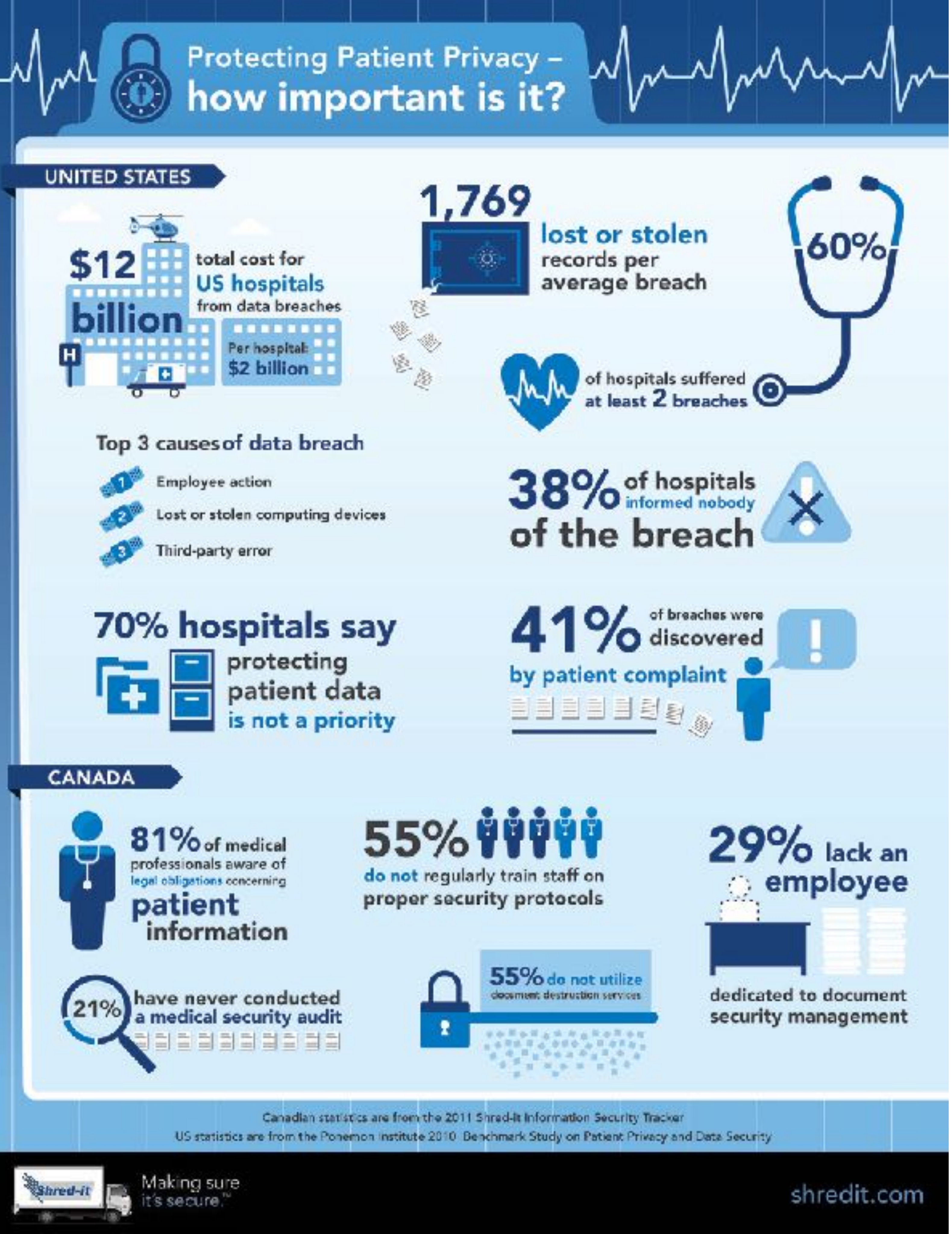
Privacy in Healthcare Data Sharing

Challenges and Opportunities

Nan Zhang

Associate Professor, The George Washington University
Program Director, National Science Foundation



Protecting Patient Privacy – how important is it?

UNITED STATES

$12 billion total cost for **US hospitals** from data breaches

Per hospital: $2 billion

1,769 lost or stolen records per average breach

60% of hospitals suffered at least 2 breaches

Top 3 causes of data breach

1. Employee action
2. Lost or stolen computing devices
3. Third-party error

38% of hospitals informed nobody of the breach

70% hospitals say protecting patient data is not a priority

41% of breaches were discovered by patient complaint

CANADA

81% of medical professionals aware of legal obligations concerning **patient information**

55% do not regularly train staff on proper security protocols

29% lack an employee dedicated to document security management

21% have never conducted a medical security audit

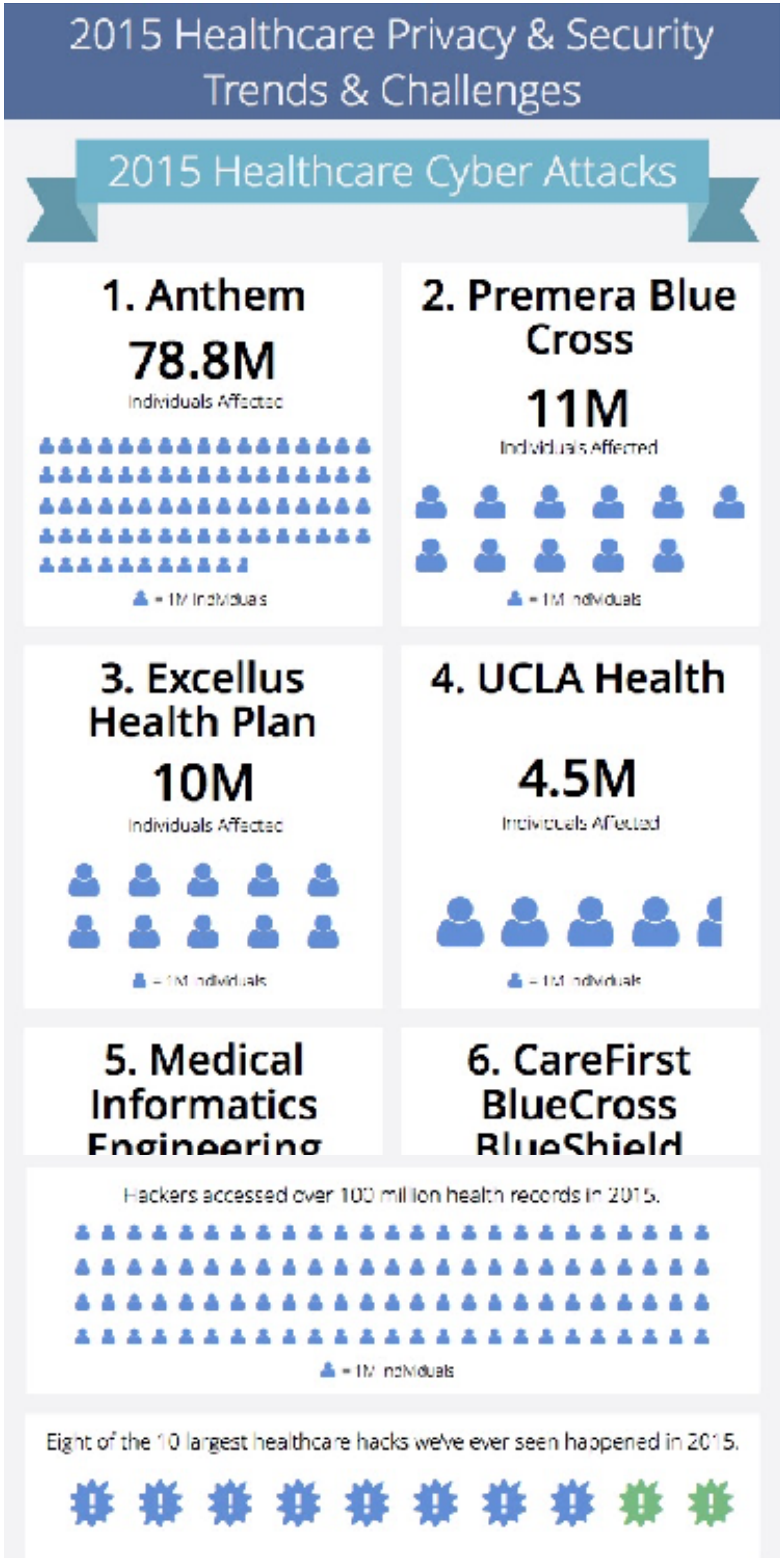
55% do not utilize document destruction services

Canadian statistics are from the 2011 Shred-it Information Security Tracker

US statistics are from the Ponemon Institute 2010 Benchmark Study on Patient Privacy and Data Security

