

CONTENT OR METADATA?

Categorisation of aspects of intercepted communications - GCHQ policy guidance.

POCs: ██████████ of GCHQ OPP-LEG ██████████@gchq.gov.uk) and

██████████ of GCHQ OPP-HQ ██████████@gchq.gov.uk).

Date last verified: 20 January 2010 by ██████████. Version 2.

Data type sorted by data class (column B; defined below)	Class (see below)	Content (c) or Metadata (m)	Note: For convenience, the terms "Metadata", "Events" and "Communications Data" are often used interchangeably, although only the last is defined in UK law (RIPA).
attachment to e-mail, eg routing diagram, picture, video	a	c	
e-mail address in the body of a message	a	c	
name of a file attached to an e-mail	a	c	
authentication data to a communications service: login ID, userid, password	ac	m	unless sent inside the body of a communication
an e-mail inside a message	c	c	
bulletin board posting	c	c	
chat-room discussions	c	c	
content of a voice call	c	c	
content of an e-mail	c	c	
cookie as a whole (some elements are listed separately below)	c	c	
DTMF data, as opposed to dialling	c	c	
keystroke logs	c	c	
search results	c	c	
search strings	c	c	
SMS or IM text	c	c	
video	c	c	
voice mail boxes	c	c	
web cam transmissions	c	c	
web forms filled in by people	c	c	
'to' 'from', 'cc', 'bcc', and 'fwd' lines within e-mail header	ca	m	
chat aliases, chat handles and other related or similar identifiers	ca	m	
e-mail address from a cookie sent to set up a communication channel	ca	m	
IMEI data	ca	m	
IMSI caller ID	ca	m	
IP addresses of the computers sending and receiving the message	ca	m	
machine ID extracted from cookies (eg Yahoo B cookies)	ca	m	

UK SECRET STRAP1
COMINT AUS/CAN/NZ/UK/US EYES ONLY ORCON

Data type sorted by data class (column B; defined below)	Class (see below)	Content (c) or Metadata (m)	Note: For convenience, the terms "Metadata", "Events" and "Communications Data" are often used interchangeably, although only the last is defined in UK law (RIPA).
personal IDs extracted from cookies, mail headers, chat sessions or buddy lists	ca	m	
telephone number	ca	m	
content summarization	cdd	cdd	
file type of an attachment to an email	cdd	cdd	
language or language fingerprinting	cdd	cdd	
speaker's gender	cdd	cdd	
speaker's ID	cdd	cdd	
e-mail subject line	ch	c	
dialling, signalling, routing, addressing or signalling information	d	m	
call charge records, including info about the time and length of a call	d	m	
Calling Line Identification, including numbers dialled	d	m	
details of routers and IP addresses that have handled the message	d	m	
DTMF dialling (not data)	d	m	
location of parties to a communication, not derived from content	d	m	
logs of visitors to chat rooms including how often they have visited/posted	d	m	
network management, eg billing, authentication or tracking of communicants	d	m	
start and finish time of an internet session or phone call	d	m	
status of chat sites, ie whether they are active and how many participants	d	m	
URLs up to and including the domain name	d	m	
session initiation protocols	d	m	
website registers including owner details; assume not UKUSA owned/registered	d	m	
creation of/access to draft message	m	m	N.B. the contents of a draft message are Content.
history of websites browsed (full URLs)	m	c	
buddy lists for web mail, instant messenger or social networking	m	m	
folders used to organise e-mails	m	m	
address books or contacts lists for web mail etc	m	c	
crypto keys	m	c	
password to internet or telephony services other than communications services	m	c	
URLs beyond the domain name, ie one that may include search terms	m	c	

Data type sorted by data class (column B; defined below)	Class (see below)	Content (c) or Metadata (m)	Note: For convenience, the terms "Metadata", "Events" and "Communications Data" are often used interchangeably, although only the last is defined in UK law (RIPA).
---	--------------------------	------------------------------------	--

Coding of data classes

a - attachment

ac - authentication of communications

c - content

ca - communications address

cdd - content derived data; some types in this class may be able to be treated as communications data due to low level of intrusion

ch - content header

d - traffic data including network management (excludes any such data sent inside the body of a communication, eg by a CSP)

m - miscellaneous; some types in this class may be able to be treated as communications data

bulk unselected: as taken from bearers without filtering or selection, save for national sensitivities