### Objectives and Goals

The main objectives of this project are to:

1. Assess the current level of compliance with Section 508 across key HUD documents.
2. Identify which policies need to be updated to meet compliance standards.
3. Determine which policies, if any, are already compliant with Section 508.
4. Provide an initial audit to gauge the scope of work required to ensure full compliance.

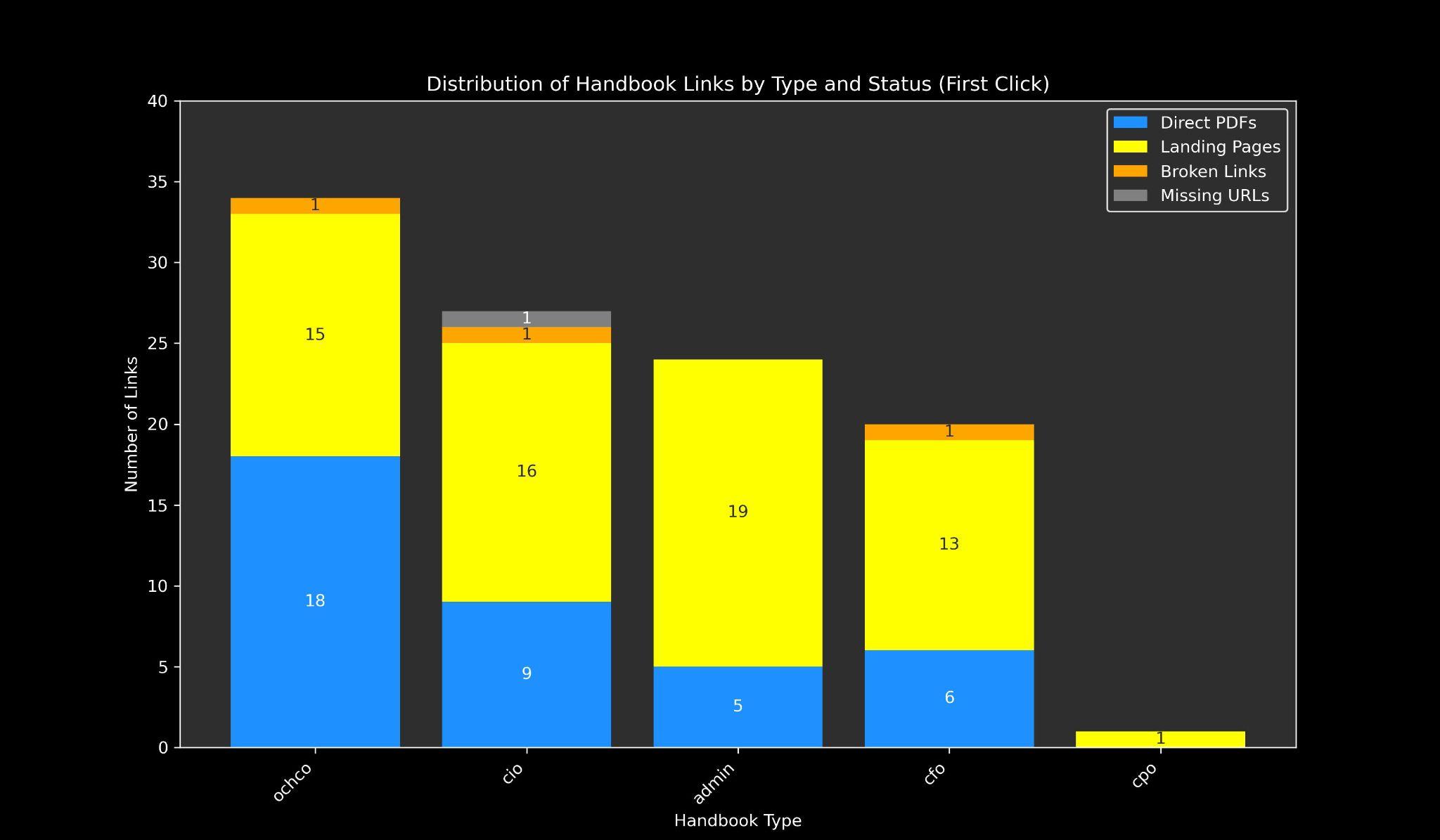
### Current Status and Findings

Completed web scraper designed to audit five key handbooks for HUD:

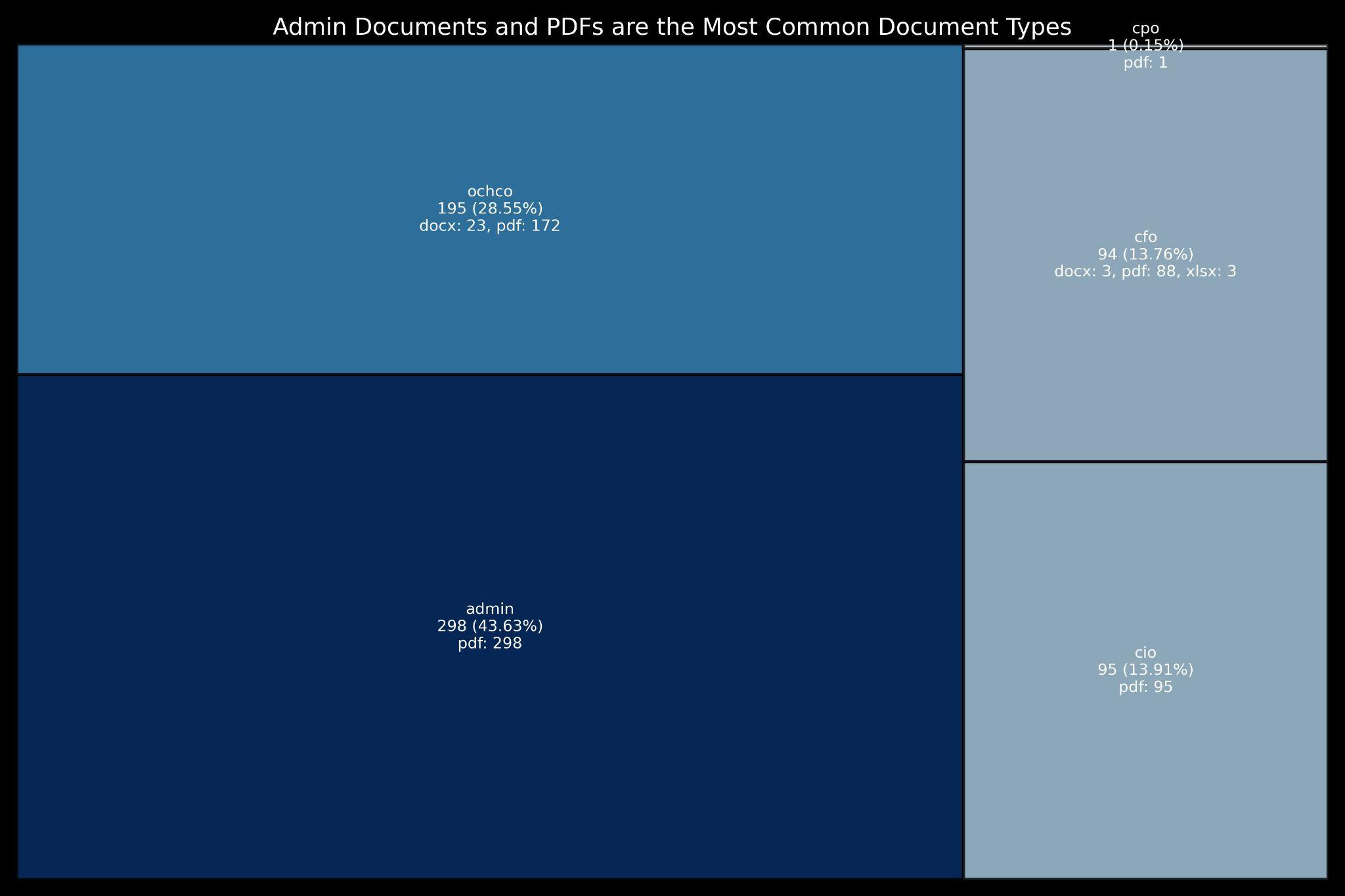
1. Administration Handbooks
2. Chief Financial Officer Handbooks
3. Chief Human Capital Officer Handbooks
4. Chief Information Officer Handbooks
5. Chief Procurement Officer Handbook

The initial audit has yielded the following key findings:

