

# 학습분석 활용을 위한 데이터 수집 및 활용방안

서울시립대학교  
교수 이재호

# 발표 내용

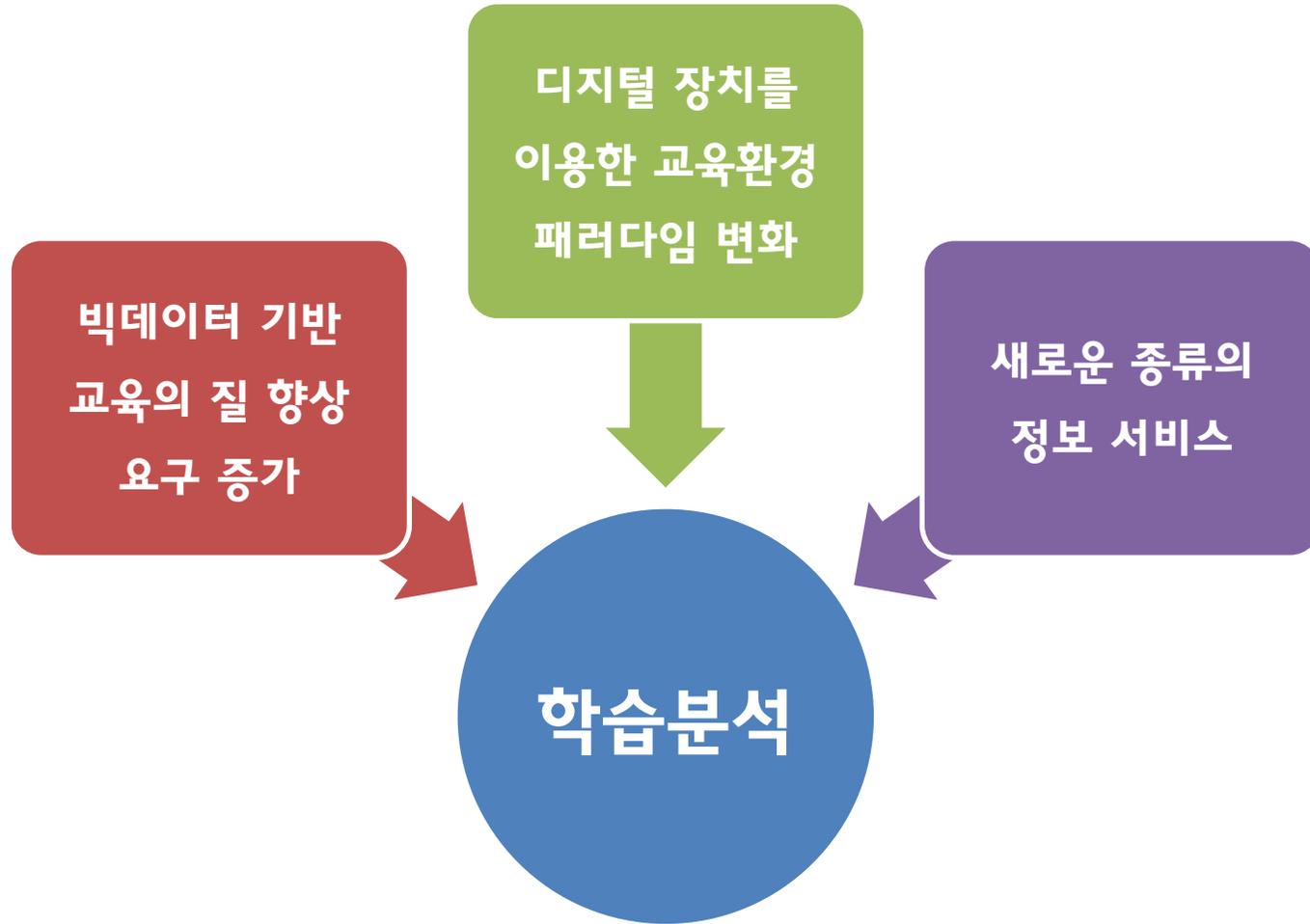
- 학습분석 개요
- 학습분석 정보 모델 및 수집
- 학습분석 분석 및 시각화



# 학습분석 개요



# 학습분석 환경



# 학습분석 - Learning Analytics

## 정의

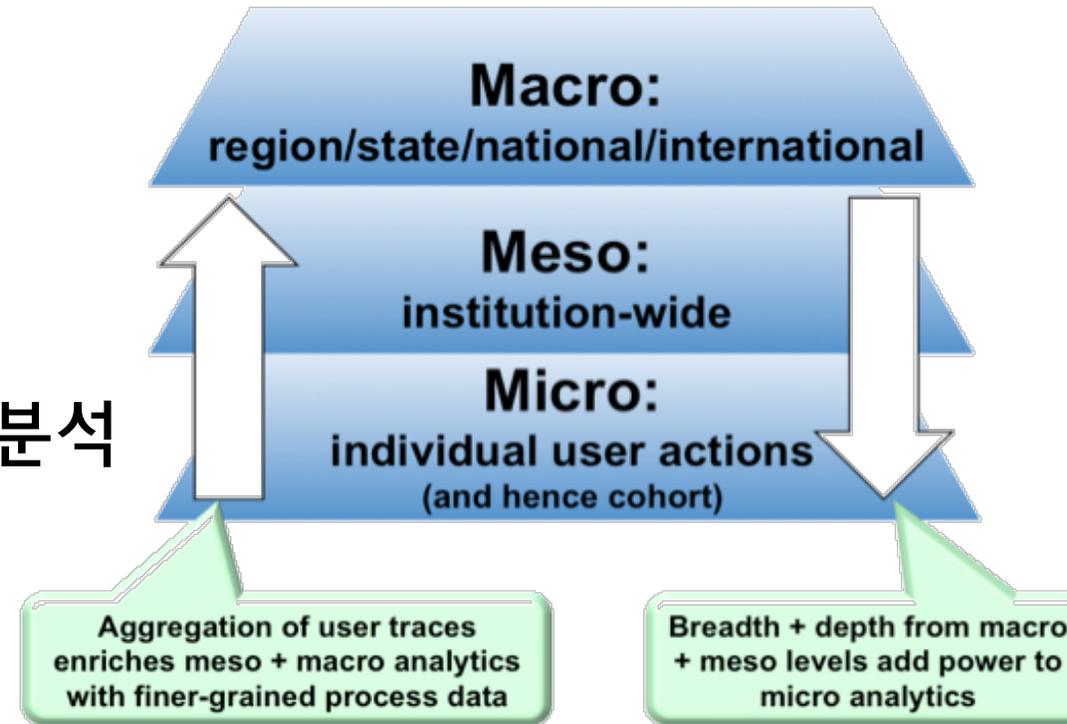
- 학습 환경에서 발생하는 이해하고 사용자에게 적합하게 제공할 수 있도록 하는 평가, 수집, 분석 및 표현 등 일련의 작업 과정

## 분석 대상

- 교육을 위한 다양한 콘텐츠 데이터
- 학습자가 생성하는 행동 데이터
- 연구자가 제시하는 분석 모델
- 교육과정에서 활용 가능한 소셜 분석 정보 등

