CPSC 69100 Fall 2024

Datapunk

Requirements Document

Ethan Epperson / [distracted.e421@gmail.com](mailto:distracted.e421@gmail.com)

Date submitted: 10.6.2024

# Introduction

**Goal of the Project**

The goal of Datapunk is to create an open-source, user-centric platform that empowers individuals to reclaim ownership of their personal data. In today’s digital landscape, personal data is fragmented across multiple platforms and shamelessly exploited by corporations for profit. Datapunk offers a solution by giving users the power to **collect, organize, analyze, and visualize** their data in a secure, decentralized environment—putting control back where it belongs: in their hands.

Datapunk will enable users to import data from various services such as Google, Spotify, Microsoft, and Meta via **APIs** and **manual uploads** of data dumps (e.g., Google Takeout). The platform supports multiple formats like **JSON, CSV, and GeoJSON** and employs **Python-based parsers** to clean and standardize the data for seamless use. Once imported, the data will be stored locally or on user-controlled databases (e.g., **CouchDB**, **PostgreSQL/PostGIS**), ensuring that users maintain complete ownership and control of their data.

Datapunk will feature a **customizable interactive dashboard** that allows users to visualize their data through dynamic charts, graphs, and geospatial maps. The dashboard will provide actionable insights into their digital activities, such as behavioral trends, music preferences, location history, and communication patterns, giving users a clear view of how their data tells the story of their life.

The project is built with **security and privacy as core principles**, incorporating **encryption at rest and in transit**, secure OAuth token handling for third-party integrations, and optional **data anonymization** for users wishing to share data externally. By prioritizing user control, privacy, and decentralization, Datapunk aims to dismantle the power dynamics that favor big tech—giving individuals the tools to understand and manage their digital footprint independently, without the surveillance or interference of corporate cloud services.

Datapunk’s ultimate mission is to help users **take back control of their personal data**, allowing them to derive value from it on their own terms, free from the clutches of exploitative big tech giants.

**Why This Project Is Important**

In today’s digital world, personal data is a goldmine, and corporations like Google, Microsoft, and Meta are cashing in. They collect, manipulate, and sell your data, often without clear consent, leaving you with no control, transparency, or say in how it’s used. Datapunk flips the script by empowering users to **import, organize, analyze, and securely store** their own data in a private, decentralized environment.

Forget relying on manipulative interfaces designed to keep you scrolling and spending. Datapunk's **data visualization** tools turn fragmented data from platforms like Google, Spotify, and Microsoft into real, actionable insights. Want to understand how your music habits sync with productivity? Curious about communication patterns across platforms? Datapunk provides a clear, customizable dashboard that respects your need for transparency and insight.

Bottom line: **Datapunk is here to disrupt the status quo**. It takes the power away from corporate giants who exploit your data and puts it back in your hands. Datapunk gives you the autonomy you deserve—an open-source, decentralized platform that prioritizes your privacy, ownership, and best interests, not theirs.

**Beneficiaries of this Project**

The primary beneficiaries of Datapunk are individuals who are tired of being exploited by big tech and want to reclaim their digital freedom. Privacy-conscious individuals will finally have the tools they need to collect, organize, and analyze their own data without relying on services that profit off their information. Tech-savvy users and data enthusiasts will find Datapunk to be a powerful ally—exploring Spotify listening trends, correlating productivity with calendar events, or diving deep into personal digital habits. No more relying on platforms designed to keep you hooked—this is your data, presented how you want, without the clutter or manipulation.

For those concerned about their digital well-being, Datapunk offers insights into how online activities affect mental, emotional, and physical health. By correlating data from different sources—calendar events, music preferences, communication trends—users can gain a better understanding of their behaviors and make positive changes. With features like trend analysis and data visualization, Datapunk gives users the tools to track habits over time and make informed decisions to improve their lifestyle.

Datapunk also caters to neurodivergent and accessibility-focused users who need more control over how they interact with data. Customizable visualizations, adjustable information density, and support for sensory preferences make Datapunk adaptable to a wide range of needs. It’s all about giving users the ability to process and explore their data in ways that work best for them, without being overwhelmed.

Professionals and businesses looking to escape the corporate cloud trap will also benefit. Datapunk provides a robust, open-source alternative for managing data independently, with secure storage and granular privacy controls. This means sensitive data stays local, and workflows can be customized without ever having to share information with big tech. Datapunk helps these users achieve true data sovereignty—keeping control firmly in their hands.