Shred-it Making sure it's secure.™

shredit.com



2015 Healthcare Privacy & Security Trends & Challenges

2015 Healthcare Cyber Attacks

1. Anthem

78.8M

Individuals Affected

● = 1M Individuals

2. Premera Blue Cross

11M

Individuals Affected

● = 1M Individuals

3. Excellus Health Plan

10M

Individuals Affected

● = 1M Individuals

4. UCLA Health

4.5M

Individuals Affected

● = 1M Individuals

5. Medical Informatics Engineering

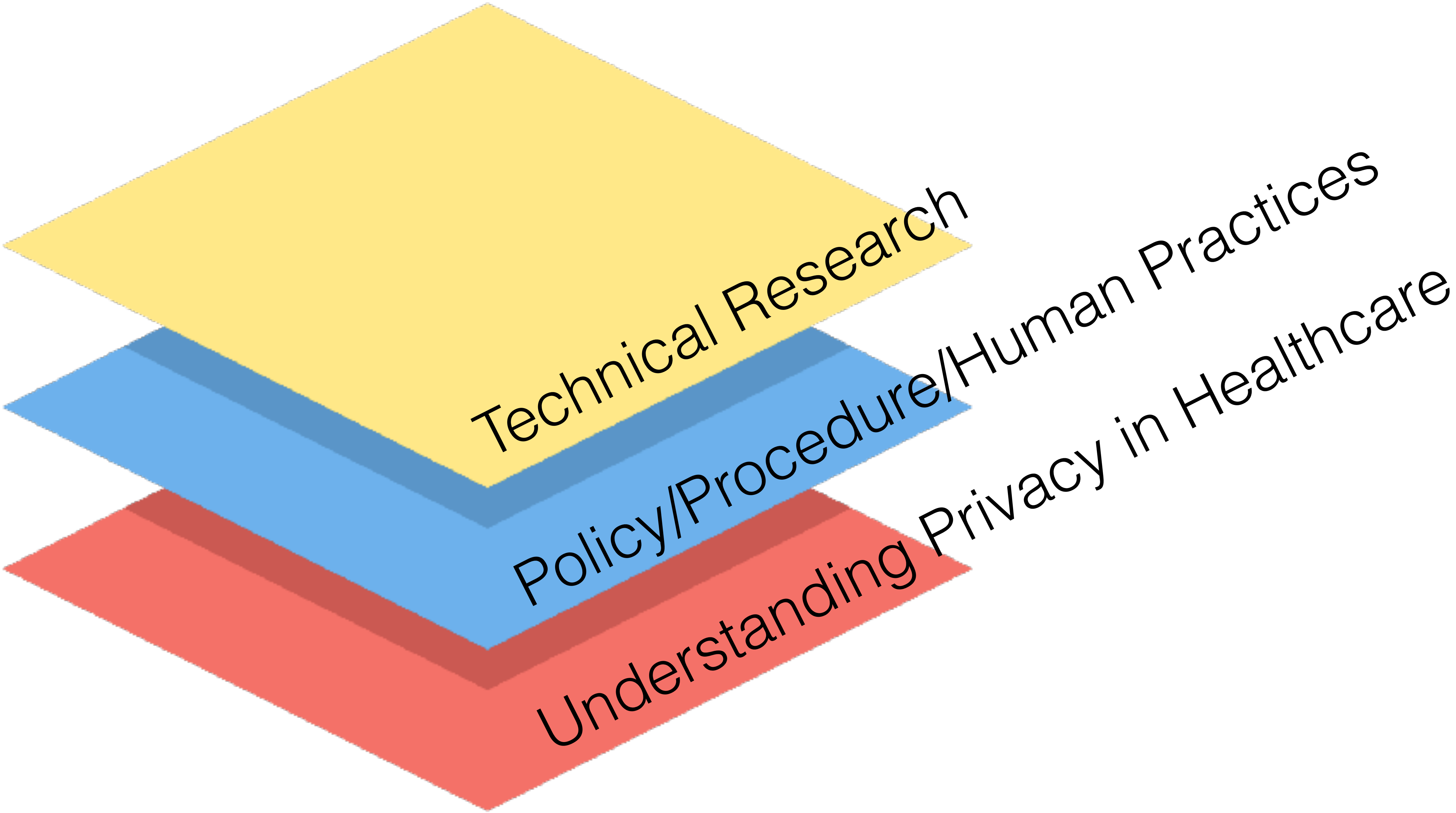
6. CareFirst BlueCross BlueShield

Hackers accessed over 100 million health records in 2015.

● = 1M Individuals

Eight of the 10 largest healthcare hacks we've ever seen happened in 2015.

Challenges


Technical Research
Policy/Procedure/Human Practices
Understanding Privacy in Healthcare

What is Privacy?

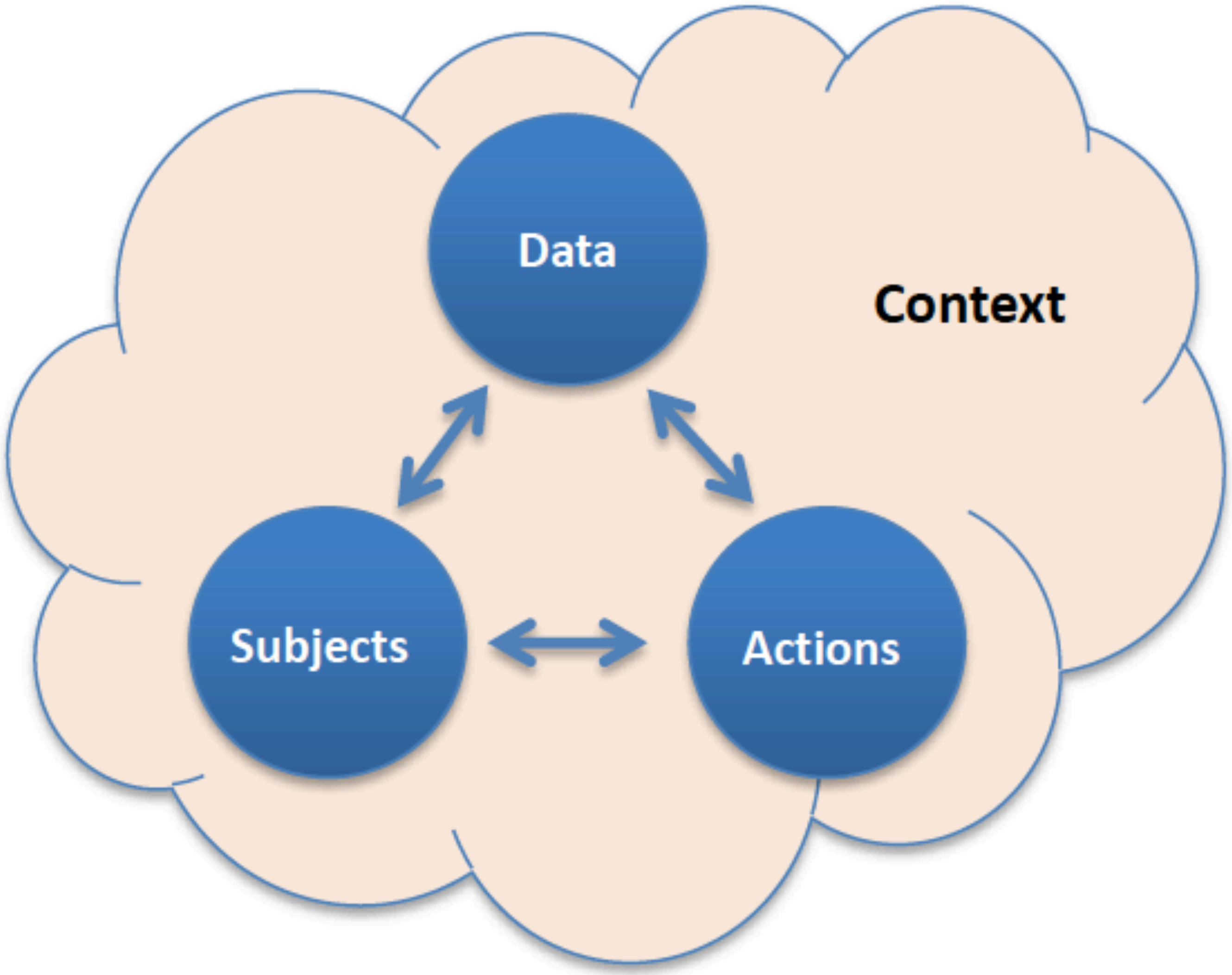
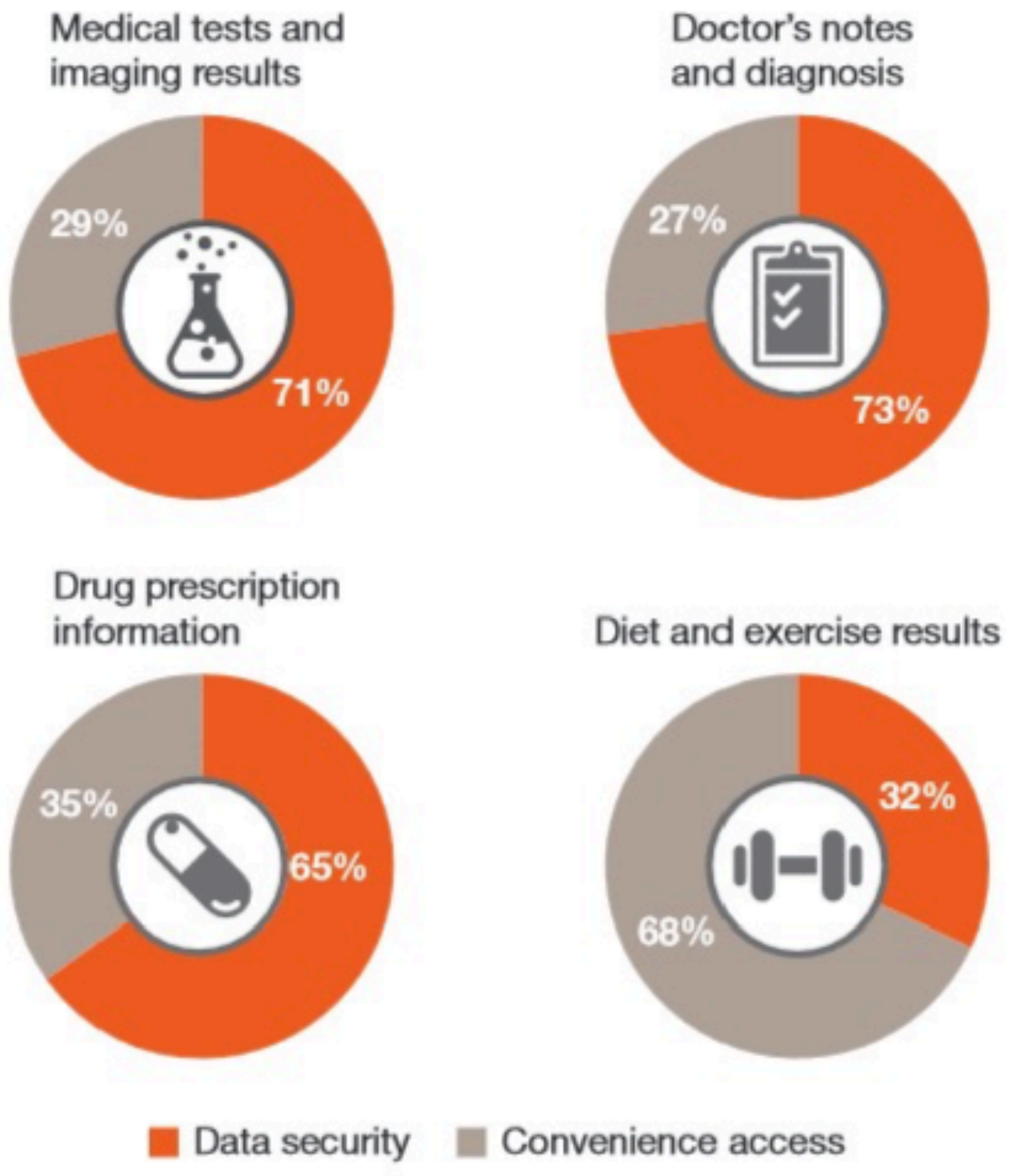