# Levels of Learning Analytics

- 수준별 데이터 수집 대상
  - Micro : 개인의 데이터 수집
  - Meso : 지역 및 기관 단위 취합 및 분석
  - Macro : 국가별 및 국제적 데이터 비교분석
- 서비스
  - 개인화된 학습 결과 분석 및 콘텐츠 제공
  - 지역별 분석 결과 수집 및 정책 결정 지원
  - 학습 정책 및 콘텐츠 연구 개발 지원



# 학습분석의 필요성

- 학습자의 성취도를 높이기 위한 기반 정보 제공
- 학습자에게 개인화된 학습 환경 및 서비스 개발
- 학습자의 집중력 유지 및 향상
- 학습 방법 및 내용의 개선 및 향상
- 학습 관련 연구를 위한 다방면의 비용 감소

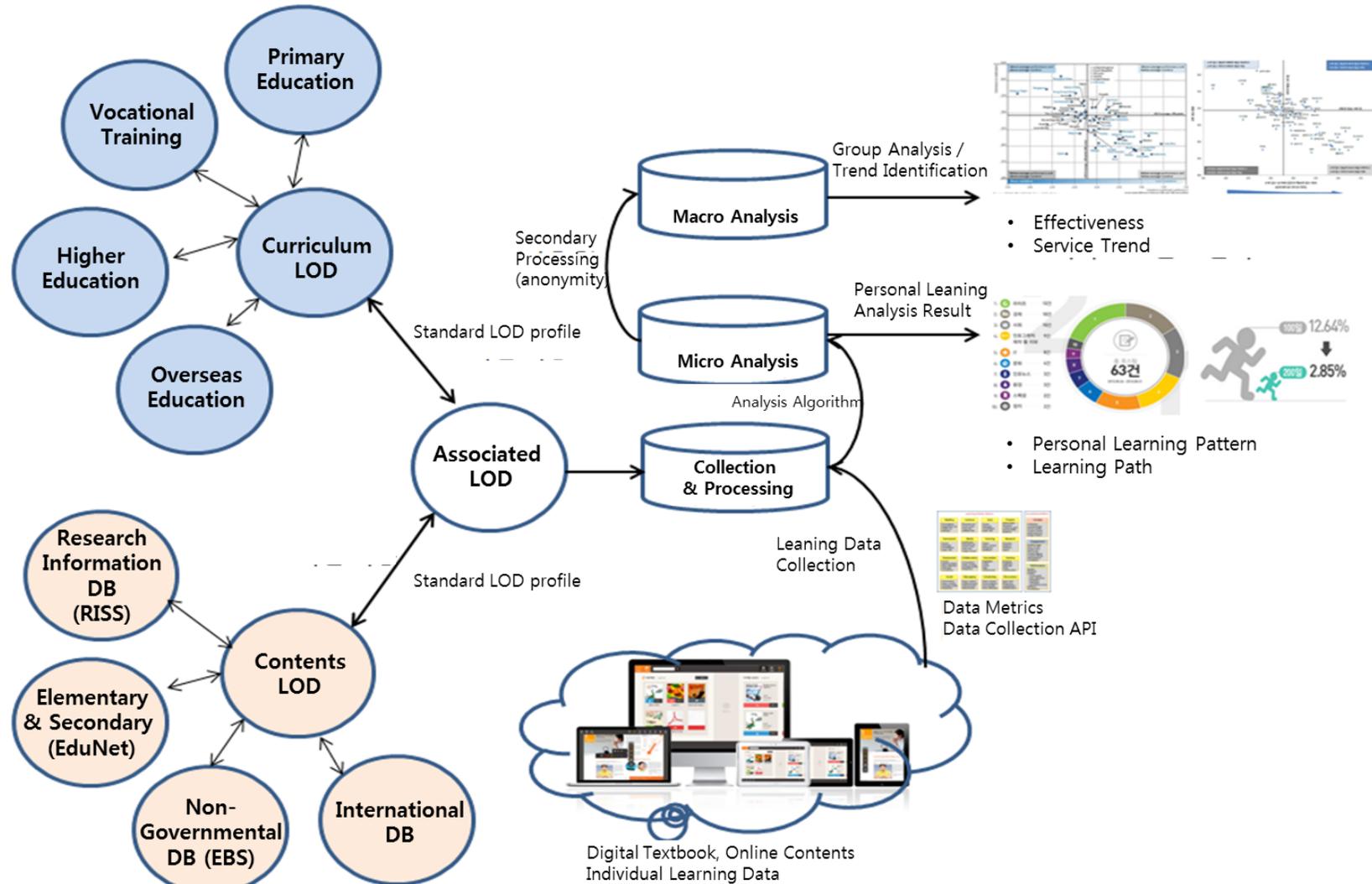


# 관련 연구

- IMS Global
  - Learning Analytics 분야를 개척하고 기술적 리더십 보유
  - 각국 학습 정보처리 분야의 기술 리더들의 지속적 논의 및 연구 수행
- 국내 과제 진행
  - 학습분석을 위한 국가적 리더십을 보유하기 위한 선도적 표준화 과제 수행



# 학습분석 개요



# 쟁점

- 정보의 수집
- 정보의 표현, 전송, 저장
- 정보의 분석
- 정보의 제공

# 학습분석 정보 모델 및 수집



# IMS Caliper

- IMS Caliper

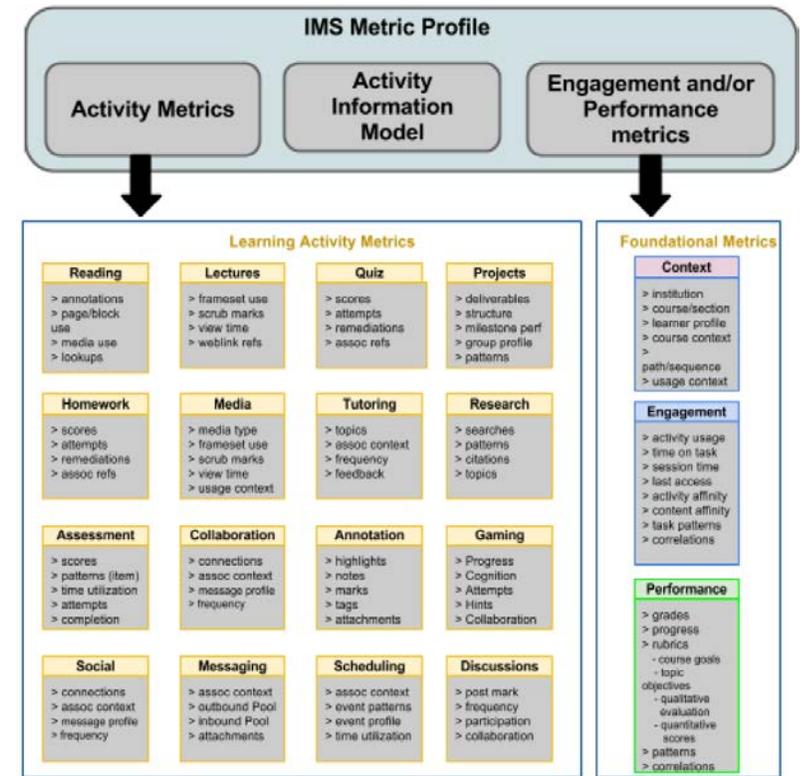
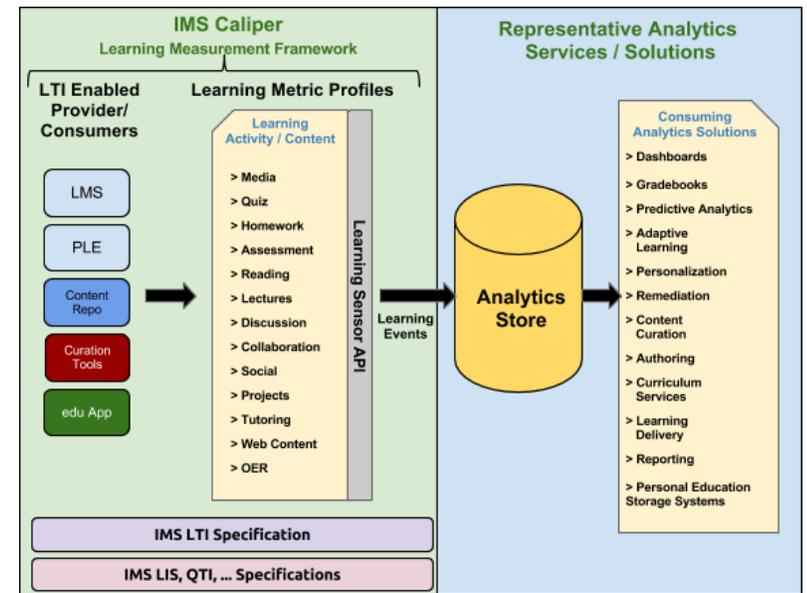
- LMS, PLE 등 표준을 준수하는 장치에서 발생하는 데이터를 수집하기 위한 기반 시스템