Lastly, Datapunk is a gift to developers and open-source enthusiasts who believe in privacy, empowerment, and the freedom to innovate. Its flexible design invites contributions, allowing developers to add custom data parsers, integrations, and visualizations. The opportunity to be part of a project that fights against data exploitation and builds tools for real user empowerment is what makes Datapunk a community-driven force for good.

**Who Datapunk Is Not For**

If you think big tech has your best interests at heart and that corporate algorithms are always looking out for you—then Datapunk might not be for you. If you love being bombarded with ads that know you better than you know yourself, or if you enjoy having your personal data traded like a commodity while you sit back with a shrug, then hey, you do you. Datapunk isn’t here to convince you otherwise—it’s here for those who are ready to take a stand, not just accept the status quo. If you're not into owning your data, making informed decisions, or taking control of your digital life, then there are plenty of other services happy to keep you in the loop—just maybe not this one.

## **Major Features of Datapunk**

**1. Data Import and Integration**

Datapunk’s data import system is a powerful tool designed to give users back control of their personal data—taking it from the clutches of big tech companies and putting it securely in their hands. The system offers both manual uploads and direct API integrations, making it easier than ever to collect personal data from sources like Google, Spotify, Microsoft, and Meta. Whether users want to analyze their Google Calendar events, search history, social media activity, or even music listening habits, Datapunk consolidates all this scattered data into a single platform, creating a clear and unified view of their digital lives. No more jumping between platforms or giving up ownership—Datapunk puts users in full control.

For those who prefer a hands-on approach, Datapunk supports manual uploads from services like Google Takeout. Users can gather comprehensive datasets in formats such as JSON, CSV, and GeoJSON, covering everything from location history and contacts to email archives and browsing data. This flexibility allows users to pull together extensive amounts of data across a wide range of services, bringing all their digital activities under one roof. Datapunk’s data cleaning, parsing, and organizing tools automatically handle the heavy lifting, preparing these datasets for analysis without requiring technical skills. Whether it's managing years of search history or diving deep into location-based insights, users can quickly clean up, organize, and explore their data, ensuring it’s ready for analysis right away.

For users who prefer a more automated experience, Datapunk integrates directly with APIs from platforms like Google, Spotify, and Meta. This eliminates the need for manual file handling, as Datapunk syncs with these services to pull the most up-to-date information, whether it’s Spotify playlists, Facebook interactions, or Google location history. Users can connect their accounts using OAuth 2.0, ensuring secure access without sharing sensitive login credentials. With automatic sync options, users can keep their data fresh with scheduled imports, or trigger updates manually whenever they need the latest insights. This flexibility offers complete control over when and how often data is updated, ensuring the platform works at the user's pace.

Datapunk is also designed to handle multiple file formats—whether it’s structured data from a CSV file, geospatial data from GeoJSON, or complex JSON exports from Google services. This adaptability ensures that users can bring together various data types and sources without worrying about compatibility issues. For those needing real-time synchronization, Datapunk supports continuous data collection from certain platforms, making sure users always have the most current insights.

To make the data import process smoother, Datapunk provides clear progress indicators during every step of the import. Users will know exactly where they are in the process, with real-time updates showing what’s happening in the background. Should anything go wrong—whether it’s a missing field in a CSV or an invalid JSON file—Datapunk offers detailed feedback and error messages to help users resolve the issue. This makes importing data less of a headache and more of a seamless experience, so users can get to the analysis part without unnecessary frustration.

**2. Customizable Dashboard and Interactive Visualization**

Datapunk’s Customizable Dashboard is the nerve center where users get to take control of their data like never before. It’s designed to transform your personal data into something you can actually understand and use—through dynamic, interactive visualizations that put you at the center of the experience. The dashboard is all about giving users the flexibility to tailor everything to their preferences. Whether you want to track activities, monitor spending, analyze productivity, or dive into your Spotify listening habits, Datapunk’s widgets make it simple to bring the most relevant insights into focus, with minimal fuss.

Customization is what makes Datapunk shine. The **drag-and-drop functionality** lets you effortlessly arrange widgets exactly how you want them. Want your productivity tracker right up front and your Spotify trends off to the side? Done. Datapunk’s **widget library** is stacked with options—covering everything from health metrics to communication data—so users can build a dashboard that fits their lifestyle, whether they’re diving deep into finances or just trying to get a handle on daily routines.

