CONTENT ETHICS – DIGITAL ETHICS IN ECM

INTRODUCTION

Enterprise content management is a field that has been evolving since its inception, when it was called document imaging and was essentially a digital replacement to microfilm. It rapidly evolved to include document management (version control), business process automation (workflow), web-based access, and integration. It is now beginning to encompass emerging technologies such as AI/ML, RPA, and blockchain and is transitioning to the cloud as its primary mode of deployment. Its names have changed through this quarter century's evolution to include 'imaging', 'document management', 'enterprise content management' (ecm), 'content services', and 'intelligent information management'. (See AIIM's The Next Wave: Moving from Ecm To Intelligent Information Management.) A significant characteristic of ecm platforms today and even from the beginning is that they have an extremely broad scope in terms of horizontal functions (human resources, finance and accounting, administration, engineering, etc.) and industries (finance, healthcare, government, manufacturing, retail, and so on). This makes ecm quite distinct from platforms built for specific functional areas such as enterprise resource programs (ERPs), financial systems, human resource applications, and customer relation management (CRM) platforms.

The evolution of ecm has paralleled the development of applied ethical fields in academic research and education as well as specializations in law. The rise of large mainframe database in the 1970s raised privacy concerns, while the creation of the internet/web and personal computing in the 90s/2000s gave rise to research areas such as computer ethics, cyber ethics, information technology ethics, and information ethics, etc. Emerging technologies such as those listed above have spurred new subdisciplines such as data ethics and AI ethics. The term **digital ethics** has also been introduced in order to encompass issues raised by emerging technologies such as those listed above and IoT, big data, and others (often associated with digital transformation). It is one of many labels used over the decades but it captures the spirit of the moment. (See my paper "Digital Ethics and other Labels" for a brief survey of the various terms used for the fields.)

As ecm and ethical research have developed in parallel, the two have generally intersected where law has codified an ethical minimum. This is particularly evident in the area of data privacy, which is a topic accounted for by ecm technology developers, implementers, and user organizations, though even here it has not had the same level of traction as one might expect. Furthermore, broader ethical concerns have not been at the forefront of ecm, even though its core technologies (databases, digital content, automation) have been their focal point in the broader academic and technical communities. As intelligence and automation capabilities increase, digital ethics needs to be a part of development and implementation roadmaps. Further, it has to be tailored to and applied to the processes and situations related to the management of (unstructured) content. To highlight this focus, I will use the term **content ethics** to describe digital ethics in relation to enterprise content management.

DIGITAL ETHICS

Given that **content ethics is envisioned as the specification of digital ethics in relation to content management** issues, this section will provide a characterization of digital ethics. In my view, 'digital ethics' includes traditional ethical issues in information technologies and information management, but also includes issues arising from big data and data analytics, intelligent automation, and artificial intelligence/machine learning, facial recognition, RPA, IoT, and more. A shorthand way of characterizing **digital ethics** is as a **combination of information ethics and AI ethics**.

The table below lists topic areas associated with the long-standing field of information ethics and the more recent field of AI ethics.

Digital Ethics	
Information Ethics	AI Ethics
▪ Data Privacy ▪ Confidentiality/Disclosure ▪ Cybersecurity ▪ Intellectual property; copyright, trade secrecy ▪ Lifecycle management / information governance.	▪ Magnified privacy concerns ▪ Bias (Discrimination) ▪ Due process ▪ Opacity (black boxes) / Explainable AI (XAI) ▪ Risk and responsibility ▪ Automation and employment ▪ Public good (government, healthcare, environment)

The topics listed on the left under the heading **information ethics** have been a matter of concern for decades. Data privacy arose as an issue in the 1970s when mainframe computers and large databases came into use by governments and large corporations. Cybersecurity and intellectual property gained importance with the rise of the internet. Lifecycle management has been the central function of records management for decades, but has recently evolved into or become a part of information governance, which addresses the other issues listed in this column from a management and compliance perspective. (Information governance is a multi-disciplinary concept, with roles falling in functional areas such as privacy law, records management, computer security, etc.) (For a more detailed account of ethics and information ethics, see my book, Ethics for Records and Information Management.)

