

아이디어 기획서

팀명	데이터 별마당 드림팀(DSDT – Data StarField Dream Team)
팀원	<ul style="list-style-type: none"> ■ 팀장 : 서명준 ■ 팀원 : 신재우, 이성현, 임대균, 안희재
아이디어 주제	공공데이터 기반 표준 메타데이터 활용 프로젝트 중개 플랫폼
제안 배경 및 필요성	<ul style="list-style-type: none"> ■ 시장현황 <ul style="list-style-type: none"> - 전 세계 데이터 산업의 시장 규모는 급격히 성장하고 있으며 2021년 10조 5,340억 달러에서 2030년까지 연평균 11.9%의 성장을 할 것으로 예상되고 있다¹⁾. - 국내에서도 위와 같은 급격한 시장형성이 이루어지고 있으며 2021년부터 2023년까지 연평균 성장률은 8.9%가 예상된다²⁾. - 이에 따라 공공데이터의 공급을 통해 시장형성에 주력 해왔던 정부는 제1차 데이터 산업 진흥 기본계획을 통해 지금까지 정부 주도로 진행되었던 <u>데이터 산업의 활성화</u> 역할을 민간 주도로 전환할 예정이다³⁾. ■ 필요성 및 차별성 <ul style="list-style-type: none"> - 현재 운영 중인 디지털 집현전은 각 정부 기관에서 생성, 보유 중인 출판기록물의 검색(library)에 좀 더 집중된 성향을 보이고 있으며, 정부 데이터 검색은 공공데이터 포털을 활용하고 있다. 즉, 지식정보의 접근이 이원화된 상황이다. - 또한, 디지털 집현전은 분산 되어있는 여러 기관들의 데이터를 한데 모은 것에 중점을 두고 있다. - DSDT는 지식정보의 접근을 일원화하여 정보접근성을 개선함과 동시에 데이터를 분석하고 활용하여, 프로젝트 빌드업, 공유까지 가능한 One-Window 데이터 플랫폼을 지향할 것이다.
유사 제품 현황 및 비교	<ul style="list-style-type: none"> - 디지털 집현전 (데이터 셋 표준화 전문) - Kaggle (데이터 셋 & 프로젝트 공유) - Wishket (프로젝트 중개) - DACON (한국의 데이터 경진대회 진행 및 코드 공유)

1) grand view research, data preparation industry data book, 2023

2) 한국데이터산업진흥원, 2023 데이터산업 현황조사 보고서

3) 대한민국정부, 제1차 ('23~'25) 데이터산업 진흥 기본계획안

<p>제안 내용</p>	<ul style="list-style-type: none"> - 디지털 집현전을 벤치마킹하되, 클라이언트가 데이터를 응용하여 프로젝트 등에 활용할 수 있는 프로젝트 중개 플랫폼을 목표로 한다. - 클라이언트가 원하는, 분야별로 흩어져 있는 공공 데이터를 프로젝트에 쉽게 활용할 수 있도록 한 공간에 모아 제공하는 데이터 통합 서비스를 제공하고자 한다. - 더 나아가 클라이언트들의 프로젝트를 쉽게 공유할 수 있는 커뮤니티 기능을 제공하고자 한다.
<p>수행 방법</p>	<ul style="list-style-type: none"> ■ 데이터 확보 방안 <ul style="list-style-type: none"> - 공공데이터 포털에서 인기 검색어를 기준으로 분야 선정 - 공공데이터 포털 API, 클라이언트 데이터 활용 - 향후 회원으로 가입한 기업들과 개인들이 업로드한 데이터까지 포함하여 활용할 예정 ■ 데이터 통합화 작업 <ul style="list-style-type: none"> - 기관, 데이터 분야마다 분리된 필드를 통합하여 DB에 저장 ■ 웹페이지 구현 <ul style="list-style-type: none"> - 프로젝트에 활용될 표준화된 데이터를 시각적으로 명확하게 제공 - 클라이언트들끼리의 프로젝트 공유 커뮤니티 기능 구현
<p>기대효과 및 활용방안</p>	<ul style="list-style-type: none"> - 분산 되어있는 공공데이터의 접근을 일원화하여 사용자의 편의를 향상시킬 수 있다. - 하나의 사이트에서 검색한 데이터를 분석하고 활용하여 프로젝트 및 기획서 등에 바로 활용할 수 있으며, 커뮤니티 기능을 통해 유저 접근성을 더욱 높일 수 있다. - 공공데이터포털의 데이터를 연동함으로써 초기부터 양질의 데이터를 다량으로 확보할 수 있으며, 이를 통해 플랫폼의 규모 경제화를 노려볼 수 있다. - 향후 B2B, B2C 구독 서비스를 통해 추천이 많은 프로젝트의 소스 열람 기능을 유료로 제공하고 GoogleAD등을 통해 수익 모델을 만들 수 있다.

