Data Intelligence

Corso di perfezionamento post laurea
«Intelligence e sicurezza nazionale»
AA: 2020-2021

http://www.disit.org/6574

Prof. Paolo Nesi
DISIT Lab
Distributed Data Intelligence and Technologies Lab
Distributed Systems and Internet Technologies Lab
Dipartimento di Ingegneria dell'Informazione
Università degli Studi di Firenze
Via S. Marta 3, 50139, Firenze, Italia
tel: +39-055-2758515, fax: +39-055-2758570
http://www.disit.dinfo.unifi.it
paolo.nesi@unifi.it

Le Sfide

- Internet delle Cose, droni, sensori, etc.
- Infrastrutture Critiche e Sistemi Cyber-Fisici
- Organizzazione, Fattore Umano e Ingegneria Sociale
- Social Network
- Smart City
- Componenti e Sistemi Hardware
- Cloud e Internet
- Biometria

Le azioni

- Sistemi Avanzati di Crittografia
- Protezione di Internet
- Protezione dell'Informazione
- Riduzione delle superfici di attacco
- Progettazione di Sistemi Informativi Resilienti
- Poligoni Virtuali per Esercitarsi sulla Sicurezza
- Investigazioni Digitali
- Intelligence e Big Data Analytics
- Condivisione delle Informazioni
- Metriche e Valutazione del Rischio e Resilienza

sommario

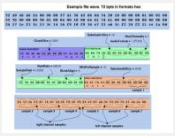
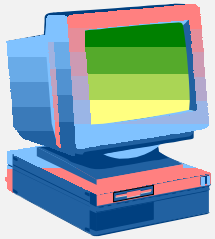
- *Dati e formati*
- *Comunicazione fra sistemi computerizzati*
- *Modelli di protezione e gestione*
- *Dati vs Metadati*
- *Social network e Smart City*
- *Motori di Ricerca, Crawling e NLP*
- *Data Mining*
- *Data Intelligence*

I dati e I formati

- Formati e informazioni
- Codifica
- Informatica ?? Informazione – automatica
- Volume dei dati
  - velocità dei flussi e computabilità:
  - static e real time
- Informazione strutturata
- entità vs relazioni: rappresentazione conoscenza
- tabelle vs reticoli
- XML

Device

DISIT Lab, Distributed Data Intelligence and Technologies
Distributed Systems and Internet Technologies
Department of Information Engineering (DINFO)
http://www.disit.dinfo.unifi.it














Programmazione

- È l'attività con cui si predispone l'elaboratore a **eseguire un *particolare insieme di azioni* su *particolari dati***, allo scopo di *risolvere un problema.*
- *Informatica: informazione - automatica*

Codifica dell'Informazione

In un calcolatore si possono (e vi è la necessità di) codificare ed elaborare vari tipi di informazioni che devono essere opportunamente codificati in modo che siano computabili:

- Numeri naturali (visto)
- Numeri reali in virgola fissa (visto)
- Numeri interi (da vedere in questa sessione)
  - In varie forme
- Rappresentazione dati: testi, visualizzazione
- …
- …

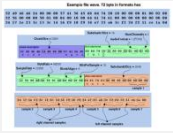


Tabella Bin, Decimale e Esadecimale

- Tabella da binario a decimale a esadecimale
- Rappresentazione VS codifica

Esadecimale	Binario	decimale
0	0000	0
1	0001	1
2	0010	2
3	0011	3
4	0100	4
5	0101	5
6	0110	6
7	0111	7
8	1000	8
9	1001	9
A	1010	10
B	1011	11
C	1100	12
D	1101	13
E	1110	14
F	1111	15

