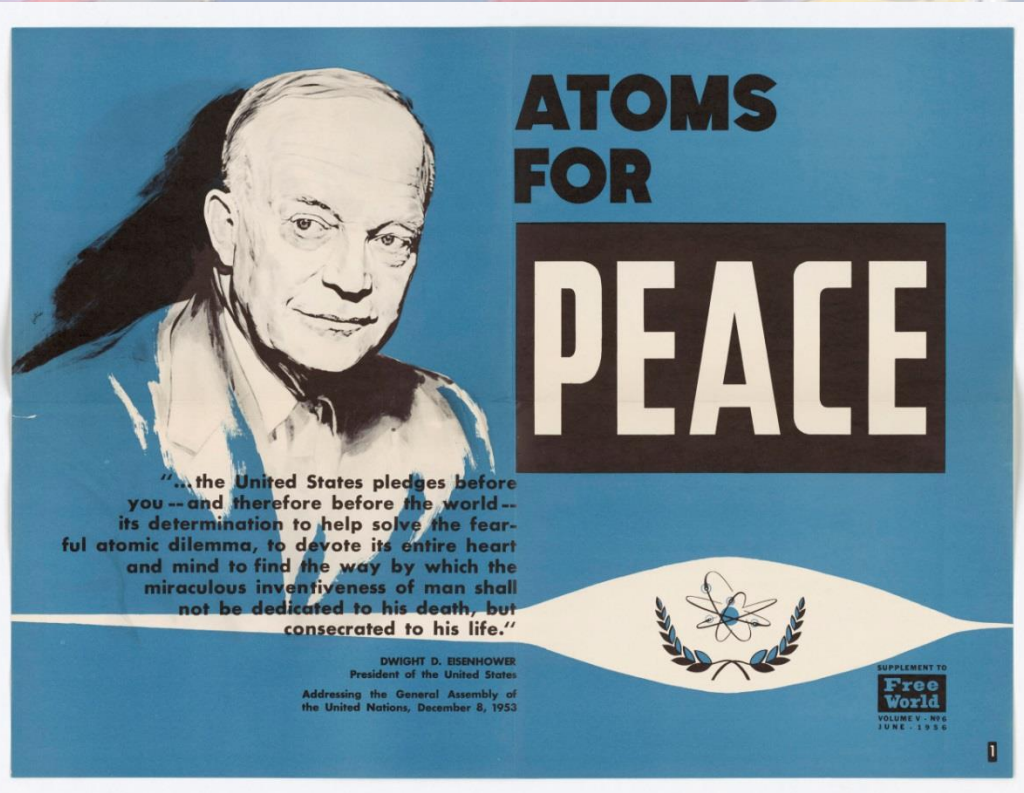
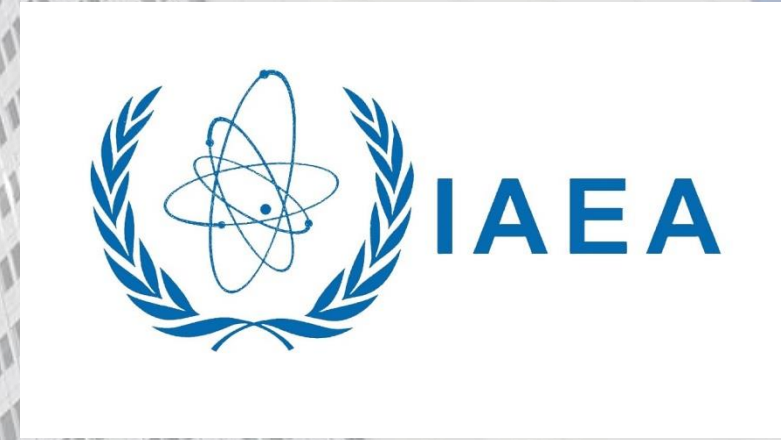


# AUTOMATED KOS-BASED SUBJECT INDEXING IN INIS

Dmitry Mironov

Nuclear Information Section

International Atomic Energy Agency (IAEA)



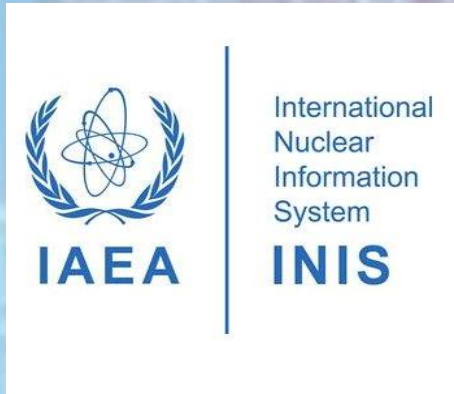
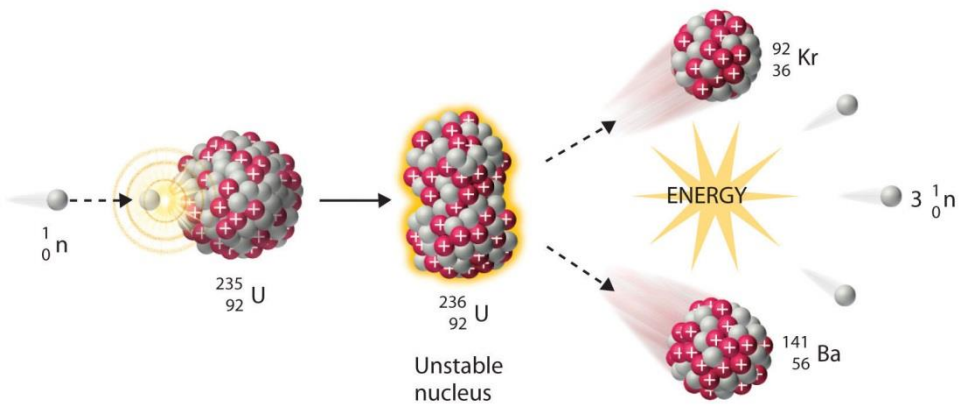
## ATOMS FOR PEACE AND DEVELOPMENT

How the IAEA supports the Sustainable Development Goals

- 13 CLIMATE ACTION
- 6 CLEAN WATER AND SANITATION
- 3 GOOD HEALTH AND WELL-BEING
- 9 INDUSTRY, INNOVATION AND INFRASTRUCTURE
- 17 PARTNERSHIPS FOR THE GOALS
- 2 ZERO HUNGER
- 15 LIFE ON LAND
- 7 AFFORDABLE AND CLEAN ENERGY
- 14 LIFE BELOW WATER



# INTERNATIONAL NUCLEAR INFORMATION SYSTEM (INIS)



## COMPARING CARBON PROFILES

Carbon dioxide emissions per megawatt-hour

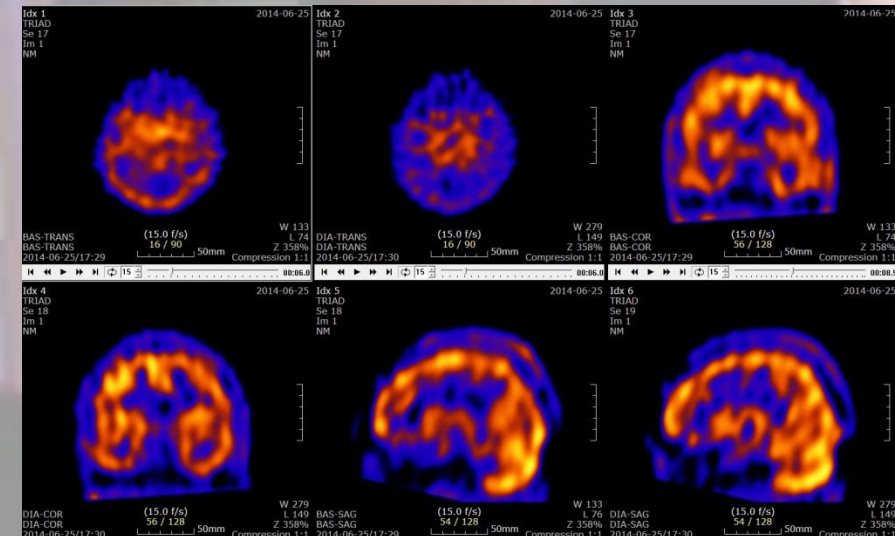
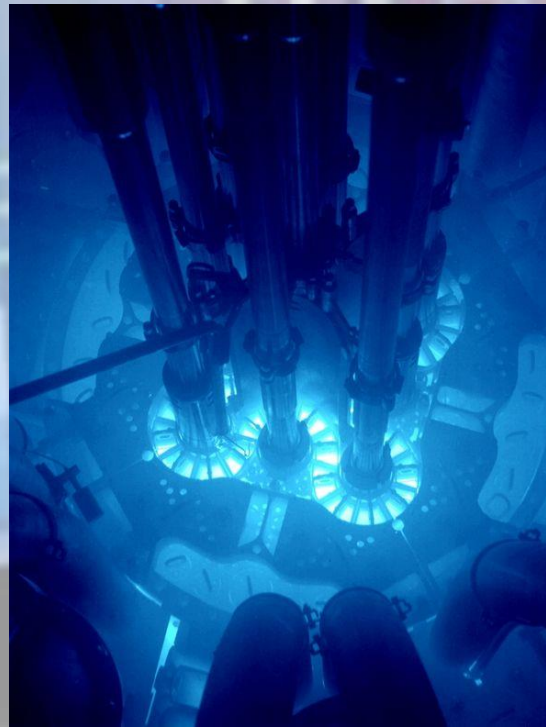
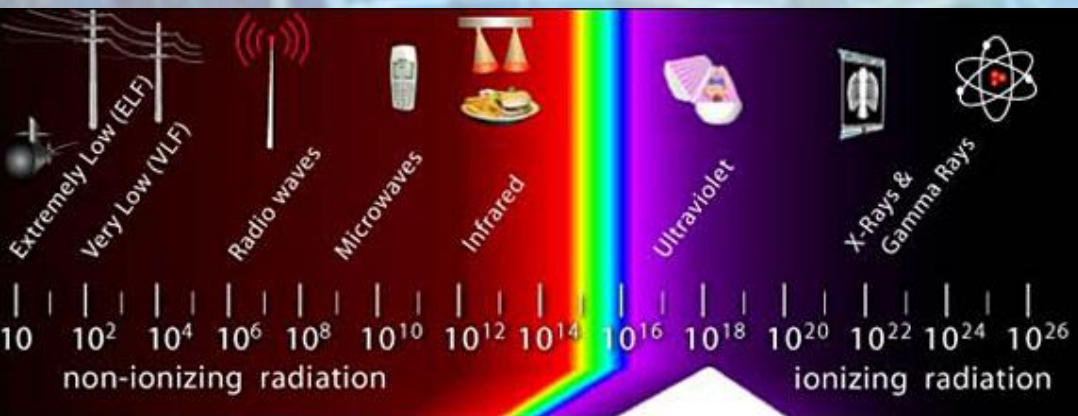
Average coal plant emits about



