

보건의료데이터 가명처리 방법 개선, 개인정보 보호하고 데이터 활용 활성화한다

- 가명처리 범위 확대 및 데이터 제공기관 책임 범위 명확화로 데이터 활용 활성화 촉진 -
- 「보건의료데이터 활용 가이드라인」 개정(안) 국민 의견수렴 -

디지털 헬스케어 분야에서 가명정보를 활용한 연구를 활성화하고 개인정보 보호를 강화하기 위해 의료데이터의 가명처리 방법 등을 개선한 「보건의료데이터 활용 가이드라인」 개정안을 마련하고 1월 19일(금)부터 1월 29일(월)까지 의견수렴 기간을 갖는다.

이번 개정안은 개인정보는 안전하게 보호하면서도 연구 목적 등 법률이 허용한 범위 내에서 보건의료데이터 활용이 활성화되도록 유전체 데이터 등 비정형데이터의 가명처리 범위를 확대하고, 데이터 제공기관의 책임 범위를 명확화하는 등 현장 애로사항을 해소하는 데에 중점을 뒀다.

먼저, 유전체 데이터는 데이터 파일 형태*에 따라 염기서열 및 메타데이터 내 주요 식별정보는 제거하거나 대체하는 방식 등으로 가명처리하여 활용할 수 있도록 개선했다. 동시에, 연구 목적으로 활용하더라도 데이터 접근권한을 통제하고 폐쇄환경에서 활용하도록 하는 등 정보보호를 위한 안전조치는 강화했다.

* FASTQ/SAM/BAM/VCF 파일 및 검사기록지

** 한국유전체학회를 통한 연구용역 추진 및 각 분야 전문가 논의('23. 6월~12월)

진료기록 등 자유입력데이터는 자연어 처리기술 등을 활용하여 정형데이터로 변환 후 식별정보 삭제, 대체 등을 거쳐 가명처리하여 활용하도록 안내하였다. 또한, 음성데이터의 경우에도 문자열로 변환하여 식별정보 삭제, 대체 등 가명처리하거나 필요시 추가로 노이즈 방식 등

을 활용하도록 규정하였다.

아울러, 가명정보 처리에 관한 개인정보 파기(개보법 제28조의7), 개인정보 처리 방침의 공개사항(제30조)에 가명정보 처리에 대한 사항 포함 등 개정된 「개인정보 보호법」 변경사항을 반영하였다.('23.9.15. 시행)

< 비정형데이터 가명처리 방안 개정 전·후 비교 >

구분	현 행	개 선
유전체 데이터	널리 알려진 질병의 유전자 변이유무 확인 등 예외적인 경우를 제외하고 가명처리 유보	NGS 기반 유전자 검사를 통해 생성된 표준화된 SAM 및 VCF 등은 메타데이터 및 염기서열에 대한 가명처리
자유입력 데이터	가명처리 유보	자연어 처리기술 등 정형데이터로 변환 후 식별정보 삭제·대체 등 방법으로 가명처리
음성 데이터	가명처리 유보	음성인식 기술 이용하여 텍스트 처리한 뒤 자유입력 데이터 가명처리 방법으로 처리

그 밖에 의료기관 등 보건의료데이터 제공기관의 과도한 부담으로 데이터 활용이 위축되지 않도록 가명정보 처리·제공과정에서의 법적책임 범위를 명확히 하는 등 현장 애로사항을 해소하기 위해 노력하였다.

가명정보 처리·활용 시	가명정보 제3자 제공 시
- 가명정보 처리·활용 과정에서 의도치 않게 특정 개인을 알아볼 수 있는 정보가 생성되었다는 사실만으로 가명정보를 처리한 자 또는 해당 가명정보를 제공한 자를 처벌하지 않음 ※ 단, 해당 정보의 처리를 즉시 중지하고, 지체없이 회수·파기하여야 함	가명정보를 제공받은 자의 안전조치 미이행 등으로 가명정보 유출 등의 문제가 발생하였거나 고의로 재식별 행위를 하는 등 그 행위 주체의 위법으로 발생한 사안에 대해서는 해당 행위자만 제재함

이번 개정안은 의료계·산업계·학계·공공기관 등 의료데이터 활용 현장 간담회*를 거쳐 마련됐다.

* 보건 의료데이터의 안전한 활용을 위한 현장 간담회('23.12.1.)

개정안에 대한 상세한 사항은 보건복지부 누리집(www.mohw.go.kr)→알림 → 공지사항 에서 확인할 수 있으며, 의견은 보건복지부 보건의료 데이터진흥과로 제출하면 된다.