Dec	Hx	Oct	Char	Dec	Hx	Oct	Html	Chr	Dec	Hx	Oct	Html	Chr	Dec	Hx	Oct	Html	Chr
0	0	000	NUL (null)	32	20	040		Space	64	40	100	@	@	96	60	140	`	`
1	1	001	SOH (start of heading)	33	21	041	!	!	65	41	101	A	A	97	61	141	a	a
2	2	002	STX (start of text)	34	22	042	"	"	66	42	102	B	B	98	62	142	b	b
3	3	003	ETX (end of text)	35	23	043	#	#	67	43	103	C	C	99	63	143	c	c
4	4	004	EOT (end of transmission)	36	24	044	$	$	68	44	104	D	D	100	64	144	d	d
5	5	005	ENQ (enquiry)	37	25	045	%	%	69	45	105	E	E	101	65	145	e	e
6	6	006	ACK (acknowledge)	38	26	046	&	&	70	46	106	F	F	102	66	146	f	f
7	7	007	BEL (bell)	39	27	047	'	'	71	47	107	G	G	103	67	147	g	g
8	8	010	BS (backspace)	40	28	050	(	(	72	48	110	H	H	104	68	150	h	h
9	9	011	TAB (horizontal tab)	41	29	051	)	)	73	49	111	I	I	105	69	151	i	i
10	A	012	LF (NL line feed, new line)	42	2A	052	*	*	74	4A	112	J	J	106	6A	152	j	j
11	B	013	VT (vertical tab)	43	2B	053	+	+	75	4B	113	K	K	107	6B	153	k	k
12	C	014	FF (NP form feed, new page)	44	2C	054	,	,	76	4C	114	L	L	108	6C	154	l	l
13	D	015	CR (carriage return)	45	2D	055	-	-	77	4D	115	M	M	109	6D	155	m	m
14	E	016	SO (shift out)	46	2E	056	.	.	78	4E	116	N	N	110	6E	156	n	n
15	F	017	SI (shift in)	47	2F	057	/	/	79	4F	117	O	O	111	6F	157	o	o
16	10	020	DLE (data link escape)	48	30	060	0	0	80	50	120	P	P	112	70	160	p	p
17	11	021	DC1 (device control 1)	49	31	061	1	1	81	51	121	Q	Q	113	71	161	q	q
18	12	022	DC2 (device control 2)	50	32	062	2	2	82	52	122	R	R	114	72	162	r	r
19	13	023	DC3 (device control 3)	51	33	063	3	3	83	53	123	S	S	115	73	163	s	s
20	14	024	DC4 (device control 4)	52	34	064	4	4	84	54	124	T	T	116	74	164	t	t
21	15	025	NAK (negative acknowledge)	53	35	065	5	5	85	55	125	U	U	117	75	165	u	u
22	16	026	SYN (synchronous idle)	54	36	066	6	6	86	56	126	V	V	118	76	166	v	v
23	17	027	ETB (end of trans. block)	55	37	067	7	7	87	57	127	W	W	119	77	167	w	w
24	18	030	CAN (cancel)	56	38	070	8	8	88	58	130	X	X	120	78	170	x	x
25	19	031	EM (end of medium)	57	39	071	9	9	89	59	131	Y	Y	121	79	171	y	y
26	1A	032	SUB (substitute)	58	3A	072	:	:	90	5A	132	Z	Z	122	7A	172	z	z
27	1B	033	ESC (escape)	59	3B	073	;	;	91	5B	133	[	[	123	7B	173	{	{
28	1C	034	FS (file separator)	60	3C	074	<	<	92	5C	134	\	\	124	7C	174			|
29	1D	035	GS (group separator)	61	3D	075	=	=	93	5D	135	]	]	125	7D	175	}	}
30	1E	036	RS (record separator)	62	3E	076	>	>	94	5E	136	^	^	126	7E	176	~	~
31	1F	037	US (unit separator)	63	3F	077	?	?	95	5F	137	_	_	127	7F	177	&#127;	DEL