Figure 1: NPRS privacy characterization.

National Privacy Research Strategy (NPRS):
https://www.whitehouse.gov/sites/default/files/nprs_nstc_review_final.pdf

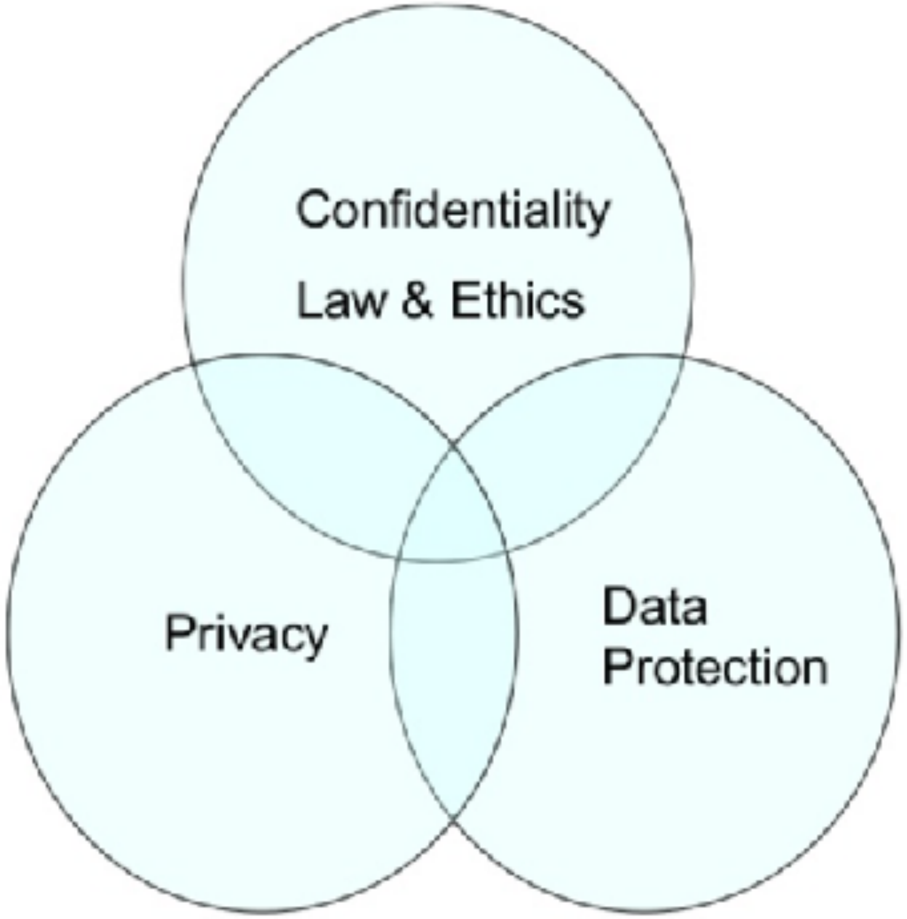
Complex Privacy Construct in Healthcare

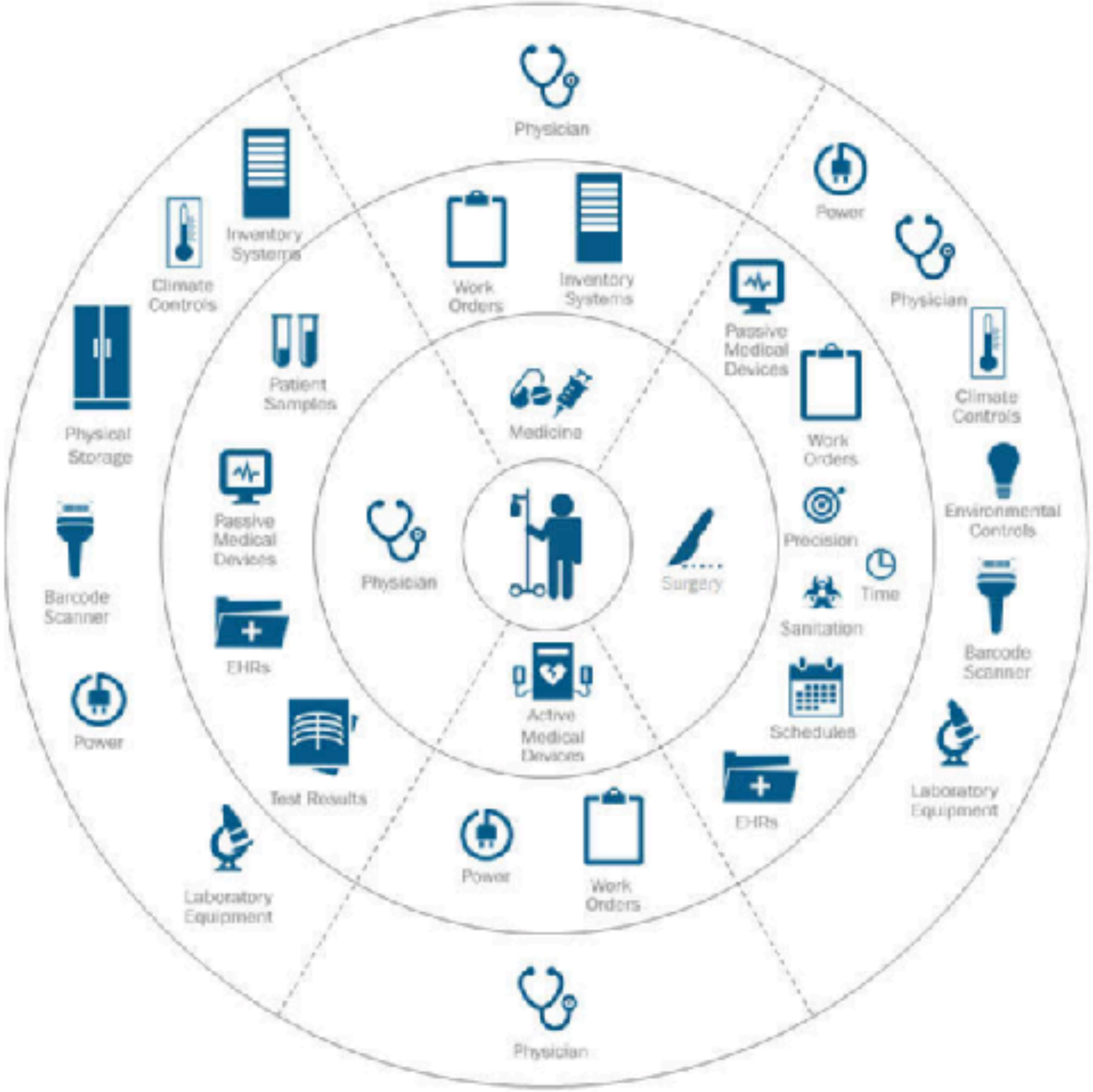
- Subjects
  - Patients, Clinical research subjects
- Actions
  - Medical treatment, Research
- Data
  - Personal info, Diagnosis, Medical tests, Prescription, Diet
- Context