When it comes to **data visualization**, Datapunk goes beyond basic charts. Users can choose between **line graphs, bar charts, and even geospatial maps**, depending on what kind of story they want their data to tell. Tracking productivity over time? A line chart is perfect. Want to visualize where you’ve been over the past year? A **heatmap** has got you covered. Switching between different visualization types is easy, letting you explore your data from different angles and uncover insights you might have missed otherwise.

And it’s not just about what’s on the surface—Datapunk’s **drill-down capability** means every widget is an entry point to deeper analysis. Say your spending widget shows you’ve blown your budget this month. Just click through for a detailed breakdown—categories, specific transactions, or even how your spending changed week by week. The same goes for any other widget: you can adjust timeframes, dig into specific metrics, and really get a grip on what your data is trying to tell you.

The customization doesn’t stop there. Datapunk knows that how information is presented matters just as much as what’s being shown. You can personalize the **color schemes**, pick **font sizes** that are comfortable for you, and even choose a **low information density mode** to keep things simple if you’re not in the mood for too much detail. This flexibility makes Datapunk accessible, whether you’re a power user who loves data overload or someone who prefers things clean and minimal. Save your favorite layouts—maybe a work-focused dashboard during the week and a wellness-focused one for the weekend—and switch between them whenever you like.

To make things even smoother, Datapunk includes **snapshot and bookmarking features**. If you find a trend in your Spotify listening that you want to keep tabs on or notice an unusual spike in your expenses, you can save a snapshot or bookmark that specific insight for later. No need to wade through pages of data—Datapunk keeps everything you care about right at your fingertips.

Lastly, the dashboard doesn’t just look good—it works for everyone. Accessibility is baked in with **color-blind-friendly palettes**, **high-contrast modes**, and even **audio narration** options for users who prefer to hear key insights rather than see them. Datapunk makes sure that everyone, regardless of their needs, can get value from their data in a way that works best for them. It’s your data, and Datapunk is here to make sure you’re in control.

**3. Data Cleaning, Parsing, and Enrichment**

Datapunk's core functionality is designed to transform raw, disorganized data from various platforms into clean, structured, and enriched datasets that are ready for deep analysis. Using a combination of pre-built and custom parsers, automated data cleaning tools, and enrichment methods, Datapunk handles the complex work of preparing data, freeing users from the need for manual intervention.

This feature is crucial for users who wish to derive meaningful insights from their data without possessing the technical skills necessary to clean or structure the data themselves. Datapunk simplifies the often tedious process of converting raw data—such as location histories, calendar events, and social interactions—into actionable information, empowering users to focus directly on analysis.

**Automated Data Parsing**

Automated Data Parsing is one of Datapunk's foundational strengths. With pre-built parsers for common data formats from platforms like Google, Spotify, and Meta, Datapunk is equipped to automatically ingest and extract key information from these datasets, supporting formats like JSON, CSV, and XML.

**Google**: Automatically parses Google Calendar events, search history, or location data exported from Google Takeout. For example, the Google Calendar parser can extract event titles, participants, and time zones, organizing this data for further insights, such as productivity analysis.

**Spotify**: Parses users' listening history, extracting details such as track names, artists, genres, and timestamps. This enables users to analyze listening patterns, preferences, and shifts in mood based on music genres.

**Meta (Facebook & Instagram)**: Extracts user interactions, posts, and media content from Facebook and Instagram exports. For instance, a Facebook parser can process comments, likes, and reactions, providing a comprehensive overview of social engagement and activity.

Once parsed, Datapunk ensures data is consistently formatted across platforms, enabling seamless integration into a user’s personal database. By transforming raw data into a standardized structure, Datapunk allows users to focus on insights rather than dealing with formatting challenges.

**Example Workflow**: When a user imports Google Takeout data, the pre-built parser immediately extracts relevant data points, such as location history, search activity, and communication logs. The tool normalizes these formats (e.g., converting timestamps to a common format) and makes the data instantly available for analysis.

**Data Enrichment Capabilities**

Beyond parsing, Datapunk excels at Data Enrichment, a process that adds critical context to raw data by appending additional metadata. This enrichment significantly enhances the value of datasets, making them more suitable for comprehensive analyses.

**Geographic Enrichment**: For instance, when a user imports location data from Google Takeout, Datapunk can augment the dataset with geographic metadata like city names, notable landmarks, and even weather conditions at the time of the visit. This enriched dataset helps users track travel patterns, analyze trips, or evaluate how weather conditions may have influenced their activities.

