ESI 관점에서의 e-Discovery

김영수* 신상욱** 홍도원***

2006 년 12 월 1 일에 발효된 미연방민사소송법(FRCP)을 통해 민사 소송의 디스커버리(discovery; 증거 공개 요구 절차) 증거물 범위가 기존 종이 문서에서 전자 문서(ESI)로 확대되면서, 최근 몇 년간 e-Discovery 가 지속적으로 이슈화되어 왔다. 여러 건의 소송에 상시 시달리고 있는 기업들은 증거 제출 의무 불이행으로 인한 소송의 패소 및 기업 이미지 실추를 막고, 관련 ESI 관리 및 저장 비용 증가에 대처하기 위해 자동화된 e-Discovery 관련 솔루션을 도입하고 있는 추세이다. 그러나, 이러한 자동화된 솔루션이 완벽한 해결책이 될 수 없고 e-Discovery 의 특성상 법과 기술 두 가지 모두에 대한 고려가 이루어져야 한다는 판단 하에, 본 고에서는 e-Discovery 의 주 대상인 ESI 에 대한 개요와 e-Discovery 과정이 진행되면서 고려해야 할 ESI 관련 내용들을 정리하였다.

목 차

* ETRI 암호기술연구팀/선임연구원
** 부경대학교 IT 융합응용공학과/부교수
*** ETRI 암호기술연구팀/팀장

Ⅰ. 서 론

미국의 민사 소송은 디스커버리(discovery)라는 "증거 공개 요구 절차"를 통해 상대방에게 증거와 정보의 공개를 요구하는 절차를 갖게 된다. 소송 당사자가 공판 전에 공판 준비를 위해 법정 외에서 법정의 방법으로 소송의 쟁점을 명확히 하고자 정보 및 증거물을 공개하고 수집하는 것으로, 소송 당사자가 서로 상대방이 보유한 증거물, 서류, 증인 등을 공개하도록 요청함으로써 서로 대등한 조건 하에서 소송을 진행할 수 있게 된다. 민사 소송 당사자는 공판 절차가 진행되기 이전에 스스로 자신이 보유한 증거를 공개함과 아울러 상대방 당사자나 제삼자에게 증거 개시 요구를 할 수 있다. 이러한 증거 공개 요구 절차는 쟁점을 명확히 하고 재판이 시작된 후 은폐될 지 모르는 증거물을 확보하는데 목적이 있는

것으로, 서로 상대방의 증거물을 자세히 알게 되므로 공판까지 가기 전에 이 과정에서 타협이 이루어지는 경우가 많이 있다. 디스커버리는 요청서, 답변서, 항의서 등의 서면으로 이루어지며 각 문서마다 변호사가 서명을 해야 한다. 이 과정은 법원 개입 없이 소송 당사자 간에 이루어지지만, 한쪽의 요청을 상대방이 거부하여 분쟁이 생기면 법원이 개입한다. 한쪽이 고의적으로 비용과 시간이 과중하게 드는 요청을 할 경우 법원이 취소 명령을 내릴 수 있으며, 반대로 한쪽이 디스커버리 의무를 다하지 않을 경우, 강제 명령을 내릴 수 있다. 2006 년 12 월 1 일 발효된 미 연방민사소송법에서 디스커버리의 대상이 되는 증거물의 범위에 ESI 가 포함되면서 e-Discovery 라는 용어가 등장하게 되었다[1].

여러 건의 소송에 상시 시달리고 있는 기업들은 증거 제출 의무 불이행으로 인한 소송의 패소 및 기업 이미지 실추를 막고, e-Discovery 대비를 위한 관리 비용 및 저장 비용 증가에 대처하기 위해 ESI 에 대한 체계적 관리 및 대응 시스템 도입을 서두르고 있다. 이러한 현실은 미국 내 기업뿐 아니라 뉴욕 증시에 상장되어 있는 국내 기업들에게도 영향을 미치게 되며, 한미 FTA 등 글로벌호가 확산될 경우 국내 다른 기업들도 이러한 법안에 대해 자유로울 수는 없을 것이다. 포춘지 선정 500 개 미국 기업들이 평균 146 개의 각종 소송에 상시적으로 시달리고 있는 현실을 감안할 때, 국내 기업들도 언제라도 위기에 내몰릴 수 있음을 인지하고 각종 소송에 대비하여 e-메일 등과 같은 전자 문서를 관리하고 정보보호 대책을 마련하는 등의 시급한 조치가 필요한 것이 현실이다.

본 고에서는 e-Discovery 의 주 대상인 ESI 에 대한 개요와 e-Discovery 과정이 진행되면서 고려해야 할 ESI 관련 내용들을 정리해 보았다. 우선 ESI 의 정의, 형태 및 특징을 정리하여 기술한 후에 EDRM 에 기반한 e-Discovery 절차에 따라 ESI 관점에서 중요한 고려 사항들을 각각 구분하여 기술한 후 8 장에서 결론을 맺는다.

II. ESI 개요

ESI(Electronically Stored Information)는 컴퓨터 하드웨어나 소프트웨어에 사용하기 위해 디지털 형태로 생성, 가공, 통신, 저장 및 사용되는 정보를 일컫는다[2]. ESI 는 컴퓨터나 컴퓨터-기반 디바이스를 통해 생성되고 또한 컴퓨터나 컴퓨터-기반 디바이스를 통해 검색 가능한 형태로 저장된다[3]. ESI 는 e-메일, 메신저, 웹 페이지, 워드 프로세싱 파일, 스프레드 쉬트, 데이터베이스, 캘린더, 디지털 팩스, 애니메이션, 비디오, 오디오 등 다양한 형태를 갖지만 일반적으로 크게 세 가지 형태로 존재한다.

