



BUTTERFLY
DATA

G-Cloud 15 | Lot 3

SERVICE DEFINITION DOCUMENT

Data science - **Advanced analytics and artificial intelligence (AI)**

Data science and AI services that deliver predictive insights, generative models, and secure, ethical analytics

Contents

1. About Butterfly Data	3
1.1. Our core propositions	3
1.2. Our technology partnerships	4
1.3. Cost-effective delivery	4
1.4. Our values, purpose and social commitment.....	5
1.5. What our clients say	6
2. Service overview	7
2.1. Core features and benefits	7
3. Detailed service description	9
3.1. Machine learning (ML)	9
3.2. ML tools	9
3.3. Generative AI	9
3.4. Network analysis	9
3.5. Prototyping.....	10
3.6. Fairness and bias	10
3.7. MLOps	10
3.8. Data governance and safety	11
4. Our approach.....	12
4.1. Assumptions	12
4.2. Outputs and deliverables	13
5. Standardised service framework.....	14
5.1. Project roles and responsibilities.....	14
5.2. Discovery phase.....	17
5.3. Setup, migration, and delivery	17
5.4. Knowledge transfer and training.....	18
5.5. Ongoing support and maintenance.....	18
5.6. Service management.....	19
5.7. Service standards	20
6. Case studies	21

Appendix 1: Other approved frameworks	23
Appendix 2: Full list of G-Cloud 15 services	24
Appendix 3: Social value response.....	25
Appendix 4: Common acronyms, abbreviations and keywords	28

SECTION 1

About Butterfly Data

An overview of who we are, our services, strategic partnerships, value proposition, social initiatives, and client testimonials



BUTTERFLY
DATA

1. About Butterfly Data

Butterfly Data is an employee-owned, [B-Corp](#) data consultancy dedicated to helping organisations turn national-scale and high-complexity citizen data into actionable intelligence. With over 20 years' experience, we deliver data solutions to public and private sector clients, including extensive experience working across multiple public sector frameworks (ref. [Appendix 1](#)), combining technical expertise with ethical, socially responsible practice.

Our security-cleared team specialises in services spanning the entire data management lifecycle, including data strategy, quality, governance, cloud adoption, data engineering, analytics, machine learning (ML) and artificial intelligence (AI) to ensure that your organisation can **put data behind every decision**. We enable your organisation to:

- Gain clarity and confidence in your data by ensuring it is accurate, governed and ready for use.
- Break down silos by unifying data sources and building scalable data platforms.
- Derive actionable insights, support decision-making, forecast risks or opportunities, and track performance.
- Automate reporting, reduce manual overheads, and embed a data-led culture across your organisation.
- Leverage AI and ML responsibly to unlock new capabilities, efficiencies, and strategic benefits.

Our client base spans central government, local and regional authorities, healthcare, education, transport and the wider public sector, as well as private sector organisations across multiple industries. We work closely with you to understand your objectives and challenges, providing tailored solutions that improve operational efficiency, transparency, and outcomes.

1.1. Our core propositions

✓ Data quality, governance and privacy

Improve data quality, compliance and confidence in your data.

☁ Data engineering and cloud adoption

Create scalable cloud platforms with robust data pipelines for multi-source integration.

🔍 Data analytics and visualisation

Uncover actionable insights and make informed decisions.

🌟 Data science and AI

Deploy custom machine learning models and ethical AI to optimise, automate & enhance.

Refer to [Appendix 2](#) for the full list of our G-Cloud services.

View more about our services



1.2. Our technology partnerships

We have delivered successful data and cloud technology projects across the UK, leveraging platforms such as [SAS](#), [AWS](#), [Microsoft Azure](#), [Snowflake](#), [Informatica](#), [Databricks](#), and other leading technologies. Our capabilities have been externally recognised through achieving [SAS Hackathon Global Grand Champion 2024](#) and other awards, demonstrating advanced applied analytics and cloud delivery expertise. Our vetted teams combine technical capability with regulatory knowledge and security best practice, enabling you to manage sensitive data confidently and meet compliance requirements.



[View more about our partnerships](#) →

1.3. Cost-effective delivery

We deliver value through expertise, operational efficiency, and a client-focused approach. Our teams are purposefully specialised and agile, minimising bureaucracy, while keeping costs competitive for global-scale innovation.

Onboarding: Our familiarity with data, processes, and project management enables faster onboarding, reducing both time and cost for you.

Breadth: With extensive experience across multiple sectors and specialisms, we efficiently deliver high-quality work. Our streamlined delivery process means new projects are executed faster, with fewer iterations.

Knowledge transfer: We work collaboratively with you to ensure our knowledge becomes yours. This builds internal capability, supporting ongoing development, maintenance, and refinement of data quality and systems.

Experience: Our consultants bring collective experience and practical insight from previous projects to ensure solutions maximise productivity, improve processes, and deliver measurable benefits.

Optimisation: We design solutions to be both optimised for your current systems and adaptable to future needs, refining deployment pipelines to support efficient updates and expansions.

Transparent costs: We only charge for the time spent delivering value. Our focus on efficiency means projects may be completed without utilising the full budget, providing additional cost-effectiveness.

1.4. Our values, purpose and social commitment

Butterfly Data believes that data can, and should, be used for good. As a B-Corp and signatory to the [Pledge 1%](#), we are committed to combining technical excellence with ethical, sustainable, and socially responsible practice.



Ethics, privacy and trust: We prioritise data quality, governance and privacy in every engagement, ensuring compliance with regulatory requirements and fostering trust in data-driven decision making.

Personalised and inclusive approach: We recognise that every organisation is unique. Our team builds a close, collaborative relationship with you to deliver bespoke solutions aligned with your goals, complexity, and operating environment.

Responsible innovation and social impact: We believe in applying data science and AI not simply for efficiency or profit, but to generate measurable, positive impact - whether that's through improving public sector services, enabling better decision-making, or supporting societal goals.

