

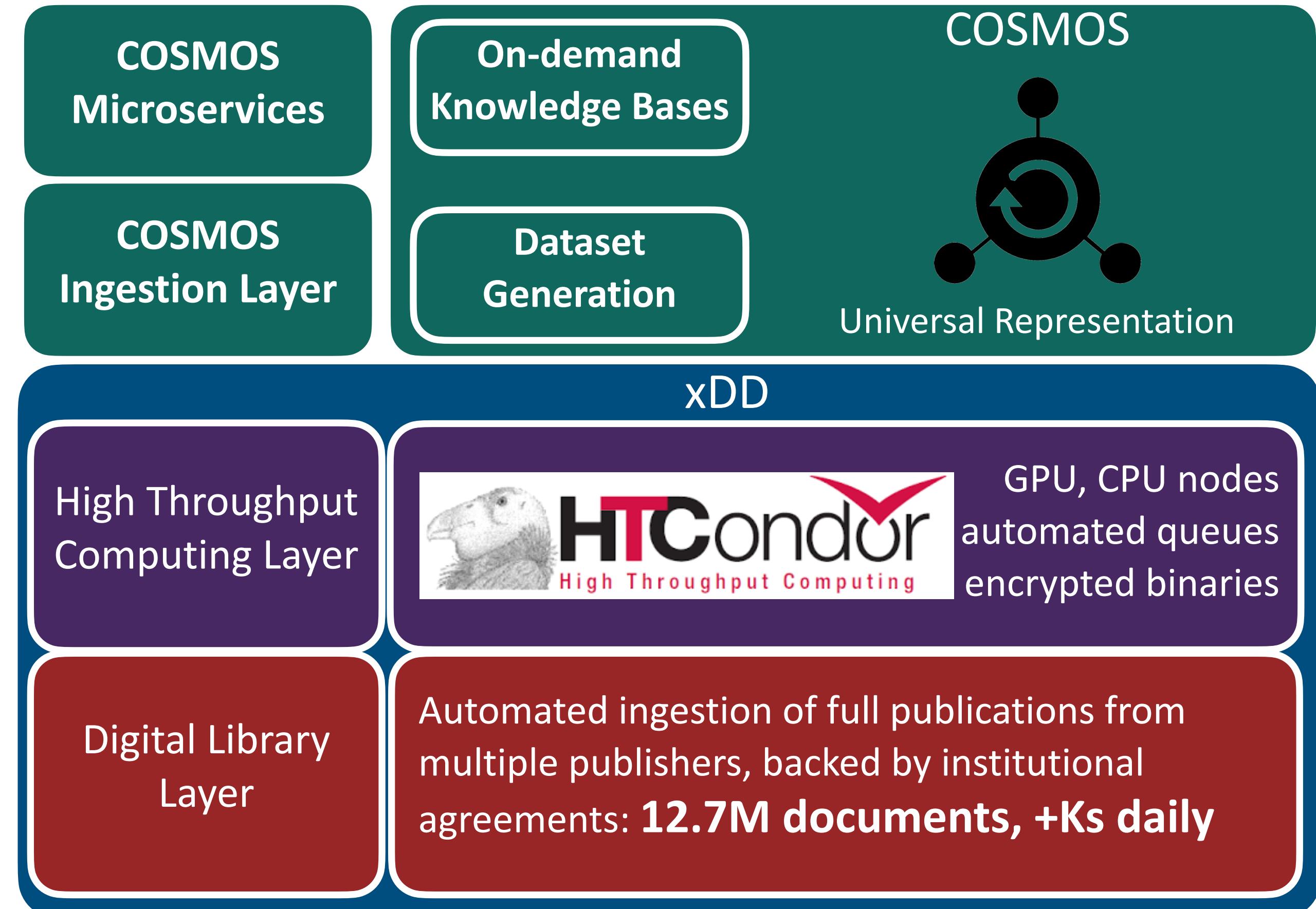
UW-COSMOS Phase 3:

A platform for extracting semantic knowledge and data from scientific publications

Shanan Peters, Theo Rekatsinas, Miron Livny



xDD & UW-COSMOS: an end-to-end stack to accelerate scientific research



- AI-powered “technical assistant” providing ecosystem of lightweight, scalable services to locate, extract, and aggregate data and information from heterogeneous sources
- HTC infrastructure to parse and analyze documents; API for full-text, fine-grained retrieval
- Automated fetching/storage of new and archival publications spanning many publishers and covering all domains of science



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

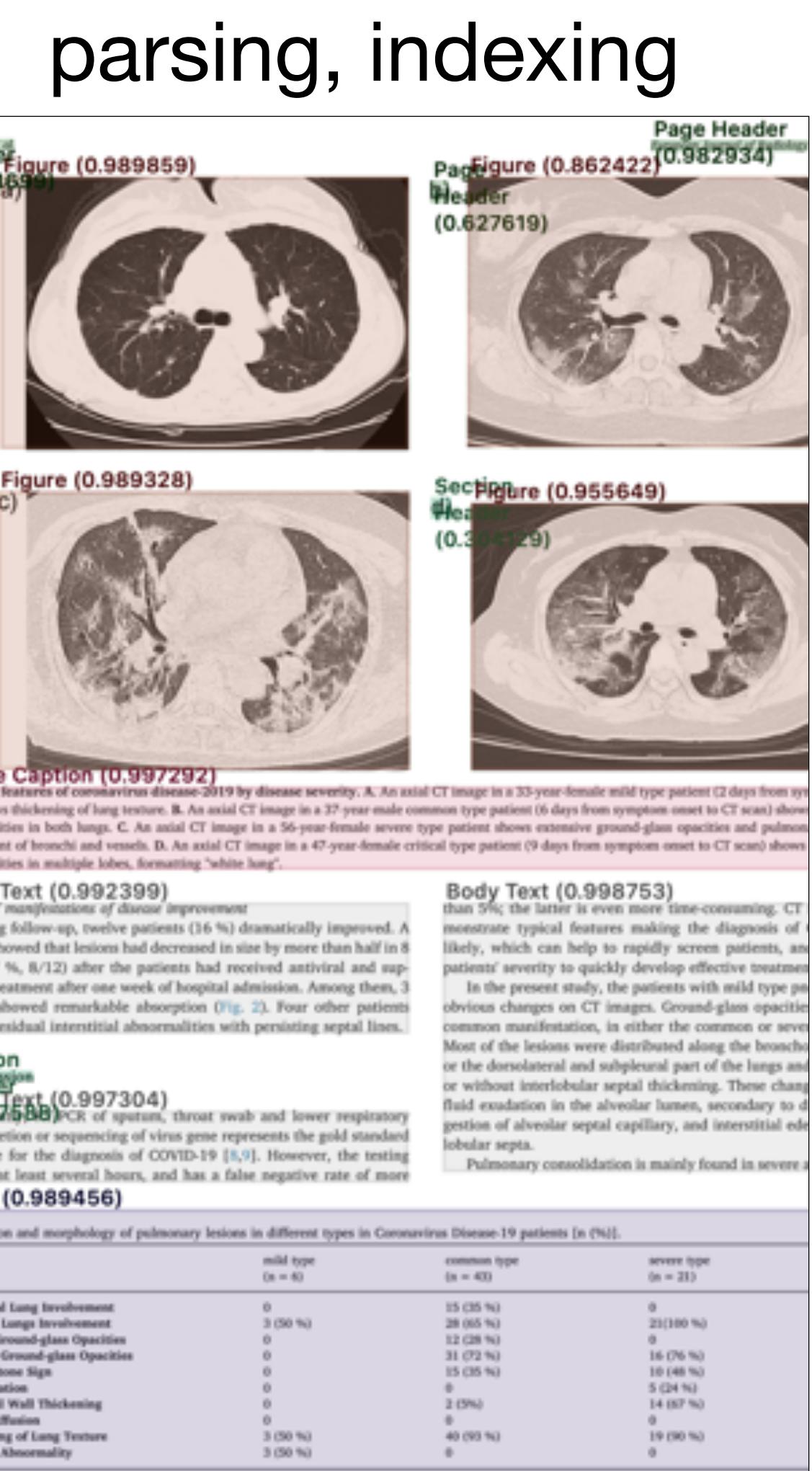
Correspondance:
Shanan Peters (peters@geology.wisc.edu)
Theodoros Rekatsinas (thodrek@cs.wisc.edu)
Miron Livny (miron@cs.wisc.edu)

UW-COSMOS Phase 3 Concept:

provide an AI-powered technical assistant and cloud infrastructure for leveraging the scientific literature in diverse, distributed workflows

Part 1:

Publication acquisition

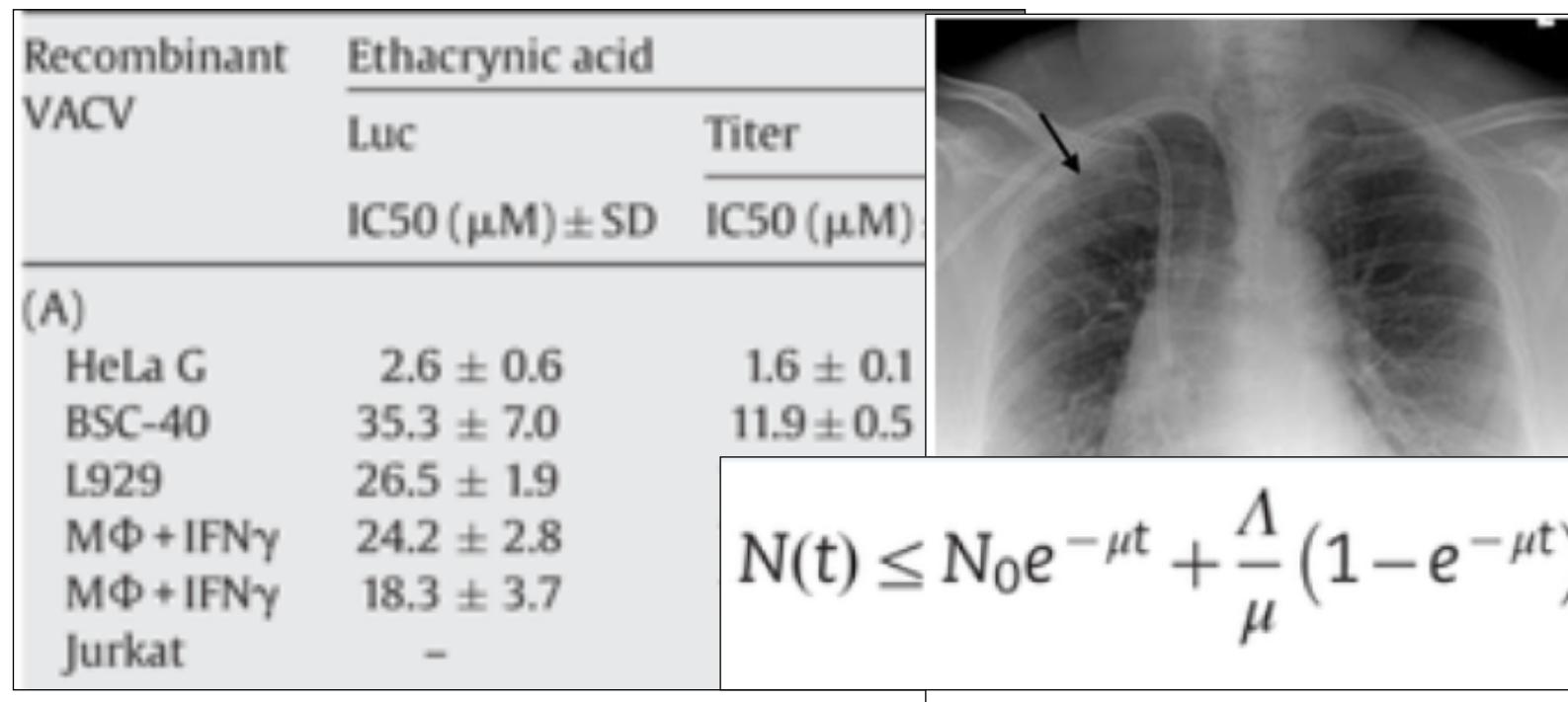


Part 2:

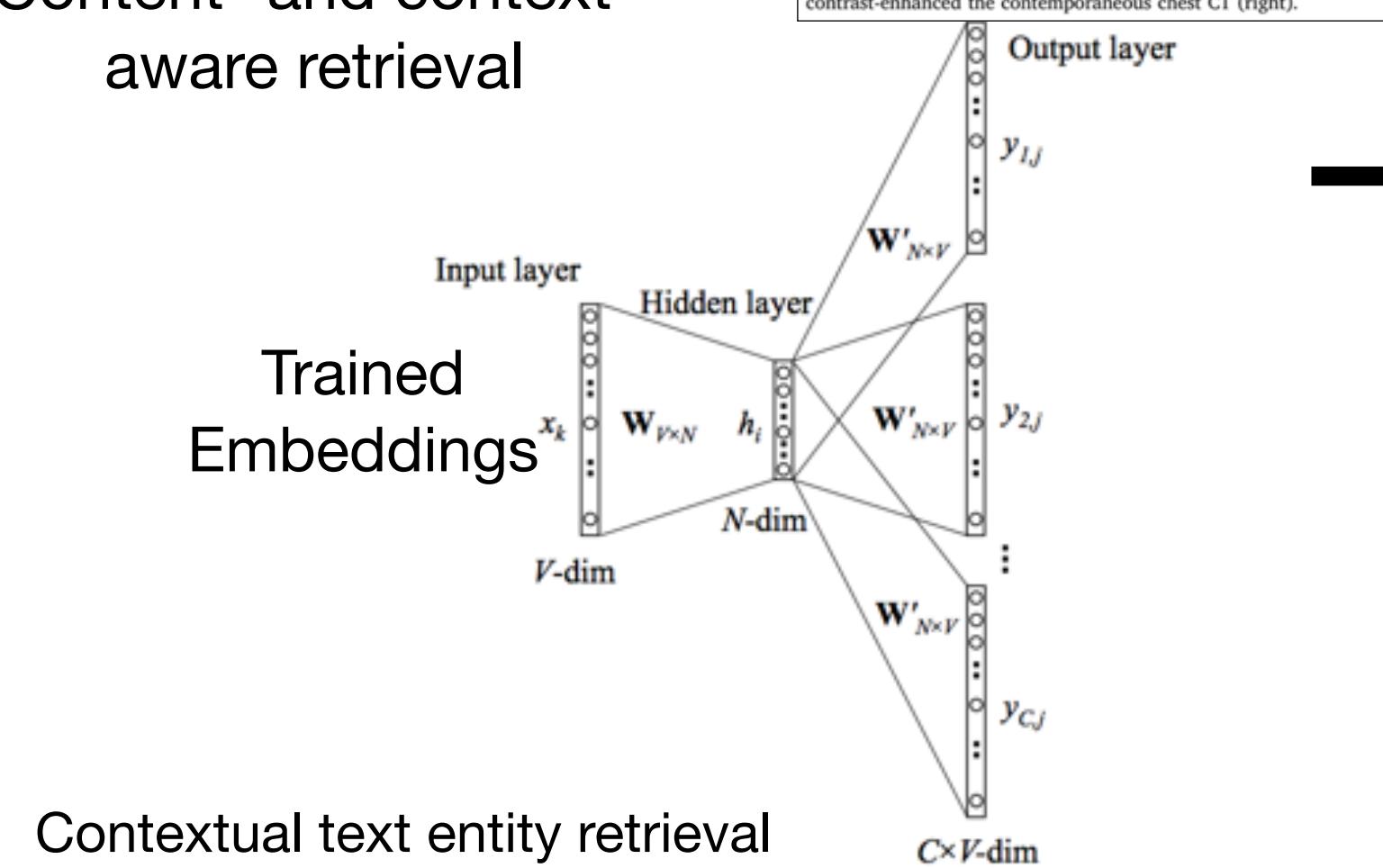
xDD/COSMOS
parsing, indexing

Part 3:

COSMOS-derived
data products



Content- and context-aware retrieval



Contextual text entity retrieval

virus binds to a ACE2 on human lung...