- Learning Metric Profile로 정의되는 다양한 데이터를

Sensor API를 통해서 수집하는 Workflow 제시

- Learning Metric Profile

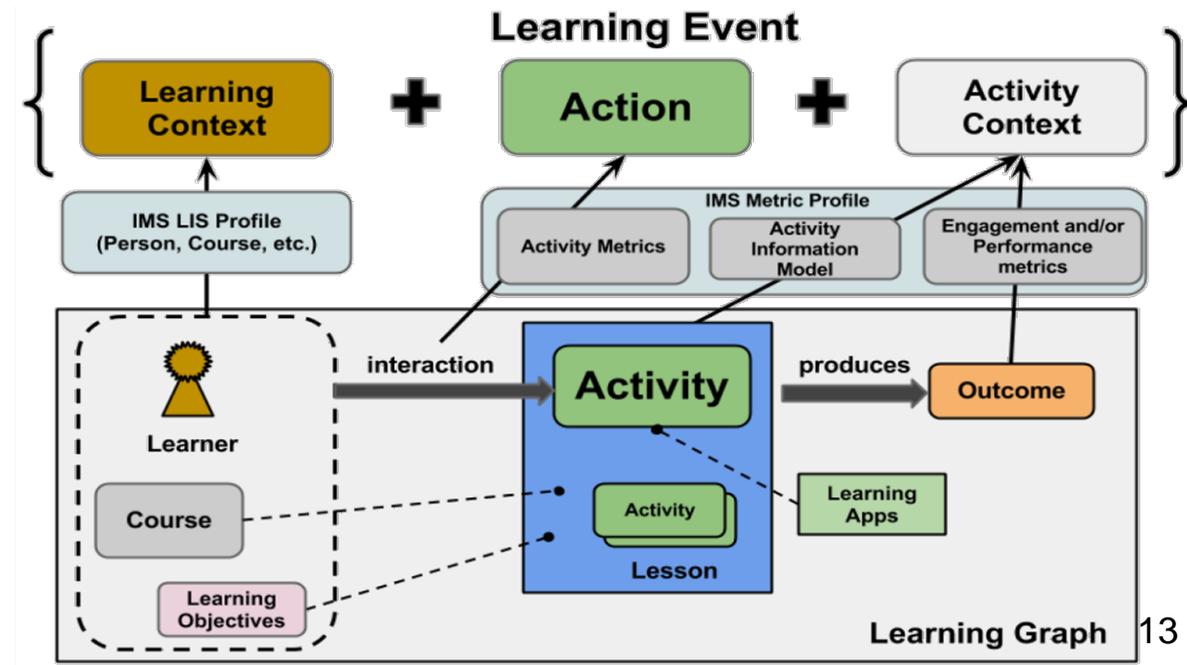
- IMS Caliper를 통해 학습 환경에서 발생하는 다양한 종류의 행동을 시스템 로그의 형태로 생성/수집하기 위한 표준 모델



# IMS Metric Profile

- 데이터 구성

- Learning Context : 학습자, 과정, 학습 활동 등 학습 행위의 주체
- Action : 읽기, 주석 달기, 문제 풀이 등 학습 중 일어나는 행위
- Activity Context : 전자책, 시험문제 등 학습 행위가 일어나게 되는 대상
- [무엇(누구)]이 [어떤 일]을 [무엇]에 수행하는 Triple 구조



# 데이터 수집을 위한 개선 방향

- IMS Caliper 기반 데이터 수집 체계의 한계
  - Log 형태의 데이터로서 다수의 Event간 중복 데이터 포함
  - Json 기반 수집을 전제로 LOD 등의 학습 콘텐츠와 연동 제한
- RDF 정보 서술을 위한 데이터 모델 정의 필요
  - 현존 IMS Caliper에서 RDF 데이터 모델에 적합하지 않은 아이템에 대한 변경 필요
    - RDF로 Mapping 되지 않는 데이터의 설계안 변경
  - Sensor API 수집 데이터의 형식 재정의
    - 현존 API를 통해 수집되는 데이터의 중복을 줄이고 수집 방법에 대한 간소화/완결성 재고

# RDF Triple 기반 데이터 수집 방안

## Json 기반 데이터

```
"@context": "http://purl.imsglobal.org/ctx/caliper/v1/AnnotationEvent",
"@type": "http://purl.imsglobal.org/caliper/v1/HighlightAnnotation",
"edApp": {
  "@id": "sample.edApp001",
  "@type": "http://purl.imsglobal.org/caliper/v1/SoftwareApplication",
  "name": "Contents Viewer",
  "description": "",
  "properties": {},
  "dateCreated": "1418283997868",
  "dateModified": "1418283997868"
},
"group": {
  "@id": "sample.group001",
  "@type": "http://purl.imsglobal.org/caliper/v1/lis/Organization",
  "name": "group",
  "description": "",
  "parentOrg": {
    "@id": "sample.group000",
    "@type": "http://purl.imsglobal.org/caliper/v1/lis/Organization",
    "name": "team a",
    "description": "",
    "parentOrg": {},
    "properties": {},
    "dateCreated": "1418283997868",
    "dateModified": "1418283997868"
  },
  "properties": {},
  "dateCreated": "1418283997868",
  "dateModified": "1418283997868"
},
"actor": {
  "@id": "sample.user001",
  "@type": "http://purl.imsglobal.org/caliper/v1/lis/Person",
  "name": "tester",
  "description": "",
  "properties": {},
  "dateCreated": "1418283997868",
  "dateModified": "1418283997868"
},
"action": "highlighted",
"object": {
  "@id": "sample.object001",
  "@type": "http://purl.imsglobal.org/caliper/v1/HighlightAnnotation",
  "name": "highlighted",
  "description": "",
  "parentOrg": {
    "@id": "sample.group001",
    "@type": "http://purl.imsglobal.org/caliper/v1/lis/Organization",
    "name": "group",
    "description": "",
    "parentOrg": {
      "@id": "sample.group000",
      "@type": "http://purl.imsglobal.org/caliper/v1/lis/Organization",
      "name": "team a",
      "description": "",
      "parentOrg": {},
      "properties": {},
      "dateCreated": "1418283997868",
      "dateModified": "1418283997868"
    },
    "properties": {},
    "dateCreated": "1418283997868",
    "dateModified": "1418283997868"
  },
  "properties": {},
  "dateCreated": "1418283997868",
  "dateModified": "1418283997868"
}
```

