Implementing Security without Inhibiting Research: Mission Impossible?

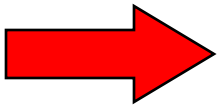
(http://www.esp.org/briite/meetings)

Robert J. Robbins

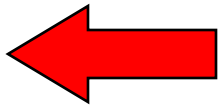
rrobbins@fhcrc.org

(206) 667 4778

Implementing Security without Inhibiting Research: Mission Impossible?



(http://www.esp.org/briite/meetings)



Robert J. Robbins

rrobbins@fhcrc.org

(206) 667 4778

Impossible?

Maybe not.

But it is very hard.

The challenge is real, yet we all need to figure out how to implement some kind of solution anyway.

The challenge is real, yet we all need to figure out how to implement some kind of solution anyway.

And, we had better be prepared to replace our solution with a better solution every few years for the next decade.

The Problem

- Culture clash between research and security.

The Problem

- Culture clash between research and security.
- Work occurs within decentralized organizations.

The Problem

- Culture clash between research and security.
- Work occurs within decentralized organizations.
- Work occurs across institutional boundaries.

The Problem

- Culture clash between research and security.
- Work occurs within decentralized organizations.
- Work occurs across institutional boundaries.
- Problem keeps changing.

The Problem

- Culture clash between research and security.
- Work occurs within decentralized organizations.
- Work occurs across institutional boundaries.
- Problem keeps changing.
- Rules keep changing.

The Problem

- Culture clash between research and security.
- Work occurs within decentralized organizations.
- Work occurs across institutional boundaries.
- Problem keeps changing.
- Rules keep changing.
- Solution keeps changing.

The Problem

- Culture clash between research and security.
- Work occurs within decentralized organizations.
- Work occurs across institutional boundaries.
- Problem keeps changing.
- Rules keep changing.
- Solution keeps changing.
- Human-subjects work is especially challenging.

Culture Clash

Culture Clash

RESEARCH	SECURITY
open	closed

Culture Clash

RESEARCH	SECURITY
open	closed
opportunistic	planned

Culture Clash

RESEARCH	SECURITY
open	closed
opportunistic	planned
creative	structured

Culture Clash

RESEARCH	SECURITY
open	closed
opportunistic	planned
creative	structured
challenge authority	respect authority

Culture Clash

RESEARCH	SECURITY
open	closed
opportunistic	planned
creative	structured
challenge authority	respect authority
one-off mentality	process driven
. . .	. . .

Decentralized Organizations

Decentralized Organizations

Would this work in your organization:

Decentralized Organizations

Would this work in your organization:

Your convenience is no reason for me to sacrifice the security of my network…

Decentralized Organizations

Would this work in your organization:

> Your convenience is no reason for me to sacrifice the security of my network…

But it does work in the military, where this quote originates.

True Story

Conversation between network administrator (N) and faculty member (F):

True Story

Conversation between network administrator (N) and faculty member (F):

N: These changes will improve the security of our network.

True Story

Conversation between network administrator (N) and faculty member (F):

N: These changes will improve the security of our network.

F: But they will make it impossible for my lab to carry out its research.

True Story

Conversation between network administrator (N) and faculty member (F):

N: These changes will improve the security of our network.

F: But they will make it impossible for my lab to carry out its research.

N: With a little effort you should be able to find a work-around.

True Story

Conversation between network administrator (N) and faculty member (F):

N: These changes will improve the security of our network.

F: But they will make it impossible for my lab to carry out its research.

N: With a little effort you should be able to find a work-around.

F: My staff and I have already devoted substantial effort to the problem and there is no work-around for us. However, we have determined that a relatively minor change in your security plan would meet your security needs while still allowing us to carry out our research.

True Story

Conversation between network administrator (N) and faculty member (F):

N: These changes will improve the security of our network.

F: But they will make it impossible for my lab to carry out its research.

N: With a little effort you should be able to find a work-around.

F: My staff and I have already devoted substantial effort to the problem and there is no work-around for us. However, we have determined that a relatively minor change in your security plan would meet your security needs while still allowing us to carry out our research.

N: What do you know about network security?

You're just an end user.

True Story

Conversation between network administrator (N) and faculty member (F):

Yes, but this end user also had a Nobel Prize and about two attractive job offers a month from other institutions.

problem and there is no work-around for us. However, we have determined that a relatively minor change in your security plan would meet your security needs while still allowing us to carry out our research.

N: What do you know about network security?

You're just an end user.

True Story

POP

QUIZ

True Story

The most likely outcome was:

True Story

The most likely outcome was:

1. The researcher totally changed his research program to meet the new security standards, or . . .

True Story

The most likely outcome was:

1. The researcher totally changed his research program to meet the new security standards, or . . .

2. The network administrator found himself with the opportunity to spend more time with his family.

Work Spans Institutional Boundaries

Work Spans Institutions

Much biomedical research is now conducted by teams of collaborators, often spanning multiple institutions.

Research that starts at one institution segues into multi-institutional work as students graduate, post-docs move on, and other changes occur.

Work Spans Institutions

Research often is accomplished by INFORMAL teams of workers, spanning multiple organizations.

These teams dynamically come into existence to meet a research need, then disband.

Work Spans Institutions

Portions of tens (or hundreds) of such teams exist at any one time within any research organization.

These teams are often not based on any formal relationships between the home institutions of the researchers.

Work Spans Institutions

Delivering high quality security across such teams either involves:

a proliferation of accounts across institutions, or

a security system designed for a totally decentralized federation

Work Spans Institutions

Delivering high quality security across

No currently available security system is designed to meet the needs of a totally decentralized federation.

Problem Keeps Changing

Changes in Problem Scope

Achieving security of research systems:

Changes in Problem Scope

Achieving security of research systems:

- within labs

Changes in Problem Scope

Achieving security of research systems:

- within labs
- across labs

Changes in Problem Scope

Achieving security of research systems:

- within labs
- across labs
- across departments

Changes in Problem Scope

Achieving security of research systems:

- within labs
- across labs
- across departments
- across campuses

Changes in Problem Scope

Achieving security of research systems:

- within labs
- across labs
- across departments
- across campuses
- across institutions

Changes in Problem Scope

Achieving security of research systems:

- within labs
- across labs
- across departments
- across campuses
- across institutions
- across state boundaries

Changes in Problem Scope

Achieving security of research systems:

- within labs
- across labs
- across departments
- across campuses
- across institutions
- across state boundaries
- across national boundaries

Changes in Problem Domain

New problems keep arising:

- financial system
- confidential data on lost laptops
- web site break-ins
- student music downloads
- termination policies
- HIPAA
- . . .