(1) 제안 배경 - 외부 환경 분석 (PEST / STEEP)

정책적 배경 ⁴⁾	<ul style="list-style-type: none"> ■ 2023. 1. 26. 제2차 국가데이터정책위원회 회의에서 제1차 데이터 산업 진흥 기본계획을 수립하여 <u>민간 주도 데이터 산업 생태계</u>를 위해 3가지 정책 방향을 제시했다. - 데이터 생태계 지원 통합인프라 구축 - 데이터 유통, 거래 제도의 혁신 - 민간 주도의 데이터 유통 거래
경제적 배경 ⁵⁾	<ul style="list-style-type: none"> ■ 2023년(추정) 기준 데이터 산업 시장은 전년 대비 4.6% 성장한 27조 1,513억원으로 계속해서 성장 중이다. ■ 산업분야 중분류 중 데이터 판매 및 제공 서비스업의 비중이 49.1% (13조 3,352억원)에 달해 가장 큰 비중을 차지하고 있다. ■ 노동 시장적 측면에서도 산업 전체에서 데이터 직무 근로 인력이 207,027명에 달하며 향후 5년 내 추가로 필요한 인력이 29,967명으로 예측될 만큼 성장 중인 산업 분야이다.
사회적 배경	<ul style="list-style-type: none"> ■ 현대 사회에서 데이터 생성 속도와 양이 폭발적으로 증가했으며, 다양한 형태의 데이터를 저장하고 분석하는 빅데이터 기술은 매우 중요하다. ■ 이러한 시대적 흐름에 맞는 핵심 역량을 개발하는 과정에서 클라이언트는 용어적 통일이 되지 않은 공공기관 데이터의 표준화 부족으로 인해 신뢰할 수 있는 자료를 찾기 어려운 경우가 많다. ■ 우리는 이러한 문제를 해결하기 위해 공공데이터의 단어를 표준화하고, 이를 기반으로 프로그래밍 프로젝트 관련 문서 작성 및 공유를 제공하는 웹사이트를 만들고자 한다.
기술적 배경	<ul style="list-style-type: none"> ■ 데이터 처리 및 분석 기술의 발전 - 각 분야에서 발생하는 빅데이터를 정리하여 각 공공 기관에서 정리하고 있고, 꾸준히 업데이트 작업을 통해 최신화되고 있다. - 이러한 데이터들은 메타데이터와 OpenAPI를 이용하여 간단하게 가져올 수 있게 되었다. ■ API 및 데이터 공유 기술의 발전 - OpenAPI는 데이터 접근성과 상호 운용성을 높이는 데 중요한 역할을 하며, 이를 통해 개발자들이 쉽게 데이터를 접근하고 사용할 수 있다. - 오픈소스 데이터 포털 플랫폼은 데이터 세트의 저장, 관리, 공유를 간편하게 할 수 있게 해주며, 공공데이터의 투명성과 접근성을 크게 향상시킨다.