Sources: HRI Consumer Survey, PwC, 2014

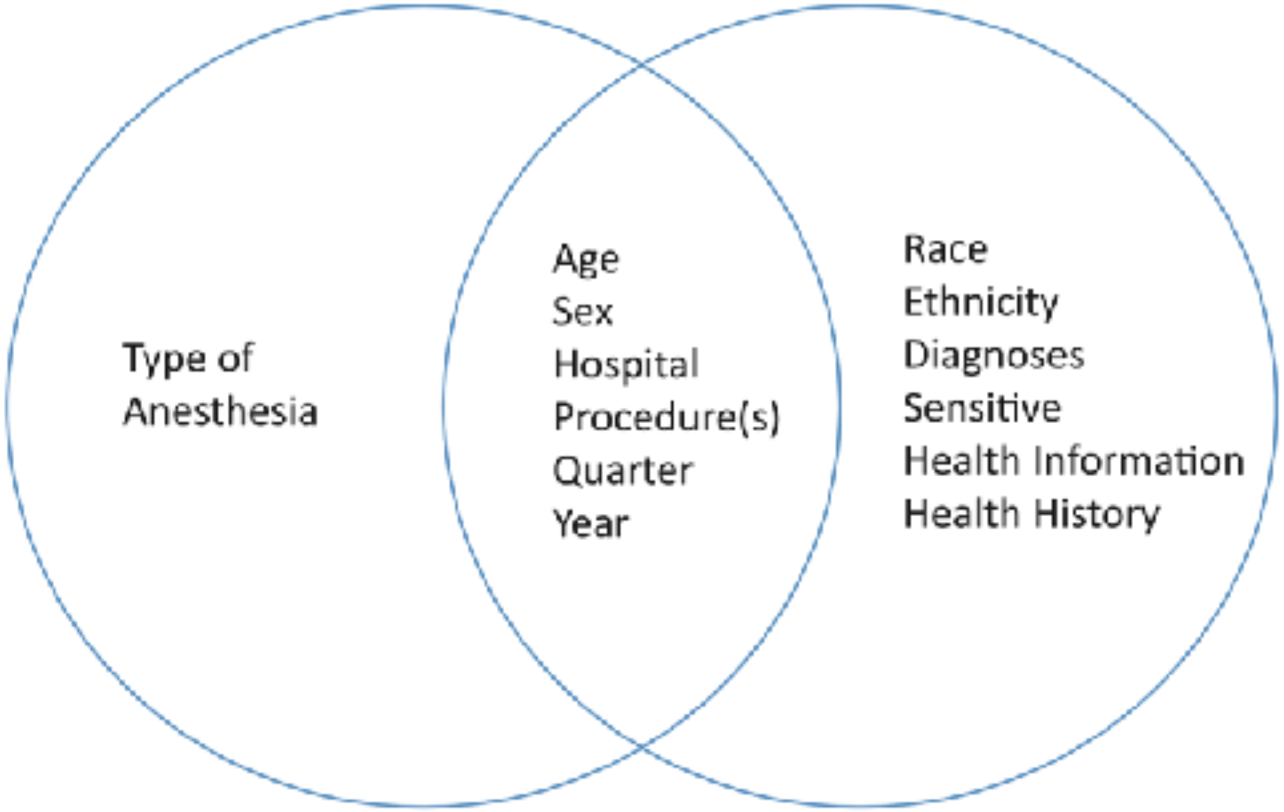
Complex Privacy Construct in Healthcare





from S. Dobridnjuk, European Standards on Confidentiality and Privacy in Healthcare
from ISE, Securing Hospitals: A research study and blueprint

Case Study 1: Clinical Anesthesia Studies



Threat: Record linkage with external data sources

L. O'Neil, F. Dexter, N. Zhang, The Risks to Patient Privacy from Publishing Data from Clinical Anesthesia Studies, Anesthesia & Analgesia, 122(6), 2016

Case Study 1: Clinical Anesthesia Studies

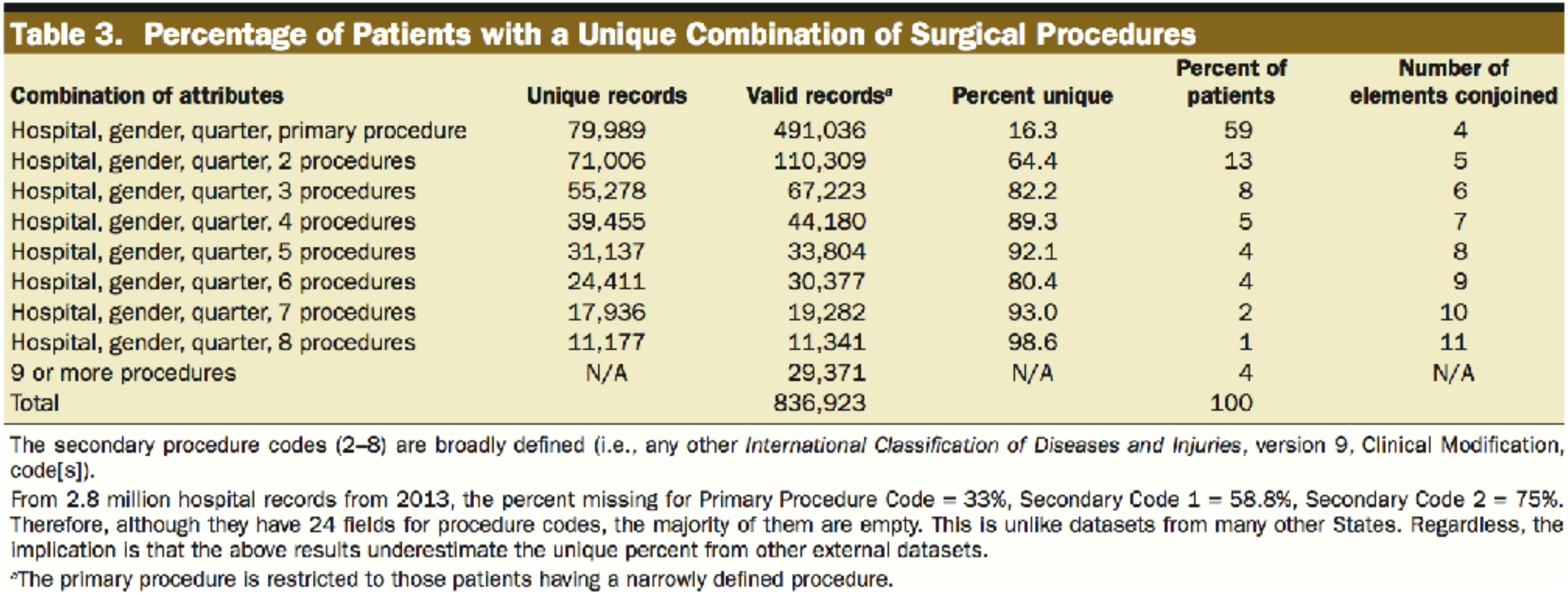


Table 3. Percentage of Patients with a Unique Combination of Surgical Procedures

Combination of attributes	Unique records	Valid records[a]	Percent unique	Percent of patients	Number of elements conjoined
Hospital, gender, quarter, primary procedure	79,989	491,036	16.3	59	4
Hospital, gender, quarter, 2 procedures	71,006	110,309	64.4	13	5
Hospital, gender, quarter, 3 procedures	55,278	67,223	82.2	8	6
Hospital, gender, quarter, 4 procedures	39,455	44,180	89.3	5	7
Hospital, gender, quarter, 5 procedures	31,137	33,804	92.1	4	8
Hospital, gender, quarter, 6 procedures	24,411	30,377	80.4	4	9
Hospital, gender, quarter, 7 procedures	17,936	19,282	93.0	2	10
Hospital, gender, quarter, 8 procedures	11,177	11,341	98.6	1	11
9 or more procedures	N/A	29,371	N/A	4	N/A
Total		836,923		100	

The secondary procedure codes (2–8) are broadly defined (i.e., any other *International Classification of Diseases and Injuries*, version 9, Clinical Modification, code[s]).