Deliverable:

AI-powered API platform

Ecosystem of REST-ful API data services providing ASKE teams and collaborators with real-time, fine-grained and aggregate data and knowledge extractions from scientific publications.

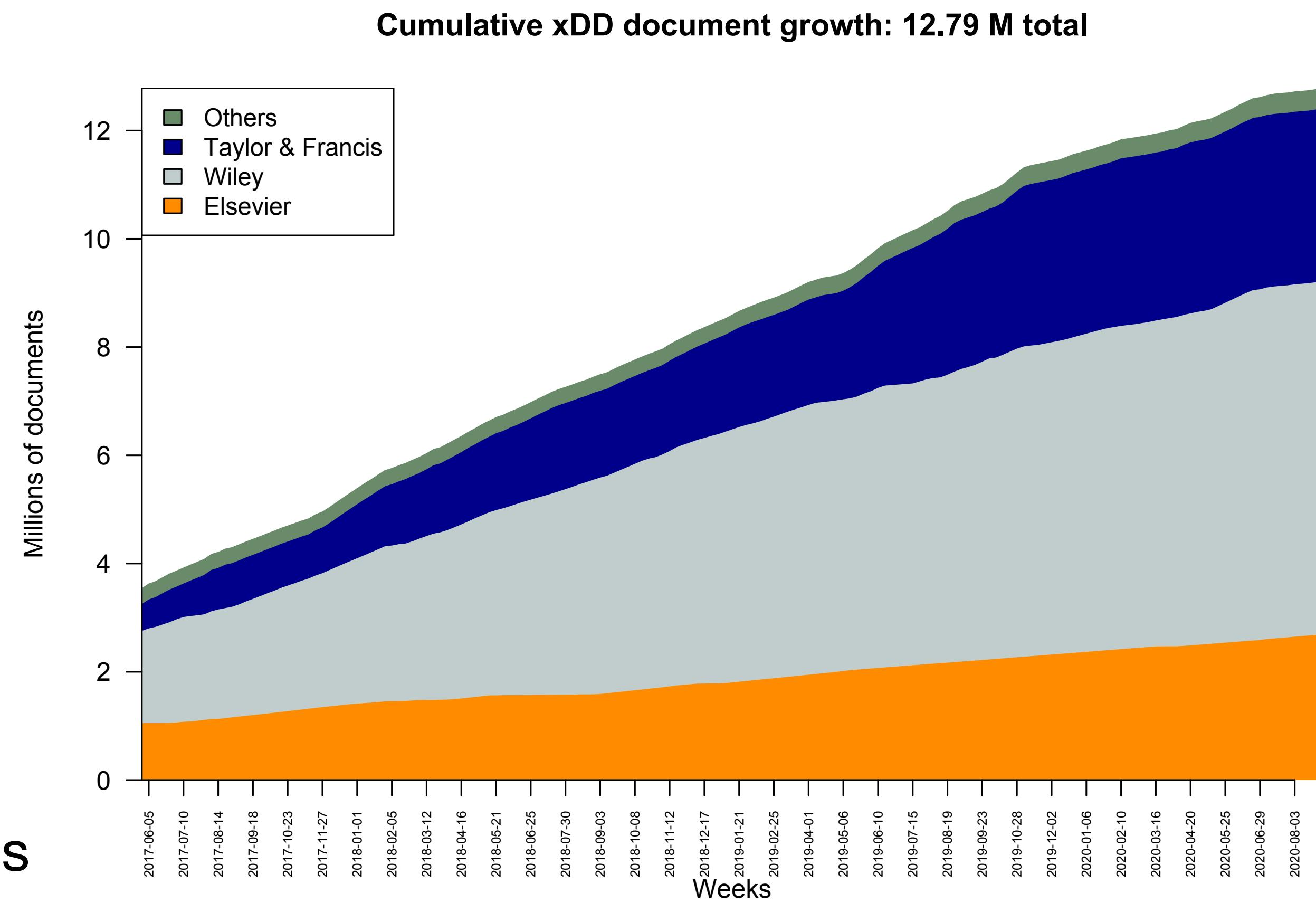
Phase 3 Tasks

- **Task 1: Scale and enhance xDD infrastructure and APIs**
 - A. Expand corpus, build new partnerships with publishers
 - B. Develop and release template for collaborator code execution
 - C. Incrementally annotate corpus with collaborator output
 - D. Improve and scale xDD APIs
- **Task 2: Scale and enhance COSMOS retrieval, KBC and APIs**
 - A. Improve and refine visual segmentation
 - B. Incorporate full-document context into text/table/figure/eqn retrieval
 - C. Automate knowledge base construction tasks
 - D. Release public API over COVID-19 COSMOS document set

Task 1A: Expand xDD Corpus

continue to grow the document set

- Ongoing work to bring in additional publishers and enhance our current agreements, extend into AKSE space
- Implement xDD “fetching as a service”
 - From a list of user-provided DOIs, journals, etc.
 - Immediately identify what's in xDD
 - Prioritize acquisition of documents already within xDD's publisher pipeline
 - Recognize and more broadly acquire open-access full-text documents
 - Influence/motivate future publisher agreements
 - Have some DOIs and/or journal runs that you need? **Let us know!**



Task 1A: Expand xDD Corpus

continue to grow the document set

The screenshot shows the Elsevier COVID-19 information center homepage. At the top, there is a navigation bar with links to About Elsevier, Products & Solutions, Services, Shop & Discover, a search icon, a shopping cart icon, and a user profile icon. The main title "Novel Coronavirus Information Center" is displayed prominently. Below the title, a subtitle reads "Elsevier's free health and medical research on the novel coronavirus (SARS-CoV-2) and COVID-19". A timestamp indicates the page was last updated on January 27, 2020. Below the subtitle, there are five categories: Clinical information, Mental health, 中文资源 (Chinese-language), Research / Drug discovery, and Public health. A large, stylized illustration of the COVID-19 virus particles is visible on the left side of the page. On the right, there is a welcome message, a statement about free access, and a link to a directory of resources. A specific article titled "On The Power of Networks, Or: Is This Our Moonshot?" by Anita de Waard is highlighted.

ELSEVIER

About Elsevier Products & Solutions Services Shop & Discover

Novel Coronavirus Information Center

Elsevier's free health and medical research on the novel coronavirus (SARS-CoV-2) and COVID-19

January 27, 2020 - Updated August 6, 2020

Clinical information Mental health 中文资源 (Chinese-language) Research / Drug discovery Public health

Welcome to Elsevier's Novel Coronavirus Information Center. Here you will find expert, curated information for the research and health community on SARS-CoV-2 (the novel coronavirus) and COVID-19 (the disease).

All resources are free to access and include guidelines for clinicians and patients.

See our directory of all Elsevier's COVID-19 resources >

On The Power of Networks, Or: Is This Our Moonshot?

Published on May 14, 2020

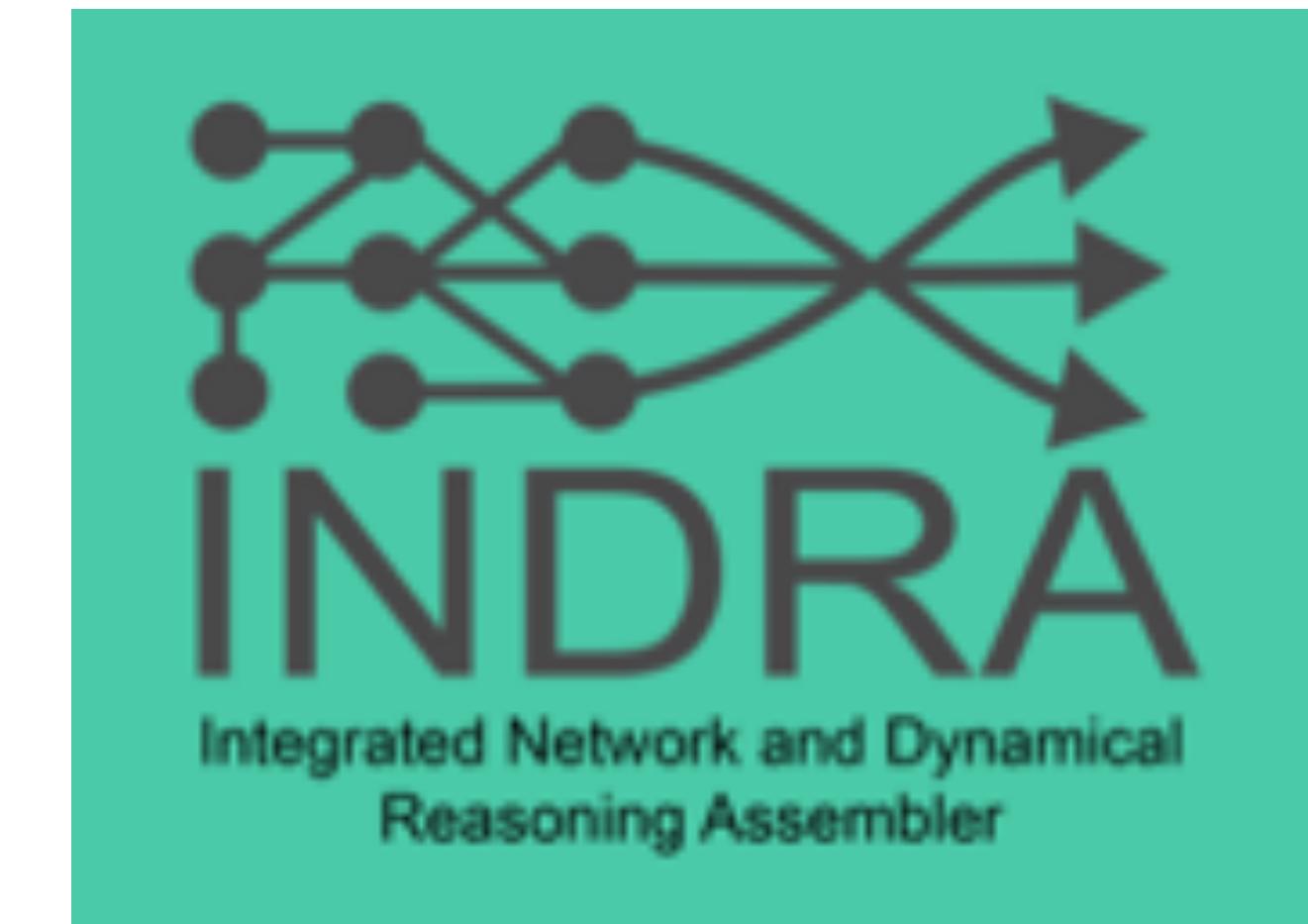
Anita de Waard
VP Research Collaborations at Elsevier

1 article + Fo

Task 1B: Container Template for Collaborator Code

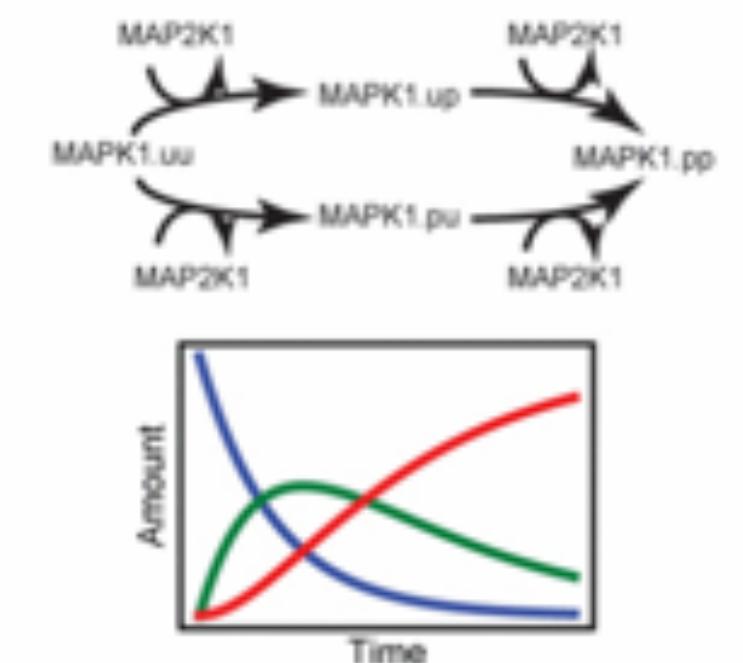
Release recipes for deploying code over xDD

- Develop an easy-to-use container template that allows ASKE collaborator code to be executed over xDD
- Collaboration with HMS provides initial use case:
 1. Define document corpus (DOIs, journals, keywords, etc.)
 2. HMS-provided docker image with logic to read input from (+ write output to) prescribed location in container
 3. Output handling: write to S3 bucket
 4. xDD infrastructure executes over 100sK of full-text documents, returns results regularly (daily?) as new documents are acquired
- Development is underway for better container support within the xDD CHTC infrastructure
- Plan to have initial template ready this month
 - Interested? Let us know!