Sustainable business practices: We strive to minimise our environmental footprint and promote sustainable operations, committed to working towards net zero, while retaining agility and responsiveness.

Refer to [Appendix 3](#) for our full social value response.

View more on our social value →

1.5. What our clients say

“ Butterfly Data have delivered a high-quality analysis of our data that shows the level of insight that can be drawn out of it, which will have an impact on future strategy and work planning for our organisation.

FOREIGN, COMMONWEALTH, AND DEVELOPMENT OFFICE (FCDO)

“ Butterfly Data have very competently demonstrated the art of the possible in tackling a really complex corner of the UK Government's data and policy landscape. They successfully brought together 3 Government departments around a shared need, developed a really valuable (and really deep!) data model / platform / visualisation, and engaged senior leaders effectively. We're really excited to see where their work goes next - and what the impact on the UK's future is as a result.

GOVERNMENT CONSULTANT

“ The extent of the dialogue and engagement with Butterfly Data has been fantastic, particularly the willingness to spend time in the office to really understand the concepts, challenges and vision. This has led to a final report which is absolutely spot on in terms of its relevance and actionability.

COUNTER TERRORISM POLICING – BORDERS

“ The invaluable work that Butterfly Data have undertaken with a key collaborator of mine will feed directly into my work, making it both simpler and faster and enabling me to better identify data gaps. Incredibly useful. Thank you.

HOME OFFICE

“ Butterfly Data conducted interviews with 15 strategic leaders in national policing to work out how they currently use data in their roles, and how they feel data use needs to change to improve policing. This was the foundation of the next stages of the data journey, which resulted in building a digital data product. We still refer to Butterfly Data's user story outputs on our data journey.

HIS MAJESTY'S INSPECTORATE OF CONSTABULARY AND FIRE & RESCUE SERVICES

SECTION 2

Service overview

A list of the core features and benefits of this specific service



BUTTERFLY
DATA

2. Service overview

We apply statistical modelling, deep learning and other methods to answer some of your biggest questions or to automate tasks like classification, forecasting and anomaly detection to power your digital services and insight generation. We utilise managed cloud services on AWS, Azure and GCP, alongside SAS and open-source frameworks like PyTorch and TensorFlow. Beyond traditional ML, we integrate generative AI to process unstructured data and generate synthetic data, including text and imagery. When you want to go beyond a proof-of-concept, we implement reliable MLOps pipelines for end-to-end lifecycle management. We prioritise responsible AI, testing for bias and fairness prior to deployment.

2.1. Core features and benefits

- **Machine learning (ML)** – Statistical modelling and deep learning approaches to automate classification, forecasting, anomaly detection and more.
- **ML tools** – Managed services such as AWS Sagemaker, Azure Machine Learning, Vertex AI on GCP and SAS or open source tools like PyTorch, Tensorflow, XGBoost and keras via Python or R.
- **Generative AI** – Integration of leading LLMs and diffusion models into services and data processing workflows to aid analysis of unstructured data, enable natural language interaction and generation of synthetic content.
- **Network analysis** – Understanding patterns and behaviours in complex systems by mapping and analysing people’s behaviours or entities and their relationships.
- **Cloud platforms** – Scalable and cost-efficient training, maintenance and deployment of ML solutions via managed services such as AWS SageMaker, Azure Machine Learning or Vertex AI on Google Cloud Platform, Databricks and Snowflake.
- **Prototyping** – Quick iteration of experimental ML approaches to evaluate their feasibility and deliver proof of concept before larger investments.
- **Model evaluation** – Testing of model performance against your business requirements.
- **Bias and fairness** – Ensuring effective and responsible ML deployment by investigating and mitigating bias of 3rd-party or bespoke machine learning models.
- **MLOps** – Automation and streamlining of your ML lifecycle, from data preparation to training, deployment, monitoring and maintenance using established DevOps principles and tools like MLflow, Kubeflow, Metaflow, BentoML or SAS Model Studio.

- **Data governance** – Holistic approach to data security, quality, auditing and compliance from Certified Data Management Professionals (CDMP)[®].
- **Agile delivery** – Reliable project, risk and change management through agile methodology for a thorough, flexible and outcome-oriented delivery model with frequent checkpoints.
- **Future-proof solutions** – We take the time to understand your organisation's requirements to design solutions that respect and support your business processes, customer-facing services and wider technology landscape.
- **Security clearance** – Vetted consultants with SC, DV or NPPV3 clearance trained to safeguard sensitive data and work in secure environments in the cloud or on your premises.

SECTION 3

Detailed service definition

An expansion on the features and benefits to uncover the tasks we typically take within this service offering



BUTTERFLY
DATA

3. Detailed service description

This document gives you an overview of tasks that we typically undertake for our clients when applying data science solutions to their business challenges and digital services. Should you require services that are not listed here we will gladly engage with you to find a suitable solution for your specific requirements, advising you on the best possible combination of tools and technologies to achieve your goals.

3.1. Machine learning (ML)

We apply a wide range of statistical modelling approaches and machine learning techniques to answer your biggest, most important questions, or to reliably automate tasks. From explainable regression that models relationships or classifies outcomes and forecasting with time series models for advanced analytics, to powerful deep learning solutions that recognise complex patterns in images, text, or audio to solve computer vision challenges, detect anomalies, transcribe audio and more. Our consultants map your business objectives to ML solutions and train, manage and serve the models using the latest tools and technologies.

3.2. ML tools

We can set up ML functionality that is built into the software you already know and own, for example integrating Power BI with Azure Machine Learning services or using tools such as SAS Visual Analytics, Visual Investigator and Intelligent Decisioning. If you prefer an open source solution, our data scientists develop bespoke models that are purpose-built for your specific business challenges using leading Python and R packages such as PyTorch, Tensorflow, scikit-learn, NetworkX, Caret, XGBoost and keras.