4) 대한민국정부, 제1차 ('23~'25) 데이터산업 진흥 기본계획안

5) 한국데이터산업진흥원, 2023 데이터산업 현황조사 보고서

(2) 제안 배경 - 내부 환경 분석 (3C)

제안자 능력	<ul style="list-style-type: none"> ■ 데이터 분석: Python, Pandas, Matplotlib, Seaborn ■ 프론트엔드: HTML, CSS, JS, JSP ■ 백엔드: Java Spring, MySQL
경쟁제품 /기술 분석	<ul style="list-style-type: none"> ■ Kaggle(데이터셋 공유 및 경진대회 진행 및 코드 공유) ■ 데이콘(한국의 데이터 경진대회 진행 및 코드 공유) ■ 위시켓(프로젝트 의뢰 및 등록, 팀원 모집)
고객 분석	<ul style="list-style-type: none"> ■ 프로젝트 작성을 위해 참조할 공공데이터를 여러 사이트를 돌아다니면서 찾고 있는 학생(안희재) ■ 25세 데이터 과학자로 경력을 시작하는 학생(임대균) ■ 각 데이터마다 다른 API사용법으로 인해 혼란을 겪는 학생(이성현) ■ 주야간 통행량과 범죄율의 상관관계 파악을 통해 치안 유지계획을 세우려 하는 경찰공무원(신재우)

(3) 제안 배경 – SWOT분석

Strength	<ul style="list-style-type: none"> - 데이터 통합화를 통한 정보 접근성 개선 - 프로젝트 기획서 단계부터 공유 단계까지의 모든 과정을 통합하는 One Window서비스 실현
Weakness	<ul style="list-style-type: none"> - 데이터를 전담할 전문인력 부족 - 시스템 초기 구축에 대한 초기 비용 발생 - 사용자 교육 및 적응 기간 필요 - 유지보수로 인한 추가 인력 필요
Opportunity	<ul style="list-style-type: none"> - 공공DB를 사용하고 있는 대한민국 모든 국민, 기업, 기관들의 데이터 전처리 작업에 소요되는 시간 및 비용 절감이 가능 - 데이터 통합화는 빅데이터 처리 및 분석의 트렌드가 될 수도 있음
Threat	<ul style="list-style-type: none"> - 카피 프로젝트로 인해 웹사이트의 정보의 질 저하가 우려됨 - 보안 문제가 발생할 경우, 사용자의 신뢰도 하락이 발생할 수 있음

(3) 필요성

필요성	<ul style="list-style-type: none"> ■ 통합화되어 있지 않은 공공 데이터를 분야별로 통합 후 저장하여, 클라이언트가 편리하게 데이터를 이용하고 공유할 수 있는 중개 서비스가 필요하다.
차별성	<ul style="list-style-type: none"> ■ 여러 분야별 공공데이터를 통합화시키고 제공한 후, 이를 중개 서비스라는 비즈니스 모델을 만들어서 서비스한다. ■ 공공 데이터만을 검색 · 조회하는 것에서 멈추지 않고, 이를 클라이언트가 쉽게 열람하고 활용할 수 있는 커뮤니티 기능을 제공한다. ■ 통합한 공공데이터를 참조하여 만든 프로젝트를 클라이언트 간 공유할 수 있는 중개 서비스를 제공한다.
현재까지 준비사항	<ul style="list-style-type: none"> ■ 기획 및 ERD작성 ■ 프로젝트의 메인페이지, 공공데이터게시판, 프로젝트게시판, 자유게시판, 자료실, 공지사항 각각의 기능들, 웹페이지 기능들을 안내하는 AI 챗봇 구현
기대효과	<ul style="list-style-type: none"> ■ 개인 및 기업의 프로젝트 제반 사항 초기사항을 지원할 수 있다. ■ 공공 DB를 사용하고 있는 대한민국 모든 국민, 기업, 기관들의 데이터 전처리 작업에 소요 되는 시간과 비용을 절감할 수 있다. ■ 데이터의 통합화는 국가가 지향해야 할 빅데이터 처리 및 분석 방향이 될 수 있다.
활용방안	<ul style="list-style-type: none"> ■ 프로젝트를 등록한 개인의 입장에서는 포트폴리오 & 이력서가 될 수 있다. ■ 사이트에 등록된 클라이언트들의 프로젝트를 바탕으로 기업의 채용 시스템과 연계할 수 있다. ■ 스타트업의 경우 새로운 사업 구상에 대한 기반을 제공받을 수 있다.