"MEK1 phosphorylates ERK2 at T185 and Y187." Knowledge

Phosphorylation(MAP2K1, MAPK1, T, 185)
Phosphorylation(MAP2K1, MAPK1, Y, 187)



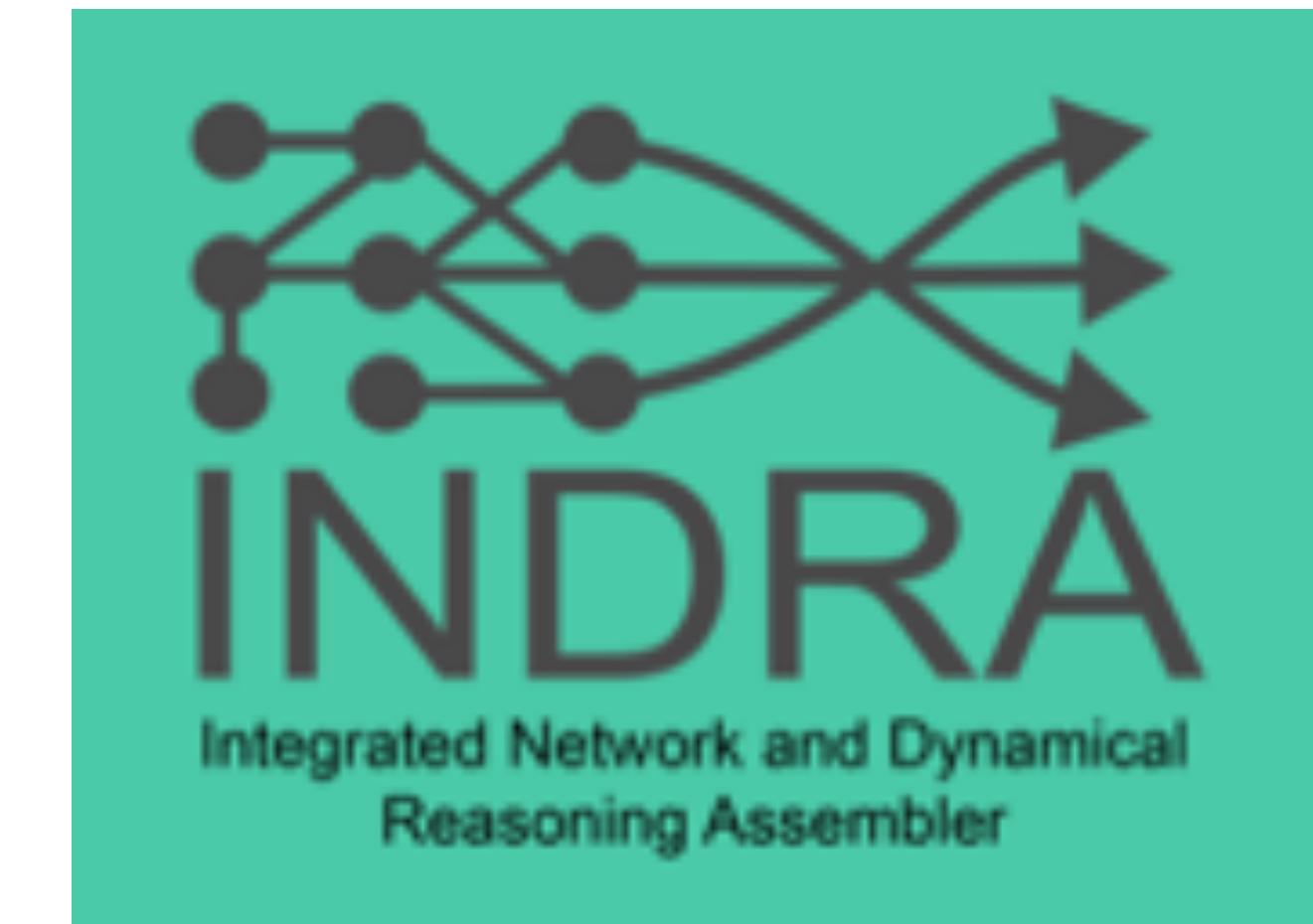
INDRA

Model

Task 1C: Augment xDD corpus with container outputs

Incrementally annotate corpus with knowledge products

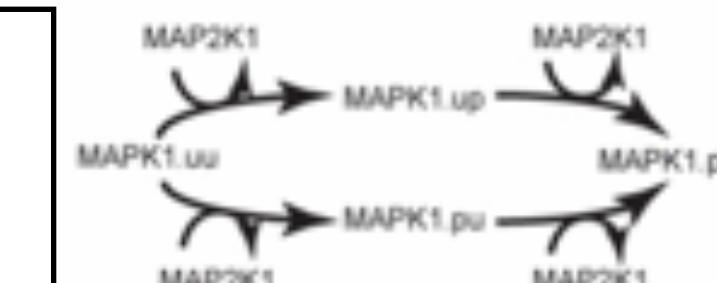
- Develop data model for and mechanisms to incorporate container code output back into xDD infrastructure.
- Expose annotations and extractions via the xDD API, even when performing arbitrary searches.
- End goal: incrementally augment knowledge structure around xDD corpus in order to automatically enrich document- and mention-level API output



"MEK1 phosphorylates ERK2 at T185 and Y187."

Knowledge

Phosphorylation(MAP2K1, MAPK1, T, 185)
Phosphorylation(MAP2K1, MAPK1, Y, 187)



INDRA

Model

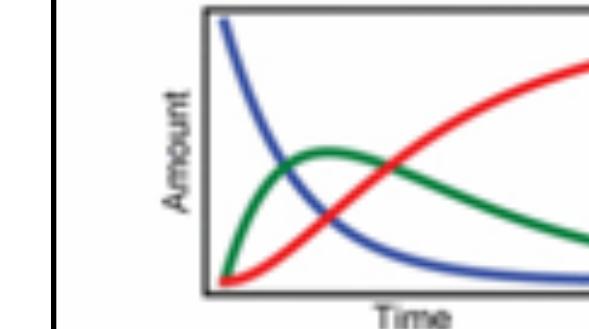
Mechanisms of Sensitivity and Resistance to CDK4/6 Inhibition

Published: 13 April 2020. [Cancer Cell](#)

- of compounds targeting PI3K, AKT, mTOR, FGFR1/2, MEK1/2, VEGFR1/2/3, PDGFRb, cKIT, and even pan-CDK
- al. (2019). A functional landscape of resistance to MEK1/2 and CDK4/6 inhibition in NRAS-mutant melanoma.
- vivo E2F reporting reveals efficacious schedules of MEK1/2-CDK4/6 targeting and mTOR-S6 resistance mechanisms.
- all mentions in [JSON](#)

GENES Mentioned: **DNMT1, SLC36A1** [see gene mentions [in context](#)]

DRUGS Mentioned: **abemaciclib, cisplatin, doxorubicin, etoposide, gemcitabine** [see drug mentions [in context](#)]



Task 1D: Improve and scale xDD API

Expand power of contextual searches, ensure responsiveness

- Additional hardware and architecture updates to enhance performance and scale
- Improve documentation for users and add powerful new retrieval/app-building features including:
 - Universal sortable scan and scroll
 - Improved query logic support (e.g., regex-like and proximity-based retrieval, such as “<term X> appearing within <30> tokens of <term Y>” and restricting results to sections within documents)
 - Incorporation of collaborator knowledge products into document- and mention-level API responses.
- Have a need? **Let us know!**

Query the xDD Corpus via [UW-COSMOS](#)

Full-text search terms/phrases: (separate multiple entries with commas):
MEK1,phosphorylates

Drug: Gene: Published: xDD Acquired

Require terms to co-occur: Case Sensitive:

[Get Results](#)

Related Terms from Trained Embedding Model:

MEK1: MEK (0.89), GSK-3β (0.89), PDK1 (0.88), ERK1 (0.88), c-Src (0.87), IKKα (0.87), PDK-1 (0.87), IKKβ (0.87), Src (0.86), c-Abl (0.86), ERK2 (0.86), GSK3β (0.86), Raf-1 (0.86), c-Jun (0.86), CDK2 (0.86), CDK9 (0.86), PAK2 (0.86), JNK1 (0.86), SHP2 (0.86), (see [similarity matrix](#) or raw [JSON](#))

phosphorylates: dephosphorylates (0.91), phosphorylate (0.87), kinase phosphorylates (0.86), directly phosphorylates (0.86), phosphatase (0.84), able phosphorylate (0.82), binds activates (0.81), ULK1 (0.81), PIP3 (0.81), binds phosphorylates (0.81), interacts (0.81), Cdc28 (0.81), exchange factor (0.81), turn phosphorylates (0.81), directly interacts (0.81), Raf-1 (0.8), serine/threonine kinase (0.8), RIP1 (0.8), (see [similarity matrix](#) or raw [JSON](#))

Full-Text Snippets **Published Since 2020-01-01:** (see all 25 results in [JSON](#))
Summary of [Mentioned Drugs & Genes](#) and [Gene-Drug Co-Occurrences](#)

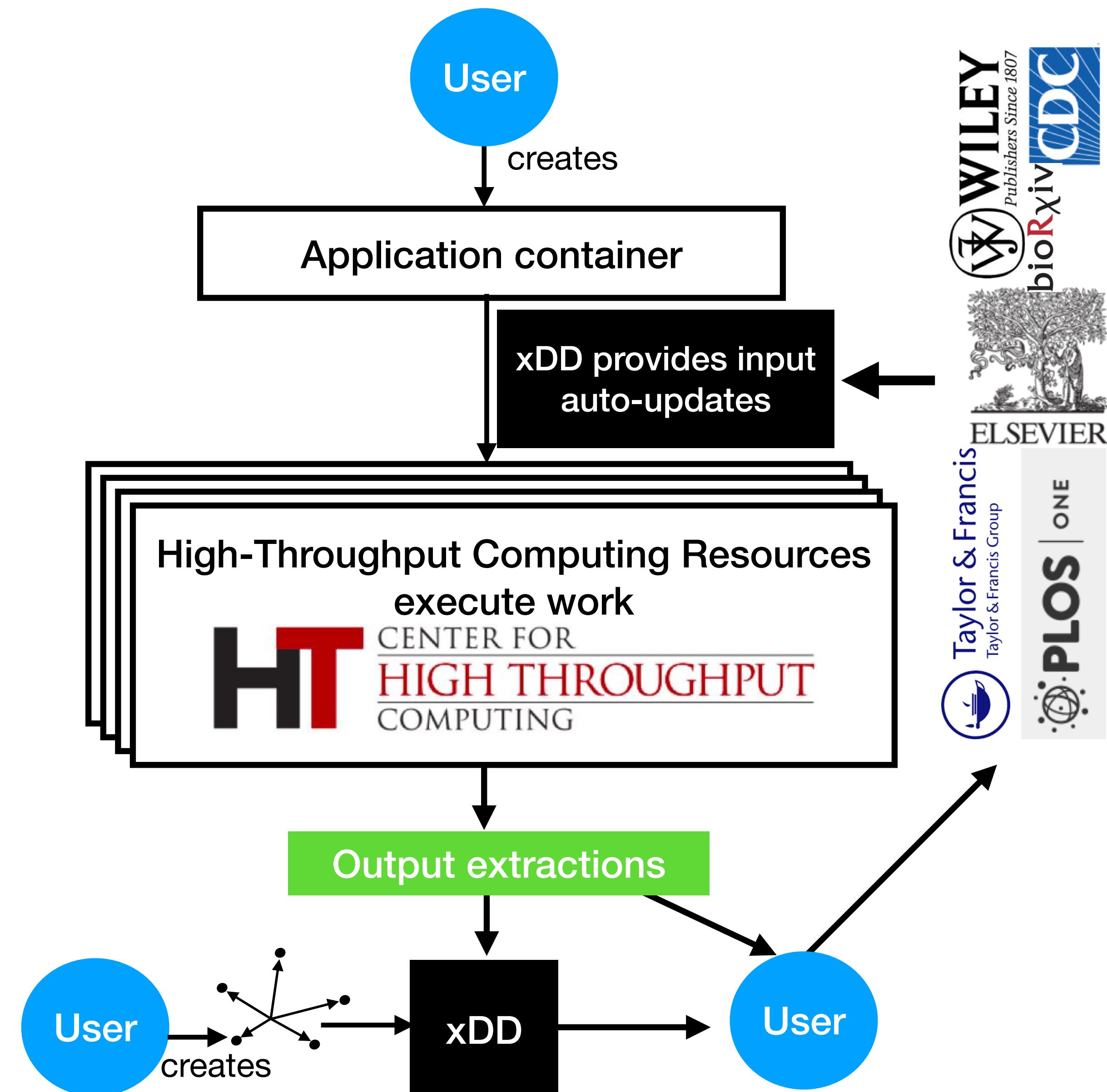
Mechanisms of Sensitivity and Resistance to CDK4/6 Inhibition
Published: 13 April 2020. [Cancer Cell](#)

- therapeutic efforts should not oversee the fact that CDK4/6 **phosphorylates** additional pocket proteins of the cell cycle, such as cyclins D1 and E1, CDK4/6 substrates, and other regulatory proteins.
- 2019; Rubio et al., 2019). In ovarian cancer, CDK6 **phosphorylates** FOXO3 and induces ATR to protect against DNA damage.
- of compounds targeting PI3K, AKT, mTOR, FGFR1/2, MEK1/2, VEGFR1/2/3, PDGFRb, cKIT, and even pan-CDK inhibitors.
- al. (2019). A functional landscape of resistance to **MEK1/2** and CDK4/6 inhibition in NRAS-mutant melanoma.
- vivo E2F reporting reveals efficacious schedules of **MEK1/2-CDK4/6** targeting and mTOR-S6 resistance mechanisms.
- all mentions in [JSON](#)

GENES Mentioned: **DNMT1, SLC36A1** [see gene mentions [in context](#)]
DRUGS Mentioned: **abemaciclib, cisplatin, doxorubicin, etoposide, gemcitabine** [see drug mentions [in context](#)]

Task 1: Increasing efficiency and working together to grow a knowledge structure around the scientific literature

