Testing with personal data standards and guidance

Introduction

It is essential that testing on systems takes place in order to ensure that when the system is moved to live as many issues as possible have been resolved. The testing needs to assess the security and robustness of the system and the data it processes. Gloucestershire County Council (GCC) is basing its approach to testing systems on British Standards Institution[1] and ICO guidance.

Key points to note:

- ✓ The council will test using dummy data where possible,
- ✓ If live personal data must be used for testing that mitigations to the risks posed must be implemented,
- ✓ Where live personal data is used, the relevant Information Asset Owner (IAO) must log decisions and testing processes within a DPIA,
- ✓ There must be a documented evaluation and justification for testing with personal data,
- ✓ Must ensure privacy notice information provided to data subjects includes that testing on the personal data will be carried out,
- ✓ Considerations must be given to all the principles of GDPR.

If you do decide to test a system using personal data then you must read this guidance and should complete a Data Protection Impact Assessment (DPIA) in order to demonstrate the reasons for doing so and any steps taken to mitigate the inherent risks.

If you need to test using personal data then you should read this guidance. If you need additional assistance then please contact the Information Governance Team.

Reasons for undertaking live testing

The reasons for live testing can include:

[1] BIP 0002:2009 Data Protection: Guidelines for the Use of Personal Data in System Testing (Second Edition)

- The particular type of data to be processed or the function of the system may require the use of the live data in order to adequately test out its capabilities.
- Test environments many not be as fully built as live environments so certain components of a system may only be adequately tested in a live environment.
- It may not be possible to replicate a particularly specialised process within the test environment due to limitations on the process itself or the data it requires.
- Test environments may not be sized in proportion to the size of the live databases, therefore live testing may be necessary to assess the scalability of a system.
- There may be configuration changes to the live environment that cannot be tested in any other way due to the limitations of the test environment.
- Project conflicts may mean that a test environment is only able to support accurate load testing for one project at a time, thus it may become essential to use a live environment.
- Practical reasons: time, tester resource and cost.

If you plan on undertaking live testing, this must be justified and reasons for documented within the DPIA.

Data Masking / anonymisation

Prior to undertaking testing, considerations must be given to identify if dummy or anonymised data can be used. If dummy test data cannot be used, there are a number of techniques which can be utilised in order to effectively mask data so the risks are mitigated;

- **Substitution**

  This technique consists of randomly replacing the contents of a cell or column of data with information that looks similar but is unrelated to the real information. For example, real names are replaced with those drawn from a random list.

- **Shuffling**

  Similar to substitution, but you replace the value of a cell with data derived from other cells within that column. This method is not so useful in smaller databases where the lack of variety between values may cause an issue.

- **Number and Date Variance**

  This method is useful for numeric and date values. An algorithm is implemented that modifies each value (such as date of birth) by a random

percentage. This technique has the advantage of being a reasonable way of masking data whilst maintaining its usefulness.

Pseudonymisation

GDPR defines Pseudonymisation as;

> 'the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject (the person the data relates to) without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person'.

An example of pseudonymisation would be to remove the names of individuals from a database, and replace them with randomly generated reference numbers. Pseudonymisation may also be achieved by scrambling data across rows.

Pseudonymisation contributes to ensuring the security of the data and reduces the risks to data subjects through the use of their data. However, pseudonymised data is still considered by the GDPR to be personal data and any use for testing purposes must still comply with the first and second principles of GDPR, and a DPIA should be completed.

Types of system testing

System testing may take one of the following forms:

- Dummy data in a test environment
- Dummy data in a live environment
- Scrambled or anonymised data in a test environment
- Scrambled or anonymised data in a live environment
- Live data in a test environment
- Live data in a live environment

The type of system that is being tested will dictate which type of testing takes place.

When considering testing with live data in a test or live environment [both of these types of testing will be referred to hereafter as 'live testing'] the business must identify a specific need for live testing.

National Identifiers

Care must be exercised in the use of national general identifiers, such as national insurance, NHS or pupil identification numbers. Their processing may be prohibited. If you intend to test on national identifiers check with the organisation they originate from as to whether that processing is permitted.

The risks of system testing

These include:

- Unauthorised access to data;
- Unauthorised disclosure of data;
- Intentional corruption of data;
- Unintentional corruption of data;
- Compromise of source system data;
- Loss of data;
- Inadequacy of data;
- Objections from customers.

Business continuity plans

System testing should be included in business continuity or disaster recovery plans for two reasons.