* 제출의견 보내실 곳: dataguideline@k-his.or.kr

보건복지부 권병기 첨단의료지원관은 “유전체 데이터 등 가명처리 방안을 구체화함으로써 정밀의료, 의료 인공지능 등 디지털 헬스케어 발전과 국민 건강증진에 도움이 되기를 기대한다” 라면서, “앞으로도 의료데이터 활용 현장의 애로사항을 지속적으로 경청하고 가명처리 관련 최신기술 동향 등을 반영하여 안전한 보건의료데이터 활용을 지원해 나가겠다” 라고 밝혔다.

- <붙임> 1. 「보건의료데이터 활용 가이드라인」 개요
 2. 「보건의료데이터 활용 가이드라인」 주요 개정내용
 3. 보건의료 가명정보 활용 관련 Q&A

담당 부서	첨단의료지원관 보건의료데이터진흥과	책임자	과 장	심은혜 (044-202-2940)
		담당자	사무관	박희정 (044-202-2941)



□ 추진배경

- 보건의료데이터의 안전한 가명정보 처리, 활용, 결합 방안을 제시하여 보건의료분야 가명정보 활용 활성화

* (최초발간) '20.9월, (1차개정) '21.1월, (2차개정) '22.12월

□ 주요내용

- (가명처리 방법) 개인 식별가능성을 낮추기 위해, 정형·비정형 데이터 유형별 개인식별요소 삭제 또는 대체 등 방안 제시
- (데이터제공심의방법) 가명처리 목적의 적합성, 가명정보 처리 적정성 평가, 안전성 확보 조치 사항 등을 검토하기 위한 기관보건의료 데이터제공심의위원회 구성·운영 절차 제시
 - * (기능) 가명정보 처리 목적의 적합성, 가명처리 적정성 등 검토
 - (구성) 5인 이상(외부위원 2인 이상)
- (제공·활용절차) 가명정보 활용 사례별(기관 내 활용, 제3자 제공, 가명정보 결합 및 반출) 절차를 제시
- (주요서식) 보건의료데이터 제공 신청 및 제공, 기간 관 계약서 및 부속 합의서 등의 표준 서식 수록
- (안전·보호조치) 법령상 규정된 안전성 확보조치 및 내부관리계획 수립·관리, 가명정보 분리보관·접근 관리, 가명정보 파기 등

구분	개정 전	개정 후
가명 처리 원칙	<신 설>	1. 개요 및 활용 원칙 ○ 보건의료 분야는 비정형 데이터가 다양하게 존재하고 있어 일률적인 가명처리 방법 및 기준 제시가 제한됨 - 따라서 개인정보처리자는 개인정보를 가명처리 하는 데에 소요되는 시간 및 비용, 인력, 기술, 방법 등을 고려하여 개인정보의 가명처리가 현실적으로 불가능하거나 개인식별 위험성을 낮추지 못하는 등의 경우에는 가명정보로 생성이 제한되는 것으로 판단하여 가명정보 활용 대상에서 제외하거나 정보주체의 동의를 받아 활용하여야 함
책임소재 명확화	<신 설>	○ 개인정보를 보호법에서 정한 목적에 따라 가명처리하고, 이에 따라 생성된 가명정보에 대해 안전조치 등 법률에서 정한 사항을 모두 준수하여 제공·활용한 경우, - 가명정보 처리·활용 과정에서 의도치 않게 특정 개인을 알아볼 수 있는 정보가 생성되었다는 사실만으로 가명정보를 처리한 자 또는 해당 가명정보를 제공한 자를 처벌하지 않음 ※ 단, 해당 정보의 처리를 즉시 중지하고, 지체없이 회수·파기하여야 함 ○ 가명정보를 제3자에게 제공할 시, 가명정보를 제공받은 자의 안전조치 미이행 등으로 가명정보 유출 등의 문제가 발생하였거나 고의로 재식별 행위를 하는 등 그 행위 주체의 위법으로 발생한 사안에 대해서는 해당 행위자만 제재함
의료 영상 데이터	④ (영상정보) 다음 중 필요한 조치를 이행하여야 함 - 영상정보 상에 환자번호·성명·생년월일 등 개인정보 일부를 삭제하거나 일부 또는 전부를 대체 - DICOM 헤더 등 메타데이터 상의 개인정보 일부를 삭제하거나 일부 또는 전부를 대체 - 영상정보 이미지 상의 특정 개인을 식별할 수 있는 신체적 특징, 문신, 보형물 등은 삭제 또는 모자이크, 표면 가장자리 삭제 등의 방법으로 처리 * 신체의 내외부를 함께 촬영한 단층촬영, 3차원 이미지(MRI, CT, 초음파 등의 경우	② 의료영상데이터 ◎ DICOM 등 메타데이터 상의 개인 식별 위험성이 존재하는 정보는 일부를 삭제하거나 일부 또는 전부를 대체하는 등 적절한 방법으로 처리 ◎ 의료영상데이터 이미지상의 개인 식별 위험성이 존재하는 정보는 일부를 삭제하거나 일부 또는 전부를 대체하는 등 적절한 방법으로 처리 ◎ 신체적 특징 및 문신, 보형물 등 의료영상데이터 이미지상의 특정 개인 식별 위험성이 존재하는 부분은 마스킹 등 식별성을 낮출 수 있는 적절한 방법으로 처리