Changes in Logical Status

Some change is so profound that jokes become reality.

Changes in Logical Status

Some change is so profound that jokes become reality.

Sarcastic comment:

DNA is inherently identifiable. Pretty soon we'll have to start putting deliberate errors into DNA sequences before we can share them…

Changes in Logi

Some change is so profound that jokes become reality.

Sarcastic comment:

DNA is inherently identifiable. Pretty soon we'll have to start putting deliberate errors into DNA sequences before we can share them…

Recent article in Science

POLICY FORUM

ETHICS

Identifiability in Genomic Research

Genomic data are unique to the individual and must be managed with care to maintain public trust.

William W. Lowrance and Francis S. Collins

Genomic research can now readily generate data that cover significant portions of the human genome at levels of detail unique to individuals. Data can now be categorized with respect to disease-related genes and linked to clinical, family, and social data. Identifiability, the potential for such data to be associated with specific individuals, is therefore a pivotal concern. Research, health care, police, military, and other DNA and genotype reference collections are growing. Members of the public and its leaders worry about risks of erroneous or malicious identity disclosure and consequent embarrassment; legal or financial ramifications; stigmatization; and/or discrimination for insurance, employment, promotion, or loans.

If the data are considered identifiable, they may be covered by informational or genetic privacy laws, with implications for consent and other rights. They may be covered by human-subjects regulations, with implications for oversight. Controlled, conditional release may be required for the data as opposed to open public release. These can all be obstacles to the conduct of health-related research.

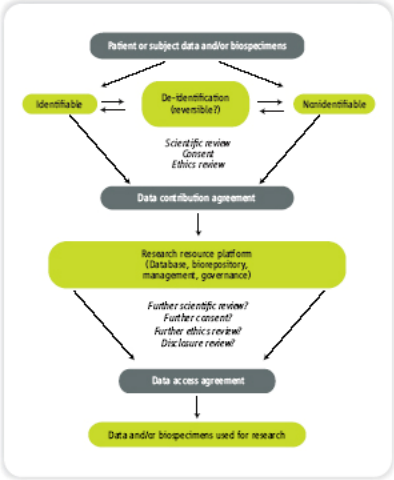
In the United States, personal data used in health care and/or research are protected by the Common Rule on Protection of Human Subjects (*1*), and the Privacy Rule under the Health Information Portability and Accountability Act (HIPAA) (*2*–*4*). They are also protected by state and other federal laws and regulations. In the European Union (*5*), informational privacy is protected by national laws that implement the Data Protection Directive, such as the U.K. Data Protection Act (1998). Most other countries have similar laws.

How these laws apply specifically, and how adequate they are in the genomic research arena, is not entirely clear. Protection of privacy was among the issues examined by the National Institutes of Health (NIH) in a recent public consultation (*6*).

New Modes of Data Flow

Until recently, most genomic research used data and biospecimens obtained fairly directly, from the data subjects themselves or clinical repositories or specialized research collections. This will continue, as it has many advantages. But now, in efforts to increase the range and quantity of data, large-scale research platforms are being built that assemble, organize, and store data, and sometimes biospecimens, and then distribute these to researchers (see figure). The advantages of such platforms, in addition to scale, are that they can be a robust staging-point for screening data quality, fostering uniformity of data format, and facilitating analysis. Some platforms accumulate data directly (as the Framingham Heart Study does); others assemble them from a variety of sources (as The Cancer Genome Atlas, the Genetic Association Information Network, and the Wellcome Trust Case Control Consortium do and U.K. Biobank will) (*7*). Among the design and governance issues are whether and how to de-identify the data and at what stages to conduct scientific and ethics review.

These new data flows, genomewide analyses, and novel arrangements such as the Informed Cohort scheme recently proposed by Kohane *et al.* (*8*) are relatively uncharted territory with respect to human subjects and privacy considerations. Precedent doesn't provide sufficient guidance. For example, the Human Genome and HapMap Projects have genotyped DNA from only a few hundred carefully selected people who prospectively consented to the analysis and to open publication after thorough explanation, discussion, and community consultation. The projects have been scrutinized closely all along. But when the data relate to more people (by orders of magnitude) or to retrospective analysis of biospecimens, then for pragmatic reasons such painstaking selection, consent negotiation, and scrutiny can't generally be achieved.



Identifiability and Identifiers

Identifiability ranges from overtly identifiable, to potentially identifiable by deduction, to absolutely unidentifiable. The concept isn't simple, as evidenced by the European Commission's publication of an elaborate "Opinion on the concept of personal data" in June 2007, 12 years after passage of the Data Protection Directive (*9*).

In legal regimens, indirect identifiability is as important as direct. For instance, the HIPAA Privacy Rule applies to "information that identifies an individual; or with respect to which there is a reasonable basis to believe the information can be used to identify the individual" (Sec. 160.103). Similarly, the U.K. Data Protection Act applies to "data which relate to a living individual who can be identified—(a) from those data, or (b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller" [Sec. I.1-(1)]. If

W. W. Lowrance is a consultant in health research policy and ethics, 72 rue de St. Jean, CH-1201 Geneva, Switzerland; e-mail: lowrance@iprolink.ch. F. S. Collins is director, U.S. National Human Genome Research Institute, Bethesda, MD 20892–2152 USA; e-mail: francisc@mail.nih.gov

Changes in Logi

Page 2:

Tactics for de-identifying genomic data:

data aren't identifiable they shouldn't be considered "personal," and a variety of rights and obligations that apply to personal data may not be relevant.

Three sorts of identifying factors can be distinguished: demographic or administrative tags (e.g., name, social security number, e-mail address, hospital name, postal code); overt descriptors (e.g., gender, eye color, height, blood type, scars, asthma); and indirect clues (e.g., medication use, number of children, spouse's occupation, circumstances of emergency-room admission). Whether particular bits of data alone or in combination should be considered sufficient to identify a person is a matter of judgment. Much may depend on whether partial identifiers can be linked with identified or identifiable data in public or other databases.

The HIPAA Privacy Rule illustrates the practical challenges. For data to be considered adequately de-identified and therefore not subject to its provisions, a number of descriptors, which it lists, must be absent [Sec. 164.514(B)(2)] (*7*). The list contains identifiers that are linked fairly directly to name and address, such as medical record numbers and hospital discharge dates. Knowing a few elements on the list may or may not allow identification, and even knowing a person-unique fact such as social security number allows identification only if it can be traced to the person through some other source.

