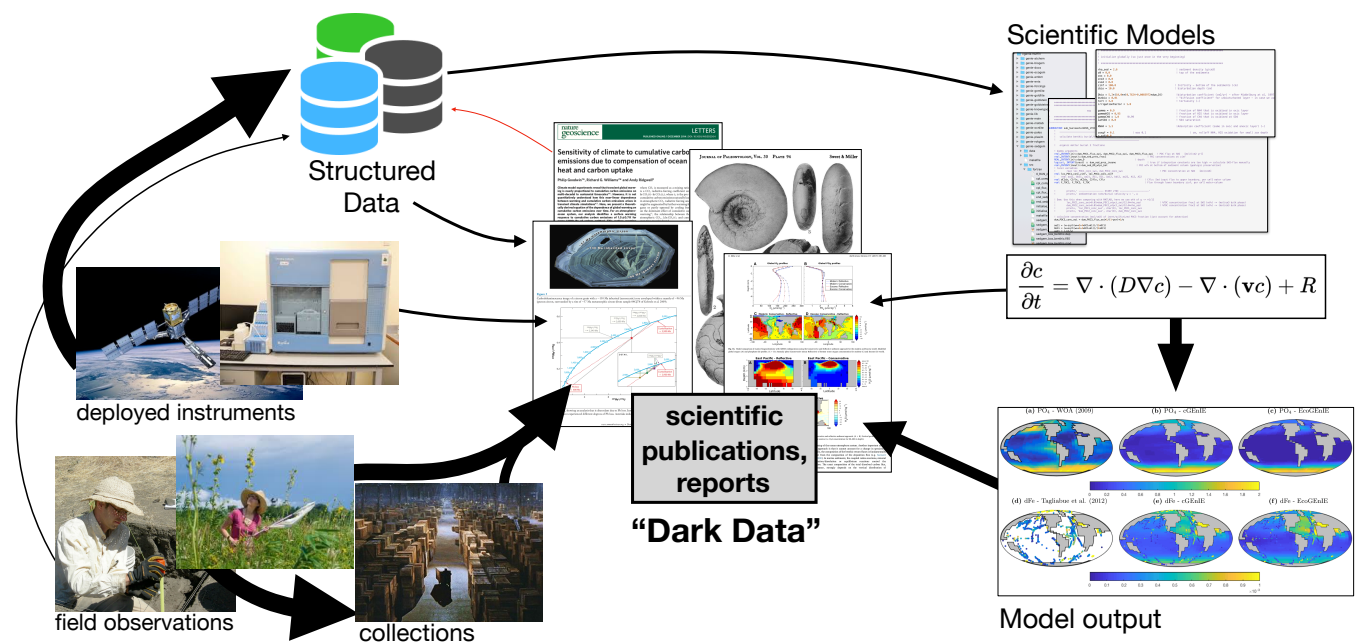


# COSMOS: An AI platform for knowledge extraction from scientific publications

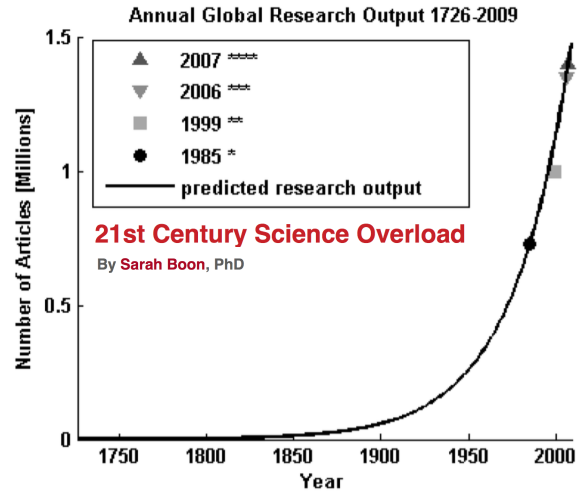
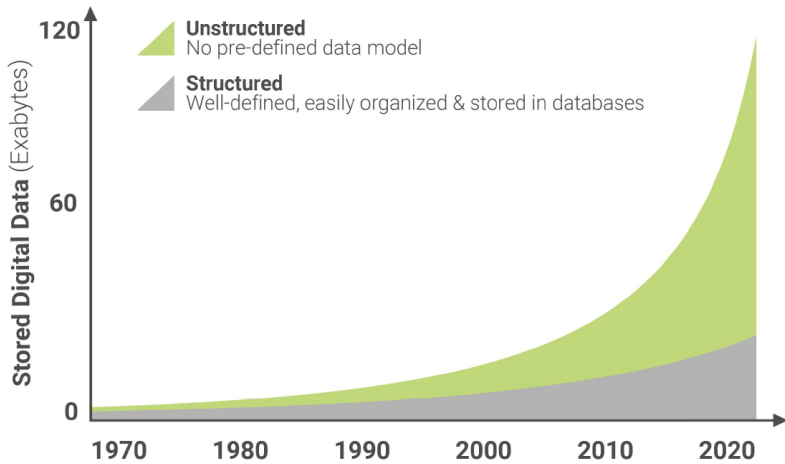
Theo Rekatsinas, Shanan Peters, Miron Livny



## Publications are central to the scientific process

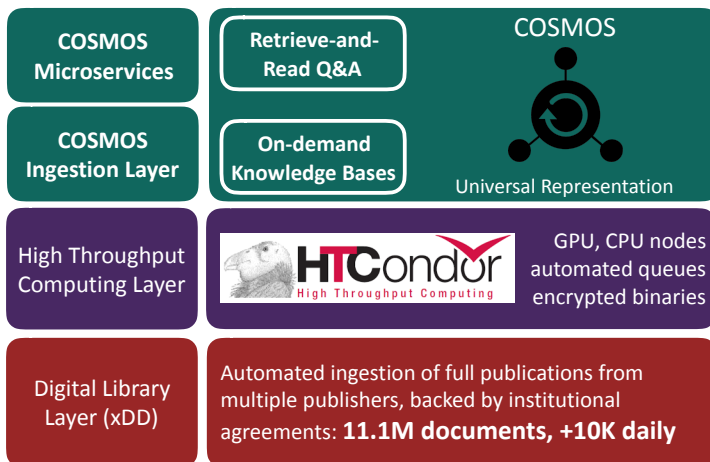


# The gap between scientists and data is increasing



Automated **knowledge extraction** from unstructured data is required to close this gap and accelerate scientific progress.

