

「금융 AI 데이터 라이브러리」 운영을 시작합니다

- 결합 데이터(가명정보)의 안전한 재사용 허용
- 효과적인 AI 학습·개발 등을 위한 양질의 금융 빅데이터 활용 촉진
 - 약 3.6억명의 고객·가입자(중복 포함)를 보유한
금융·핀테크·통신·데이터 전문기업 등 32개 기업·기관 참여
- 데이터 개방·공유를 통한 상생(相生)의 빅데이터 생태계 기반 구축

작년 하반기 발표한 「금융권 인공지능(AI) 활용 활성화 및 신뢰확보 방안」의 후속조치로 오늘 6월 21일(수), 금융위원회는 「금융 AI 데이터 라이브러리」를 혁신금융서비스로 지정하였다.

「금융 AI 데이터 라이브러리」는 가명정보의 안전한 재사용을 허용하는 데이터 인프라이다. 현재 금융회사, 핀테크 기업, 통신사, 신용정보회사, 데이터 전문기업 등 총 32개 기업·기관이 데이터 개방·공유를 통한 상생의 금융 빅데이터 생태계 구축을 위하여 동 컨소시엄에 참여하기로 하였다. 추가 참여 문의가 있는바, 향후 동 컨소시엄에 참여하기를 희망하는 기업·기관은 언제든지 컨소시엄에 참여할 수 있다.

< 현행 데이터 결합과 「금융 AI 데이터 라이브러리」 비교 >



2020년 신용정보법 개정을 통하여 가명정보와 데이터(가명정보) 결합 제도를 도입하였다. 가명정보는 특정 개인을 알아볼 수 없도록 처리한 정보로, 통계작성, 연구, 공익적 기록보존을 위해서는 정보주체의 동의 없이 가명정보를 이용할 수 있다. 아울러, 데이터 결합에 따른 재식별 우려 등을 해소하기 위하여 금융위원회가 지정한 데이터전문기관이 데이터 결합을 수행하도록 규정하였다.

AI 학습, 혁신적 서비스 개발, 신용평가 및 리스크관리 고도화 등을 위해서는 데이터 결합이 필수적이다. 현행 신용정보법령 및 「금융분야 가명·익명처리 안내서」에 따르면 데이터전문기관은 데이터 결합을 수행한 후, 결합 전·후 데이터를 모두 즉시 파기하여야 한다. 데이터 이용기관도 데이터 이용목적 달성 후, 결합 후 데이터를 파기하여야 한다. 이에 따라 중소 핀테크, 금융회사들이 대량·양질의 데이터셋을 구축·이용하는데 큰 애로가 발생하고 있다.

구체적으로, 데이터전문기관은 동일한 데이터를 여러 번 결합에 사용하는 경우에도 데이터 보유기관으로부터 매번 데이터를 전송받아 결합해야 한다¹⁾. 데이터 이용기관은 데이터 결합 신청시 명시한 이용목적만을 위해서만 해당 데이터를 사용할 수 있으므로, 데이터 결합을 통해 이미 확보한 데이터를 새로운 이용목적으로 이용하고자 할 경우 다시 데이터 결합을 신청해야 한다²⁾. 데이터 결합 신청시 명시한 기업·기관만 결합 데이터를 이용할 수 있으므로 결합 신청시 명시하지 않은 새로운 기관이 동일한 결합 데이터를 이용하고자 하는 경우에도 다시 데이터 결합을 신청해야 한다.

- 1) (예시) 데이터전문기관이 A+B, A+C, A+D를 수행하려는 경우에도 A를 저장하고 있을 수 없어 데이터 보유기관으로부터 A를 매번 전송받아야 함
- 2) (예시) '신용평가모형 고도화'를 목적으로 결합한 데이터를 '대출심사 전략'에 활용 불가

이러한 불필요한 데이터 결합에 과도한 시간과 비용이 소요되고 있다. 데이터 결합 신청부터 최종 데이터 결합 및 제공까지 평균 약 2개월이 소요되어 사업 추진 등이 지연될 우려가 있다. 또한, 데이터전문기관 및 데이터 보유기관의 운영비·인건비 등이 중복 발생한다. 반복적인 데이터 송·수신에 따른 정보유출 등 보안위험의 발생도 증가할 수 있다.

다만, 데이터 재사용을 전면 허용하는 경우 재식별 우려 등 개인정보보호 문제가 초래될 수 있어 우선 혁신금융서비스를 통해 테스트를 진행할 예정이다. 고도의 데이터 보호체계를 갖춘 신용정보원·금융보안원이 데이터를 저장·관리 하며, 데이터 재사용시 적정성 평가를 거칠 계획이다. 이를 통해 동일한 데이터 라도 데이터 이용목적, 데이터 이용기관의 개인정보 보호수준 등을 고려하여 가명처리 수준을 차등 적용*하는 등 재식별 우려를 해소할 수 있다.

