Requirements Engineering

(Summer 2022)

Prof. Nan Niu (nan.niu@uc.edu)
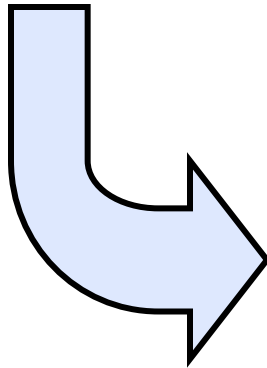
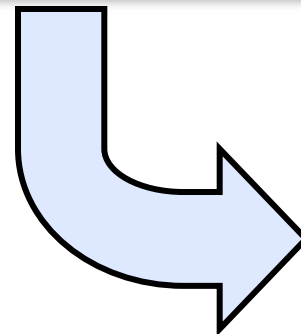## https://github.com/nanniu/RE-Summer2022

Today's Menu

Tuesday (July 19)
Visual modeling notations

More RE Stories

Wednesday (July 20):
ASN3 Due
Req.s Traceability
ASN4 Release

Thursday (July 21):
Tracing effectiveness
Learning-based methods
ASN5 release

What's "req.s traceability"?

⇨ **the ability to describe and follow the life of a requirement**

Have we seen *traceability* already?

An SRS should be

a)   Correct;
b)   Unambiguous;
c)   Complete;
d)   Consistent;
e)   Ranked for importance and/or stability;
f)   Verifiable;
g)   Modifiable;
h)   Traceable.

Source: IEEE STD 830

Why caring about "traceability"?

→ **Many standards consider it a quality indicator**

    ↳ **IEEE STD-830-1998, "Guides to Software Requirements Specifications"**

    ↳ **CMMI**

    ↳ **U.S. Federal Aviation Administration (FAA)**

    ↳ **...**

→ **It is indispensable for carrying out many software engineering activities**

    ↳ **Verification, e.g., whether code satisfies design**

    ↳ **Validation, e.g., whether stakeholders' goals have been fulfilled**

    ↳ **Change impact analysis, e.g., how much code will be affected if this requirement changes**

    ↳ **System-level test coverage analysis**

    ↳ **Risk assessment**

    ↳ **...**

Story about ChoicePoint

→ **ChoicePoint**

↳ **Headquarters: Alpharetta (near Atlanta), Georgia, USA**

↳ **A data aggregation company**

➢ Combined personal data sourced from multiple public and private databases for sale to the government and the private sector
➢ Maintained more than 17 billion records of individuals and businesses

→ **Security breach**

↳ **In 2006, records on more than *163,000* consumers were acquired by identity thieves**

→ **Review by the US FCRA (Fair Credit Reporting Act)**

↳ **Revealed that ChoicePoint has developed the software products without proper controls mandated by the FCRA**

↳ **ChoicePoint was fined *$15 million* in civil penalties**

↳ **ChoicePoint must undergo biennial security audits for the next *20 years***

Story

→ **ChoicePoint**
- ⤷ Headquarters: Alp
- ⤷ A data aggregation
  - ➢ Combined personal
    databases for sale
  - ➢ Maintained more t

→ **Security breach**
- ⤷ In 2006, records on more than *163,000* consumers were acquired by identity thieves

→ **Review by the US FCRA (Fair Credit Reporting Act)**
- ⤷ Revealed that ChoicePoint has developed the software products without proper controls mandated by the FCRA
- ⤷ ChoicePoint was fined *$15 million* in civil penalties
- ⤷ ChoicePoint must undergo biennial security audits for the next *20 years*

Tracing/Relating FRs to NFRs

↪ **NFR1 (Operational): The system shall interface with the Choice Parts System. This provides the feed of recycled parts data.**

↪ **NFR2 (Usability): Users shall feel satisfied using the system; 85% of all users will be satisfied with the system.**

↪ FR_i: The user shall search for the preferred repair facility using vehicle location and radius in miles.

↪ FR_j: The estimator shall search for available recycled parts using damaged vehicle parts information.

↪ FR_k: The display shall have two regions: left of the display is graphical right of the display is a data table.

Correct Answers

	NFR1 (Operational)	NFR2 (Usability)
FR_i	0	1
FR_j	1	1
FR_k	0	0

Spreadsheet (tracing manually in general) doesn't work

→Tedious, time-consuming, & error-prone

⇨Scalability, e.g., easily hundreds of requirements & other artifacts

⇨Evolving, i.e., keeping up with the changing software in a spreadsheet is not always a good use of your time

In practice, traceability is often dropped, or performed as needed (as opposed to systematically).