- Native 파일: 소프트웨어 애플리케이션을 통해 생성된 문서나 파일 그 자체 형태를 의미하는 것으로 스프레드 쉬트나 워드 프로세싱 문서를 예로 들 수 있다. 고유 파일은 사용자가 볼 수 있는 텍스트나 스프레드 쉬트 번호 같은 내용뿐 아니라 작성자나 생성 날짜 같은 사용자에게 보이지 않는 문서 정보(메타데이터)를 모두 담고 있다.
- 메타데이터(Metadata): 생성 날짜, 작성자, 수정 날짜, 파일 크기, 파일 타입 등 문서의 다양한 특성을 나타내주는 정보를 일컫는 것으로 ESI 파악을 위한 메인 소스가 된다.
- 이미지 파일(Image File): 문서에 대한 신속한 리뷰 기능 제공을 위해 표준화된 이미지 형태로 변형된 TIFF(Tagged Image File Format)나 PDF(Portable Document Format)를 예로 들 수 있다[4]. 이러한 이미지 파일은 표준화되고 프린터가 가능하지만 수정은 불가능한 특징을 갖는다. 종이로 된 문서는 그 형태상 ESI 가 아니지만, 이를 디지털이나 이미지 포맷으로 변환하게 되면 ESI 로 간주한다. 종이로 된 문서는 OCR(Optical Character Recognition) 소프트웨어를 통해 검색이 가능한 디지털 형태로 변환되거나 TIFF 나 PDF 같은 이미지로 변환될 수 있다.

ESI 는 종이로 된 문서와 비교했을 때 다음과 같은 특징을 갖는다.

- 기하급수적인 양적 증가: ESI 는 종이 문서에 비해 생성하거나 배포하는 방법이 쉽기 때문에 그 양이 급속도로 증가하게 된다. 개인당 생성되는 ESI 의 양은 대략 1 년에 800MB 정도로 추산하고 있다.
- 높은 복잡도: ESI 는 메타데이터가 내포되어 있기 때문에 종이 문서보다 많은 문서 보존 정보를 제공한다. 메타데이터는 보이지 않는 문서의 히스토리로서 문서의 제목, 주제, 작성자, 저장된 위치, 문서 생성 시간, 액세스 시간, 수정 시간, 인쇄 시간, 주석, 개정 번호, 총 에디팅 시간, 생성하기 위해 사용된 템플릿 등을 포함한다. 문서에 내장된 메타데이터를 찾고 보는 것은 데이터 마이닝 관점에서 매우 중요하며, 많은 e-Discovery 분쟁들이 이러한 메타데이터로 인해 발생되므로 문서 관리 시 손상되지 않은 메타데이터를 포함할 수 있도록 문서를 주의하여 관리해야 한다[5].
- 보존의 어려움: ESI 는 종이 문서에 비해 수정하기가 더 용이하다. ESI 는 저장 매체에만 존재할 수 있고 그 저장 매체가 덮어씌워지거나, 손상되거나 또는 읽을 수 없게 될 수 있으므로 ESI 를 보다 적극적으로 보존해야 한다. ESI 를 보존하기 위한 대비가 없다면 파괴되거나 수정될 것이다.
- 삭제의 어려움: ESI 는 손상되기 쉽지만 종이 문서에 비해 폐기하기가 어렵다. ESI 가 서버

에 저장되거나 백업 또는 e-메일로 전송되었다면, 문서 삭제는 그 효력이 없게 된다. 또한 데이터 손실을 방지하기 위한 소프트웨어에 탑재된 자동 복구나 자동 저장 기능은 주기적으로 현재 사용중인 문서의 백업 사본을 자동적으로 생성하므로 ESI의 삭제를 더욱 어렵게 만든다.

- 소프트웨어나 하드웨어에 대한 의존성: 파일을 정확하게 열어서 보기 위해서는 특정 소프트웨어가 필요하며, 원래의 소프트웨어 환경으로부터 떨어져 나왔을 때 데이터는 읽을 수 없게 되므로 그 의미를 상실하게 된다.

III. ESI 저장 및 관리

정보 관리(Information Management) 단계는 회사나 조직 내의 ESI 관리/보존 정책을 통해 특정 ESI를 유지하고 관리하는 단계이다[6]. 본 장에서는 데이터 접근성 정도에 따른 ESI 저장 유형과 e-Discovery 관점에서의 ESI 보유 방법에 대해 살펴본다.

ESI는 시간이 지남에 따라 저장 및 보관해야 하는 양이 급속도로 증가하는 반면 접근 빈도수는 점점 적어지므로, 저장하는 방법과 장소가 바뀌게 된다. 컴퓨터 데이터 저장소(Storage), 간단히 저장소 또는 메모리는 차후에 조회하기 위해 일정시간 동안 ESI를 보관하는 장치와 기록 매체를 의미한다. E-Discovery 관점으로 보자면 컴퓨터 데이터 저장소에 대한 접근성이 감소함에 따라 발견, 복구, 읽기 가능 형태로 가져오는 비용이 빠르게 증가하게 된다. 법원이 대부분의 기업에 의해 사용되는 데이터를 5가지로 분류하고 접근성이 감소되는 순서로 이를 열거한 판례가 존재한다.

① 액티브, 온라인 데이터: 액티브 단계는 ESI가 생성, 수신, 처리되는 단계 또는 빠르고 빈번하게 액세스되어야 하는 단계로, 온라인 데이터에 대한 저장소는 하드드라이브와 같은 마그네틱 디스크를 포함한다.

② Near-line 데이터: 이 범주에 해당하는 데이터에 대한 저장소는 자동화된 저장 시스템으로 탈부착이 가능한 광학 디스크를 예로 들 수 있다. 액세스 속도는 수 msec 에서 2 분 정도의 범위에 있다[7].