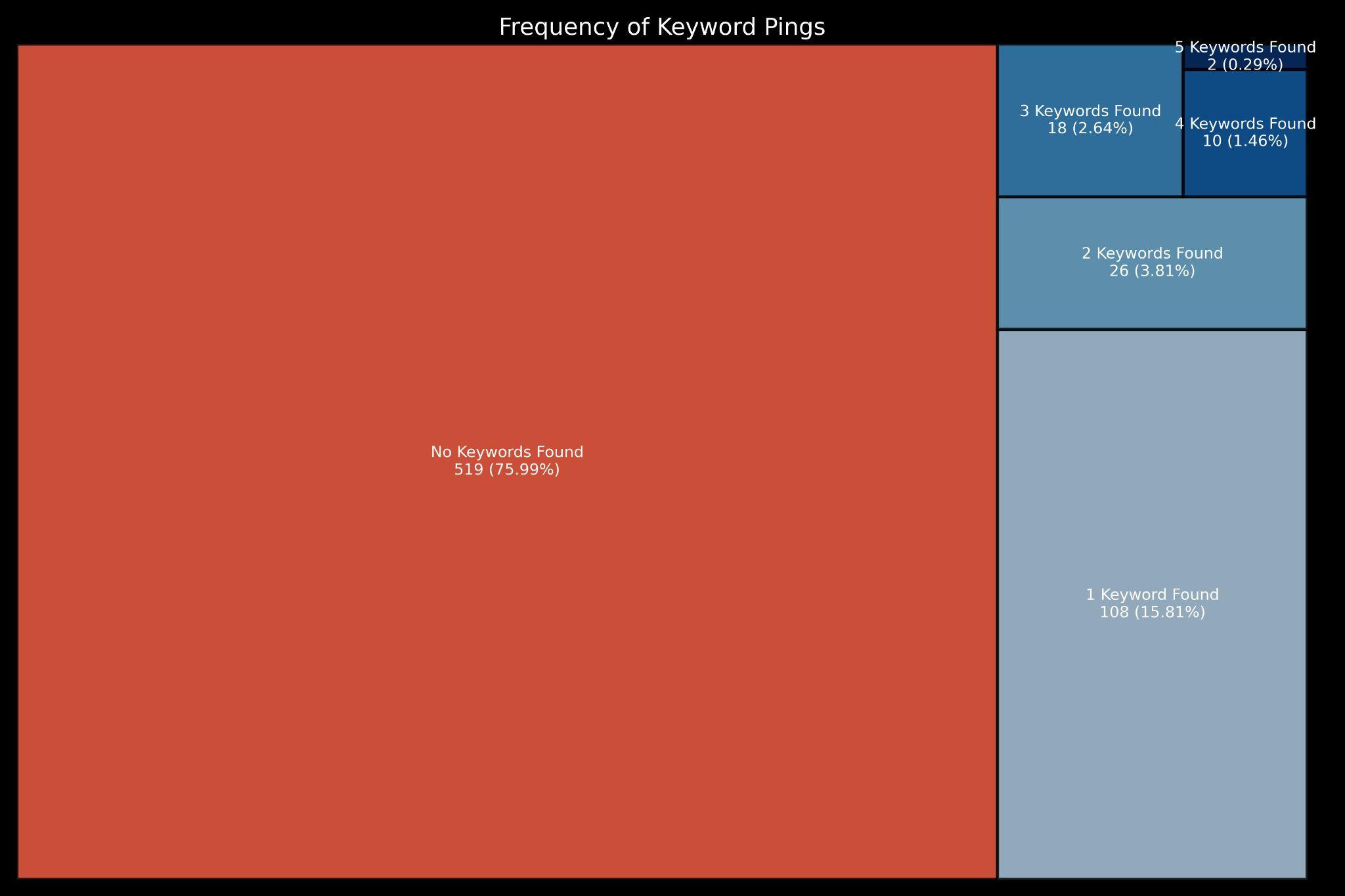
* About ⅓ of the “first clicks” into these handbooks lead straight to the policies intended to be scraped while the remainder have an intermediary landing page.
* 3 of these first clicks had broken links. They all had URLs beginning with “hudatwork.”