## xDD and COSMOS: an end-to-end stack for accelerating scientific discovery



- Ecosystem of lightweight, scalable services to locate, extract, and aggregate data and information from heterogeneous sources
- Supporting HTC infrastructure to parse and analyze documents, expose data via API
- Principled, automated access to new and archival publications spanning publishers



support 2014-2018 by NSF-ICER 1343760, DARPA ASKE  
partial current support from USGS

**xDD API:**

<https://geodeepdive.org/api>

**Code available at:**

<https://github.com/UW-COSMOS>

**Correspondance:**

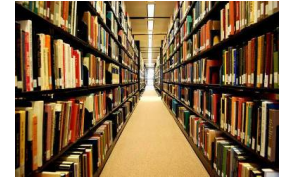
Shanan Peters ([peters@geology.wisc.edu](mailto:peters@geology.wisc.edu))

Theodoros Rekatsinas ([thodrek@cs.wisc.edu](mailto:thodrek@cs.wisc.edu))

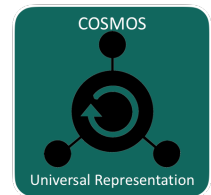
Miron Livny ([miron@cs.wisc.edu](mailto:miron@cs.wisc.edu))

# Outline

1. xDD: A portal to scientific publications and HTC



2. COSMOS: Knowledge extraction as a service

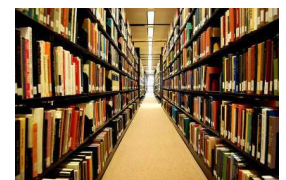


3. Demo: analyzing model code with COSMOS

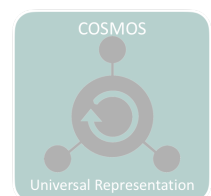


# Outline

1. xDD: A portal to scientific publications and HTC

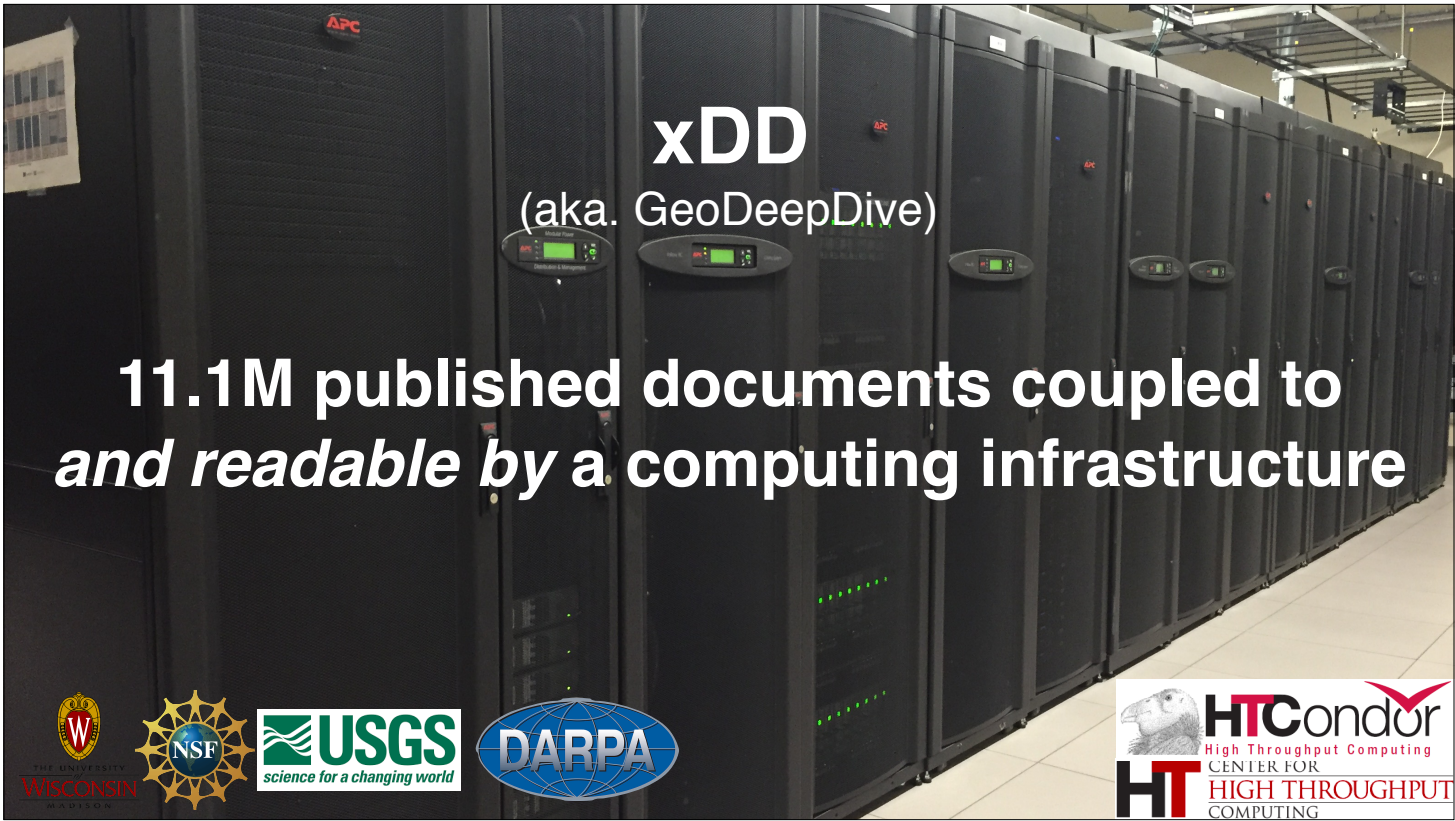


2. COSMOS: Knowledge extraction as a service



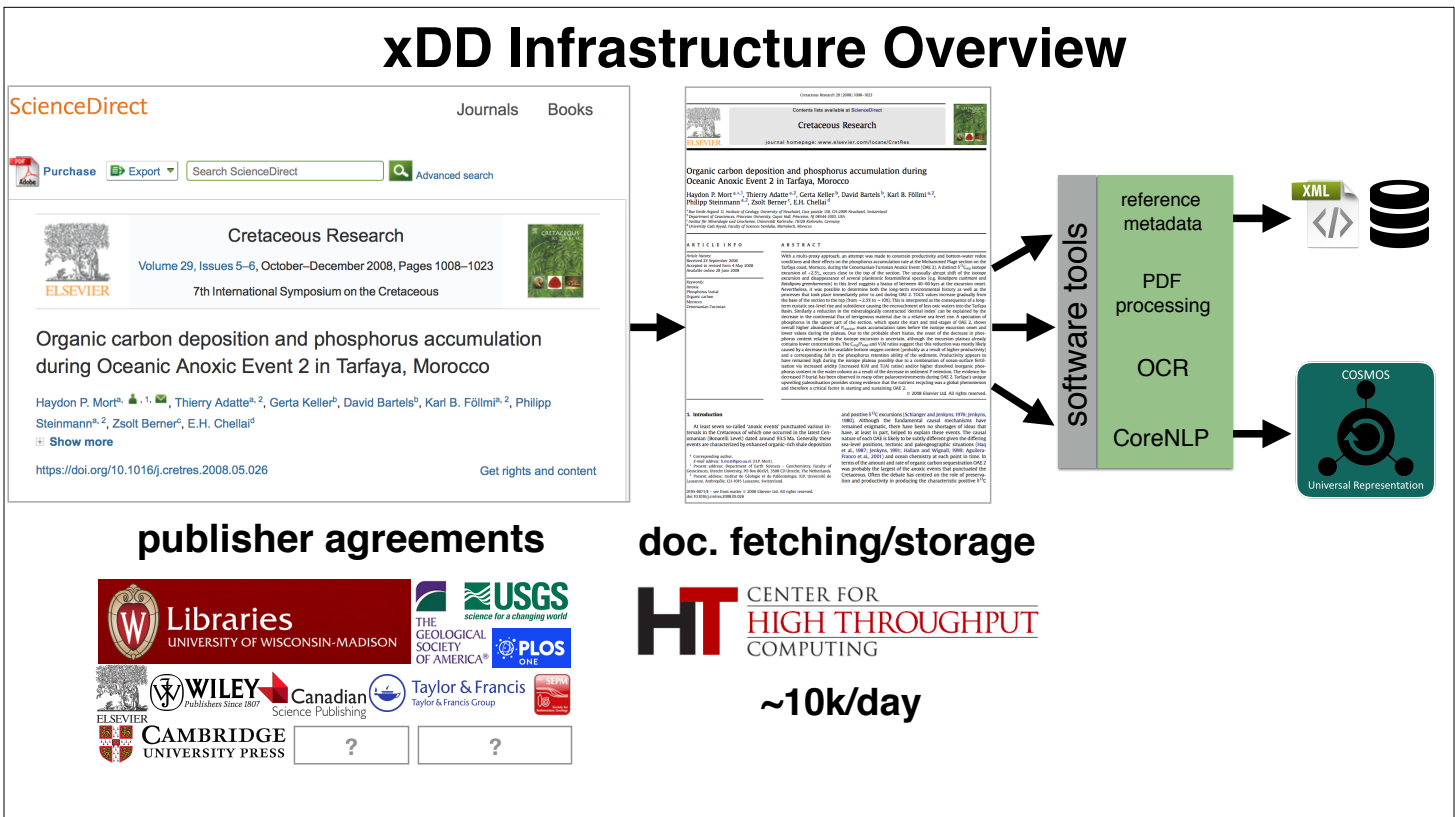
3. Demo: analyzing model code with COSMOS





**xDD**  
(aka. GeoDeepDive)

**11.1M published documents coupled to  
and readable by a computing infrastructure**



# Vocabulary ingestion and labeling

geodeepdive.org/api/dictionaries

curated lists of terms  
with hierarchy/context:



The Paleobiology Database  
revealing the history of life

→ 350k  
taxonomic names/  
bio classification



mindat.org

→ 6k  
mineral names/  
chemistry



Macrostrat

→ 45k  
stratigraphic names/  
hierarchy, rock types

GDD-supplied full text  
and indexing capability

Pyritization of soft-bodied fossils: Beecher's Trilobite Bed, Upper Ordovician, New York State

Derek E.G. Briggs  
Department of Geology, University of Bristol, Wills Memorial Building, Queen's Road  
Bristol BS8 1RJ, England  
Simon H. Bottrell, Robert Raiswell  
Department of Earth Sciences, University of Leeds, Leeds LS2 9JT, England

**ABSTRACT**  
Although pyrite is ubiquitous in fine-grained, organic-bearing marine sediments, it is only rarely involved in the preservation of soft-bodied organisms. Beecher's Trilobite Bed in Upper Ordovician strata of New York State is an exception—it is a classic locality for trilobites having appendages and other soft tissues preserved in pyrite. The relative timing and duration of the formation of pyrite associated with the fossils and their host sediments were determined by use of sulfur isotope ratios. The coexistence and separation