Tracing manually (e.g., Spreadsheet) in general doesn't work
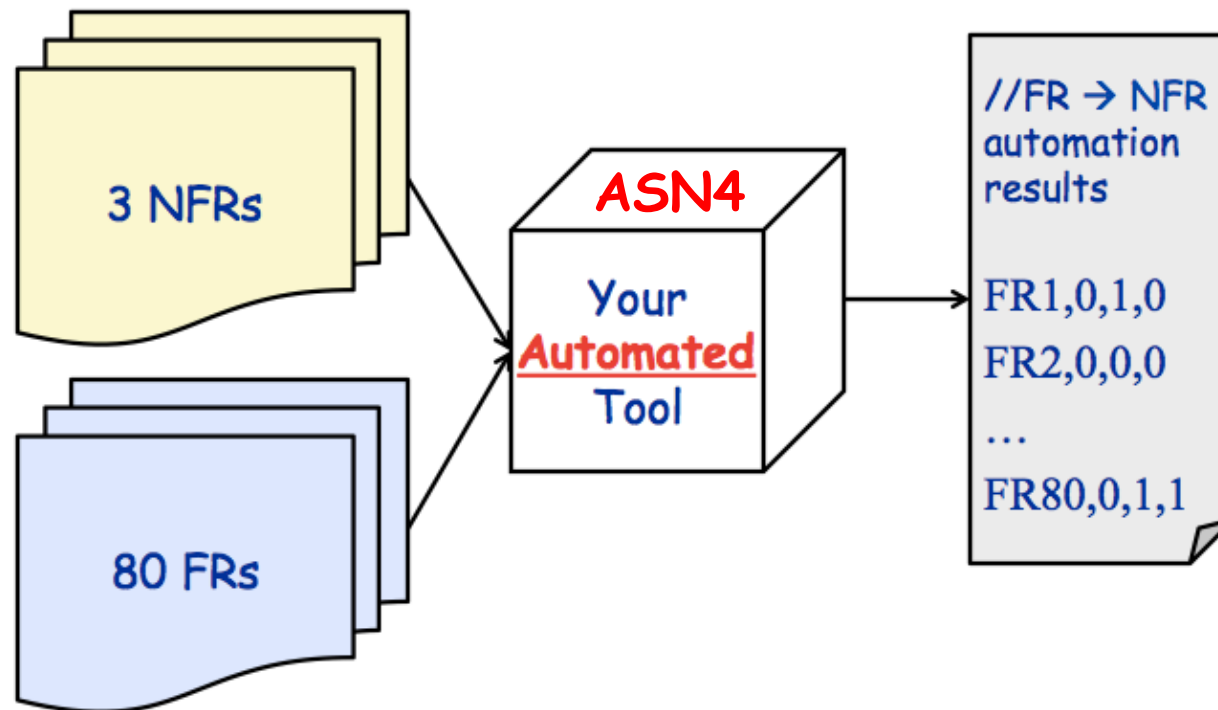
→Tedious, time-consuming, & error-prone

→What we want = what **ASN4** is about
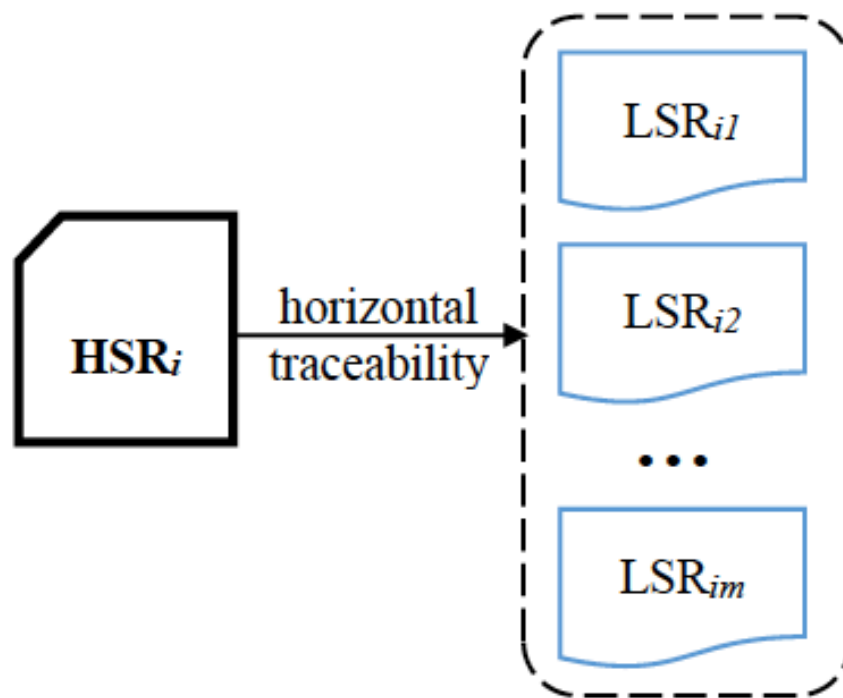
⤷Scalable ⬅ (fully) automatic

⤷On-the-fly ⬅ no need to store & maintain any static traceability information, so we can trace at any time without worrying about software evolution

Correct Answers

	NFR1 (Operational)	NFR2 (Usability)	NFR3 (Security)
FR_i	0	1	0
FR_j	1	1	0

3 NFRs

80 FRs

ASN4

Your **Automated** Tool

//FR → NFR automation results

FR1,0,1,0

FR2,0,0,0

...