The topics listed on the right under the heading **AI ethics** have arisen as issues in response to proliferation of "intelligent" technologies such as next generation AI (machine learning) as well as other related emerging technologies such as big data, IoT, RPA, and cloud platforms. There is overlap between the issue sets. For example, privacy issues are magnified as increasing quantities and types of data are captured and analyzed (big data analytics) and as AI algorithms (e.g., facial recognition, machine learning) generate new information and allow the interpretation of data that would otherwise be too voluminous to be useful. AI-based automation raises new issues as well, in particular, data driven, automated decision making. Machine learning (ML) algorithms have raised questions of bias and discrimination in their representations of groups and

decisions about individuals. They have also raised questions of due process and transparency, as the ability to understand and challenge decisions is made more difficult by "black box" algorithms. Other issues raised by AI that are either new or magnified include the attribution of risk and responsibility to autonomous systems that depend on data sets, and the impact of automation on employment and the character of work.

So, digital ethics can be understood to include in its scope traditional information ethics issues as well as issues arising from or magnified by AI / intelligent technologies. I exclude robot ethics and machine ethics from digital ethics because of special issues with autonomous systems. Such systems can be conventional robots, but also autonomous vehicles, drones, and weapons. These technologies are a bit outside of the ambit of ecm, as are the ethical problems they pose. That said, capturing records of autonomous systems is within the scope of professional responsibility and technical challenge for information professionals managing digital content, so there is an intersection point between digital ethics as I am defining it and robot ethics. (See my article, AI, Records, and Accountability, ARMA Magazine.)

CONTENT ETHICS: DIGITAL ETHICS IN ECM

Digital ethics and the issues central to it can be understood on a general level and will continue to be addressed by policy and legislation at such a level. But operationalizing digital ethics requires working out organizational policies, practices and procedures in the context of **specific application areas**. Developers, implementers, and end user organizations need to operationalize digital ethics within the context of their technologies and business models. Unstructured content in the form of repositories, knowledge bases, and processes forms such a context. For this reason, I use the label **'content ethics' to denote this specification and operationalization of digital ethics for this context of unstructured content**.

Figure 1 depicts the current state scope of content ethics. Traditional information ethics issues constitute the major focus of ethically managing unstructured content. Managing unstructured content raises the same types of concerns as other information systems (e.g., privacy), with the added complication that its being unstructured makes it more difficult to control and manipulate. While sensitive data may be included in data fields and certain document types may be designated as confidential, much sensitive data exists within the body of documents. This makes identification and control more difficult technically and requires more nuanced judgements and actions.

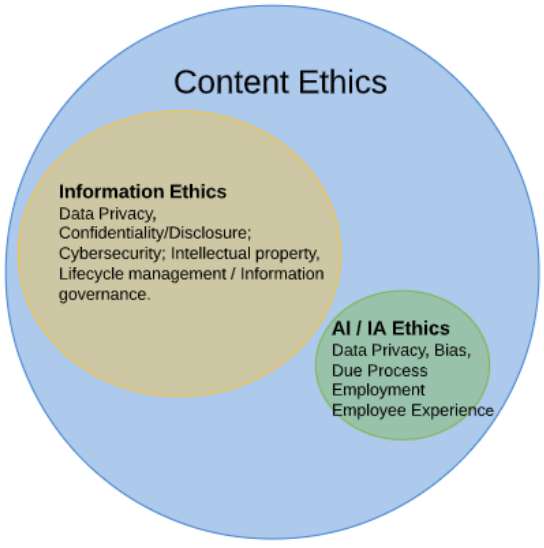

Figure 1 also illustrates that newer ethical concerns driven by AI / intelligent automation (IA) are within the scope of content ethics. These issues include potential job loss and the degradation of employee experience in environments where professional judgement is replaced by automated decision making. The smaller circle in the figure represents the more minor role these issues currently play in content ethics. They have a less central place at this time because the emerging technologies that raise these issues are just beginning to be incorporated into the ever-evolving stack of technologies that comprise ecm. Further, cognitive technologies such as machine learning and RPA are being used in lower risks context such as document content classification, data extraction, and systems integration. (See AIIM's How to Fit Artificial Intelligence into Your Information Management Strategy). Ecm workflow/business process automation continues (at this time) to be rules-based, and therefore (relatively) transparent. Nevertheless, as technologies develop and are incorporated into our stacks, riskier use cases will become more common. Such use cases include mortgage application approval processes, employment applications/hiring, benefits case management solutions, etc. The prominence of issues driven by cognitive technologies can only be expected to grow. The smaller circle can be expected to expand to at least equal the circle representing information ethical issues.