구분	개정 전	개정 후
	3차원 재건 등의 기술을 활용하면 신체 이미지 복구 가능성이 존재하여 식별 위험 존재 ** 가명처리에 소요되는 시간, 비용 및 위험도, 기술 수준 등을 고려하여 식별성이 존재하는 특이정보를 포함한 정보는 제외하거나 정보주체 동의를 받아 처리할 것을 권장	◎ 얼굴 등 입체 영상 재건 기술을 통한 복원으로 개인식별 위험성이 존재하는 컴퓨터 단층촬영 등의 이미지 집합체는 목적 달성에 필요한 이미지만 선별하여 제공하거나 이미지 상의 신체 표면을 마스킹하는 등 적절한 방법으로 처리
유전체	⑥ (유전체 정보) 아래와 같은 몇가지 예외적인 경우를 제외하고 가명처리 가능유무 유보, 본인 동의 기반으로만 사용 ☑ 널리 알려진 질병에 관한 유전자 변이 유·무 또는 변이 유형 - 구체적 변이정보(예:Loc)가 아닌 큰 단위의 유전자(gene) 단위 정보로 제공함으로써 개인 재식별 가능성을 크게 낮춤 * (예시) A항암제 사용 시 B유전자 변이 환자의 치료반응 연구 ☑ 생식세포 변이 정보를 제거한 신생물 고유(neoplasm)*의 신규변이 정보 - 생식세포 변이(정상조직 변이)를 제거한 신규 생성 변이 정보는 암의 원인이 되는 변이 정보만 포함하게 되므로 개인 식별 가능성 없음 * 신생물(neoplasm): 세포의 이상증식 현상으로 종양(tumor)으로 알려져 있음	③ 유전체 데이터 ※ 본 가이드라인에서 안내하고 있는 방법은 인체유래물 연구 동의 또는 기증 동의를 받아 수집, 처리하고 있는 인체유래물등에 해당하지 않음 ◎ 의료기관에서 채취한 인체유리물을 NGS 기반 유전자 검사를 통해 생성된 SAM/BAM/VCF 파일 및 검사기록지는 아래와 같은 적절한 방법으로 처리 1) 염기서열 정보: 개인식별 위험성이 존재하는 회귀 변이정보(생식세포) 및 단연쇄 반복 정보 등은 처리 목적과 직접적으로 관계되지 않은 경우 삭제하거나 일부 또는 전부를 대체 하는 등 적절한 방법으로 처리 2) 위 염기서열을 제외한 정보: 데이터 형태별 메타데이터 또는 구조화 되지 않았으나 대상 파일 및 기록에 나열된 문자열 (또는 코드)이 기록된 개인식별 위험성이 존재하는 정보, 특이정보는 일부를 삭제하거나 일부 또는 전부를 대체하는 등 적절한 방법으로 처리 ◎ 의료기관에서 채취한 인체유래물을 NGS 기반 유전자 검사를 통해 생성된 원시 데이터인 FASTQ 파일을 이용하고자 할 때는 정보주체 동의를 받아 처리하는 것을 권고 - FASTQ 파일은 누구든지 염기서열 정보의 맵핑작업을 통해 각 시퀀싱 리드별로 표준참조염기서열에서 서열 정리된 염색체 번호 및 위치, 변이정보가 기록된 SAM/BAM/VCF 등의 파일 생성이 가능 ◎ 유전체 데이터는 염기서열 등 그 안에 담긴 정보의 내용을 모두 해석해 내지 못하고 있어 데이터 자체 식별 위험성을 낮추는데 한계가 존재하고, 부모·형제·자매 등 제3자의 정보를 담고 있을 수 있으므로 처리 환경에 대한 위험성 검토를 강화한 처리 환경의 제한(접근권한 관리 및 폐쇄 환경 등) 권고