## RDF 기반 Property 데이터 + Activity 데이터

```
@prefix caliper: <http://purl.imsglobal.org/caliper/v1/>.
@prefix lis: <http://purl.imsglobal.org/caliper/v1/lis/>.
@prefix : <#>

:annotation.highlightAnnotation.examp001
  :context <http://purl.imsglobal.org/ctx/caliper/v1/AnnotationEvent>;
  :type caliper:HighlightAnnotation;
  :action "highlighted";
  :start "1418283990000";
  :end "1418283999999";
  :duration "9999".

:sample.epApp001
  :type caliper:SoftwareApplication;
  :name "Content Viewer";
  :created "1418283997868";
  :modified "1418283997868".

:sample.group001
  :type lis:Organization;
  :name "group";
  :created "1418283997868";
  :modified "1418283997868".

:sample.group000
  :type lis:Organization;
  :name "team a";
  :created "1418283997868";
  :modified "1418283997868".

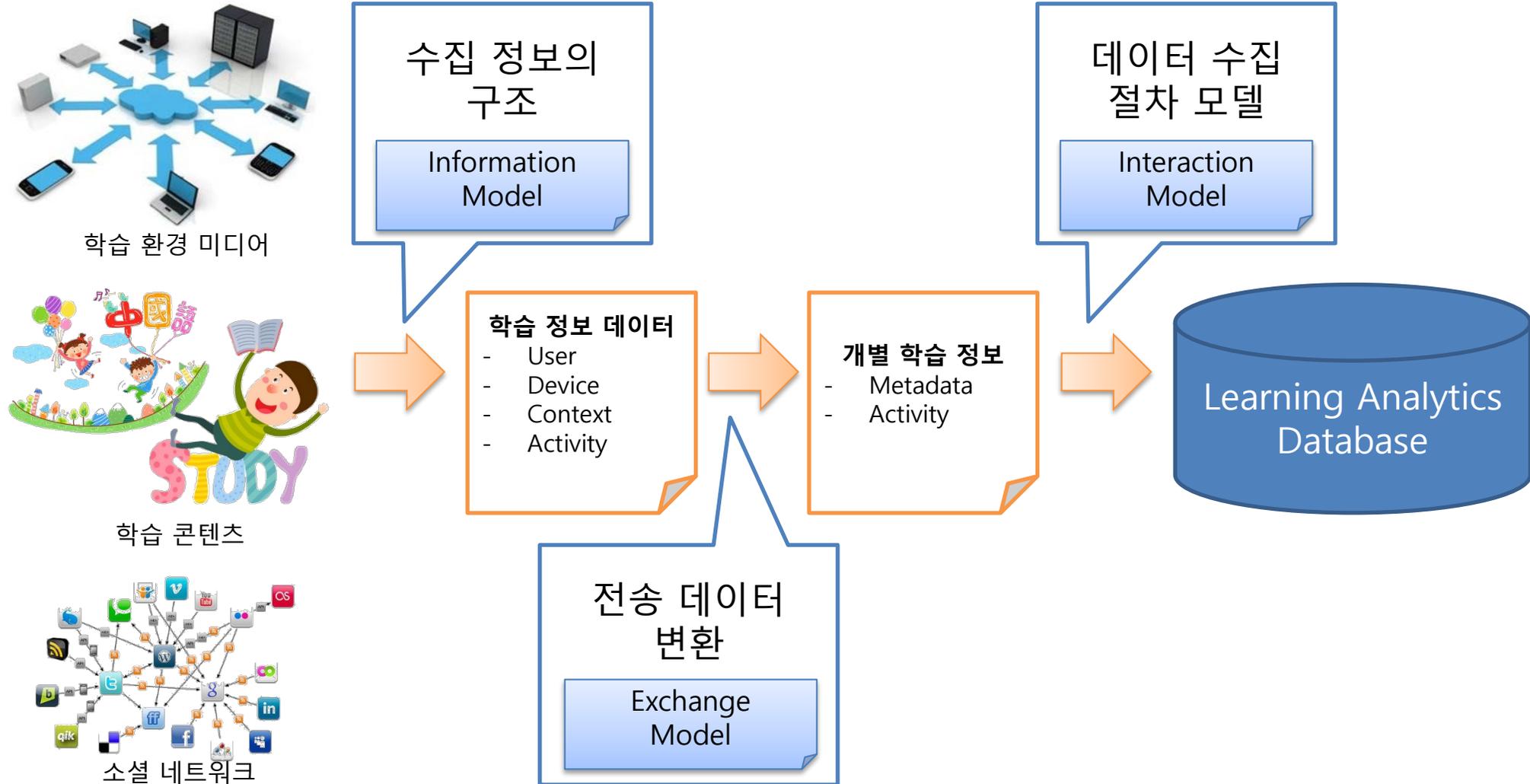
:sample.user001
  :type lis:Person;
  :name "tester";
  :created "1418283997868";
  :modified "1418283997868".

:sample.object001
  :type caliper:HighlightAnnotation;
  :name "highlighted";
  :created "1418283997868";
  :modified "1418283997868".

:annotation.highlightAnnotation.examp001.selection
  :selected :sample.object001.
```

```
@prefix : <#>
:sample.user001
  <http://purl.imsglobal.org/ctx/caliper/v1/AnnotationEvent>
  "annotation.highlightAnnotation.examp001";
  :group :sample.group001.
:sample.group001
  :part_of :sample.group000.
:annotation.highlightAnnotation.examp001
  :event_occured :sample.edApp001;
  :target :sample.digitalResource001;
  :object :sample.object001.
:sample.object001
  :selected :annotation.highlightAnnotation.examp001.selection.
```

