

# '20년 제3차 기록관리 정책포럼 개최 결과 보고

## □ 정책포럼 개요

- 주 제 : 기록서비스 확대를 위한 개인정보 문제 공동 해결 방안

· 주제발표 ① 디지털 문서의 개인정보 필터링 및 마스킹 기술: 임석중 (KISTI)  
② 전자기록물 공개재분류를 위한 비공개정보 필터링 및 마스킹 기술 김진아(국가기록원)  
③ 국립중앙도서관의 온라인 서비스와 개인정보보호 최윤경(국립중앙도서관)  
④ 개인정보 가명처리 정책동향: 박윤식(한국인터넷진흥원)  
· 종합토론 이영도(국가기록원) / 오용석(한국인터넷진흥원) / 김순석(한라대학교)

- 일시/장소 : '20. 11. 12. (목) 14:00~17:00 / 행정기록관 2층 세미나실
- 참석 : 총 74명  
(내부) 기록정책부장, 기록서비스 및 개인정보보호관련 업무담당자 등 19명  
(외부) 개인정보보호 관련 업무 및 기록관리 업무담당자 등 55명  
※ 행정기관(12), 교육행정기관(7), 군(軍)행정기관(2), 준정부기관(5), 지방자치단체(9), 기타공공기관(6), 공기업(6), 대학기관(8)

## □ 추진 성과

- 기록서비스 확대를 위한 개인정보보호 필요성 인식 및 처리기술 공유
- 개인정보보호관련 기술개발 현황과 한계점 및 향후 연구방향에 대한 공감대 형성
- 연구세미나 발표자료\* 및 영상콘텐츠\*\* 등 성과물 지속 활용  
\* 정책포럼 자료집(PDF) 및 개최 결과(PDF) 국가기록포털 게시를 통한 정보 공유 확대  
\*\* 정책포럼 발표·토론 현장 영상 국가기록포털 게시
- 세미나 개선 필요사항 발굴 및 만족도 향상을 위한 참여자 만족도 조사\*  
- 12개 문항(객관식 11, 주관식 1) 41명 참여, 만족도 총점 80.09  
\* '20년 행정안전부 성과관리시행계획 반영(목표 80점 대비 약105.13%(84.10) 달성)  
↳(1회) '기록의 날' 기념 학술회의 80.52 / (2회) 제1차 기록관리 연구세미나 94.09 /  
(3회) 제2차 기록관리 정책포럼 81.71 / (4회) 제3차 기록관리 정책포럼 80.09

## □ 주요 논의 내용

- 개인정보 보호법, 정보공개법 등 개인정보 관련 법령의 한계와 개선 필요성
- 현재 개인정보보호 관련 기술 개발 현황 및 향후 개선방향
- 정보의 활용 측면에서 본 개인정보의 범주 및 활용 시 검토 범위

- (국가기록원 이영도 학예연구관) 현재 개인정보 보호법은 살아계신 분들의 개인정보를 다루는데, 일제강점기 강제동원 명부 DB구축·서비스를 담당하면서 명부에 수록된 분들의 생사를 증명할 수 있는 방법의 법적 규정이 필요하다고 느낌. 개인의 사상·신념·건강·유전 정보 등 민감 정보, 고유식별 정보 등 개인정보의 범주와 그에 따른 처리 방법, 그리고 정보공개법상 예외적으로 공개 가능한 개인정보를 어떤 범주에서 어떻게 기술적으로 해결할 지도 논의가 필요함. 첫 번째 발표문 8페이지의 원본파일 삭제 기능 관련, 기록물법상 원본 삭제는 민감하기 때문에 어떤 의미인지 설명바람. 그리고 개인정보 보호법에서 개인정보가 포함된 파일은 암호화하도록 되어 있는데, 암호화 된 파일에도 필터링 기술을 적용할 수 있는지, 필터링 된 개인정보도 법적 혹은 다른 목적으로 공개 가능할 경우 수작업 처리해야 되는 것인지 궁금함. 두 번째 발표문 관련, 비공개정보를 기술적으로 어떻게 처리할 것인가에 앞서 어느 범주로, 어디까지 할 것인지 먼저 고민해야 할 듯. R&D 사업사례에서, 공개기록물에 대해서도 비공개정보를 필터링 할 수 있는 대책마련이 필요하다는 결론은 이해되지 않음. 주민번호나 이메일이 많이 나온 것은 시행공문에 포함된 공개 가능한 이메일 주소가 66%나 필터링되지 않을까 생각함. 특히 개인정보가 필터링 된 공개기록물이 88%라고 하는데 의구심이 듭. 국가안전보장과 관련된 정보공개법 2호에 대해서도 학습 분석 시도했는데, 비공개 대상 정보 1호부터 8호 중 2호를 선택한 이유가 궁금함. 인명을 마스킹하는 것은 법원에서도 판결문 제공을 위해 시도 하고 있으나, 사람의 이름은 정형화 되지 않고, 가명도 섞여 있어, 정형화 된 위치를 잡아 처리하는 기술 등을 연구하고 있지만 기술적 한계에 부딪힘. 따라서 인명에 대한 마스킹 시도는 굉장히 뜻깊으며 좀 더 개발되길 바람. 기록에 대한 마스킹은 설계단계에서 그친 것 같았음. 필터링을 할 때 PDF나 원문 파일을 텍스트 파일로 변환해서 필터링 한 것이 거의 100%에 가깝게 검출 되었는데, 마스킹은 실제 어느 파일에 하려했던 것인지 의문임. 우리나라의 개인정보 보호법은 일률 보편적으로 적용되다보니, 도서관이나 국가기록원 등 국민의 정보 접근을 확대보장 해야 하는 기관의 활용 측면에는 한계가 있음. 이 부분은 지속적으로 공동의 논의점이 되길 바람.