③ 오프라인 저장소: 마그네틱 테이프나 광학 디스크가 이 범주에 해당한다. 저장 매체 라벨이 붙고 선반이나 랙으로 구조화되고 수동으로 액세스된다는 점에서 앞의 두 범주와는 다르다. 오프라인 저장소는 재난 복구나 액세스되지 않을 것으로 예상되는 기록물의 저장을 위해 사용되며, 어느 경우든 검색은 가능하지 않다. 재난 복구가 필요할 시에는 전체 테이

프나 디스크가 로드된다.

④ 백업 테이프(일반적으로 데이터 압축 사용): 순차적 액세스 매체로 데이터는 구조화되지 않으며, 데이터 조회를 위해서는 전체 테이프에 대한 내용 복원이 이루어져야 한다. 디스크에 추가적으로 데이터를 넣기 위해서는 사용된 압축을 해제해야 한다. e-Discovery 시에 소송 상대방에게 백업 테이프에 담긴 ESI 의 제출을 요구하기 위해서는, 법원에 해당 ESI 와 사건의 관련성이 ESI 조회 및 처리 비용보다 더 크다는 증거를 제시해야 한다. 이에 법원은 타당성을 검토하고, 이러한 검색이 업무와 정보 관리 활동을 방해하는 정도를 고려하여 최종 결정을 하게 된다.

⑤ 삭제(Erased), 파편화(Fragmented), 손상(Corrupted)된 데이터: 삭제된 파일의 경우 새로운 파일에 의해 덮어씌워지지 않았다면 사라지지 않는다. 파편화된 파일은 잘게 잘려서 연속적이지 않은 분리된 영역에 저장된 경우이고, 손상된 파일은 컴퓨터 바이러스, 하드웨어, 소프트웨어 오동작에 의해 파손된 파일을 의미한다.

위의 5 가지 데이터 소스는 2 원(Two-tier) 디스커버리를 위한 기초를 형성하는 것으로 두 개의 그룹으로 구분되는데 이는 FRCP Rule 26(b)(2)에 정의되어 있다. 2 원 디스커버리는 합리적으로 액세스 가능한 소스와 과도한 비용 부담 때문에 합리적으로 액세스 가능하지 않은 소스에서 발견되는 ESI 들을 구분한다.

- 1 원(First-tier): 처음 3 가지 데이터 소스(액티브, near-line, 오프라인)로 구성되며, 합리적으로 액세스 가능한 소스로 정의된다.
- 2 원(Second-tier): 마지막 2 개의 데이터 소스(백업 테이프와 삭제, 파편화, 손상된 데이터)로 구성되며, 합리적으로 액세스가 가능하지 않는 소스로 정의된다.

IV. ESI 식별

식별(Identification) 과정은 보존의 의무가 있는 ESI 나 소송 발생 시 필요한 모든 관련 정보의 위치를 확인하는 단계로 디스커버리에 사용해야 하는 ESI 의 범위를 결정하게 된다. e-Discovery 의 범위는 세가지 요소로 나타낼 수 있다[8]. 식별 과정 동안 일반적으로 이들 3 가지 범위 내에서 ESI 를 판단하여 식별한다. 관련된 ESI 의 범위는 재판까지 소송의 전체 과정 동안에 유동적으로 변경될 수 있고, 이에 따라서 식별할 ESI 도 달라지게 된다.

- 데이터 관리자(Data Custodian) 및 키 플레이어: 대상 ESI 들에 대한 개별 소유자들 또는 관리자들의 범위를 파악하는 것은 ESI 식별을 위해 가장 중요한 것 중 하나이다. 데이터

관리자는 ESI 가 어떻게 유지되는지, 어디에 유지되는지, 어떻게 액세스되는지를 알고 있는 사람들이고, 키 플레이어는 소송의 요인 또는 ESI 에 어느 정도 관계가 있는 사람들이다. 사업의 의뢰인, 고객, 다른 외부인이 키 플레이어가 될 수도 있다. 증인이 될 수 있는 사람들은 직접 만나보거나 또는 화상 회의, 전화 통화와 같은 방법으로 인터뷰하여 파악해야 한다. 인터뷰는 인터넷 또는 문서의 형태로 설문 조사와 같은 방법으로 수행될 수도 있다.

- 해당 ESI 위치: 합리적 또는 잠재적인 소송에 관련된 내용으로 일반 PC 뿐 아니라 모바일 장치의 텍스트, 주소록, 일정 등을 포함할 수 있다. 잠재적으로 대응할 ESI 의 위치, 활용성, 접근성, 형태 등을 식별할 필요가 있다.
- 타임 프레임: 조사에 관계가 있는 시간 범위를 정의하는 시작 날짜와 종료 날짜로, 정확한 산출이 어렵지만 식별에 있어서 매우 중요한 범위이다.

위의 3 차원 범위를 통해 범위 내에 있는 것을 식별할 뿐만 아니라, 범위 밖에 있는 ESI 를 배제시키는 제한 요인들도 식별할 수 있다. 3 가지 범위를 통한 ESI 식별을 위해서는 다음 단계들을 따라야 한다.

① ESI 데이터 관리자 및 소유자에 대한 초기 또는 조기 소송 평가(Early Case Assessment): 소송의 요인에 기반하여 가장 대응될 가능성이 있는 ESI 의 제어 또는 관리 권한을 누가 가지는지 대개 파악할 수 있으므로 업무 기능, 작업 프로젝트, 부서, 사무실, 지리적 위치, 다른 범주의 조합에 의해 핵심 인물을 식별한다[9]. 또는 이름에 의해 사람을 식별할 충분한 정보를 가질 수도 있다. 소수의 데이터 관리자들이 대량의 관련 ESI 를 가지는 경우가 많은데, 일반적인 가이드라인을 참조하면, 5~10 명의 데이터 관리자가 문서의 80% 정도를 가지게 된다고 한다. 만일 데이터 관리자가 다수라면, 높은 비율의 관련 ESI 를 가진 관리자들인 대량 데이터 관리자(Major Data Custodian) 그룹과 낮은 비율의 관련 ESI 를 가진 관리자들인 소량 데이터 관리자(Minor Data Custodian) 그룹으로 구분된다.