Average natural gas plant emits about



Nuclear plant emits



- **Introduction**
- **International Nuclear Information System**
- **INIS workflow**
- **Challenges**
- **Two-pass indexing**
- **Results and conclusions**
- **Outcomes and future work**



# INTERNATIONAL NUCLEAR INFORMATION SYSTEM (INIS)

**155** INIS members



**131** Member States

**24** international organizations

**155** INIS members



**131** Member States

**24** international organizations

**4** objectives



Acquire, process, preserve literature on peaceful uses of nuclear science and technology



Provide free and open access to nuclear information

Develop and maintain a nuclear knowledge organization system



Assist IAEA Member States in building their scientific information management capacities

## 155 INIS members



**131** Member States

**24** international organizations

## 4 objectives



Acquire, process, preserve literature on peaceful uses of nuclear science and technology



Provide free and open access to nuclear information

Develop and maintain a nuclear knowledge organization system



Assist IAEA Member States in building their scientific information management capacities

## INIS repository



**4.2 million** bibliographic records  
108 000 in 2018



**570 000** full-text documents  
19 500 in 2018

**3.2 million** web page views



**1.6 million** unique searches

**1 million** unique visitors



[inis.iaea.org](http://inis.iaea.org)

[inis.iaea.org/search](http://inis.iaea.org/search)



## INIS thesaurus

A knowledge organization system

8 languages

Arabic, Chinese, English, French,  
German, Japanese, Russian, Spanish

31 309 descriptors

150 in 2018





## INIS thesaurus

A knowledge organization system

8 languages

Arabic, Chinese, English, French,  
German, Japanese, Russian, Spanish

31 309 descriptors

150 in 2018



International Nuclear Information System  
(INIS)

JOINT THESAURUS

ETDE/INIS Joint Reference Series No. 1 (Rev. 2.2)

Nuclear reactors and reactor safety • Energy efficiency • Nuclear instrumentation  
Power transmission and distribution • Materials and physical sciences research

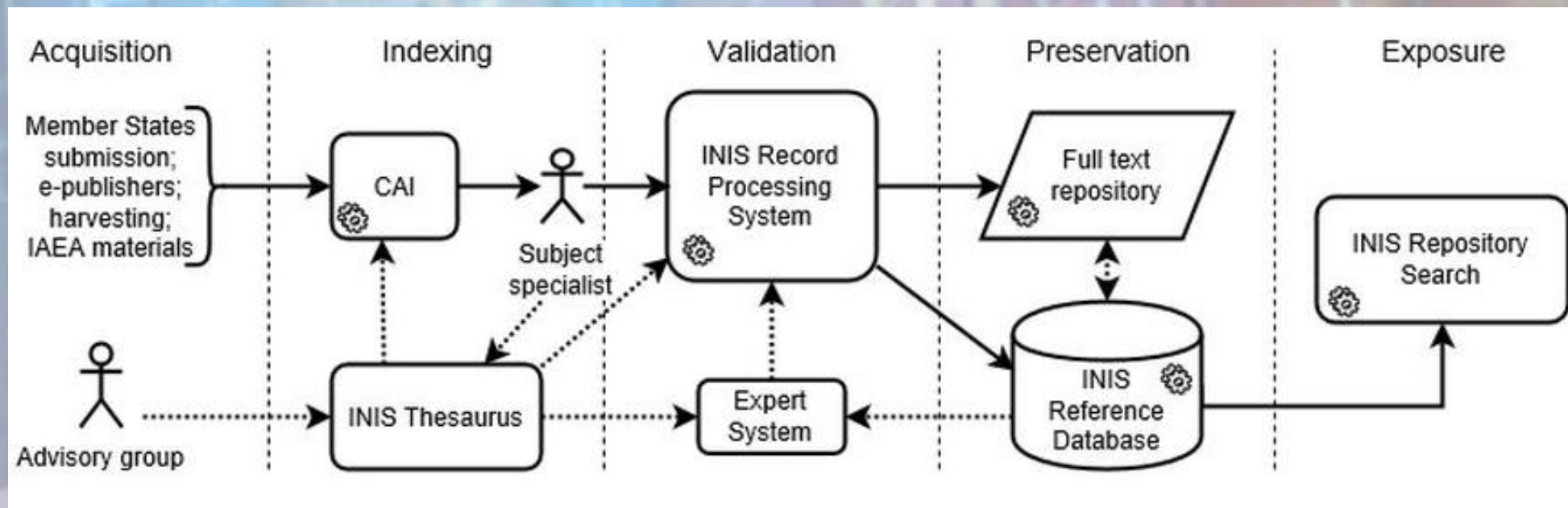


Renewable energy technologies • Radiation protection • Energy storage, conversion, and consumption  
Radioactive waste management • Energy policy • Radiation effects on living organisms • Fossil fuels