of the trilobites show relatively light sulfur isotope values in contrast to the enclosing sediments, which is characteristic for a substantial excursion to heavy isotope values. Preservation of soft parts requires rapid burial of carcasses in sediments otherwise lost to metabolizable organic matter. In these circumstances, pyrite formation within the sediments is suppressed, thus concentrations of sulfate and oxidative iron are initially high enough to promote early, rapid, and extensive pyritization of nonmineralized tissue.

**INTRODUCTION**  
Fossils that preserve soft tissues provide critical evidence of the morphology and paleobiology of animal organisms—no contrast to mineral shells fossil assemblages, which yield only limited information. Soft tissues (i.e., those lacking an mineral component in life) may be preserved in a variety of ways. Those that are particularly decay resistant (cuticles composed of keratin, sporopollenin, chitin, sclerotized shells, for example) may become fossilized as stable kerogen compounds in certain environments (Friggoli et al., 1992; Jovan et al., 1995). Tissues more susceptible to bacterial breakdown (e.g., muscle, internal organs, thin cuticles) survive only when they are replaced by very early autogenic mineralization (Cohen, 1990). This normally involves one of three groups of diagenetic minerals: phosphate, carbonate, or pyrite. Pyrite is commonly a component of fine-grained, organic-rich marine sediments, formed by reactions between dissolved iron minerals and the H<sub>2</sub>S generated by anaerobic sulfide-reducing bacteria (Goldhaber and Kaplan, 1974). In marine sediments, iron and sulfur are normally present in abundance, and pyrite formation is apparently controlled by the concentration of metabolizable organic carbon (Berner, 1970, 1984).

Although pyrite is widespread in marine sediments, and commonly a fossil in association with fossils, there are usually the remains of mineralized (Hollister, 1922) or at least refractory, mineralized, or mineralized (Kornick and Edwards, 1985; Beecher's Trilobite Bed (named after the Yale paleontologist who worked extensively on the trilobites in the 1930s) is one of the very rare examples where pyrite formed early enough to contribute to the preservation of soft tissues. Only the Chevronea (Green Branch) Haverhill locality of western Germany (Shaner et al., 1980; Kott and Witzke, 1982; Bottrell and Briggs, 1995), which preserves the soft tissues of trilobites (Chalmers and Bergstrom, 1973), cephalopods (Chalmers, 1983), and nonarthropods (Chalmers and Bergstrom, 1973), for example, is comparable.

**FIGURE 1.** Trilobite specimen, 20 mm long, from Beecher's Trilobite Bed (photograph by A. E. Almond, provided by H. B. Whittington).