FR80,0,1,1

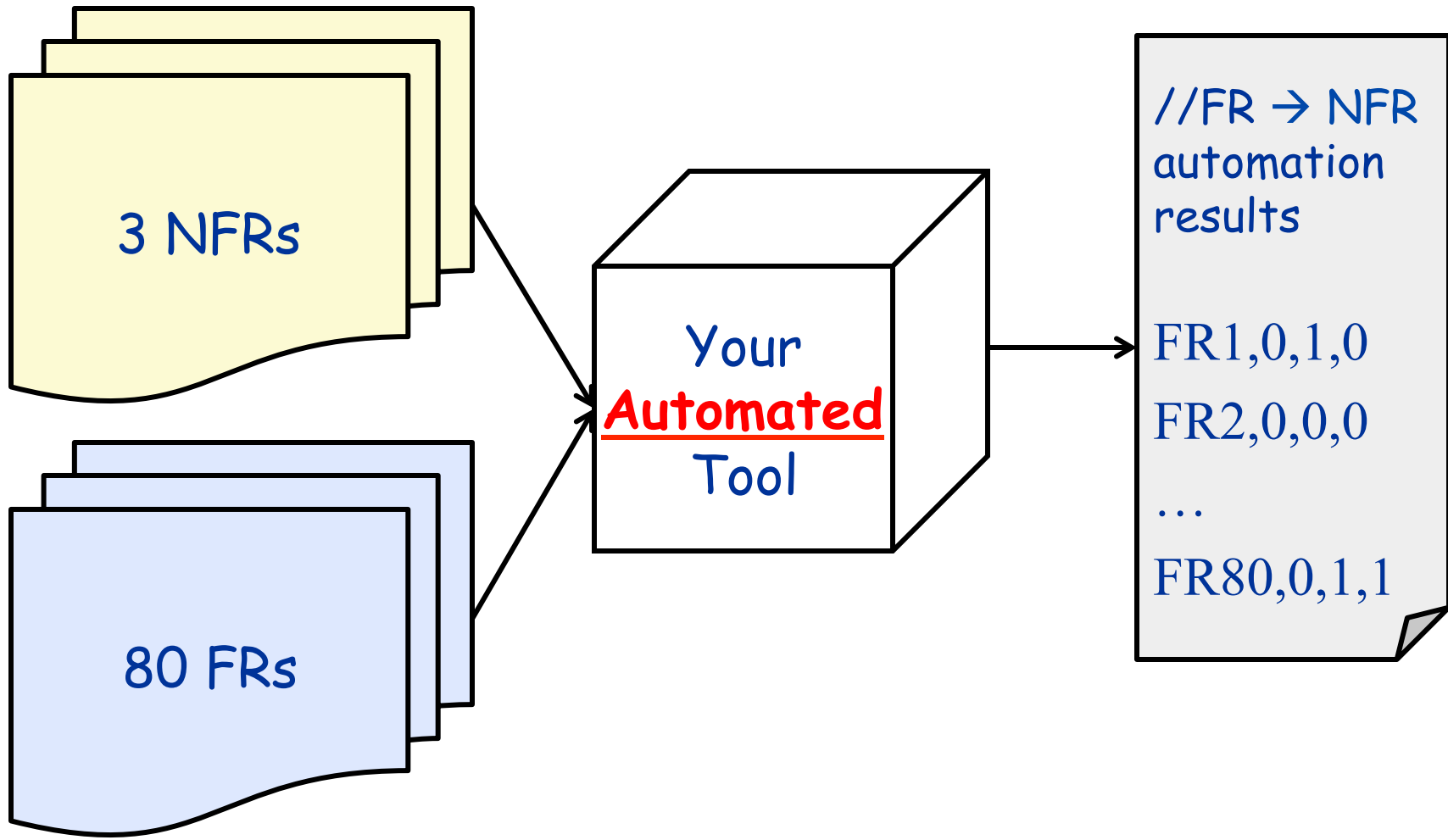# Horizontal and vertical traceability

artifacts of *different* stages & granularities
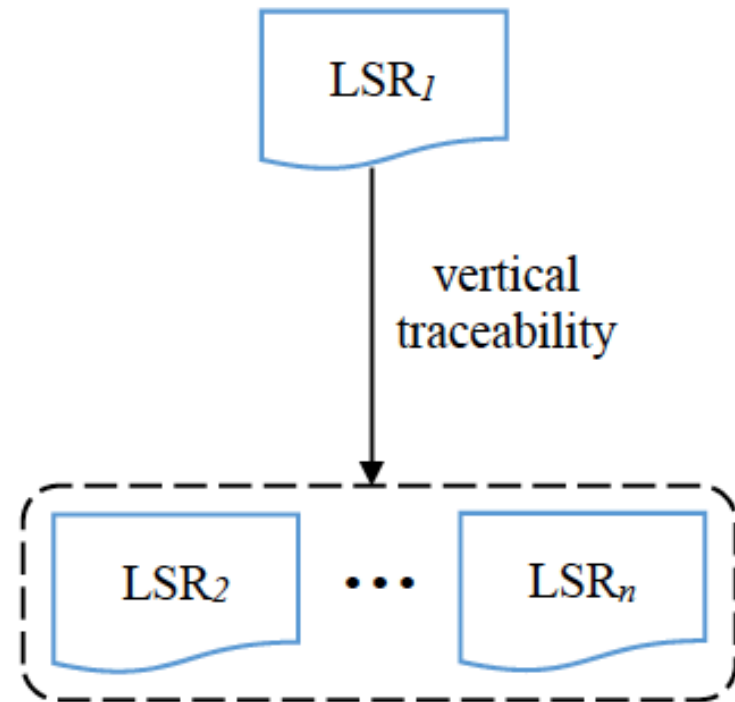
artifacts of *same* stage & granularity