Source: www.asciitable.com

128	Ç	144	É	160	á	176	▒	193	┴	209	╤	225	ß	241	±
129	ü	145	æ	161	í	177	▒	194	┬	210	╥	226	Γ	242	≥
130	é	146	Æ	162	ó	178	▓	195	├	211	╙	227	π	243	≤
131	â	147	ô	163	ú	179	│	196	─	212	╘	228	Σ	244	⌠
132	ä	148	ö	164	ñ	180	┤	197	┼	213	╒	229	σ	245	⌡
133	à	149	ò	165	Ñ	181	╡	198	╞	214	╓	230	µ	246	÷
134	å	150	û	166	ª	182	╢	199	╟	215	╫	231	τ	247	≈
135	ç	151	ù	167	º	183	╖	200	╚	216	╪	232	Φ	248	°
136	ê	152	ÿ	168	¿	184	╕	201	╔	217	┘	233	Θ	249	∙
137	ë	153	Ö	169	⌐	185	╣	202	╩	218	┌	234	Ω	250	·
138	è	154	Ü	170	¬	186	║	203	╦	219	█	235	δ	251	√
139	ï	156	£	171	½	187	╗	204	╠	220	▄	236	∞	252	ⁿ
140	î	157	¥	172	¼	188	╝	205	═	221	▌	237	φ	253	²
141	ì	158	₧	173	¡	189	╜	206	╬	222	▐	238	ε	254	■
142	Ä	159	ƒ	174	«	190	╛	207	╧	223	▀	239	∩	255	
143	Å	192	└	175	»	191	┐	208	╨	224	α	240	≡		

Source: www.asciitable.com

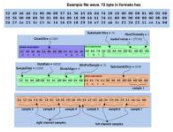
Conversioni con Esadecimali

Hexadecimal-Binary-Decimal Conversion

Hex Number	Binary Number	Decimal Digit 000X	Decimal Digit 00X0	Decimal Digit 0X00	Decimal Digit X000
0	0000	0	0	0	0
1	0001	1	16	256	4,096
2	0010	2	32	512	8,192
3	0011	3	48	768	12,288
4	0100	4	64	1,024	16,384
5	0101	5	80	1,280	20,480
6	0110	6	96	1,536	24,576
7	0111	7	112	1,792	28,672
8	1000	8	128	2,048	32,768
9	1001	9	144	2,304	36,864
A	1010	10	160	2,560	40,960
B	1011	11	176	2,816	45,056
C	1100	12	192	3,072	49,152
D	1101	13	208	3,328	53,248
E	1110	14	224	3,584	57,344
F	1111	15	240	3,840	61,440