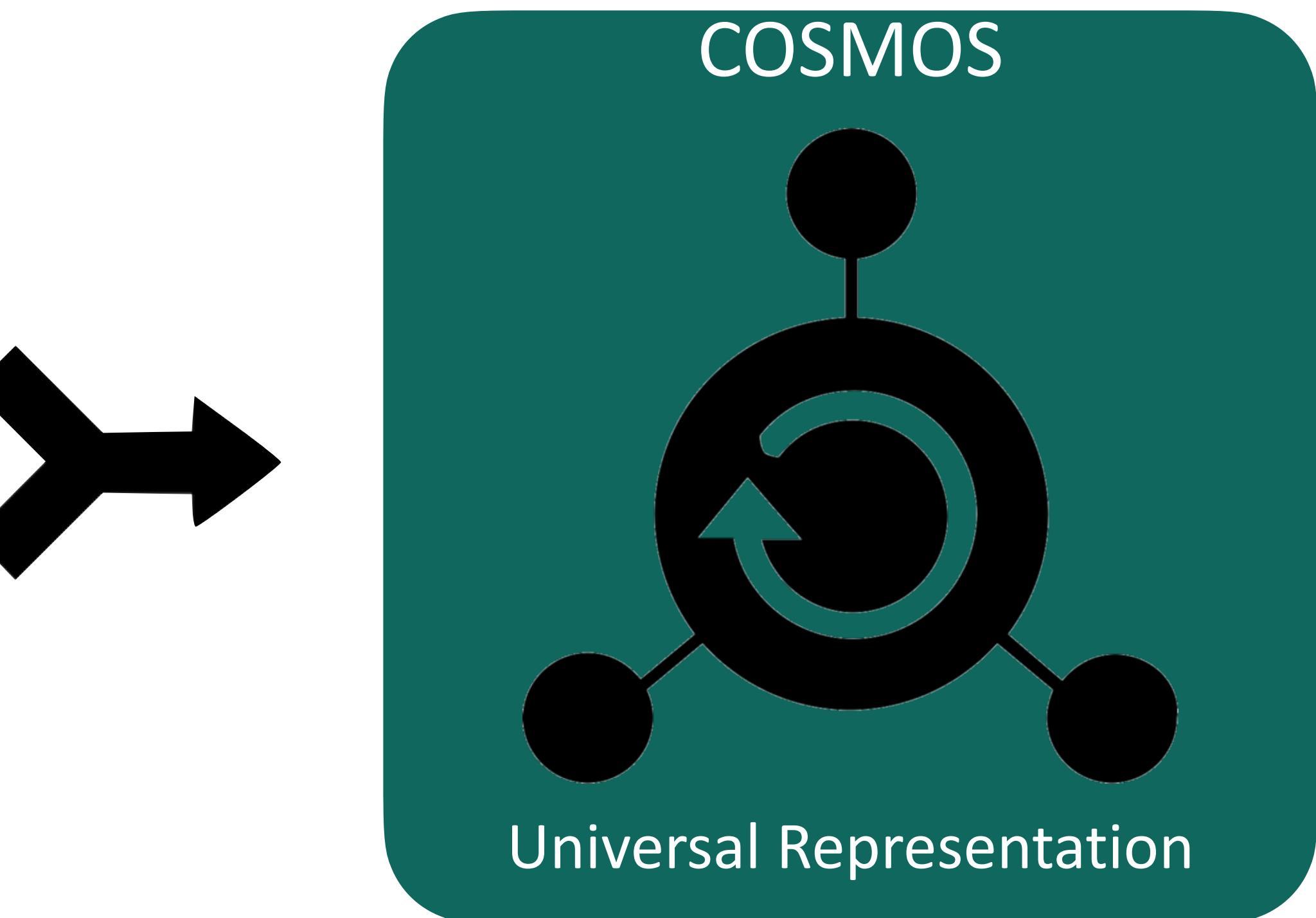
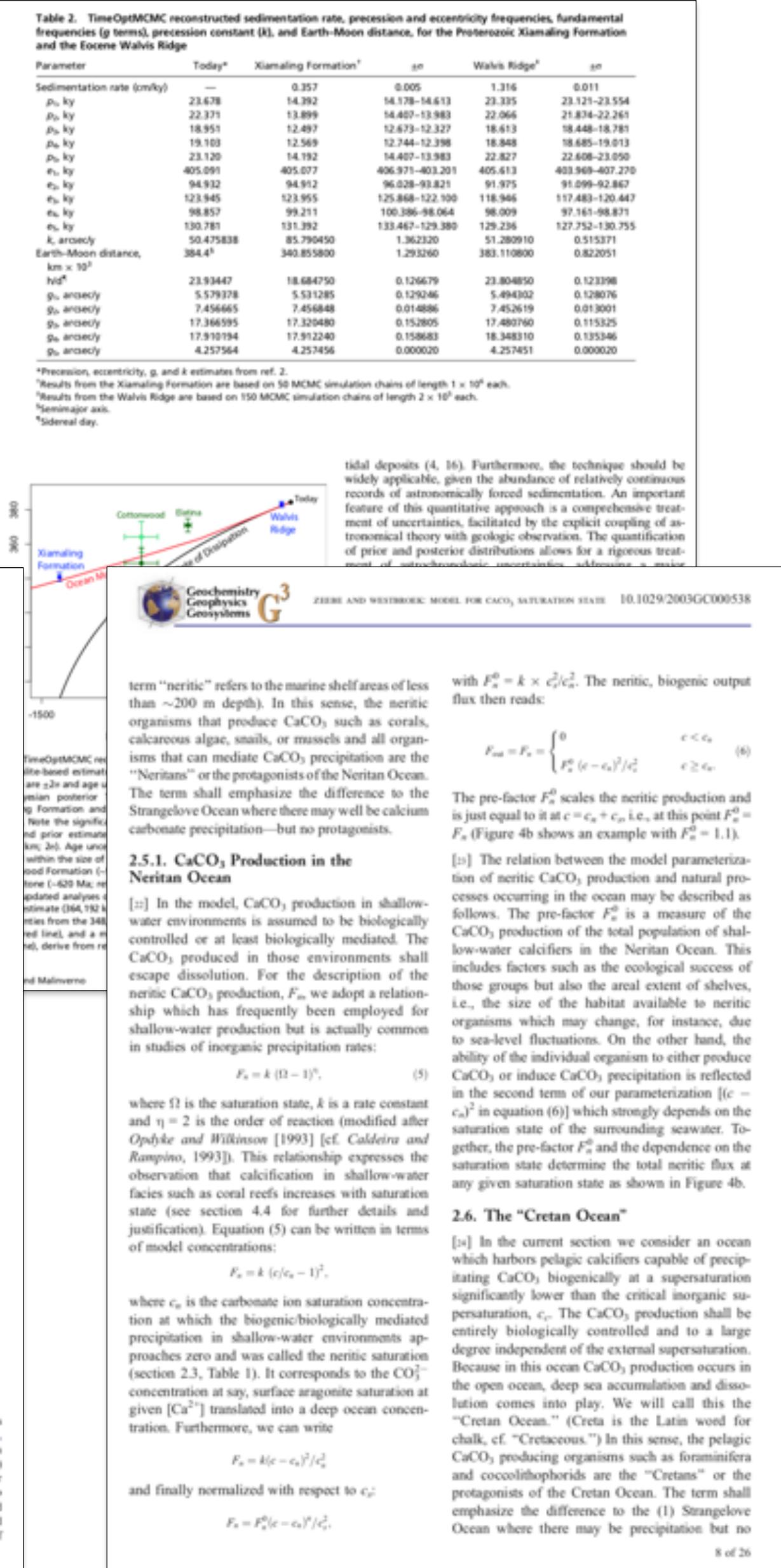
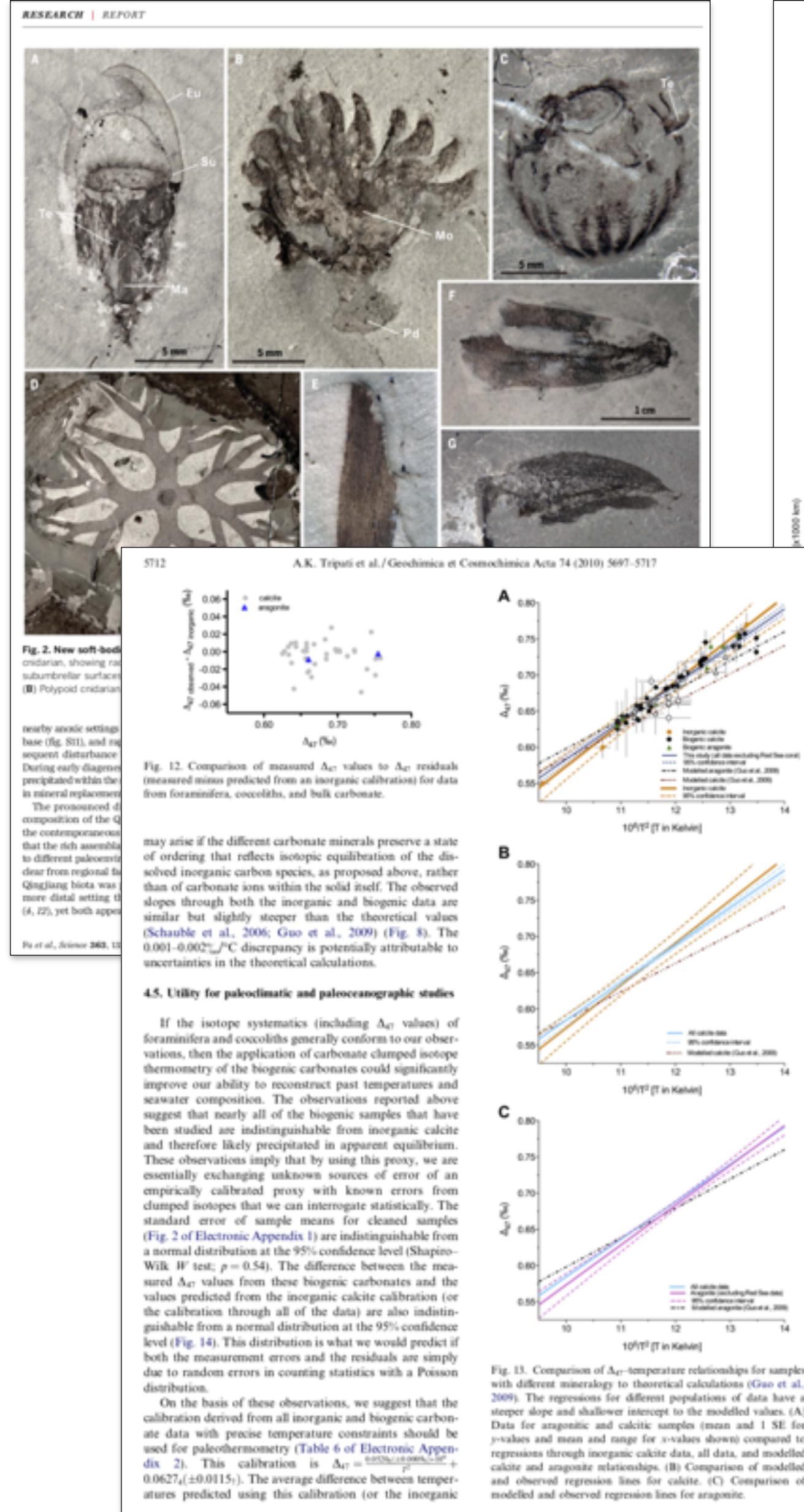
- Example:
 - User develops an application using container template that extracts entity relationships from xDD documents ("x causes y").
 - Output extractions are sent to the user and fed back into xDD to provide additional context and knowledge at document/mention levels, allowing, e.g.:
 - “Give me all tables mentioning <x> or <y> from the literature which provide evidence of <x causes y>.”
 - “Give me all extracted <y> caused by <x> within the context of <covid-19>.”
 - Links back to full document- and mention-level context are never lost! Entire end-to-end updated daily.



Phase 3 Tasks

- **Task 1: Scale and enhance xDD infrastructure and APIs**
 - A. Expand corpus, build new partnerships with publishers
 - B. Develop and release template for collaborator code execution
 - C. Incrementally annotate corpus with collaborator output
 - D. Improve and scale xDD APIs
- **Task 2: Scale and enhance COSMOS retrieval, KBC and APIs**
 - A. Improve and refine visual segmentation
 - B. Incorporate full-document context into text/table/figure/eqn retrieval
 - C. Automate knowledge base construction tasks
 - D. Release public API over COVID-19 COSMOS document set

COSMOS dataset generation



**Fine-grained retrieval,
word/document embeddings,
extraction-ready tables,
equations, figures**

UW-COSMOS Visual Document Segmentation

Page Header (0.971699)

Figure (0.989859)

Page Header (0.982934)

Figure (0.862422)

Page Header (0.627619)

Figure (0.989328)

Section Header (0.304129)

Figure (0.955649)

Figure Caption (0.997292)

Fig. 1. CT features of coronavirus disease-2019 by disease severity. A. An axial CT image in a 33-year-female mild type patient (2 days from symptom onset to CT scan) shows thickening of lung texture. B. An axial CT image in a 37-year-male common type patient (6 days from symptom onset to CT scan) shows multiple ground-glass opacities in both lungs. C. An axial CT image in a 56-year-female severe type patient shows extensive ground-glass opacities and pulmonary consolidation, enlargement of bronchi and vessels. D. An axial CT image in a 47-year-female critical type patient (9 days from symptom onset to CT scan) shows extensive ground-glass opacities in multiple lobes, formatting "white lung".

Body Text (0.992399)

3.2.2. CT manifestations of disease improvement

During follow-up, twelve patients (16 %) dramatically improved. A new CT showed that lesions had decreased in size by more than half in 8 cases (67 %, 8/12) after the patients had received antiviral and supportive treatment after one week of hospital admission. Among them, 3 patients showed remarkable absorption (Fig. 2). Four other patients showed residual interstitial abnormalities with persisting septal lines.

Section Header (0.997304)

Body Text (0.997304)

(0.987588)

RT-PCR of sputum, throat swab and lower respiratory tract secretion or sequencing of virus gene represents the gold standard technique for the diagnosis of COVID-19 [8,9]. However, the testing requires at least several hours, and has a false negative rate of more

Table (0.989456)

Table 2

The location and morphology of pulmonary lesions in different types in Coronavirus Disease-19 patients (n = 149).

	mild type (n = 40)	common type (n = 40)	severe type (n = 21)	critical type (n = 3)
Unilateral Lung Involvement	0	15 (35 %)	0	0
Bilateral Lung Involvement	3 (50 %)	28 (65 %)	23 (100 %)	3 (100 %)
Unique Ground-glass Opacities	0	12 (28 %)	0	3 (100 %)
Multiple Ground-glass Opacities	0	31 (72 %)	16 (76 %)	3 (100 %)
Paving Stone Sign	0	15 (35 %)	10 (48 %)	3 (100 %)
Consolidation	0	6	5 (24 %)	3 (100 %)
Bronchial Wall Thickening	0	2 (5%)	14 (67 %)	3 (100 %)
Pleural Effusion	0	0	0	3 (100 %)
Thickening of Lung Texture	3 (50 %)	40 (90 %)	19 (90 %)	3 (100 %)
No Lung Abnormality	3 (50 %)	0	0	0

Page Header (0.987905)

W. Tang, Q. Cai and L. Qin et al. / Journal of Infection 80 (2020) 388–393

Table (0.776724)

Table 1

Demographics, baseline characteristics of overall 149 patients.

