**Initial Narrative (Before Selections):**

To accomplish our research objectives, we will utilize **[1: data sources/types]** and collect **[2: identifiability levels]** datasets that may include **[3: sensitive data categories]**, obtained from **[4: data origin]**. These datasets primarily consist of **[5: data formats]**, and we anticipate a dataset size and complexity of approximately **[6: data volume & complexity]**.

**Clickable Areas and Selection Options:**

1. **[1: data sources/types]** (Multi-select checkboxes)
   * Internally sourced clinical data (Mayo EHR)
   * Externally sourced data (under Data Use Agreement)
   * Publicly available datasets
   * Prospectively collected patient data at Mayo Clinic
   * Other (free-text)
2. **[2: identifiability levels]** (Multi-select checkboxes)
   * Fully identifiable PHI
   * Limited dataset (some indirect identifiers)
   * De-identified dataset (no direct or indirect identifiers)
   * Other (free-text)
3. **[3: sensitive data categories]** (Multi-select checkboxes)
   * Genetic or genomic information
   * Mental/behavioral health data
   * Reproductive health data
   * Substance use data
   * None
   * Other (free-text)
4. **[4: data origin]** (Multi-select checkboxes)
   * Retrospective (existing Mayo EHR data)
   * Prospective (newly collected from Mayo patients)
   * External collaborator data (under DUA)
   * Publicly available data source
   * Other (free-text)
5. **[5: data formats]** (Multi-select checkboxes)
   * Structured clinical data (labs, vitals)
   * Imaging data (radiology scans, pathology slides)
   * Genomic/molecular data
   * Textual clinical notes (unstructured)
   * Other (free-text)
6. **[6: data volume & complexity]** (Drop-down single-select)
   * Small dataset (<1GB)
   * Moderate dataset (1GB-100GB)
   * Large dataset (>100GB)
   * Not yet determined
   * Other (free-text)

Below are examples of how each selection might trigger additional sentences or fields. Each trigger is limited to expanding details under the “Study Resources & Data Types” scope, and does not address other submodules (like Data Security & Protection). The idea is that for each selection made, additional text or small input fields are revealed inline, allowing the user to refine their narrative in a guided, natural language fashion.

**Expansion Logic by Selection**

**1. Data Sources/Types (Multi-Select)**  
For each data source/type selected, add a sentence or fields for clarification:

* **Internally sourced clinical data (Mayo EHR)**
  + Add a sentence: “The internal EHR data will be derived from Mayo Clinic’s secure clinical repositories.”
  + Field: [Specify any particular clinical domain or department (optional free-text)]
* **Externally sourced data (under Data Use Agreement)**
  + Add a sentence: “This external data will be acquired from collaborating institutions under a formal Data Use Agreement.”
  + Field: [Name of external collaborator / institution (free-text)]
  + Field: [Type of DUA (drop-down: Standard DUA, Custom Institutional DUA, Other)]
* **Publicly available datasets**
  + Add a sentence: “We will also incorporate publicly available datasets that are already de-identified and open for research use.”
  + Field: [Name/URL of the dataset source (free-text)]
* **Prospectively collected patient data at Mayo Clinic**
  + Add a sentence: “We intend to collect new patient data from Mayo Clinic participants following IRB-approved consent or waiver processes.”
  + Field: [Estimated number of participants (numeric)]
  + Field: [Type of clinical unit or setting where data is collected (free-text)]
* **Other (free-text)**
  + Add a sentence: “In addition, we will utilize other data sources as described below.”
  + Field: [Describe this other source (free-text)]

**2. Identifiability Levels (Multi-Select)**  
For each identifiability level chosen, prompt for further details:

* **Fully identifiable PHI**
  + Add a sentence: “We will be handling fully identifiable PHI.”
  + Field: [Type(s) of identifiable information included (multi-select: patient name, MRN, DOB, address, etc.)]
  + Field: [Rationale for needing PHI (free-text)]
* **Limited dataset (some indirect identifiers)**
  + Add a sentence: “We will include a limited dataset that may contain indirect identifiers, such as dates of service.”
  + Field: [Specify which indirect identifiers (multi-select: dates, zip codes, etc.)]
* **De-identified dataset (no direct or indirect identifiers)**
  + Add a sentence: “We will use a de-identified dataset, ensuring all direct and indirect identifiers have been removed according to HIPAA Safe Harbor standards.”
  + Field: [Method of de-identification (drop-down: manual review, automated de-identification tool, other)]
* **Other (free-text)**
  + Add a sentence: “The dataset identifiability will differ from standard categories, as specified below.”
  + Field: [Describe identifiability approach (free-text)]