* (예시) 정보보호 수준이 높은 기관에 대해서는 생년월일을 출생년도로 가명처리하고, 정보보호 수준이 상대적으로 낮은 기관에 대해서는 10세 단위 연령으로 가명처리

「금융 AI 데이터 라이브러리」 운영을 7월 중 시작할 계획이다. 중소 핀테크, 금융회사 등이 안전한 가명정보의 재사용을 통해 데이터 결합에 소요되는 시간·비용 등을 절약하고 금융권이 데이터를 적시성 있게 활용할 수 있을 것으로 기대된다. 아울러, 데이터 생태계 활성화와 양질의 빅데이터 확대를 통해 AI 학습·개발이 촉진되고 금융 AI 성능이 개선될 것으로 기대된다.

담당 부서 <총괄>	금융위원회 금융데이터정책과	책임자	과 장	신장수 (02-2100-2620)
		담당자	사무관	조윤수 (02-2100-2625)
<공동>	금융감독원 금융데이터실	책임자	실 장	김충진 (02-3145-7160)
		담당자	팀 장	심은섭 (02-3145-7162)
	한국신용정보원 빅데이터센터	책임자	센터장	이철흠 (02-3705-5433)
		담당자	팀 장	조근상 (02-3705-5966)
	금융보안원 데이터혁신센터	책임자	센터장	오중효 (02-3495-9900)
		담당자	팀 장	이광우 (02-3495-9910)



□ 규제특례 주요내용

❶ 데이터전문기관의 업무는 가명정보 결합, 익명처리 적정성 평가 등으로 열거(신정법 §26의4②, 영 §22의4⑦)

→ 데이터전문기관인 신용정보원이 데이터 저장·관리 등 라이브러리 구축·운영 업무를, 금융보안원이 결합키 관리기관 업무를 수행할 수 있도록 허용

❷ 데이터전문기관은 결합한 데이터를 결합의뢰기관에 전달한 후 결합전·후 데이터를 지체없이 삭제할 의무(영 §14의2③, 안내서 p.69)

→ 신용정보원은 결합 전·후 데이터를, 금융보안원은 결합키를 보관할 수 있도록 허용

❸ 결합키 생성방식은 결합의뢰기관 사이에만 공유되어야 하며 결합의뢰기관 외에는 공개 불허(영 §14의2③, 규정 §15의2②)

→ 컨소시엄 참여기관의 데이터가 효율적으로 결합될 수 있도록 금융보안원이 결합키 생성 방식을 조율할 수 있도록 허용

❹ 결합 신청시 결합목적과 이용기관을 구체적으로 명시할 의무(규정 별표8Ⅱ④, 안내서 p.94)

→ 결합 신청시 데이터 활용목적·주체를 특정하기 곤란하므로 결합목적은 '라이브러리 내 저장·재사용', 이용기관을 '라이브러리 참가기관 또는 참가기관이 데이터를 제공하기로 합의한 기관'으로 작성 허용

❺ 결합 신청시 결합데이터 이용기관으로 명시하지 않은 제3자에게 결합 데이터 제공 금지(안내서 p.70)

→ 결합 신청시 이용기관으로 명시하지 않은 기관도 금융 AI 데이터 라이브러리에 저장된 결합데이터 활용* 허용

* 적정성 평가 및 필요시 추가 가명처리 수행.

- ⑥ 신용정보회사, 신용정보집중기관, 금융회사 등은 가명정보 활용 관련 사항을 신용정보활용체제에 포함하여 공시할 의무(법 §31①, 영§27①,②)

→ 신정원은 가명정보를 직접 활용하는 것이 아니므로 라이브러리 관련 가명정보의 가명정보 활용 관련 공시의무를 배제

□ 부가조건 주요내용

- ① 데이터와 결합키가 함께 저장되어 발생할 수 있는 재식별 우려 등을 해소하기 위해 신용정보원은 결합키를 제외한 결합 전·후 데이터를, 금융보안원은 결합키를 저장·관리
- ② 신용정보원은 결합전 데이터를 최장 1년간 보관하고, 결합에 활용할 때 외에는 해당 데이터에 대한 접근·열람·가공 등을 금지
- ③ 신용정보원은 결합 전 데이터를 라이브러리에 저장하기 전, 가명처리 적정성 여부를 확인하고, 이용기관이 가명처리되지 않은 개인신용정보를 전송한 경우 즉시 삭제
- ④ 신용정보원·데이터전문기관은 데이터 결합 중 재식별 가능성이 발생할 경우, 결합을 중단하고 데이터를 즉시 삭제
- ⑤ 신용정보원은 결합 후 데이터를 이용기관에 전송하기 이전, 이용목적이 통계작성, 연구, 공익적 기록보존에 해당하는지 확인
- ⑥ 금융보안원은 결합키 관리업무와 데이터 결합업무를 수행하는 부서(팀)를 분리하고, 두 업무를 동시에 수행하여 발생할 수 있는 재식별 우려 등을 해소하기 위한 관련 보안대책을 수립