Beecher's Bed is additionally important as the only major occurrence of soft-bodied organisms (Kornick-Lapostolle) known from the Ordovician (Almond and Briggs, 1995). In this paper we analyze the mineralization of the trilobites in Beecher's Bed and present a model for the preservation of soft tissues in the fossil record.

GEOLOGY, v. 19, p. 1271-1276, December 1995

labeled  
entities, tuples

DOI:  
10.1130/0091-7613(1991)019<1221:POSBFB>2.3.CO;2

Trilobita  
Triarthrus  
Climacograptus

pyrite

Frankfort Shale

exposed  
via API

"term\_hits": { ▼ 189470 properties, 6 MB

"Navajosuchus novomexicanus": 7,  
"Ptilocolepidae": 2,  
"Geotrupidae": 652,  
"Macropoma lewesiensis": 11,  
"Simia morio": 7,  
"Anhanguera robustus": 4,  
"Montipora verrilli": 36,  
"Geffenina wangi": 6,  
"Shuvosaurus": 198,  
"Dryorhizopsidae": 1,  
"Ostrea antarctica": 12,  
"Pholidophorus dentatus": 10,  
"Oochorista": 6,  
"Sciurus arizonensis": 18,  
"Prosbole biexcisa": 3,  
"Toxopatagus": 271,  
"Sulcavitus": 169,  
"Brontops amplus": 2,  
"Melonella": 1087,  
"Bathrotomaria": 688,  
"Placochelyanus": 64,  
"Ilioichione": 555,  
"Attenosaurus subulensis": 10,  
"Miccylyotyrans": 4,

"Deryeuma": 4,  
"Chaetabraeus": 2,  
"Cardinalis cardinalis": 1113,  
"Odontaster": 2532,  
"Coeloma": 384,  
"Megatrema": 36,  
"Xinjiangtitan": 4,  
"Tephrodytes brassicarvalis": 74,  
"Prolagus crusafonti": 42,  
"Streblascopora germana": 2,  
"Dichobunoidea": 11,  
"Cetotherium hupschi": 1,  
"Hauericeras (Gardeniceras) gardeni": 10,  
"Parevania": 15,  
"Histriobdellidae": 130,  
"Berosus (Berosus)": 13,  
"Dinornis elephantopus": 8,  
"Sternoxi": 11,  
"Bellimurina (Bellimurina)": 4,  
"Raphignathoidea": 76,  
"Cybelurus occidentalis": 4,



Define your own dictionaries:

[https://github.com/UW-Deepdive-Infrastructure/dictionary\\_example](https://github.com/UW-Deepdive-Infrastructure/dictionary_example)

```

"term_hits": {
  "Navajosuchus novomex": 189470,
  "Ptilocolepidae": 2,
  "Geotrupidae": 652,
  "Macropoma lewesiensis": 1,
  "Simia morio": 7,
  "Anhanguera robustus": 1,
  "Montipora verrilli": 1,
  "Geffenina wangi": 6,
  "Shuvosaurus": 198,
  "Dryorhizopsidae": 1,
  "Ostrea antarctica": 1,
  "Pholidophorus dentatus": 1,
  "Oochorista": 6,
  "Sciurus arizonensis": 1,
  "Prosbole biexcisa": 1,
  "Toxopatagus": 271,
  "Sulcavitus": 169,
  "Brontops amplus": 2,
  "Melonella": 1087,
  "Bathrotomaria": 688,
  "Placochelyanus": 64,
  "Ilioichione": 555,
  "Attenosaurus subulensis": 10,
  "Miccyлотyrans": 4,
  "pubname": "Earth-Science Reviews",
  "publisher": "Elsevier",
  "title": "The Cambrian palaeontological record of the Indian subcontinent",
  "coverDate": "Available online 11 June 2016",
  "URL": "http://www.sciencedirect.com/science/article/pii/S0012825216301179",
  "authors": "Hughes, Nigel C.",
  "hits": 4,
  "highlight": [
    " represented is the hyolithomorph Sulcavitus wynnei (Waagen, 1885) collected from several horizons within",
    ") Hy Sulcavitus wynnei (Waagen), dorsum, Khussak Formation, Salt Range, SH, GSI 4118 (CMCIP 71490",
    ",) Kruse and Hughes in press, fig. 5C, scale bar: 2.5 mm; T) Hy Sulcavitus wynnei (Waagen), venter",
    " 71490), Kruse and Hughes in press, fig. 5A, scale bar: 2.5 mm; U) Hy Sulcavitus wynnei (Waagen"
  ]
},
{
  "pubname": "Alcheringa: An Australasian Journal of Palaeontology",
  "publisher": "Taylor and Francis",
  "title": "Biostratigraphic potential of Middle Cambrian hyoliths from the eastern Georgina Basin",
  "coverDate": "2002 01",
  "URL": "http://www.tandfonline.com/doi/abs/10.1080/03115510208619263",
  "authors": "Kruse, Peter D.",
  "hits": 1,
  "highlight": [
    "it is noteworthy that Sulcavitus possesses a distinctive deep median sulcus on the dorsum"
  ]
}

```

Define your own dictionaries:  
[https://github.com/UW-Deepdive-Infrastructure/dictionary\\_example](https://github.com/UW-Deepdive-Infrastructure/dictionary_example)

Main Menu
About
Download
Search

Search the database

Login

Basic info	Taxonomic history	Classification	Relationships
Morphology	Ecology and taphonomy	External Literature Search	Age range and collections

## Pithecanthropus erectus

Mammalia - Primates - Hominidae

GeoDeepDive matched this taxon in 229 documents from 73 journals/publications:

- Soriano, M.. *The fluoric origin of the bone lesion in the Pithecanthropus erectus femur. American Journal of Physical Anthropology* January 1970.

...The Fluoric Orig- in of the Bone Lesion in the *Pithecanthropus erectus* Femur Facultad de Medicina...

... that the *Pithecanthropus erectus* suffered a bone fluorosis of the t m e of the Periostitis deformans...

.... discovered by Dr. Eugene Dubois in 1891-1892 and attributed to the *Pithecanthropus erectus*, have...

... a bone lesion very similar to that of the *Pithecanthropus erectus*. Both bones present several...

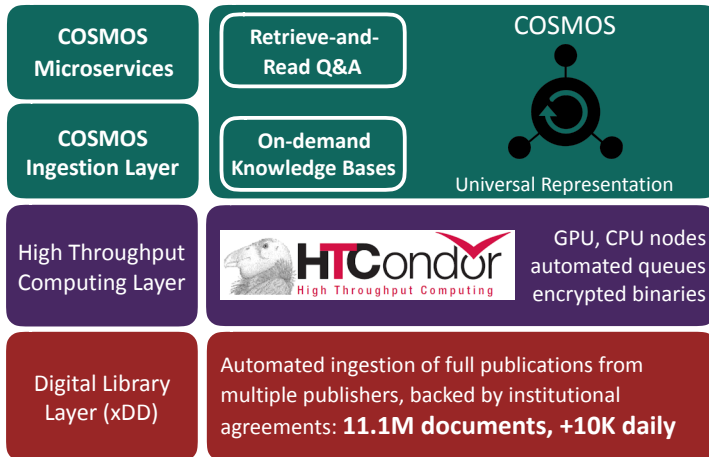
.... PHYS.ANTHROP.,32: 49-58. 49 t Fig. 1 Femur of the *Pithecanthropus erectus*. (Taken from Julien, '69...