General data	Overall (N = 149)
Age	45.11 ± 13.35
Sex	Male/Female(81/68)
Height (cm)	167 (11.25)
Weight (kg)	65.92 (18.75)
BMI	23.75 (4.54)
Exposure history	
Stay in Wuhan	80 (53.09%)
stay in Hubei Province except Wuhan	5 (3.36%)
Contact with people from Hubei Province	49 (32.89%)
No relation with Hubei Province	15 (10.06%)
Chronic disease	
cardio-cerebrovascular disease	28 (18.79%)
digestive system diseases	8 (5.37%)
endocrine diseases	9 (6.04%)
malignant tumor	2 (1.34%)
neural system diseases	0 (0%)
respiratory system diseases	1 (0.67%)
others	4 (2.68%)
Temperature (°C)	37.86 ± 0.87
pH	7.11 (0.04)
increased	21 (14.09%)
decreased	4 (2.68%)
PaO ₂ (kpa, range 80–100)	96.05 (25.58)
decreased	23 (15.44%), 56.25 ± 6.58
SaO ₂ (range ≥ 95%)	97.43% (2%)
decreased	14 (9.41%)
Heart rate (beats per minute)	88.63 ± 14.27
Systolic pressure (mmHg)	129.98 ± 15.44
Diastolic pressure (mmHg)	81.69 ± 10.16

BMI: body mass index; pH: hydrogen ion concentration.

Table (0.990325)

Table 2

Signs and symptoms at admission, treatment and clinic outcome of overall 149 patients.

General data	Overall (N = 149)
Onset-inpatient interval (day)	6.83 (5)
Signs and symptoms at admission	
fever	114 (76.51%)
cough	87 (58.39%)
expectoration	48 (32.21%)
dyspnea	2 (1.34%)
muscle pain	5 (3.36%)
headache	13 (8.72%)
sore throat	23 (14.99%)
snotty	5 (3.36%)
chest pain	5 (3.36%)
chest tightness	16 (10.24%)
chill	23 (14.99%)
diarrhea	13 (7.38%)
nausea and vomiting	2 (1.34%)
CURB-65 score	0/1 (134/15)
PSL grading	I/II/III/IV/V (82/54/10/3/0)
Treatment	N = 149
Oxygen therapy	134 (89.93%)
NEV	2 (1.34%)
IMV	0 (0%)
ICU	0 (0%)
Antibiotic treatment	34 (22.82%)
Antifungal treatment	0 (0%)
Antiviral treatment	140 (93.96%)
Interferon administration	144 (96.64%)
Glucocorticoids	5 (3.36%)
Immunoglobulin therapy	19 (12.75%)

ALT: alanine aminotransferase; AST: aspartate aminotransferase; CK: creatine kinase.

Page Header (0.970255)

Journal of Paediatrics and Child Health

Page Header (0.972339)

Figure (0.969785)

Page Header (0.800970)

BRIEF COMMUNICATION

Body Text (0.808039)

Three children who recovered from novel coronavirus 2019 pneumonia

Body Text (0.997827)

In December 2019, a cluster of acute respiratory illness, now known as novel COVID-19, occurred in Wuhan, Hubei Province, China.^{1–3} The disease rapidly spread from Wuhan to other areas. Previous studies suggest that COVID-19 is more likely to infect older adult men, particularly those with chronic comorbidities.^{4–6} In the isolation ward of the children's hospital affiliated to Zhengzhou University, three children were hospitalised with pneumonia caused by 2019 novel coronavirus (COVID-19). Two were sisters, aged 6 and 8 years old and one was a 6-month-old boy. All three patients had fever, and two had nasal congestion and rhinitis, associated with fatigue, diarrhoea and headache. The 6-year-old girl mainly had cough. None had dyspnoea or cyanosis. Their computerised tomographic scans are shown in Figures 1–3. None of the children required intensive care or mechanical ventilation or had any severe complications.

The time between admission and diagnosis was 2 days. Families of all three children had at least one infected relative, with the children's infection occurring after the parents' infection. The two sisters had family members who visited Wuhan. The other child had no direct link to Wuhan.

Throat swab, sputum, stool and blood samples were tested for COVID-19 nucleic acid using reverse-transcriptase polymerase chain reaction. All three children are confirmed cases. According to the current conditions of the reported cases, the three children mainly belong to family cluster cases.

The children were closely monitored in hospital. The two sisters, but not the infant, were treated with nebulised interferon- α 2b 100 000 IU/kg, twice daily for 7 days.

The children's fever resolved within 3 days and after 1 week their symptoms improved significantly. The children were discharged after 10 days, when they had two consecutive negative polymerase chain reaction tests of respiratory specimens at least 1 day apart. However, they required home isolation for 14 days after discharge. The girls were given psychological support by nursing staff. Strict infection control measures were instituted, although the 6-month-old needed special isolation because he was unable to wear a mask.

Very few cases have been reported of children infected with COVID-19.^{1–3} The Children's Hospital affiliated to Zhengzhou University is a first-class Children's Hospital in the province. It was designated as one of the regional medical centres by the state in 2019. It is also the designated hospital for medical treatment of children with COVID-19 pneumonia, and is the children's medical centre of Henan Province. At present, the hospital has sent more than medical staff to support Wuhan. The three children were the first children to be managed in Henan Province. Through everybody's effort, careful treatment, psychological nursing support and infection control measures in hospital, these

Figure Caption (0.986687)

Fig. 3 Computerised tomographic scan of 6-month-old boy.

Figure (0.990268)

Figure Caption (0.992492)

Fig. 1 Computerised tomographic scan of 8-year-old girl.

Figure (0.981846)

COSMOS AI-powered results

COSMOS AI-constructed KB

Not Secure — cosmos3.ctc.wisc.edu:80

COSMOS View extractions Usage Permalink Next image >

Page Header (0.975626)
bioRxiv; Genetics and Evolution 81; 2020; 394298

Figure (0.983779)
(0.983223) China/Wuhan/2020/B71_20L_40K70 00TTSAGTCCTTAACTGACTTTGGACCTTCTTGTAGGAGCTTCAACATTTAAGATTC
Australia/VIC01/2020/B71_20L_4 00TTSAGTCCTTAACTGACTTTGGACCTTCTTGTAGGAGCTTCAACATTTAAGATTC
USA/NC1/2020/B71_20L_40K70 00TTSAGTCCTTAACTGACTTTGGACCTTCTTGTAGGAGCTTCAACATTTAAGATTC
Japan/ICU-004/2020/B71_20L_40 00TTSAGTCCTTAACTGACTTTGGACCTTCTTGTAGGAGCTTCAACATTTAAGATTC
France/IDF073/2020/B71_20L_40 00TTSAGTCCTTAACTGACTTTGGACCTTCTTGTAGGAGCTTCAACATTTAAGATTC
Germany/Berlin/2020/B71_20L_4 00TTSAGTCCTTAACTGACTTTGGACCTTCTTGTAGGAGCTTCAACATTTAAGATTC

Deletion

5' — ORF1ab polyprotein — Spike — Envelope — Membrane — Nucleocapsid — 3'

Deletion

China/Wuhan/2020/B71_20L_40K70 TGGTGAATTCTTGAACATATAATTTAGAAGCTATGCTTGAACATATTTAAGATTC
Australia/VIC01/2020/B71_20L_4 TGGTGAATTCTTGAACATATAATTTAGAAGCTATGCTTGAACATATTTAAGATTC
USA/NC1/2020/B71_20L_40K70 TGGTGAATTCTTGAACATATAATTTAGAAGCTATGCTTGAACATATTTAAGATTC
Japan/ICU-004/2020/B71_20L_40 TGGTGAATTCTTGAACATATAATTTAGAAGCTATGCTTGAACATATTTAAGATTC
France/IDF073/2020/B71_20L_40 TGGTGAATTCTTGAACATATAATTTAGAAGCTATGCTTGAACATATTTAAGATTC
Germany/Berlin/2020/B71_20L_4 TGGTGAATTCTTGAACATATAATTTAGAAGCTATGCTTGAACATATTTAAGATTC

Figure Caption (0.994004)
Fig. 1. Genomic organization of SARS-CoV-2 and pairwise nucleotide sequence alignment showing deletions in the ORF1ab polyprotein and in the 3' end of the genome.

Table (0.992867)
Table 1

Mutations found in the entire genome of SARS-CoV-2 strains. The number in the parentheses indicated the location of amino acid in its protein.

Genomic region	No. of mutations	Misense mutation	SARS-CoV-2 strain
5' UTR	8	N/A	
ORF1ab polyprotein	48	29 A (337) → T P (309) → S S (420) → N T (649) → I A (3179) → V I (1599) → F I (1467) → V M (2194) → T I (2293) → I L (2244) → T	USA/CA3/2020/B71_20L_408008 USA/CA4/2020/B71_20L_408009 France/TDF0115/2020/B71_20L_408430 USA/CA1/2020/B71_20L_406034 USA/CA5/2020/B71_20L_408010 Japan/TY-WK-012/2020/B71_20L_408643 Korea/ICDC03/2020/B71_20L_407193 USA/CA3/2020/B71_20L_408008 USA/CA4/2020/B71_20L_408009 Shenzhen/SDTH-004/2020/B71_20L_408395 Wuhan/WHR1/2019/B71_20L_406796 Wuhan/SPBCAN-BH-03/2019/B71_20L_409093

Table: page_objects Table: object_contexts

Select data Show structure Alter table New item Select data Show structure Alter table

Column	Type	Comment	Column	Type	Comment
id	int(11) Auto Increment		id	int(11) Auto Increment	
page_id	int(11) NULL		pdf_id	int(11) NULL	
context_id	int(11) NULL		cls	varchar(200) NULL	
bytes	longblob NULL		header_id	int(11) NULL	
content	varchar(10000) NULL		header_content	text NULL	
bounding_box	json NULL		content	longtext NULL	
init_cls_confidences	json NULL				
cls	varchar(200) NULL				
pp_rule_cls	varchar(200) NULL				
annotated_cls	varchar(200) NULL				
confidence	decimal(9,6) NULL				
classification_success	tinyint(1) NULL				
proposal_success	tinyint(1) NULL				

Indexes

PRIMARY	id
INDEX	pdf_id
INDEX	header_id

Alter indexes

- Convert all modalities in heterogeneous PDFs into a structured representation
- Throughput of approximately one page every 2 seconds (GPU accelerated)
- Ability to deploy in parallel over distributed (CPU and/or GPU) nodes via HTCondor

Task 2A: Improve COSMOS visual segmentation

Incorporate user annotations into continual model refinement

- Improve and deploy document-tagging functionality for model training and evaluation
- Leverage our support for tagging complex data types and relations (e.g. linked information hierarchies)

geochimical observations based on sediment P/AES during dissolved, associated with the surface of black carbon, or partitioned into NOC. Thus, the distribution of PAHs can be expressed as

$$K_F = f_{BC}K_{OC} + f_{BC}K_{BC}$$
 Equation (4)

where K_F is now the PAH partition coefficient based on two particulate phases, NOC and BC; f_{BC} is the fraction BC in the soil or sediment; and K_{BC} is the BC-normalized partition coefficient. Recently, Bücheli and Gustafsson [21] and Accardini-Dey and Gschwend [17] revised this equation to include a nonlinear term:

$$K_F = f_{BC}K_{OC} + f_{BC}K_{BC}C_B^{n-1}$$
 Equation (5)

in which n is the Freundlich exponent. This term recognizes the nonlinear adsorption interaction between the HOC and BC rather than the purely linear absorptive mechanism in Equation

COSMOS View training data Usage Permalink Next image >

Page Header
4 P. LI ET AL

Table Caption
Table Effects of vitamin B6 on plasma AChE activity and PLP concentrations in rats fed with isocarbophos.

	AChE activity (nmol/min/ml)	PLP concentration (nmol/l)	sBP (mmHg)	dBP (mmHg)
Saline	2.11 ± 0.37	20.4 ± 5.3	128 ± 21	76 ± 25
Vitamin B6	2.16 ± 0.36	28.9 ± 2.4*	124 ± 20	72 ± 13
Eliprodil	2.09 ± 0.29	20.1 ± 4.7	130 ± 17	86 ± 9
Isocarbophos	1.98 ± 0.30	22.9 ± 3.5	137 ± 29	84 ± 10
Isocarbophos + vitamin B6	2.07 ± 0.31	27.6 ± 3.7*	126 ± 23	77 ± 13
Isocarbophos + vitamin B6 + Eliprodil	2.15 ± 0.27	26.1 ± 5.1#	138 ± 15	79 ± 17

Table Note After isocarbophos treatment, blood pressure was determined by the tail cuff method. Blood was collected to measure the levels of plasma AChE activity and PLP concentrations in rats. Data were expressed as means ± SD. N was 6 in each group. One-way ANOVA was used for statistical analyses. * $p < 0.05$ VS Saline, # $p < 0.05$ VS Isocarbophos.

Table Caption
Table Effects of vitamin B6 on the blood flow of posterior cerebral artery (PCA) in rats fed with isocarbophos.

Groups	Right PCA		Left PCA	
	V_L (cm/s)	V_d (cm/s)	V_L (cm/s)	V_d (cm/s)
Saline	26.32 ± 4.21	12.35 ± 2.41	26.85 ± 4.66	13.54 ± 2.55
Vitamin B6	24.63 ± 5.39	11.26 ± 3.41	27.52 ± 5.98	12.54 ± 3.63
Eliprodil	21.58 ± 4.26	9.66 ± 1.58	20.51 ± 3.64*	10.49 ± 2.53
Isocarbophos	15.74 ± 3.18*	8.57 ± 1.36*	17.41 ± 3.15*	8.26 ± 1.64*
Isocarbophos + vitamin B6	26.79 ± 3.54*	13.66 ± 2.08*	26.12 ± 4.66*	13.52 ± 2.36*
Isocarbophos + vitamin B6 + Eliprodil	15.25 ± 3.64†	7.65 ± 1.54†	15.88 ± 3.32†	7.24 ± 1.62†

Table Note After isocarbophos treatment, blood flow of PCA was determined by TCD before sacrifice. Data were expressed as means ± SD. One-way ANOVA was used for statistical analyses. N was 6 in each group. * $p < 0.05$ VS Saline, * $p < 0.05$ VS Isocarbophos, † $p < 0.05$ VS Isocarbophos + vitamin B6.

Body Text
Supplementation of vitamin B6 significantly recovered the blood flow of PCAs in isocarbophos-treated rats. Furthermore, the addition of eliprodil, which is a selective NR2B antagonist [27], abolished the effects of vitamin B6 on the blood flow of PCAs. There were no obvious differences of vascular thickness among groups as determined by HE staining (Figure 1B). Taking these data together, it can be suggested that vitamin B6 via NR2B activation preserved the cerebral blood flow in isocarbophos-treated rats.