참고2

주요 용어 설명

- (가명정보) 특정 개인을 알아볼 수 없도록 이름·주민등록번호 등 개인을 식별할 수 있는 정보에 대해 삭제·범주화·암호화 등의 처리를 한 정보
 - 가명정보는 통계작성, 연구, 공익적 기록보존 목적을 위하여 정보주체의 동의 없이 활용할 수 있어, 양질의 빅데이터를 필요로 하는 AI 학습·개발 등에 활용하기 용이

- (결합키) 이름·전화번호 등을 암호화하여 생성한 정보로, 동일한 암호화 기법을 사용하는 경우, 동일한 사람에 대해서는 동일한 결합키가 생성됨

성명	전화번호	생년월일	대출금(만원)	→ 가명 처리	결합키	연령대	대출금(만원)
이순신	010-123-4567	1990.4.28.	3,600		HBD5919	30대	3,600
유관순	010-987-6543	1964.12.16.	15,200		QIG1846	50대	15,200
정약용	010-546-3728	1977.8.5.	5,300		ZOQ8842	40대	5,300

* 성명·전화번호를 사용해 결합키 생성, 재식별 우려가 있는 생년월일은 연령대로 범주화

- (데이터전문기관) 기업간 가명정보 결합을 지원하는 기관으로, 신정법에 따라 금융위가 지정한 데이터전문기관을 통해서만 가명정보 결합 가능
- (가명정보 결합) 두 개 이상의 기업이 가명정보를 데이터전문기관에 송부시, 데이터전문기관은 결합키를 활용하여 가명정보를 결합
 - 데이터전문기관은 결합키를 삭제 또는 다른 대체값으로 변환한 후 결합을 신청한 기업에게 결합된 가명정보를 전달

< 기업A >			< 기업B >		
결합키	가입보험	사고빈도(회/년)	결합키	주행거리(km/년)	과태료(원/년)
BQE1845	◇◇보험	2	BQE1845	5,400	110,000
GMZ5091	□□보험	0	GMZ5091	8,900	30,000
ENW3029	△△보험	5	ENW3029	18,700	52,000



< 데이터전문기관 >				
결합키 대체값	가입보험	사고빈도(회/년)	주행거리(km/년)	과태료(원/년)
AAA0001	◇◇보험	2	5,400	110,000
AAA0002	□□보험	0	8,900	30,000
AAA0003	△△보험	5	18,700	52,000

□ (결합키 관리기관) 결합키를 안전하게 생성·공유·관리하는 기관으로
신정법에 규정되어 있지 않으나, 라이브러리를 위해 도입

- 신정원에 결합키와 가명정보를 모두 저장하는 경우, 정보유출 등 발생시 재식별 우려가 높아지므로 결합키는 금보원이 그 외 데이터는 신정원이 저장·관리하도록 이원화 체계를 마련

< 이용기관A >

일련번호	결합키	대출액(만원)
1	ABC123	8,700
2	EDG423	15,762
3	GEC943	300

< 이용기관B >

일련번호	결합키	통신료(원)
1	GEC943	15,000
2	ABC123	45,000
3	EDG423	67,000


< 라이브러리 >

이용기관A		이용기관B	
일련번호	대출액(만원)	일련번호	통신료(원)
1	8,700	1	15,000
2	15,762	2	45,000
3	300	3	67,000

< 결합키 관리기관 >

이용기관A		이용기관B	
일련번호	결합키	일련번호	결합키
1	ABC123	2	ABC123
2	EDG423	3	EDG423
3	GEC943	1	GEC943

< 데이터전문기관 >

< 이용기관A Dataset>		< 이용기관B Dataset>		 데이터결합	< 결합키 연계정보>	
일련번호	대출액(만원)	일련번호	통신료(원)		이용기관A 일련번호	이용기관B 일련번호
1	8,700	2	45,000		1	2
2	15,762	3	67,000		2	3
3	300	1	15,000		3	1



< 결합후 데이터 >

대출액(만원)	통신료(원)
8,700	45,000
15,762	67,000
300	15,000