CP1 CP2 CP3 CP4 CP5 CP6 CP7 CP8 CP9

CT1 CT2 CT3 **CT4** CT5 CT6 CT7

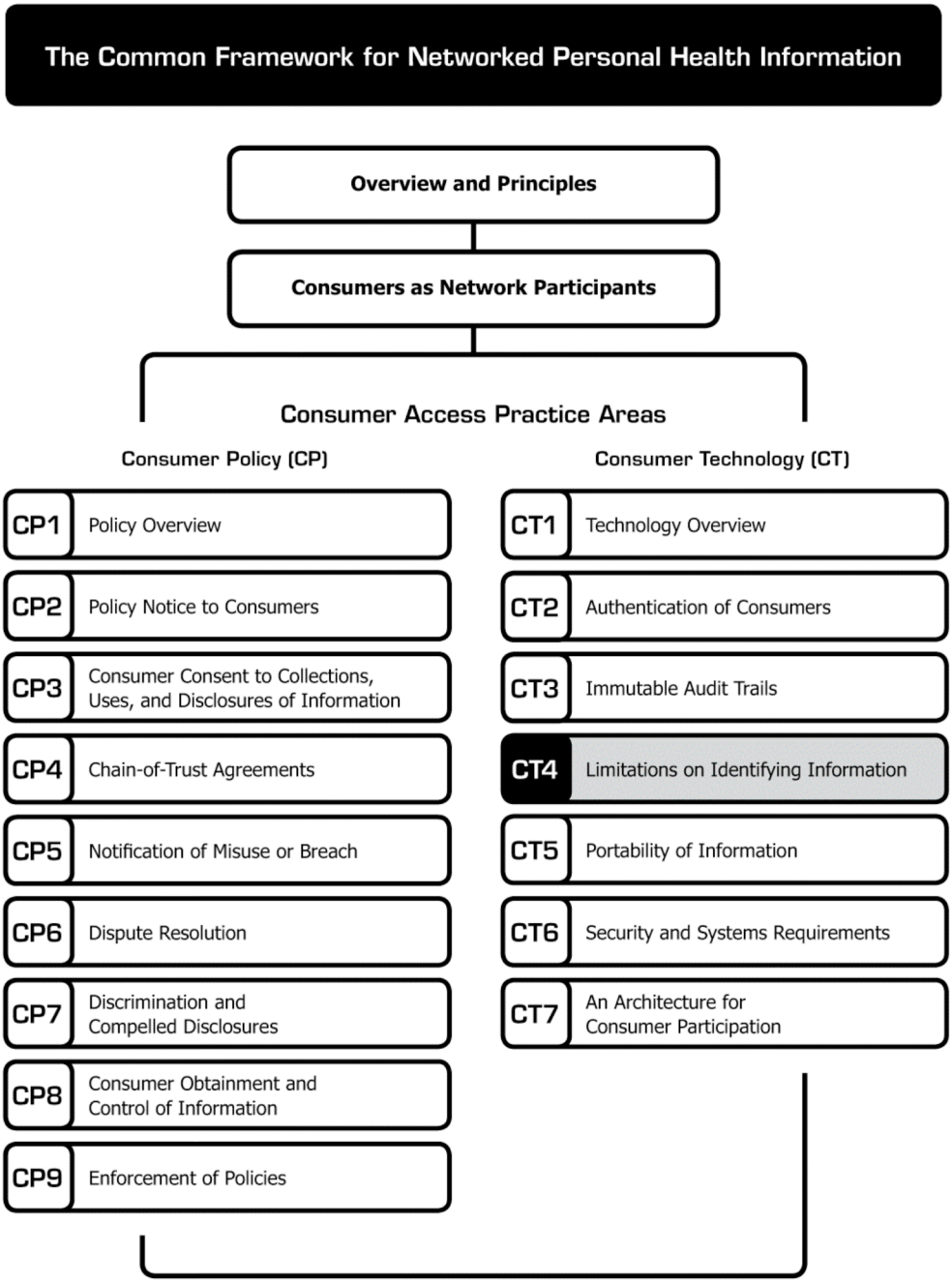
Limitations on Identifying Information

Limitations on Identifying Information

The document you are reading is part of the ***Connecting for Health*** *Common Framework for Networked Personal Health Information*, which is available in full and in its most current version at http://www.connectingforhealth.org/.

This framework proposes a set of practices that, when taken together, encourage appropriate handling of personal health information as it flows to and from personal health records (PHRs) and similar applications or supporting services.

As of June 2008, the Common Framework included the following published components:



Limitations on Identifying Information*

This practice area addresses the following **Connecting for Health** Core Principles for a Networked Environment*:

2. Purpose specification

3. Collection limitation and data minimization

4. Use limitation

7. Security safeguards and controls

* "The Architecture for Privacy in a Networked Health Information Environment," **Connecting for Health**, June 2006. Available at: http://www.connectingforhealth.org/commonframework/docs/P1_CFH_Architecture.pdf.

