

Head of Content Capture

Coherent Digital is looking for an innovative and highly organized *Head of Content Capture* to lead the identification, harvesting, and ingestion of high-quality, hard-to-find digital content. This role will develop and manage automated and semi-automated tools and techniques—leveraging AI, advanced web searches, third-party tools, and open web databases—to ensure that content is captured comprehensively, accurately, and ethically.

The ideal candidate will combine technical fluency, strong research instincts, and project management skills. This is a critical position supporting the scalability and sustainability of Coherent’s mission to preserve and enrich at-risk and born-digital materials.

Key Responsibilities

- **Tool & Technique Development**
Design and refine methods for discovering and capturing content using AI-driven search, web crawlers, public APIs, link analysis, and scraping tools—within legal and ethical boundaries.
- **Content Source Discovery**
Proactively identify new sources of high-value content relevant to Coherent’s databases (e.g., from NGOs, government agencies, academic grey literature, defunct websites, and emerging web collections).
- **Workflow Design & Oversight**
Manage pipelines for automated and semi-automated content ingestion. Ensure the fidelity, formatting, and completeness of ingested materials.
- **Contractor & Vendor Management**
Oversee third-party contractors and outsourced teams involved in content capture. Provide clear specifications, QA oversight, and performance monitoring.
- **Data Quality Assurance**
Develop and enforce standards for quality, format, and metadata expectations in captured content. Work closely with enrichment teams to ensure downstream usability.
- **Monitoring & Reporting**

Track and report on capture metrics such as:

- *Volume of content captured*
 - *Coverage of targeted source domains*
 - *Cost per unit of content captured*
 - *Capture accuracy and comprehensiveness*
 - *Ethical compliance and tool transparency*
- **Cross-Team Collaboration**
Collaborate with engineering, product, and content teams to successfully deliver content to the platform.

Skills and Qualifications

- Background in library and information science, digital preservation, or open data initiatives.
- Experience with web crawling, automated harvesting, and content aggregation technologies.
- Familiarity with tools like Heritrix, Scrapy, Webrecorder, or AI-based discovery and extraction tools.
- Strong understanding of content rights, ethical web scraping, and data licensing issues.
- Experience managing distributed or third-party teams on technical and content-driven projects.
- Knowledge of digital libraries, grey literature, or born-digital collections is a plus.
- Comfort working with structured and unstructured data; ability to troubleshoot ingestion errors and formatting inconsistencies.
- Strong documentation, workflow design, and project management skills.

Success Metrics

- **Scale:** Volume of high-quality content captured and ingested.
- **Efficiency:** Cost per item or source ingested, and time to ingest.
- **Accuracy:** Completeness and correctness of captured material.
- **Ethics:** Use of responsible, transparent, and legal capture methods.
- **Impact:** Support for downstream metadata enrichment and product development.

Reporting and Classification

- **Reports to:** André Avorio, Chief Product Officer
- **Classification:** Exempt

Coherent Digital is an equal opportunity workplace. We are committed to equal employment opportunities regardless of race, color, ancestry, religion, sex, national origin, sexual orientation, age, citizenship, marital status, disability, gender identity, veteran status or any other protected characteristic as outlined by federal, state, or local laws.

Apply now!

To apply for this position, please email aavorio@coherentdigital.net with your resume and cover letter.