**Musical Metadata**: Similarly, Spotify listening history can be enriched by adding metadata such as track mood, genre, tempo, and energy levels. This helps users discover patterns in their listening habits during specific activities, such as upbeat music during workouts or calming music while working.

**Contextual Metadata for Events**: Calendar events or task lists imported from Google or Microsoft can be enriched with additional metadata like event duration, location type (virtual or physical), or participant details. This enables users to analyze how much time they spend in virtual meetings versus in-person ones, providing valuable insights into work-life balance or productivity trends.

**Example Use Case**: A user may want to analyze their Google Takeout location history. After importing the data, Datapunk enriches it with nearby landmarks, city names, and the weather conditions during each travel event. This enriched dataset can then be analyzed to uncover trends, such as how certain weather patterns affect the user’s travel habits.

**Cross-Source Correlation**

One of Datapunk's most powerful features is its ability to correlate datasets from different sources. This allows users to uncover relationships between various datasets, providing deeper insights and a holistic understanding of their digital lives. By connecting diverse data points, Datapunk reveals correlations that are often overlooked when analyzing datasets in isolation.

**Example Correlation**: A user might correlate their Google Calendar events with their Spotify listening habits to determine whether certain types of music improve productivity. For example, users can observe if they tend to focus better while listening to specific genres during work hours.

**Behavioral Insights**: Another example could involve correlating fitness data from wearable devices with social media activity from Meta or Instagram. This could help users understand how their physical activity influences their social media engagement, or vice versa.

Cross-source correlation transforms how users interact with their data, enabling them to ask complex questions like, "Does my mood change based on the type of music I listen to after certain meetings?" or "How do my daily activities affect my sleeping patterns?" By providing these insights, Datapunk empowers users to develop a comprehensive understanding of their behavior and make informed decisions based on their data.

**4. Analysis, Reporting, and Trend Detection**

Datapunk’s Analysis, Reporting, and Trend Detection feature(s) is designed to empower users by unlocking valuable insights from their digital data. It’s not just about looking at numbers—it’s about helping users take control of their story by finding the patterns that define their habits and behaviors. With advanced analysis, custom reports, and trend detection, Datapunk turns what might feel like an overwhelming sea of data into clear, actionable information. Whether you’re curious about how productive your days have been, how often you're scrolling through social media, or what your music choices say about your mood, Datapunk provides the tools you need to see and shape your digital footprint.

At the core of this capability lies Behavioral Insights and Trend Analysis. Datapunk is constantly at work behind the scenes, processing data from Google Calendar, Spotify, wearable devices, and more, to paint a real-time picture of a user's habits and trends. Imagine being able to clearly see how your productivity shifts throughout the day based on your calendar data or uncover how your music tastes have evolved over the past year. With Datapunk, these insights come to life, making it easy to notice changes that might affect your well-being—like productivity dips, boosts in social engagement, or even subtle mood fluctuations tied to different genres of music. Datapunk isn't just about showing users what happened; it’s about helping them understand why, so they can make informed decisions to enhance their lives.

Custom Reports and Summaries are another cornerstone of Datapunk’s power. These reports are crafted to give users a clear and detailed look at their activity over time. Whether you want to know how much time you've spent in meetings this month, how your communication habits stack up week over week, or which songs have been on repeat the most, Datapunk has you covered. Reports can be scheduled to run automatically or generated on-demand, making it easy to stay informed. Each report is visually rich—filled with graphs, timelines, and other visual elements that make interpreting data a breeze. For instance, a monthly productivity report might show your peak working hours, your most frequently contacted people, or how much focus time you’re losing to interruptions. These visual narratives allow users to see patterns without getting lost in the details.

One of Datapunk’s most distinctive abilities is Correlation and Pattern Detection. This feature goes beyond simply tracking your activities by revealing how different aspects of your life are interconnected. It provides a holistic view of your digital and physical experiences by correlating data from various sources. For example, you could analyze how your fitness routines influence your work productivity by comparing your step count data with calendar events. Or you might explore how different music genres align with your mood by linking Spotify listening habits with mood tracking data. By identifying these hidden relationships, Datapunk empowers users to understand the cause-and-effect dynamics within their daily routines, enabling them to optimize and create more balanced, productive lives.

