Data Protection Impact Assessment (DPIA) - Full Assessment

Guidance for the Project Manager and Sponsor

Use the pre-screening template first. If that shows a high risk in processing the data then you must carry out this full DPIA. **Do not complete this form unless you have already completed the pre-screening and it shows high risk and the DPO as advised you to do a full DPIA.**

The Data Privacy Impact Assessment (DPIA) will enable you to systematically and thoroughly analyse how your project or system will affect the privacy of the people whose data you are dealing with and show how you will minimise the privacy risks. This template has been designed to incorporate the legal requirements of the General Data Protection Regulation (GDPR) and the Data Protection Act 2018. Conducting a DPIA is a legal requirement under the GDPR particularly if the proposed processing is using new technologies and poses a high-risk to people's data. Further information and guidance on the DPIA is also available on the ICO website here: ICO's PIA code of practice and the Article 29 Working Party here.

GOVERNANCE ARRANGEMENTS

This DPIA will be submitted to the Corporate Information Governance Group (CIGG) and the advice of the Data Protection Officer (DPO) will be sought as part of that process. You must keep the signed DPIA and all supporting documents with your project file for audit purposes.

1. **PROJECT SUMMARY**

Project Name	Data Lake	**Directorate and Service**	Corporate Services ICT
Project Sponsor and Position	Sudip Trivedi Head of Data and Analytics	**Project Manager and Position**	, Project Manager
Project Start Date **Project End Date**	April 2019 March 2020	**Project Go Live Date (anticipated/planned)**	End of April 2019

Third parties involved/associated with the Project:	Amazon Web Services (AWS)	**Does this DPIA cover multiple projects?**	Yes (Will produce DPIA Pre-screening assessment form for every use cases which will include details of the data)

High Level description of the Project:

- This project is about delivering a cloud based data storage capability (Data Lake) for the council, which will provide a unified mechanism to systematically store structured data from any council systems or database, semi-structured data (CSV, logs, XML), unstructured data (emails, documents, PDFs) and binary data (images, audio, video) allowing for its processing and making it available for reporting and analysis.
- Purpose of completing this DPIA is to have a framework in place which covers the Data lake as a whole. We will continue to complete pre-screening assessment form for individual use cases which will contain details about the project, what data will be stored and processed, who will have access and how long the data will be kept.

2. DESCRIPTION OF THE PROJECT

Include here a plain English description of:

- Over the years, Camden has had a sophisticated approach to the use of data and working with the partners. Most of that data is still held in 'silos' and it can be costly and time consuming to bring together a view of citizen, place and service data.
- We have now developed a Data lake solution that will allow us to bring a mass amount of data from multiple sources in one place to be able to better understand the needs of residents, businesses and partners.

What is a Data Lake from the Council's Perspective?

- A cost effective solution to capture, store, analyse, protect and manage mass amount of data in one place.
- Ability to store all types of structured and unstructured data, such as Adult Social Care data from Mosaic, PDFs from FOI requests, sensors data from SMART meters etc.
- Flexible infrastructure on the Cloud

    - Ability to increase storage and compute capability as needed
    - Ability to move data back from the cloud to on premise data centre if required
- A layered solution to data management
    - Ability to consume raw data from multiple sources, blend and analyse the data and present findings through visualisation tool.
    - Lab area within the Data Lake will allow analysts/data scientists to work with the raw data, save results and either purge or promote.
- Allow analysts with SQL skills the ability to query data stored in the Lake
    - Data analysts across the council can explore and interrogate data using SQL query language
    - For the non-analysts or general users data output can be presented through a dashboard or visualisation tool
- Data Lake will provide ability to share data securely with external partners such as NHS or other councils.

- The purpose of having a data lake is to ensure that the Council have a cost effective and flexible data storage and processing capability that is currently not available through any other tools. As an example we are already processing thousands of Freedom of Information responses to extract text and provide Google style search capability and intending to store thousands of audio files (customer call record) from the Contact centre and produce trend analysis.
- Data can be uploaded in the Data Lake either manually by the system administrator or automated schedule job. Once the data is stored in the Data Lake developers and the data analysts will have role based access to transform the data and prepare in the required format then present findings using Qlik Sense dashboards.
- Data will be stored in the secure storage area within the AWS platform. Access to the data will be controlled by the role based security set by the system administrator.
- Data Lake provides the capability to collect and store mass amount of data from multiple sources. So far we have stored Freedom of Information responses files in the data lake which are available on the Open Data platform. At this point it is not clear which data sets will be stored and their sources. Project will produce DPIA pre-screen assessment form detailing the data sets and source locations for each of the use cases as it happens.
- Data will be stored in the Data lake according to the use case requirements which will be detailed in the pre-screening assessment form. Nevertheless Data Lake provides the capability to automatically delete data after a set period.