3.3. Generative AI

Whether you want to enhance your products or operations through a Large Language Model (LLM), or explore the current state of the art in image generation with diffusion models, we can help you integrate and evaluate generative AI solutions. Our team has experience working in Natural Language Processing with commercial LLMs via APIs, as well as open source options that can be deployed locally or on your cloud infrastructure for added privacy if your data is highly sensitive.

3.4. Network analysis

Network analysis offers you powerful methods to understand the hidden patterns and behaviours of people or entities in complex systems. By mapping and analysing

the relationships in your data we can help you to identify communities that hold shared beliefs from survey data, make visible behavioural patterns in communication data, model the spread of diseases or misinformation, optimise logistics and travel, and more.

3.5. Prototyping

There are many machine learning algorithms to choose from, and not every option will be a good fit for the challenge at hand or the data available to you. We help you understand the potential and limitations of AI for your specific business context before committing to large investments. We do this by iterating quickly to evaluate the most promising algorithms. Usually, this also involves the creation of a proof-of-concept model that demonstrates what level of capability is realistic to achieve.

3.6. Fairness and bias

We thoroughly test your models to ensure that the model's performance meets your business needs. For example, requiring reasonable compute power and delivering predictions with reasonable latency. Special attention is also paid to ensuring that predictions are accurate, including minimising the types of errors that would have an outsized negative impact on the usefulness and safety of the model for end users. A model can perform well across the entirety of your data but fail to provide good predictions for some subsets (e.g., the classification of jobseekers of minority ethnicities could be disproportionately inaccurate). That is why we take steps to identify bias and fairness issues and take corrective action, where possible.

3.7. MLOps

We streamline your entire machine learning lifecycle, from data preparation and model training to deployment, monitoring and retraining in production using open-source tools like MLflow, Kubeflow, Metaflow and BentoML or commercial enterprise solutions like SAS Model Studio. We do this by applying DevOps principles like continuous integration and continuous development (CI/CD) to ensure ML models are reliable, scalable and deliver business value efficiently and consistently. Enhanced automation and collaboration speed up training and maintenance while robust systems and processes ensure reliability and scalability of the ML solutions you integrate into your operations and services. Our MLOps approach helps you manage your models and related data consistently, ensuring governance, compliance and quality.

3.8. Data governance and safety

Our consultants are experts in data governance. All solutions we develop are secure by design through role-based access control, with more granular approaches available. Optionally, we integrate data quality solutions in your machine learning workflows to ensure that your predictive models perform as expected. Our team can also work with you to conduct a Data Protection Impact Assessment (DPIA) to ensure compliance with UK GDPR.

SECTION 4

Our approach

The approach we take to this service, including the inputs, contributions, outputs and deliverables



BUTTERFLY
DATA

4. Our approach

We pride ourselves on closely collaborating with you and your existing suppliers, acting as an extension of your team. We are multidisciplinary, spanning data engineering, data analytics, data visualisation, data science, business analysis, project management, plus individual specialisms. This lets us approach your challenges from a variety of perspectives, combining our expertise to implement holistic solutions.

We typically kick off a project with an exploration phase. In collaboration with your relevant stakeholders, requirements are defined and prioritised, operational constraints understood and aspirations for future developments and integrations considered. In addition, we investigate relevant data and metadata that may already be available within your organisation, for example logs and usage data.

This results in a project roadmap and delivery plan which is updated throughout the project. We conduct frequent sprint reviews, usually bi-weekly, accompanied by status update presentations and work-in-progress demonstrations to collect feedback often and early on. This allows us to quickly iterate, learn from insights uncovered along the way and adapt to changing priorities while delivering within agreed timeframes and budgets. Throughout, a dedicated project manager serves as your central point of contact for clear communication and swift escalation, if required.

The code base is managed via Git repositories, optionally with connection to project management and knowledge management tools, for example work tickets tracked via Jira and synchronised with Bitbucket and Confluence. We are tool-agnostic and happy to work with your preferred collaborative coding platform and project management tools. All code is thoroughly commented, and we provide documentation that illustrates the overall solution design.

We offer long-term support but are equally happy to hand over the system to your internal teams upon project completion. To ensure a smooth transition, we typically conduct a series of handover and training sessions with your core users to give them time to get to grips with the solution and resolve any questions they may have. This process is supported by the frequent sprint reviews and technical documentation.

4.1. Assumptions

- Access to the relevant decision-makers, product owners and end users, for example for discovery workshops, feedback during development, signing off deliverables and handover or training.
- Access to relevant data stores and/or APIs, where applicable via dedicated accounts for each of our team. We always aim to minimise access to a necessary minimum and work with your team to mitigate risks, for example by masking sensitive data where appropriate.

- Access to adequate cloud computing resources on AWS, Azure or GCP where required by the scale of data and chosen ML techniques. We either use your existing cloud instance or set up a temporary solution with your preferred cloud provider for the duration of the project (until handover to you or migration of the final solution to your cloud instance). Where possible, experimentation will be carried out on local compute.

4.2. Outputs and deliverables

During the project

- Project roadmap and delivery plan.
- Frequent sprint reviews and accompanying update presentations and demonstrations.
- Any documents relating to legal requirements that are required at the start of the project, for example a DPIA if sensitive data is processed (DPIA drafted by Butterfly Data but owned and signed off by the client as the data controller).

Upon project completion

- Any machine learning features that were developed and models trained using your organisation's data. Where the chosen cloud platform supports this, storage in managed services like Amazon SageMaker Feature Store, Azure Machine Learning Managed Feature Store or Feast and Vertex AI on Google Cloud Platform.
- Any machine learning models and MLOps pipelines developed.
- For larger projects, written reports that document crucial findings and insights uncovered during the development phase and in user research, often including recommendations for future developments that go beyond the scope of the active project.
- Git repository containing all relevant code on your preferred collaborative coding platform.
- Archive of all presentations, written reports and status updates from the project delivery phase.
- Training sessions to handover the project to your internal team.
- Technical documentation detailing the ML model architecture, MLOps pipeline, code base and underlying rationale.