Section Header
Vitamin B6 improves isocarbophos-impaired learning and memory, which is suppressed by eliprodil in rats

Body Text
The typical phenotype of dementia is the impairment of the spatial memory [28,29]. Thus, the learning and memory behaviors were evaluated by the MWM test. As shown in Table 3, after 12-week isocarbophos administration, swimming distance in quadrant

Other

University of Wisconsin - Madison at 16:03 04 January 2018

Section Header
Vitamin B6 increases CaMK-II expression in rat hippocampus

Body Text
CaMK-II has been demonstrated to potentiate NMDA receptor function by increasing the fraction of receptors in the synaptosomal membranes [32]. Thus, we investigated whether isocarbophos altered the CaMK-II protein expression in the

Task 2A: Improve COSMOS visual segmentation

Incorporate user annotations into continual model refinement

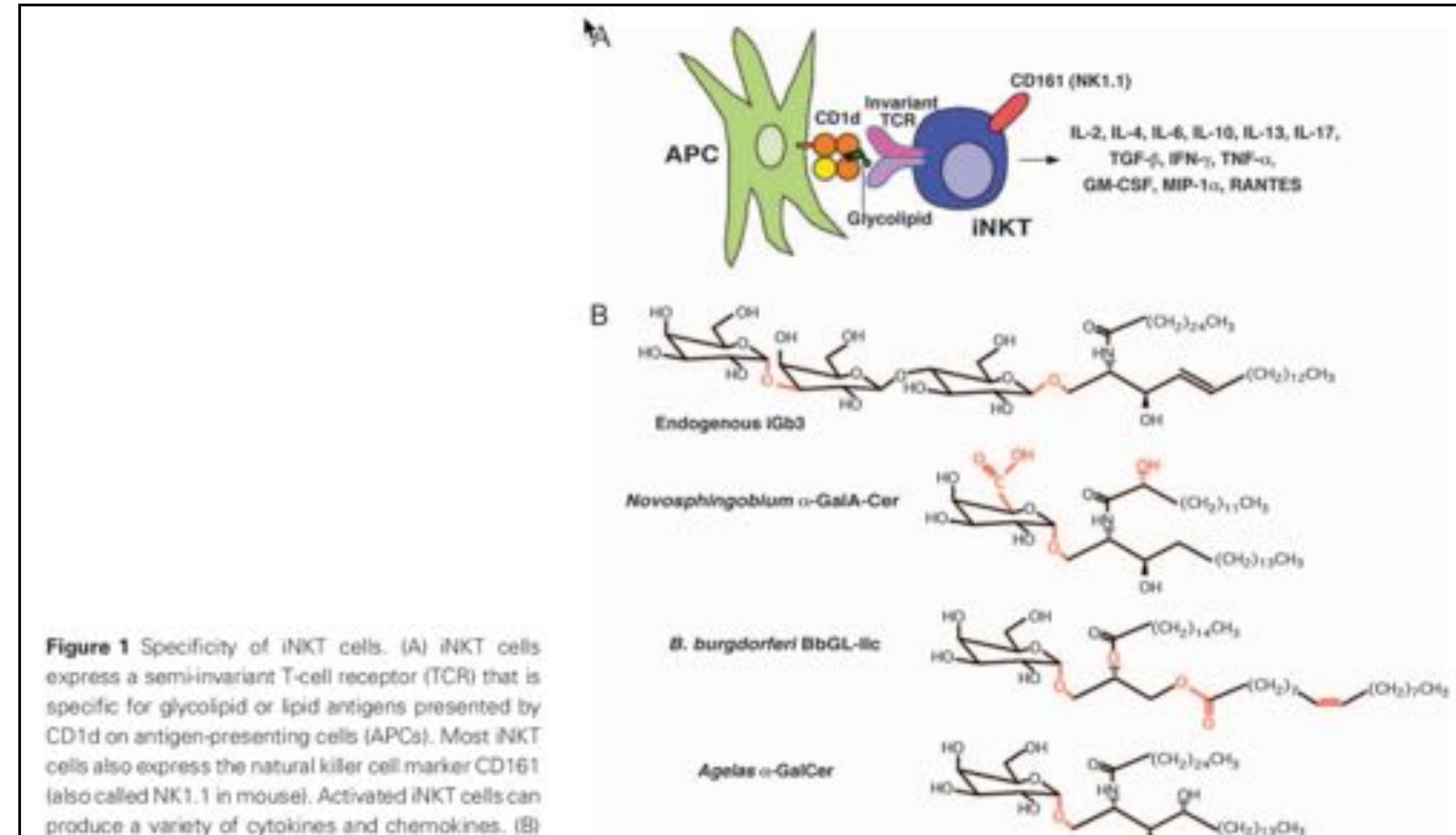
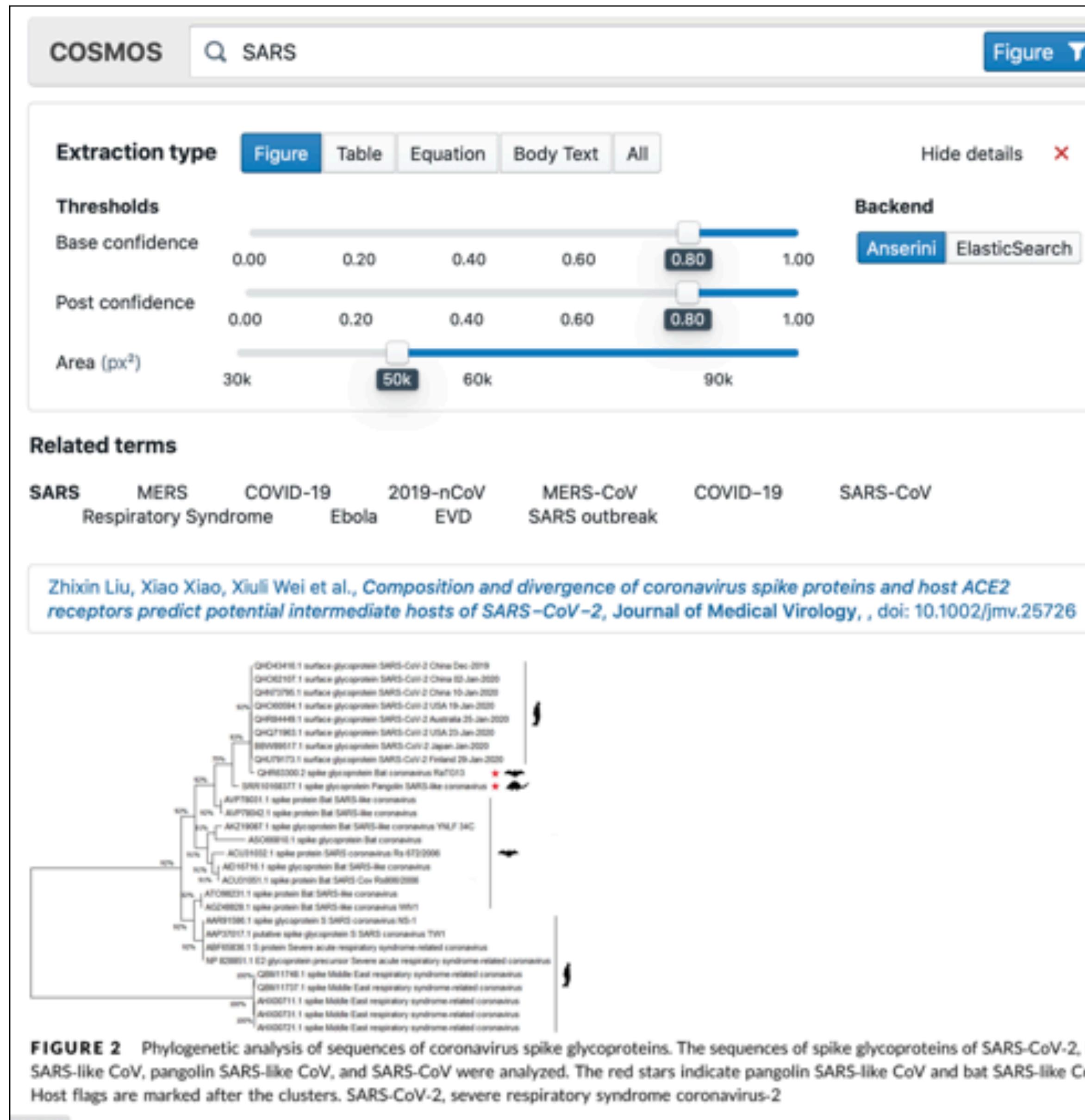


Figure 1 Specificity of iNKT cells. (A) iNKT cells express a semi-invariant T-cell receptor (TCR) that is specific for glycolipid or lipid antigens presented by CD1d on antigen-presenting cells (APCs). Most iNKT cells also express the natural killer cell marker CD161 (also called NK1.1 in mouse). Activated iNKT cells can produce a variety of cytokines and chemokines. (B)

- COSMOS web browser-based software for manual tagging of objects in a visual reference space (deployable over any image for any purpose); assessing AI proposals
- Interface to make explicit links between tagged objects (e.g., variables to equations)
- Need to leverage this in order to improve visual model and add modalities

Task 2B: Add document-level contextual retrieval

Incorporate body text descriptions into table and figure retrieval



- Frontend web app surfacing COSMOS KB and micro services, including
 - word embedding model trained over target corpus
 - two modes for retrieval: ElasticSearch index over extracted objects, TF-IDF based retrieval
 - preliminary data-frame extraction for tables

<https://cosmos.wisc.edu/sets/covid>

Task 2B: Add document-level contextual retrieval

Incorporate body text descriptions into table and figure retrieval

Table Reference

TABLE. Hospitalization, intensive care unit (ICU) admission, and case-fatality percentages for reported COVID-19 cases, by age group — United States, February 12–March 16, 2020

Age group (yrs) (no. of cases)	Hospitalization	ICU admission	Case-fatality
0–19 (123)	1.6–2.5	0	0
20–44 (705)	14.3–20.8	2.0–4.2	0.1–0.2
45–54 (429)	21.2–28.3	5.4–10.4	0.5–0.8
55–64 (429)	20.5–30.1	4.7–11.2	1.4–2.6
65–74 (409)	28.6–43.5	8.1–18.8	2.7–4.9
75–84 (210)	30.5–58.7	10.5–31.0	4.3–10.5
≥85 (144)	31.3–70.3	6.3–29.0	10.4–27.3
Total (2,449)	20.7–31.4	4.9–11.5	1.8–3.4

* Lower bound of range = number of persons hospitalized, admitted to ICU, or who died among total in age group; upper bound of range = number of persons hospitalized, admitted to ICU, or who died among total in age group with known hospitalization status, ICU admission status, or death.

Table Note

Table Caption

Table

Table Reference Context

aged ≤19 years. Percentages of ICU admissions were lowest among adults aged 20–44 years (2%–4%) and highest among adults aged 75–84 years (11%–31%) (Table).

Table Reference

Among 44 cases with known outcome, 15 (34%) deaths were reported among adults aged ≥85 years, 20 (46%) among adults aged 65–84 years, and nine (20%) among adults aged 20–64 years. Case-fatality percentages increased with increasing

age, from no deaths reported among persons aged ≤19 years to highest percentages (10%–27%) among adults aged ≥85 years (Table) (Figure 2).

Table Reference Context

Table Reference

Among 508 (12%) patients known to have been hospitalized, 9% were aged ≥85 years, 36% were aged 65–84 years, 17% were aged 55–64 years, 18% were 45–54 years, and 20% were aged 20–44 years. Less than 1% of hospitalizations were among persons aged ≤19 years (Figure 2). The percentage of persons hospitalized increased with age, from 2%–3% among persons aged ≤9 years, to ≥31% among adults aged ≥85 years. (Table).

Table Reference Context

Table Reference

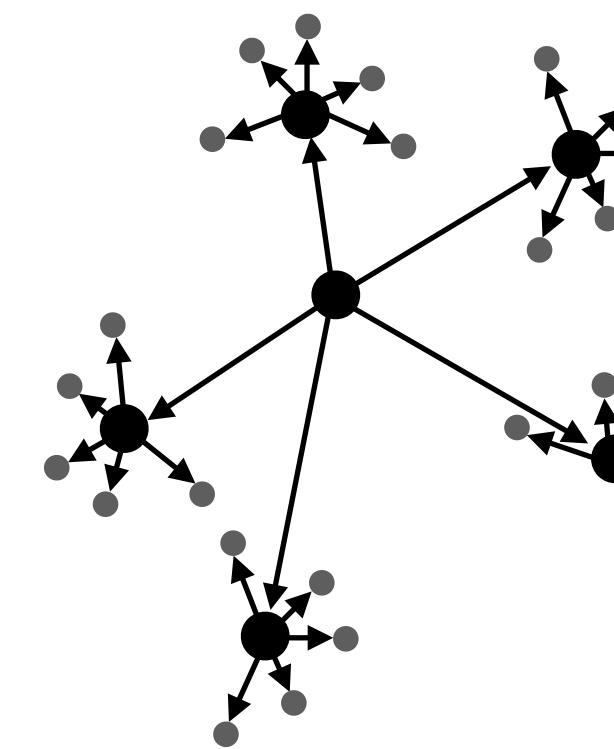
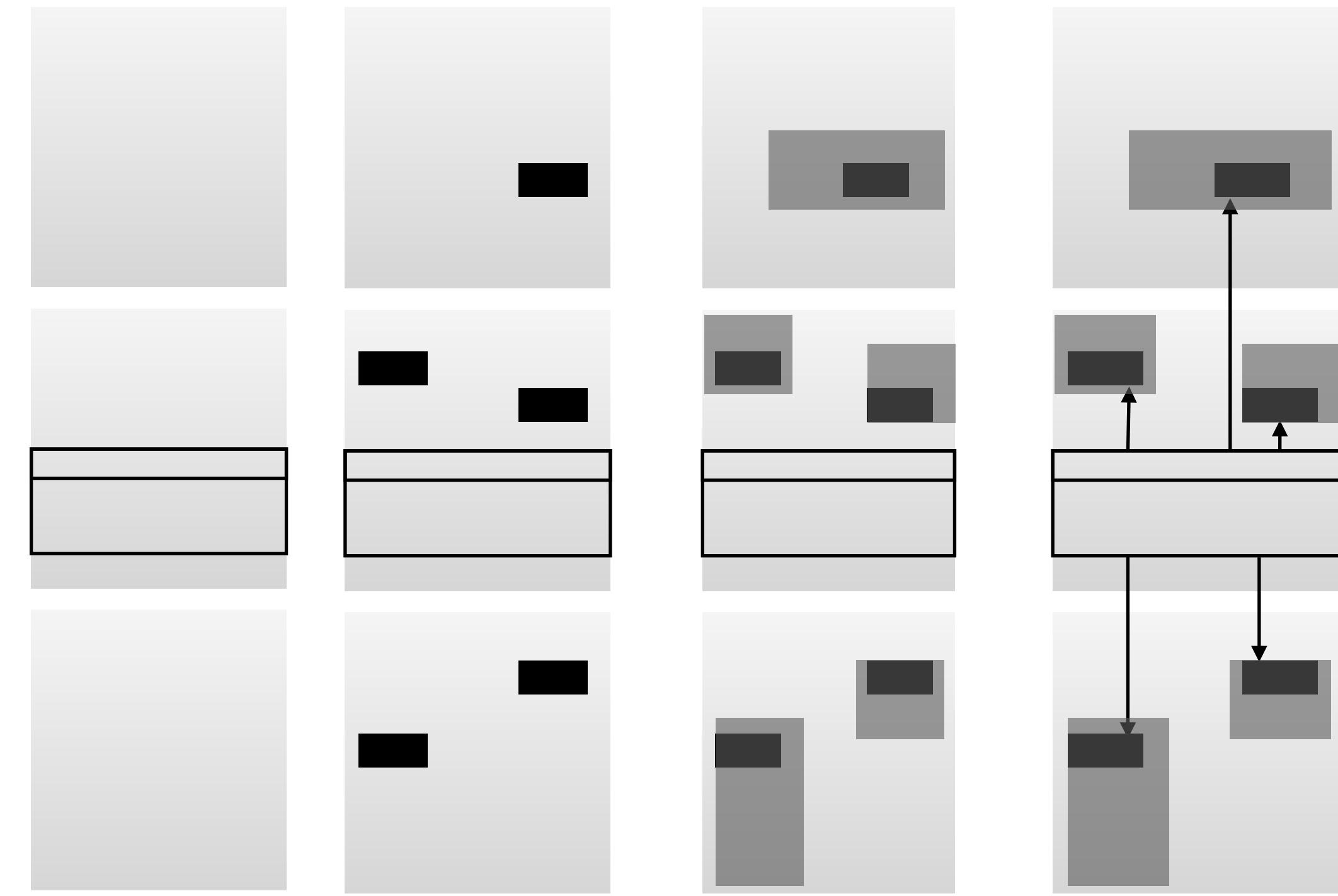
COVID-19
Admissions

