls to detect any alteration to the data.

Archive & Disposal

After use, make sure the data is still stored, archived and later disposed of safely.

CDs and DVDs should be sent to the Information Security Officer for safe destruction.

APPENDIX . ECA data confidentiality levels.

CONFIDENTIALITY LEVELS

COURT	COUNCIL	COMMISSION	Explanation	Comment for the specific Court case
PUBLIC	PUBLIC	PUBLIC	Information whose public disclose would not damage the interest of the Institutions, Member States or other parties.	Unclassified information accessible to everybody inside or outside the Court.
INTERNAL		LIMITED + Commission Internal	Information reserved for a limited number of persons on a need to know basis whereby the information is only useful within the Institution (internal policies, procedures, rules, etc).	Information available to staff and to contractors.
LIMITED	LIMITED	LIMITED	Information reserved for a limited number of persons on a need to know basis whereby the unauthorised disclosure would be prejudicial to the Institutions, Member States or other parties.	Information accessible to selected groups of business people without naming specific persons (ex. people part of a recruitment panel).
RESTREINT	EU RESTREINT	EU RESTREINT	Cause substantial distress to individuals or financial loss or facilitate improper gain or advantage for individuals or companies or prejudice the investigation of crime or impede the effective development of operation of EU policies or undermine the proper management of the EU and its operations.	Information accessible to identified individuals (this will cover nearly all DEC-C). Some information maintained by the Legal service might also belong here.
CONFIDENTIAL	IN-CONFIDENTIAL	IN-CONFIDENTIAL	Information reserved for a limited number of persons on a need to know basis whereby the unauthorised disclosure would be prejudicial to an individual person (medical data, disciplinary investigation dossier, etc).	Cases where records should be accessible by very few persons. This category should be used mostly for confidential data concerning individuals (ex.medical records, subject to medical secret, disciplinary investigation dossier, etc).
N/A	EU CONFIDENTIAL	EU CONFIDENTIAL	Materially damage diplomatic relations or prejudice individual security or substantially undermine the financial viability of major organisations or serious impede the development or operation of major EU policies or shut down or substantially disrupt significant EU activities. Unauthorised disclosure of the information could harm the essential interests of the EU or of one or more of its Member. States.	N/A
N/A	EU SECRET	EU SECRET	Raise international tension or threaten life directly or cause substantial material damage to EU or a Member State. Unauthorised disclosure of the information could seriously harm the essential interests of the EU or of one or more of its Member States.	N/A
N/A	EU TOP SECRET	EU TOP SECRET	Information that can threaten directly the stability of the EU or one of its Member States or direct loss of life or sever damage to Member State economy. Unauthorised disclosure of the information could cause exceptionally grave prejudice to the essential interests of the EU or of one or more of its Member States.	N/A

Sources:
- COM Decision 29/11/2001 (2001/844/EC)
- COM Decision 03/02/2005 (2005/94/EC)
- COM Decision 16/08/2006 (2006/3602)

- Council Decision 19/03/2001 (2001/264/EC)
- Council's Guide on the security of information (September 2006)