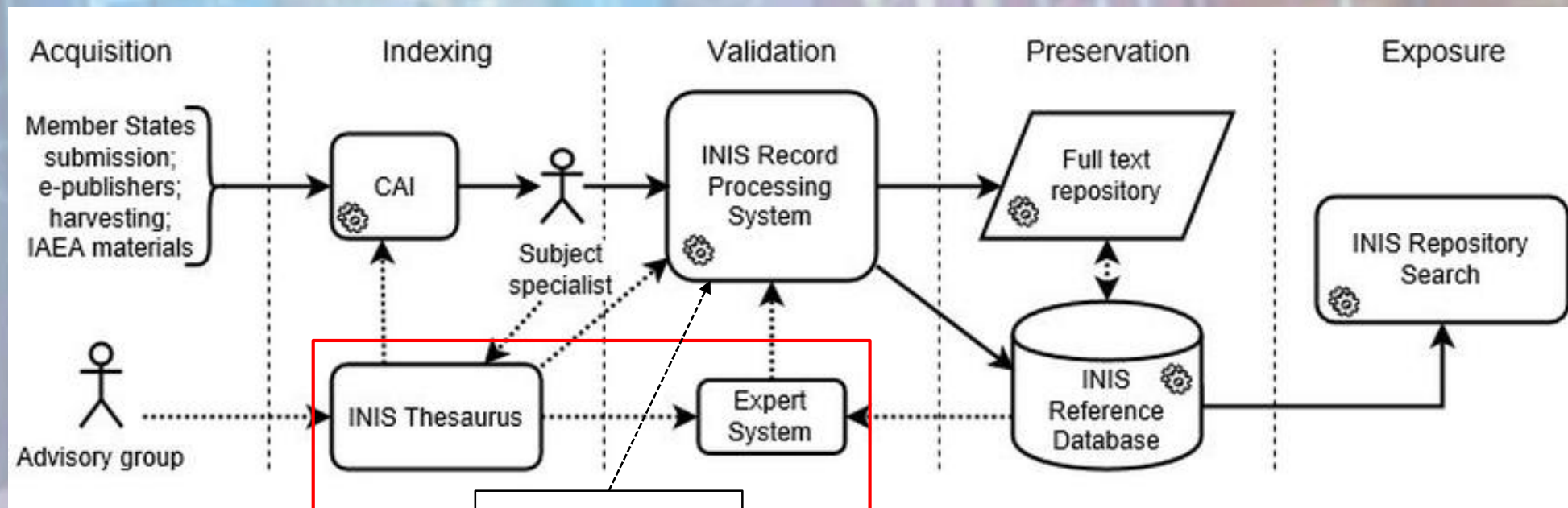
Vienna, April 2009

# INIS WORKFLOW





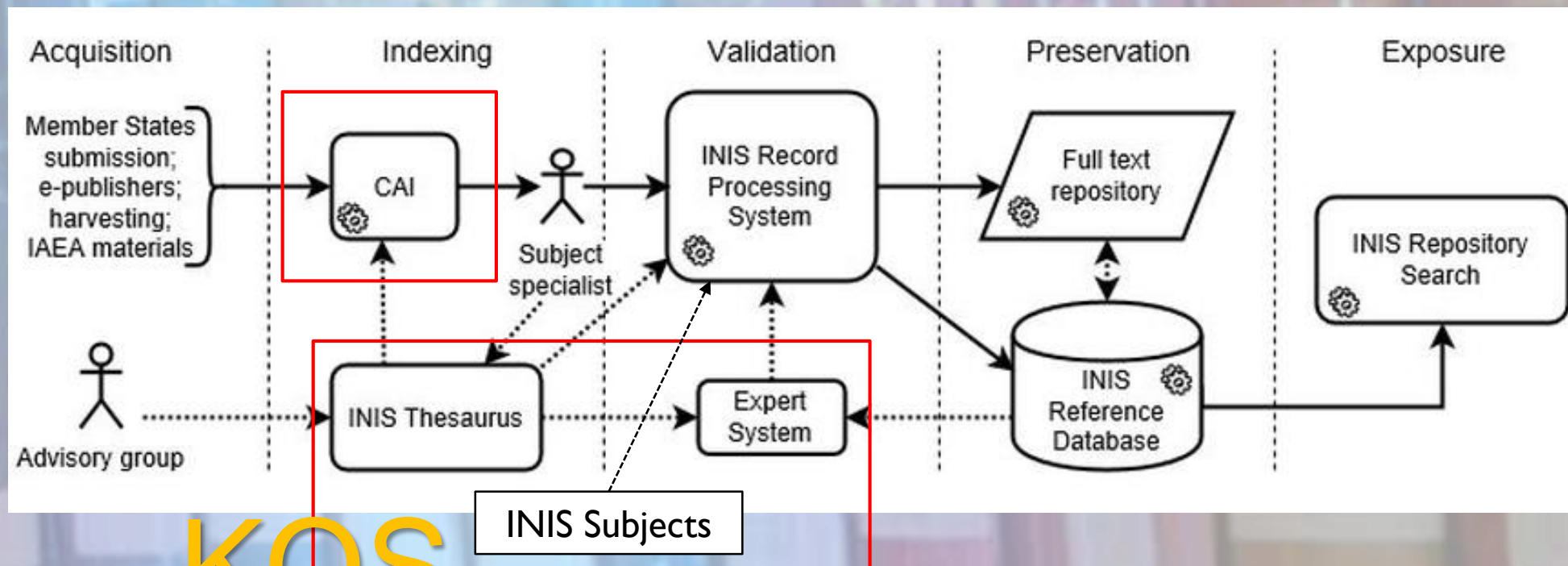
# INIS WORKFLOW



KOS

INIS Subjects

# INIS WORKFLOW

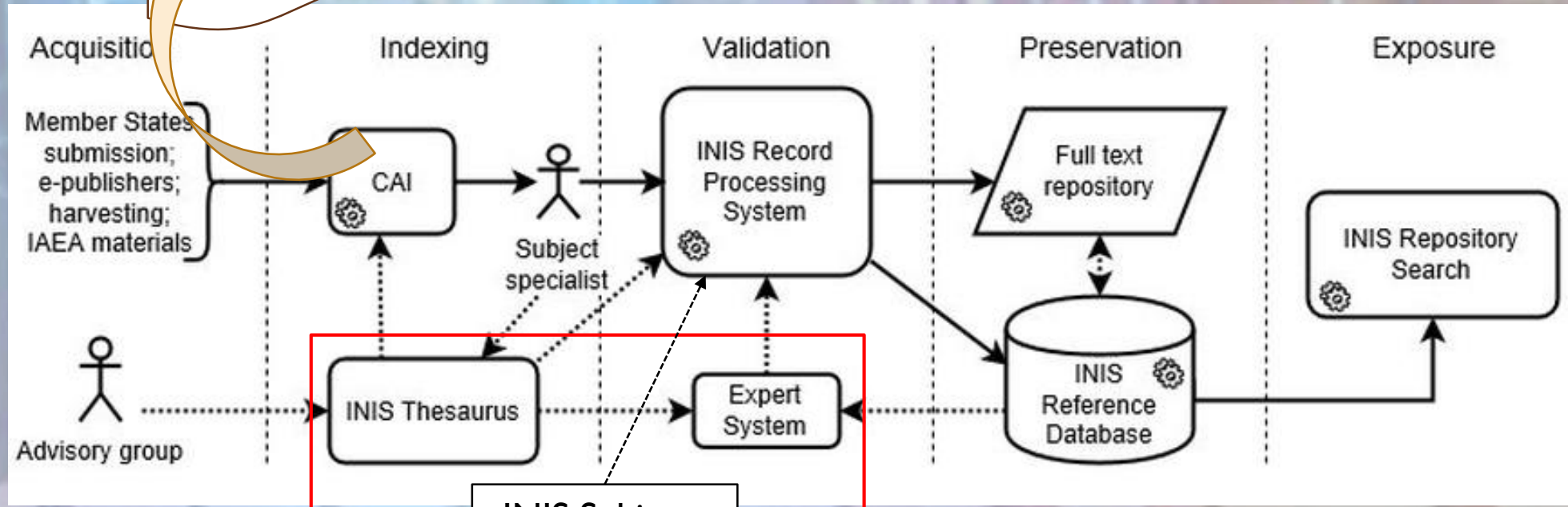


KOS



# INIS WORKFLOW

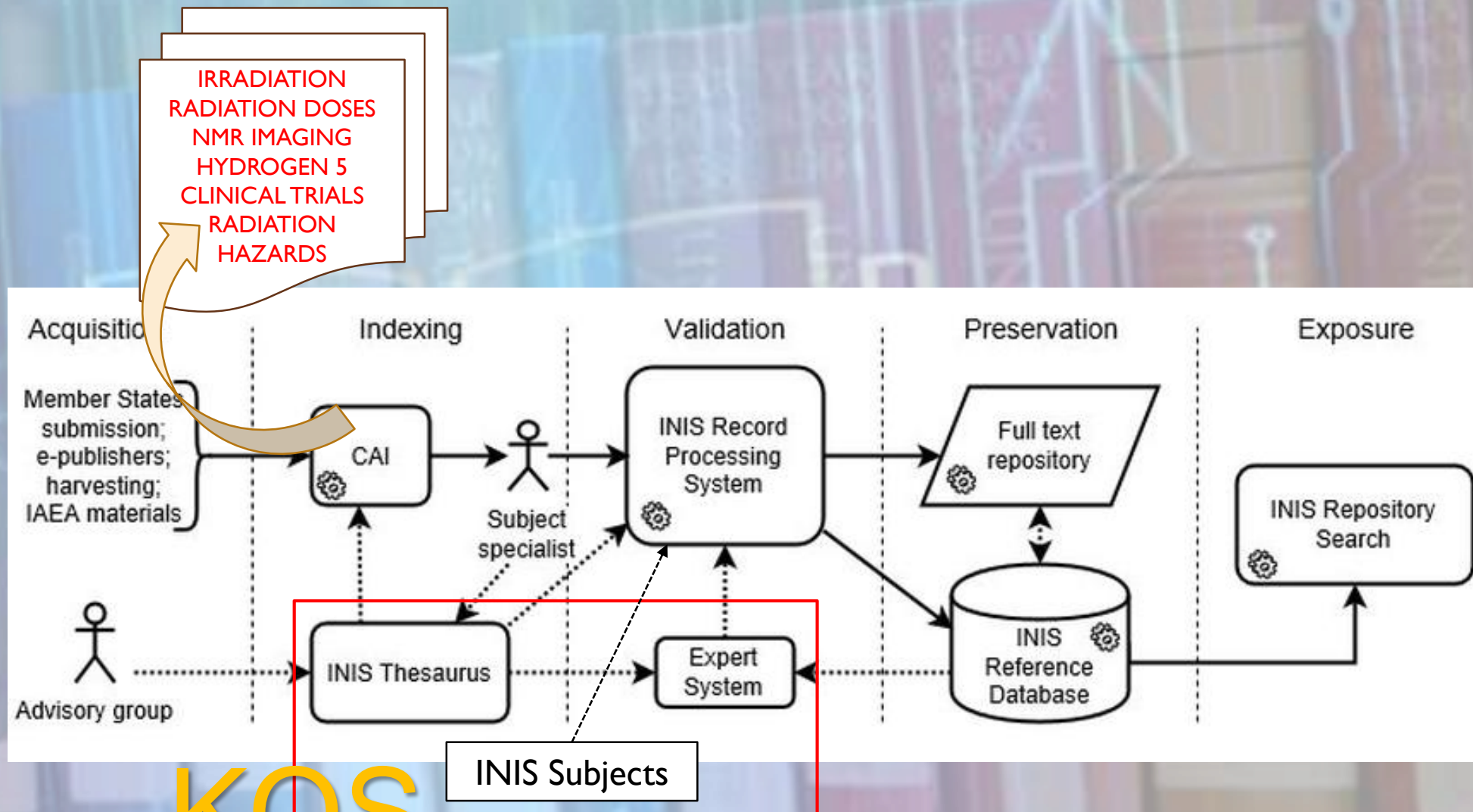
mentioned domain inde  
fulfil. A substantial num  
r a class of descriptors dep  
nt of new rules can become  
ck. To mitigate this, we sh  
ombination with a purely n  
to retrain the algorithm, co  
tion we foresee is impleme  
se in deep learning, speci  
wed their efficiency i