From 2.8 million hospital records from 2013, the percent missing for Primary Procedure Code = 33%, Secondary Code 1 = 58.8%, Secondary Code 2 = 75%. Therefore, although they have 24 fields for procedure codes, the majority of them are empty. This is unlike datasets from many other States. Regardless, the implication is that the above results underestimate the unique percent from other external datasets.

[a]The primary procedure is restricted to those patients having a narrowly defined procedure.

S71.041A: Puncture wound with foreign body, right hip, initial encounter

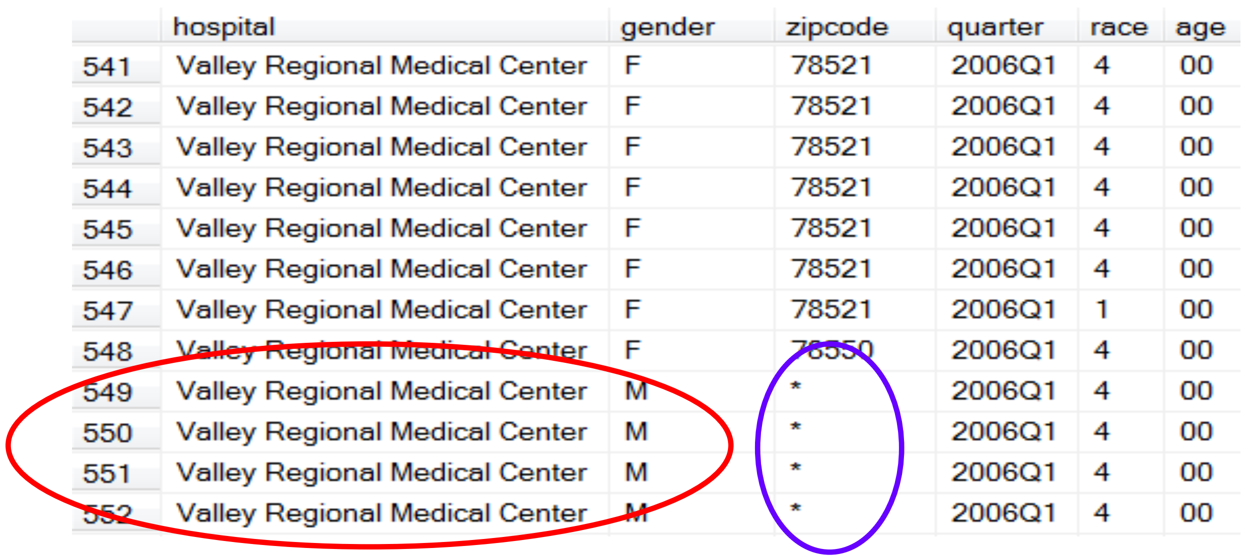
Implications on Policy / Procedure

Case Study 2: Public Health Data Sharing

- The last two digits of the patient's ZIP code are suppressed if there are fewer than thirty patients included in the ZIP code.
- The entire ZIP code is suppressed if a hospital has fewer than fifty discharges in a quarter.
- The entire ZIP code and gender code are suppressed if the ICD-9-CM code indicates alcohol or drug use or an HIV diagnosis.
- The entire ZIP code and provider name are suppressed if a hospital has fewer than five discharges of a particular gender, including ‘unknown’. The provider ID is changed to '999998'.
- The country code is suppressed if the country field has fewer than five discharges for that quarter .
- The county code is suppressed if a county has fewer than five discharges for that quarter .
- Age is represented by 22 age group codes for the general patient population and 5 age group codes for the HIV and alcohol and drug use patient populations.
- Race is changed to ‘Other’ and ethnicity is suppressed if a hospital has fewer than ten discharges of a race.
- If a hospital has fewer than fifty discharges in a quarter, the provider ID is changed to ‘999999’.

Texas Inpatient Public Use Data File (PUDF), https://www.dshs.texas.gov/thcic/hospitals/Inpatientpudf.shtm

Case Study 2: Public Health Data Sharing



	hospital	gender	zipcode	quarter	race	age
541	Valley Regional Medical Center	F	78521	2006Q1	4	00
542	Valley Regional Medical Center	F	78521	2006Q1	4	00
543	Valley Regional Medical Center	F	78521	2006Q1	4	00
544	Valley Regional Medical Center	F	78521	2006Q1	4	00
545	Valley Regional Medical Center	F	78521	2006Q1	4	00
546	Valley Regional Medical Center	F	78521	2006Q1	4	00
547	Valley Regional Medical Center	F	78521	2006Q1	1	00
548	Valley Regional Medical Center	F	78550	2006Q1	4	00
549	Valley Regional Medical Center	M	*	2006Q1	4	00
550	Valley Regional Medical Center	M	*	2006Q1	4	00
551	Valley Regional Medical Center	M	*	2006Q1	4	00
552	Valley Regional Medical Center	M	*	2006Q1	4	00

Example: If a hospital has fewer than five discharges of a particular gender, then suppress the zipcode of its patients of that gender.

hospital, gender $\not\rightarrow$ zipcode

"It may be possible in rare instances, through complex analysis and with outside information, to ascertain from the PUDF the identity of individual patients. Considerable harm could result if this were done. PUDF users are required to sign and comply with the DSHS Hospital Discharge Data Use Agreement in the Application before shipment of the PUDF. The Data Use Agreement prohibits attempts to identify individual patients."

M. F. Rahman, W. Liu, S. Thirumuruganathan, N. Zhang, G. Das, Privacy Implications of Database Ranking, VLDB 2015.
X. Jin, M. Zhang, N. Zhang, G. Das, Versatile Publishing for Privacy Preservation, KDD 2010

NSF Opportunities for Healthcare Privacy Research

- **Privacy Research**
  - In August 2013 and in February 2014, the White House Office of Science and Technology Policy (OSTP) issued two Requests For Information (RFI) on privacy research activities pursued by the agencies
  - NSF: Approximately $25M per year is invested in privacy research activities
    - Approximately 35% of the Secure and Trustworthy Cyberspace (SaTC) program
- **Healthcare**
  - NITRD: The Federal Government, under the leadership of NSF and Health and Human Services (NIH, ONC, AHRQ) should invest in a national, long-term, multi-agency research initiative on NIT for health that goes well beyond the current national program to adopt electronic health records.
  - NSF Smart and Connected Health (SCH) Program

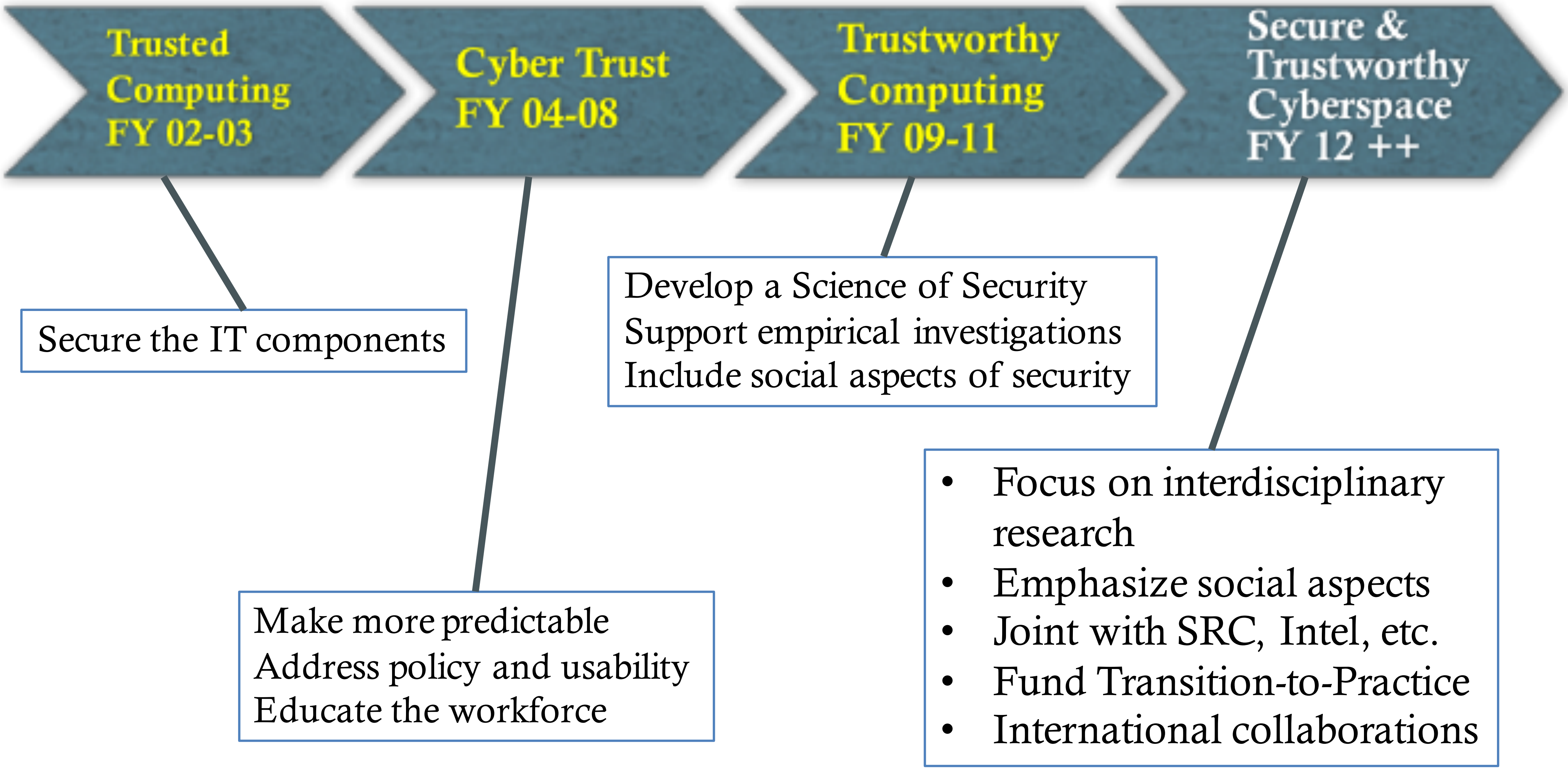
NSF Secure and Trustworthy Cyberspace (SaTC) Program