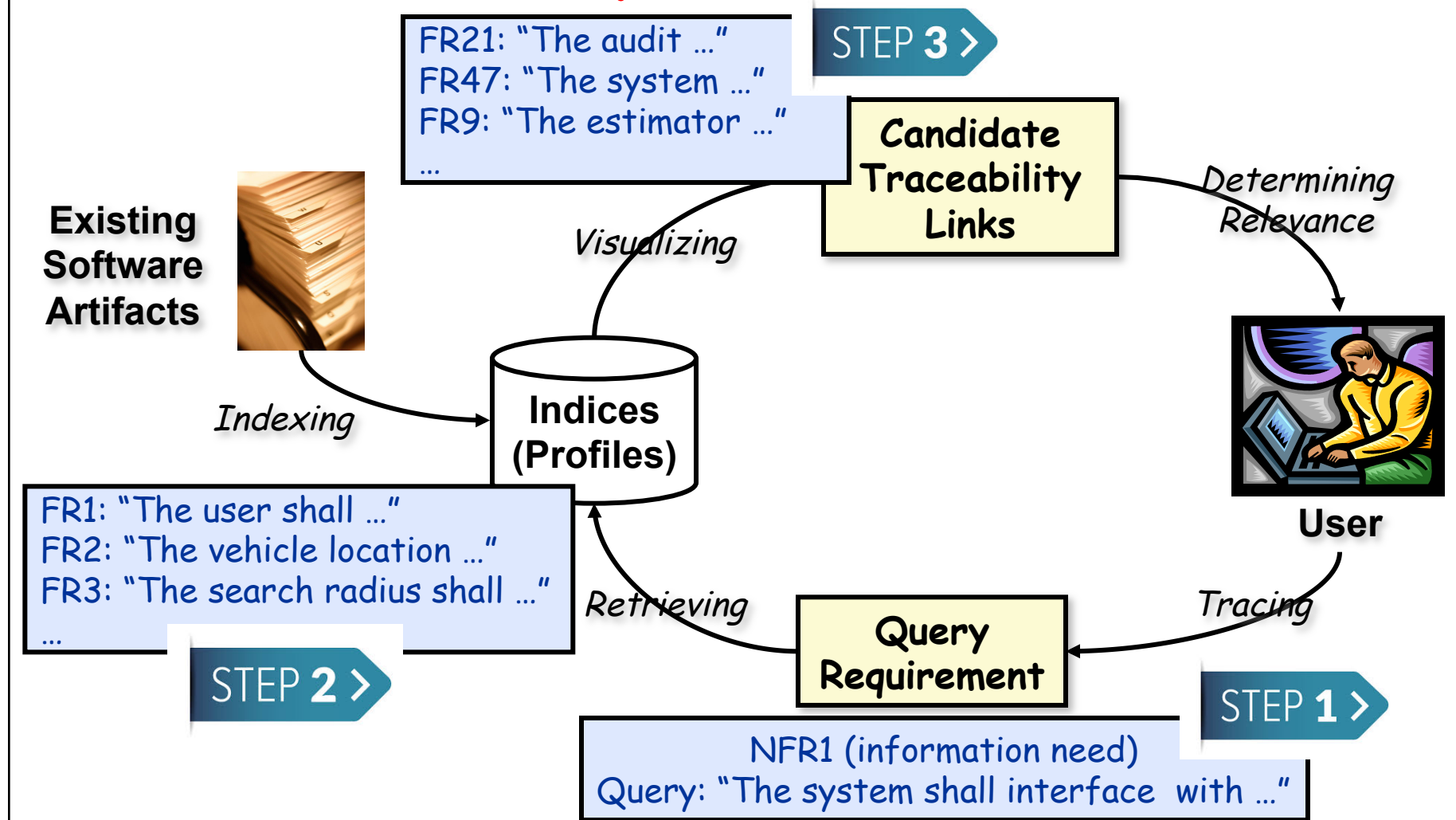# ASN4: A conceptual picture & also an instance of *"req.s tracing"*

3 NFRs

80 FRs

Your
Automated
Tool

//FR → NFR
automation
results

FR1,0,1,0

FR2,0,0,0

…

FR80,0,1,1

IR-Based Requirements Traceability
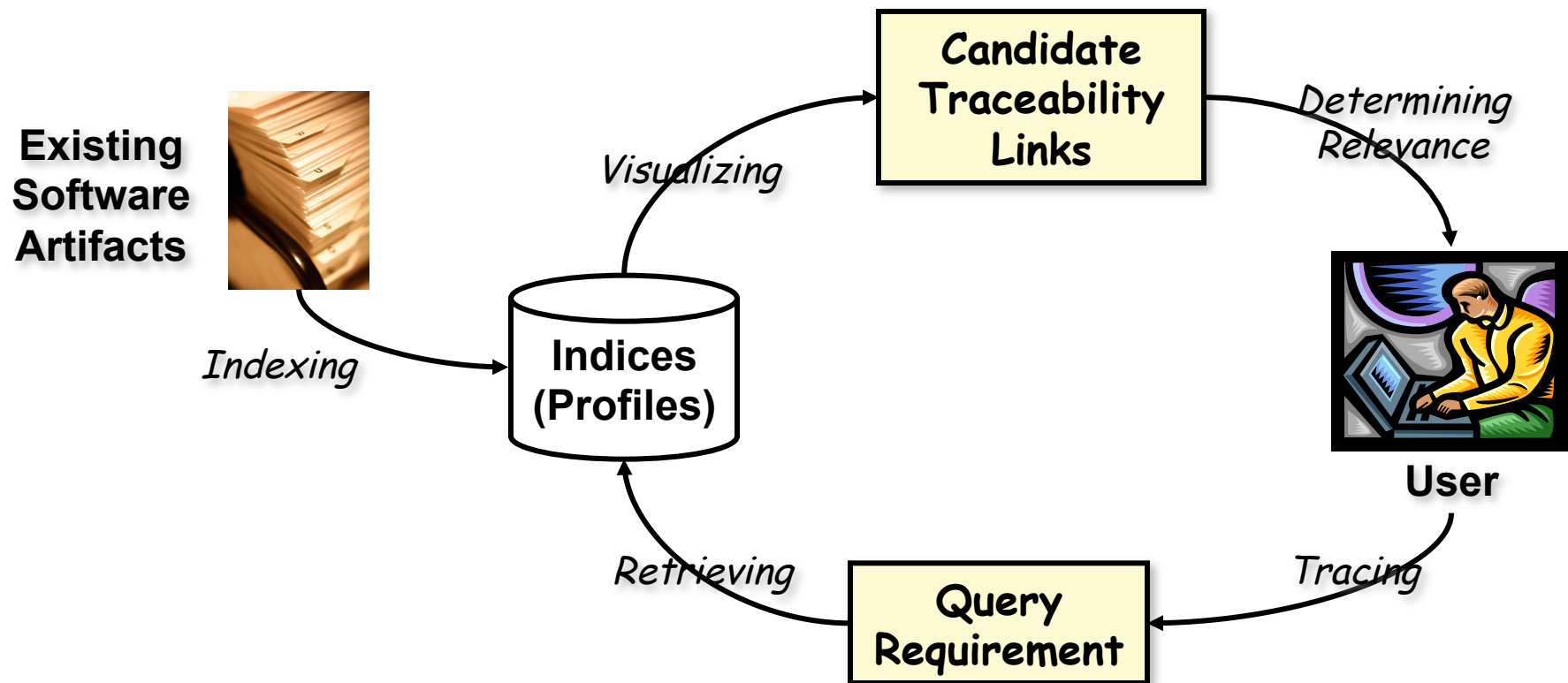
IR-Based ASN4 Solution
(fully automatic)

FR21: "The audit ..."
FR47: "The system ..."
FR9: "The estimator ..."
...