② 관련된 ESI 가 어떤 것이고 어디에 위치하는지에 대한 광범위한 초기 평가: 실적과 금융 기록, 계약서와 계약서 초안, 구매 주문서, 문자 메시지, e-메일 등의 소송에 적용되는 모든 유형의 ESI 를 고려한다. 차후에 좀더 완전하게 정렬하기 위해 ESI 들을 넣어두는 일종의 분류 폴더 개념을 고려한다. 유용한 ESI 폴더는 다음의 세 종류로 구분된다.

* 확실한 범위 내에 있는 ESI 를 담는 폴더: 시간 범위 내에 있고 소송의 요인과 명확히 관련된 ESI 는 이 폴더에 담는다. 기업의 정보 처리 애플리케이션이나 데이터베이스가 업그레이드되었다면 날짜를 처리하는 것이 매우 복잡해진다. 산출할 필요가 없는 ESI 를 수집, 검토, 산출하지 않고 관련된 ESI 만을 산출하기 위해서는 날짜가 정확하다는

것을 확인해야 한다. MAC 시간이 확인되어야 한다면, 컴퓨터 포렌식 전문가의 도움을 필요로 한다[10]. MAC 시간이나 MAC 데이터는 잘못 될 수 있고, 이 때문에 상대방이 ESI 의 신빙성에 이의를 제기하거나 판사가 증거로 허용하지 않을 수도 있으므로 컴퓨터 포렌식 기술을 활용하여 이를 정확히 하는 것은 큰 의미가 있다.

* 범위 내에 있을 수도 있는 ESI 를 담는 폴더: ESI 를 위한 임시 폴더와 같은 것으로, 산출할 필요가 있을 수도 있다고 생각되면 ESI 를 이 폴더에 넣는다. 확실히 알지 못한다면, 지나치다 싶을 정도로 주의해야 한다. 소송의 범위가 확대되거나 축소될 수도 있고, 이 폴더에서의 ESI 가 다른 두 폴더 중 하나로 옮겨질 수도 있다.
* 범위 밖에 있는 ESI 를 담는 폴더: 디폴트로 또는 의도적으로 이 폴더에 옮겨진 것은 관련되지 않았다는 것을 나중에 변호해야 할 수도 있는 ESI 이다. 이 ESI 는 법적 유지 대상이 아니다.

③ 사건의 타임 프레임을 확인하거나 연관성 제시: 특정한 시작 또는 종료 날짜가 명시되지 않을 수 있고, 날짜가 열거되더라도 사람들이 실수하거나 진실을 왜곡할 수 있으므로 그 유효성을 검증해야 한다.

고전적인 ESI 저장 공간인 하드드라이브 외에도 최근 스마트 환경이 구축되고 소셜 네트워킹이 활성화되면서 네트워크 드라이브, 스마트 디바이스, 소셜 네트워크, 소셜 미디어 사이트, 협업 플랫폼 등 도처에 ESI 가 존재하게 되었다. 이러한 잠재적인 ESI 들은 e-Discovery 관점에서 매우 가치가 높다고 할 수 있다[11]. 데이터의 소유자나 독점자 없이 누구나 손쉽게 데이터를 생산하고 인터넷에서 공유할 수 있도록 한 사용자 참여 중심의 인터넷 환경인 웹 2.0 은 여러 가지 이유로 e-Discovery 관점에서 매우 가치가 있다[12].

- 많은 콘텐츠들이 무기한의 시간 동안 아카이빙된다.
- 직원들은 자신들의 업무 범위 내에서 다양한 형태로 이를 사용한다.
- 소송에 관련된 포스트와 파일들이 Web 2.0 공간에서만 존재할 수 있다.

트위터, 페이스북, 마이스페이스와 같은 소셜 네트워킹 사이트에서 사용자의 활동은 e-Discovery 관점에서 매우 의미가 있고 litigation-hold 의 대상일 수 있기 때문에 기업에서 중요한 관심의 대상이다[13]. 사전 소송 준비 목적으로 이러한 잠재적 ESI 들을 식별하기 위해서는 다음과 같은 준비 작업이 필요하다.

- 목록 로그 작성: 기업, 부서, 팀, 직원에 의해 사용되는 미디어와 네트워크를 파악하고, 이름, 유형, 날짜, 사용 이유에 대한 로그를 기록해 둔다.

- 정보 흐름 문서화: 누가 콘텐츠를 저장하는지, 그 데이터 파괴 정책이 무엇인지, 콘텐츠의 인증된 사본을 어떻게 획득하는지 등을 파악하여 문서화한다.
- 소셜 네트워크: 개인 또는 기업이 제한된 시스템에서 공개적으로 또는 준-공개적으로 프로파일을 구성하는 것을 허용하는 웹 기반 서비스로 블로그, 오프라인으로 보거나 듣기 위해 다운로드된 오디오/비디오 콘텐츠인 podcasts, 디지털 장치에 자동적으로 업데이트된 정보를 전송하는 RSS(Real Simple Syndication), 그리고 협업 도구로 사용되는 웹 사이트인 위키(Wikis) 등을 예로 들 수 있다.
- 소셜 미디어 사이트와 플랫폼: 소셜 미디어는 소셜 웹 사이트에서 원하는 사람에 의해 생성 및 공개된 콘텐츠를 의미하는 것으로, 콘텐츠가 온라인으로 장기간 남아있기 때문에 다른 형태의 사용자 생성 콘텐츠에 비해 쉽게 파악될 수 있다. 대표적인 소셜 미디어 사이트로는 유튜브(Youtube), 페이스북(Facebook), 마이스페이스(Myspace), Linkedin, Yelp, 그리고 트위터(Twitter) 등이 있다. 점점 많은 기업들이 내부 블로그와 위키를 가능하게 하는 Jive 또는 SocialText 와 같은 소셜 미디어 플랫폼을 사용하고 있는데, 이러한 소셜 미디어 네트워크 또는 플랫폼을 사용하고 있다면, litigation-hold 기간 동안 데이터가 보존, 검색 및 산출될 수 있다는 것을 보장하기 위한 단계들을 반드시 취해야 한다.
- 협업 플랫폼(Collaboration Platforms): 소셜 미디어는 소셜 웹 사이트에서 원하는 사람에 의해 생성 및 공개된 콘텐츠를 말하는 것으로, 이러한 플랫폼은 파일 저장을 위해 하드드라이브 대신 웹을 사용한다. 가장 널리 사용되는 웹 기반 협업 도구로는 구글 Docs 와 마이크로소프트의 MS 오피스 Live Workspace 등이 있다.