- NSF’s flagship research program for research in cybersecurity
  - SaTC is the largest unclassified cybersecurity research program in the world
- Primarily targeted at US colleges & universities
- Also open to US non-profits, and sometimes for-profits
- \$75M+ in FY16 grant cycle, ~200 new grants (FY15), ~900 active grants

Sizes / Schedule / Results (core program 16-580)

	Amount & duration	Submission Deadline	# FY15 funded
Small	Up to $500k, 3 years	November 16, 2016	74 proposals/ 60 projects
Medium	Up to $1.2M, 4 years	October 19, 2016	38 proposals/ 23 projects
Large	Up to $3M, 5 years	October 19, 2016	10 proposals/ 3 projects
Cybersecurity Education	Up to $300K, 2 years	Dec 15, 2016	8 proposals/ 6 projects

SaTC



Trusted Computing FY 02-03

Cyber Trust FY 04-08

Trustworthy Computing FY 09-11

Secure & Trustworthy Cyberspace FY 12 ++

Secure the IT components

Make more predictable
Address policy and usability
Educate the workforce

Develop a Science of Security
Support empirical investigations
Include social aspects of security

- Focus on interdisciplinary research
- Emphasize social aspects
- Joint with SRC, Intel, etc.
- Fund Transition-to-Practice
- International collaborations

SATC Frontiers Portfolio: 2012-2014

Data Privacy

- Privacy Tools for Sharing Research Data (2012)
- Harvard University
- $4.8M for 4 years

Socio-economic

- Beyond Technical Security: Developing an Empirical Basis for Socio-Economic Perspectives (2012)
- UCSD, Berkeley, GMU
- $10M for 5 years

Healthcare

- Enabling Trustworthy Cybersystems for Health and Wellness (2013)
- Dartmouth, UIUC, JHU, Michigan
- $10M for 5 years

Web Privacy

- Towards Effective Web Privacy Notice and Choice: a Multi-disciplinary Perspective (2013)
- CMU, Fordham, Stanford
- $3.75M for 4 years

Trust in Cloud

- Rethinking Security in the Era of Cloud Computing (2013)
- UNC, NCSU, Stony Brook, Duke, Wisconsin-Madison
- $6M for 5 years

Outsourced Computation

- Modular Approach to Cloud Security (2014)
- BU, MIT, Northeastern, U. Connecticut
- $4.9M for 5 years

Program Obfuscation

- Center for Encrypted Functionalities (2014)
- UCLA, Stanford, Columbia, UT Austin, JHU
- $10M for 5 years

SBE/SaTC

- SBE / SaTC seeks to fund cutting edge SBE research proposals that
  - Have the potential to enhance the trustworthiness and security of cyberspace AND
  - contribute to theory or methodology of basic SBE sciences
- Researchers are encouraged to include SBE science and collaborate with SBE scientists as needed
- Uses the domain of cybersecurity to explore, develop or "push the boundaries" of SBE science.
  - Make theoretical or methodological contributions to the SBE sciences
  - Seek generalizable theories
  - Proposals will be reviewed by SBE scientists

Transition to Practice Option/Perspective

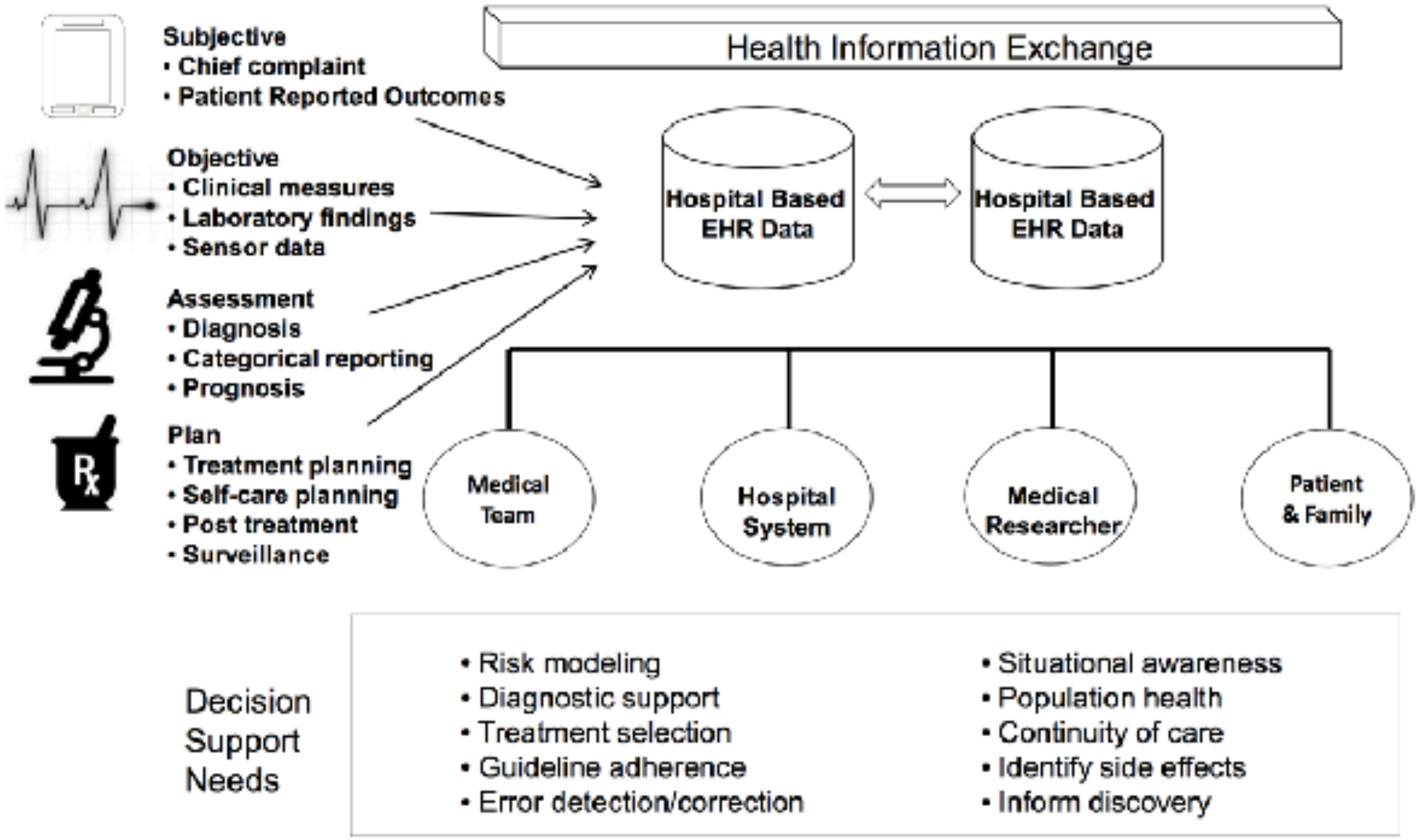
- Supports later stage activities in the research and development lifecycle such as prototyping and experimental deployment
- **Exclusively** on transitioning existing research results to practice

- In FY15, was an option (up to $167K extra for Small, up to $400K extra for Medium in addition to research grant)
- In FY16, was a perspective (up to $500K/Small or $1.2M/Medium)
- For FY17, is a designation (up to $500K/Small or $1.2M/Medium)