세 번째 발표 관련, 도서관 이용자의 개인정보를 수집하고 보호 하는 방법과 개인의 의견에 관한 이력도 개인정보 보호법으로 보호되는 개인정보에 포함되는 것인지를 검토한 바 있는지 의견 부탁드립니다. 두 번째는 개인정보가 포함 된 자료의 비공개요청은 당사자가 하는지, 제3자가 특정 도서를 보다가 개인정보를 발견해도 가능한 것인지 궁금함. 온라인 자료는 개인정보 일부 삭제 후 서비스가 가능한데, 개인정보가 포함 된 오프라인 자료의 대출·열람 제한은 어떤지 궁금함. 세 번째는 소장도서에 열람제한 요청이 들어오면 비공개 결정 전 별도의 내부 검토 프로세스가 있는지, 개인정보를 마스킹하고 제공하는 온라인 원문 서비스 비율도 궁금함. 국립중앙도서관 사례에서 보면 세부주소 외 개인의 이력은 대부분 공개 하는데, 보호해야 할 개인정보의 범주는 어디까지인지 답변 부탁드립니다.

마지막 발표 관련, ‘공익적 기록보존’의 의미가 궁금하고, 가명처리의 목적이 통계나 공익적 기록보존 ‘등’인데, 이외에도 추가로 처리할 수 있는 여지가 있는 것인지 의견 부탁드립니다. 개인정보 보호법이 개정되면서 가명정보와 익명정보 정의가 비슷한 것으로 이해되는데, 추가 설명 부탁드립니다.

가명처리 방식에는 삭제와 대체가 있는데, 선택 시 기준이 궁금함. 개인정보 보호법에서는 민감 정보나, 고유 식별 정보처럼 개인정보의 범주를 넓게 보는데, 가명정보 처리 범주도 모든 것을 포괄하는지 궁금함. 전문기관 사례와, 공공법인·기관·단체·일반기업까지 전문기관이 될 수 있는지 설명 부탁드립니다.

- (한국과학기술정보연구원 임석종 박사) 원본파일 삭제 옵션은 해당 문서에서 원본파일을 유지 할 수도 있고 삭제할 수도 있다는 단순한 선택적 기능임. 공공기관이나 개인의 경우 지침이나 법률에 따라 개인정보를 폐기하게 되기 때문에, 원본에 있는 개인정보 삭제 후 마스킹 한 자료만 사용할 수도 있는 상황을 고려해 기술적으로 선택 기능을 제시함.

두 번째는 암호화 목적 자체가 개인정보 유출 방지이기 때문에, 개인정보가 암호화 된 파일을 처리하지는 않고, 해제 된 문서 파일을 대상으로 처리함. 비공개 개인정보와 공개 가능한 개인정보가 혼합된 전자문서에서 공개가능한 개인정보에 대한 판단은 사용자의 선택이기 때문에, 검출 내용을 보고 마스킹할지와 어떤 범위까지 할 것인지는 단순한 선택 기능을 제공하는 부분으로, 수작업으로 검토 후 결정해야 함.

○ (국가기록원 김진아 사서주사) 첫 번째, 비공개정보 검수 기준을 만들기 위해 기존의 기준서에서 정의하던 비공개정보 대상과, 비공개정보를 가진 기록물 철을 분석했고, 결과는 21페이지의 표로 정리함. 두 번째, 기록관리시스템을 사용하고 있는 기관의 개인정보 필터링 사례에 대한 질의에 답변하려면, 결과 값과 타기관 소장 기록물을 전부 매칭 해 봐야 하는데, 타기관 소장 기록물이라 확인의 어려움이 있음.

이메일 주소는 말씀대로 공개기록물임에도 불구하고, 본문에 있는 공무원의 이메일이 많이 검출되지 않았나 생각함. 이번 보고서에서는 다 담지 못했지만, 연구를 진행하면서 이런 부분이 검출되지 않도록 제외처리 등 장치를 마련함. 비공개 대상 2호가 추가 된 이유는 연구 진행 과정에서 받은 여러 의견 중 민감하게 생각할 수 있는 제2호에 대한 사항도 추가해보자는 의견이 있어, 내용 분석을 통해 단어사전에 넣은 케이스임.