Identifying Through Genomic Data

Matching against reference genotype. The number of DNA markers such as single-nucleotide polymorphisms (SNPs) that are needed to uniquely identify a single person is small; Lin *et al*. estimate that only 30 to 80 SNPs could be sufficient (*10*). Thus, such data can be used, with high certitude, to confirm that two samples come from the same person; whether this can identify anybody in the usual sense depends on whether the reference data are personally identified.

Collections that can be used for matching continue to grow. Identified biospecimens from millions of people are held by criminal justice systems and armed services (*11,12*). Biospecimens and a growing number of genomic analyses are held by health-care, public health, and health research institutions. To be clear, the risk is not that a match might be found but that a de-identified data set will become linkable to a specific person because the matched data set contains personal identifiers.

Linking to nongenetic databases. A second route to identifying genotyped subjects is deduction by linking and then matching genotype-plus-associated data (such as gender, age, or disease being studied) with data in health-care, administrative, criminal, disaster response, or other databases (*10,13,14*). There is no shortage of public and commercial databases about people's lives, especially in the United States. If the nongenetic data are overtly identified, the task is straightforward. Even if such data are not fully identified, inferential narrowing-down may be possible. Statisticians have many techniques for identifying data subjects from partial data (*15,16*).

Profiling from genomic data. A number of physical attributes can now be inferred from DNA analysis, such as gender, blood type, approximate skin pigmentation, and manifestations of Mendelian disorders. Reliability of predictions will likely increase regarding height or other aspects of skeletal build, hair color and texture, eye color, and even some craniofacial features. Soon many chronic disease susceptibilities will be predictable and, before long, some behavioral tendencies will be. In 5 to 10 years, many attributes will be profilable.

Tactics for De-identifying Genomic Data

Limiting the proportion of genome released. The first option is to release only limited segments of genomes, such as sequence traces or a few variants, along with minimum necessary phenotypic or other data. But "how much" is sufficient for identifying, by any route, depends on the region and extent of genome covered, the density of mapping, the rarity of variants, the degree of linkage disequilibrium, and other factors (*17*). This makes it difficult to develop general guidance on how much to expose publicly.

Many projects do limit the portion of genome they release, especially if the release is unrestricted. Precautions can be taken, such as releasing sequence traces in such a separated manner that no individual's data can be reassembled by overlaps. But releasing too-few SNPs or too-short snippets of sequence may thwart research.

Statistically degrading data. This is possible, for example, by lumping all purines and all pyrimidines. Unfortunately, the occurrence of a T instead of a C in one data cell can mean the difference between disease and health. So for many lines of genomic research, degrading data degrades usefulness.

Sequestering identifiers via key-coding (reversibly de-identifying) (*7*). This is the method most widely used in health research. Administrative or other overt identifiers are separated from data, but a link is maintained between them via an arbitrary numerical key-code (*18*). Held securely and separately, the key allows reassociation of substantive data with identifiers if necessary. The key and responsibility for its use can be delegated to a trusted party; its use can be guided by agreed-upon criteria and subjected to oversight.

Provision of Access to Data

Open versus controlled release. A cultural habit of rapid, open release of genomic data has been pursued by the involved scientists and institutions since the beginning of the Human Genome Project (*19–20*). There is no question about the research advantages of such principles and policies. But almost certainly, the principles will have to be modified now for databases that include extensive genotypic information, to heighten the protection of identifiability (*21*).

Open data release, as with deposition in a publicly accessible Web site, is acceptable only if either: (i) the data are for all practical purposes not identifiable; or (ii) consent to the release is ethically legitimate and is granted by the data subjects, or the necessity for consent is waived by a competent ethics body. Most projects now take three precautionary steps: sequestering the standard identifiers via key-coding; performing disclosure risk-reduction (such as by rounding birth date to year of birth); and providing access to the de-identified data under conditional terms.

Terms of agreements. Data-access agreements (alternatively called "certifications" or "use agreements") cover many matters. Legally they amount to contracts, and they may have to be entered into by researchers' institutions as well as the researchers.

Agreements may set limitations on purposes and uses, allowed users, or other matters covered by consent, either for the whole dataset or for particular data-subjects, and may address how data will be released. They should refer to physical, organizational, and information technology security. They may specify who will be responsible for de-identifying data and may cover key-coding, safeguarding of the key, and criteria for use of the key. They should always state that researchers will make no attempt to identify nonidentified data. They should restrict unauthorized passing on of data and should extend the chain of custody and the accompanying obligations if data are passed on. They may address linking, if linking to other datasets is contemplated that might increase identifiability. Invariably they require that derived data on individuals be protected at least as carefully as the data being provided. They may make access contingent on Institutional Review Board or other ethics committee approval and may specify the stage(s) at

Changes in Logi

Page 2:

Tactics for de-identifying genomic data:

Statistically degrading data. This is possible, for example, by lumping all purines and all pyrimidines. Unfortunately, the occurrence of a T instead of a C in one data cell can mean the difference between disease and health. So for many lines of genomic research, degrading data degrades usefulness.

POLICYFORUM

data aren't identifiable they shouldn't be considered "personal," and a variety of rights and obligations that apply to personal data may not be relevant.

Three sorts of identifying factors can be distinguished: demographic or administrative tags (e.g., name, social security number, e-mail address, hospital name, postal code); overt descriptors (e.g., gender, eye color, height, blood type, scars, asthma); and indirect clues (e.g., medication use, number of children, spouse's occupation, circumstances of emergency-room admission). Whether particular bits of data alone or in combination should be considered sufficient to identify a person is a matter of judgment. Much may depend on whether partial identifiers can be linked with identified or identifiable data in public or other databases.

The HIPAA Privacy Rule illustrates the practical challenges. For data to be considered adequately de-identified and therefore not subject to its provisions, a number of descriptors, which it lists, must be absent [Sec. 164.514(B)(2)] (*7*). The list contains identifiers that are linked fairly directly to name and address, such as medical record numbers or hospital discharge dates. Knowing a few elements on the list may or may not allow identification, and even knowing a person-unique fact such as social security number allows identification only if it can be traced to the person through some other source.

Identifying Through Genomic Data

Matching against reference genotype. The number of DNA markers such as single-nucleotide polymorphisms (SNPs) that are needed to uniquely identify a single person is small; Lin *et al.* estimate that only 30 to 80 SNPs could be sufficient (*10*). Thus, such data can be used, with high certitude, to confirm that two samples come from the same person; whether this can identify anybody in the usual sense depends on whether the reference data are personally identified.