# KOS

INIS Subjects

# INIS WORKFLOW

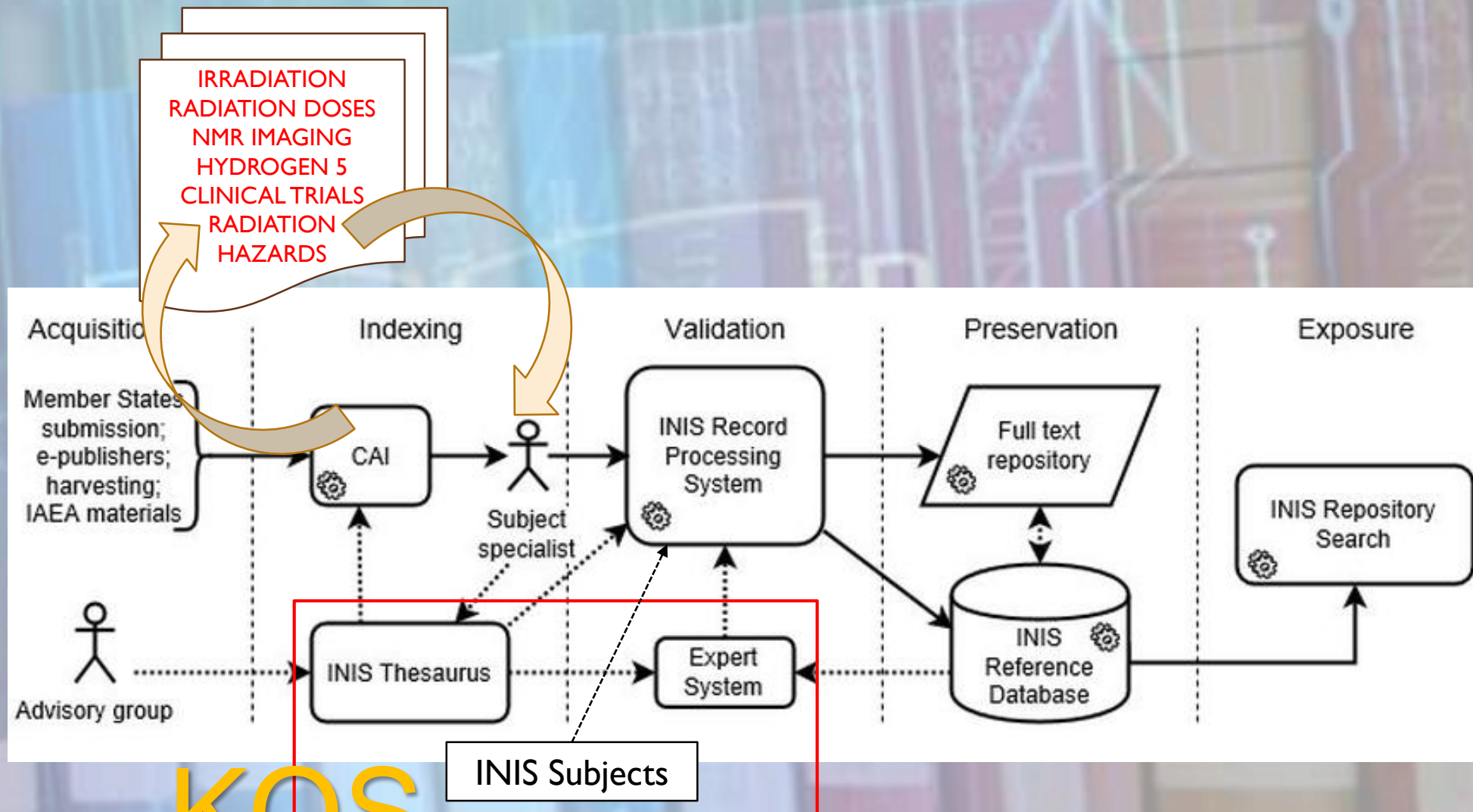


KOS

INIS Subjects



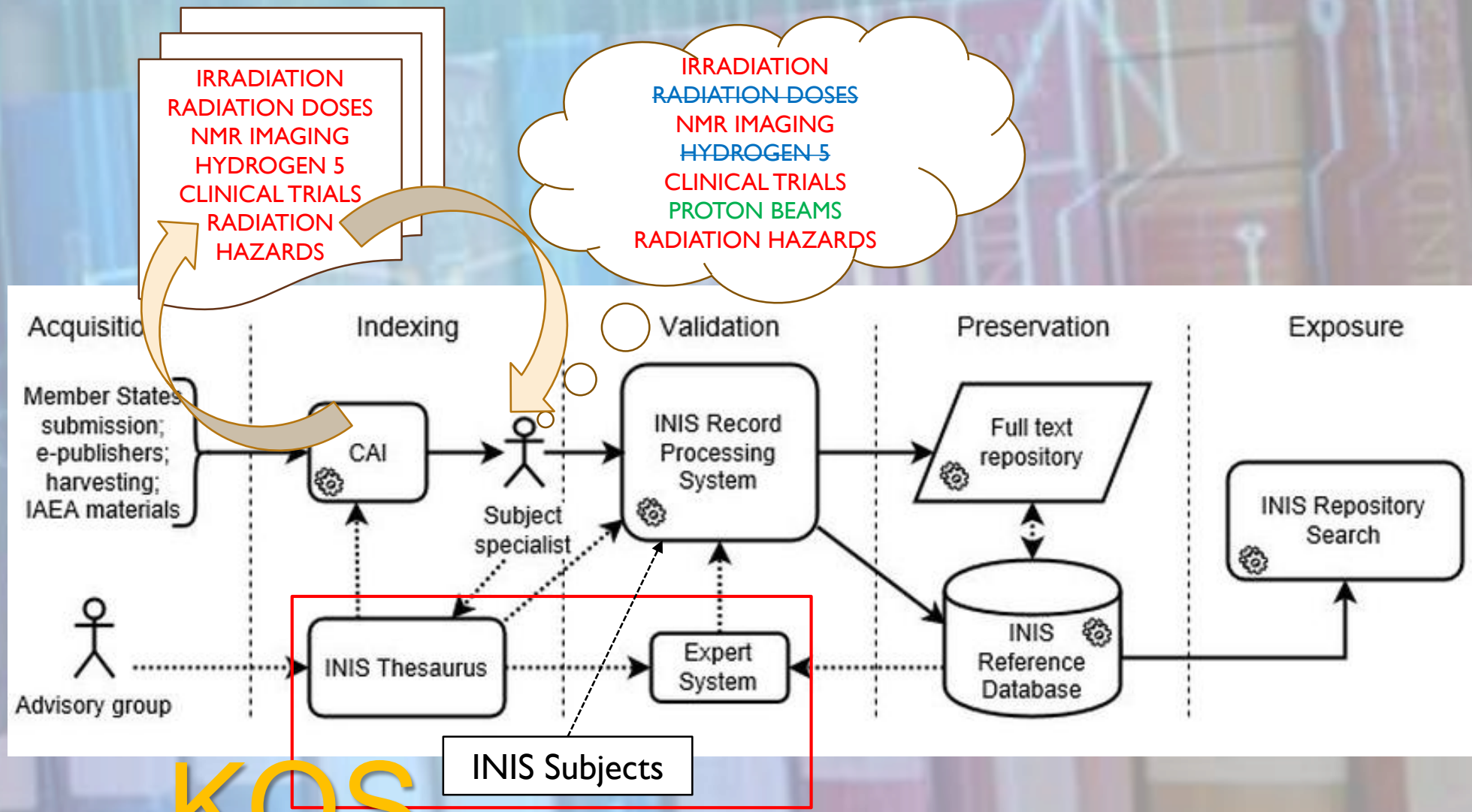
# INIS WORKFLOW



KOS

INIS Subjects

# INIS WORKFLOW



KOS

## TRIGA research reactors: A pathway to the peaceful applications of nuclear energy

BY DONALD M. FORTNEY,  
ERNEST RAYNE, AND  
WILLIAM L. WERTSCHER

*The first three TRIGA reactors were placed in operation just two years after the idea for such a reactor was conceived.*

Less than two years after the idea of President Dwight D. Eisenhower, in December 1953, "Atoms for Peace" proposed to the United Nations General Assembly, TRIGA is a new kind of inherently safe, training, research, and development reactor. It is a small, self-contained, fuel, and operating in the General Assembly Conference of Disarmament Commission in the Diego Hall, New York City, TRIGA (Training, Research, Development, and Education) has evolved into the most widely used research reactor in the world, with operating power levels up to 250 kilowatts (kW) and with an installed base of 33 reactors in 14 countries in five continents.

TRIGA is a small, self-contained, fuel, and operating in the General Assembly Conference of Disarmament Commission in the Diego Hall, New York City, TRIGA (Training, Research, Development, and Education) has evolved into the most widely used research reactor in the world, with operating power levels up to 250 kilowatts (kW) and with an installed base of 33 reactors in 14 countries in five continents.

TRIGA (Training, Research, Development, and Education) has evolved into the most widely used research reactor in the world, with operating power levels up to 250 kilowatts (kW) and with an installed base of 33 reactors in 14 countries in five continents.

TRIGA (Training, Research, Development, and Education) has evolved into the most widely used research reactor in the world, with operating power levels up to 250 kilowatts (kW) and with an installed base of 33 reactors in 14 countries in five continents.

TRIGA (Training, Research, Development, and Education) has evolved into the most widely used research reactor in the world, with operating power levels up to 250 kilowatts (kW) and with an installed base of 33 reactors in 14 countries in five continents.