Unicode Character Code

- **Unicode is a 16-bit code.**

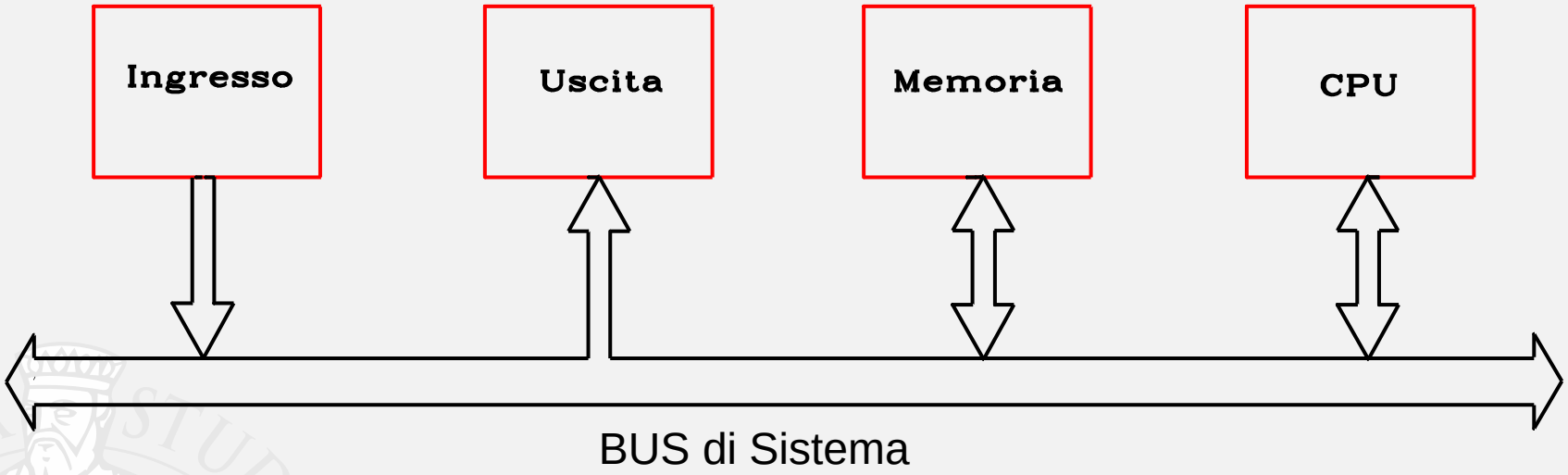


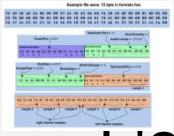
0000	NUL	0020	SP	0040	@	0060	`	0080	Ctrl	00A0	NBS	00C0	À	00E0	à
0001	SOH	0021	!	0041	A	0061	a	0081	Ctrl	00A1	¡	00C1	Á	00E1	á
0002	STX	0022	"	0042	B	0062	b	0082	Ctrl	00A2	¢	00C2	Â	00E2	â
0003	ETX	0023	#	0043	C	0063	c	0083	Ctrl	00A3	£	00C3	Ã	00E3	ã
0004	EOT	0024	$	0044	D	0064	d	0084	Ctrl	00A4	¤	00C4	Ä	00E4	ä
0005	ENQ	0025	%	0045	E	0065	e	0085	Ctrl	00A5	¥	00C5	Å	00E5	å
0006	ACK	0026	&	0046	F	0066	f	0086	Ctrl	00A6	¦	00C6	Æ	00E6	æ
0007	BEL	0027	'	0047	G	0067	g	0087	Ctrl	00A7	§	00C7	Ç	00E7	ç
0008	BS	0028	(	0048	H	0068	h	0088	Ctrl	00A8	¨	00C8	È	00E8	è
0009	HT	0029	)	0049	I	0069	i	0089	Ctrl	00A9	©	00C9	É	00E9	é
000A	LF	002A	*	004A	J	006A	j	008A	Ctrl	00AA	ª	00CA	Ê	00EA	ê
000B	VT	002B	+	004B	K	006B	k	008B	Ctrl	00AB	«	00CB	Ë	00EB	ë
000C	FF	002C	´	004C	L	006C	l	008C	Ctrl	00AC	¬	00CC	Ì	00EC	ì
000D	CR	002D	-	004D	M	006D	m	008D	Ctrl	00AD	–	00CD	Í	00ED	í
000E	SO	002E	.	004E	N	006E	n	008E	Ctrl	00AE	®	00CE	Î	00EE	î