# 데이터 수집 절차



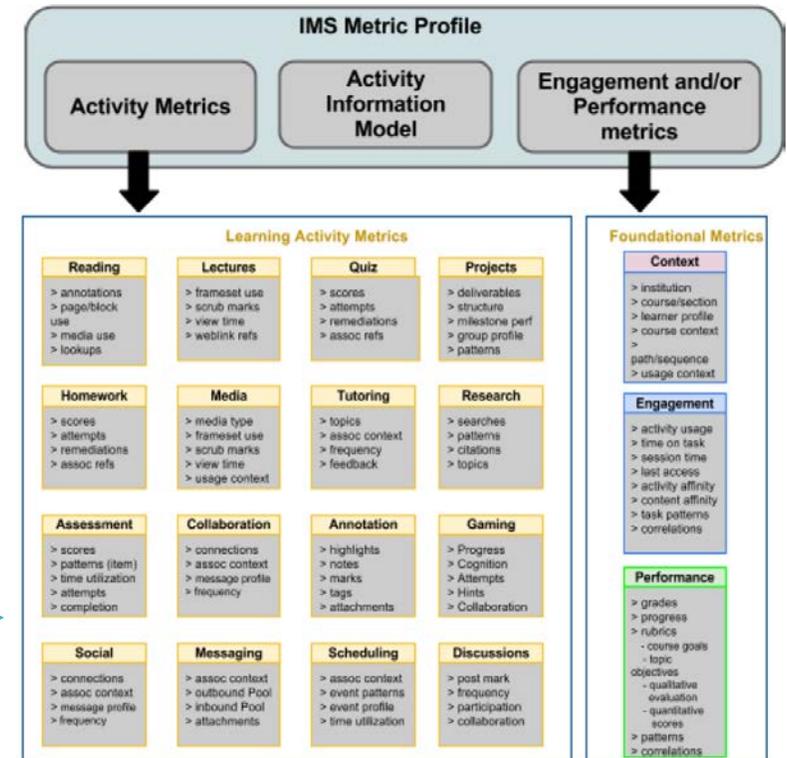
# 정보 모델 - Information Model

언어와 무관하게 개발하고 활용 가능한 표준적인 정보 모델

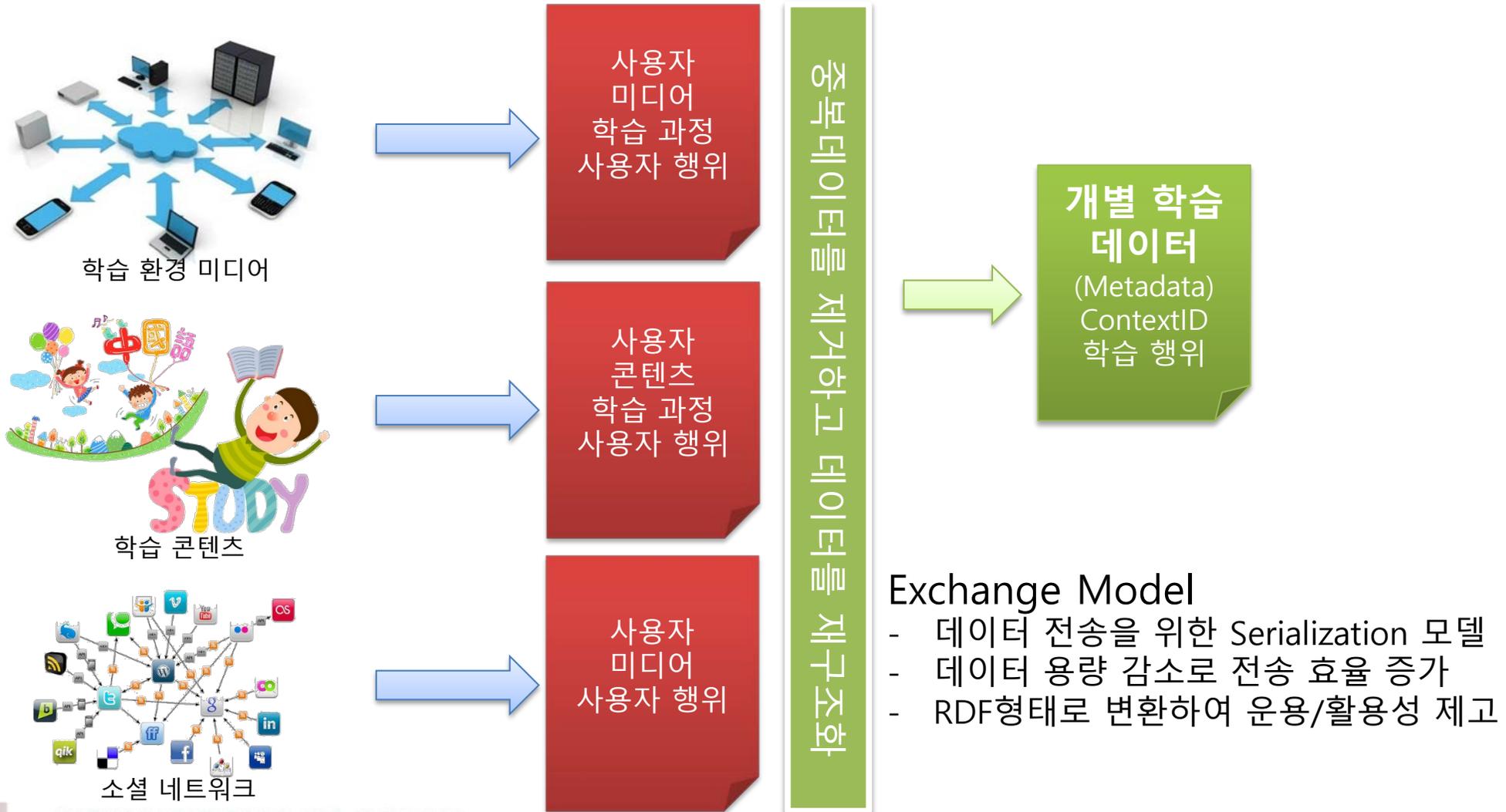
어떤 데이터를 생성하고 수집해야 할 지의 가이드라인 역할

데이터를 구조와 완결성을 위한 검증 수단으로 활용

생성/수집/검증을 위한 표준 정보 모델



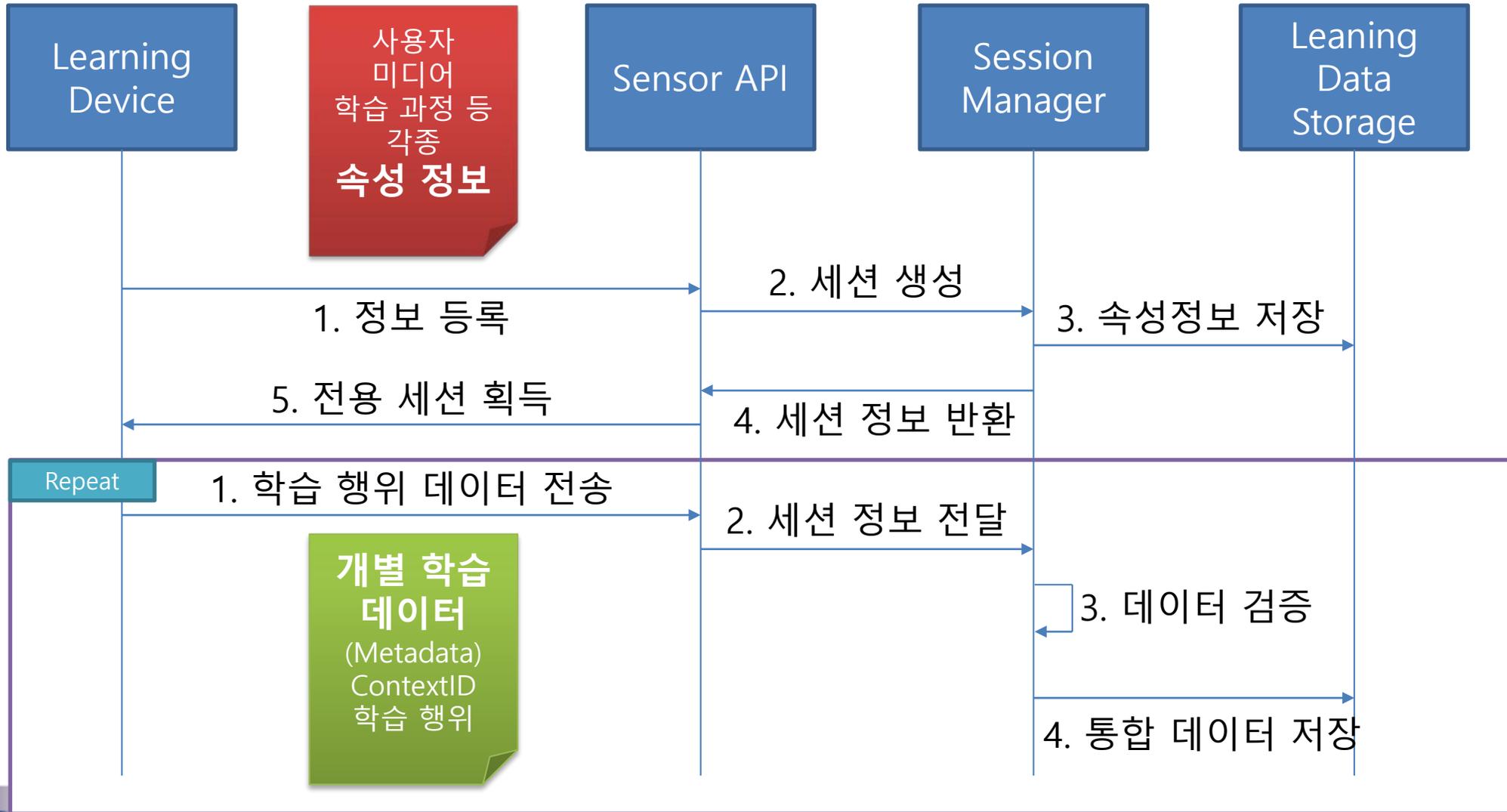
# 교환 모델 - Exchange Model



## Exchange Model

- 데이터 전송을 위한 Serialization 모델
- 데이터 용량 감소로 전송 효율 증가
- RDF형태로 변환하여 운용/활용성 제고

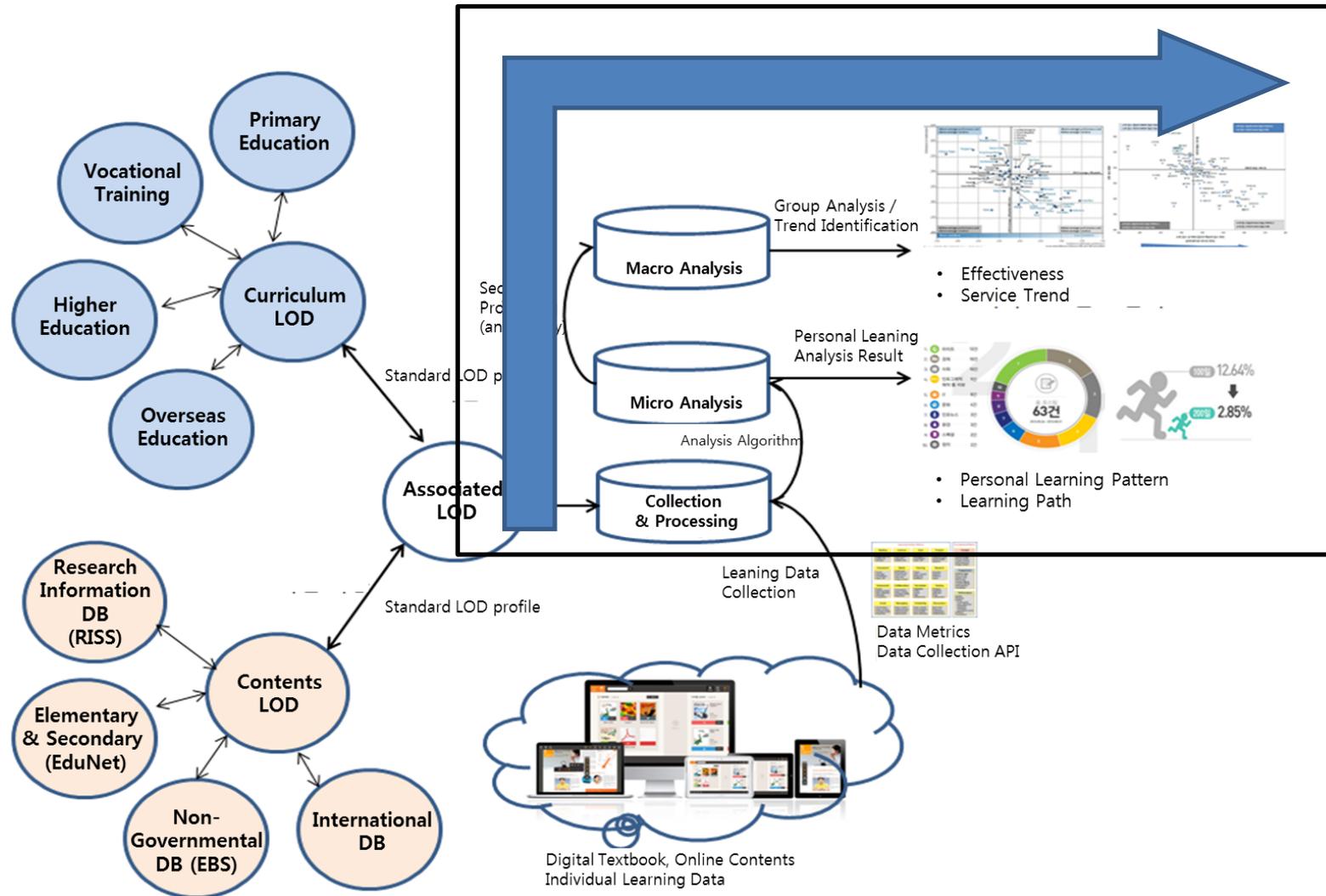
# 상호작용 모델 - Interaction Model



# 학습분석 분석 및 시각화



# Analysis & Visualization



# 학습 빅데이터 분석 시스템

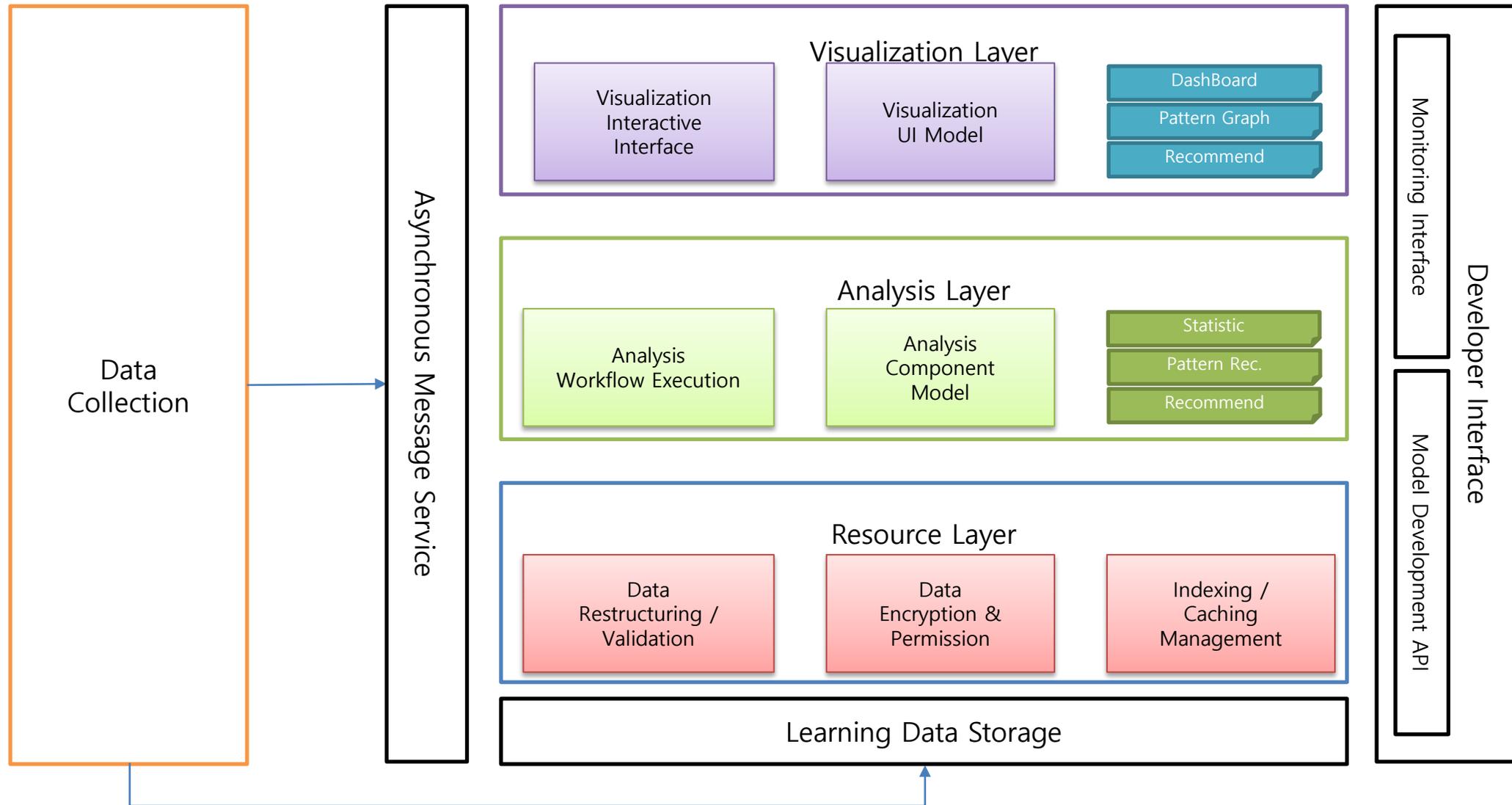
- 학습 분석 활용 예시

- 학습 플랫폼 분석 대쉬보드 : 학습자의 성취도 확인 및 타인과의 비교분석 자료 제공
- 예측 분석 : 학습 패턴을 통해 성취도를 예측하여 학습 가이드 라인 생성
- 개인화 학습 추천 : 학습자의 성취 경향을 통해 추천 학습 방법 및 콘텐츠 추천
- 소셜 네트워크 분석 : 소셜 환경내에서의 학습 연관성 분석 및 연구 지원
- 담화 분석 : 학습 환경내의 대화, 지도, 교습 등의 자연어를 통해 학습 과정 분석

# 학습 분석 요구사항

- **개방성 및 확장성:** 기존 시스템의 수정 없이 데이터 수집 방법 / 분석 알고리즘 확장 가능
- **분산환경:** 시스템이 활용하는 데이터 및 기능을 분산된 환경에서 저장/실행 가능
- **상호운용성:** 기존의 교육 환경이나 가상 교육 환경(VLE) 등의 시스템과 통합 운용 가능
- **재사용성:** 기능 컴포넌트나 데이터를 조합/재사용하여 복잡한 기능 설계 및 수행 가능