TRIGA (Training, Research, Development, and Education) has evolved into the most widely used research reactor in the world, with operating power levels up to 250 kilowatts (kW) and with an installed base of 33 reactors in 14 countries in five continents.



Fig. 1. World map showing TRIGA reactor locations.

48

RESEARCH NEWS

November 2007





# CHALLENGES

TRIGA TYPE REACTORS  
REACTORS  
NEUTRONS  
NUCLEAR FUELS  
MAGNESIUM 20  
REACTOR DESIGN  
RESEARCH REACTORS  
LEAD  
FISSION PRODUCTS  
NEUTRON BEAMS  
RADIOGRAPHY  
MELANOMAS

TRIGA TYPE REACTORS  
REACTORS  
NEUTRONS  
NUCLEAR FUELS  
MAGNESIUM 20  
REACTOR DESIGN  
RESEARCH REACTORS  
LEAD  
FISSION PRODUCTS  
NEUTRON BEAMS  
RADIOGRAPHY  
MELANOMAS

**Resulting subject analysis often contains**



TRIGA TYPE REACTORS

REACTORS ←

NEUTRONS

NUCLEAR FUELS

MAGNESIUM 20

REACTOR DESIGN

RESEARCH REACTORS

LEAD

FISSION PRODUCTS

NEUTRON BEAMS

RADIOGRAPHY

MELANOMAS

## Resulting subject analysis often contains



Too broad descriptors



TRIGA TYPE REACTORS

REACTORS

NEUTRONS

NUCLEAR FUELS

MAGNESIUM 20

REACTOR DESIGN

RESEARCH REACTORS

LEAD

FISSION PRODUCTS

NEUTRON BEAMS

RADIOGRAPHY

MELANOMAS



“...leads to a melting”

## Resulting subject analysis often contains



Too broad descriptors

Misleading suggestions

TRIGA TYPE REACTORS

REACTORS

NEUTRONS

NUCLEAR FUELS

MAGNESIUM 20



20 mg

REACTOR DESIGN

RESEARCH REACTORS

LEAD

FISSION PRODUCTS

NEUTRON BEAMS

RADIOGRAPHY

MELANOMAS

## Resulting subject analysis often contains



Too broad descriptors

Misleading suggestions

Incorrect interpretations

TRIGA TYPE REACTORS  
REACTORS  
NEUTRONS  
NUCLEAR FUELS  
MAGNESIUM 20  
REACTOR DESIGN  
RESEARCH REACTORS  
LEAD  
FISSION PRODUCTS  
NEUTRON BEAMS  
RADIOGRAPHY  
MELANOMAS