SECTION 5

Standardised service framework

Our framework that ensures consistent, high-quality delivery through defined roles, Agile processes, standards, training and ongoing support

5. Standardised service framework

All our services are built on a common service framework that drives consistency, quality, and continuous improvement. This framework covers:

- Project roles and responsibilities for effective service delivery.
- Discovery phase to define deliverables, architectural requirements, risks, and opportunities.
- Delivery approach, using Agile methods across defined phases and supported by modern collaboration tools.
- Standards that shape our capabilities, culture, and service excellence.
- Training and user support to ensure that your teams maximise the value of data services and tools.
- Support and maintenance for dependable operation of live services.

5.1. Project roles and responsibilities

Butterfly Data can provide a fully independent, self-managing delivery team that works closely with your stakeholders to understand requirements and deliver high-quality outcomes. Our teams are tailored to your needs and budget, typically combining the following roles:

Accountable Executive (AE)

Responsible for the overall success of the engagement, ensuring the solution meets requirements and providing senior support when needed. Key responsibilities:

- Building strong, collaborative stakeholder relationships.
- Providing vision, governance, quality assurance and service standards.
- Leading risk, issue, and assumption management.
- Handling escalations and resolving concerns.
- Overseeing contracts, financials, timesheets and invoicing.

Delivery Lead (DL)

The primary interface between the Butterfly Data team and your project stakeholders, ensuring a smooth and effective delivery. Responsibilities include:

- Ensuring Agile working practices and team effectiveness.
- Aligning resources, removing blockers, and maintaining focus.
- Planning, tracking, and reporting delivery progress.

- Coordinating interdependent teams.
- Clarifying requirements and maintaining organisational knowledge.

Technical Lead (TL)

The bridge between the delivery team and your technical stakeholders, combining skills in analysis, architecture design, and project management. Responsibilities include:

- Ensuring deliverables fully meet technical requirements.
- Managing and resolving technical challenges.
- Designing and evolving robust architectures and prototypes.
- Balancing performance, security, scalability and other system qualities to manage trade-offs effectively.
- Providing developers with essential context and domain understanding.

Design Lead (UL)

The link between the team and the end-users, ensuring user needs are fully represented in all outputs. Responsibilities include:

- Understanding and documenting user capabilities, needs, and feedback.
- Capturing SME knowledge.
- Running workshops and demonstrating work-in-progress features.
- Designing accessible, intuitive UI/UX.
- Promoting and enhancing data literacy.

Data Analysts (DA)

Analysts connect business understanding to technical delivery. Responsibilities include:

- Interviewing key stakeholders and users to understand your business, project requirements and wider operational considerations.
- Collect, clean, analyse and visualise your data.
- Exploring data sources to assess quality and value.
- Present analyses and the resulting options and recommended actions based on thorough analysis.
- Documenting insights through reports and visualisations.
- Collaborating with your technical teams to improve data systems and processes (e.g., data collection).

Business Analyst (BA)

Business analysts identify business needs and recommend solutions to improve processes, systems, and profitability:

- Documenting and translating your business needs into functional specifications, backlogs and roadmaps.
- Ensuring that solutions align with the defined business requirements.
- Improving business processes through analysis.

Data Engineers (DE)

Engineers design and build the infrastructure required for analytics and data science. Responsibilities include:

- Setting up, configuring or integrating cloud platforms.
- Creating data pipelines to integrate and transform diverse data sources.
- Developing and optimising data structures, databases, algorithms, and models.
- Ensuring performance, reliability, and scalability.
- Managing testing, deployment, and release processes.
- Developing custom functionality to extend your infrastructure's capabilities.
- Collaborating with Data Analysts and Data Scientists to deploy dashboards or machine learning models.

Data Scientists (DS)

Data scientists develop statistical and machine learning models to generate insights and automate tasks. Responsibilities include:

- Working with your subject matter experts to develop hypotheses, experiments and research questions that address your business challenges.
- Analysing complex datasets, using statistical models to quantify and explain trends and relationships in your data.
- Building and validating predictive models.
- Training machine learning models to automate tasks, such as classification.
- Working with cross-functional teams to translate insights into strategy and to operationalise experiments, e.g., by deploying machine learning models to live services.

5.2. Discovery phase

Most projects begin with a discovery phase, designed to fully understand your challenges, risks, and opportunities before refining the delivery plan. Butterfly Data's planning approach follows the [Information Technology Infrastructure Library \(ITIL\) framework](#) with a structured Discovery Phase using a comprehensive assessment of requirements by:

- Establishing clear business objectives, desired outcomes and success criteria, supported by an initial business case.
- Understanding your current capabilities, constraints, risks and dependencies across people, processes and technology.
- Identifying your stakeholder needs and expectations through structured engagement and user research.
- Assessing challenges and opportunities to determine high-level requirements and value drivers.
- Exploring and evaluating potential solution options to determine feasibility, suitability and alignment with your organisation's goals.
- Defining recommended approaches and priorities to inform subsequent service design and transition activities.

We are experienced in Agile delivery, ensuring value, control, and flexibility without compromising on our [service standards](#).

The discovery phase also establishes the detailed specifications for deliverables, including:

- **Deliverable types** (e.g., datasets, models, reports), each with unique risks, assumptions, and dependencies.
- **Acceptance criteria** and performance metrics to be met.
- **Architecture characteristics** (non-functional requirements) that balance performance, availability, security, portability and other key qualities.