STEP **3** >

Candidate Traceability Links

Determining Relevance

Existing Software Artifacts

Visualizing

Indexing

Indices (Profiles)

User

FR1: "The user shall ..."
FR2: "The vehicle location ..."
FR3: "The search radius shall ..."
...

STEP **2** >

Retrieving

Query Requirement

Tracing

STEP **1** >

NFR1 (information need)
Query: "The system shall interface with ..."

Example

→ **Two requirements**

↳ **r1 = "create and deactivate patients profile"**

↳ **r2 = "patients create and edit profile"**

→ **In this lecture, we introduce some basic retrieval methods: set-based, Jaccard, tf-idf.**

→ **Assumption of IR-based ASN4 solution**
  ↳ **the more textual similarity there is between the two requirements, the more likely one is linked with (traceable to) the other**
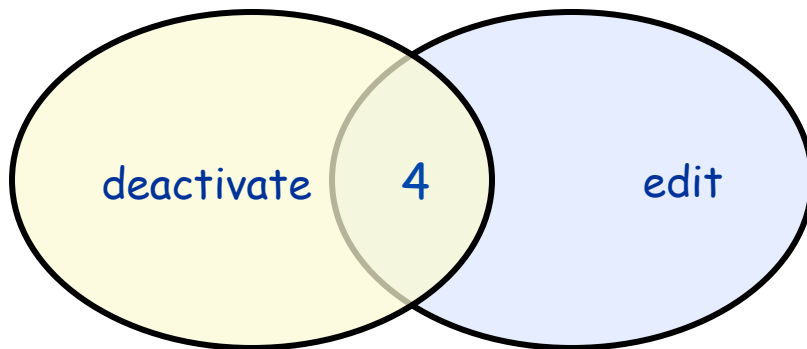
Similarity based on set overlapping

→ Basic formula

$$S(R1, R2) = \frac{2\,|R1 \cap R2|}{(|R1| + |R2|)}$$

> ↳ r1 = "create and deactivate patients profile"
>
> ↳ r2 = "patients create and edit profile"

deactivate   4   edit

→ Resulting similarity

↳ S(r1, r2) = (2x4) / (5+5) = 0.8

↳ Suppose the threshold is 0.5, then {r1, r2} would be regarded as traceable to each other

Similarity based on Jaccard index

→ **Basic formula**   The Jaccard similarity coefficient, J, is given as

$$J = \frac{M_{11}}{M_{01} + M_{10} + M_{11}}.$$

M_{11} represents the total number of attributes where A and B both have a value of 1.
M_{01} represents the total number of attributes where the attribute of A is 0 and the attribute of B is 1.
M_{10} represents the total number of attributes where the attribute of A is 1 and the attribute of B is 0.
M_{00} represents the total number of attributes where A and B both have a value of 0.
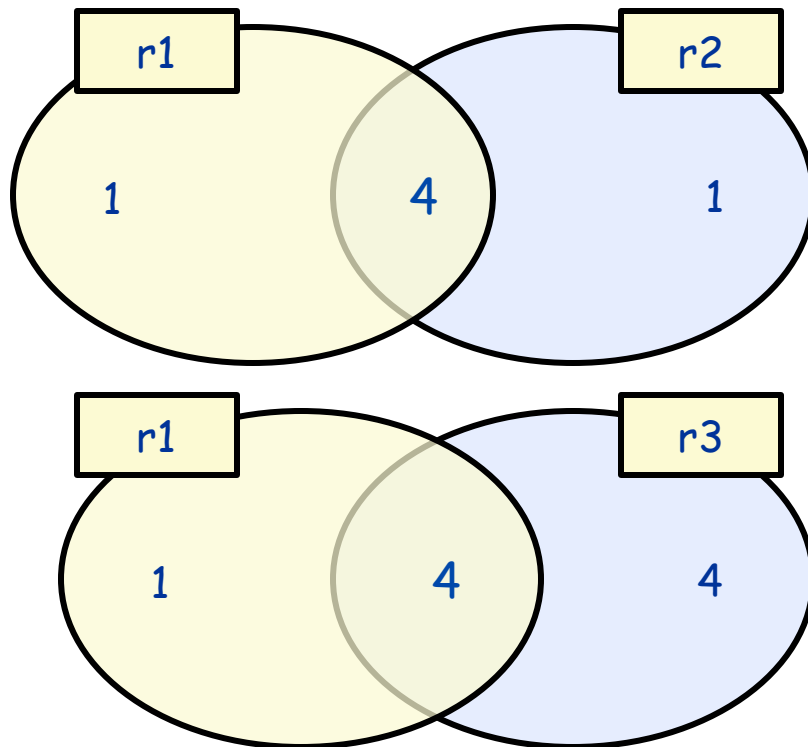
→ **In our example**

	create	and	deactivate	patients	profile	edit	a	including	photo
r1	1	1	1	1	1	0	0	0	0
r2	1	1	0	1	1	1	0	0	0
r3	1	1	0	1	1	1	1	1	1

Jaccard (cont'd)

↳ r1 = "create and deactivate patients profile"

↳ r2 = "patients create and edit profile"