구분	개정 전	개정 후
전 사체	⑦ (유전체를 제외한 오믹스 정보) - 유전체 정보와는 달리 대사체, 단백질체 등은 유전체 정보를 복구해 내는 것이 불가능 하므로 별도의 조치가 불필요함. 다만, 전사체 등은 유전체 정보가 복구될 가능성이 있으므로 가명처리 가능여부를 유보함	④ 유전체데이터를 제외한 오믹스 데이터 ○ 대사체, 단백질체는 유전체 정보를 복구해 내는 것이 불가능하므로 별도 조치 불필요. 또한 의료기관에서 진료목적으로 채취한 인체 유래물을 NGS 기반 유전자 검사를 통해 생성된 전사체의 발현값 데이터를 활용하는 경우 별도 조치 불필요 ○ 의료기관에서 진료목적으로 채취한 인체 유래물을 NGS 기반 유전자 검사를 통해 생성된 발현값 외의 데이터 중 개인식별 위험성이 존재하는 정보는 일부를 삭제하거나 일부 또는 전부를 대체 하는 등 적절한 방법으로 처리
자유 입력 데이터	② (의료인의 관찰·입력* 정보) - 의료 환경에서 의료인이 입력한 정보 중 정형화된 정보로써, 개인 식별자는 제외하거나 삭제하는 것을 전제로 함. - 정형화되지 않은 자유입력 정보는 안전한 가명처리 방법이 개발될 때까지 가명처리 가능여부를 유보함 * 예 : 진단코드, 주호소, 알레르기·과거력, 처방전 상 약제코드·용량·기간, 시술코드	⑤ 자유입력데이터(문자열) ○ 자연어 처리 기술 등을 통해 정형데이터로 변환 후 ① 정형데이터 가명처리 기준에 따라 적절한 방법으로 처리 ○ 정형데이터로 처리하지 않은 문자열 데이터는 전수 또는 키워드 검사, 정규 표현식 검사 등을 통해 개인 식별 위험성이 존재하는 정보를 확인하고 일부를 삭제하거나 일부 또는 전부를 대체 등 적절한 방법으로 처리 ※ 식별정보 등을 제외한 문자열 데이터에도 단어나 문장의 의미, 전후 사정, 별칭 등으로 인한 식별 위험성이 존재할 수 있으므로, 문자열 검사를 위한 키워드 목록 및 정규 표현식에 충분한 검토가 필요하며, 가명처리 적정성 평가 시에도 이에 대하여 검토하여야 함
음성 데이터	⑤ (음성정보) 가명처리 가능여부 유보 (본인 동의 기반으로만 사용 가능) * 예 : 진료 과정 음성녹음, 판독 과정 음성녹음, 환자 진단용 음성녹음 등	⑥ 음성 데이터 ○ 음성데이터를 음성인식기술을 이용하여 자유입력데이터(문자열)로 변환 후 ⑤ 자유입력데이터(문자열) 가명처리 기준에 따라 적절한 방법으로 처리 ○ 문자열로 처리하지 않은 음성데이터는 성문, 성조, 말투, 고유의 음성 주파수 등 음성 자체의 특성 및 음성에 담긴 내용으로 인해 개인 식별 위험성이 존재할 수 있으므로 일부를 삭제하거나 노이즈 추가 등 적절한 방법으로 처리 ※ 음성은 인공지능 화자인식 기술 등을 통해 개인 식별 위험성이 존재할 수 있어 주의가 필요함
보호법 개정 내용 반영	<신 설>	6. [5단계] 안전한 관리 ○ 가명정보 처리에 관한 사항 공개: 보호법 제3조제5항 및 동법 제30조, '개인정보 처리방침 작성지침(일반)'에 따라 가명정보 처리에 관한 사항을 기관 개인정보 처리방침에 기재하여야 함 ※ (참조) 「개인정보 보호법」 제3조, 제30조,