마지막으로, 마스킹은 한글이나 MS오피스 등 여러 가지 파일을 대상으로도 진행했었지만, 현재 기술로는 PDF 표준화 된 규격 안에서 마스킹 하는 기술만 가능하기 때문에, PDF의 비공개 대상을 마스킹 하는 것으로 설계함.

○ (국립중앙도서관 최윤경 사무관) 수집하는 개인정보는 개인정보 보호법상 접근보다, 도서관에서 수집하는 개인정보에 대해 말씀드린 것으로, 개인에게 동의를 받는 정보 정도로 생각바람. 온·오프라인 자료에 포함된 개인정보는 당사자 요청도 가능하고, 제3자 요청도 가능한데, 담당부서에서는 제3자가 요청하는 경우보다 당사자 혹은 해당 출판사의 검색 및 열람 제한 요청 사례가 대부분이라고 함. 개인정보가 포함 된 오프라인 자료는 정보를 가리기 위해 종이를 붙일 경우 열람 시 떼어내고 확인하거나 파손될 위험이 있기 때문에, 아예 검색되지 않도록 열람 제한함. 비공개 요청은 법적으로 명시되어 있고, 내부 규정도 있기 때문에 담당자와 관련 과에서 확인 후 처리하며, 위원회는 별도로 없음. 온라인 원문 제공에 대한 개인정보보호처리 비율은 따로 관리되지 않아 정확한 수치를 말씀드릴 수는 없지만, 적은 비율이라고 함. 두 번째, 성명·생년월일·학력·경력 등 책의 판권지나 표지 쪽에 들어가는 저자의 이력에 해당하는 경우, 본인이 노출한 정보라면 가리지 않는 것으로 알고 있음.

○ (국가기록원 조이형 연구협력과장) 토론 답변과 더불어 개인정보보호위원회가 출범하면서 개인정보보호관련 제도와 정책을 더 구체화 시키고 확대하는 분위기인데, 한국인터넷진흥원에서 개인정보 보호법과, 정보공개법 비공개대상정보 6호 등에 관한 내용도 큰 틀에서 진행하고 있는 것이 있으면 함께 말씀 부탁드립니다.

○ (한국인터넷진흥원 박윤식 데이터활용지원팀TF 팀장) 공익적 기록보존의 구체적 의미와 사례는 아직 없음. 실제 결합은 전문기관을 통해야 하지만, 아직 결합 전문기관으로 지정된 곳이 없음. 새로운 제도이기 때문에 시범사례로 여러 연구를 진행하고 있음. 예를 들어 사회보장위원회에서 사회보장 사각지대를 찾는 연구와 관련, 어떤 식으로 결합 정보를 사용할 지 함께 연구 중임. 통계, 과학적 연구, 기록보존 외 다른 사례도 가명처리의 목적이 될 수 있을지는, 향후 비슷한 무게의 공익적 사례가 나올 때 ‘…등’으로 유권 해석 할 수도 있겠지만, 아직은 앞선 세 가지 목적으로만 사용할 수 있다고 봐야 함.

두 번째, 가명정보와 익명정보의 가장 큰 차이점은 어떤 방법을 통해서도 개인을 식별할 수 없으면 익명정보이고, 법으로는 금지 돼 있지만 결론적으로 누구인지 찾아낼 수 있는 정보는 가명정보임. 예를 들어, 우리나라에 남자 50%, 여자 50%가 있다는 정보는 익명정보에 해당함. 가명처리기술 중 삭제와 대체 중 어떤 것을 사용할지에 대한 기준은 가명처리 수준 정의를 할 때 결정하면 되는데, 가명정보처리 가이드라인 부록에 상세히 되어 있으니 참고 바람.

세 번째, 개인정보보호관련 조치에서 고유식별정보, 바이오정보, 비밀번호는 암호화해서 보관하도록 되어 있고, 그 세 가지 정보는 가명처리 대상에 포함되지 않음. 그 외에, 가명처리를 하지 못하도록 법에 특별히 명시 된 것을 제외한 민감정보와 개인정보는 가명처리가 가능함.

마지막으로 결합 전문기관은 개인정보보호위원회나 정부부처의 장이 심사를 통해 지정 하는데, 공공기관, 사기업 구분 없이 모두 가능함. 다만 기업은 대부분 영리추구를 목적으로 두기 때문에, 공공기관 등 목적이 부합할 경우 정부부처로부터 심사받고 결합전문기관 지정을 받으면 됨. 만약 여러 목적을 같이 가졌다면, 개인정보보호위원회의 지정심사를 받으면 됨. 사기업 중에서도 결합전문기관 지정을 준비하고 심사받는 곳도 있음.