↳ r3 = "patients create and edit profile including a photo"

r1   1   4   1   r2

r1   1   4   4   r3

→ **Set-based similarity**

   ↳ S(r1, r2) = (2×4) / (5+5) = 0.8
   ↳ S(r1, r3) = (2×4) / (5+8) = 0.62

→ **Jaccard-based similarity**

   ↳ S(r1, r2) = 4 / 6 = 0.67
   ↳ S(r1, r3) = 4 / 9 = 0.44

Results So Far (threshold=0.5)

→ Our example

  ↳ r1 = "create and deactivate patients profile"

  ↳ r2 = "patients create and edit profile"

  ↳ r3 = "patients create and edit profile including a photo"

  ↳ r4 = "patients create and and edit edit edit profile"

→ Set-based overlap   → Jaccard index   → tf-idf

Set-based overlap		Jaccard index		tf-idf	
2,4	(1.00)	2,4	(1.00)	2,4	(0.89)
1,2	(0.80)	1,2	(0.67)	1,2	(0.56)
1,4	(0.80)	1,4	(0.67)	2,3	(0.50)
2,3	(0.77)	2,3	(0.63)		
3,4	(0.77)	3,4	(0.63)	3,4	(0.45)
1,3	(0.62)			1,4	(0.36)

VSM (vector space model)
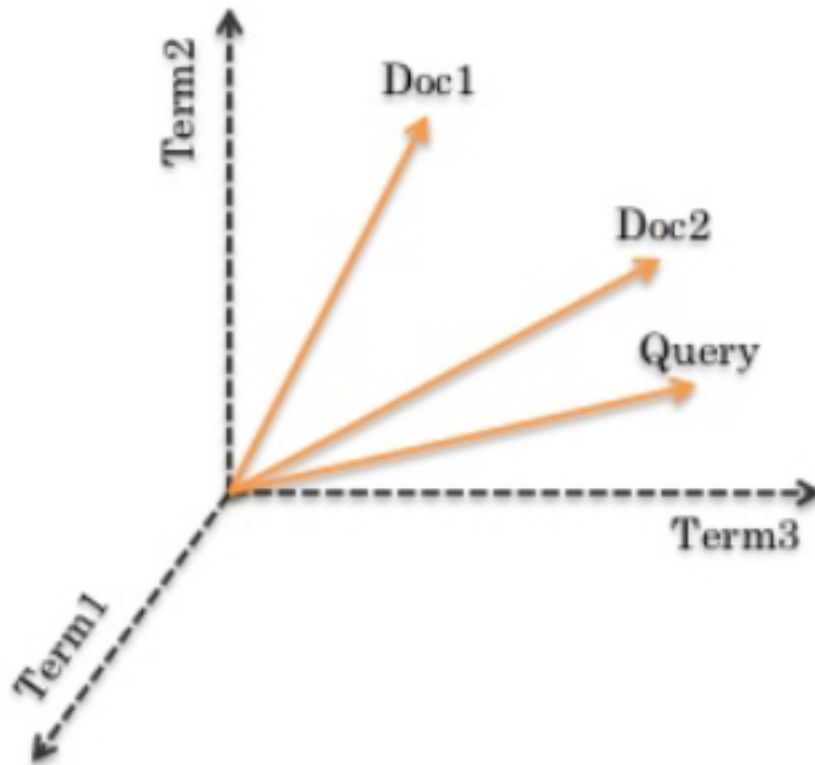
$$Cos(D_i, Q) = \frac{\sum_{j=1}^{t} d_{ij} * q_j}{\sqrt{\sum_{j=1}^{t} d_{ij}^2 * \sum_{j=1}^{t} q_j^2}}$$

tf-idf

	create	and	deactivate	patients	profile	edit	a	...	photo
r1	1	1	1	1	1	0	0	...	0
r2	1	1	0	1	1	1	0	...	0
r3	1	1	0	1	1	1	1	...	1
r4	1	2	0	1	0	3	0	...	0

$$sim(d,q) = \cos(d,q) = \frac{\sum_{i=1}^{N} w_i \cdot q_i}{\sqrt{\sum_{i=1}^{N} w_i^2 \cdot \sum_{i=1}^{N} q_i^2}} \cdot$$

$$w_i = tf_i(d) \cdot idf_i,$$

$$idf_i = \log_2\left(\frac{n}{df_i}\right),$$

sim(r2, r4) =

[1*log(4/4+1)]*[1*log(4/4+1)]   //create
+ [1*log(4/4+1)]*[2*log(4/4+1)]   //and
+ [1*log(4/4+1)]*[1*log(4/4+1)] //patients
+ [1*log(4/3+1)]*[0*log(4/3+1)]   //profile
+ [1*log(4/3+1)]*[3*log(4/3+1)]   //edit
...   // denominator