CONTENT ETHICS ROADMAP

Below is a roadmap for content ethics in relation to the current state ecm functional stack. Ecm functional areas will continue to evolve and merge with other technologies. (This roadmap will be periodically updated.) The roadmap identifies functional areas and relevant ethical risks, responsibilities, and opportunities. Subsequent pieces will address specific areas in the roadmap.

FUNCTIONAL AREA	RISK/RESPONSIBILITY/OPPORTUNITY AREAS
Core Repository	
Metadata / database search **Full text search** **File storage management** **Collaboration** **Security**	Data privacy, confidentiality/disclosure issues are central to the management of content repositories. Privacy/confidentiality applies at the content/document type level, metadata level, and to data elements within digital documents. Intellectual property; copyright, trade secrecy also arise for content, especially at the content/document type level. Cybersecurity risks such as data breaches attend stored content in networks and in the cloud. Internal access controls such as "lease privilege" and "minimum necessary" apply within solutions.
Lifecycle Management	
Capture **Records Management** **Retention Management**	Lifecycle management / information governance capabilities align with key responsibilities and opportunities. Policies and functional controls should be applied to what is captured, who has access during ingestion, and quality measure.

	Records management controls, including retention management, are key factors in supporting accountability and minimizing data privacy risks such as unauthorized access and unjustified usage.
Integration and Access	
Systems Integrations **Cloud computing** **Public Access**	Data privacy, confidentiality/disclosure and cybersecurity issues attend to systems integrations. Access rights for systems on both sides of the integration need to be evaluated. Security risks at integration points need to be assessed. Cybersecurity risks for on-promise and network systems, as well as protocols and security controls of cloud service providers are within the scope of ethical and legal assessment. Jurisdictional differences need to be considered. Policies and mechanisms for making document accessible on a self-serve basis to constituents (of public agencies) is a matter of legal and ethical consideration.
Business Process Automation	
Digital forms **Workflow** **Case Management** **Reporting / Visualization**	Data privacy, confidentiality/disclosure issues attend to data and documents accessed and acted on within active business processes based in workflow and case management solutions. Reporting and visualization output should be evaluated where a broader user base is the target audience. Questions of transparency, fairness, bias and due process attend to automated business processes that include decisions and actions that affect the rights of persons. Traditional rules-based workflow platforms provide greater transparency than ML-based systems, but still need to be evaluated for and designed to eliminate bias and promote fairness.
Intelligent Automation	
RPA	RPA technologies provides systems integration functions and therefore inherit data privacy issues relevant to integration. RPA technologies automate routine tasks normally performed by knowledge users. Worker displacement is a risk, as is further routinizing of the workplace and making worker experience less meaningful. Deployment of RPA should consider worker reassignment to roles that take advantage of human judgement. It should also be accompanied by or incorporate richer sets of information that inform and compliment human judgement.
AI/ML Content analytics & semantics **Document classification & PII identification Data Recognition and Extraction**	Machine learning (ML) automation and intelligent automation are used to identify document types (auto-classification), extract data, and improve search/retrieval by surfacing content. Data privacy issues arise for generative information and need to be addressed. As with RPA, AI-driven automation may reduce headcount and hollow out knowledge worker experience. Consideration should be given to worker re-assignment to roles requiring professional judgement and to creating information rich environments that enhance and complement human knowledge work. As AI/ML is incorporated into automated business processes impacting data subjects' fundamental interests (financial, juridical, governmental-administrative), issues of bias, due process and fairness become salient. Systems should be designed to mitigate these risks and enhance decision-making. Such measures include design that puts humans in the loop and providing better, relevant information to systems and humans decision makers. Systems should also be designed to provide reports explaining/justifying decisions, and should provide degrees of system autonomy in line with risk. Low risk decisions can be fully

	automated, while for higher risk decisions recommendations should be generated for human decision makers.