결합사례는 시작 된지 얼마 되지 않았기 때문에, 실질적으로 어떻게 결합이 돼야 목적을 달성할 수 있는가 연구 중임. 사회보장 외에도 인명구조를 할 때, 전화로 긴급신고가 들어오면 데이터 분석을 통해 긴급 상황인지 아닌지를 빨리 인식하고 출동 여부를 결정하는데 사용하는 결합도 추진하고 있고, 불법스팸의 경우 정보 결합을 통해 스팸 발송 패턴 등을 연구해 어떻게 빨리 막을 것인지도 일종의 결합 사례로 연구 중임. 기록관리 등 여러 가지 분야의 새로운 정책개발에도 결합이 훌륭하게 이용될 수 있을 것이라고 생각함.

○ **(한라대학교 김순석 교수)** 부연설명으로, 개인정보 보호법은 일반법이고 정보공개법은 특별법이기 때문에, 정보공개법이 우선순위임. 정보공개에 필요한 개인정보에 대한 부분은 정보공개법을 참고하면 되고, 그 외에 일반적인 사항들에 대해서는 일반법이 적용된다고 보면 됨. 가명정보와 익명정보의 가장 큰 차이점은 개인정보 보호법 제58조의2를 참고하면 되는데, ‘시간·비용·기술 등을 합리적으로 고려할 때 다른 정보를 사용하여도 더 이상 개인을 알아볼 수 없는 정보’가 익명정보를 정의 한 것임. 가명정보는 과학적 연구, 공익적 기록보존, 통계작성을 위해 개인정보를 쓸 수 있도록 특례로 허용한 것 일뿐, 가명정보 또한 개인정보로서, 시간·비용·기술을 조금만 들여도 언제든지 식별 가능하고, 민감 정보들이 대부분 살아있기 때문에 주의가 필요함. 고유식별정보인 주민등록번호는 그 자체만으로도 누구인지 바로 알아볼 수 있기 때문에 삭제가 원칙임. 다만 민감정보는 가명처리 대상이 될 수 있으며, 처리 후 세 가지 목적으로만 활용할 수 있음. 예를 들어, 몸무게나 혈압, 심전도 같은 건강정보는 측정할 때마다 달라지기 때문에 민감정보이긴 하지만, 공개거나 하지 않고 그대로 처리함. 삭제와 대체에 있어 삭제는 쉽지만, 대체의 경우 범위가 다양함. 예를 들어 삭제의 경우 1950년 9월 16일이면 제일 쉬운 삭제는 일 단위를 지우는 것임. 1950년 9월. 그 다음 더 강도를 높이면 1950년. 이런 식으로 위험도에 따라 식별 가능성 정도를 처리하도록 가이드라인에서 제시하고 있음. 외부제공과 내부사용의 경우 위험도가 다르기 때문에 외부 제공 시 추가처리가 필요함. 환경적 위험도 검토 후 강도를 정해야하기 때문에, 일괄적용 할 수는 없음.

○ **(한국인터넷진흥원 오용석 개인정보정책단장)** 개인정보 보호법 6조 관련, 다른 법률에 특별한 규정이 있는 경우를 제외하고는 이 법에서 정하는 바에 따르기 때문에 일반법과 특별법의 지위가 부여되었고, 정보공개법의 개인정보에 대해 특수하게 적용하기 위해서는 법 개정이나 추가적인 배경·활용 논의가 필요함. 국가기록물에 포함된 개인정보도 활용을 위해서는 관련법에 명시가 되어야 할 듯. 가명정보가 개인정보 보호법에 추가 된 것도 4차 산업 혁명의 데이터 산업 발전을 위해 반영된 것도 있고, 우리나라의 개인정보 보호법과 같은 EU의 GDPR(General Data Protection Regulation)에서의 허용 사례도 반영 된 것임. GDPR에는 개인정보에 관한 권리가 절대적인 것이 아니라 타 권리와 조화롭게 이용되어야 된다고 명시 돼 있음.

우리나라도 헌법재판소에서 개인정보 자기결정권을 판시하면서, 개인이 갖고 있는 개인정보에 대한 권리가 헌법상 권리로 작용을 하고 있고, 국가기록물의 알 권리와 거의 동등한 수준임. 여러 가지 권리가 상충될 때는 법원에서 이익형량을 통해 판단하고, 사회적, 행정부에서도 판단하게 됨. 최근 어린이 학대 문제로 어린이집 CCTV가 의무화 되었는데, 어린이의 권리를 위해 도입된 데 반해 유치원 교사의 사생활은 침해하는 경향이 있음. 법이 통과 된 것은 어린이의 권리가 이익형량에서 우선시 된다는 것이 사회적으로 합의됐기 때문임. GDPR처럼 국가기록물의 개인정보를 일정부분 활용하기 위해서는 관련법 상에 명시가 돼야 할 필요가 있음.

첫 번째, 개인정보 필터링 및 마스킹 기술과 관련 개인정보 보호법에 가명정보가 추가 됐고, 익명정보에 대한 58조의2가 신설되었음. 비식별, 마스킹처리 테크닉이 가명정보에 쓰이는지 익명정보에 쓰이는 지에 대한 구분은 정확히 없음. 이름을 오\*\*이라고 했을 때, 가명정보인지 익명정보인지 구분하기 위해서는 김순석 교수님 말씀처럼 위험도분석까지 같이 고려해야 됨. 우리나라에 몇 없는 성만 남기고 이름을 마스킹 처리 했다면 개인정보가 아닌지? 문맥과 처리 환경을 같이 고려해야 됨. 개인정보에 대한 판단도 시간, 비용, 기술 등의 결합가능성과 입수가능성 등을 같이 봐야 함.