Returns the original table, and the natural language descriptions:

- Percentage hospitalized increases with age
- Fatalities increase with age
- ICU admissions lowest among adults aged 20–44
- ICU admissions highest among adults aged 75–84

Task 2B: Document-level contextual retrieval

Incorporate body text descriptions into table and figure retrieval



1. —
2. —
3. —
4. —
5. —

1. —
2. —
3. —
4. —
5. —

1. —
2. —
3. —
4. —
5. —

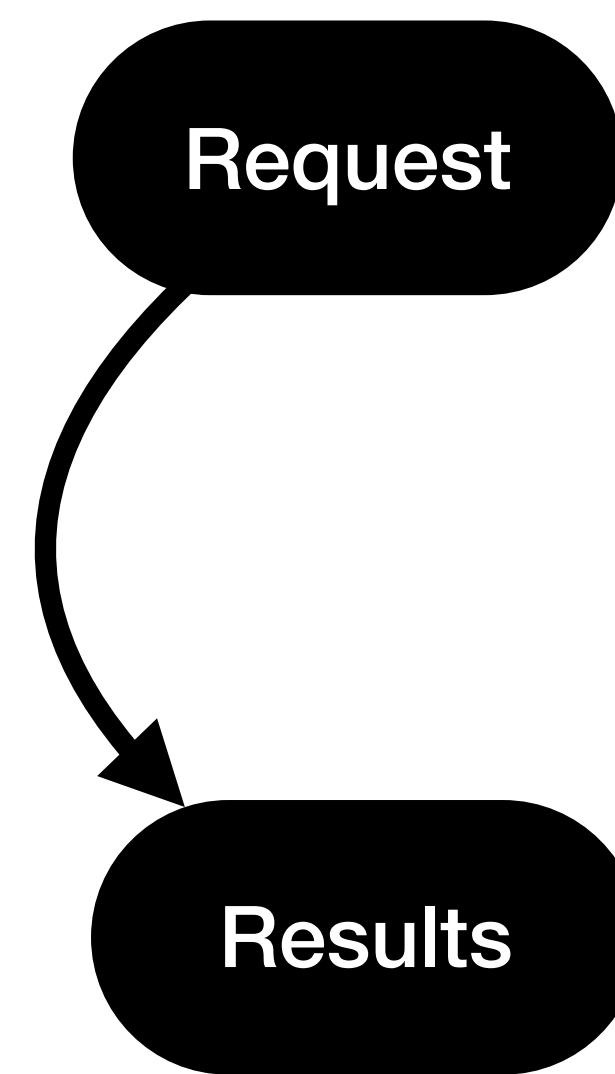
1. —
2. —
3. —
4. —
5. —

1. —
2. —
3. —
4. —
5. —

index the graph structures in a single document

1. —
2. —
3. —
4. —
5. —

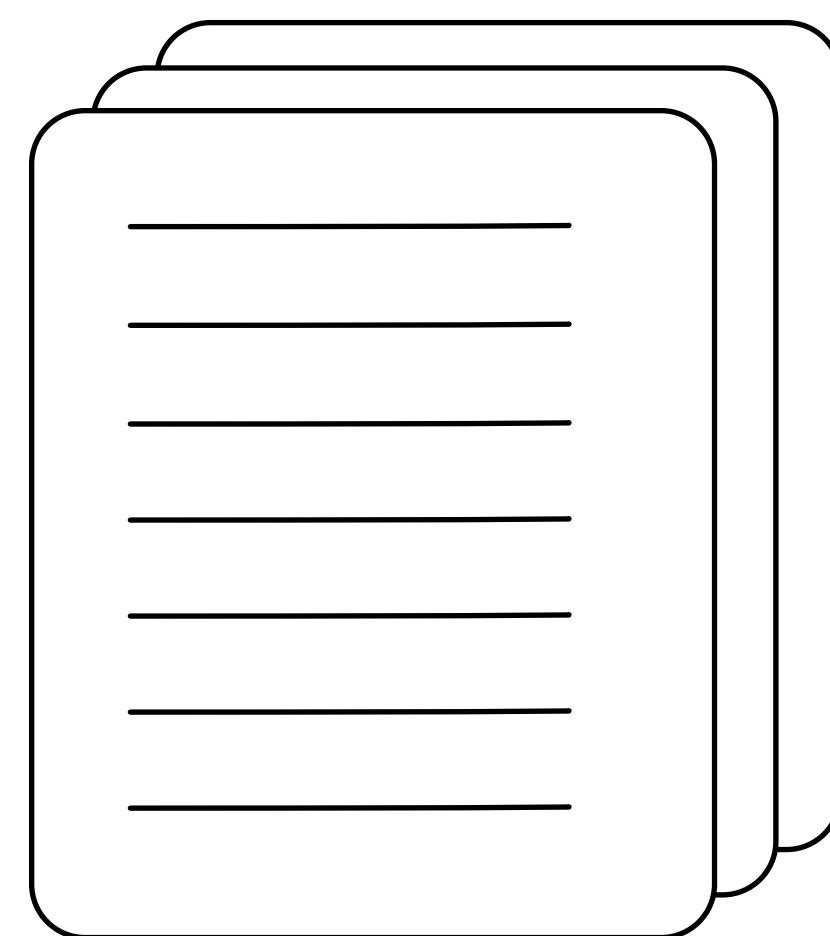
index the graph structures across all documents in corpus



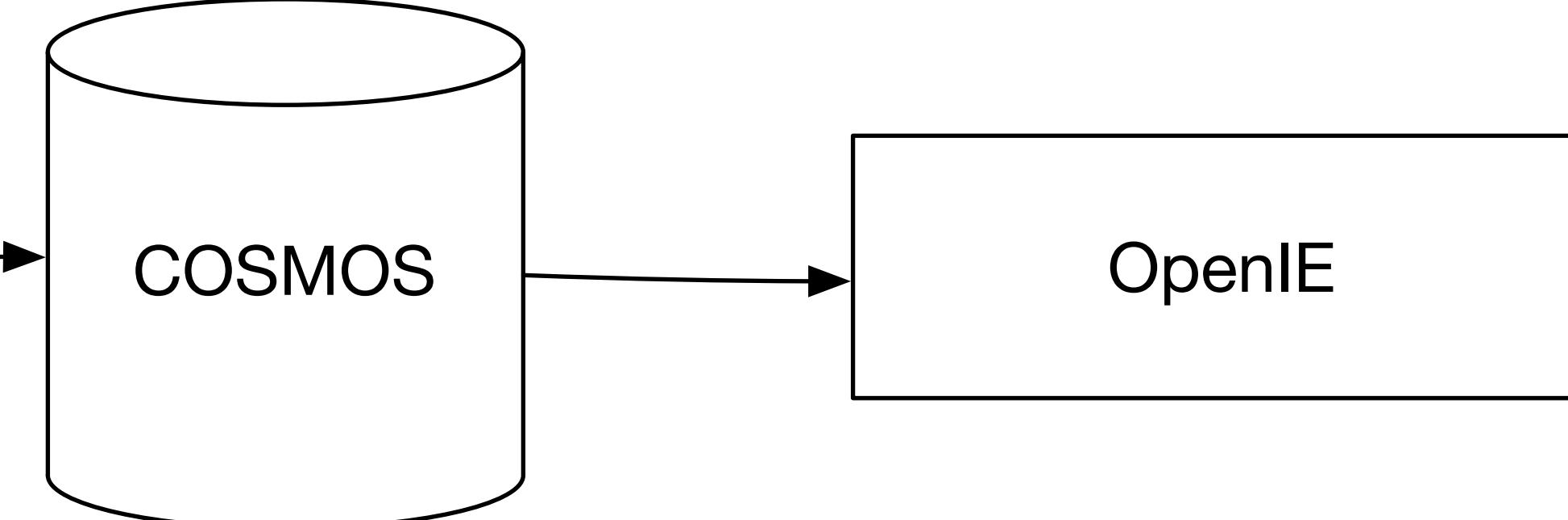
Retrieve object plus body text content with query

Task 2C: Automated KBC with CosmosIE and Marius

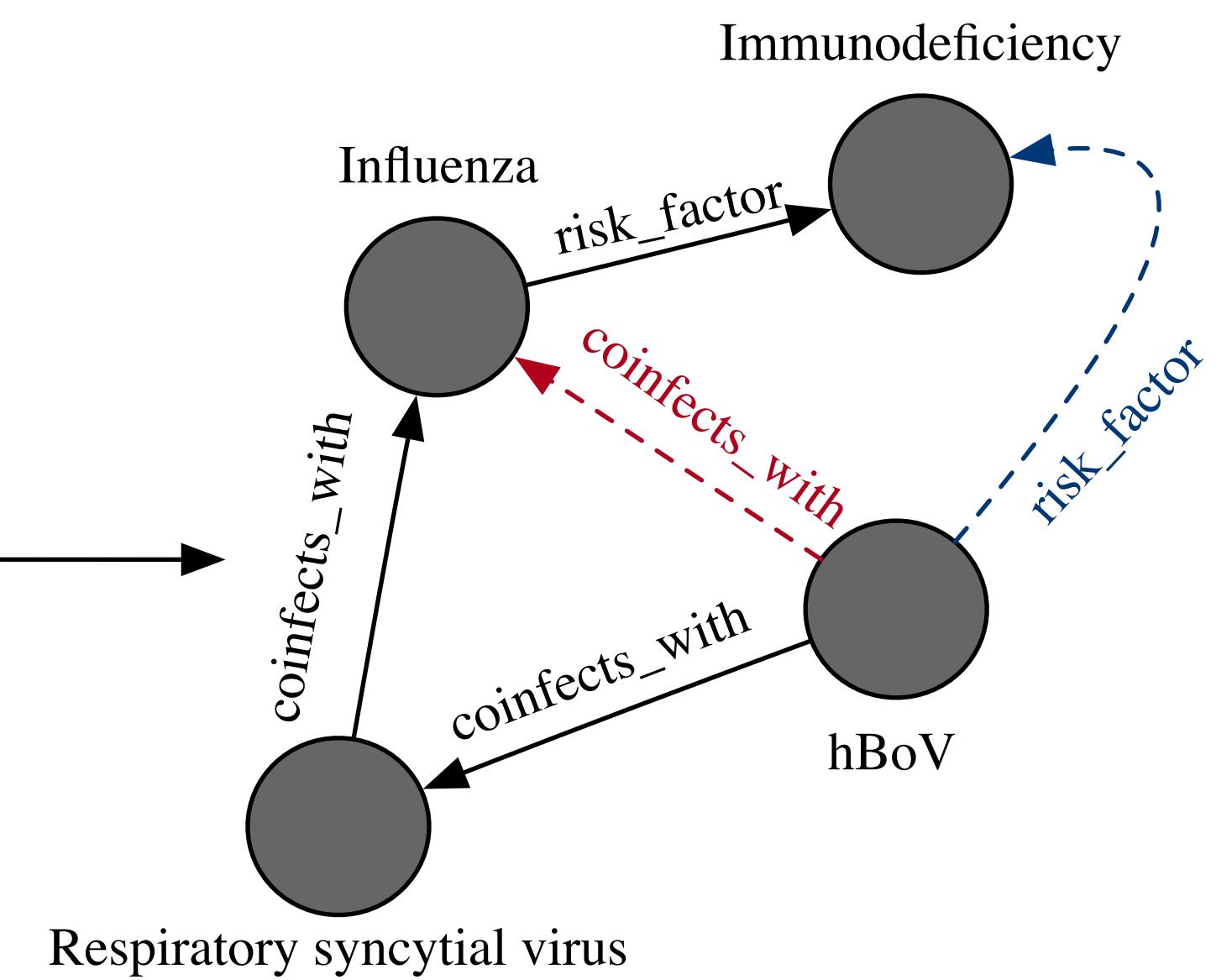
- Use state of the art natural language models to extract knowledge graphs from target corpora, at scale
- Deploy state of the art scalable graph embedding framework, Marius, for knowledge base completion