= 0.89

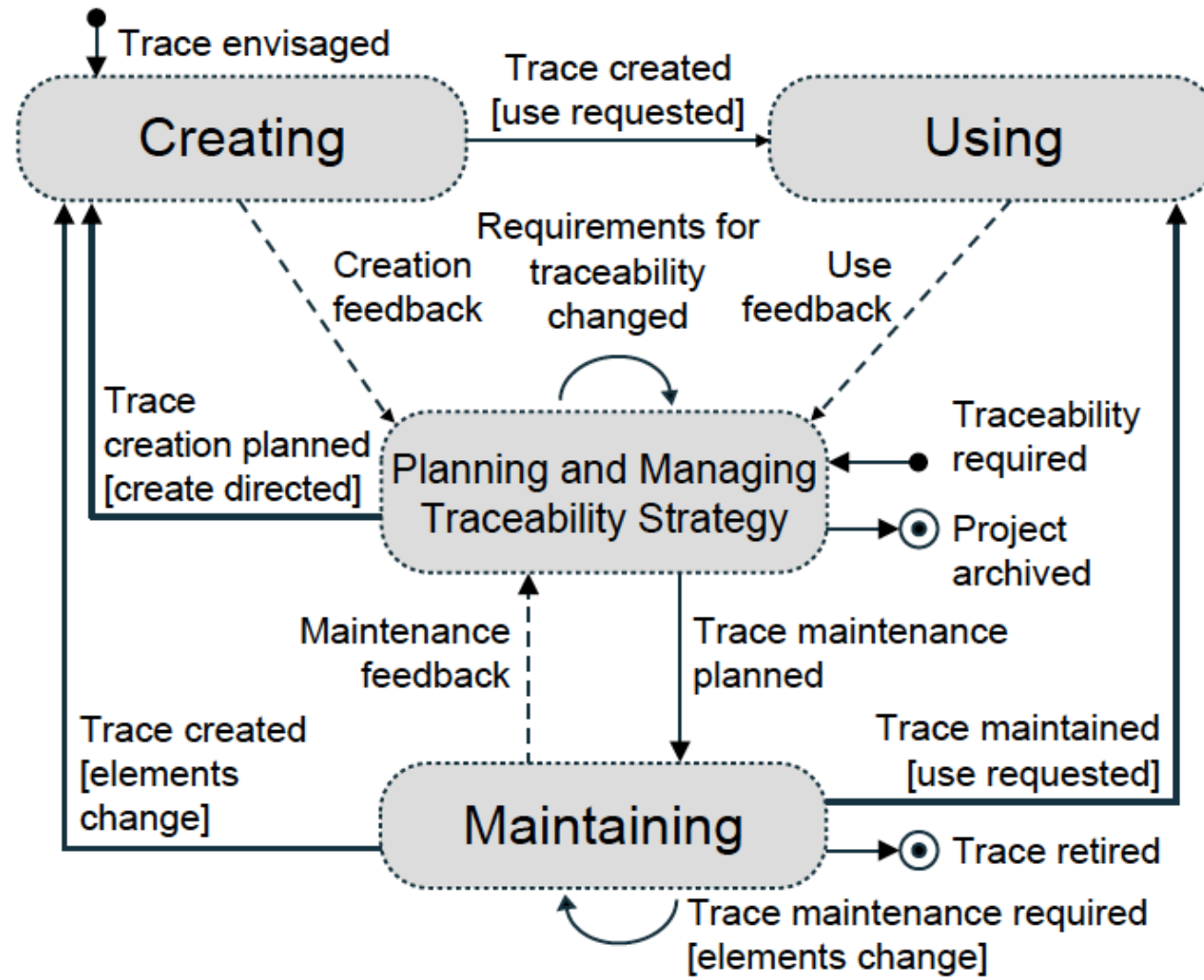# Demo Jaccard & tf-idf
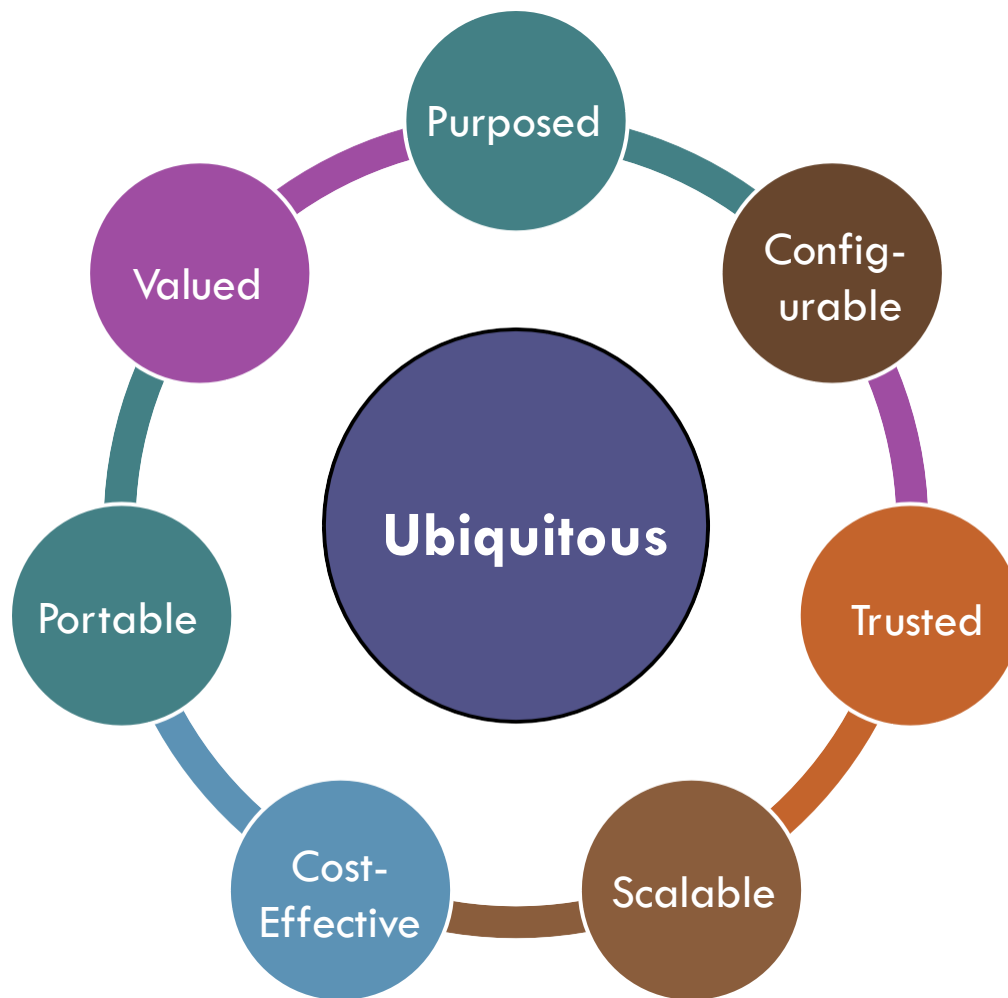
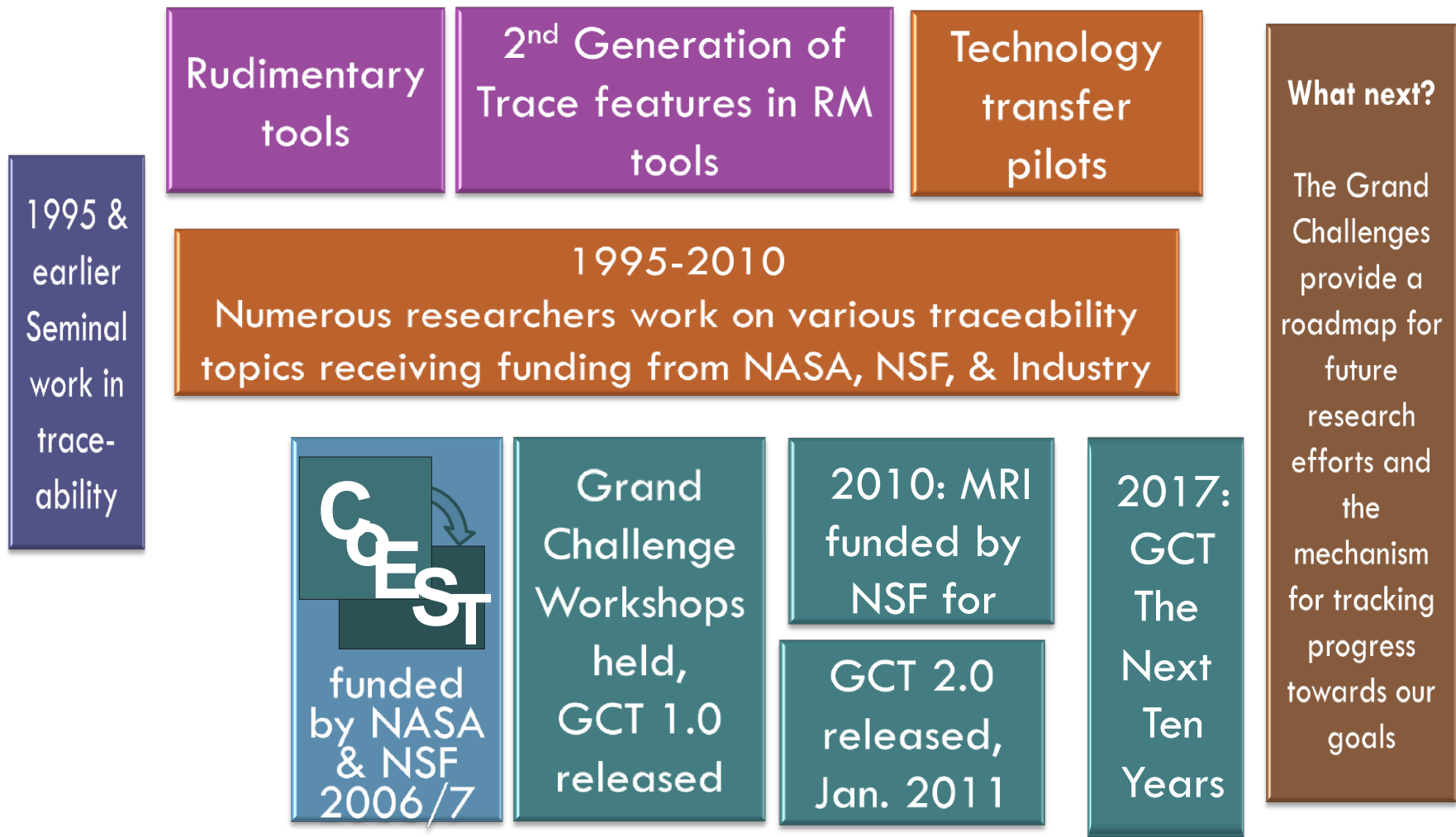# Requirements Traceability: Activities