There are significant risks if business partners of Consumer Access Services are permitted to combine data with other databases to identify individuals or create a more complete profile of the consumer's health. Such practices have the potential to create unauthorized third party relationships of which the consumer may be completely unaware. Chain-of-trust agreements should prohibit this type of activity. (*See* ***CP4: Chain-of-Trust Agreements***.) In addition, Consumer Access Services can further protect consumers — as well as themselves — by ensuring that the identifying information they expose to partners is the minimal amount necessary. For example, in some cases, a Consumer Access Service could share a consumer's age, but not date of birth, with a third party because age is less potentially revealing of identity than a specific date of birth.

In the Internet Age, information is increasingly difficult to classify as "identified" or "de-identified," particularly as it is copied, exchanged, or recombined with other information. With rapidly evolving technologies and databases, it is more appropriate to describe a spectrum of "identifiability," rather than a binary classification of information as identifiable or not. The question could then become not whether de-identified information *might be* made re-identifiable, but rather *which entities* would be able to re-identify the information, *how much effort* they would have to expend, and *what limits* are placed on their doing so.

HIPAA Regulations (45 C.F.R. § 164.514) provide standards for de-identification, including a list of 18 "identifier" data elements that must be stripped out in order for a limited data set to qualify as "de-identified."[1]

The Privacy Rule also allows a second way to de-identify information by having a qualified statistician determine, using generally accepted statistical principles and methods, that the risk is very small that the information could be used, alone or in combination with other reasonably available information, by the anticipated recipient to identify the subject of the information. The qualified statistician must document the methods and results of the analysis that justify such a determination.

This HIPAA regulation remains a reasonable industry standard for defining information as "de-identified" in many circumstances today. However, it may not be fully identity-protective in some contexts, such as when applied to very small subsets of populations, or with the ever-increasing amounts of "partially identifying information" gathered in electronic environments. (*See* ***Appendix A*** *for more on partially identifying information*.) This reality will necessitate frequent monitoring of risk by policymakers in both the public and private sectors.

* **Connecting for Health** thanks Matt Kavanagh, independent contractor, and Josh Lemieux, Markle Foundation, for drafting this paper.

[1] Accessed online on January 2, 2008, at the following URL: http://www.hhs.gov/ocr/combinedregtext.pdf.

Recommended Practice:

Consumer Access Services should limit disclosures of identifying data to only those data that are necessary to perform the specified function(s) that the recipient is authorized to perform.

Care should be taken to limit the release or exposure of information that can be directly or indirectly tied to an individual, including electronic identifiers such as IP address, cookies, and web beacons.

Any release of such indirectly or directly identifying information should be consistent with all nine **Connecting for Health** Privacy Principles and all of the Practice Areas of this Common Framework, particularly specification of purpose, limitation of use to only specified purpose, and no unauthorized combining of data to create a more complete profile of individuals.

Appendix A: "Partially Identifying Data"

In today's web environment, much of what consumers do is recorded and tracked by the sites they visit. Even when consumers are not logged in, various pieces of information are collected about them. These little bits of data are often not personally identifying at the time and point of collection. But in some cases, these bits of information can be combined with other bits of information to build a more complete profile of each user. When enough information is collected and combined, it can be used to identify individuals. Hence, we call this information "partially identifying." Examples include cookies, web beacons, and even search keywords.

For illustration, "persistent cookies" are little pieces of text deposited in the web browsers of consumers by the web sites they visit. In a similar way that a ticket from the dry cleaner lets the proprietor link the customer out front with the right clothes held in the back, cookies contain lookup information that lets a web site link a user to other information held about him in a database, such as preferences, search history, or checkout items for purchase on the site.

When the consumer returns to a web site at a later time, persistent cookies such as these can tell the web browser to display the user name, show whatever the user has specified to appear on the site's homepage, allow for access to previously entered search queries, or display information about items the user had previously added to a shopping cart.

When search engine companies collect user search query history "anonymously" (i.e., not tied to a specific user identity), the partially identifiable information the user provides can be identifying in and of itself if a consumer searches for information about her name, address, telephone number, and/or personal identifiers. When this information is combined with additional search queries that detail the user's interests, hobbies, health conditions, etc., a very personal picture can be elicited quite easily. For example, America Online in the summer of 2006 released 20 million "de-identified" search queries of more than 650,000 of its users with the intention to help researchers design better search engines. AOL initially claimed the search data had been made anonymous by replacing each search query's associated AOL username with a different unique user ID. But for those search queries that included identifying information along with personal interests, not only were some users' identities revealed, but also intimate details about their personal lives.