[Hide recognized terms from this document](#)

Taxonomic names	Homo Homo erectus Pithecanthropus Pithecanthropus erectus
Lithologies	ash volcanic

The Paleobiology Database  
revealing the history of life

# xDD and COSMOS: an end-to-end stack for accelerating scientific discovery



- Ecosystem of lightweight, scalable services to locate, extract, and aggregate data and information from heterogeneous sources
- Supporting HTC infrastructure to parse and analyze documents, expose text via API
- Principled, automated access to new and archival publications spanning publishers

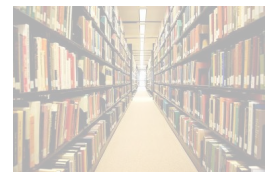


**xDD API:**  
<https://geodeepdive.org/api>  
**Code available at:**  
<https://github.com/UW-COSMOS>

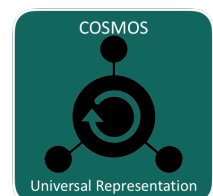
**Correspondance:**  
Shanan Peters (peters@geology.wisc.edu)  
Theodoros Rekatsinas (thodrek@cs.wisc.edu)  
Miron Livny (miron@cs.wisc.edu)

## Outline

1. xDD: A portal to scientific publications and HTC



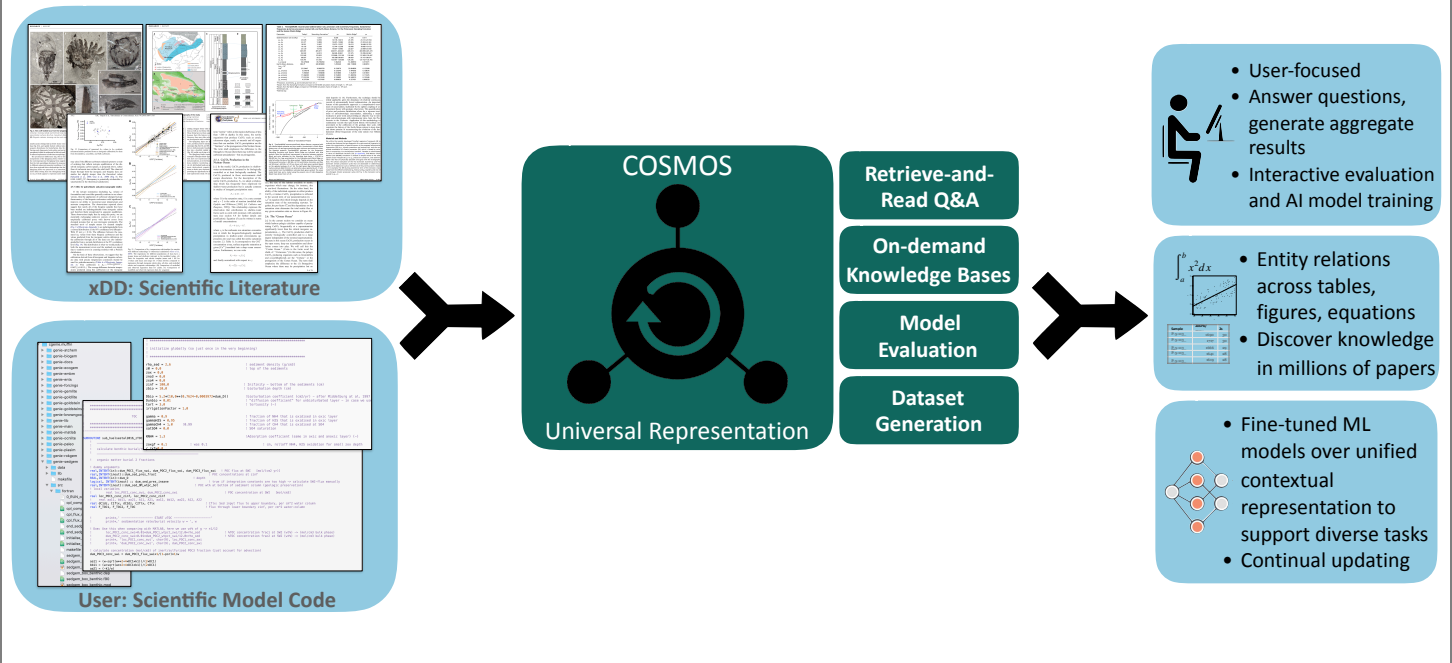
2. COSMOS: Knowledge extraction as a service



3. Demo: analyzing model code with COSMOS



# Accelerating scientific discovery with COSMOS



## Knowledge extraction from high-variety input

**Equation**

$$\delta^{13}\text{C} = (R_{\text{sample}}/R_{\text{standard}} - 1) \times 10^3$$

**Body Text**

where  $R$  is  $^{13}\text{C}/^{12}\text{C}$ . The standard is Pee Dee Belemnite limestone that has been assigned a value of  $0.0\text{‰}$ . The precisions of  $\delta^{13}\text{C}$  determination were less than  $0.2\text{‰}$ . POC and PON concentrations were determined using a TCD detector attached to the elemental analyzer.

For Chl  $a$  and pheophytin concentrations, POM samples were extracted in the dark for 12 h by 90% acetone, and their concentrations were measured by the fluorometric method (Japan Meteorological Agency, 1970), using a calibrated Turner Designs TD700 fluorometer. In this study, chlorophyll (Chl) was determined as the total pigment including pheophytin.  $\text{PO}_4\text{-P}$  was extracted filtrate by the ascorbic acid–Mo blue method (Strickland and Parsons, 1965), using a Technicon Auto Analyzer.