Collections that can be used for matching continue to grow. Identified biospecimens from millions of people are held by criminal justice systems and armed services (*11,12*). Biospecimens and a growing number of genomic analyses are held by health-care, public health, and health research institutions. To be clear, the risk is not that a match might be found but that a de-identified data set will become linkable to a specific person because the matched data set contains personal identifiers.

Linking to nongenetic databases. A second route to identifying genotyped subjects is deduction by linking and then matching genotype-plus-associated data (such as gender, age, or disease being studied) with data in health-care, administrative, criminal, disaster response, or other databases (*10,13,14*). There is no shortage of public and commercial databases about people's lives, especially in the United States. If the nongenetic data are overtly identified, the task is straightforward. Even if such data are not fully identified, inferential narrowing-down may be possible. Statisticians have many techniques for identifying data subjects from partial data (*15,16*).

Profiling from genomic data. A number of physical attributes can now be inferred from DNA analysis, such as gender, blood type, approximate skin pigmentation, and manifestations of Mendelian disorders. Reliability of predictions will likely increase regarding height or other aspects of skeletal build, hair color and texture, eye color, and even some craniofacial features. Soon many chronic disease susceptibilities will be predictable and, before long, some behavioral tendencies will be. In 5 to 10 years, many attributes will be profilable.

Tactics for De-identifying Genomic Data

Limiting the proportion of genome released. The first option is to release only limited segments of genomes, such as sequence traces or a few variants, along with minimum necessary phenotypic or other data. But "how much" is sufficient for identifying, by any route, depends on the region and extent of genome covered, the density of mapping, the rarity of variants, the degree of linkage disequilibrium, and other factors (*17*). This makes it difficult to develop general guidance on how much to expose publicly.

Many projects do limit the portion of genome they release, especially if the release is unrestricted. Precautions can be taken, such as releasing sequence traces in such a separated manner that no individual's data can be reassembled by overlaps. But releasing too-few SNPs or too-short snippets of sequence may thwart research.

Statistically degrading data. This is possible, for example, by lumping all purines and all pyrimidines. Unfortunately, the occurrence of a T instead of a C in one data cell can mean the difference between disease and health. So for many lines of genomic research, degrading data degrades usefulness.

Sequestering identifiers via key-coding (reversibly de-identifying) (*7*). This is the method most widely used in health research. Administrative or other overt identifiers are separated from data, but a link is maintained between them via an arbitrary numerical key-code (*18*). Held securely and separately, the key allows reassociation of substantive data with identifiers if necessary. The key and responsibility for its use can be delegated to a trusted party; its use can be guided by agreed-upon criteria and subjected to oversight.

Provision of Access to Data

Open versus controlled release. A cultural habit of rapid, open release of genomic data has been pursued by the involved scientists and institutions since the beginning of the Human Genome Project (*19–20*). There is no question about the research advantages of such principles and policies. But almost certainly, the principles will have to be modified now for databases that include extensive genotypic information, to heighten the protection of identifiability (*21*).

Open data release, as with deposition in a publicly accessible Web site, is acceptable only if either: (i) the data are for all practical purposes not identifiable; or (ii) consent to the release is ethically legitimate and is granted by the data subjects, or the necessity for consent is waived by a competent ethics body. Most projects now take three precautionary steps: sequestering the standard identifiers via key-coding; performing disclosure risk-reduction (such as by rounding birth date to year of birth); and providing access to the de-identified data under conditional terms.

Terms of agreements. Data-access agreements (alternatively called "certifications" or "use agreements") cover many matters. Legally they amount to contracts, and they may have to be entered into by researchers' institutions as well as the researchers.

Agreements may set limitations on purposes and uses, allowed users, or other matters covered by consent, either for the whole dataset or for particular data-subjects, and may address how data will be released. They should refer to physical, organizational, and information technology security. They may specify who will be responsible for de-identifying data and may cover key-coding, safeguarding of the key, and criteria for use of the key. They should always state that researchers will make no attempt to identify nonidentified data. They should restrict unauthorized passing on of data and should extend the chain of custody and the accompanying obligations if data are passed on. They may address linking, if linking to other datasets is contemplated that might increase identifiability. Invariably they require that derived data on individuals be protected at least as carefully as the data being provided. They may make access contingent on Institutional Review Board or other ethics committee approval and may specify the stage(s) at

Changes in Logi

Page 2:

Tactics for de-identifying

When reality starts to resemble parody, things are getting too complex for comfort.

POLICYFORUM

data aren't identifiable they shouldn't be considered "personal," and a variety of rights and obligations that apply to personal data may not be relevant.

Three sorts of identifying factors can be distinguished: demographic or administrative tags (e.g., name, social security number, e-mail address, hospital name, postal code); overt descriptors (e.g., gender, eye color, height, blood type, scars, asthma); and indirect clues (e.g., medication use, number of children, spouse's occupation, circumstances of emergency-room admission). Whether particular bits of data alone or in combination should be considered sufficient to identify a person is a matter of judgment. Much may depend on whether partial identifiers can be linked with identified or identifiable data in

type-plus-associated data (such as gender, age, or disease being studied) with data in health-care, administrative, criminal, disaster response, or other databases (*10, 13, 14*). There is no shortage of public and commercial databases about people's lives, especially in the United States. If the nongenetic data are overtly identified, the task is straightforward. Even if such data are not fully identified, inferential narrowing-down may be possible. Statisticians have many techniques for identifying data subjects from partial data (*15, 16*).

Profiling from genomic data. A number of physical attributes can now be inferred from DNA analysis, such as gender, blood type, approximate skin pigmentation, and manifestations of Mendelian disorders. Reliability of predictions will likely increase

key allows reassociation of substantive data with identifiers if necessary. The key and responsibility for its use can be delegated to a trusted party; its use can be guided by agreed-upon criteria and subjected to oversight.

Provision of Access to Data

Open versus controlled release. A cultural habit of rapid, open release of genomic data has been pursued by the involved scientists and institutions since the beginning of the Human Genome Project (*19–20*). There is no question about the research advantages of such principles and policies. But almost certainly, the principles will have to be modified now for databases that include extensive genotypic information, to heighten the protection of identifiability (*21*).

tions. To be clear, the risk is not that a match might be found but that a de-identified data set will become linkable to a specific person because the matched data set contains personal identifiers.