5.3. Setup, migration, and delivery

Butterfly Data is an established delivery partner to major UK Government departments including HMRC, DWP, FCDO and the Home Office, supporting the secure management and operation of highly sensitive data and critical services. We have extensive experience working within regulated public sector environments where security, resilience and compliance are essential.

We provide specialist cloud services across AWS, Microsoft Azure, Google Cloud Platform and SAS alongside open-source and proprietary technologies. Our teams design, migrate and operate cloud environments in line with recognised best

practice including the AWS, Azure and Google Cloud Architecture Frameworks and ensure compliance with GDPR and the UK Government Cloud Security Principles.

Butterfly Data supports organisations at all stages of cloud adoption including migration from on-premise infrastructure, movement between cloud providers and the modernisation of legacy systems. Our approach identifies and mitigates technical, security and operational risks associated with decoupling legacy platforms while maintaining service continuity and data integrity.

We provide end-to-end support covering cloud architecture design, infrastructure selection and environment configuration. This enables the delivery of secure, scalable and resilient cloud platforms that meet operational requirements and support future change. Our services prioritise automation, strong data governance and cost-effective use of cloud resources to deliver reliable, sustainable and compliant cloud services for the public sector.

5.4. Knowledge transfer and training

We ensure effective knowledge transfer throughout every engagement, helping client teams build capability, adopt industry best practices and independently run the new platforms and systems after handover.

Training can be tailored to specific needs, especially where specialist tools or methods are involved, and delivered through courses, videos, presentations, one-to-one sessions, or workshops online, on-site or at suitable venues.

Most engagements include some form of coaching or training, such as:

- **Project team training:** Our consultants share expertise through peer support and on-the-job mentoring to embed best practices and drive successful outcomes.
- **Stakeholder training:** We provide specialist insights to help stakeholders understand our services, align terminology and avoid miscommunication.
- **End-User training:** We design and deliver bespoke training and documentation that supports adoption of new analytical tools and platforms (e.g., SAS Viya), identifying issues early and defining configuration or onboarding steps to ensure a smooth user experience.

5.5. Ongoing support and maintenance

Beyond development and implementation, we provide ongoing support and maintenance for both new and existing services.

Butterfly Data deploys security-cleared consultants and developers across multiple Public Sector projects and can recruit or train specialists to meet flexible data management needs, including:

- SAS and Linux administrators
- SAS, R and Python developers
- Data engineers, data analysts and data scientists
- Agile project managers
- Cloud solution architects
- Business and test analysts

We support a wide range of data management services and project activities:

- **Consultancy:** Feasibility studies, business process modelling, strategy, and execution.
- **Development:** Requirements specification, data modelling, interface design, functional specification, bespoke tools, prototyping, code review.
- **Provisioning:** Cloud architecture design, IaaS, migration strategies, tool selection, data and model migration.
- **Training:** Strategy, learning pathways and delivery.

5.6. Service management

We aim for complete customer satisfaction. As part of our standard service, key stakeholders receive dedicated support from the accountable executive and project lead.

Standard support is provided remotely via email during UK working hours (Mon–Fri, 09:00–17:00). Support outside these hours for critical deadlines, system maintenance, or other project activities can be arranged remotely or on-site at additional cost.

Emails received during standard hours will be acknowledged and triaged (or resolved) within 4 working hours. Issues requiring more time will be assigned to the Project Lead, who will classify and prioritise them:

- **Critical:** Affects service integrity or availability; highest priority, resolved as quickly as possible.
- **Major:** Mildly affects all users or severely affects a few; resolved after critical issues.
- **Minor:** Affects few users; resolved after major issues.
- **Information requests:** No user/service impact; addressed as priorities allow.

Our triage system ensures critical issues receive immediate attention, resources are allocated efficiently, and queries are answered promptly.

5.7. Service standards

Butterfly Data operates according to a structured framework that ensures consistent service quality:

- Incremental development through Agile delivery with regular client feedback.
- Adherence to [DAMA-DMBOK](#) best practices, B-Corp ethical standards, and UK Government Functional Standards including [Technology Code of Practice \(TCOP\)](#), [Government Service Standard](#), and [Data Ethics Framework](#).
- Consultants are authentic, collaborative, innovative, and proactive, consistently demonstrating autonomy, influence, expertise, and commerciality.
- Ensuring our team have appropriate skills and knowledge to deliver high quality work, by aligning to recognised frameworks including Skills Framework for the Information Age ([SFIA](#)) and the Government Digital and Data Profession Capability Framework ([DDaT](#)).
- Built-in compliance with GDPR, Secure Record Handling, Official Secrets Act, Computer Misuse Act and other regulatory requirements. All staff undergo annual GDPR training.

SECTION 6

Relevant case studies

A showcase of how we have effectively implemented this service, or similar solutions, in real-world projects, delivering measurable results



BUTTERFLY
DATA

6. Case studies

Synthetic face data for policing using Generative AI: Butterfly Data worked with a UK police authority to address ethical and privacy challenges associated with testing facial recognition (FR) systems by developing a synthetic face generation tool using generative AI. The solution leveraged an open-source model with a custom Python prompt engine to produce large, diverse datasets of realistic synthetic faces with controlled demographic attributes (such as age, ethnicity, gender), varied **lighting**, and expressions. This enabled repeatable and controlled evaluation scenarios without the use of real-world images, mitigating privacy risk while supporting rigorous system assessment.

The synthetic data platform empowered the client to comprehensively evaluate FR system performance and detect potential algorithmic biases across a wide range of demographic variables, informing adjustments that enhanced system fairness and accuracy. By facilitating ethical testing in a sensitive domain, the project demonstrated both practical utility and a foundation for future responsible AI deployment in policing and other public sector contexts.