Another example of unintentional identification occurred as a result of an airline's practice of printing customers' frequent-flyer numbers on boarding passes in addition to names and seat numbers. An investigative reporter doing a story on identity theft retrieved a passenger's discarded ticket stub and used the information to purchase another ticket from the same airline (in this case from British Airways). In doing so, the reporter was granted access to additional pieces of the passenger's identity, including "passport number, date of birth, and nationality."

The above cases, in which partially identifying information is used by external parties to identify an individual, occurred outside of contractual agreements. However, they do illustrate how the identifiability of information can change over time.

Acknowledgements

This framework is a collaborative work of the **Connecting for Health** Work Group on Consumer Access Policies for Networked Personal Health Information — a public-private collaboration operated and financed by the Markle Foundation. **Connecting for Health** thanks Work Group Chair David Lansky, PhD, Pacific Business Group on Health, for leading the consensus development process for this framework, and Josh Lemieux, Markle Foundation, for drafting and editing the documents. We thank Carol Diamond, MD, MPH, managing director at the Markle Foundation, for developing the conceptual structure for this approach to networked personal health information. We particularly thank the members of the Work Group, whose affiliations are listed below for identification purposes only, for reviewing several drafts of these documents and improving them invaluably each time.

Jim Dempsey, JD, Center for Democracy and Technology; Janlori Goldman, JD, Health Privacy Project and Columbia University School of Public Health; Joy Pritts, JD, Center on Medical Record Rights and Privacy, Health Policy Institute, Georgetown University; and Marcy Wilder, JD, Hogan & Hartson LLP, made important contributions to the policy framework. Matt Kavanagh, independent contractor, and Clay Shirky, New York University Graduate Interactive Telecommunications Program, made important contributions to the technology framework. Stefaan Verhulst of Markle Foundation provided excellent research, and Jennifer De Pasquale and Michelle Maran of Markle contributed to this framework's final proofreading and production, respectively.

Connecting for Health Work Group on Consumer Access Policies for Networked Personal Health Information

Lead

David Lansky, PhD, Pacific Business Group on Health (Chair)

Staff

Matt Kavanagh, Independent Contractor
Josh Lemieux, Markle Foundation

Members

Wendy Angst, MHA, CapMed, A Division of Bio-Imaging Technologies, Inc.

Annette Bar-Cohen, MPH, National Breast Cancer Coalition

Jeremy Coote, InterComponentWare, Inc.

Maureen Costello, Ingenix

Diane Davies, MD, University of Minnesota

James Dempsey, JD, Center for Democracy and Technology

Stephen Downs, SM, Robert Wood Johnson Foundation

Joyce Dubow, AARP

Thomas Eberle, MD, Intel Corporation and Dossia

Lisa Fenichel, Health Care For All

Stefanie Fenton, Intuit, Inc.

Steven Findlay, Consumers Union

Mark Frisse, MD, MBA, MSc, Vanderbilt Center for Better Health

Gilles Frydman, Association of Cancer Online Resources (ACOR.org)

Melissa Goldstein, JD, School of Public Health and Health Services Department of Health Sciences, The George Washington University Medical Center

Philip T. Hagen, MD, Mayo Clinic Health Solutions

Robert Heyl, Aetna, Inc.

David Kibbe, MD, MBA, American Academy of Family Physicians

Jerry Lin, Google Health

Kathleen Mahan, MBA, SureScripts

Ken Majkowski, PharmD, RxHub, LLC

Philip Marshall MD, MPH, WebMD Health

Deven McGraw, Center for Democracy and Technology

Kim Nazi*, FACHE, U.S. Department of Veterans Affairs

Lee Partridge, National Partnership for Women and Families

George Peredy, MD, Kaiser Permanente HealthConnect

Joy Pritts, JD, Center on Medical Record Rights and Privacy, Health Policy Institute, Georgetown University

Scott Robertson, PharmD, Kaiser Permanente

Daniel Sands, MD, MPH, Cisco Systems, Inc.

Clay Shirky, New York University Graduate Interactive Telecommunications Program

Joel Slackman, BlueCross BlueShield Association

Anna Slomovic, PhD, Revolution Health

Cynthia Solomon, Follow Me

Ramesh Srinivasan, MedicAlert Foundation International

Michael Stokes, Microsoft Corporation

Susan Stuard, New York-Presbyterian Hospital

Paul Tang, MD, Palo Alto Medical Foundation/ Sutter Health

Jeanette Thornton, America's Health Insurance Plans

Frank Torres, JD, Microsoft Corporation

Tony Trenkle*, Centers for Medicare & Medicaid Services

Jonathan Wald, MD, Partners HealthCare System

James Walker, MD, FACP, Geisinger Health System

Marcy Wilder, JD, Hogan & Hartson LLP

Anna Wong, Medco Health Solutions, Inc.

Matthew Wynia, MD, MPH, CAPH, American Medical Association

Teresa Zayas-Caban, PhD*, Agency for Healthcare Research and Quality

**Note: State and Federal employees participate in the Personal Health Technology Council but make no endorsement.*