Linking to nongenetic databases. A second route to identifying genotyped subjects is deduction by linking and then matching geno-

ing data degrades usefulness.

Sequestering identifiers via key-coding (reversibly de-identifying) (7). This is the method most widely used in health research. Administrative or other overt identifiers are separated from data, but a link is maintained between them via an arbitrary numerical key-code (*18*). Held securely and separately, the

address linking, if linking to other datasets is contemplated that might increase identifiability. Invariably they require that derived data on individuals be protected at least as carefully as the data being provided. They may make access contingent on Institutional Review Board or other ethics committee approval and may specify the stage(s) at

Rules Keep Changing

Rules Keep Changing

HIPAA

Sarbanes Oxley

News stories of lost laptops

Internal audit departments

Non-research savvy auditors

Engaged boards of directors

. . .

Solution Keeps Changing

Solution Keeps Changing

We need comprehensive support for implementing security in a totally decentralized federation.

No such solution exists.

So we keep implementing the approximation du jour (or maybe de jure).

Human Subjects Research

What is Human Subjects Research?

Certain activities are obviously human subjects research, appropriately covered by IRB rules and procedures.

But, where are the limits? What activities are covered and what are not?

- Effect of food additive?
- Price of popcorn in movie theaters?
- Production of recipe book?

HSR Criteria

Project:

MBA student wants to interview theater managers about price of popcorn at different times and for different features.

Problem:

Should this activity be considered research involving human subjects covered by 45 CFR part 46?

Answer:

HSR Criteria

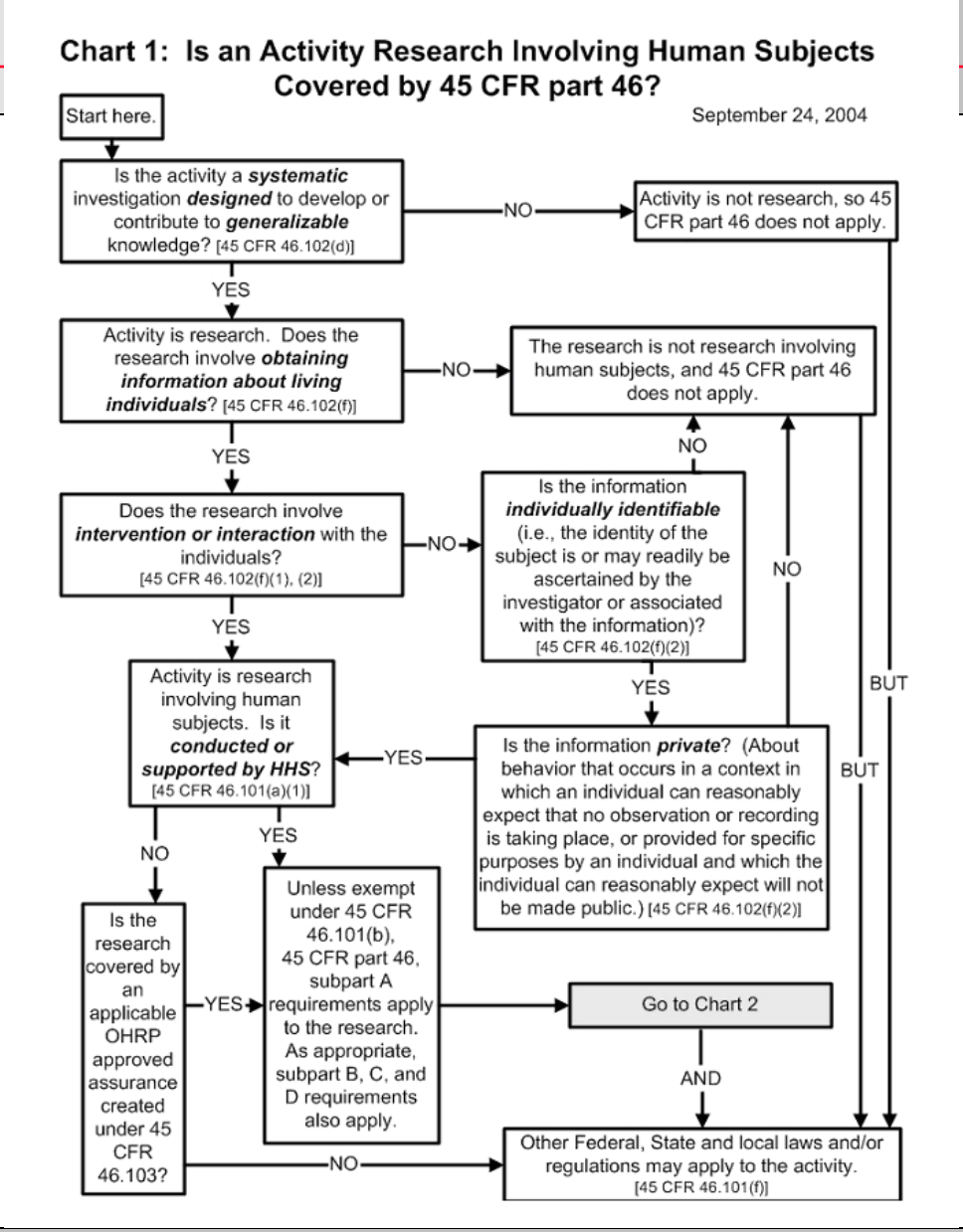
Project:

MBA student wants to interview theater managers about price of popcorn at different times and for different features.

Problem:

Should this activity be considered research involving human subjects covered by 45 CFR part 46?

Answer:



HSR Criteria

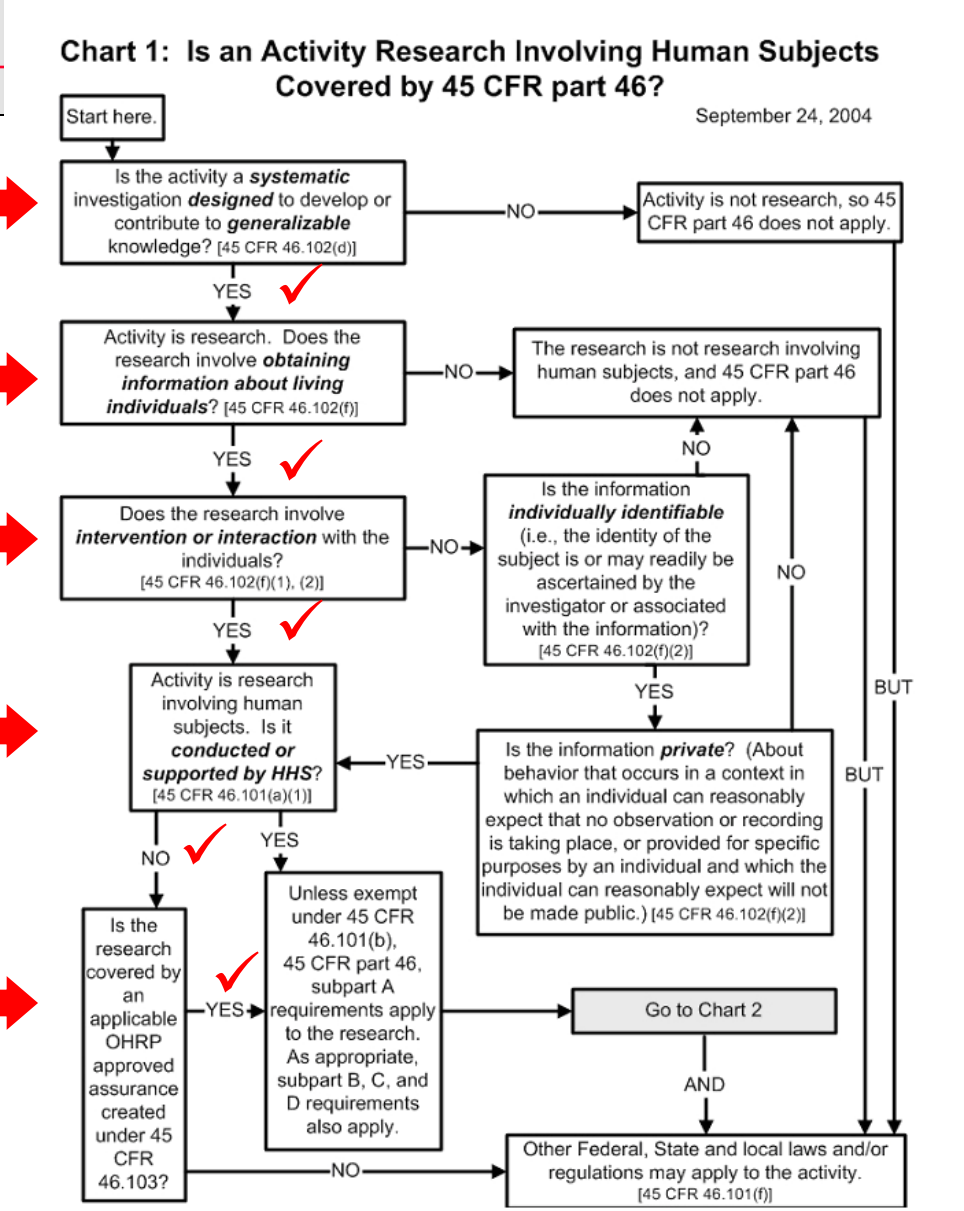
Project:

MBA student wants to interview theater managers about price of popcorn at different times and for different features.

Problem:

Should this activity be considered research involving human subjects covered by 45 CFR part 46?

Answer:



HSR Criteria

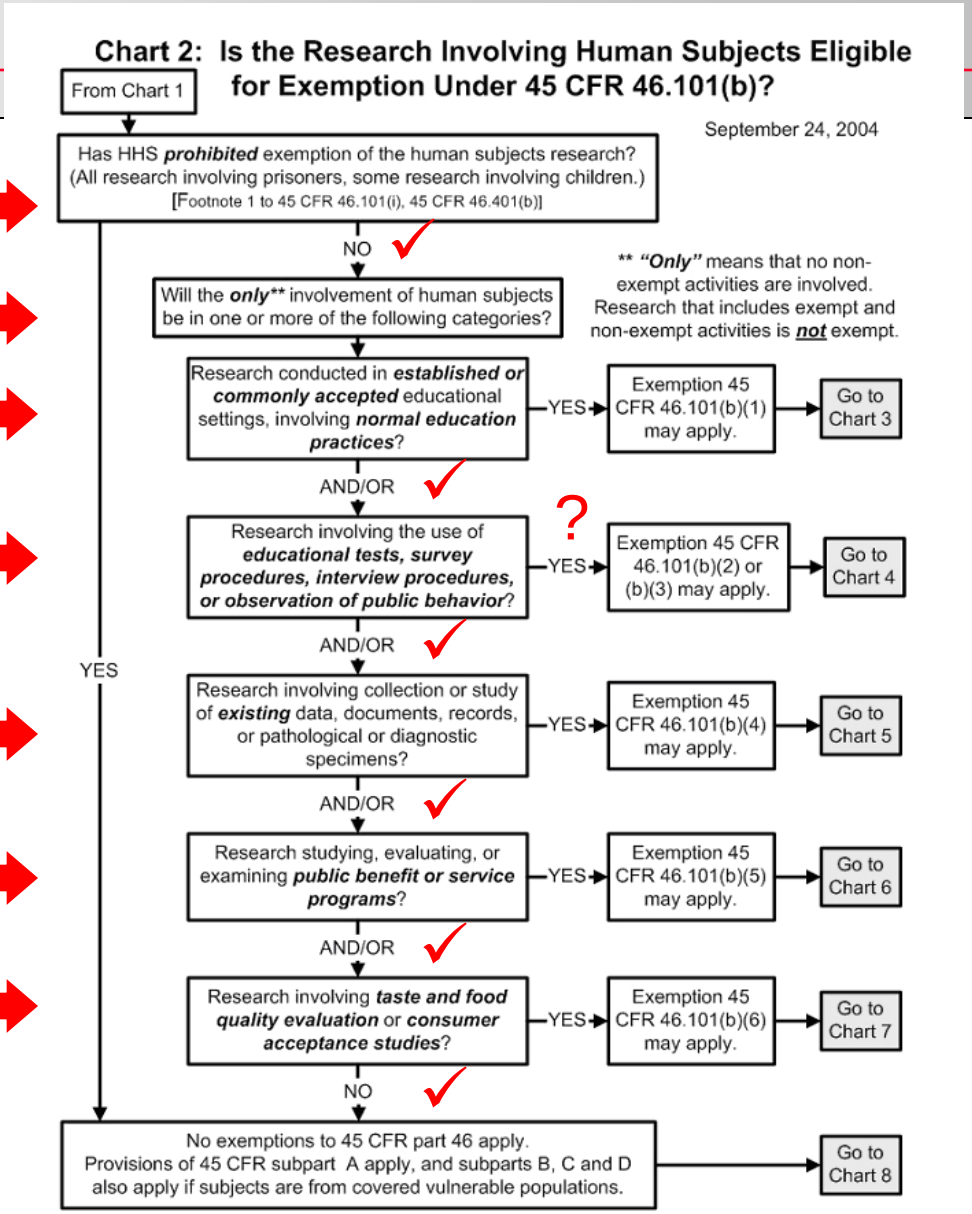
Project:

MBA student wants to interview theater managers about price of popcorn at different times and for different features.