**Section Header**

**3. Results**

**Section Header**

**3.1. Variations in river discharge and riverine POM composition**

**Body Text**

River discharge of the Kiso Rivers changed considerably during the observation period (Fig. 3). Discharge was low ( $< 500 \text{ m}^3 \text{ s}^{-1}$ ) until 22 June, and suddenly increased on 24 June (the first flood,  $\sim 2000 \text{ m}^3 \text{ s}^{-1}$ ), reaching a peak flood on 28 June (the second flood,  $\sim 3000 \text{ m}^3 \text{ s}^{-1}$ ). After that, it

during normal discharge. However, the concentration in the Nagara River at high discharge was the same level as that at normal discharge. After discharge, POC concentrations decreased in all rivers.  $\delta^{13}\text{C}$  of POM in the Kiso River and the Nagara River varied from  $-27.3\text{‰}$  to  $-23.1\text{‰}$  and from  $-29.7\text{‰}$  to  $-25.9\text{‰}$ , respectively. On the other hand,  $\delta^{13}\text{C}$  of POM in the Ibi River remained fairly constant (ca.  $-30\text{‰}$ ). The C/N ratios varied from 7.8 to 22.3 and reached the highest values during high discharge in all rivers.

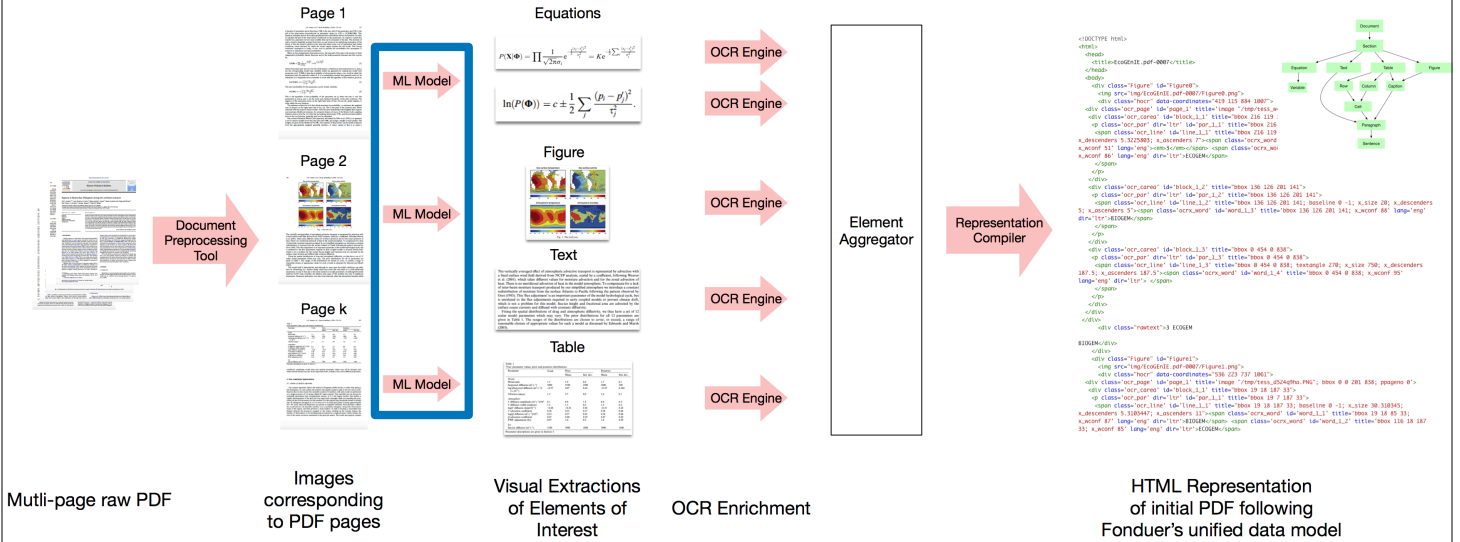
**Table 1**

Summary of physical and chemical variables in the Kiso rivers collected at  $\sim 15 \text{ km}$  upstream from the river mouth

	Discharge ( $\text{m}^3 \text{ s}^{-1}$ )	POC ( $\text{mg l}^{-1}$ )	PON ( $\text{mg l}^{-1}$ )	$\delta^{13}\text{C}$ (‰)	C/N (mol ratio)
<b>Kiso River</b>					
20 June	155	0.61	0.06	-27.3	12.6
28 June	1257	1.78	0.09	-25.5	22.3
4 July	269	0.30	0.03	-23.1	12.5
<b>Nagara River</b>					
20 June	63	2.28	0.34	-27.7	7.8
28 June	1072	2.11	0.13	-25.9	18.3
4 July	129	0.44	0.06	-29.7	8.7
<b>Ibi River</b>					
20 June	21	1.21	0.14	-30.9	9.8
28 June	622	2.53	0.15	-29.5	20.9
4 July	63	0.60	0.10	-29.0	7.9



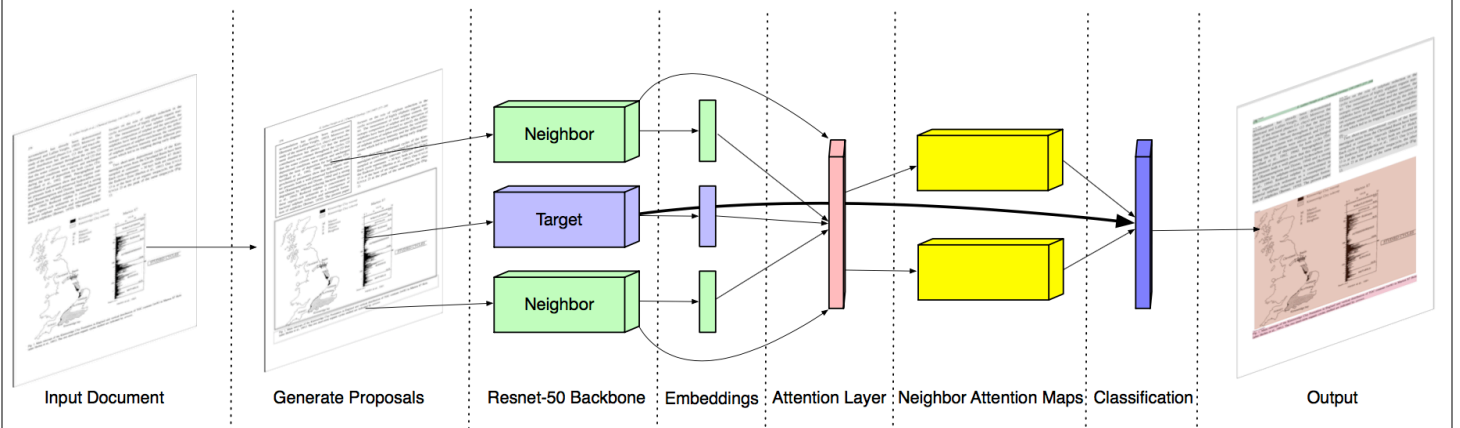
# From PDF to XML



How do we get these elements from a scanned image?

We need support for images to address the format heterogeneity across publications.

# The COSMOS Attentive RCNN Model



New distributed representation (in the visual space) for each element in the page.