- **실시간성:** 실시간으로 분석이 수행되며 일정 시간 내에 분석이 완료 가능
- **활용성:** 사용자 상호작용을 통한 효과적인 시각화를 통해 분석 데이터 표현
- **보안 및 모니터링:** 개인 정보를 보호하기 위한 암호화 및 보호 수단이 필요하며, 작업이 수행되는 상황을 추적할 수 있어야 함

# 개념 아키텍처



# 컴포넌트 구성

- Resource Layer

- Distributed Task Execution

- 분산 환경에서의 작업 수행을 제어하고, 비동기적 / 병렬 작업 시 자원 분배 기능을 제공

- Resource Monitoring

- 분산 자원에 대한 작업 수행 상황을 추적하고 모니터링 할 수 있는 방법 제공
    - Development API를 통해 모니터링 정보를 확인 할 수 있는 Interface 제공

- Data Encryption

- 개인정보 등 보안이 필요한 정보를 관리하고 암호화하는 기능 제공



# 컴포넌트 구성

- Analysis Layer

- Analysis Workflow Execution

- 사용자의 요청에 따라 제한 시간, 데이터 범위 등의 제약조건에 따라 분석 workflow를 실행하고 결과를 종합하는 기능 제공

- Analysis Component Model

- 분석 알고리즘 컴포넌트를 모델화 하여 개발자가 새로운 모델을 추가하거나, 기존의 모델을 재사용하여 복잡한 workflow를 구성하는 등의 관리 기능 제공
    - Development API를 통해 개발자가 시스템에 모델을 관리할 수 있는 API 제공

- Built-in Models

- 데이터 통계, 패턴 추출, 관계 데이터 마이닝 등 기본 알고리즘 모델 제공



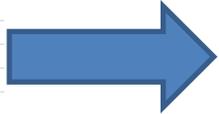
# 컴포넌트 구성

- Visualization Layer
  - Visualization Interactive Interface
    - 사용자 상호작용이 가능한 인터페이스 구성 제공
  - Visualization Component Model
    - 데이터의 시각화 방법을 모델링하여 사용자가 원하는 시각화 구성 선택기능 제공
    - Development API 를 통해 개발자가 모델을 추가/관리할 수 있는 API 제공