구분	개정 전	개정 후
	<신 설>	- 가명정보 파기 ○ 보호법 제28조의4제2항에 따라 가명정보의 처리 기간을 별도로 정하였다면, 처리 기한이 지난 이후에는 지체없이 보호법 제21조에 따라 가명정보를 파기해야 함 ※ 처리 기간을 별도로 정하지 아니한 경우에는 가명정보 처리 목적 달성 이후 지체없이 보호법 제21조에 따라 가명정보를 파기해야 함 ※ (참조) 「개인정보 보호법」 제21조, 제28조의4, 동법 시행령 제16조
서식 추가	<신 설>	부록 1. 참고자료 1. 관련 서식(예시) [가명정보 이용·제공 신청서 및 첨부서식] - 가명정보 이용·제공 신청서 - 활용데이터 선정 사유 및 요구 수준표 - 가명정보에 대한 안전성 확보 조치 계획서 - 가명정보에 대한 안전조치 의무이행 약속서 - 가명정보 처리 목적 증빙자료 서식 - 식별 위험성 검토 결과보고서 - 항목별 가명처리 계획서 - 가명정보 제공 및 활용 계약서 - 비밀유지의무 서약서 - 기관 보건 의료데이터심의위원회 심의관련 서식 보완 등 [가명정보 처리 관련 실무 서식] - 가명정보·추가정보 관리대장 - 가명정보·추가정보 접근권한 관리대장 - 가명정보·추가정보 파기 관리대장
FAQ	붙임 6. FAQ <신 설>	부록 5. 자주 묻는 질문(FAQ) Q12. 국민건강보험공단 등 공공기관에서 제공하는 가명정보는 해당 기관 내 분석센터(안심구역 등)에 방문해야만 이용 가능한지? A. 공공기관은 가명정보 처리와 관련하여 내부 관리계획을 수립해 안전한 가명정보 제공 절차를 마련하고 있으며, 가명정보를 기관 내 조성된 안전한 처리 환경에서 이용하도록 하거나 기관데이터 제공심의위원회 심의를 통해 기관 외 반출하여 제공하고 있음 * 온·오프라인(안심구역 등) 폐쇄망 활용을 우선 하되, 개인 식별위험이 낮은 데이터부터 점진적으로 기관 외 반출 권장

Q. 가명정보란?

○ 개인정보보호법(제2조1의2)에 따른 가명처리*를 거쳐 생성된 정보로서 그 자체로는 특정 개인을 알아볼 수 없도록 처리한 정보로서,

* 개인정보의 일부를 삭제하거나 일부 또는 전부를 대체 하는 등의 방법으로 추가정보가 없이는 특정 개인을 알아볼 수 없도록 처리하는 것

- 보건의료데이터도 가명정보의 처리에 관한 특례에 따라 과학적 연구, 통계 작성 및 공익적 기록보존 목적으로 활용될 수 있습니다.

Q. 이번 가이드라인 개정으로 달라지는 점은?

○ 기술발전 수준과 데이터 처리 환경 등을 고려하여 그간 가명처리에 제한이 있었던 유전체, 음성 및 자유입력(text) 데이터의 가명처리 방법을 구체화하였습니다.

- 유전체데이터의 가명처리와 관련하여, 전문가 검토(한국유전체정보학회)와 현장 의견수렴을 거쳐 유전체 데이터 내 가명처리 요소(생식세포 돌연변이, 희귀변이 및 단연쇄반복 정보 등) 처리 방법을 제시 하였습니다. 아울러, 정보보호를 위해 데이터 분석환경(접근 권한 관리 및 폐쇄 환경 등)을 강화할 것을 권고하였습니다.

Q. 모든 유전체 데이터를 가명처리하여 연구 목적으로 활용할 수 있는 것인지?

○ 본 가이드라인에 따른 유전체 데이터 가명처리는 주로 의료기관의 진료과정에서 발생하는 유전체 데이터에 활용될 수 있습니다.

생명윤리법에 따라 동의 기반하에 실시되는 유전체분석 연구나 인체유래물 등 별도의 연구 동의를 받는 데이터는 생명윤리법을 준수해야 합니다.

Q. 가명정보의 안전한 활용방안은?

- 개인정보처리자는 개인정보 보호법(제28조의2)에 따라 보건의료데이터를 가명처리하여 활용하고자 하는 경우 보호법 제28조의4에 따른 가명정보에 대한 안전조치 의무를 준수하여야 합니다.
- ① 개인이 식별되지 않도록 가명처리, ② 추가정보 분리보관 및 불필요시 파기 ③ 재식별 금지 ④ 그 외 안전한 분석 공간에서 활용 및 가명정보 활용 목적 달성 시 파기 등 물리적·관리적·기술적 안전조치 후 활용

Q. 가명정보 활용으로 무엇이 좋아지는지?

- 개인 식별성이 없도록 가명처리한 데이터를 과학적 연구 등에 활용함으로써 개인 맞춤 정밀의료, 의료 인공지능 기술개발 등에 활용될 수 있습니다.
- 데이터에 기반한 의료연구가 활성화됨으로써 궁극적으로 의료서비스의 질 향상과 국민건강증진에 기여할 것으로 기대됩니다.