1. Following a major incident, it may be necessary to test existing systems or data before starting to use them again.
2. Test systems may need to be periodically recovered and reloaded.

Key principles

Prior to undertaking testing with live personal data you will need to work through and document how the testing will comply with the data protection principles. Guidance on how to meet some of the requirements is set out below. More information can be found on the IMS staffnet pages.

Principle 1 – Lawful, fair and transparent processing

System testing is considered as processing as defined in GDPR and the Data Protection Act (DPA) 2018 and as such organisations must have a GDPR Article 6 lawful basis in order to carry this out.

You must consider whether or not the testing of that personal data is compatible with the original lawful basis (e.g. when seeking consent from the data subject did you get express consent for their data to be used in testing or when issuing a privacy notice did you advise them that testing using their data would occur?). If you have any

concern about this then you should log it as a risk in the DPIA and seek advice from IMS.

GDPR places stronger emphasis on transparency and the need to inform data subjects about how their data will be used. This is usually done through Privacy Notices. If personal data is to be used in testing situations then the privacy information must explain that is going to happen.

Principle 2 – Purpose limitation

GDPR places a strong emphasis on the need to only process personal data for specific purposes and the council treats testing of personal data as a purpose in itself.

If you do not include testing as a specific purpose at the time you provide privacy information you will need to consider what GDPR says about compatibility. To decide on whether a new purpose is compatible you should consider:

- any link between your original purpose and the new purpose;
- the context in which you originally collected the personal data – in particular, your relationship with the individual and what they would reasonably expect;
- the nature of the personal data – e.g. is it particularly sensitive;
- the possible consequences for individuals of the new processing; and
- whether there are appropriate safeguards in place – e.g. encryption or pseudonymisation.

You should undertake a Compatibility Assessment to identify if the purpose is compatible. Contact the Information Governance Team for advice on how to do this.

Where testing with live personal data is unlikely to have any detrimental affect on the data subject (for instance, testing that a new module for a system can pick up data brought in from an existing part of the system) then it is easier to argue that it is compatible with the original purposes.

If live personal data is to be disclosed for testing purposes to a third party then you must ensure that data subjects are aware of this and that you have an appropriate information sharing agreement with the partners in place governing how the data may and may not be used.

Principle 3 – Personal data is adequate, relevant and limited to what is necessary.

All personal data used in testing must be strictly relevant to the purpose of testing. Individual data items should be listed and identified as non-personal, personal or sensitive personal. Once data has been classified, reasons for inclusion in testing should be provided. Where a reason for inclusion cannot be found, the personal data must not be used.

Principle 4 – Personal data is accurate

This principle is most important when testing any system that matches, cleanses or in any way changes data. The method of testing needs to ensure the accuracy of data is upheld and any risks to accuracy are mitigated.

At the council it is important that a system testing regime is implemented to maintain an audit trail to capture errors during the testing process and allow for immediate correction. Accuracy checks must be carried out on the data being fed into the system especially if there is a possibility of that data being merged with other data or fed back into the source system.

Principle 5 – Personal data is not kept for longer than is necessary for that specified purpose or those purposes

You must have mechanisms in place to ensure that test data is not retained for longer than is required to fulfil legal, regulatory or business requirements.

The service area must consult with IMS on an appropriate retention period for the testing data. Where test data is retained after testing is complete and still constitutes personal data, it may be provided as part of the response to a subject access request and this must be considered when establishing retention schedules.

Principle 6 – Personal data is kept secure (integrity and confidentiality)

The use of dummy data or test accounts should be used where possible. There should be tight controls on creation, activity and closure of such dummy data and the number of dummy accounts that exist at any time. Such accounts should be fictitious and clearly distinguishable from live accounts so as to not cause more risks.

Organisational measures to render testing safer include:

- User accountability is vital: maintenance of secure audit trails, surveillance and tracking methods as appropriate.
- Any activity carried out with personal data will then take place against a well-established background in DPA compliance.
- Must use a secure IT infrastructure and place safeguards to ensure that sensitive information is not viewable by those who should not have access to it.
- The Information Asset Owner should be made accountable for the use (and testing) of relevant personal data within the council.
- A system testing process must be created, with input from all key business areas and should clearly detail roles, responsibilities and requirements in respect of system testing.
- A Data Protection Impact Assessment (DPIA) should be carried out by the service area in order to document the testing process, information risks and mitigations for those risks.
- Segregation should be employed where ever possible as a safety measure. For example, those who authorise a task should not undertake it or check it. Segregation of location would mean different functions should not be carried out in the same location at the same time.
- Test environments - where data cannot be scrambled or anonymised, it should be tested on a totally separate, secure and isolated test system wherever possible.
- Procedures for bug fixing must be in place.
- Backup facilities must be available where data from two or more systems is being compared or where testing takes place continuously on the test system.
- If purchasing a product “off the shelf”, or using a testing environment that is part of a suite of products already in use by the council, that environment should be subject to the same scrutiny and monitoring as a tailor-made solution. It should not be assumed to be compliant simply because it is provided by a reputable vendor.
- Remote working – where remote working has to be carried out (e.g. internet or extranet applications) this must be allowed only after a full assessment of the additional risks that result from transferring test data to and from an external location. It should be regarded as a very high-risk activity and assessed, managed and monitored accordingly.

The accountability principle

It is vital that documentation exists that covers the processing of the personal data for testing, identifies how the council complied with the principles of GDPR, what risks to the data exist and how the council mitigated those risks.

A completed Data Protection Impact Assessment should be used to demonstrate compliance with this principle.

Compliance is demonstrated through a number of ways;

- Completion of a Data Protection Impact Assessment (DPIA) which reviews the nature, context and content of the processing and sets out the mitigation of risks;
- Review of the processing by the council's Data Protection Officer (DPO) where there is a high risk;
- Ensuring that the information asset being tested appears on the corporate Information Asset Register, and that the entry is up to data and reflects that testing of live personal data takes place;
- If the testing is to be carried out by a third-party supplier, then ensuring that the Cyber and Information Management (Procurement) Policy has been followed and that appropriated GDPR clauses exist within the contract with that supplier;
- Retention of audit and access logs from when the testing took place.

Failure to demonstrate compliance, even if no breach has taken place, could result in the ICO undertaking an investigation and either ordering a halt to the processing or potentially issuing a fine.

Personal data is processed in accordance with the rights of data subjects

Please refer to the Information Rights Policy for further information on the rights given to data subjects by GDPR and DPA 2018.

Where there is a likelihood of test data entering live systems, there is a possibility of loss, damage or corruption which may lead to claims of substantial damage and distress. If there is any likelihood of this or any other aspect of testing that could potentially give rise to claims against the council, then alternative methods of testing must be found which do not involve using live personal data.

International transfer to countries outside of the EEA

Where testing takes place by a third party on behalf of the council within the EEA that organisation will be required to meet its obligations under GDPR and the agreement the council has in place with them. Testing with personal data taking

place outside of the EEA should be prohibited. If an Information Asset Owner feels this is necessary then they should contact dpo@gloucestershire.gov.uk and liaise with the Information Assurance Team.

Breaches of the DPA: What to do if things go wrong

In the event of a breach of data protection occurring during system testing, you are required to follow the Incident Response and Escalation Procedure and immediately email informationsecurity@gloucestershire.gov.uk.

Important Note

The ICO (Information Commissioners Office) advises that the use of 'live' personal data for system testing should be avoided. System administrators should wherever possible develop alternative methods of system testing. Should the Information Commissioner receive a complaint about the use of personal data for system testing, their first question to the council would be why no alternative to the use of personal data had been found.

Breaches of personal data used in live system testing could result in the ICO investigating and fining the council, as well as having an impact on the council's reputation with our customers.

Examples of the impact of using personal data for testing purposes when things have gone wrong can be found below;

- *A pupil was away from home at boarding school. The pupil's parents received a letter from the local hospital informing them that their daughter had been involved in a road accident. In fact, there had been no accident, but the hospital had been using live patient data to test a system for sending out letters to patients.*
- *A US pension service revealed that an unauthorised person had gained access to one of its databases hosted in a test environment. The system security was not as robust as that in the live system, despite holding information such as names, addresses, dates of birth and beneficiary details.*

Document information and review

Owner:	Jenny Grodzicka, Head of Information Management (DPO)
Last Reviewer:	Nick Holland, Callum Crossan & Ben Crow
Date created:	December 2020
Next review date:	December 2021
Approval:	Information Board, 05/02/2021
Version:	2
Classification:	**UNCLASSIFIED**

Version History

Version	Version date	Summary of Changes
0.2	September 2015	First version
1	July 2020	Revised version – updated in light of GDPR requirements
2	December 2020	Reviewed and amended by NH/CC/BCr