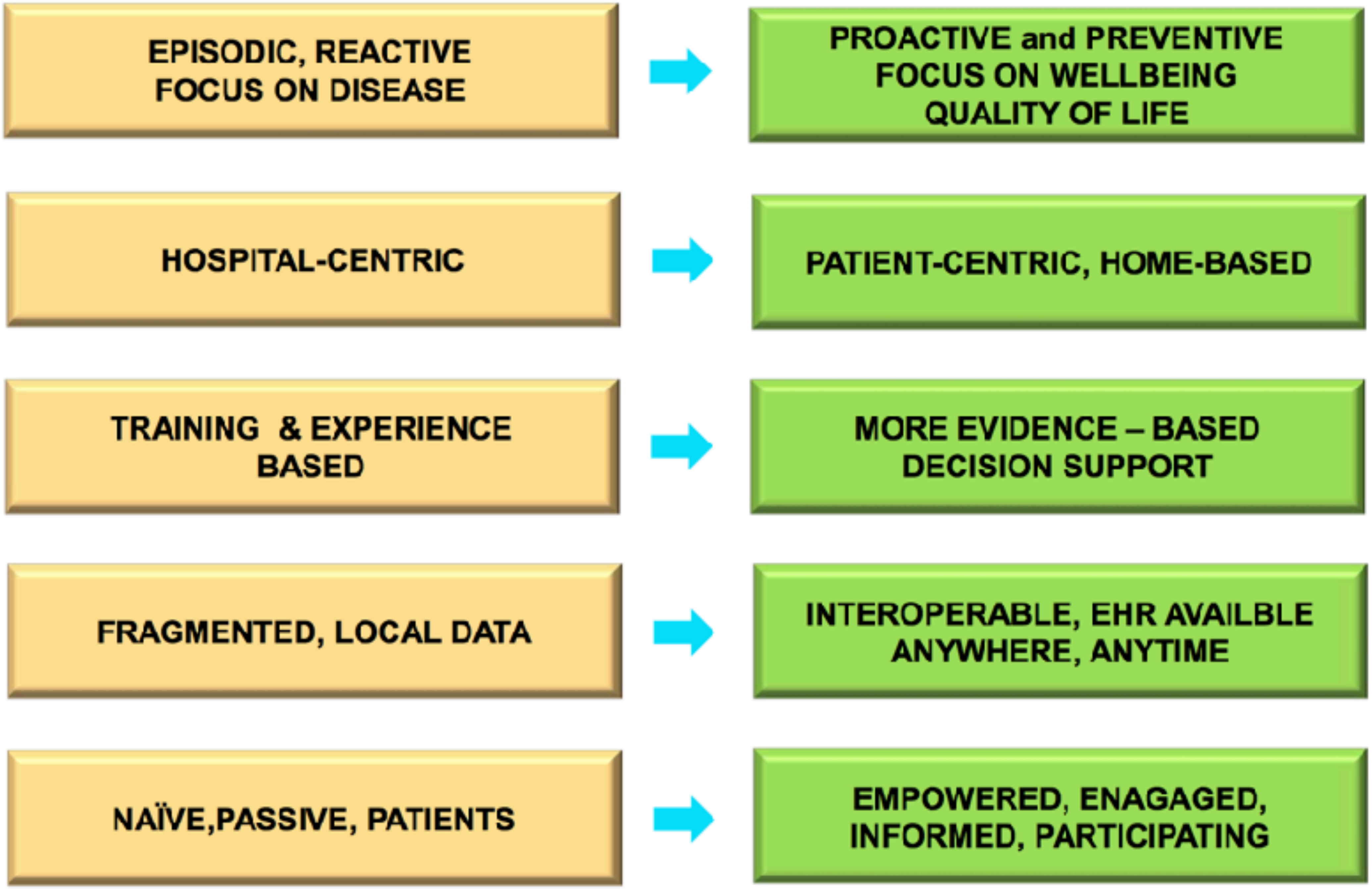
- Software developed must be released under an open source license or justify why not

NSF Smart and Connected Health (SCH) Program

- To fill in research gaps that exist in science and technology in support of health and wellness
- To advance the fields of health, wellness, improve quality of care and reduce cost by leveraging the fundamental science research

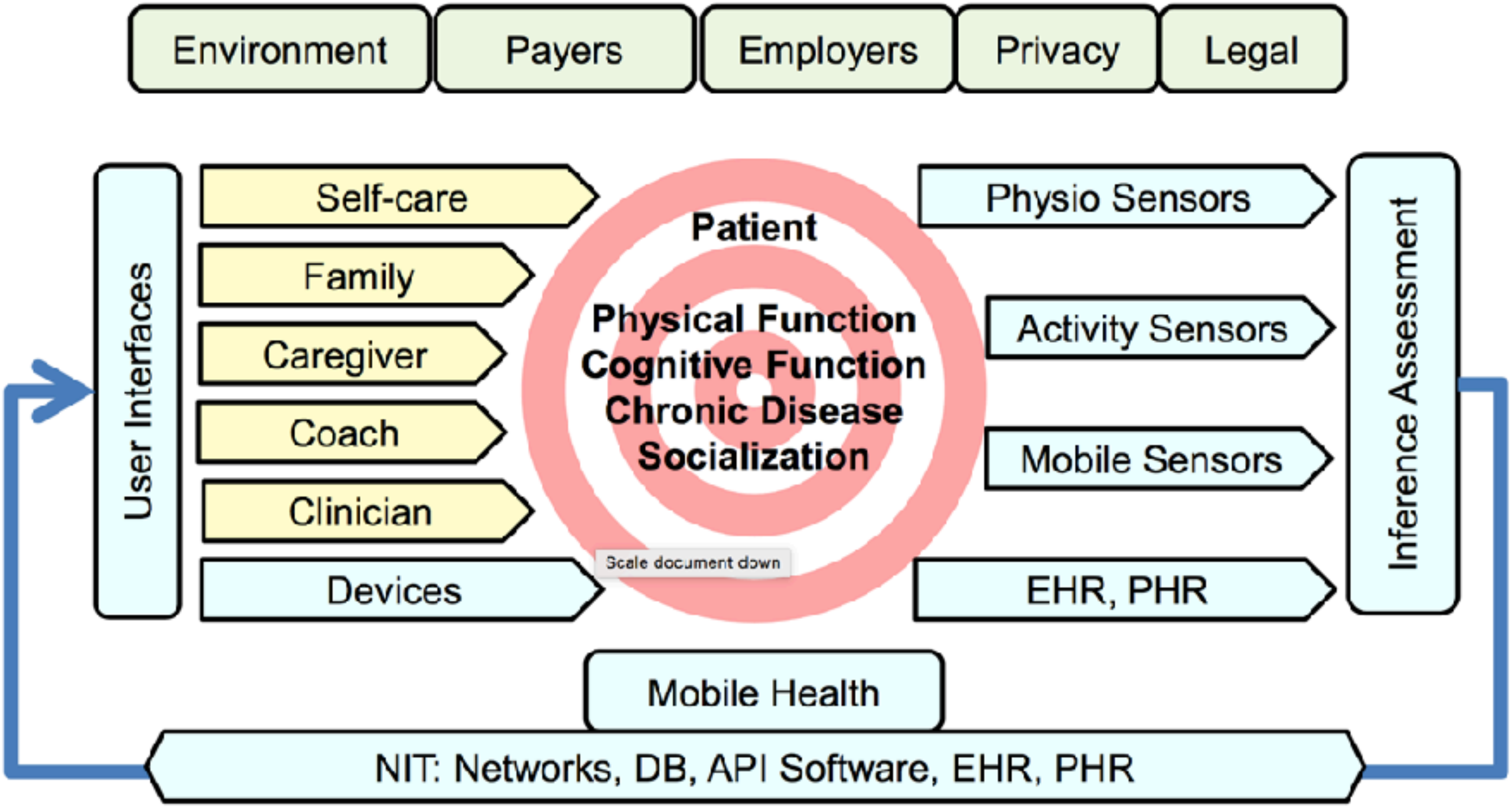


Traditional Medicine ⇨ SCH



EPISODIC, REACTIVE FOCUS ON DISEASE	➡	PROACTIVE and PREVENTIVE FOCUS ON WELLBEING QUALITY OF LIFE
HOSPITAL-CENTRIC	➡	PATIENT-CENTRIC, HOME-BASED
TRAINING & EXPERIENCE BASED	➡	MORE EVIDENCE – BASED DECISION SUPPORT
FRAGMENTED, LOCAL DATA	➡	INTEROPERABLE, EHR AVAILBLE ANYWHERE, ANYTIME
NAÏVE,PASSIVE, PATIENTS	➡	EMPOWERED, ENAGAGED, INFORMED, PARTICIPATING