## Resulting subject analysis often contains

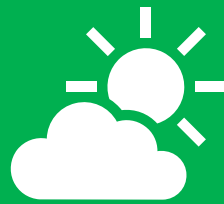


Too broad descriptors

Misleading suggestions

Incorrect interpretations

## 50 years of INIS



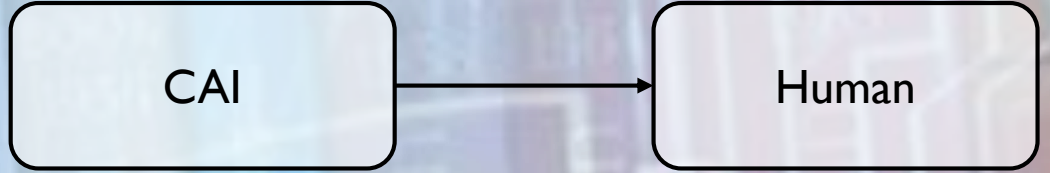
4 million indexed records

Experienced specialists

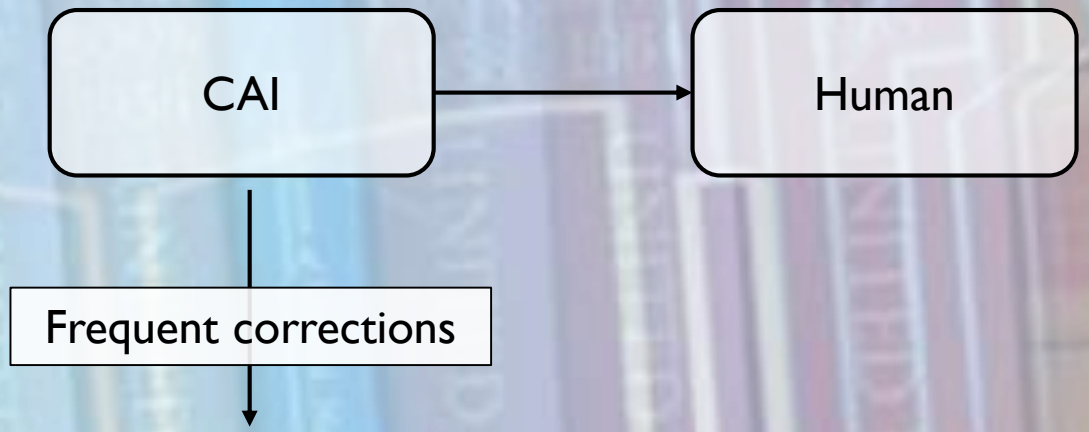
Unique KOS and workflow



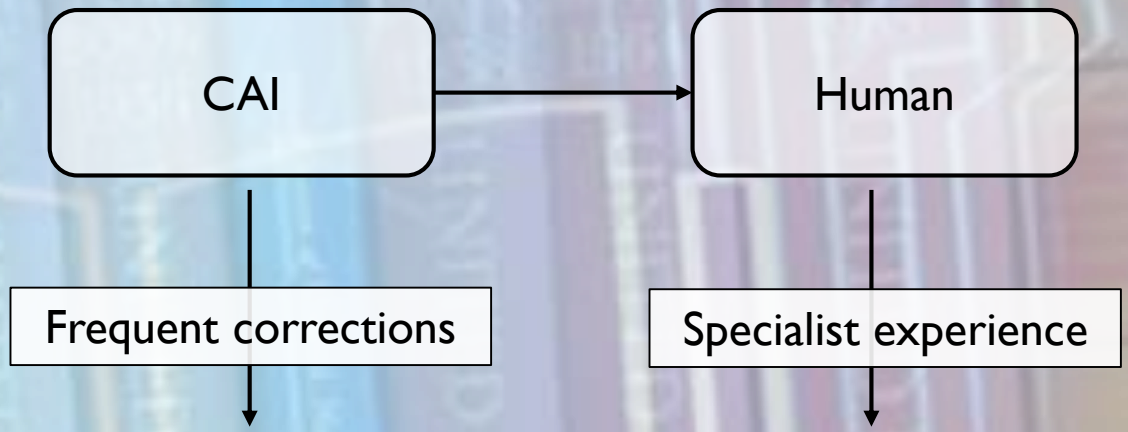
# TWO-PASS INDEXING, TIER I



# TWO-PASS INDEXING, TIER I

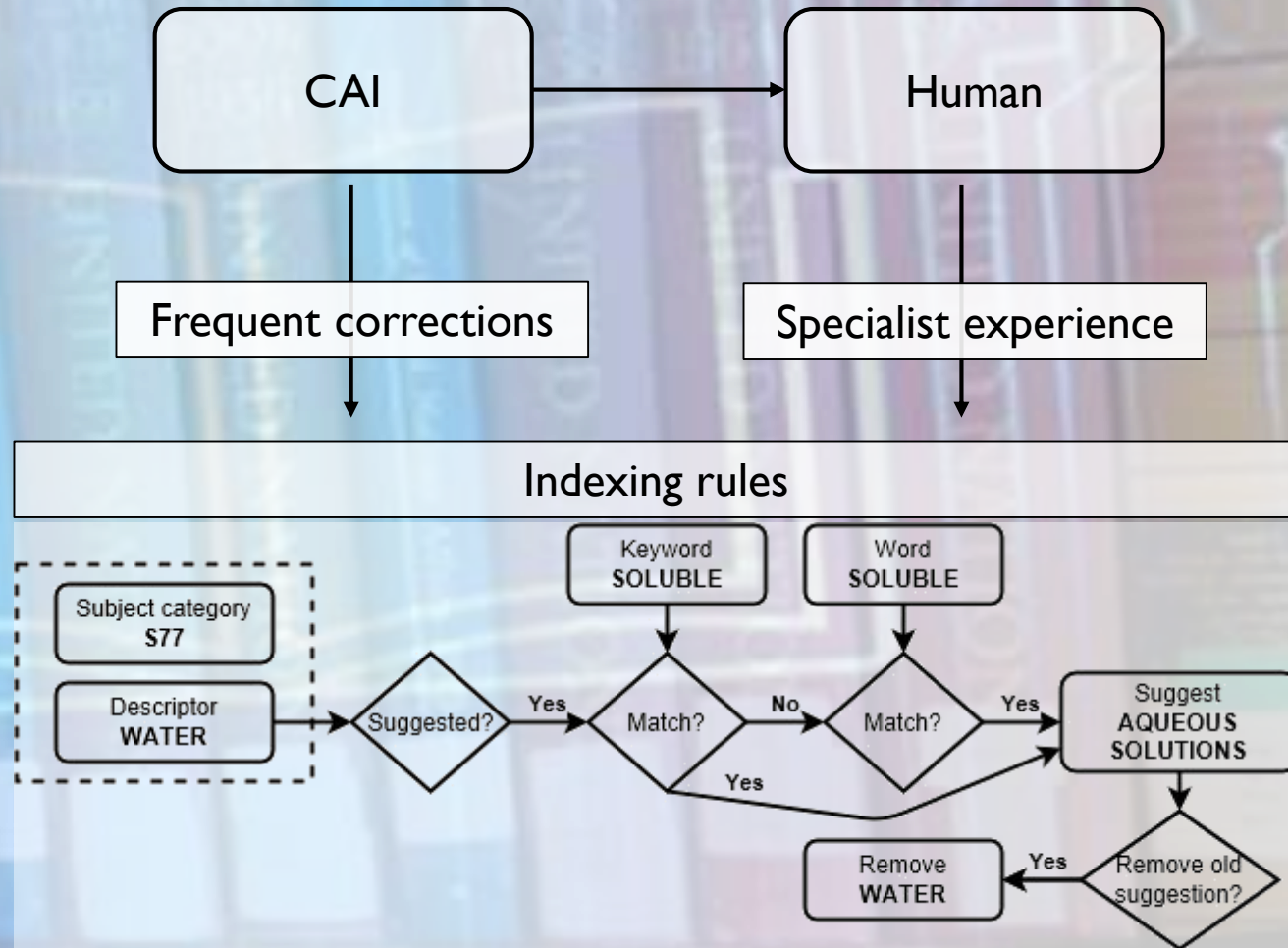


# TWO-PASS INDEXING, TIER I

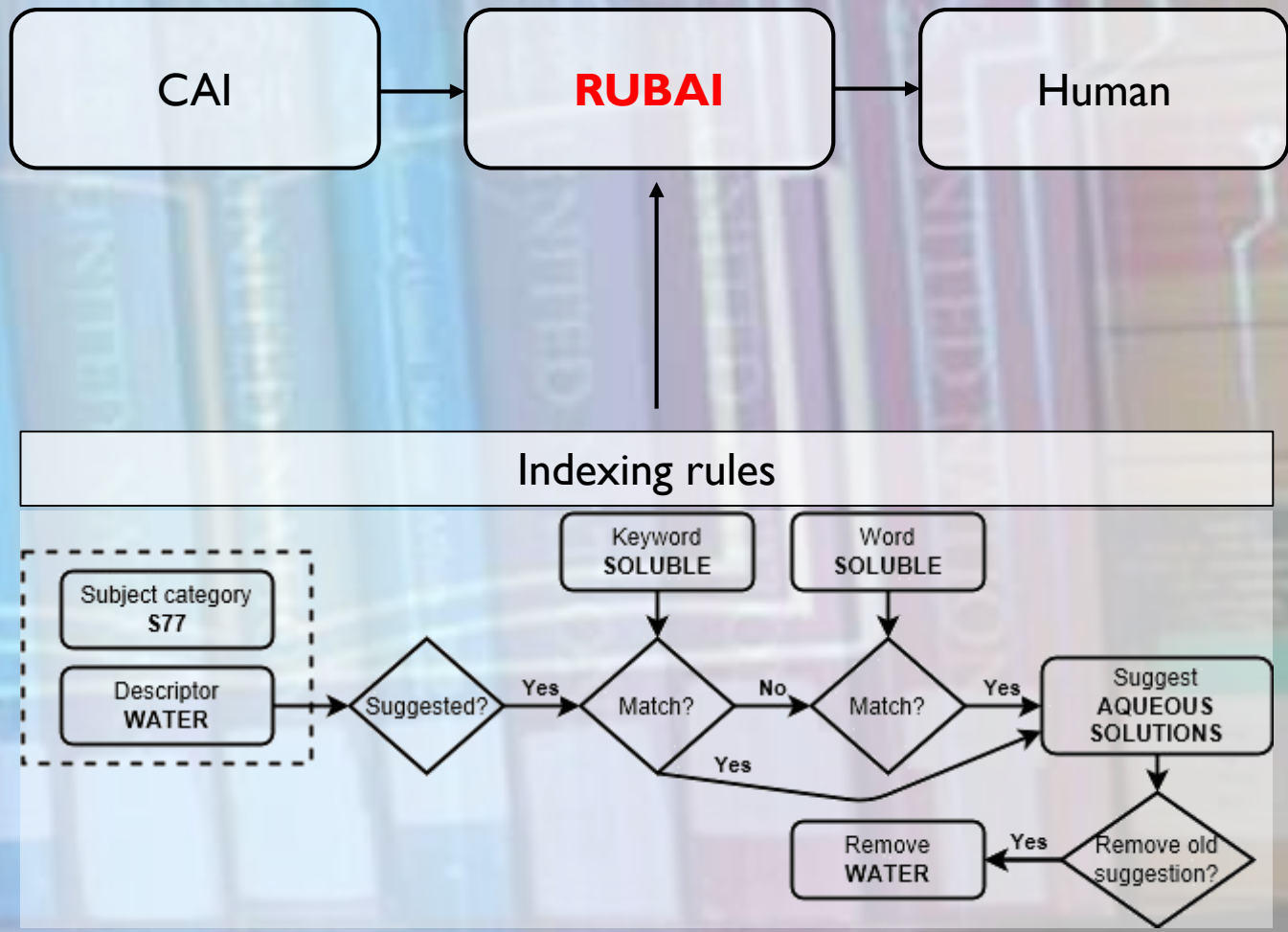




# TWO-PASS INDEXING, TIER I



# TWO-PASS INDEXING, TIER I



# TWO-PASS INDEXING, TIER I



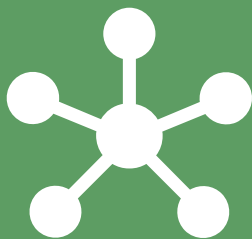
**940** unique rules

Add

Replace

Remove

Keep





# TWO-PASS INDEXING, TIER I



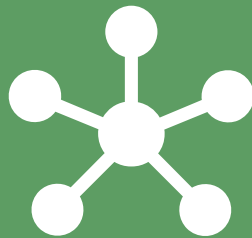
**940** unique rules

Add

Replace

Remove

Keep



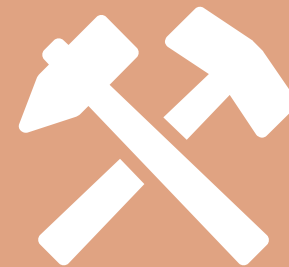
**KOS** integration

Thesaurus

INIS subjects

Expert knowledge

Quality control



# RUBAI: RESULTS

ID	Number of records	Descriptors							Records where more than one descriptor were				Operations per record performed by human after...	
		Total after...			added by human after...		removed by human after...		added after...		removed after...			
		CAI	RUBAI	Human	CAI	RUBAI	CAI	RUBAI	CAI	RUBAI	CAI	RUBAI	CAI	RUBAI
1	94	2636	1154	1032	144	3	1748	125	42	0	94	36	<i>20.13</i>	<i>1.36</i>
2	94	2570	1177	984	86	4	1672	197	24	1	94	67	<i>18.7</i>	<i>2.14</i>
3	97	2473	875	845	121	70	1749	100	37	18	97	32	<i>19.27</i>	<i>1.75</i>
4	95	2361	900	900	92	185	1553	185	19	46	95	49	<i>17.31</i>	<i>3.89</i>
5	96	2398	1154	916	118	102	1600	340	27	28	96	76	<i>17.9</i>	<i>4.6</i>
6	91	2166	1050	929	114	83	1351	204	26	24	91	52	<i>16.1</i>	<i>3.15</i>
7	97	3030	1111	971	114	17	2173	157	31	4	97	39	<i>23.58</i>	<i>1.79</i>
8	62	2019	745	658	69	8	1430	95	15	1	62	23	<i>24.18</i>	<i>1.66</i>
9	93	1799	782	716	99	50	1182	116	22	10	92	35	<i>20.41</i>	<i>2.31</i>
10	98	1912	802	760	121	70	1273	112	36	14	97	30	<i>14.22</i>	<i>1.86</i>
11	77	1712	753	788	171	105	1095	70	53	30	77	21	<i>16.44</i>	<i>2.27</i>

# RUBAI: RESULTS

ID	Number of records	Descriptors						Records where more than one descriptor were				Operations per record performed by human after...		
		Total after...			added by human		removed by human	added after...		removed after...				
<b>Average</b>														
<b>Descriptors</b>										<b>Operations per record performed by human after...</b>		<b>RUBAI</b>		
1												<i>1.36</i>		
2										<i>2.14</i>				
3		Total after...			added by human after...		removed by human after...				<i>1.75</i>			
4											<i>3.89</i>			
5		<b>CAI</b>	<b>RUBAI</b>	<b>Human</b>	<b>CAI</b>	<b>RUBAI</b>	<b>CAI</b>	<b>RUBAI</b>	<b>CAI</b>	<b>RUBAI</b>	<i>4.6</i>			
6		<b>2280</b>	<b>955</b>	<b>864</b>	<b>114</b>	<b>63</b>	<b>1530</b>	<b>155</b>	<b>18.93</b>	<b>2.43</b>	<i>3.15</i>			
7	97	3030	1111	971	114	17	2173	157	31	4	97	39	<b>23.58</b>	<i>1.79</i>
8	62	2019	745	658	69	8	1430	95	15	1	62	23	<b>24.18</b>	<i>1.66</i>
9	93	1799	782	716	99	50	1182	116	22	10	92	35	<b>20.41</b>	<i>2.31</i>
10	98	1912	802	760	121	70	1273	112	36	14	97	30	<b>14.22</b>	<i>1.86</i>
11	77	1712	753	788	171	105	1095	70	53	30	77	21	<b>16.44</b>	<i>2.27</i>



# RUBAI: RESULTS

Operations per record needed by human after



# ARE WE DOING WELL?

Domestic Corporation in San Diego, Calif. Over the years, TRIGA (Training, Research, Development, General, Atomic) has evolved into the more widely used acronym TRIGA, which operated as the General Atomic Corporation for the Peace Act of 1958 and Atomic Energy International in 1972.