(4) 개발 내용

개발 목표	<ul style="list-style-type: none"> ■ 공공데이터포털에서 제공하는 오픈API를 활용하여, 데이터를 수집하고 통합화하여 DB에 저장한다. ■ 이렇게 통합화 작업을 진행한 데이터를 활용하여 클라이언트에게 제공할 수 있는 프로젝트 중개 서비스 플랫폼을 구축한다.
개발 내용	<ol style="list-style-type: none"> 1. 공공데이터 포털의 오픈 API를 활용하여 데이터를 추출한다. 2. 추출한 데이터의 통합화 작업을 진행하고, 서버 DB에 업데이트한다. <ul style="list-style-type: none"> - 원본에서 증분, 수정데이터가 발생 시, 서버 DB에 주기적으로 동기화 3. DB에 저장된 데이터를 바탕으로 프로젝트 중개 커뮤니티 서비스를 구현하고, 배포한다. <ul style="list-style-type: none"> ■ 회원 관리 기능 <ul style="list-style-type: none"> - 회원가입, 로그인, 로그아웃 ■ 프로젝트 관리 기능 <ul style="list-style-type: none"> - 기획서 표준 양식 지원 - 프로젝트 게시판 구현 - 프로젝트 좋아요, 댓글 등의 커뮤니티 기능 구현 4. 작동 여부를 테스트하고, 추가 서비스를 기획한다.

(5) 달성 목표 및 달성 전략

<p>달성 목표</p>	<ul style="list-style-type: none"> ■ 1차 목표 <ul style="list-style-type: none"> - 각 분야의 공공데이터를 추출하여 통합화 작업 수행 - 증분, 수정데이터의 업데이트 확인 - 통합화한 데이터를 조회할 수 있는 웹 페이지 구현 ■ 2차 목표 <ul style="list-style-type: none"> - 로그인 및 게시판 기능 구현 - 프로젝트 공유, 리뷰 서비스 기능 구현 - 회원 등급에 따른 개인과 기업의 서비스 구현 - 최종적으로 BM 모델 확립 및 테스트, 운용
<p>달성 전략</p>	<ul style="list-style-type: none"> ■ 공공데이터 통합화 <ul style="list-style-type: none"> - 공공데이터를 오픈API와 Python을 활용하여 데이터 통합화 작업 - Python으로 데이터를 통합화하여 MongoDB에 저장 - 증분 & 수정데이터가 발생 시, S3에 주기적으로 업데이트 ■ Spring Framework와 MVC패턴을 사용하여 웹 사이트 서비스 구현

(6) 개발 일정

	1주차(5/31 ~ 6/6)	2주차(6/7 ~ 6/13)	3주차(6/14 ~ 6/20)	4주차(6/21 ~ 6/28)
계획 및 자료조사	○			
기존 제품 분석	○			
설계	○	○		
구현 및 점검		○	○	
시연 및 완성				○

■ 최종 발표 : 6.27(목)

(7) 참여 인원

이름	역할 및 능력
서명준	<ul style="list-style-type: none"> ■ 공공데이터 가공 및 통합화 작업 ■ DB 설계 및 관리 ■ 증분, 수정 데이터 업데이트 ■ 프로젝트 버전 관리
임대균	<ul style="list-style-type: none"> ■ 회원 관리 기능 - 회원가입, 로그인
안희재	<ul style="list-style-type: none"> ■ 회원 관리 기능 - 로그아웃, 회원 등급 관리 ■ AWS로 서버 배포 및 관리 담당 - 인프라 작업
신재우	<ul style="list-style-type: none"> ■ 게시판 기능 - 메인페이지 및 각 게시판 Back-End 구현 - 게시판 데이터베이스 설계 및 관리
이성현	<ul style="list-style-type: none"> ■ 게시판 기능 - 메인페이지 및 각 게시판 Front-End 구현