Problem:

Should this activity be considered research involving human subjects covered by 45 CFR part 46?

Answer:



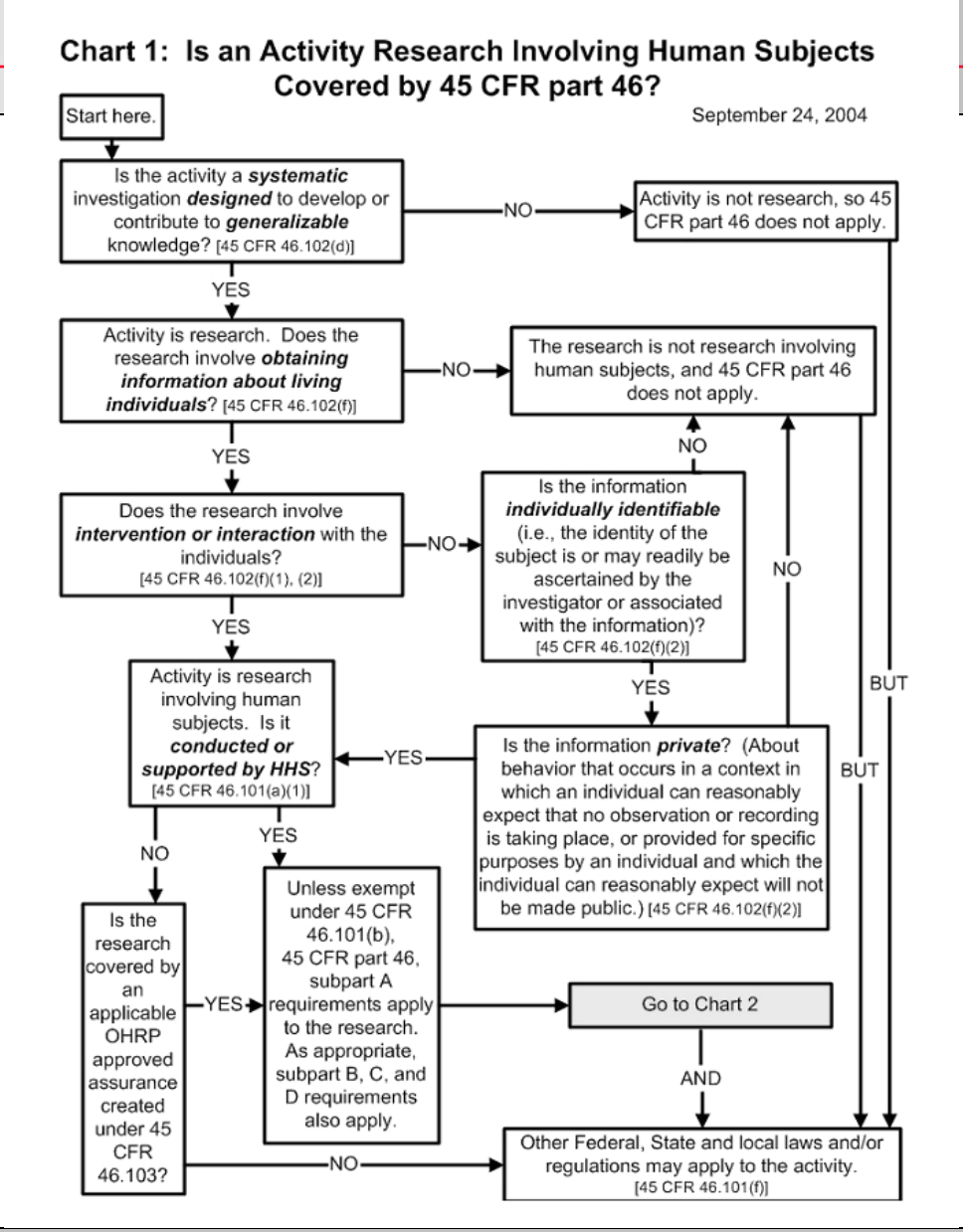
HSR Criteria

Project:

Research team wants to interview IRB heads, security officers, other institutional leaders to determine the policy requirements governing the deployment of multi-site digital security systems.

Problem:

Should this activity be considered research involving human subjects covered by 45 CFR part 46?