Domestic Corporation in San Diego, Calif. Over the years, TRIGA (Training, Research, Development, General, Atomic) has evolved into the more widely used acronym TRIGA, which operated as the General Atomic Corporation for the Peace Act of 1958 and Atomic Energy International in 1972.

TRIGA TYPE REACTORS

REACTORS

NEUTRONS

NUCLEAR FUELS

MAGNESIUM 20

REACTOR DESIGN

RESEARCH REACTORS

LEAD

FISSION PRODUCTS

NEUTRON BEAMS

# ARE WE DOING WELL?

Domestic Corporation in San Diego, Calif.  
Over the years, TRIGA (Training, Re-  
search, Development, General, Acce-  
ss) has evolved into the most widely used

reactors with the prototype TRIGA reactor  
at General Atomics (GA) in San Diego. GA  
is a TRIGA, which operated as the  
General Atomics Corporation for the Peace  
Act 1954 and is now owned by General Atomics.

As the result of an agreement TRIGA  
GA is OGE's core research reactor in San Diego.  
There are also the Equilibrium of Veterans as  
research reactor of an agreement TRIGA  
GA is a research reactor.

Domestic Corporation in San Diego, Calif.  
Over the years, TRIGA (Training, Re-  
search, Development, General, Acce-  
ss) has evolved into the most widely used

reactors with the prototype TRIGA reactor  
at General Atomics (GA) in San Diego. GA  
is a TRIGA, which operated as the  
General Atomics Corporation for the Peace  
Act 1954 and is now owned by General Atomics.

As the result of an agreement TRIGA  
GA is OGE's core research reactor in San Diego.  
There are also the Equilibrium of Veterans as  
research reactor of an agreement TRIGA  
GA is a research reactor.

TRIGA TYPE REACTORS

NUCLEAR FUELS

REACTOR DESIGN  
RESEARCH REACTORS

FISSION PRODUCTS  
NEUTRON BEAMS



# ARE WE DOING WELL?

Domestic Corporation in San Diego, Calif. Over the years, TRIGA (Training, Research, Development, General, Atomic) has evolved into the more widely used acronym TRIGA, which operated as the General Atomic Corporation for the Peace Act of 1954 and Atomic Energy Administration. As the result of an agreement TGA, the AEC's commercial reactor is known as the Department of Energy's National Institute of Standards and Technology (NIST).

Domestic Corporation in San Diego, Calif. Over the years, TRIGA (Training, Research, Development, General, Atomic) has evolved into the more widely used acronym TRIGA, which operated as the General Atomic Corporation for the Peace Act of 1954 and Atomic Energy Administration. As the result of an agreement TGA, the AEC's commercial reactor is known as the Department of Energy's National Institute of Standards and Technology (NIST).

TRIGA TYPE REACTORS

NUCLEAR FUELS

REACTOR DESIGN  
RESEARCH REACTORS

FISSION PRODUCTS  
NEUTRON BEAMS

RADIOGRAPHY  
MELANOMAS

# ARE WE DOING WELL?

Domestic Dependence in San Diego, Calif. Over the years, TRIGA (Training, Research, Development, General, Academic) has evolved into the most widely used reactor system in the world. It is the foundation of an advanced TRIGA, and ORNL's commitment to research and development of advanced TRIGA reactors is a key part of the program's mission.

Domestic Dependence in San Diego, Calif. Over the years, TRIGA (Training, Research, Development, General, Academic) has evolved into the most widely used reactor system in the world. It is the foundation of an advanced TRIGA, and ORNL's commitment to research and development of advanced TRIGA reactors is a key part of the program's mission.

Domestic Dependence in San Diego, Calif. Over the years, TRIGA (Training, Research, Development, General, Academic) has evolved into the most widely used reactor system in the world. It is the foundation of an advanced TRIGA, and ORNL's commitment to research and development of advanced TRIGA reactors is a key part of the program's mission.

Domestic Dependence in San Diego, Calif. Over the years, TRIGA (Training, Research, Development, General, Academic) has evolved into the most widely used reactor system in the world. It is the foundation of an advanced TRIGA, and ORNL's commitment to research and development of advanced TRIGA reactors is a key part of the program's mission.

TRIGA TYPE REACTORS

NUCLEAR FUELS

REACTOR DESIGN  
RESEARCH REACTORS

FISSION PRODUCTS  
NEUTRON BEAMS  
RADIOGRAPHY  
MELANOMAS



TRIGA TYPE REACTORS

NUCLEAR FUELS

REACTOR DESIGN

RESEARCH REACTORS

FISSION PRODUCTS

NEUTRON BEAMS

INDUSTRIAL RADIOGRAPHY

MELANOMAS

NEUTRON CAPTURE THERAPY

ISOTOPE PRODUCTION

TRIGA-2-MUSASHI REACTOR

# FROM A SINGLE DECISION TO A DATASET

<b>Descriptor</b>	<b>RUBAI Operation</b>	<b>Human Action</b>	<b>Human Decision</b>
TRIGA TYPE REACTORS	Keep	Keep	Approve
REACTORS	Remove		Approve
RESEARCH REACTORS	Keep	Keep	Approve
LEAD	Remove		Approve
RADIOGRAPHY	Add	Remove	Reject
MELANOMAS	Add	Remove	Reject
NEUTRON CAPTURE THERAPY		Add	
ISOTOPE PRODUCTION		Add	
INDUSTRIAL RADIOGRAPHY		Add	



# FROM A SINGLE DECISION TO A DATASET

<b>Descriptor</b>	<b>RUBAI Operation</b>	<b>Human Action</b>	<b>Human Decision</b>
TRIGA TYPE REACTORS	Keep	Keep	Approve
REACTORS	Remove		Approve
RESEARCH REACTORS	Keep	Keep	Approve
LEAD	Remove		Approve
RADIOGRAPHY	Add	Remove	Reject
MELANOMAS	Add	Remove	Reject
NEUTRON CAPTURE THERAPY	Not Add	Add	
ISOTOPE PRODUCTION	Not Add	Add	
INDUSTRIAL RADIOGRAPHY	Not Add	Add	

# FROM A SINGLE DECISION TO A DATASET

<b>Descriptor</b>	<b>RUBAI Operation</b>	<b>Human Action</b>	<b>Human Decision</b>
TRIGA TYPE REACTORS	Keep	Keep	Approve
REACTORS	Remove		Approve
RESEARCH REACTORS	Keep	Keep	Approve
LEAD	Remove		Approve
RADIOGRAPHY	Add	Remove	Reject
MELANOMAS	Add	Remove	Reject
NEUTRON CAPTURE THERAPY	Not Add	Add	Reject
ISOTOPE PRODUCTION	Not Add	Add	Reject
INDUSTRIAL RADIOGRAPHY	Not Add	Add	Reject

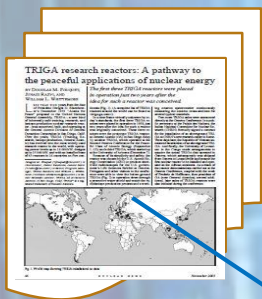
# FROM A SINGLE DECISION TO A DATASET

Descriptor	RUBAI Operation	Human Action	Human Decision
TRIGA TYPE REACTORS	Keep	Keep	Approve
REACTORS	Remove		Approve
RESEARCH REACTORS	Keep	Keep	Approve
LEAD	Remove		Approve
RADIOGRAPHY	Add	Remove	Reject
MELANOMAS	Add	Remove	Reject

**DATASET**

Features					Label
Category	CMV	Weight	Title Match	Operation	Decision
60	-0.0211	0.2474	0	2: remove	1: approve
36	0.5676	0.4759	1	3: keep	0: reject
37	2.5355	0.7849	0	1: add	1: approve

# FROM A SINGLE DECISION TO A DATASET



Descriptor	RUBAI Operation	Human Action	Human Decision
TRIGA TYPE REACTORS	Keep	Keep	Approve
REACTORS	Remove		Approve
RESEARCH REACTORS	Keep	Keep	Approve
LEAD	Remove		Approve
RADIOGRAPHY	Add	Remove	Reject
MELANOMAS	Add	Remove	Reject

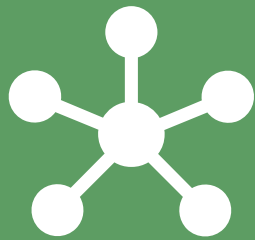
**DATASET**

Features					Label
Category	CMV	Weight	Title Match	Operation	Decision
60	-0.0211	0.2474	0	2: remove	1: approve
36	0.5676	0.4759	1	3: keep	0: reject
37	2.5355	0.7849	0	1: add	1: approve