**2.2.2. Ocean model**  
 The ocean model is divided into two submodels: physical and biological-chemical submodels.  
**Physical submodel:** The governing equations of the submodel are as follows: the equation of motion for rotational fluid under the assumption of hydrostatic and Boussinesq approximations (Eq. (1)), the equation of continuity of incompressible fluid (Eq. (2)), and equations of advection-diffusion of the salinity (Eq. (3)), and heat included in water (Eq. (4)).

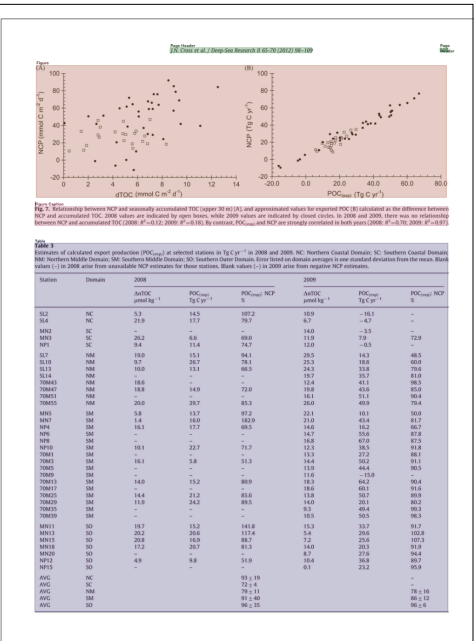
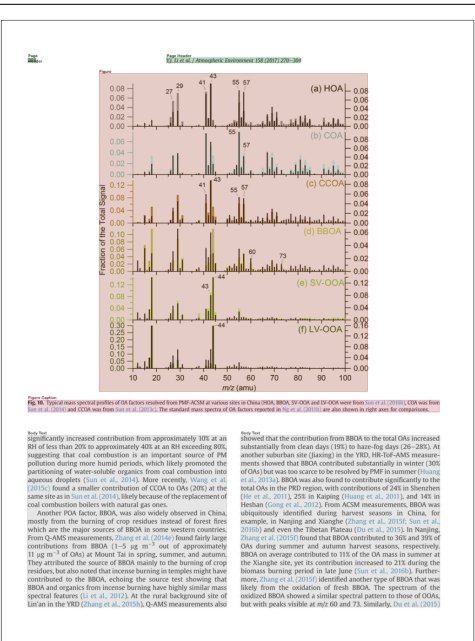
$$\rho_0 \frac{D\mathbf{v}}{Dt} + \nabla_{\text{horiz}} \cdot \mathbf{v} = \rho_0 \mathbf{g} - \nabla_{\text{horiz}} \pi + \mathbf{K} \nabla_{\text{horiz}}^2 \mathbf{v} + \mathbf{K}_v \nabla_{\text{vert}}^2 \mathbf{v} \quad (1)$$

$$\nabla_{\text{horiz}} \cdot \mathbf{v} = 0 \quad (2)$$

$$\frac{D\theta}{Dt} + \nabla_{\text{horiz}} \cdot (\mathbf{v} \theta) = \frac{\partial \theta}{\partial t} + \mathbf{K} \nabla_{\text{horiz}}^2 \theta + D_{\text{adv}} \left( \frac{\partial^2 \theta}{\partial x^2} + \frac{\partial^2 \theta}{\partial y^2} \right) + D_{\text{diff}} \left( \frac{\partial^2 \theta}{\partial z^2} + \frac{\partial^2 \theta}{\partial x^2} + \frac{\partial^2 \theta}{\partial y^2} \right) + \text{RHEF} \quad (3)$$

$$\frac{D\theta}{Dt} + \nabla_{\text{horiz}} \cdot (\mathbf{v} \theta) = \frac{\partial \theta}{\partial t} + \mathbf{K} \nabla_{\text{horiz}}^2 \theta + D_{\text{adv}} \left( \frac{\partial^2 \theta}{\partial x^2} + \frac{\partial^2 \theta}{\partial y^2} \right) + D_{\text{diff}} \left( \frac{\partial^2 \theta}{\partial z^2} + \frac{\partial^2 \theta}{\partial x^2} + \frac{\partial^2 \theta}{\partial y^2} \right) \quad (4)$$

where  $\rho_0$  denotes Coriolis parameter, and  $\rho_0 \mathbf{g}$  denotes the reference density of sea water. Free surface elevations are computed by calculating the convergence or divergence of barotropic components of water flow. Vertical diffusion coefficients are estimated using the turbulence model for ocean boundary layer proposed by Noh and Kim (1999), which is a simplified and improved version of the turbulent closure scheme by Mellor and Yamada (1982). The advective-diffusion equations are described in the Cartesian coordinate system. Ocean bottom topography is assumed to have a partial-plane, as noted by Adcroft et al. (1997) for representing the bottom slopes realistically in the Cartesian coordinate system. These equations are numerically solved by the finite-difference methods. In our computational code, for spatial differences, the Uniform Third Order Polynomial Interpolation Algorithm (TIO-PIA) and the second order central difference is used.



# The output of COSMOS's object detection module: tables, figures, equations, and associated text (captions, body text)

# Knowledge extraction from high-variety input

## Input: Raw High-Variety Inputs

Table 2. List of uncharged organic compounds used for calculation of the relationship  $\Delta K_d$  versus  $\log K_{ow}$  from Fig. 3

Compound	Parameter	$\log K_{ow}$
Glucose	Context	-2.57
Glycerol		-0.74
Methanol	Context	-0.42
1,4-Dioxane		-0.32
Ethanol	Values	-0.24
Acetone		-0.1
2-Propanol		0.74
2-Butanol		1.25

for hydrophobic, nonionogenic analytes ( $f(A_{COOH})$  approaches 0) a value of  $\Delta K_d$  is a function of

$$\Delta K_d = -0.38 \log K_{ow} - 0.26 \quad (5)$$

The found relationship displays an existence of the specific sorption on the gel (negative slope of the plot), the effect of which is strengthened with increasing hydrophobicity of the analyte. To establish the form of the relationship between  $\Delta K_d$  and  $A_{COOH}$ , a set of carboxylic acids with  $\log K_{ow} < -0.5$  was used (Table 3). The range of  $\log K_{ow}$  values close to or less than  $-0.5$  were chosen according to the observations described above. The obtained relationship is given in Fig. 4. It can be seen that the plot of  $\Delta K_d$  versus  $A_{COOH}$  is well

## Output: Extracted Structured Knowledge

EID	Expression
E1	$\Delta K_d = -0.38 \log K_{ow} - 0.26$
E2	$G_0 = \left[ 1 + (a + b(RH+0.01) + c(RH+0.01)^2) \frac{RH}{100 - RH} \right]^{1/3}$