Automated Trend Detection and Anomaly Alerts take Datapunk’s analysis a step further by actively monitoring your data to identify emerging patterns or unusual activities. Datapunk does the heavy lifting, highlighting notable trends—such as increases in screen time, sudden drops in physical activity, or changes in sleep quality—so that users don’t have to spend hours sifting through data. The anomaly detection feature is there to ensure that you’re alerted whenever something looks out of the ordinary, whether it’s a significant change in behavior or an unexpected pattern. For example, if there's a sudden drop in sleep quality, Datapunk can help you connect the dots—maybe it’s tied to an increase in late-night work sessions.

But it’s not just about the data; it’s also about how users interact with that data. Datapunk offers Interactive Data Exploration Tools that put the power back into users’ hands, allowing them to engage with their data visually and intuitively. With dynamic visualizations—like timelines, bar charts, and scatter plots—you can zoom in on specific data points, adjust time frames, and explore correlations. And for those who want to dig even deeper, the Custom Query Builder lets users create detailed, personalized analyses without needing technical know-how. It’s all about making the data work for you and answering the unique questions you have about your life.

To make insights more accessible, Datapunk also features Summarized Insights with Narrative Storytelling. Instead of overwhelming users with raw numbers or complex charts, Datapunk provides summaries in straightforward, conversational language. Imagine receiving a monthly summary that reads, “Your productivity peaked during the afternoons while listening to instrumental music—consider continuing that trend to stay focused.” These narrative insights help users quickly grasp the story their data is telling, empowering them to take action without needing to dive into the details.

By expanding the features of behavioral insights, custom reports, correlation detection, trend analysis, and interactive exploration, Datapunk helps users turn raw data into a rich, meaningful understanding of their digital habits. This isn’t just data analysis—it’s a movement towards personal empowerment, helping users take control of their information to build a better, more informed version of their digital selves.

# **Production Needs for Datapunk**

Data

Datapunk is all about taking back control of your personal data, and that means collecting data from the major players that have been holding it hostage. We’re pulling data from **Google**, **Spotify**, **Microsoft**, and **Meta** so users can create a comprehensive and unified view of their digital lives.

**Google Data** will be accessed via Google Takeout and APIs, gathering information from Google Calendar, Contacts, Location History, and Search History. Google Calendar data includes event titles, locations, participants, and recurrence patterns, which users can use to analyze their time management and social interactions. Contacts data provides names, phone numbers, and emails, allowing users to monitor their communication networks. Location History, with its timestamped GPS coordinates, offers insights into movement patterns and travel habits—giving users the ability to see where they’ve been and why it matters.

**Spotify Data** will be collected through Spotify’s API, including listening history, playlists, and top tracks. Users will get insights into their music habits, from how often they’re listening to that favorite song to broader patterns like which genres help them focus or relax. With this data, users can start to make connections between their music preferences and their productivity or mood, giving them even more control over their routines.

**Meta Data**, sourced from Facebook and Instagram, will be accessed through exports and APIs to retrieve posts, comments, reactions, and shared media. These insights will help users track their social media engagement, understand what kind of content they interact with most, and visualize their connections. Whether it’s looking at how much time you spend reacting to cat videos or exploring which posts spark the most engagement from friends, Meta data adds depth to understanding social habits.

**Microsoft Data**, such as Outlook and Teams, will be accessed via Microsoft Graph API. This data will include calendar events and contacts from Outlook, giving insight into work-life balance, while Teams collaboration data will provide a view into professional productivity and meeting habits. By combining these datasets, users can understand how they’re spending their time professionally, spot inefficiencies, and improve how they interact with colleagues.

## Software

**Backend Development**: The backend of Datapunk will be built using **Python**, specifically with the **Django** framework. Django is here to keep things secure, robust, and make sure everything flows smoothly. With powerful libraries like **Pandas** and **ijson**, we’ll be handling data imports, parsing, and cleaning from services like Google, Spotify, and Meta like pros. These tools ensure that even large volumes of JSON or CSV data are processed seamlessly, meaning users get their data without the headaches.

**Database Management**: **PostgreSQL**, combined with **PostGIS** for geospatial data capabilities, will be our main database. PostgreSQL is solid, reliable, and scalable, which is exactly what we need for handling everything from location history to communication records. With **PostGIS**, we’ll add geospatial smarts—allowing users to analyze where they’ve been, map out their routines, and dig into their own movement data. Plus, **Docker** will be used to containerize the whole setup, making sure deployments are consistent and setup is hassle-free for users.