Machine learning for data classification with defence industry: Butterfly Data supported a large defence organisation to explore how machine learning can enhance data classification and analysis for complex operational datasets by delivering a rapid proof-of-concept (POC) model. Working within a tight four-week timeframe, our team evaluated multiple modelling approaches including engineered feature models, pre-trained model **adaptation**, and tailored transformation-based models, with rigorous pre-processing and testing in a secure environment to meet demanding accuracy and performance criteria.

The resulting POC demonstrated promising classification performance even on limited training data, showing that robust machine learning workflows can be developed rapidly under constraint and providing valuable insights to inform the client's broader machine learning strategy. This engagement not only delivered technical outcomes but also enhanced the organisation's data-driven evaluation processes and laid a foundation for potential future operational deployment of advanced classification capabilities.

Automating reports for using an AI Co-Pilot: Butterfly Data partnered with the Gloucestershire & Oxfordshire Community Rail Partnership (GOCR) to streamline and automate their reporting processes, which were previously manual and resource-intensive. The engagement involved co-designing and implementing an AI-driven "Co-pilot" solution that accepts raw stakeholder feedback data, applies reusable templates and **prompts**, and produces branded reports and visualisations (e.g., graphs, word clouds, slide decks) within a Microsoft environment. This

automation significantly reduced the time required to generate formal reports, addressing inefficiencies in data entry, **formatting**, and compilation.

The solution delivered substantial efficiency gains, reducing report production from several hours of manual effort to under 30 minutes, and **freeing** up GOCRPs staff to focus on higher-value activities such as analysis, community **engagement**, and project delivery. The system also improved consistency and quality of outputs, helping the organisation better evidence its impact, respond to funders and **partners**, and invest more of its limited capacity into direct community work.

[View more case studies](#)



APPENDICES

Appendix

References including our full G-Cloud list, social value response, and other frameworks we are present on

Appendix 1: Other approved frameworks

Butterfly Data also operates through the following frameworks (correct as of December 2025):

Digital Outcomes and Specialists 7 (DOS7) (CCS ID: RM1043.9)
Spark DPS (CCS ID: RM60-49)
Digital & Legacy Application Services (DALAS) <i>As part of Netcompany's subcontracting ecosystem</i> (CCS ID: RM6335)
Artificial Intelligence (AI) (CCS ID: RM6200)
Technology Services 4 (TS4) (CCS ID: RM6190)
Constellia Neutral Vendor Framework for Innovation (NVfI)
Naval Design Partnering (NDP) <i>As part of Frazer Nash's subcontracting ecosystem</i>
Home Office Accelerated Capability Environment (ACE)

Appendix 2: Full list of G-Cloud 15 services

This is the full list of G-Cloud services provided by Butterfly Projects Ltd t/a Butterfly Data:

End-to-End Data Management Solutions
Cloud Strategy, Migration & Optimisation
Legacy System Migration
Cloud Migration - Enterprise Workload Migration to Amazon Web Services (AWS)
Cloud Migration - Enterprise Workload Migration to Microsoft Azure
Cloud Migration - Enterprise Workload Migration to Google Cloud Platform (GCP)
Data Platform Modernisation
Snowflake Data Platform Set-Up, Optimisation & Support
SAS Cloud Deployment
SAS Legacy Services
Data Transformation, Enrichment and Modernisation
Multi-Cloud Data Platform Engineering and Architecture Design
Informatica Cloud Integration Services
Data Governance and Compliance Frameworks
Data Integration and Interoperability Services
Data Capability and Maturity Assessment
Data Protection and Privacy Services
Synthetic Data Services
Data Quality Management and Assurance
SAS Viya Data Preparation and Data Quality
Data Science - Advanced Analytics and Artificial Intelligence (AI) Services
Data Analytics, Insights and Visualisation

Appendix 3: Social value response

Tackling economic inequality	Butterfly Data was founded on the principle that data should be used for public good and that economic opportunity should be shared fairly. Established in 2003 and transitioning to an Employee Ownership Trust (EOT) in 2021, every current and future Butterfly Data employee has a direct stake in the success of the organisation. This model ensures long-term, inclusive growth and fair wealth distribution, directly addressing economic inequality. In 2023, we became one of fewer than 1,500 certified UK B Corporations, legally committing us to operate for the benefit of people, communities and the environment. We are a growing organisation with over 30 staff across the UK and offices in Cardiff and Manchester, enabling us to recruit nationally without geographic barriers and to create high-value digital jobs across the UK. Our workforce reflects our commitment to diversity and social mobility. Many of our colleagues are in second careers, having transitioned from sectors such as teaching, marketing, translation and the armed forces, helping to address digital skills shortages while widening access to well-paid technology roles. We actively support pathways into employment through collaborations with universities, offering synthetic projects for students, paid summer internships, and graduate opportunities that frequently lead to permanent roles. We also provide placements for veterans that have resulted in long-term employment, knowledge transfer and upskilling. Our Reservist Policy includes an additional 10 days of paid leave for training and we have been recognised with the Gold Armed Forces Employer Recognition Scheme Award for our sustained support to Defence personnel. Our benefits and reward structures are deliberately designed to remove barriers for priority groups. These include flexible and part-time working for all roles from day one, remote or hybrid working as standard, enhanced maternity / paternity / adoption pay, private healthcare for employees and their immediate families, and funded training with pay progression linked to skills acquisition. Consistent remuneration across all roles reinforces fairness and transparency, supporting retention and long-term career progression. We have also delivered pro bono and low-cost data support to third-sector organisations such as the Gloucestershire and Oxfordshire Community Rail Partnership (GOCR), where we developed AI-enabled reporting solutions to translate community engagement data into actionable insights for local authorities and transport planners. Our work with The Nelson Trust, a women's charity supporting vulnerable groups, strengthened their data capability through robust data models, Power BI dashboards, and data quality frameworks, enabling them to evidence impact, secure funding and improve service delivery. Beyond delivering projects, our staff actively engage in outreach and STEM initiatives. We support schools and charities through
-------------------------------------	--