### DataFrame1

Context	Values
Glucose	-3.24
Glycerol	-2.57
Methanol	0.74
1,4-Dioxane	-0.42
Ethanol	0.32
Acetone	-0.24
2-Propanol	-0.1
2-Butanol	0.74
2-Pentanol	1.25

EID	Parameter	Context	Data
E1	logKd	Compound	DataFrame1
E2	a	Composition	DataFrame2
E2	b	Composition	DataFrame3
E2	c	Composition	DataFrame4

# COSMOS extracts knowledge from multi-modal unstructured data (text, tables, images, equations, diagrams)

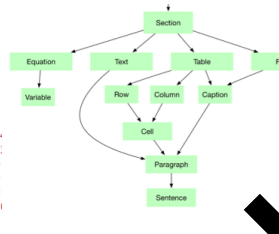
# Knowledge as a service in COSMOS

From the intermediate XML representation to knowledge bases and open-domain extraction

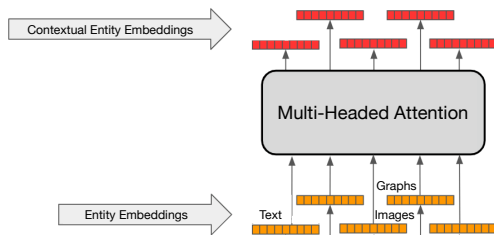
## Semi-structured data representation

```
<!DOCTYPE html>
<html>
<head>
<title>EcoGenIE.pdf-0007</title>
</head>
<body>
<div class="Figure" id="Figure0">

<div class="hocr" data-coordinates="419 115 884 1007">
<div class="ocr_page" id="page_1" title="image 7/tmp/tess_w">
<div class="ocr_carea" id="block_1_1" title="bbox 216 119">
<p class="ocr_par" dir="ltr" id="par_1_1" title="bbox 216 119">
<span class="ocr_line" id="line_1_1" title="bbox 216 119">
x_descenders 5.3225803; x_ascenders 7</span>
<span class="ocr_word">
x_wconf 51' lang="eng"><em>3</em></span>
<span class="ocr_x_woi">
x_wconf 86' lang="eng" dir="ltr">ECOGEN</span>
</div>
</div>
</div>
</div>
</div>
</body>
</html>
```



## Contextual data representation



- Retrieve-and-Read Q&A
- On-demand Knowledge Bases
- Model Evaluation
- Dataset Generation

- User-focused
- Answer questions, generate aggregate results
- Interactive evaluation and AI model training
- Entity relations across tables, figures, equations
- Discover knowledge in millions of papers
- Fine-tuned ML models over unified contextual representation to support diverse tasks
- Continual updating

## From Semi-structured data to Equation - Variable Models

Equation

$$C = C_0 + F \times \frac{t}{H} \quad (2)$$

where  $C$  is the headspace concentration of  $CO_2$  at time  $t$ ,  $C_0$  is its initial head-space concentration,  $F$  is the  $CO_2$  flux,  $H$  is the height of the head-space layer in the chamber. The fluxes of

- The parse tree for extracted equation (red-colored bounding box) reveals symbols **C**, **t**, **Co**, **F**, and **H**
- The same symbols are recognized in the text below this equation (purple “Variable” tokens).
- Variable tokens are linked to descriptions using the output of Open-IE (using CoreNLP), linking the Variable tokens and the phrase tokens.
- This method can be further improved (e.g., F here not automatically associated with  $CO_2$  flux).

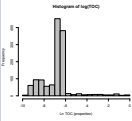
# Open-domain retrieve and read interaction

**Model Factoid-Q&A:** What is "TOC" in my source code?

**Definition**

Total Organic Carbon (TOC) is the amount of organic carbon in soil or a geological formation, particularly the source rock for a petroleum play.

**Empirical Distribution**

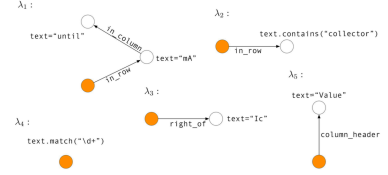


**Entity-Aware Code Analysis**

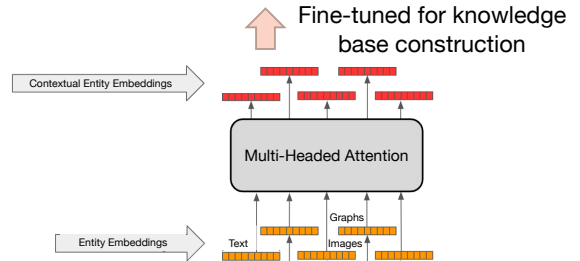
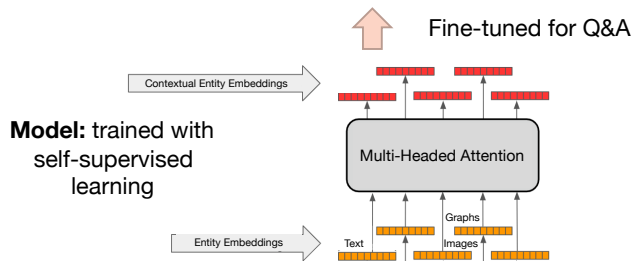
**Input**  
Annotated examples on PDF

MAXIMUM RATINGS	Symbol	Value	Unit
Collector - Emitter Voltage	V <sub>CE(sat)</sub>	40	Vdc
Collector - Base Voltage	V <sub>CB(sat)</sub>	40	Vdc
Emitter - Base Voltage	V <sub>EB(sat)</sub>	6.2	Vdc
Collector Current - Continuous	I <sub>C</sub>	800	mA
Base Current - Continuous	I <sub>B</sub>	100	mA
Power Dissipation (T <sub>C</sub> = 25°C)	P <sub>D</sub>	500	mW
Storage Time (T <sub>C</sub> = 25°C)	T <sub>S</sub>	1.2	µs
Turn-Off Time (T <sub>C</sub> = 25°C)	T <sub>OFF</sub>	0.5	µs
Operating and Storage Ambient Temperature Range	T <sub>stg</sub>	-55 to +150	°C

**Output**  
Ensemble of programs synthesized from the input examples

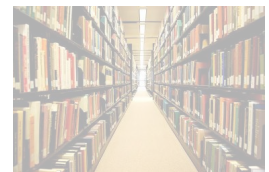


**Task**  
Extract the value of "Collector Current" for each model of transistor

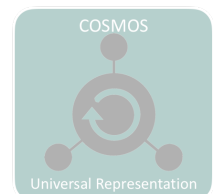


## Outline

1. xDD: A portal to scientific publications and HTC



2. COSMOS: Knowledge extraction as a service



3. Demo: analyzing model code with COSMOS