000F	SI	002F	/	004F	O	006F	o	008F	Ctrl	00AF	¯	00CF	Ï	00EF	ï
0010	DLE	0030	0	0050	P	0070	p	0090	Ctrl	00B0	°	00D0	Ð	00F0	¶
0011	DC1	0031	1	0051	Q	0071	q	0091	Ctrl	00B1	±	00D1	Ñ	00F1	ñ
0012	DC2	0032	2	0052	R	0072	r	0092	Ctrl	00B2	²	00D2	Ò	00F2	ò
0013	DC3	0033	3	0053	S	0073	s	0093	Ctrl	00B3	³	00D3	Ó	00F3	ó
0014	DC4	0034	4	0054	T	0074	t	0094	Ctrl	00B4	´	00D4	Ô	00F4	ô
0015	NAK	0035	5	0055	U	0075	u	0095	Ctrl	00B5	µ	00D5	Õ	00F5	õ
0016	SYN	0036	6	0056	V	0076	v	0096	Ctrl	00B6	¶	00D6	Ö	00F6	ö
0017	ETB	0037	7	0057	W	0077	w	0097	Ctrl	00B7	·	00D7	×	00F7	÷
0018	CAN	0038	8	0058	X	0078	x	0098	Ctrl	00B8	¸	00D8	Ø	00F8	ø
0019	EM	0039	9	0059	Y	0079	y	0099	Ctrl	00B9	¹	00D9	Ù	00F9	ù
001A	SUB	003A	:	005A	Z	007A	z	009A	Ctrl	00BA	º	00DA	Ú	00FA	ú
001B	ESC	003B	;	005B	[	007B	{	009B	Ctrl	00BB	»	00DB	Û	00FB	û
001C	FS	003C	<	005C	\	007C	|	009C	Ctrl	00BC	1/4	00DC	Ü	00FC	ü
001D	GS	003D	=	005D	]	007D	}	009D	Ctrl	00BD	1/2	00DD	Ý	00FD	Þ
001E	RS	003E	>	005E	^	007E	~	009E	Ctrl	00BE	3/4	00DE	ý	00FE	þ
001F	US	003F	?	005F	_	007F	DEL	009F	Ctrl	00BF	¿	00DF	ß	00FF	ÿ