  - Built-in Models
    - Dashboard, Recommendation, Pattern Graph 등 기본 시각화 모델 제공



# 실행 예시

S	P	O
s01	solved	p101
s01	wronged	p102
s01	wronged	p103
s01	solved	p104
s01	wronged	p105
s01	wronged	p106
s01	solved	p107
s01	solved	p108
s01	solved	p109
s01	solved	p110
p101	part_of	ex1
p102	part_of	ex1
p103	part_of	ex1
p104	part_of	ex1
p105	part_of	ex1
p106	part_of	ex1
p107	part_of	ex1
p108	part_of	ex1
p109	part_of	ex1
p110	part_of	ex1
p101	about	number
p102	about	algebra



```

solved ~> read.csv("data/solved.csv")
wronged ~> read.csv("data/wronged.csv")
part_of ~> read.csv("data/part_of.csv")
about ~> read.csv("data/about.csv")

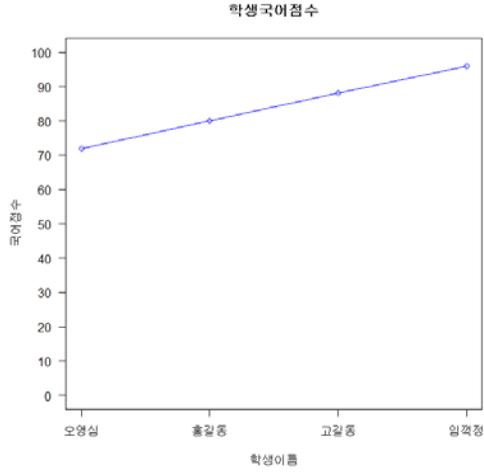
# Join identifier data
joined_data = merge(solved, wronged, by="id", all=TRUE)
joined_data = merge(joined_data, part_of, by="id", all=TRUE)
joined_data = merge(joined_data, about, by="id", all=TRUE)

# Analyze student performance
student_scores = aggregate(score ~ student_id, data=joined_data, FUN=mean)

# Plotting
plot(student_scores, xlab="학생이름", ylab="점수", main="학생국어점수", col="blue")