	sponsorship and volunteering, such as Race to the Line STEM challenges, Girl Tech Manchester innovation sprints, and enterprise programmes like The Bumbles of Honeywood, delivered in schools with high levels of free school meal eligibility or additional learning needs. Through our Pledge 1% commitment, we also support initiatives e.g., Magic Breakfast, helping to tackle child hunger and improve educational attainment. These activities promote aspiration, social mobility and early exposure to digital and data careers for underrepresented groups.
Fighting climate change	Butterfly Data is a sustainable organisation with a long-standing commitment to environmental stewardship. We introduced a Carbon Reduction Plan in 2020 and achieved B-Corp certification in 2023, legally committing us to operate for the benefit of all stakeholders. As a predominantly virtual company, our scope 1 and scope 2 emissions are zero, and our limited scope 3 emissions are measured annually and fully offset. By supporting staff to work from home as standard, we significantly reduce carbon emissions associated with commuting. Where travel is required, rail is the preferred option, with car sharing encouraged when trains are not suitable. We are members of the Octopus EV scheme to support staff in transitioning to electric vehicles, and all company cars are electric. We also actively encourage colleagues to adopt green energy tariffs at home, supporting lower-carbon lifestyles beyond the workplace. Our services directly contribute to environmental sustainability in the public sector. Through our data maturity, integration, and transformation services, we promote sustainable data practices, including efficient data architectures, optimised code, and reduced data storage and compute requirements. These approaches reduce energy consumption associated with cloud infrastructure and support public sector organisations in minimising the environmental impact of data and analytics. Our Carbon Reduction Plan is publicly available on our website.
Equal opportunity	Butterfly Data's policy promotes respect, inclusivity, professionalism, integrity and transparency. We uphold ethical business practices, with zero tolerance for modern slavery, human trafficking or exploitation across our operations and supply chain. We are proud that our gender pay ratios are closely aligned for all consultant grades, and for roles with more than five employees, we aim to maintain pay within $\pm 5\%$ across genders, reflecting our commitment to fair and equitable treatment. All employees are expected to uphold our values and report concerns, and we work collaboratively with suppliers to promote responsible and ethical practices. Our services support equal opportunity by improving how public sector organisations manage and use data. Through our Data Maturity Assessment and Data Transformation services, we help standardise data accessibility and governance, reducing disparities in access to information and services. This creates a more level

	playing field for diverse communities to engage with public services, improving inclusion, innovation, and economic participation. Internally, as a Disability Confident Employer Level 2, we embed equality through transparent pay structures, consistent remuneration across roles, inclusive recruitment practices and training programmes designed to upskill staff from diverse backgrounds. By reducing structural barriers to progression and ensuring access to learning opportunities, we promote a fair and inclusive working environment that reflects the communities our public sector clients serve.
Wellbeing	Employee wellbeing is central to Butterfly Data's culture and long-term sustainability. We have three qualified Mental Health First Aiders (two male and one female), providing additional, confidential support alongside our structured 1-to-1 mentoring programme. This encourages open conversations about wellbeing and ensures staff have multiple routes to access support. We actively promote work-life balance and connection through regular in-person team events across the UK. These range from social and family-friendly gatherings to outdoor activities such as walking, canoeing, and shared challenges, helping to build trust, resilience and strong relationships beyond day-to-day project work. This sense of belonging supports mental wellbeing, collaboration and retention. All employees receive comprehensive private healthcare, including enhanced mental health support and wellbeing programmes. This ensures timely access to care and empowers individuals to take an active role in managing their health. Beyond our internal practices, our Data Integration services support improved management of health and wellbeing data for public sector organisations, enabling better monitoring, earlier interventions and more effective delivery of health and social care services that benefit communities at scale.

Appendix 4: Common acronyms, abbreviations and keywords

Term/Abbreviation	Explanation
Agile delivery	A project management methodology emphasising reliable project, risk, and change management through a thorough, flexible, and outcome-oriented delivery model with frequent checkpoints.
AI (Artificial Intelligence)	The application of statistical modelling, deep learning, and other methods to answer complex questions or to automate tasks like classification, forecasting, and anomaly detection. Often used synonymously with the term “Machine Learning”.
AKS (Azure Kubernetes Service)	A managed service used in Azure migration for refactoring and modernisation of legacy applications to improve scalability and reduce technical debt.
API-led connectivity	Standardised, reusable API interfaces that enable rapid integration of new applications and services while maintaining security and governance standards.
Apache Arrow	An open-source library providing columnar data structures for in-memory computing for high performance applications that process and transport large data sets. It is designed to both improve the performance of analytical algorithms and the efficiency of moving data from one system or programming language to another.
Apache Kafka	An open-source event streaming platform used for real-time, high-performance data synchronisation and processing, ensuring timely information flow and real-time insights.
Apache Superset	An open-source tool used for data analytics and visualisation, supporting insights gathering via a no-code interface to build charts and web-based SQL editor for advanced querying. .
AWS (Amazon Web Services)	A major public cloud platform supported for multi-cloud, migration, and data platform engineering, with services aligned to frameworks like the AWS Well-Architected Framework.
Azure	A major public cloud platform supported for multi-cloud, migration, and data platform engineering, including native services like Azure Migrate and Azure Policy.
CDMP® (Certified Data Management Professionals)	A globally recognised certification program by DAMA International. Professionals with this certification deliver services related to data security, privacy, quality, auditing, and compliance.
CI/CD (Continuous Integration/Continuous Deployment)	A set of DevOps practices automating the deployment of software changes and data pipelines used for automated builds, tests and deployment, ensuring reliable platforms and enhancing reduced risk.