V. ESI 보존

산출할 필요가 있는 ESI 에 대해 식별하고 나면 이러한 ESI 가 우연히 또는 고의로 삭제 및 변경되지 않도록 보존(Preservation) 과정을 거쳐야 하고 때로는 해당 ESI 들을 수집(Collection) 해야 한다. 여기서는 보존에서 가장 중요하다고 할 수 있는 ESI 에 대한 litigation-hold 에 대해 살펴본다.

Litigation-hold 는 소송 또는 정부 조사에 관련된 종이 문서 또는 ESI 의 파괴를 방지하기 위해 기업에 의해 적극적으로 행해지는 보존 유지 행위로서 목적은 잠재적인 연관성이 있는 ESI 에 대하여 특정 기간 동안 안전하게 보존된다는 것을 보장하는 것이다. Litigation-hold 의 실패는 곧 고의적 증거 폐기인 문서 파기(spoliation)로 갈 가능성이 크므로 매우 심각한 문제가

될 수 있다. 보존 방법이나 보존 매체가 합리적이라면 어떤 방법으로 문서를 보존할지는 문서 관리자의 재량이기는 하나, 액세스하기 힘든 백업 테이프에 담는다거나 검색을 용이하게 하는 인덱스 없이 저장하여 ESI 를 보존한다면, 차후에 ESI 가 합리적으로 액세스될 수 없다라고 법원에 얘기할 수 없게 되므로 반드시 액세스를 가능하게 유지해야 한다.

- Litigation-hold 의 시작. 다음과 같은 일이 발생할 경우 litigation-hold 를 시행해야 할 필요성을 느끼게 된다.
  * 사실상 "우리는 당신을 고소할 것이다"라고 얘기하는 상대 변호사 또는 정보 기관으로부터의 보존 편지를 받았을 경우
  * "우리는 당신에 대한 고소를 진행 중이다"라는 말로 고소가 제기된 경우
  * 기업들이 고소당하거나 조사받는 것과 같은 소송과 조사의 위협이 있을 경우
  * 소송으로 번질 수 있는 문제가 있다는 것을 알게 된 경우
- Litigation-hold 의 적용 시기: Litigation-hold 발효를 위한 가장 합당한 시기는 보존 편지를 수신한 날짜이거나 고소가 제기된 날짜이지만, 편지나 고소가 없다면 소송이 예상될 때 이를 시작하는 것이 바람직하다. Litigation-hold 에 관한 문제는 상황의 사실에 기반하므로, litigation-hold 를 왜 발효했고 어떤 조치를 취했는지 반드시 문서화하고 이것이 합리적으로 받아들여질 수 있도록 설득력 있게 주장할 필요가 있다. 발효일 이전에 파괴된 관련 증거들은 당신이 증거를 보존할 필요성을 인식하지 못했기 때문에 피해가 없을 것이나, 발효일 이후에 발생하는 잠재적 증거의 파괴는 결코 허용되지 않는다.
- Litigation-hold 의 지체와 처벌: Litigation-hold 를 뒤늦게 발효했지만 소송에 관련된 ESI 가 손실되지 않았다면 다행한 일이다. 법적 이슈는 상대 당사자가 litigation-hold 의 지연이나 지체에 의해 해를 입었는지 여부이지 hold 의 타이밍이 아니다. 부정한 목적을 가지고 고의로 litigation-hold 를 지연시키게 되면 법적 처벌로 이어질 가능성이 많다. 한 예로, 2008 년 한 여성 변호사 보조원은 그녀의 이전 고용주 및 파트너 등에 대해 성희롱과 해고 조치가 부당하다며 소송을 제기했다. 법률 사무소는 그 파트너가 컴퓨터에서 관련 ESI 를 보존하는 어떠한 행위도 취하지 않고 1 년 이상 컴퓨터 작업을 계속하도록 했다. 그 파트너는 명백한 증거가 될 수 있는 e-메일을 삭제했고 컴퓨터를 계속 사용했기 때문에 삭제된 e-메일이 아마도 덮어씌워져서 복원할 수 없었다고 증언했다. 그러나, 법원은 법률 사무소와 파트너가 파트너의 컴퓨터에서 전자 기록 또는 e-메일을 보존하기 위한 단계들을 취하지 않았다라고 판단했다.
- 인력 이동과 litigation-hold: 기업 내 인원이 변경되면, 퇴직하거나 부서를 이동하는 직원

들의 컴퓨터, 휴대폰, PDA에 포함된 파일들은 재분배된다. 이러한 과정에서 사용하던 디지털 장치에 대한 관리가 제대로 이루어지지 않는다면 litigation-hold가 어려워지는 상황이 발생하게 된다. 그러나, 직원이 승진 또는 해고되었기 때문에 증거의 핵심적인 부분이 손실되었다고 주장하는 것은 법원으로부터 동정을 얻지 못한다. Litigation-hold 절차를 효과적으로 관리할 의무는 직원이 떠났다고 해서 없어지는 것이 아니다.