주민등록번호, 카드번호처럼 정규식 또는 정규 표현식(Regular Expression)이라고 하는 일정한 패턴의 개인식별정보는 100%가 아니더라도 대부분 자동검출 되고 있으나, 이름과 같이 정규식으로 표현되지 않는 정보를 필터링할 경우 솔루션들은 보통 DB화합. KISTI에서는 대부분 정규식으로 표현 가능한 개인정보의 검출 및 처리기술 현황에 대해 말씀하셨는데, 자연어 처리와 관련된, 비정형데이터에서의 개인정보 검출에 대한 연구 방향성이 있는지 궁금함. 국가기록원에서도 이미지 파일, 스캔자료처럼 다양한 비정형 데이터의 개인정보검출에 대한 연구가 필요하다고 생각함.

두 번째 발표에서, 소송정보도 개인정보로 본 것인지 궁금함. 인물명 중심의 기계학습 연구에서는 70%의 결과 값이 나왔다고 하는데, 연구의 타당성과 효과성 검증을 위해 기계학습 전 DB화라든지, 비교 연구를 한 것이 있는지 궁금함. 별도로 찾아 본 Kobert 모델 자료에는 익명관련 정확도가 93~94%로 나와 있는데, 국가기록원 데이터는 Kobert 모델을 썼는데 70%로 나온 것에 대해 분석한 바 있는지 궁금함. 개인정보 예외처리 대상에 해당한다는 의사결정은 수동처리 했는지 궁금하고, 이에 대한 자동화률이 있는지 궁금함.

국립중앙도서관의 개인정보 침해요인 평가는 개인정보 보호법에서 법령의 제·개정 사항에 대해 개인정보 침해 요인이 있는지 개인정보보호위원회에서 보는 절차고, 법 개정 외, 개인정보 보호법 제18조제2항제5호의 개인정보를 목적 외의 용도로 이용하거나 이를 제3자에게 제공하지 아니하면 다른 법률에서 정하는 소관 업무를 수행할 수 없는 경우는 보호위원회의 심의의결을 거친 경우 판단해야 될 것이라고 생각함. 한국인터넷진흥원에서는 개인정보에 대한 감수성에 관해 많은 고민을 하고 있음. 어린이도서관의 개인정보는 아이용과 부모용으로 분리하는 등 처리방침에 특수성을 반영해도 좋을 듯. 도서관의 개인화 서비스는 특정 책의 대여 등 사상적 판단을 가능하게 하는 민감 정보에 대해 주의 할 필요가 있다고 생각함.

개인정보 가명처리에서는 마스킹 기술이 가명처리인지 익명처리인지 부가적으로 말씀 바라며, 마스킹처리 후에는 공개 가능한지 궁금함. 제대로 가명처리 한 것인지 개인정보 처리자의 판단에 따라 외부전문가로 구성된 적정성 평가단이 검토할 수 있다고 하셨는데, 결합이나 가명처리는 개인정보처리기관이 내부적으로 할 수도 있기 때문에, 기관에서 자의적으로 외부전문가를 구성하면 가명처리의 적정성에 대한 공정성 있는 평가가 가능할지 의문임. 이 부분에 대해 개인정보보호위원회나 KISA 차원에서 인력풀을 운영 할 계획이 있는지 궁금함. 정형데이터에 대한 적정성 검증은 DB분석 도구 등 여러 가지 도구가 있을 텐데, 국가기록원에는 보고서 등 텍스트나 데이터, 이미지 등 비정형데이터에 대한 가명·익명처리 적정성 검토·검증이 필요하며, 이에 대한 도구나 기준이 있는지 궁금함.

○ (한국과학기술정보연구원 임석종 박사) KISTI의 기술 현황에서 정규식 개인정보에 대한 처리기술만 있는 것 같다는 말씀에 대체로 동의함. 지금까지 보고서 원문이 개인정보처리의 주요대상이었고, 특별히 텍스트환경에서의 자연어에 대한 처리까지 할 필요성을 느끼지 못했었는데 문제의식을 느끼게 됨. 다만, 향후 연구과제에서도 말씀드렸드시피, 이미지 안에 포함 돼 있는 보고서가 텍스트라 하더라도 표나 그림 안에 있는 개인정보가 상당히 많아, 그 정보들을 처리하는 것이 중요과제임. 기계학습데이터를 구축하면서 이미지 형태로 돼 있는 정보들에 대한 패턴을 조사를 하고, 물리적으로 좌표를 인식하도록 하는 게 아주 어려움. 주민등록번호와 같이 정형화 된

이미지면 비교적 쉽고 간단하지만, 이미지 형식이나 내용이 상당히 다양하기 때문에 딥러닝 학습모델을 개발해서 적용하는 방법을 연구 중에 있고, 자연어 처리와 관련해서는 시범적으로 특수문자라든지 수식, 디지털변환과정에서 깨지는 문서 등에 대해 내부적으로 작업을 진행하고 있음.