NUL	Null	SOH	Start of heading	CAN	Cancel	SP	Space
STX	Start of text	EOT	End of transmission	EM	End of medium	DEL	Delete
ETX	End of text	DC1	Device control 1	SUB	Substitute	Ctrl	Control
ENQ	Enquiry	DC2	Device control 2	ESC	Escape	FF	Form feed
ACK	Acknowledge	DC3	Device control 3	FS	File separator	CR	Carriage return
BEL	Bell	DC4	Device control 4	GS	Group separator	SO	Shift out
BS	Backspace	NAK	Negative acknowledge	RS	Record separator	SI	Shift in
HT	Horizontal tab	NBS	Non-breaking space	US	Unit separator	DLE	Data link escape
LF	Line feed	ETB	End of transmission block	SYN	Synchronous idle	VT	Vertical tab

Il Calcolatore

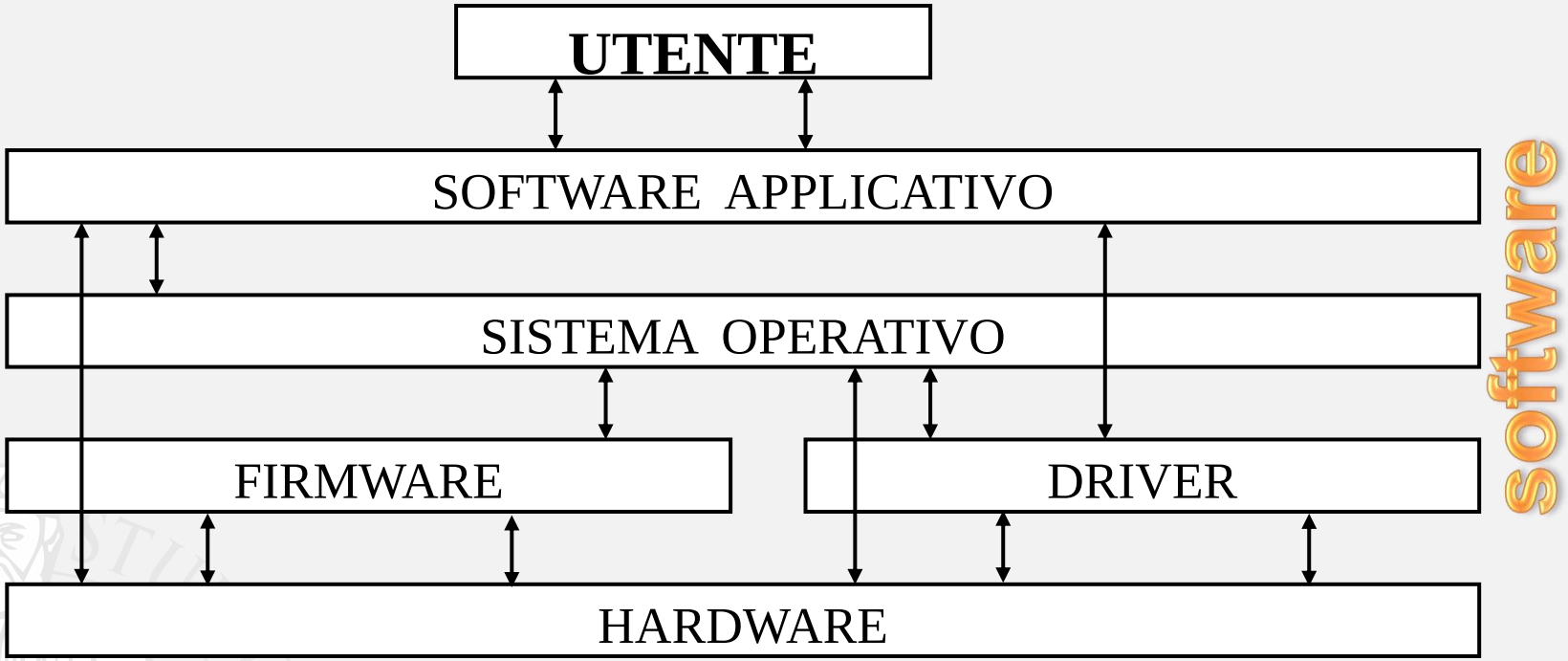
- Corrisponde allo schema dei PC anni 80
- Tuttora in largo uso nei sistemi di controllo
- Le unità sono collegate da un bus di sistema (insieme di fili che portano informazioni in parallelo: per esempio, 8 bit → 8 fili



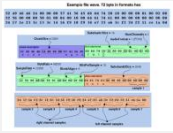


Hardware e Software

L'elaboratore è composto dall'hardware, i dispositivi fisici che lo costituiscono, e dal software, quelle procedure e istruzioni che ne dirigono le operazioni.
Visione Funzionale vs interazione reale dell'utente

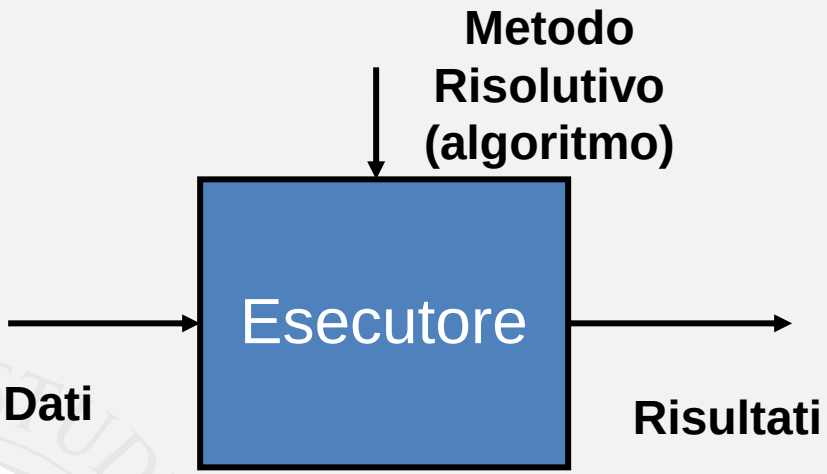


Algoritmo



Un algoritmo è una sequenza **finita** di mosse (passi) che risolve ***in un tempo finito*** una *classe* di problemi.