**Frontend Development**: The frontend will be built with **React** to ensure the user experience is interactive and intuitive. React lets us craft a highly customizable dashboard where users can view trends and insights from their data in real time. Libraries like **Recharts** and **React-Leaflet** will help bring the data to life with dynamic visualizations—whether it's creating beautiful charts or mapping your latest road trip. It’s all about making data both insightful and visually appealing, so users can see both the big picture and the small details.

**Security Tools**: Security is key. We’ll be using **LUKS** for encrypting data at rest, particularly for local storage on Linux systems like **NixOS**. For sensitive information stored in the PostgreSQL database—think OAuth tokens and anything personally identifiable—we’ll rely on **pgcrypto** for encryption. This layered approach ensures that user data remains secure both in storage and in transit. On top of that, **OAuth 2.0** will be used to securely manage integrations with platforms like Google and Spotify, ensuring data transfer is as safe as possible every step of the way.

## Hardware

Datapunk is designed to run on **local user machines**, primarily desktop environments, and that means no cloud dependency or hidden infrastructure. Users need machines capable of handling Docker, as the entire Datapunk environment will be containerized to keep everything consistent and simple to set up. The suggested system requirements include a **multi-core CPU** (like an Intel i5 or equivalent AMD Ryzen) to handle data parsing, database management, and real-time visualizations with ease.

To keep things running smoothly, we recommend at least **8GB of RAM**, especially when dealing with larger datasets or running multiple containers. For storage, **SSD** is the way to go, with at least **50GB of free space** to accommodate imported data, Docker volumes, and databases. The speed of an SSD makes a big difference when you’re processing large files or generating visualizations on the fly.

The platform will primarily support **Linux** operating systems, with a preference for **NixOS** because of its security and reliability. That said, **Windows** and **macOS** will also be supported, making Datapunk accessible across different user systems. As for graphics, a standard integrated or dedicated GPU will do the trick—while data visualizations aren't heavy on graphics, having a decent GPU ensures everything renders smoothly, especially in the interactive dashboard.

## Funding

**Development Costs**: Datapunk may be open-source, but that doesn’t mean it’s free to build. **API access fees** for Google, Spotify, and Microsoft are a real expense, particularly if we need to bump up to higher usage tiers to support more users. These costs are essential to keep data imports smooth and reliable, so users aren’t stuck waiting or dealing with limited features.

**Hosting and Distribution**: Although Datapunk is all about running locally, there’s still a need to host things like the code repository, documentation, and support resources. We’ll likely use **GitHub** for code hosting, and services like **Netlify** or **Vercel** for documentation and user guides. While the free tiers of these platforms may work for a while, we’ll need funding to scale if the project grows and demands more resources.

**Encryption Tools and Security**: While tools like **LUKS** for disk encryption and **pgcrypto** for database security are free to use, implementing them correctly and ensuring best practices may require consulting services. We’re skipping the full **third-party security audit** for now—that’s something to consider when Datapunk’s user base is big enough to justify the cost—but we’re making sure the basics are airtight from day one.

**Future Development and Scalability**: Datapunk’s main focus is staying local, but we’re not ruling out future scalability. Adding advanced analytics or AI-driven features down the line could mean moving to cloud infrastructure, and that’s where things get expensive. If we want to offer real-time analytics and need additional processing power or storage, we’ll be looking at cloud services like **AWS** or **Azure**, which could significantly increase costs. Funding will be critical for that kind of growth, but for now, we’re keeping it lean and local.

# User needs

**End User Groups**

The end users of Datapunk are a diverse group of individuals who all share a common goal: taking control of their personal data. These users range from tech-savvy data enthusiasts to privacy-conscious individuals who are tired of having their personal information collected, stored, and monetized by big tech corporations without their explicit consent.

**Privacy Advocates** are highly concerned about the lack of control they have over their personal data. They are looking for a solution that allows them to reclaim ownership and organize their information in a secure, decentralized environment. They want to stop relying on platforms that profit from their data without providing transparency. Datapunk empowers these users by giving them the tools to securely store and analyze their personal data without relying on cloud services controlled by corporations like Google, Facebook, or Microsoft.

**Tech-Savvy Users** include individuals who are comfortable working with data and technology but want more transparency and control over how their digital information is handled. They appreciate features like the ability to import data from multiple sources (Google, Spotify, Microsoft) and analyze trends across different datasets. These users enjoy exploring insights through Datapunk’s customizable visualizations and dashboards, correlating data from multiple platforms to reveal hidden patterns in their behavior.