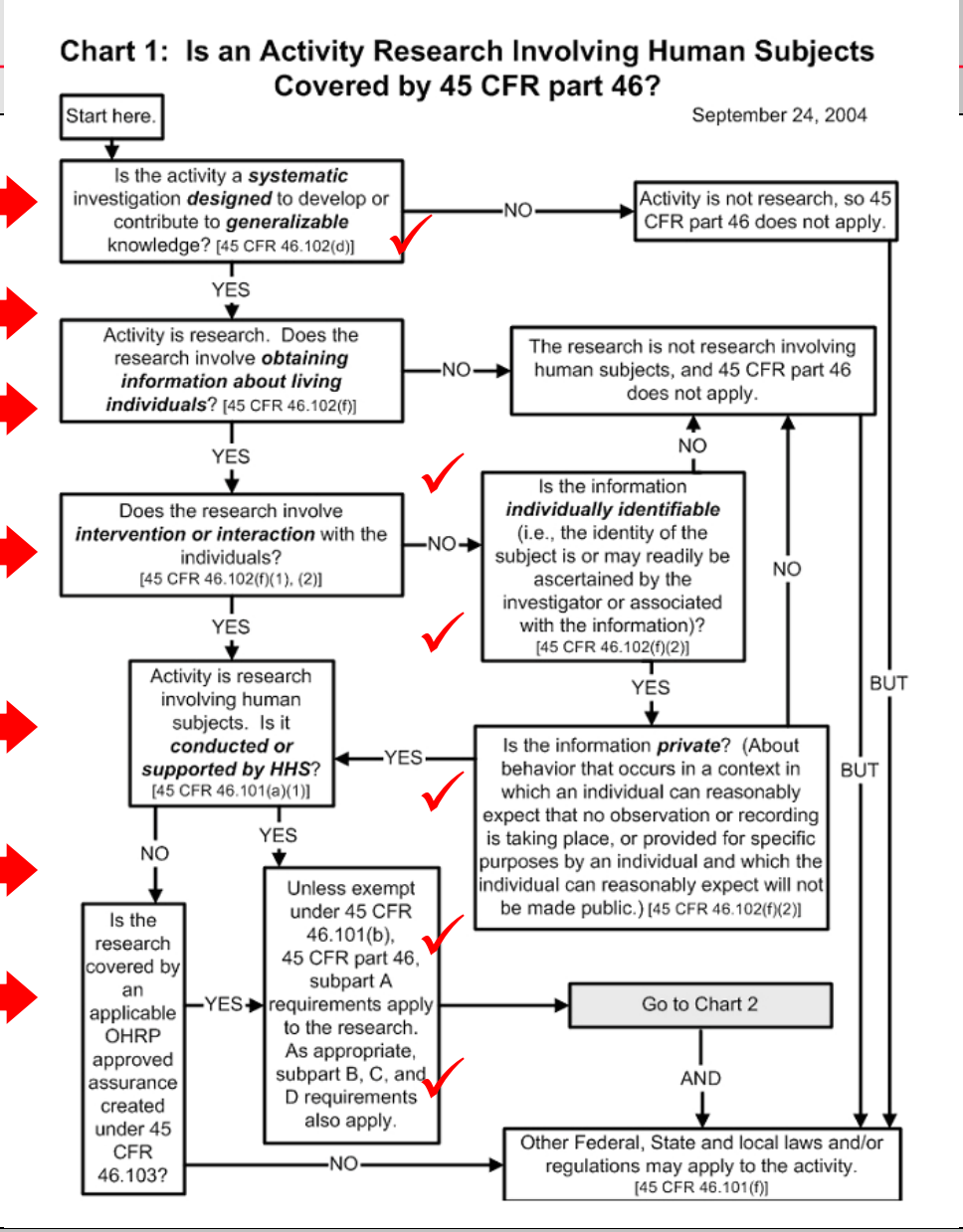
HSR Criteria

Project:

Research team wants to interview IRB heads, security officers, other institutional leaders to determine the policy requirements governing the deployment of multi-site digital security systems.

Problem:

Should this activity be considered research involving human subjects covered by 45 CFR part 46?



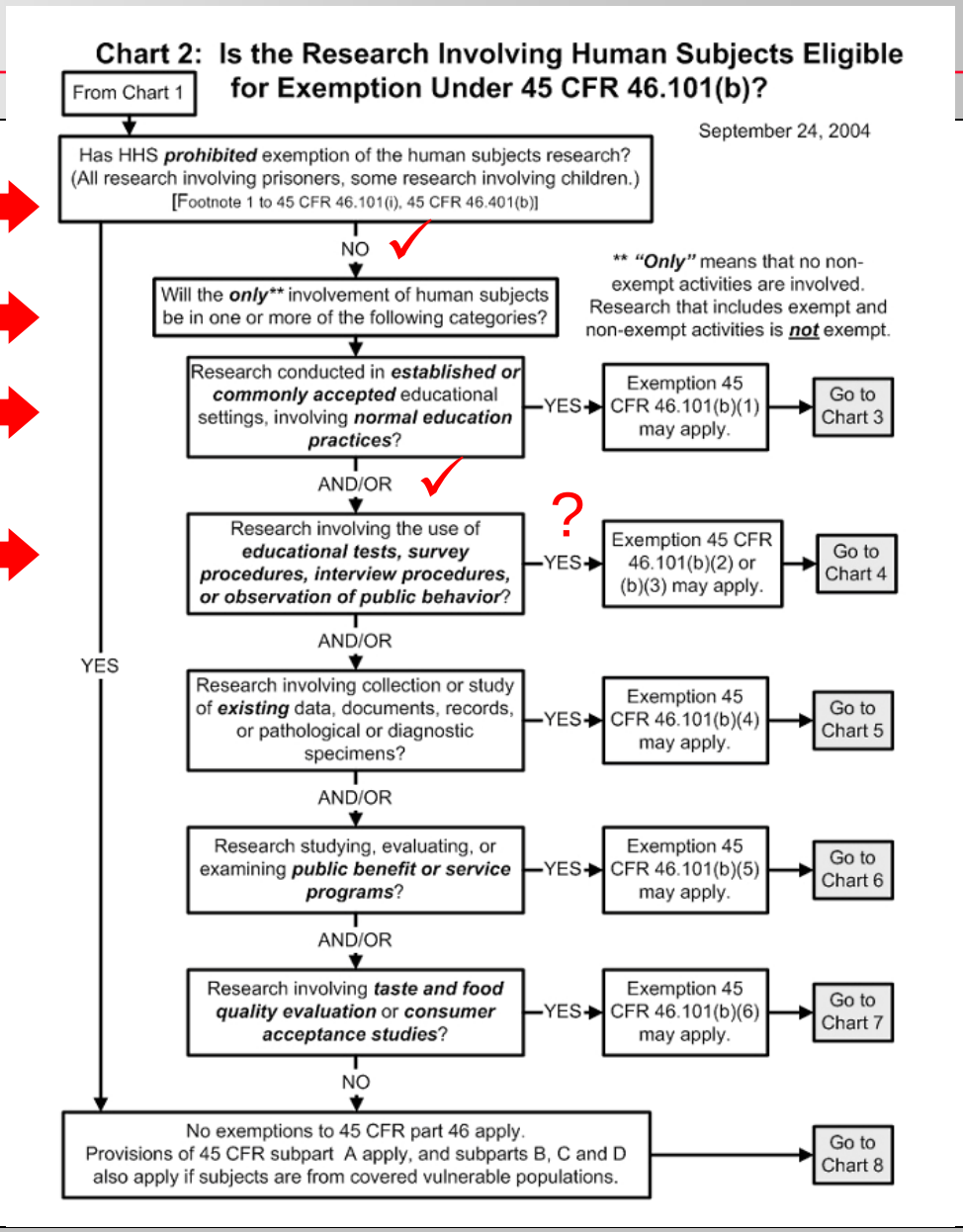
HSR Criteria

Project:

Research team wants to interview IRB heads, security officers, other institutional leaders to determine the policy requirements governing the deployment of multi-site digital security systems.

Problem:

Should this activity be considered research involving human subjects covered by 45 CFR part 46?



END