**3. Sensitive Data Categories (Multi-Select)**  
For each sensitive category selected, add clarifications:

* **Genetic or genomic information**
  + Add sentence: “Genomic information will be included, requiring careful consideration of re-identification risk.”
  + Field: [General type of genomic data (drop-down: WGS, WES, targeted panel, SNP array)]
* **Mental/behavioral health data**
  + Add sentence: “Data related to mental/behavioral health will be included, necessitating careful handling and additional protections.”
  + Field: [Specify data type (e.g., clinical notes, psychiatric evaluations) (free-text)]
* **Reproductive health data**
  + Add sentence: “Reproductive health-related information will be part of our dataset.”
  + Field: [Type of reproductive data (e.g., fertility treatments, obstetric records) (free-text)]
* **Substance use data**
  + Add sentence: “Substance use information will be included, which may be subject to additional regulatory considerations (e.g., 42 CFR Part 2).”
  + Field: [Substance type or general category (free-text)]
* **None**
  + Add sentence: “No sensitive health categories will be included in these datasets.”
* **Other (free-text)**
  + Add sentence: “Other sensitive health categories, as described below, will be included.”
  + Field: [Describe the sensitive data category (free-text)]

**4. Data Origin (Single-Select)**

* **Retrospective (existing Mayo EHR data)**
  + Add sentence: “The data will be retrospectively extracted from existing EHR records at Mayo Clinic.”
* **Prospective (newly collected from Mayo patients)**
  + Already covered in #1 expansions if selected, but here you could add:
  + Add sentence: “Data will be collected prospectively from patients during their standard clinical visits.”
* **External collaborator data (under DUA)**
  + Already addressed above in #1 expansions, but you might add a reinforcing sentence:
  + Add sentence: “This data will be obtained from external sources that have been secured with a DUA.”
* **Publicly available data source**
  + Add sentence: “Publicly available data will be sourced from reputable repositories (e.g., NIH, PhysioNet).”
* **Other (free-text)**
  + Add sentence: “We will obtain the data from other origins, as specified below.”

**(In many cases, origin specifics are tied closely to the sources chosen in #1, so you may keep #4 minimal if #1 already captures enough detail.)**

**5. Data Formats (Multi-Select)**  
For each chosen format, prompt for basic details:

* **Structured clinical data (labs, vitals)**
  + Add sentence: “Our datasets include structured clinical fields, such as laboratory results and vital signs.”
  + Field: [Specify any key clinical domains (e.g., oncology labs, cardiology vitals) (free-text)]
* **Imaging data (radiology scans, pathology slides)**
  + Add sentence: “We will incorporate imaging data, which may include radiology scans or pathology slides.”
  + Field: [Modality selection (drop-down: MRI, CT, X-ray, Histopathology)]
  + Field: [Approx. number of images (numeric)]
* **Genomic/molecular data**
  + Add sentence: “Genomic or molecular data will be included for analysis.”
  + Field: [Data type (e.g., WES, WGS, gene expression data) (free-text)]
* **Textual clinical notes (unstructured)**
  + Add sentence: “Unstructured textual data, such as physician notes, will be analyzed.”
  + Field: [General data cleaning steps planned (free-text)]
* **Other (free-text)**
  + Add sentence: “Additional data formats not listed above will be included.”
  + Field: [Describe format (free-text)]

**6. Data Volume & Complexity (Single-Select)**

* **Small (<1GB)** or **Moderate (1GB-100GB)** or **Large (>100GB)** or **Not yet determined**
  + Typically no further expansion needed here. This information is usually enough for high-level IRB planning. However, you could add:
  + Add sentence: “We anticipate the total data volume to be [selected option], which influences our resource planning.”

**Summary of Expansion Behavior**

* **For Multi-select fields (#1, #2, #3, #5):**  
  Each selection triggers a sentence and possibly a short set of fields (dropdowns or free-text) to elaborate.
* **For Single-select fields (#4, #6):**  
  Generally one sentence of elaboration, possibly with a small prompt for details if needed.

This approach ensures that the narrative dynamically updates with targeted, relevant details without overburdening the user. Each expanded section remains focused on “Study Resources & Data Types,” while additional complexities (e.g., security measures, sharing agreements, access controls) can be handled in the subsequent modules.