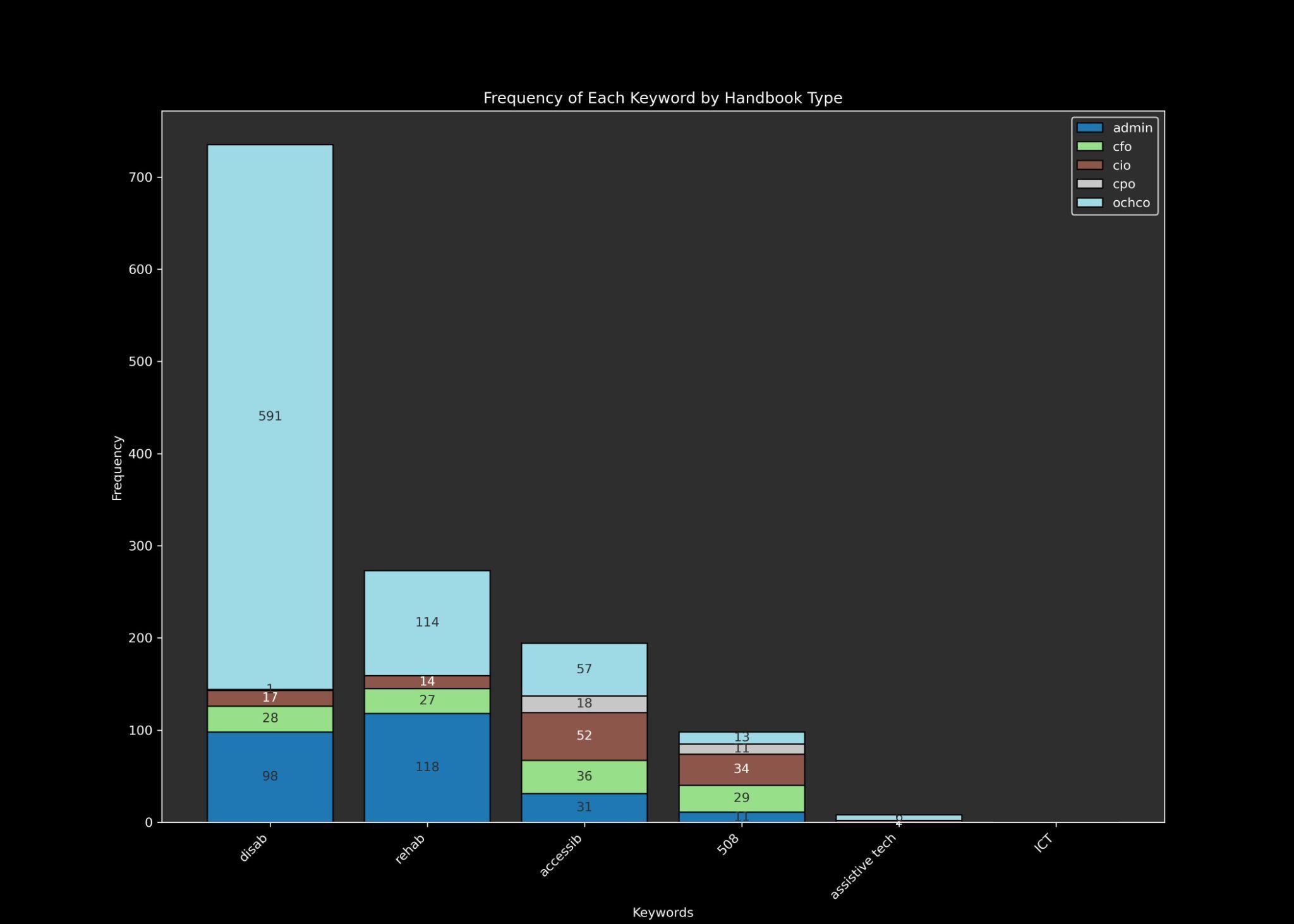
* The most commonly referenced documents are from the Administration Handbooks.
* PDFs are the most common document type.



* Approximately 519 (76%) of documents scraped contained no relevant keywords, indicating potential non-compliance or lack of explicit accessibility mentions.
* A total of 164 instances were identified where at least one keyword related to Section 508 compliance was found.
* Of these, the majority (15.81%) contained only one keyword, while 56 documents contained multiple keywords.



* The keyword "disab," a stem word for "disabled" or "disability," was the most frequently flagged term, predominantly in the Chief Human Capital Officer Handbooks (OCHCO).
* Other common keywords in descending order of frequency included "rehab," "accessib," and "508."
* Keywords such as "assistive tech" and "ICT" returned negligible results.
* The most keyword pings overall were found in Administrative (admin) and OCHCO handbooks.



* A total of 55 (8.05%) of documents were identified as unsearchable, likely meaning they are photocopies or images of some type.

### Proposed Next Steps

To continue the progress made in this project, the following next steps are proposed:

1. **Refine Keyword Relevance**:
   * Develop a process to identify and filter out false positives by analyzing the context in which keywords appear.
   * Define language parameters that indicate whether a keyword occurrence is relevant to Section 508 compliance.
2. **Enhance Data Collection**:
   * Extend the web scraper to include additional HUD handbooks or other relevant documents not covered in the initial audit.
   * Include metadata such as publication dates to provide more context.
3. **Improve Searchability Assessment**:
   * Implement more robust techniques to determine the searchability of documents, including a process for scanned documents or image recognition / analysis.
   * Develop a system to categorize documents based on their searchability and format.
4. **Stakeholder / Policy Engagement**:
   * Collaborate with other teams or interns to review and validate findings.
   * Engage with policy makers and document owners to gather feedback and plan updates.
5. **Automate Compliance Reporting**:
   * Create automated tools to generate compliance reports based on the audit findings.
   * Develop dashboards to visualize compliance status and track progress over time.
6. **Expand Keyword Library**:
   * Update the keyword list to include more terms related to accessibility and compliance as identified in the ongoing review.
   * Use machine learning techniques to identify new relevant keywords from the document corpus.

### Conclusion

The next phase of this project aims to build on the initial audit's success by refining the analysis, improving data collection, and identifying who to engage with to refine both the data work and the policies themselves.