**Neurodivergent Users** find value in having control over their digital environment to manage sensory overload and improve their understanding of personal habits and routines. Datapunk is designed to cater to these users with customizable dashboards, accessible data visualizations, and low information density modes. They can personalize the system to suit their preferences, making the tool adaptable and user-friendly.

**Data Privacy Enthusiasts and Open-Source Advocates** are drawn to Datapunk for its commitment to the principles of open-source software and personal data sovereignty. They value tools that align with their philosophy of decentralization, privacy, and transparency. These users are also likely to contribute to the project, building custom integrations or improving existing features within the Datapunk community.

These user groups represent the broad spectrum of people who will benefit from Datapunk’s mission of empowering individuals to reclaim control over their digital footprint in a private, secure, and user-friendly way.

**Training Needs for Effective Use**

To effectively use Datapunk, end users will require training across several key areas. While the platform is designed to be user-friendly, the variety of data sources, integrations, and customizable features means users will need a solid understanding of how to manage and analyze their data effectively.

Users will need training on how to import data from services like Google, Spotify, and Microsoft, both via direct API integration and manual file uploads. The training should explain how to authorize third-party services using OAuth and navigate Datapunk's data source management system. Clear instructions on how to schedule automatic imports and manually trigger data updates will be essential to ensuring users can keep their datasets current. For manual uploads, users will need guidance on which formats are supported (e.g., JSON, CSV), how to map uploaded data to the appropriate fields, and how to troubleshoot errors during the import process. This includes understanding error messages and the steps to resolve common issues, such as incorrect file formats or missing data.

Users must be trained on how to navigate and customize their interactive dashboard. Training should include how to add and remove widgets, adjust layouts using the drag-and-drop interface, and configure specific visualizations like pie charts, bar graphs, and heatmaps. Instruction on how to use drill-down features to explore data in more depth will be vital to unlocking the full analytical power of Datapunk. Customizing themes, managing low-information-density modes for accessibility, and adjusting the time ranges for different datasets will also be covered in the training. This will empower users to tailor their dashboards to their personal preferences and specific use cases.

Users will also need a clear understanding of how Datapunk parses and cleans data, as well as how it enriches datasets with additional metadata. This training should explain how to use pre-built parsers for services like Google and Spotify and what to do in cases where data requires custom parsing solutions. Tutorials should also cover data cleaning tools, including how to handle redundant or incomplete data, manage duplicates, and enrich records with geographic or contextual metadata, like adding city names to location history or genres to music data.

Lastly, users will benefit from training on how to perform in-depth analysis and generate custom reports. This includes learning how to use Datapunk’s trend detection, clustering, and correlation features to uncover hidden patterns in their data. Users should be guided on how to set up reports that automatically summarize key metrics such as time spent in meetings or spending habits. Additionally, they will need to understand how to manage permissions for third-party integrations, encrypt their data, and control data retention settings. This will include managing OAuth tokens, setting up secure backups, and configuring data masking for exporting or sharing insights.

**Documentation Provided for Datapunk**

To help the intended users benefit from Datapunk, we will provide comprehensive and user-friendly documentation. This documentation will guide users through every feature and function of the platform, ensuring that they can navigate and use Datapunk effectively, regardless of their technical expertise.

The **User Guide** will serve as the foundational document, explaining each feature step-by-step. It will cover essential tasks such as setting up and configuring data sources, importing data manually via file uploads, managing data imports, and customizing the dashboard. Users will learn how to add, remove, and adjust widgets to personalize their experience, making their dashboard a true reflection of their preferences and needs.

The **Data Import and Parsing Documentation** will provide detailed instructions on how to integrate data from multiple platforms, whether through real-time API connections or manual uploads. It will also cover how data is parsed, cleaned, and enriched with metadata, so users understand how their information is transformed into meaningful insights.

The **Data Visualization and Dashboard Customization** section will focus on the various visualization options available. Users will learn how to create and customize different types of visualizations, including charts, graphs, heatmaps, and geospatial maps, to explore trends and patterns. Additionally, it will explain the drill-down functionality, allowing users to delve deeper into specific metrics and uncover more granular insights.

**Data Security and Privacy Controls** will be covered in a dedicated section to ensure users are fully informed about how their data is protected. This part of the documentation will explain managing permissions for integrated services, setting up encryption for local data storage, and using privacy settings to anonymize data when needed.