pes of personal data to be processed and data flow map(s):

Personal data:

List the types of data that you intend to process and the types of data subject (for example, names, addresses of residents, service users etc):

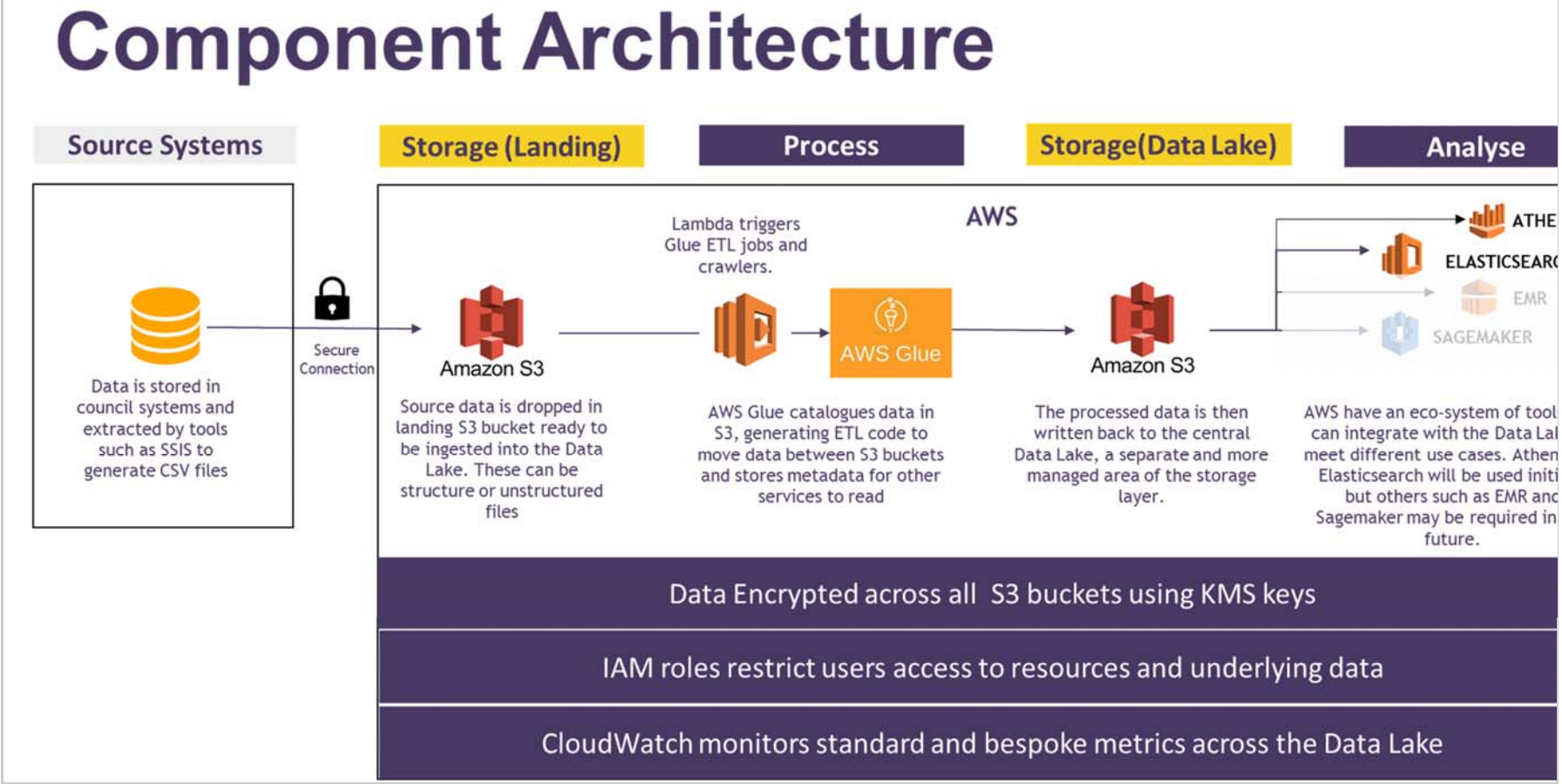
- We are intending to process below personal data in the Data Lake, and further details of the data will be provided in the pre-screening assessment form.
    - Names
    - Addresses
    - Contact Number
    - Email Addresses

Special category data:

List the types of special category data and the types of data subject:

- We are not intending to process any special category or criminal conviction data at this stage

Data Flows:


Component Architecture
Source Systems
Storage (Landing)
Process
Storage(Data Lake)
Analyse
AWS
Data is stored in council systems and extracted by tools such as SSIS to generate CSV files
Secure Connection
Amazon S3
Source data is dropped in landing S3 bucket ready to be ingested into the Data Lake. These can be structure or unstructured files
Lambda triggers Glue ETL jobs and crawlers.
AWS Glue
AWS Glue catalogues data in S3, generating ETL code to move data between S3 buckets and stores metadata for other services to read
Amazon S3
The processed data is then written back to the central Data Lake, a separate and more managed area of the storage layer.
ATHE
ELASTICSEARC
EMR
SAGEMAKER
AWS have an eco-system of tool
can integrate with the Data Lal
meet different use cases. Athen
Elasticsearch will be used initi
but others such as EMR and
Sagemaker may be required in
future.
Data Encrypted across all S3 buckets using KMS keys
IAM roles restrict users access to resources and underlying data
CloudWatch monitors standard and bespoke metrics across the Data Lake

3. DATA PROTECTION PRINCIPLES

This section demonstrates how the project meets the data protection principles.

- How will you make sure that you only process the data that is necessary and proportionate for the purpose of the project, and no more than is necessary?
- If the data was originally collected for one purpose and you intend to use it for another purpose, explain how you will inform the data subjects.
- How will you make sure that the data is kept accurate and up to date?
- How long will you keep the data for and how will you destroy it at the end of the retention period?

• In the Data Lake we have role based access to the data fields, this will ensure that access is permitted to the required data sets only. • Data retention will be based on the use case and in line with the GDPR regulation

- Have you cleared the information security arrangements with the Information Security Manager? YES

- **Record the Information Security manager's comments here:**

The technical design was approved by the Shared Digital Security Manager in May 2018, I have checked with the Camden IT security Lead

4. BASIS OF PROCESSING

- Which legal basis in Article 6 are you relying on? See this guide to help you identify the legal basis https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/
- If you think you need to rely on legitimate interests then ask the Information and Records Management Team for advice.
- If you are processing special category data, you will also need a legal basis under Article 9 to process this. See this guide to help you identify the legal basis **https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/lawful-basis-for-processing/special-category-data/**
- If you are processing criminal convictions data or data for law enforcement reasons then you should speak to the Legal team as you need an additional legal basis to do this.

Article 6(1)(e) public task: Processing is necessary for you to perform a task in the public interest or for your official functions, and the task or function has a clear basis in law.

5. DISCLOSURES OF DATA

- Will you be transferring/ sharing/giving this data to a data processor or a sub-processor? **NO**
- Tick here to agree that you will be entering into a data processing agreement with them **[√]**
- Will you be sharing data with any other third party? **NO**
- List the third parties that you propose to share with:
- Tick here to agree that you will be entering into a data sharing agreement with the third parties **[ ]**

6. TRANSFERS OF DATA OUTSIDE OF THE EEA

Will any personal data be processed outside of the EEA? NO
See a list of countries here: https://www.gov.uk/eu-eea

If your answer is yes, you must consult the DPO straight away, and see the guidance here:
https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-gdpr/international-transfers/

If there WILL be a transfer out of EEA enter comments of the data protection advisor:

To contain comments detailing the safeguards that need to be put in place.

7. DATA SUBJECT RIGHTS AND COMPLIANCE WITH CORPORATE POLICIES

Information in Camden contains the Council's policies and procedures on data protection compliance, including how to respond to requests from people to enforce their rights under data protection law.

- You must comply with the requirements in Information in Camden. Tick here to agree that you will be complying with IIC on Data Subject Rights **[√]** If there is a reason why you cannot do this, please explain why here:

8. CONSULTATION WITH INTERESTED PARTIES

Is one of the outcomes of your project going to make a change which will have a direct effect on data subjects?, For example: introducing CCTV into a library? If so, contact the Information and Records Management Team for advice at dpa@camden.gov.uk about whether you need to consult with stakeholders. N/A

Record the comments of the data protection adviser here:

To include advice on whether consultation is necessary and the steps to take.

9. RISK ASSESSMENT AND MITIGATION

Risk is a combination of **impact**- how bad the effect of the risk would be- and **probability** – the likelihood of the risk happening**.** Risk is assessed from the perspective of the data subject (as opposed to risk to the Council) and what the impact could be on them as a result of the proposed data processing. For each of the risks you identify:

1. think about how likely they are to occur and categorise them according to **Table 1 in the appendix (e.g., rare, unlikely etc)**.

2. Then consider the impact each risk will have and categorise them according to **Table 2 in the appendix (e.g., minor, moderate etc)**.