Cloud-agnostic	Our cloud-agnostic services let you select your preferred platform from a range of major cloud platforms, including AWS, Azure, and Google Cloud Platform. Can also apply to tools, for example, Snowflake can be run on AWS, Azure or GCP.
DAMA (Data Management Association)	A recognised framework, along with its DMBok (Data Management Body of Knowledge), used to align services, ensuring industry-standard management of information assets and supporting data quality principles.
Data Clean Rooms	A Snowflake feature that enables privacy-preserving data collaboration and supports secure data practices.
Data lineage	A comprehensive, end-to-end view of data flows, transformations and points of consumption across an organisation, which facilitates data auditability, and simplified compliance.
Data masking/redaction	Strategies that keep data functional while enabling compliant and controlled sharing and collaboration by obfuscating Personally Identifiable Information (PII).
Data Vaults	A proven framework for schema modelling and redesigned data structures, used to achieve optimised query performance and reduced costs.
Denodo	A technology used for data virtualisation, which provides a unified view of organisational data without moving or copying it, prioritising speed and cost-efficiency.
DevOps	DevOps is a set of practices, tools, and cultural philosophies that automates and integrates the processes between software development (Dev) and IT operations (Ops).
DPA (Data Protection Act)	UK legislation that forms part of the regulatory data protection and privacy compliance focus alongside GDPR.
DPIA (Data Protection Impact Assessment)	A risk assessment conducted to support compliance with UK GDPR, particularly prior to data migration or processing sensitive data.
ELT (Extract, Load, Transform)	A data integration approach where data is loaded into the target system before transformation, typically using cloud-native pipelines.
ETL (Extract, Transform, Load)	A data integration approach where data is transformed before being loaded into the target system.
Future-proof solutions	Solutions designed to be maintainable and effective long-term by respecting and supporting an organisation's business processes, customer-facing services, and wider technology landscape.
GCP (Google Cloud Platform)	A major public cloud platform supported for multi-cloud, migration, and data platform engineering.
GDPR (General Data Protection Regulation)	A key regulation for data protection and privacy compliance, which is a core focus of governance and privacy services.
Generative AI	The integration of Large Language Models (LLMs) and diffusion models to programmatically create synthetic text, image, audio, and video, or to aid in the analysis of unstructured data.

IDMC (Intelligent Data Management Cloud)	Informatica's cloud platform used to deliver secure, scalable, and governed data integration services.
IaaS (Infrastructure as a Service)	A cloud computing service model (like compute and storage) that allows for state-of-the-art, scalable, and cost-effective cloud platforms.
LLMs (Large Language Models)	Cutting-edge models deployed to programmatically create synthetic data and content or integrated into services for natural language interaction.
MDM (Master Data Management)	A structured approach involving the creation of a centralised "golden record" for critical data entities (e.g., citizens, patients, assets).
ML (Machine Learning)	Statistical modelling and deep learning approaches used to automate tasks such as classification, forecasting, and anomaly detection.
MLOps	The use of established DevOps principles, like CI/DC, to automate and streamline the full machine learning lifecycle, from data preparation to training, deployment, monitoring, and maintenance.
OSINT (Open Source Intelligence)	Trusted third-party data sources used for data enrichment to maximise the value of data and unlock new insights during analytics.
PaaS (Platform as a Service)	A cloud computing service model (like a database or web server) that allows for state-of-the-art, scalable, and cost-effective cloud platforms.
Parquet	A modern columnar file format for data serialisation by decompressing and decoding its contents into memory. It is designed to be space/IO-efficient at the expense of CPU utilization for decoding. It does not provide any data structures for in-memory computing.
PII (Personally Identifiable Information)	Sensitive data (e.g., names, addresses, medical information) that should be protected through privacy-focused services like data masking, redaction, and synthetic data generation and is subject to data privacy and protection regulations like GDPR and the Data Protection Act.
Power BI	An enterprise platform tool for data analytics and visualisation. Often included in existing Office 365 subscriptions and integrates well with Microsoft services like Azure.
QKB	Abbreviation for Quality Knowledge Base, which is part of the SAS Data Quality tooling set (DataFlux, Loqate) used for profiling, cleansing, and validation.
RBAC (Role-Based Access Control)	Secure data practices using customisable access policies based on the "least privilege" principle, enhancing security.
SAR (Subject Access Requests)	An operational aspect of GDPR compliance that the data protection services support through features like data lineage tracking.

SAS	A suite of analytics, data management, and business intelligence software. SAS 9.4 and Viya 3.x are considered legacy, with customers transitioning to the SAS Viya (4) platform.
Snowflake	A cloud-based data platform that unifies data warehousing, data lakes, data engineering, and data sharing into a single, managed service, allowing organisations to store and analyse vast amounts of data without managing infrastructure.
SSO (Single-Sign-On)	An easy-to-use authentication method that simplifies access control and enhances secure data practices through a single user account that can be used across multiple tools and services.
Star Schema	A proven framework for schema modelling and redesigned data structures, used to achieve optimised query performance and reduced costs.
Synthetic Data	Data created via rule-based simulations or Generative AI to mimic real-world data without exposing PII, enabling safe innovation, collaboration, and GDPR compliance.
Tableau	An enterprise platform tool for data analytics and visualisation.
TCO (Total Cost of Ownership)	A measure of the total long-term cost, which is targeted for reduction through platform optimisation and the use of serverless compute technologies.
UX (User Experience)	The process of capturing user requirements to design intuitive, self-service dashboards and visualisations that are meaningful and aligned with business operations.
Viya Platform	The native cloud platform for SAS, which is the target for modernisation and migration from legacy SAS platforms. Often called SAS Viya 4 to differentiate from (legacy) Viya 3.x.



BUTTERFLY
DATA

Ready to transform your data?

Book your free discovery call and find out how our bespoke data services and solutions could help you uncover untapped potential and maximise ROI.

ccs@butterflydata.co.uk

+44 29 212 00140

butterflydata.co.uk

Get in touch