```

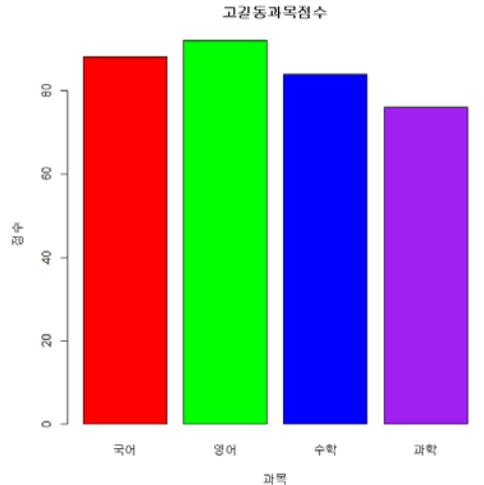
교사용 학생별 시험 결과



Rstudio Plotting Tool 이용한 시각화  
- 목적에 따른 Plotting 모델 적용



학생용 시험별 점수



## Triple 데이터 생성 및 저장

- 학생 시험 결과
- 시험문제 정보
- 학생 개인정보 일부

## R 기반 분석 스크립트 적용

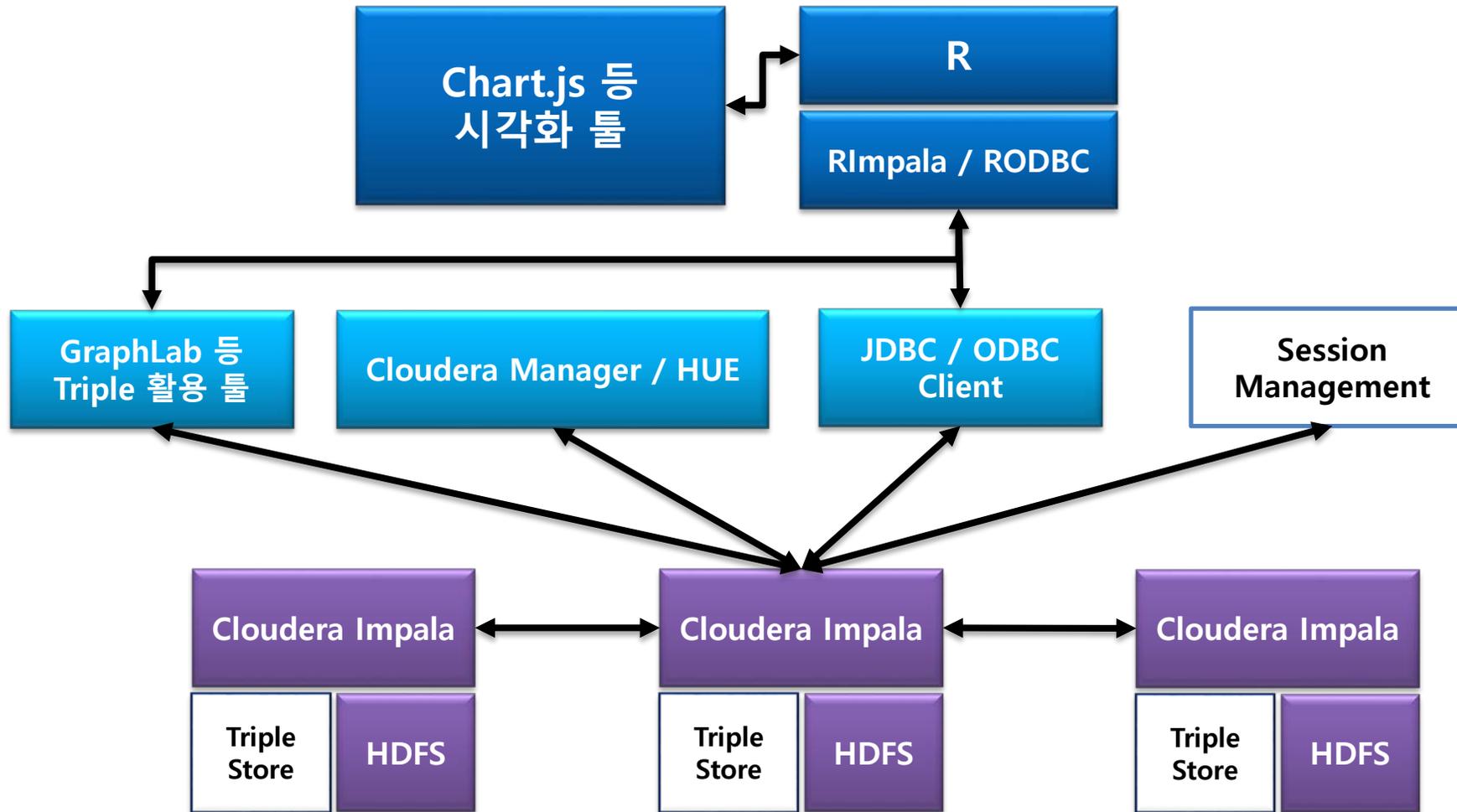
- Identifier 데이터 Join
  - 시험 정보 / 개인 정보
  - 개인 성적 등
- 분석 데이터 정렬 및 시각화 전처리

# 향후 계획

- 기존 시스템에 대한 개선점 분석 및 설계안 개선
  - 기존의 시스템에 대한 주요 미비점 보완 및 개선
    - 활용 데이터의 현실화 및 중규모 이상(수만~수십만건)의 데이터 처리 수행
  - 새로운 데이터 형식에 적합한 요구사항 적용 및 실험
    - RDF 변경에 따른 Multi Session 기반 데이터 수집 시스템 설계 및 분석 적용
    - 수집데이터에 적합한 DB 등 기능 컴포넌트 재구성



# POC 시스템 설계 개선안



# POC 시스템 설계 및 구축

- 신규 요구사항 적용 시스템에 대한 파일럿 데모시스템 구축
  - 기존 시스템+개선 RDF 구조에 적합한 설계 변경안 적용
- 중규모 이상 데이터에 대한 시스템 구축 및 실험
  - Multi-Node 기반 대용량 데이터 저장 및 분석을 위한 시스템 구성
  - 관련 실험 결과를 통한 표준안 기여

# 결론

- 학습 분석 시스템의 글로벌 표준화 제시
  - 학습 분석 표준 모델을 구축하여 표준안 고안 및 검증
  - 새로운 시스템에 대한 국제 표준 기여 및 확산
  - 오픈 소스 시스템으로 개발하여 국내외 산업 진흥 기여
- 개방적이고 활용성 높은 학습 분석 체계를 마련하여 학습 수준을 향상하고 새로운 환경의 학습 체계 연구 확산