Goal-Oriented View

→ Traceability is always there... is neither consciously established nor sought; it is built-in and effortless.

It has effectively 'disappeared without a trace'.

Roadmap of "req.s traceability"

1995 & earlier Seminal work in trace-ability

Rudimentary tools

2nd Generation of Trace features in RM tools

Technology transfer pilots

1995-2010
Numerous researchers work on various traceability topics receiving funding from NASA, NSF, & Industry

CoEST funded by NASA & NSF 2006/7

Grand Challenge Workshops held, GCT 1.0 released

2010: MRI funded by NSF for

GCT 2.0 released, Jan. 2011

2017: GCT The Next Ten Years

What next?

The Grand Challenges provide a roadmap for future research efforts and the mechanism for tracking progress towards our goals

Working on traceability (papers) is fun

SST'19 —— Software and Systems Traceability

10th International Workshop at the 41st International Conference on Software Engineering (ICSE), May 27, 2019

Journals & Magazines > IEEE Access > Volume: 6 ❓

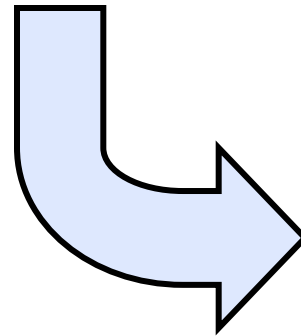## Recommending Refactoring Solutions Based on Traceability and Code Metrics

Publisher: IEEE   Cite This   📄 PDF

Ally S. Nyamawe 🆔 ; Hui Liu ; Zhendong Niu ; Wentao Wang ; Nan Niu   **All Authors**

Tomorrow, we're headed to …

Thursday (July 21):
Tracing effectiveness
Learning-based methods
ASN5 release