○ (국가기록원 김진아 사서주사) 국가기록원에는 판결문관련 정보와 기록물이 굉장히 많음. 판결문 안에서의 소송번호는 이름, 죄명, 판결문에 대한 내용과도 연관되기 때문에 지금까지 공개재분류를 하거나 열람처리를 할 때 소송번호도 개인정보로 취급했었고, 정규식 패턴으로 추가한 부분들이 있음. 두 번째, Kobert에서 인명정보에 대한 결과 값이 0.93이나 0.94가 나왔는데, 기록원은 0.7이 나온 것은 일반적으로 인터넷상의 텍스트에 비해 공문서나 보고서 양식에는, 중간에 스페이스바가 많이 들어있다든지 줄 바꿈이 있는 등 인명검색을 할 때 오류요소들이 많기 때문임. 외자로 된 이름도 찾기 힘든 부분 중 하나이다보니, 평균적인 결과 값보다 적은 결과 값이 나옴. 수동처리를 한 사항은 문서관리카드 기안문에 있는 공무원의 이름과 직급, 이메일, 주소, 전화번호였는데, ‘문서가 끝나고 몇 단락 이하까지의 문장 안에 있는 것들은 제외 한다’는 식으로 관련 룰을 만들어 알고리즘을 만들었고, 최종적으로는 수동으로 판단하는 구조로 처리함.

○ (국립중앙도서관 최윤경 사무관) 이용자 데이터의 확대수집에 동의하는 것이 가능하도록 개인정보처리방침을 변경할 계획인데, 만약 한 번도 모으지 않은 데이터를 모으게 될 경우 개인정보침해요인 평가를 거쳐야 한다고 알고 있음. 도서관에서 개인정보나 이용자의 데이터를 활용해 추천 서비스 등 개인화서비스를 한다면 어느 수준까지 해야 되는지에 대해 잘 지적 해 주신 것 같고, 다만 사상적으로나 청소년유해물과 같은 자료는 현재 이용제한을 시키고 있음. 이런 위험한 부분은 서비스를 기획 할 때 염두에 두겠음.

○ (한국인터넷진흥원 박윤식 데이터활용지원팀TF 팀장) 마스킹 자체를 익명처리로 볼 수는 없고, 상황에 따라 고려해야 함. 두 번째, 적정성 평가에 대해 현재 외부전문가로 반드시 구성해야 된다는 규정이 있진 않지만, 권고 사항임. 최종적으로 가명처리 된 결과물을 사용 시 개인이 식별된다면,

가명처리 한 개인정보처리자 쪽에 문제가 생기고, 책임져야하기 때문에, 이 문제에서 자유롭기 위해서는 심사단을 잘 구성해야 할 것임. 가명정보제도가 시행 된지 얼마 안 되었기 때문에 아직 외부전문가 풀을 별도로 운영하진 않지만, 문제가 예상될 때 한국인터넷진흥원이나 개인정보보호위원회로 요청하시면 가지고 있는 작은 풀이라도 지원 해 드릴 용의가 있음.

이미지나 동영상 같은 비정형데이터에 대해서는 아직까지 가명처리 기술과 적정성검토기준이 명확하지 않음. R&D를 통해 가명처리나 적정성 평가 기술들을 연구 중에 있으니, 향후 기술이 어느 정도 완성되면 공개하겠음.

○ **(한라대학교 김순석 교수)** KISTI의 개인정보 민감도 분류 기준과 마스킹 외 다른 기술도 있는지 궁금하고, 가명처리를 통한 활용에 대해서도 준비하는 게 있으신지, 그리고 국가기록원에서는 혹시 후속연구 진행 계획이 있으신지 궁금함. 국립도서관의 경우, 외국 도서관 사례가 흥미가 있었는데, 미국처럼 수집 된 정보들을 개인을 식별하기 위한 용도로는 사용하지 않되, 이용자 분석이나 웹사이트 개선을 위한 용도로 활용하는 사례가 우리나라에도 있는지 여쭙봄. 영국의 경우 비개인화 및 익명처리를 통해 이용자 데이터의 심층적 분석과 다른 정보와의 결합이 가능하도록 하였는데, 우리나라에도 이런 사례가 있는지 궁금함. 향후 솔루션 도입 계획이 있는지도 여쭙봄. 마지막 발표에서는 개인정보 보호법 개정 이후 개인정보보호위원회를 비롯해 향후 공익적 기록보존에 대한 활용 관련 정책이나 행·재정적인 지원 방향이 있는지 여쭙고자 함.