Target Corpus



OpenIE

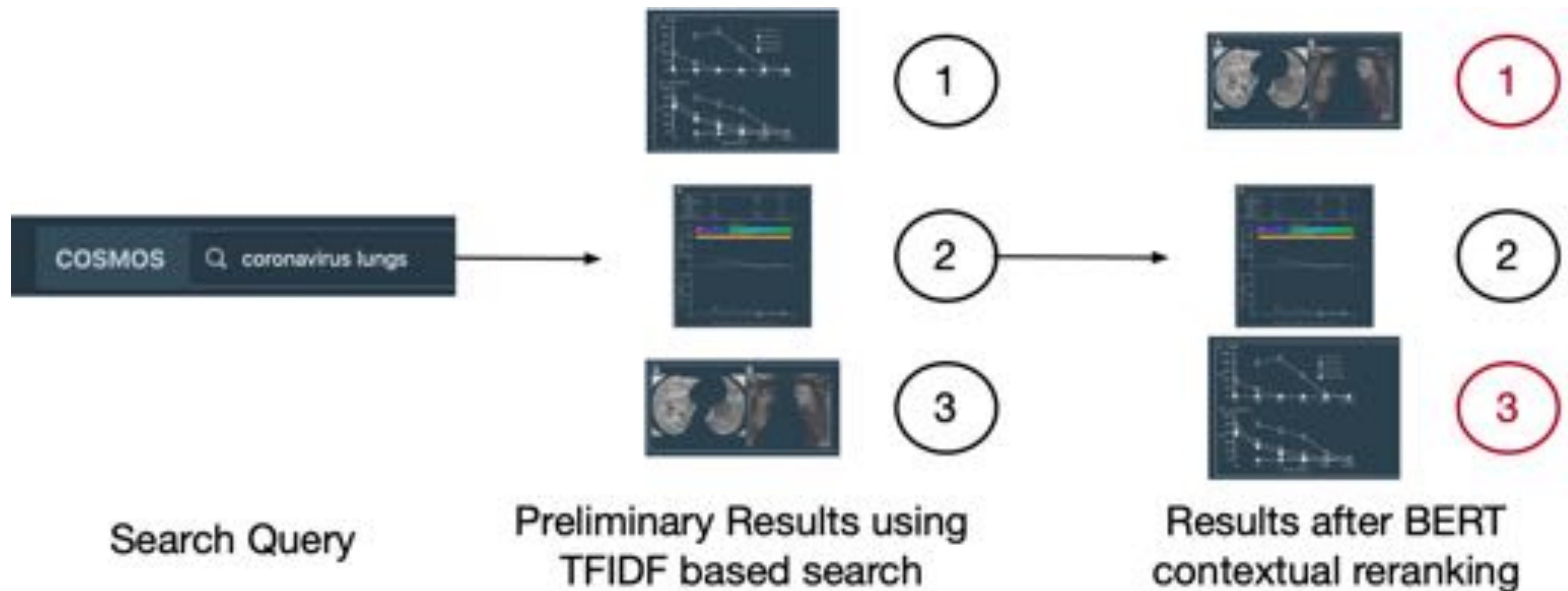


Marius Knowledge Completion

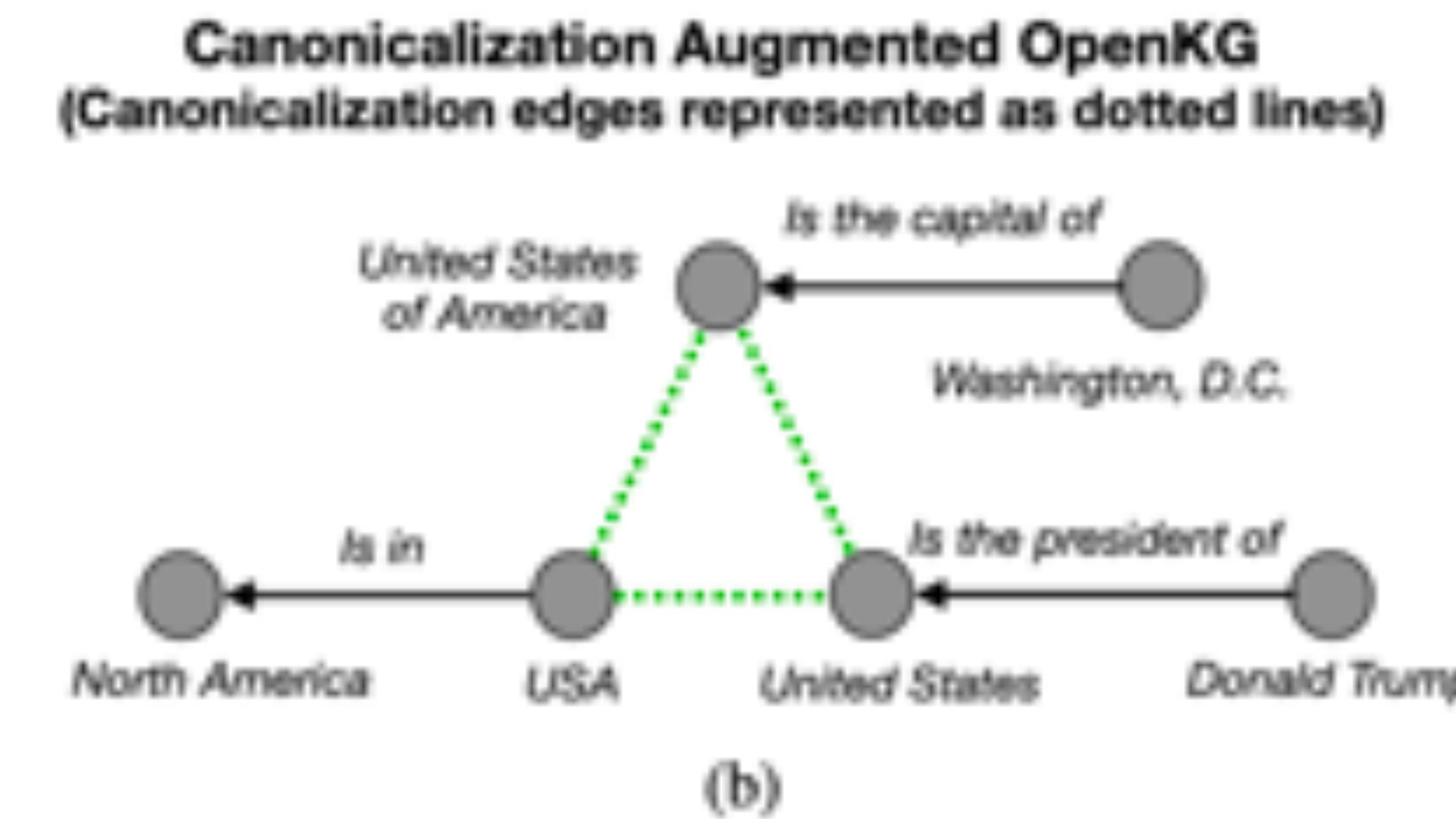
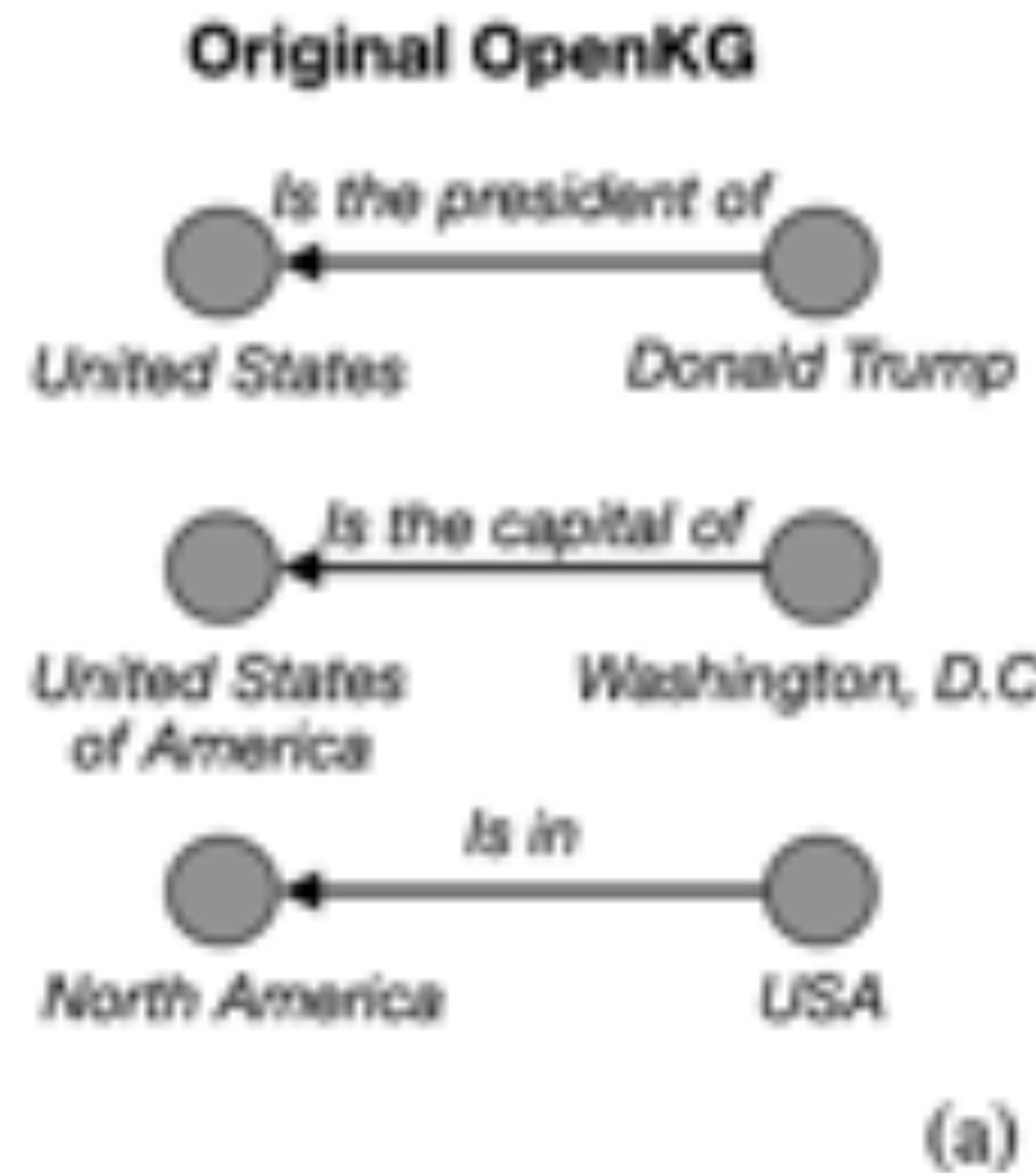
Task 2C: Automated KBC with CosmosIE and Marius

Improving COSMOS Search using Contextual Reranking

- Use deep learning to predict the most relevant search results using natural language understanding, improving over basic string-based retrieval.



Task 2C: Automated KBC with CosmosIE and Marius



[1] Gupta et. al: <https://www.aclweb.org/anthology/D19-1036/>

- Construct an Open Knowledge Base from COSMOS, then use Canonicalized-infused Representations [1] to canonicalize extracted entities
- Enhance COSMOS with question-answering, KB completion capabilities

Task 2D: Release public COSMOS API over COVID-19 set

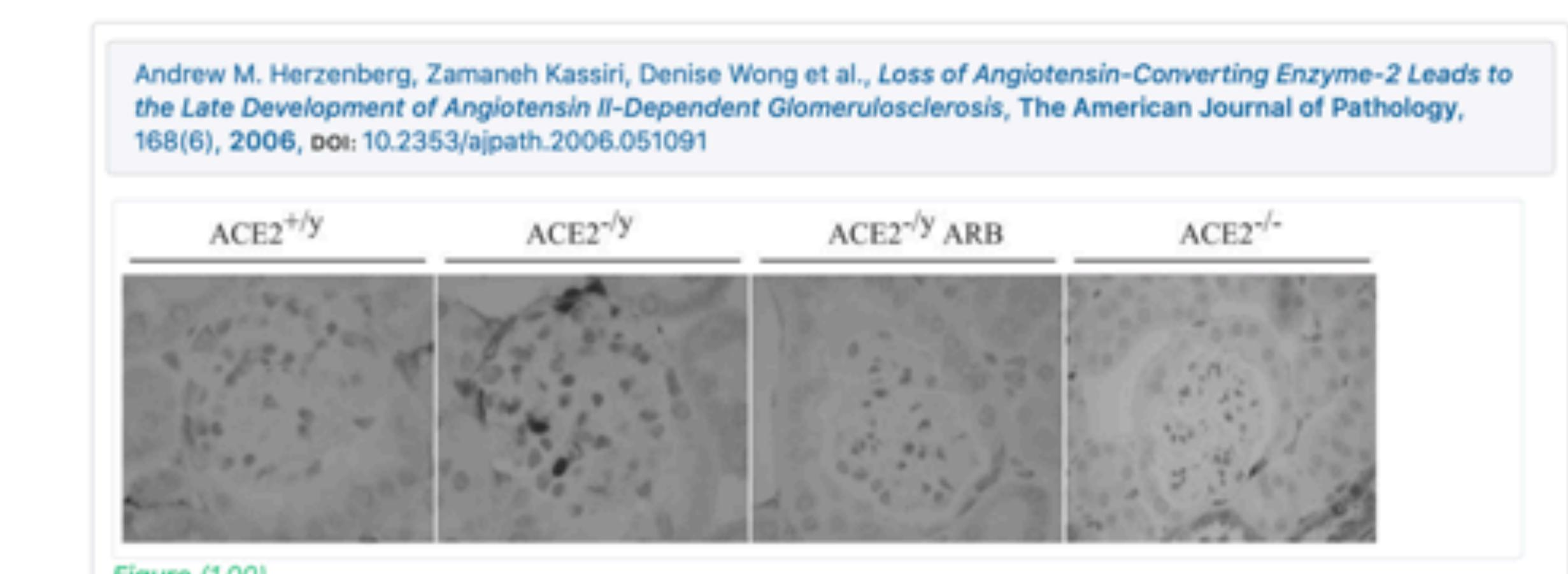
Allow third-party search/retrieval of table, figure, eqns

- Current interface sits on the COSMOS REST-ful API service
- Expand capabilities to allow team interfacing with COSMOS extractions, including:
 - Direct search of visual objects and contexts, multiple retrieval/rank methods
 - Stable URLs, with versioned access to past, current outputs:
 - table, figure, equations
 - custom trained word embedding with normalization variants (“raw” vs. lower cased, ligature-cleaned)



FIGURE 3 Analysis of relative protein levels of ER stress in the liver of ACE2^{+/-} and Ad-ACE2-treated db/db mice. A, Relative protein levels of GRP78, ATF4, and CHOP in the liver of ACE2^{+/-} mice. B, Relative ACE2 protein levels in the liver of Ad-ACE2-treated db/db mice. C, Immunofluorescence analysis of ACE2 in the liver of Ad-ACE2-treated db/db mice. Blue represents DAPI, red represents ACE2. D, Relative protein levels of GRP78, ATF4, and CHOP in the liver of Ad-ACE2-treated db/db mice. The data are presented as the mean ± SD of n = 3 in ACE2^{+/-} and Ad-ACE2-treated db/db mice. *P < 0.05 versus WT or Ad-GFP-treated by Student t test

Extracted data OCR text JSON object



Our vision: service-oriented knowledge extraction



Agent



Applications



Services



Infrastructure
and data

- COSMOS: An intelligent (AI-driven) virtual assistant that can consider and reason about semantics when linking/synthesizing artifacts from xDD-automated publication stream
- REST-ful APIs that enable users to build custom web and analytical applications over the scientific literature (simple to complex)
- Services that bring fine-grained knowledge extraction from publications within reach of every scientist – without requiring programming expertise (interactions based on natural language and point-and-click interfaces) or need to find and aggregate documents
- Compute infrastructure and document acquisition pipeline to support continual updating over diverse domains of science and engineering