Patient-Centered Framework



Environment

Payers

Employers

Privacy

Legal

User Interfaces

Self-care

Family

Caregiver

Coach

Clinician

Devices

Patient

Physical Function

Cognitive Function

Chronic Disease

Socialization

Scale document down

Physio Sensors

Activity Sensors

Mobile Sensors

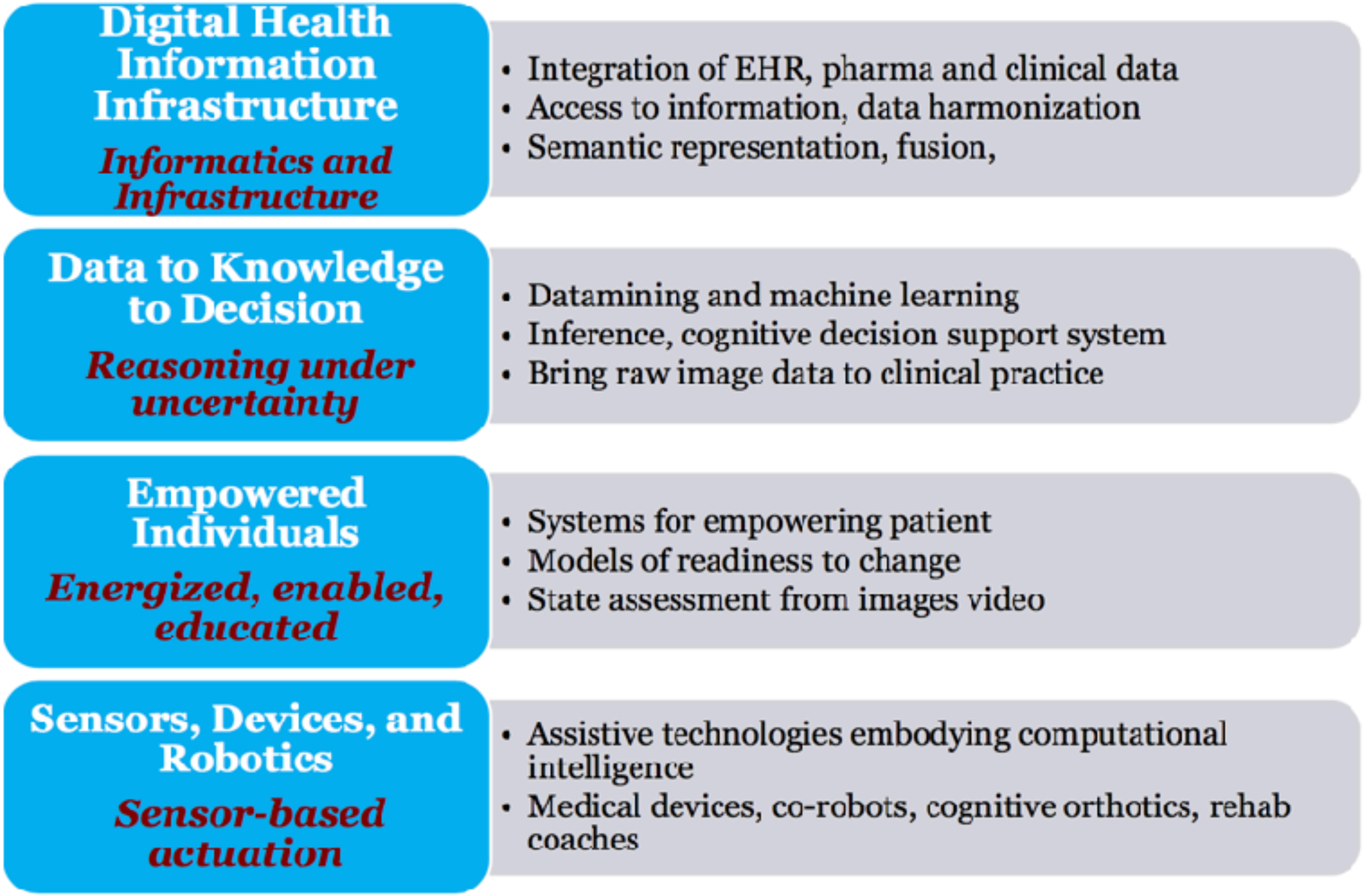
EHR, PHR

Inference Assessment

Mobile Health

NIT: Networks, DB, API Software, EHR, PHR

SCH Research Areas



Digital Health Information Infrastructure

Informatics and Infrastructure

- Integration of EHR, pharma and clinical data
- Access to information, data harmonization
- Semantic representation, fusion,

Data to Knowledge to Decision

Reasoning under uncertainty

- Datamining and machine learning
- Inference, cognitive decision support system
- Bring raw image data to clinical practice

Empowered Individuals

Energized, enabled, educated

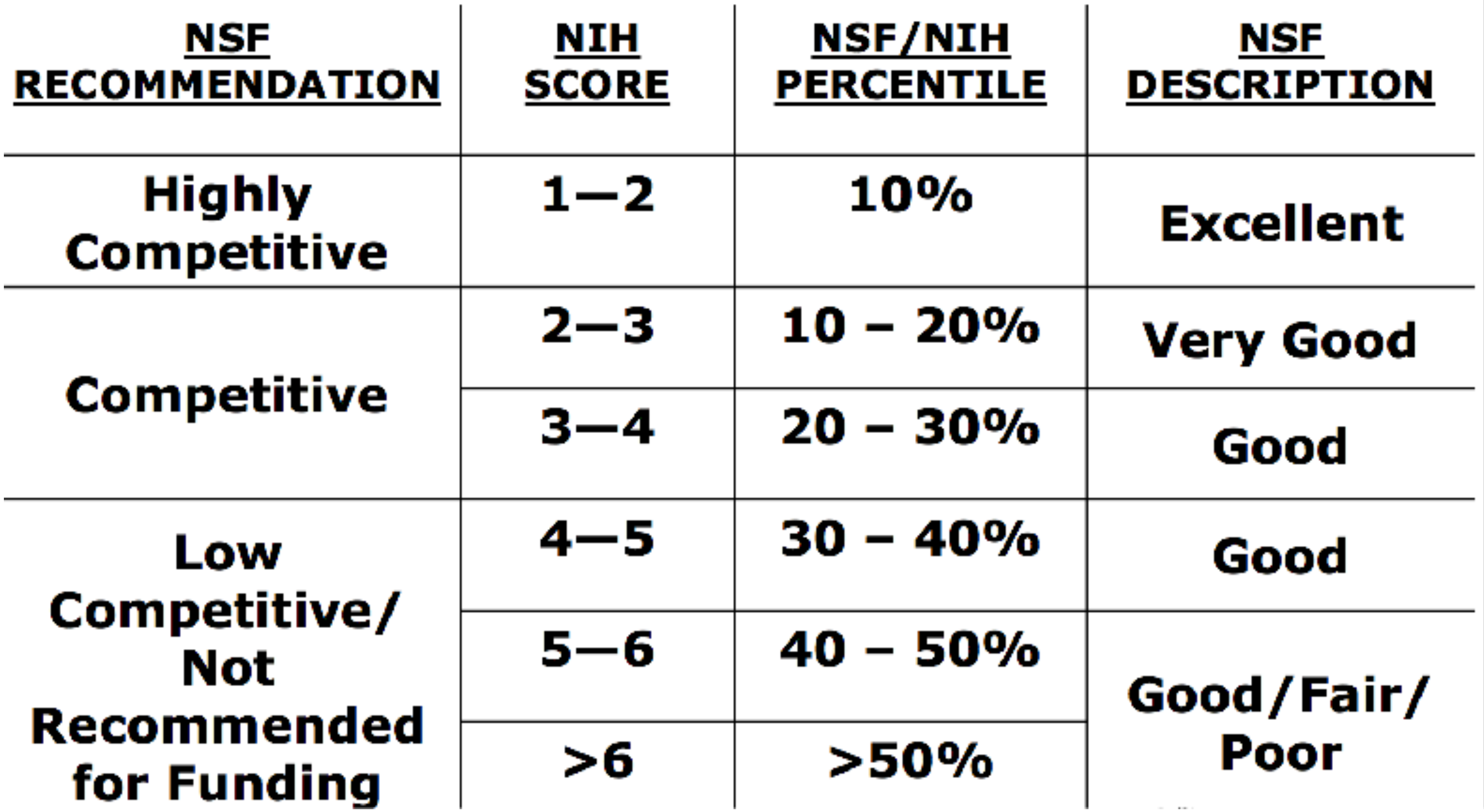
- Systems for empowering patient
- Models of readiness to change
- State assessment from images video

Sensors, Devices, and Robotics

Sensor-based actuation

- Assistive technologies embodying computational intelligence
- Medical devices, co-robots, cognitive orthotics, rehab coaches

NSF v NIH Review Scores



NSF RECOMMENDATION	NIH SCORE	NSF/NIH PERCENTILE	NSF DESCRIPTION
Highly Competitive	1—2	10%	Excellent
Competitive	2—3	10 – 20%	Very Good
	3—4	20 – 30%	Good
Low Competitive/ Not Recommended for Funding	4—5	30 – 40%	Good
	5—6	40 – 50%	Good/Fair/ Poor
	>6	>50%	

Thank you