Ⅵ. ESI 처리와 분석

보존 또는 수집된 ESI들은 처리(Processing), 리뷰(Review), 그리고 분석(Analysis) 과정을 거치게 된다. 중복되거나 연관성이 없는 ESI들을 제외하고 효과적인 리뷰를 할 수 있는 형태의 포맷으로 변경하는 단계가 처리 과정이고, 리뷰와 분석 과정은 각각 ESI 들을 검토하여 비밀성을 요구하거나 특화(Privileged)된 민감한 문서를 선별해내는 것과 사건의 주제, 주요 인물, 중요한 문서 등 ESI에 대한 요약 정보를 작성하는 것이다. 리뷰와 분석은 순차적으로 이루어지기보다는 서로 상호 보완적으로 반복적인 피드백을 거쳐 진행된다. 본 장에서는 ESI에 대한 상세한 처리, 필터링 및 검토 과정에 대해 살펴본다. 처리, 필터링, 검토 단계에서, 아래의 단계들을 수행한다.

① 원하는 결과와 이를 달성하기 위해 필요한 것을 정의한다. 계획에 의해 시작하고, 수행될 필요가 있는 것과 대안을 결정하고, 그리고 어떻게 진행할 것인지를 결정한다.

② 초기 소송 평가를 수행한다. 소송의 요소들, ESI의 유형, 강점, 쌍방 모두의 의도와 열의의 정도 등을 현실적으로 살펴본다. 초기 소송 평가는 소송과 관련하여 기업이 처한 상황을 결정하도록 하며, 이를 위해 수집한 정보는 meet-and-confer session 준비를 위해 변호사가 필요로 하는 중요한 소스가 된다. 정확한 초기 소송 평가는 소송 전략과 소송을 계속 진행할 것인지 타협할 것인지를 결정하는데 직접적으로 영향을 줄 수 있다. 많은 e-discovery 업체들은 초기 소송 평가를 위한 다양한 지원 서비스와 자동화 도구들을 제공한다[14]-[16]. 이러한 업체들은 자사의 e-Discovery 관련 솔루션들을 통해 문서가 어디에 저장되어 있는지, 어떤 포맷으로 저장되어 있는지, 수정하지 않는 방법으로 어떻게 조회될 수 있는지 등을 파악할 수 있고, 다양한 포맷의 e-메일을 포함한 ESI에 대한 분석 기능이 있으며, 수집된 ESI를 변호사가 검토하고 산출하는 포맷으로 변환하는 기능도 제공한다.

③ 보존된 ESI 를 처리하고 인덱싱한다. 보존된 ESI 는 검색 가능한 상태가 아니므로 검색

가능한 포맷으로 처리될 필요가 있다. 처리는 기업 내부 기술자나 제 3 의 업체에 의해 특별한 소프트웨어를 사용하여 이루어져야 하며, 법적으로 변호 가능한 절차로 진행되어야 한다. 모든 ESI 가 처리될 필요는 없다. 소송의 유형, 데이터 소스, 데이터 관리인의 수 등을 통해 처리가 필요한 ESI 를 결정한다. 처리할 ESI 를 선별할 때에는 소송에 이로운 잠재적인 핵심 증거를 배제하지 않도록 유의해야 한다.

④ 처리된 ESI 를 필터링한다. ESI 는 키워드, 날짜, 관리인, 시간 프레임을 포함한 다양한 검색과 내용 필터링(Conceptual Filtering), 중복 제거(Deduping), 유사 중복(Near-duplicate) 처리 등을 거쳐서 관련된 대응 ESI 와 그렇지 않은 것으로 분리된다. 소프트웨어는 RDS (Reference Data Set)을 통해 ESI 를 필터링할 수도 있다[17].

⑤ 필터링된 ESI 를 검토하고 분석한다. 검토는 irrelevant, privileged, work product 등 공시로부터 보호되는 다른 ESI 들을 식별하기 위한 ESI 의 미세조정 필터링이라 볼 수 있다. 첫 번째 단계 검토를 위해 태깅(Tagging) 작업을 거친 후 실제 리뷰에 들어가게 된다. ESI 에 대한 태깅에는 요구된 소송 범주 셋과 일치하거나 하나 이상의 필터에 의해 선택된 ESI 를 표시하는 Responsive, 중요하지 않거나 관련되지 않는 ESI 를 표시하는 Irrelevant, 변호사-고객 특권에 의해 공개에 대해 보호되기 때문에 제공되는 것이 요구되지 않는 ESI 들에 붙는 Privileged, 개인적인 내용이지만 e-Discovery 의 대상일 수도 있는 Private, 변호사가 소송 당사자를 변호하는 동안 작성한 서면 기록, 소송 당사자나 증인과의 대화 기록, 조사 및 기밀 문서에 붙이는 보호특례(Work Product), 수정(편집)된 부분은 기밀로 보호된 상태를 유지하도록 문서로부터 제거되거나 알 수 없게 된 텍스트를 일컫는 Redact, 훼손되거나 열 수 없게 된 파일들에 붙이는 Corrupted, 암호 키를 사용하여 복호화될 때까지 내용이 숨겨지고 읽을 수 없게 된 파일에 태깅하는 Encrypted, 내용을 보기 위해 패스워드를 필요로 하는 파일인 Password Protected, 소송의 결과에 영향을 줄 수 있는 관련된 내용을 포함하는 문서, e-메일, 텍스트 메시지 등에 대한 속어로서, 결정적 증거가 되는 문서 또는 e-메일을 일컫는 Hot, 그리고 추가적인 주의와 관심이 필요한 파일들인 추가 검토(For Further Review) 등이 있다. 태깅 작업 후, 실제 리뷰에서는 고려해야 하는 사항들로는, 문서 또는 e-메일 메시지와 첨부 파일 간의 연관성을 보존하는 것, 중복 파일을 연결시켜 일관성 있게 태깅을 하는 것, 고유 파일 포맷을 유지함으로써 메타 데이터를 보존하는 것, 그리고 공시를 위한 ESI 를 준비하기 전에 이를 위한 태깅과 편집(교정, Redaction)이 일관성이 있다는 것을 검증하는 것 등을 들 수 있다.