○ **(한국과학기술정보연구원 임석종 박사)** 과학기술정보 특성상 개인정보를 민감도에 따라 특별히 분류하지는 않음. 비식별 기술은 암호화나 DRM 기술, 소소하게는 텍스트나 폰트 깨짐에 대한 노하우나 처리 기술들을 가지고 있음. 최근 이미지를 OCR로 바꾸는 작업도 해 보았으나 용이하지는 않았음. 그래서 방대한 기계학습데이터를 구축하고 있다는 장점에 접근하여, 논문이나 보고서 안에 들어있는 표·그림 330만 건에 대한 데이터를 구축하고 학습시킴으로써, 물리적·논리적 좌표를 통해 이미지 안에 포함 돼 있는 개인정보를 추출하고, 마스킹 하는 연구를 진행하고 있음.

- **(국가기록원 김진아 사서주사)** 샘플링 검사도 하고 모델도 만들어야 하는데 비해 연구 기간이 짧아 아쉬움이 남음. 후속 연구가 진행된다면 지금의 학습모델을 더 똑똑하게 만드는 반복학습이 필요할 것 같고, 인명뿐 아니라 그 외에 정보공개법에 있는 다른 후속과제들을 포함하는 반복학습을 진행할 계획임. 두 번째로는, 실제 각급 기관에서 사용하고 있는 기록관리 시스템과 국가기록원에서 사용하고 있는 중앙영구관리기록시스템 안에서 구현되어야 할 기능이기에 때문에, 이 안에서의 연계방법과 최종적으로 시스템 안에서의 마스킹 기술까지 연구해야 될 부분으로 생각하고 있음.
  
- **(국립중앙도서관 최윤경 사무관)** 우리나라의 경우 수집정보를 이용자분석, 서비스 개선이나 웹서비스 개선을 위한 용도로 활용할 수 있다는 자체는 명시되어 있는데, 수집하는 정보 자체가 너무 다름. 미국은 로그라든지 이용기록 자체를 더 상세하게 기술해서 수집정보를 확대했는데, 아직 우리는 아이디, 주소 등 개인정보 위주로만 수집하고 있어서, 활용 측면에서는 좀 미흡함. 영국 유사 사례는 향후 검토 해 보겠음. 향후 솔루션 도입 계획은 갖고 있긴 하지만, 기술적인 부분에 대한 이해가 부족한 것 같아서, 오늘 KISTI 사례 등을 통해 더 검토 해 보겠음.
  
- **(한국인터넷진흥원 박윤식 데이터활용지원팀TF 팀장)** 공익적 기록보존 등 가명정보 활용과 관련 된 부분은 아직 연구된 바가 많이 없기 때문에, 향후 정부의 여러 가지 정책이나 행정·재정적인 지원 방향에 대해서는 함께 논의를 하되, 관련예산 등의 부분은 지금도 적극적인 지원을 받고 있어서, 함께하면 가명처리 제도에 대해 서로 같이 발전 할 수 있는 기회가 될 것임.



**붙임 3**

**정책포럼 참여 만족도 설문조사 결과(요약)**

문항	총계	만족도(점수)					미응답
	80.09	107	244	91	7	0	40 (통계포함)
문항1. 포럼은 전반적으로 만족하는가?	소계	매우만족(100)	만족(80)	보통(60)	불만족(40)	매우불만족(20)	미응답
	79.02	10	19	12	0	0	0
문항2. 포럼 운영은 적절했는가? (장소, 안내, 자료)	소계	매우적절(100)	적절(80)	보통(60)	부적절(40)	매우부적절(20)	미응답
	81.46	10	24	7	0	0	0
문항3. 시간배분은 적절했는가?	소계	매우적절(100)	적절(80)	보통(60)	부적절(40)	매우부적절(20)	미응답
	76.59	6	22	13	0	0	0
문항4. ★ <b>최고</b> 포럼 주제는 적절했는가?	소계	매우적절(100)	적절(80)	보통(60)	부적절(40)	매우부적절(20)	미응답
	84.88	15	23	1	2	0	0
문항5. 기조연설 및 발표내용은 주제와 적합했는가?	소계	매우적합(100)	적합(80)	보통(60)	부적합(40)	매우부적합(20)	미응답
	81.46	10	24	7	0	0	0
문항6. 주제에 따른 발표자섭외는 적절했는가?	소계	매우적절(100)	적절(80)	보통(60)	부적절(40)	매우부적절(20)	미응답
	81.95	10	25	6	0	0	0
문항7. 발표내용에 따라 토론이 적절했는가? (내용,시간 등)	소계	매우적절(100)	적절(80)	보통(60)	부적절(40)	매우부적절(20)	미응답
	81.50	9	25	6	0	0	2
문항8. 토론자 섭외는 적절했는가?	소계	매우적절(100)	적절(80)	보통(60)	부적절(40)	매우부적절(20)	미응답
	81.50	9	25	6	0	0	1
문항9. 본 포럼이 기록관리 정책에 대한 관심을 끌기에 충분했는가?	소계	매우만족(100)	만족(80)	보통(60)	불만족(40)	매우불만족(20)	미응답
	77.07	9	18	13	1	0	0
문항10. ★ <b>최저</b> 본 포럼이 기록관리 업무에 도움을 주는가?	소계	매우도움됨(100)	도움됨(80)	보통(60)	도움안됨(40)	매우도움안됨(20)	미응답
	74.63	9	15	14	3	0	0
문항11. 향후 정책포럼에 참여할 의사가 있는가?	소계	매우있다(100)	있다(80)	할 수도 있다(60)	없다(40)	매우없다(20)	미응답
	80.98	10	24	6	1	0	0