Finally, a **Troubleshooting and Error Handling** section will be included to help users resolve any issues they encounter. This will cover common errors and their solutions, including step-by-step guidance on how to fix data import problems or resolve authorization issues, ensuring users are never stuck without help.

# Security and privacy

Datapunk ensures **data encryption at rest and in transit**. All data stored locally on users' devices is encrypted using **LUKS (Linux Unified Key Setup)** for partition-level encryption, which means that even if a device is compromised, the data remains protected. Sensitive fields within the PostgreSQL database, such as OAuth tokens, are encrypted using **pgcrypto** for an additional layer of security. Communication between Datapunk and third-party services, or between local components (such as the UI and API), is secured using **TLS (Transport Layer Security)** to prevent unauthorized access during data transfers.

Datapunk uses **OAuth 2.0** for secure authentication, managing user access to third-party services like Google, Microsoft, and Spotify. This means that Datapunk only accesses the data that users explicitly consent to share. OAuth tokens are encrypted at rest and securely stored to prevent unauthorized access.

Users have full control and transparency over what data is imported and how it is stored. They can manage permissions for each integration, such as Google Calendar or Spotify, revoke access, and delete specific datasets when necessary. Datapunk provides a clear permissions dashboard, allowing users to review and adjust data access settings at any time.

When users wish to share data reports or insights, Datapunk includes a **data masking** feature to anonymize personally identifiable information—such as names, email addresses, and locations—before exporting. This feature ensures that users can share insights without compromising their privacy.

For added security, users can create encrypted backups of their data using **pg\_dump** with **GPG encryption**. These backups can be securely stored and restored when needed, providing strong protection for sensitive data during the backup process.

Datapunk’s **privacy-centered design** ensures that it runs locally on the user’s machine, reducing reliance on third-party cloud services. This means that user data remains in their control, rather than being stored on corporate servers. Additionally, users can configure **automatic data retention policies** and have the option to delete their data entirely from the system, ensuring full ownership and privacy.

# Summary

The **Datapunk project** is all about putting power back where it belongs—in the hands of individuals. This open-source platform will let users reclaim, organize, and take control of their personal data from services like Google, Spotify, Microsoft, and Meta. Datapunk consolidates all that scattered data into a user-controlled, decentralized environment, making sure users can store, visualize, and analyze their data privately, without big tech breathing down their necks.

**Datapunk** will be a game-changer for **privacy-conscious individuals**, **tech-savvy data enthusiasts**, and **neurodivergent users**. It empowers them to derive meaningful insights from their digital lives through features like data imports, customizable dashboards, interactive visualizations, and secure storage. Datapunk gives users a clear, detailed, and unified picture of their online habits—whether that's media preferences, productivity trends, or social interactions—all without relying on exploitative corporate cloud services.

This project **does** involve working with private data, but we’ve got the safeguards covered to keep it safe. We're talking **encryption at rest and in transit**—using **LUKS** for local storage and **TLS** for secure communication. Integrations with third-party services are managed through **OAuth 2.0**, making sure that Datapunk only accesses data users explicitly consent to share. Plus, when users need to share insights, Datapunk offers **data masking** to anonymize all personally identifiable information.

To pull this off, we’re using a range of powerful tools. The backend is built with **Python** and **Django** for a secure and robust system, while **PostgreSQL/PostGIS** handles data storage—including geospatial data for users’ location history. The frontend is powered by **React** for an interactive, smooth experience, with **Pandas** managing data analysis under the hood. Security is taken care of with **pgcrypto** for database encryption, ensuring that sensitive data is always protected. And since everything runs locally, there’s no need for any special hardware—just a computer that can handle **Docker**, and you're good to go.

If funding is needed, it’ll go towards **API access fees** for services like Google and Spotify, and to support documentation and hosting services such as **GitHub** or **Netlify** for when the community starts to grow. For now, though, the focus is on keeping things simple, secure, and in the hands of the user.

And honestly, let’s face it—Datapunk is the antidote to all the nonsense we’ve come to expect from the digital world. Tired of companies making billions off your data while you get nothing but targeted ads? Yeah, us too. Datapunk is here to flip the script. We’re giving you the tools to look at your own data, draw your own conclusions, and use it how you see fit. No more mystery algorithms or surprise ads that know more about you than your best friend. With Datapunk, you’re in control—no strings, no fine print, no corporate overlords. Just you and your data, the way it should be.