VII. ESI 산출

E-Discovery 의 최종 단계로서 검토와 분석이 끝난 ESI 를 쌍방간에 협의된(주로 meet-and-confer 세션에서 논의됨) 포맷으로 산출(Production)하고 제출(Presentation)한다. 본 장에서는 비교적 순조롭고 효율적으로 산출과 제출을 달성하기 위해 최종 e-Discovery 단계에 관하여 기술한다. 산출은 보존하고 처리된 후 많은 필터들과 철저한 특권 검토를 통해 살아 남은 ESI 로 시작한다. 산출될 관련 ESI 데이터 셋을 남겨두고, 관련이 없거나 규칙 또는 법원 명령에 의해 보호되는 모든 ESI 는 필터링 된다. 디지털 이전 시대에 산출은 박스로 종이 문서들을 포장하여 요청 당사자에게 전달하는 것으로 이루어졌지만, e-Discovery 로 인해 다음 두 가지의 기본적인 산출 방법을 가진다.

- 인계(Deliver): ESI 를 CD, DVD 와 같은 물리적 마그네틱 매체 또는 휴대용 하드드라이브로 인계한다. 종이와 같이 ESI 는 포장되어 요청 당사자에게 전달된다.
- 공개(Release): 대응 당사자에게 ESI 를 활용할 수 있게 하는 것으로 ESI 는 서버와 같은 소송 서비스 플랫폼에 저장될 수 있다. 인가된 사용자는 VPN 등 안전한 방법을 통해 서버에 액세스할 수 있다.

공개 방법을 사용한다면 ESI 에 액세스를 보호하고 제공하기 위해 사용하는 여러 가지 단계들이 있다.

- 체계적인 보안 정책 및 프로토콜을 구현한다. 악성 소프트웨어 및 바이러스 등 악의적 공격으로부터 ESI 가 손상되거나 파괴되는 것을 막을 필요가 있다.
- 쌍방 간에 ESI 를 저장하기 위한 비용과 안전한 액세스를 공유한다. 산출된 데이터 셋에 안전하고 개별적인 로그인 액세스를 가지고, 수신 당사자는 그들이 원하는 문서와 그들이 문서를 수신할 포맷을 검토하고 지명할 수 있다.
- 액세스 로그로 액세스와 활동을 추적한다. 모든 네트워크는 설정에 따라 액세스와 활동 내역을 기록하는 로깅 기능을 가지고 있다. 로그는 액세스나 해제가 문제시되면 필요로 하는 네트워크 메타데이터와 같다. 예를 들면, 한 쪽에서 공개했다고 주장하는 콘텐츠에 대해서 상대방이 이를 부인하고 자신이 액세스했다고 주장한다면, 이러한 로그를 통해서 상대방의 주장이 거짓이라는 것을 증명할 수 있다.
- 변호 가능성을 보장하기 위해 각 파일에 대해 산출 히스토리를 추적한다. 데이터 셋이 공개됨에 따라 산출 결과가 로깅되는데, 이들 로그에 대한 공개 타이밍이 분쟁의 이슈가 된다면 입증 근거로 활용할 수 있게 된다.

물리적 인계이든 디지털 공개이든 간에 산출 타이밍은 동시 산출과 규칙적 산출 중의 하나로 행해진다. 동시 산출은 한번에 포장하여 산출하는 것으로, 작은 규모의 ESI 를 수반하는 소송의 경우에는 이 방법이 의미가 있다. 최종 마감 데드라인까지 산출을 시작하지 않는 소송 전략에 관련된 다른 이유들이 있을 수 있다. 규칙적 단계를 통한 산출(Rolling Production)은 설정된 스케줄에 따라 ESI 를 산출하는 것으로, 문서들이 작은 꾸러미로 처리된다면, 이들 문서들이 준비되는 대로 활용 가능하게 할 수 있다.

ESI 에 대한 문서화(Documentation)는 무엇인가를 했는지 또는 하지 않았는지를 증명하라는 법원의 명령을 받았을 경우, 사실 증명을 위해 문서를 남기는 것으로 매우 중요한 작업이다. 매체의 모든 부분을 다루는 각각의 개인 또는 회사에 의해 e-Discovery 의 단계를 이동함에 따라 매체의 모든 부분을 추적해야 한다. EDRM 이 권고하는 산출 히스토리 로그 포함 정보는 전달된 날짜, 연락처를 포함하여 전달된 사람의 정보, 운송업자의 추적 정보를 포함하여 전달에 사용된 수단, 라벨 사본을 포함한 전달된 매체의 세부 사항, 산출된 요소의 식별 정보, 산출된 문서 ID, 전달된 매체의 사본 위치, 산출이 어느 문서 요청에 대한 응답인지의 정보, 그리고 주석 등이다.

로드 파일은 소송 지원 데이터베이스 애플리케이션에 산출물을 업로드 시 이용되는 형식의 파일로, native(예; 엑셀 파일), near-native(예; CSV 같은 콤마로 구분된 엑셀 파일), 또는 near-paper 파일(예; 엑셀 파일의 이미지나 TIFF) 등을 문서 ID 에 연결한다. Summation 의 .dii, IPRO 의 .lfp, Concordance 의 .opt 등 소프트웨어 애플리케이션에 따라 다양한 로드 파일 포맷이 존재한다.