	소계	주관식	미응답
	41건	12	29건
문항12. 향후 개최 시 반영할 개선점이 있다면?	<ul style="list-style-type: none"> <li>○ 개별 기관에서 고민하고 있는 좋은 주제였습니다. 정책방향이나 논의주제로써 적절하지만 뭔가 구체적인 방안 등은 조금 미흡했던 것 같습니다. 추가적이고 후속조치 성격의 구체 사례나 방안에 대한 논의 자리가 있으면 좋겠습니다.</li> <li>○ 직접적인 업무에 도움이 안 돼 아쉬움이 있었으나 앞으로 적용 시 매우 좋을 것 같습니다.</li> <li>○ 사전계획을 통해 주제의 일관성 유지 필요. 다루는 정보의 대상(보존기록, 통계, 행정정보데이터셋, 도서)이 다른데 "개인정보보호"라는 주제를 논의하다보니 각기 다른 상황만 보여주는 것 같음</li> <li>○ 주제별 발표시간이 좀 짧은 것 같습니다.</li> <li>○ 토론자의 토론 질문 시간에 대한 조정이 잘 되었으면 좋겠습니다. 좋은 질문 및 답변을 많이 들을 수 있으리라 생각합니다.</li> <li>○ 17시까지 진행되면, 지방에 사는 사람은 너무 늦게 귀가함.</li> <li>○ 1,3주제와 같이 단순 업무소개, 연구사항이 아닌 것은 토론 할 필요가 없다고 생각함. 1,3주제를 묶어 진행, 2,4주제를 묶어 심층토론 할 수 있도록 진행변경 필요.</li> <li>○ 발표자료 화면에 같이 띄워주면서 발표해주시면 좋겠습니다.</li> <li>○ 시간표에 시간표기 오차가 있습니다. 마스크 때문일 수도 있는데 발표소리가 잘 들리지 않아 좀 불편했습니다</li> <li>○ 수고많으셨습니다.</li> <li>○ 환기를 좀 더 했으면 좋겠네요. Online 중계도 했으면 좋겠습니다.</li> <li>○ 참여인원 확대. 기록관리기관 확대.</li> </ul>		

### ○ 총 평 : 성과목표(평균 80점) 105.13% 달성

\* '20년 행정안전부 성과관리시행계획 반영 목표치: 평균 80점(만족도 총점÷개최횟수)

연번	개최명	개최일	총점
총계	총 4회		평균 84.10
1	'기록의 날' 기념 학술회의(제1차 기록관리 정책포럼)	'20. 6. 9.	80.52
2	제1차 기록관리 연구세미나	'20. 7. 8.	94.09
3	제2차 기록관리 정책포럼	'20. 7. 28.	81.71
4	제3차 기록관리 정책포럼	'20. 11. 12.	80.09

- 문항10(기록관리 업무에 도움)이 최저점 항목이었으며, '주제' 관련 주관식 의견이 있었음.
  - ⇒ 개별 기관에서 정책방향을 위해 고민해야 하고 있는 좋은 주제였으나, 구체적 해결 방안이나 실무에 적용할 수 있는 직접적인 사례 등을 포함하여 논의 할 필요가 있음.
    - ↳ 정책·실무적으로 적용 가능한 사례가 포함 된 주제 선정을 위해 노력하겠음. 제도 초창기이기 때문에 사례가 부족한 주제를 논의할 경우 후속논의가 가능토록 하겠음.
- 주제발표 및 토론시간의 조정이 필요함.
  - ⇒ 주제별 발표 시간이 짧고, 토론 질문 시간의 불균형에 대한 의견이 있었음.
    - ↳ 유사 발표수를 줄여 발표 시간을 재조정 하고, 토론자 시 시간안배 등 증재토록 하겠음.
- 진행 미흡 개선 필요.
  - ⇒ 개선의견: 자료집 오기, 음향 크기, 개최장소 환기, 온라인 중계 필요, 참여인원 확대 등
    - ↳ 자료집 오기 검수 철저 및 개최 시 음향이나 환기 등 현장진행에 심열토록 하고, 현재 실시간 중계가 불가능하여 녹화 후 홈페이지를 통해 사후 공개 중. 개최 사전 사후 영상 및 자료집도 공개 된다는 점을 더욱 홍보토록 하겠음. 참여인원 확대는 거리두기 단계지침에 따르므로, 대규모 공간 확보가 어려울 시 신청자를 대상으로 양해토록 하겠음.

○ 학술회의 참여 만족도 그래프 요약



○ 설문조사 응답자 통계 분석