3. Then look at **Table 3** and see the risk level. Where the level says mitigations are needed, think about what these will be and how they will reduce the risk level down.

4. Enter the details in the grid below

There is more information on the council's approach to risk here
https://lbcamden.sharepoint.com/sites/intranet/finance/Pages/Risk_Management.aspx

Risk 1 *[include as many rows as necessary to identify each risk individually]*	**Risk Level Before any Mitigations**	**Risk Level After Mitigations**
Source of risk: *Governance and access control to the data in the Data lake* **Potential impact on individuals:** *Personal Data is accessed by someone who is not permitted to view such data* **Threats that could lead to illegitimate access, undesired modification and disappearance of data:** *Unauthorised access to the data can lead to illegal processing of data, modification and loss of data.* **Any compliance or corporate risks?** *Risk of breach or GDPR and corporate procedures on information security.* **Where mitigations are required what are these?** *To mitigate the risk we have implemented* • *Role based access control to the relevant data only, this will also guarantee access control at the granular level of data.* • *We have also configured audit function, which will record who have accessed which data.*	6 Possible and minor (depending on the nature of the data)	4 Unlikely and minor.

Risk 2	Details and Risk Level Before any Mitigations	Risk Level After Mitigations
Source of risk: Data will be hosted on a Cloud based platform which is outside of Camden network **Potential impact on individuals:** Due to the fact that data is residing outside of the Camden network, personal data can be hacked, lost or stolen and misused. **Threats that could lead to illegitimate access, undesired modification and disappearance of data:** *This could lead to data breach, incompliance with GDPR, Financial penalty and reputational damage.* **Where mitigations are required what are these?** As a mitigation to this risk we have ensured • A robust governance process • Contractual agreement with Amazon to ensure robust data security control are in place • General awareness for the data Analysts while accessing the data	6 Possible and minor (depending on the nature of the data)	Unlikely and minor.

10. OVERALL RISK RATING FOR THE PROJECT AS A WHOLE ONCE THE MITIGATING MEASURES HAVE BEEN PUT IN PLACE:

LOW	**MODERATE**	**MEDIUM/ HIGH**	**HIGH**

ANNEX A: DATA FLOW MAPS

ANNEX B Risk Assessment Tables

Table 1 Likelihood of Risk Occurring

Rare	One-off failure
Unlikely	Possible that it may reoccur but not likely
Possible	Might happen or reoccur on a semi-regular basis (no more than once a quarter)
Likely	Will reoccur on a regular basis, pointing to some failure in controls
Almost Certain	Wilful act, systemic failure in controls

Table 2 Impact of Risk if it occurs

Negligible	No personal data involved, or risk won't have any impact.
Minor	• Short-term, minimal embarrassment to an individual • Would involve small amounts of sensitive personal data about an individual • Minimal disruption or inconvenience in service delivery to an individual (e.g. an individual has to re-submit an address or re-register for a service)
Moderate	*More than a minimal amount of sensitive personal data is involved at this level* • Short-term distress or significant embarrassment to an individual or group of individuals (e.g. a family) • The potential of a financial loss for individuals concerned • Minimal disruption to a group of individuals or significant disruption in service delivery or distress to an individual (e.g. availability to a set of personal information is lost, requiring resubmission of identity evidence before services

Major	Significant amount of HR, or resident personal, and / or sensitive data released outside the organisation leading to significant actual or potential detriment (including emotional distress as well as both physical and financial damage) and / or safeguarding concerns
Catastrophic	Catastrophic amount of HR or service user personal and or sensitive data released outside the organisation leading to proven detriment and / or high-risk safeguarding concerns. Data subjects encounter significant or irreversible consequences which they may not overcome (e.g. an illegitimate access to data leading to a threat on the life of the data subjects, layoff, a financial jeopardy)

Risk Assessment: Table 3

IMPACT	Score:	PROBABILITY				
		Rare	Unlikely	Possible	Likely	Almost Certain
	Catastrophic	5	10	15	20	25
	Major	4	8	12	16	20
	Moderate	3	6	9	12	15
	Minor	2	4	6	8	10
	Negligible	1	2	3	4	5

Level of risk	
1-3 Low Risk	Acceptable risk No further action or additional controls required Risk at this level should be monitored and reassessed at appropriate intervals
4-6 Moderate Risk	A risk at this level may be acceptable, if so no further action or additional controls required If not acceptable, existing controls should be monitored or adjusted
8-12 Medium / High Risk	Not normally acceptable Efforts should be made to reduce the risk, provided this is not disproportionate Determine the need for improved control measures
15-25 High Risk	Unacceptable Immediate action must be taken to manage the risk A number of control measures may be required

Annex C:
Any DPO Advice or comments not included above