산출 히스토리 로그는 또한 연계 관리(Chain of Custody) 관점에서도 매우 중요하다[18]. e-Discovery 에서의 연계 관리는 ESI 취급자, 컴퓨터 포렌식 전문가, 또는 다른 조사관에 의한 ESI 보존에 관한 것으로, 전자 증거가 적절하게 다루어지고 보존되었는지 그리고 손상될 위험이 없었는지를 기록하는 과정이라 할 수 있다. ESI 가 어디에 저장되었었는지, 누가 ESI 에 액세스 했었는지, 또는 ESI 에 무엇이 행해졌었는지 등을 빠짐없이 기록해 두어야 할 필요가 있다. 소송이 법원으로 가게 되면 조사가 진행되면서 누구도 ESI 를 변경하지 않았다는 것을 변호사가 보일 수 있도록 각 단계를 꼼꼼하게 문서화해야 한다. 문서화된 연계 관리 없이, 그 사실 이후에 ESI 가 변경되지 않았다는 것을 증명하는 것은 불가능하다. 연계 관리는 파일이나 문서 수준에서만 ESI 에 적용되는 것은 아니라, 랩탑, PDA, 휴대폰, 다른 디지털 장치와 같은 하드웨어, 전체 데이터 셋 등 다양한 수준에서 적용될 수가 있다. ESI 의 보존과 수집에 대해 전자적인 연계 관리의 일부로 하드웨어에 행해진 모든 작업들을 문서화해야 한다.

VII. 결 론

본 고에서는 e-Discovery 의 주요 대상이라고 할 수 있는 ESI 에 대한 개요와 e-Discovery 과정이 진행되면서 고려해야 할 ESI 관련 내용들을 정리해 보았다. 우선 ESI 의 정의, 형태 및 특징을 정리하여 기술한 후에, EDRM 에 기반한 e-Discovery 절차에 따라 ESI 관점에서 중요한 고려 사항들을 각각 구분하여 기술하였다. E-메일을 포함한 ESI 관리 솔루션이나 e-Discovery 관련 제품들이 출시되고 있고 이 중 몇몇 제품들은 시장점유율을 높여가고 있는 현실이지만, 기업들이 이러한 시스템 및 제품들을 도입한다고 해서 e-Discovery 에 대한 모든 대비가 이루어진 것은 아니다. EDRM 에서 제시한 e-Discovery 절차는 법적으로 진행해야 하는 과정들을 기술한 것으로, e-Discovery 관련 제품들은 이러한 과정들을 수행하면서 검색 및 분석 등 자동화를 통해 도움을 줄 수 있는 기능들을 제공하는 것이 목적이다. E-Discovery 특성상, 관련 종사자들은 법과 기술 두 가지 모두에 대한 동일한 수준의 지식이 필요하므로, 모든 상세 과정을 두 가지 관점에서 고려하는 것이 e-Discovery 대비를 위한 올바른 방법이라 할 수 있다. 이에 본 고에서 기술한 것과 같이 e-Discovery 의 각 과정을 ESI 관점으로 고려하는 것이 의미가 큰 작업 중의 하나인 것은 분명하다. 향후 e-메일, 메신저, 웹 페이지, 워드 프로세싱 파일, 스프레드 쉬트, 데이터베이스, 캘린더, 디지털 팩스, 애니메이션, 비디오, 오디오 등 다양한 ESI 들에 대하여 각 종류별 분석이 이루어진다면, 이를 통해 e-Discovery 과정에 대한 기술적 이해력이 증가되는 동시에, 기존 e-Discovery 솔루션에 대한 추가적인 요구 사항들을 도출할 수 있을 것으로 기대된다.

<참 고 문 헌>

[1] Federal Rules of Civil Procedure(FRCP), http://www.law.cornell.edu/rules/frcp/

[2] ESI, “Wikipedia”, http://en.wikipedia.org/ wiki/ESI

[3] Adam I. Cohen & G. Edward Kalbaugh, “ESI Handbook: Sources, Technology and Process”, 2010 Edition, Aspen Publishers, 2010. 7. 19.

[4] R.Wiggins, C.Davidson, R.Harnsberger, J.Lauman, and P.Goede, “Image File Formats: Past, Present, and Future”, RadioGraphics, 21, 2001, pp.789-798.

[5] K.Jones, et al, “Real Digital Forensics”, Addison-Wesley, 2006.

[6] Electronic Discovery Reference Model, http://edrm.net/

[7] S.Lau and J.Lui, “Scheduling and Data Layout Policies for a Near-line Multimedia Storage Architecture”, Multimedia Systems, Vol.5, No.5, 1997, pp.310-323.

[8] L.Volonino and I.Redpath, “e-Discovery for Dummies”, Wiley, 2010.
[9] W.Chesher and B.Botta, “Early Case Assessment”, CEIC 2010, 2010.
[10] MAC(Modification, Access, Change/Creation) time, http://en.wikipedia.org/wiki/MAC_times
[11] S.Golder, D.Wilkinson, and B.Buberman, “Rhythms of Social Interaction: Messaging within a Massive Online Network”, The 3rd International Conference on Communities and Technologies, 2007.
[12] J.Kleinberg, “The Convegence of Social and Technological Networks”, Communications of the ACM, Vol.51, No.11, 2008, pp.66-72.
[13] S.Wasserman and K.Faust, “Social Network Analysis: Methods and Applications”, Cambridge University Press, 1993.
[14] B.Babineau, “Leveraging Analytics to Lower e-Discovery Costs: A Study of Clearwell Systems’ Customers”, Clearwell Whitepaper, 2007.
[15] Guidance Software, “Defending Your e-Records Retention Policies with Active Enforcement”, Guidance Software Whitepaper, 2006.
[16] AccessData, “Digital Investigation”, AccessData Presentation Source, 2010.
[17] Reference Data Set, http://www.nsrl.nist.gov/
[18] Chain of Custody, http://en.wikipedia.org/wiki/Chain_of_custody

* 본 내용은 필자의 주관적인 의견이며 NIPA 의 공식적인 입장이 아님을 밝힙니다.