

## 장기보존 전자기록의 위험평가에 관한 연구

송치호<sup>○</sup>, 차현철<sup>\*</sup>

<sup>○</sup>주식회사 제이씨원

<sup>\*</sup>동양대학교 컴퓨터정보전학과

e-mail: chsong@jcone.co.kr<sup>○</sup>, hccha@dyu.ac.kr<sup>\*</sup>

## A Study on the risk evaluation of electronic records for long-term preservation

Chi-Ho Song<sup>○</sup>, Hyun-Chul, Cha<sup>\*</sup>

<sup>○</sup>JC One Co. Ltd.,

<sup>\*</sup>Dept. of Computer Information Warfare, Dong-Yang University

### ● 요약 ●

본 논문에서는 전자기록의 장기보존전략을 선정하기 위해 전자기록 위험평가를 위한 위험요소 평가 방안을 제안한다. 이 방안은 전자기록의 포맷 정보를 기반으로 전자기록의 위험도를 평가하여 향후 장기보존을 위한 전략을 제시하므로 30년 이상의 준 영구 보존이 필요한 장기보존 전자기록에 대한 위험도 평가에 적합하다. 따라서다양한 소프트웨어를 사용하여 생성되는 전자기록을 단지 저장하는 것이 아니라 향후 활용이 가능하도록 전자기록물의 장기보존 전략 수립에 활용이 가능하다.

**키워드:** 파일포맷(File Format), 위험평가(Risk Evaluation), 마이그레이션(Migration)

### I. Introduction

전자기록의 위험을 평가하기 위해서는 우선 파일포맷에 대한 분석이 필요한데 본 연구에서는 전자기록을 파일로 저장하기 위해 사용한 파일포맷에 대한 위험도 평가를 위해 필요한 위험요소 항목을 도출한 후 각 위험요소 항목에 대한 평가 기준을 정립한 후 위험평가를 수행하였다.

### II. Preliminaries

#### 1. Related works

##### 1.1 해외 동향

파일포맷 기반의 위험평가에 대한 연구는 해외에서 주로 연구되고 있으며 <Table 1>은 해외에서 수행한 파일포맷에 대한 유사 연구 사례로서 PRONOM은 다양한 파일포맷에 대한 기술정보를 서비스하고 있으며 AONSII 와 DipRec은 장기보존을 위한 전자기록의 위험평가를 수행하고 있다.

Table 1. 해외 유사 사례 연구

프로젝트	수행	주요 내용
PRONOM	영국 국가기록원	• 파일포맷, 구동S/W 등 기술정보 저장소 구축
AONS II	호주 국립도서관	• 전자기록 장기보존을 위한 평가 지원
DipRec	오스트리아 공과대학	• 시맨틱웹/온톨로지 기반의 지식베이스 시스템
PANIC	호주 퀸즐랜드대학	• 예술, 과학 분야의 뉴미디어 보존방안 제시
SPPT	뉴질랜드 통계청	• 디지털 자료의 장기보존 전략 제시

### III. The Proposed Scheme

본 연구에서는 우선 파일포맷의 위험도 평가를 위해 필요한 위험요소 항목을 도출한 후 각 위험요소 항목에 대한 평가기준을 정립하고 이를 기반으로 의사결정 지원 프로세스를 제시하였다. 그간 전자기록의 장기보존에서 가장 큰 문제로 지적되는 파일포맷은 보석클, 하나워드, 훈민정음, 한글 워드프로세스 등 국내에서만 사용되던 워드프로세서의 개발이 중단되면서 기존 응용프로그램에서 작성된 다수의

전자기록에 대한 재현이 불가능해지면서 발생되어진 영향이 크다.

## 1. 위험평가

### 1.1 위험요소 항목 정의

전자기록에 대한 위험 평가는 전자기록에 적용된 파일포맷의 위험도를 평가하는 것으로 위험도를 판단하기 위해 7개의 기준을 제시하였다. 첫 번째 요소는 파일포맷의 표준 여부이고 두 번째 요소는 파일포맷 정보의 공개 여부이며 세 번째 요소는 파일포맷의 사용 유무이다. 네 번째 요소는 파일포맷의 버전에 대한 평가이고 다섯 번째 요소는 파일포맷을 지원하는 소프트웨어에 대한 종속성이며 여섯 번째 요소는 파일포맷이 저장매체에 종속적인가를 평가하는 것이다.

마지막 일곱 번째 요소는 파일포맷이 운영체제 및 하드웨어 종속적인가를 평가하는 것이다.

### 1.2 위험평가 알고리즘

앞 절에서 제시한 평가 방식을 적용할 수 있도록 학습을 통해 위험요소항목에 대한 지수와 가중치 기반의 전자기록의 위험도 평가를 위한 알고리즘을 <Fig 1>에 제시하였다.

$$RF_i = f(S_i) * wS_i + f(I_i) * wI_i + (1 - f(D_i)) * wD_i + f(R_i) * wR_i$$

$$(wS_i + wI_i + wD_i + wR_i = 1)$$

$f(S_i)$ : 표준성 지수,  $wS_i$ : 표준성가중치  
 $f(I_i)$ : 정체성 지수,  $wI_i$ : 정체성가중치  
 $f(D_i)$ : 종속성 지수,  $wD_i$ : 종속성가중치  
 $f(R_i)$ : 재현가능성 지수,  $wR_i$ : 재현가능성가중치

Fig. 1. 위험도 평가 알고리즘

### 1.3 위험평가 결과

본 연구에서 수립한 알고리즘을 이용하여 국내에서 사용 중인 포맷에 대한 위험도 평가를 수행한 결과 국내에서 더 이상 사용되지 않는 GUL(훈민정음) 포맷이나 SWF 포맷은 위험으로 평가되었음을 알 수 있었다. 본 연구에서는 포맷에 대한 위험요소 항목에 대한 조사 및 알고리즘 개발에 목표를 두었으나 향후 알고리즘에 대한 보정 및 학습을 위해서는 과거에 사용되었던 포맷의 기술정보에 대한 체계적인 학습을 통해 알고리즘에 대한 가중치와 요소 지수값을 보정해 가는 과정이 대단히 중요하다고 판단된다.

## IV. Conclusions

현재까지 작성되어 보존중인 전자기록물의 위험 평가를 위해서는 다양한 위험요소 항목에 대한 기술정보 관리가 필요하며 특히 국내에서 개발되어 사용된 포맷에 대한 기술정보는 국가 기관 주도의 포맷 기술정보 관리 체계를 강화할 필요가 있다고 판단되며 위험요소 항목에 주어진 가중치에 대해서도 반복적인 학습을 통해 가중치에 대한 세부 조정이 필요하다.

## Acknowledgment

본 논문은 “2016년 안전행정부 국가기록원의 기록물 보존기술 연구개발 사업”의 일환으로 진행된 “전자기록 기술정보기반 장기보존 의사결정 지원체계에 관한 연구” 과제 결과를 토대로 작성됨

## References

- [1] Adrian Brown, Head of Digital Preservation, “Automatic Format Identification Using PRONOM and DROID”, The National Archives 2006.
- [2] Myongji University, “Research & Development on Application Technology of the Next Generation Infrastructure for Electronic Records Management”, National Archives of Korea (11-1311153-000192- 01) 2011.
- [3] David Pearson, APSR Project Manager, “AONS II : continuing the trend towards preservation software Nirvana”, iPRES2007.
- [4] David Pearson, “AONS II : continuing the trend towards preservation software Nirvana”, Web Archiving and Digital Preservation Branch, National Library of Australia.