**840** unique rules



Add

Replace

Remove

Keep

**KOS** integration



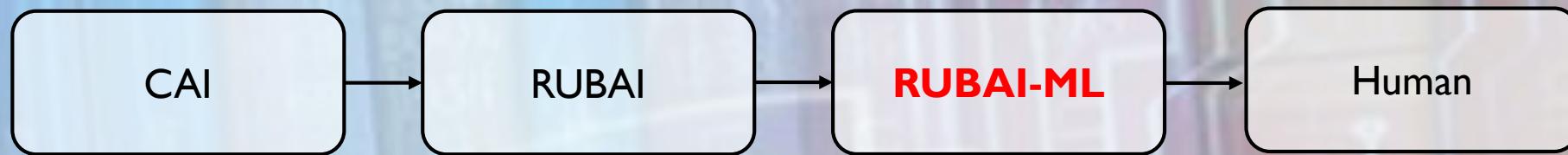
Thesaurus

INIS subjects

Expert knowledge

Quality control

# TWO-PASS INDEXING, TIER 2



**840**



**KOS integration**



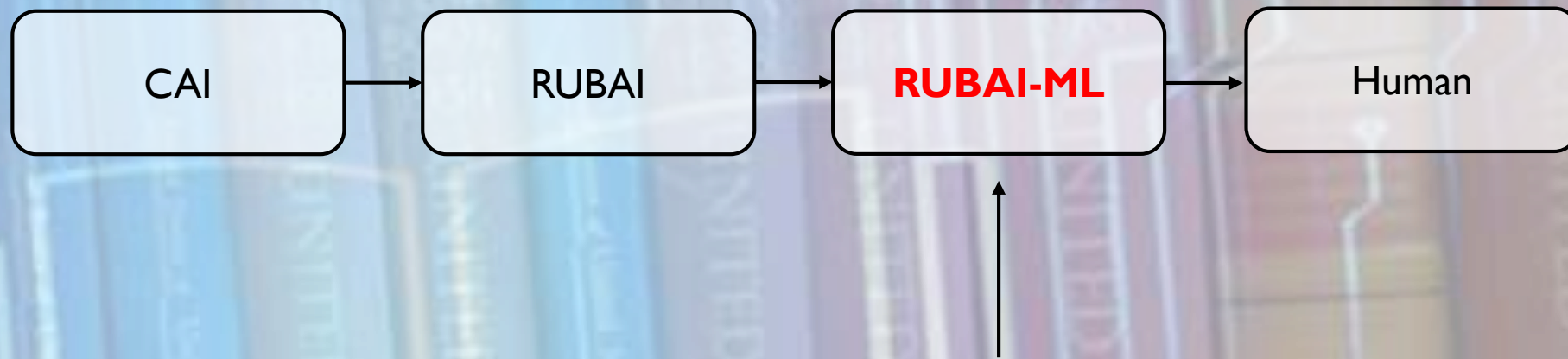
Thesaurus

INIS subjects

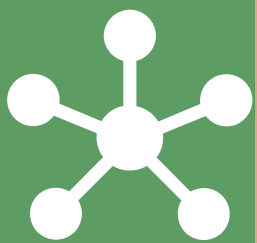
Expert knowledge

Quality control

# TWO-PASS INDEXING, TIER 2



**840**



**KOS**



inter

The

INIS

Exp

Qu

**DATA**



**analysis**

Machine Learning

KOS

Modeling human  
cognition

**DATA** set

17,500 documents

84% accuracy

130,000 unique  
decisions



# RUBAI-ML RESULTS: BRINGING THE OUTCOME...

<b>Model evaluation results</b>				
<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>Specificity</b>	<b>F<sub>1</sub></b>
0.837	0.838	0.793	0.873	0.815



# RUBAI-ML RESULTS: BRINGING THE OUTCOME...

Model evaluation results				
Accuracy	Precision	Recall	Specificity	F <sub>1</sub>
0.837	0.838	0.793	0.873	0.815

ID	Number of records	Descriptors										Operations per record performed by human after...	
		Total after...				added by human after...			removed by human after...			CAI	RUBAI
		CAI	RUBAI	RUBAI-ML	Human	CAI	RUBAI	RUBAI-ML	CAI	RUBAI	RUBAI-ML		
1	120	2519	1032	1040	1094	205	262	257	1626	200	202	3.85	3.83
2	111	2104	859	862	936	175	232	231	1343	155	157	3.49	3.49
3	109	2217	916	915	960	160	216	220	1417	172	175	3.56	3.62
4	22	474	169	172	198	44	53	52	320	24	26	3.5	3.54

# RUBAI-ML RESULTS: BRINGING THE OUTCOME...

Model evaluation results				
Accuracy	Precision	Recall	Specificity	F <sub>1</sub>
0.837	0.838	0.793	0.873	0.815

ID	Number of records	Total after...		
		CAI	RUBAI	RUBAI-ML
1	120	2519	1032	10
2	111	2104	859	8
3	109	2217	916	9
4	22	474	169	1

Operations per record performed by human after...	
RUBAI	RUBAI-ML
3.60	3.62

ID	Number of records	Operations per record performed by human after...	
		CAI	RUBAI
1	120	3.85	3.83
2	111	3.49	3.49
3	109	3.56	3.62
4	22	3.5	3.54

# RUBAI-ML RESULTS: BRINGING THE OUTCOME...

Model evaluation results				
Accuracy	Precision	Recall	Specificity	F <sub>1</sub>
0.837	0.838	0.793	0.873	0.815

		Total descriptors after...					
ID	Number of records	CAI	RUBAI	RUBAI-ML	HUMAN	per record	uman after...
		2519	1032	1040	1094		
1	120	2104	859	862	936	3.83	
2	111	2217	916	915	960	3.49	
3	109	474	169	172	198	3.62	
4	22	<b>1829</b>	<b>744</b>	<b>747</b>	<b>797</b>	3.54	

# RUBAI-ML RESULTS: BRINGING THE OUTCOME...

Model evaluation results				
Accuracy	Precision	Recall	Specificity	F <sub>1</sub>
0.837	0.838	0.793	0.873	0.815

		Total descriptors after...					
ID	Number of records	CAI	RUBAI	RUBAI-ML	HUMAN	per record human after...	RUBAI
1	120	2519	1032	1040	1094	3.83	
2	111	2104	859	862	936	3.49	
3	109	2217	916	915	960	3.62	
4	22	474	169	172	198	3.54	
		<b>1829</b>	<b>744</b>	<b>747</b>	<b>797</b>		

... CLOSER TO THE HUMAN CHOICES



## **Identifies missing descriptors in case of:**

- **Specific semantic relations**
- **Narrower descriptors**
- **Broader document scope**

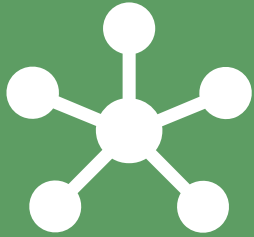
**840** unique rules

Add

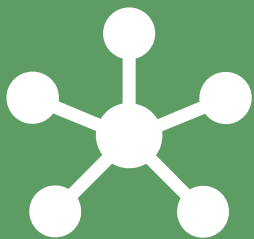
Replace

Remove

Keep



**840**



**KOS**



**integration**

Thesaurus

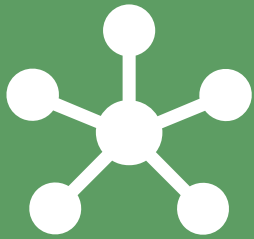
INIS subjects

Expert knowledge

Quality control

# CONCLUSIONS

**840**



**KOS**



**DATA**



**analysis**

Machine Learning

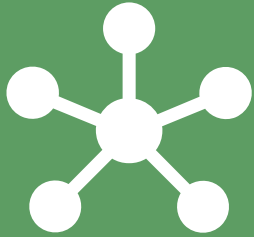
KOS

Modeling human  
cognition



# CONCLUSIONS

**840**



**KOS**



**DATA**



**DATA set**

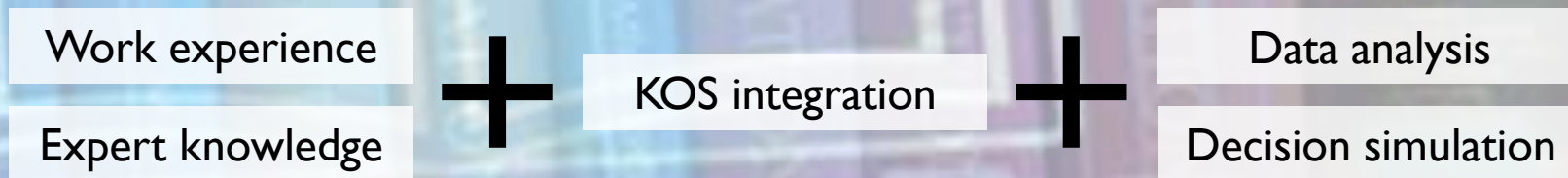
17,500 documents

84% accuracy

130,000 unique  
decisions



# CONCLUSIONS



## **Main outcomes:**

- **Model human decision-making strategies**
- **Gradual optimization of the indexing quality**
- **Automating the indexing process**

## Future work:

- **Extend the coverage and consistency of rules**
- **Improve ML validation component**
- **OR:** replace rules and ML validation with a single ML solution
- **DREAM:** completely data-driven classification algorithm based on (convolutional) neural networks



**THANK YOU !**

**International Nuclear Information System (INIS)**  
**[inis.iaea.org](http://inis.iaea.org)**

**INIS Repository**  
**[inis.iaea.org/search](http://inis.iaea.org/search)**

**INIS: The First Forty Years**  
**[www.iaea.org/sites/default/files/inis-40-anniversary.pdf](http://www.iaea.org/sites/default/files/inis-40-anniversary.pdf)**