- L'esecuzione delle azioni *nell'ordine specificato dall'algoritmo* consente di ottenere, a partire dai dati di ingresso, i risultati che risolvono il problema.
- Algoritmi complessi possono risolvere problemi semplici e viceversa



ESECUTORE/elaboratore
una ***macchina astratta***
capace di ***eseguire le azioni***
specificate dall'algoritmo.

I dati e I formati

- Formati e informazioni
- Codifica
- Informatica ?? Informazione – automatica
- Volume dei dati
  - velocità dei flussi e computabilità:
  - static e real time
- Informazione strutturata
- entità vs relazioni: rappresentazione conoscenza
- tabelle vs reticoli
- XML

Bit e byte

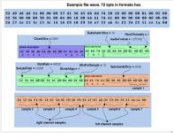
- 1 bit: V/F, 1/O, H/L, acceso/spento
- 1 Byte = 8 bit
- 1 Word = 16 o 32 bit (e.g., word a 16…)
- 1 nibble = 4 bit
- Un numero binario a 10 bit ha una dinamica (numero di combinazioni) di 1024, da 0 a 1023:
- Il massimo numero rappresentabile con n bit è pari alla dinamica meno 1 cioè (2^n-1)

Prefissi Scientifici – la memoria

Scientific Prefixes

- For computer memory, 1K = 2^{10} = 1024. For everything else, like clock speeds, 1K = 1000, and likewise for 1M, 1G, *etc.*

Prefix	Abbrev.	Quantity	Prefix	Abbrev.	Quantity
milli	m	10^{-3}	Kilo	K	10^{3}
micro	μ	10^{-6}	Mega	M	10^{6}
nano	n	10^{-9}	Giga	G	10^{9}
pico	p	10^{-12}	Tera	T	10^{12}
femto	f	10^{-15}	Peta	P	10^{15}
atto	a	10^{-18}	Exa	E	10^{18}



Prefissi del Sistema Internazionale

10^n	Bit	Prefisso	Simbolo	Nome	Equivalente decimale
10^{24}	80	yotta	Y	Quadrilione	1 000 000 000 000 000 000 000 000
10^{21}	70	zetta	Z	Triliardo	1 000 000 000 000 000 000 000
10^{18}	60	exa	E	Trilione	1 000 000 000 000 000 000
10^{15}	50	peta	P	Biliardo	1 000 000 000 000 000
10^{12}	40	tera	T	Bilione	1 000 000 000 000
10^{9}	30	giga	G	Miliardo	1 000 000 000
10^{6}	20	mega	M	Milione	1 000 000
10^{3}	10	kilo o chilo	k	Mille	1 000
10^{2}	7	etto	h	Cento	100
10	4	deca	da	Dieci	10

Flussi dati

- bps: bit per secondo
- Bps: Byte per secondo pari a numero di bps/8
- KByte, KB: 1024 Byte
- 700 kbps → 87 kBps (pari a 700/8)
- 87 kBps → 89,600 Bps (pari a 87*1024)
- 700 kbps → 716800 bps
  (pari a 700*1024, 89600*8)

I dati e I formati

- Formati e informazioni
- Codifica
- Informatica ?? Informazione – automatica
- Volume dei dati
  - velocità dei flussi e computabilità:
  - static e real time
- Informazione strutturata
- entità vs relazioni: rappresentazione conoscenza
- tabelle vs reticoli
- XML

Tabelle

- Tipi diversi
- Semantica definita

Installation neue PC			
Abteilung	Geplante PC	Installations -termin	Kosten
Einkauf	2 Desktops	August	2.200€
Verkauf	2 Laptops	September	4.600€
Lager	1 Desktop 1 Laptop	Oktober	3.400€

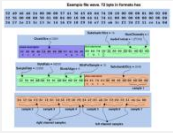
Esempio di Informazione strutturata

- Persona
  - Nome: massimo 25 caratteri, codifica ASCII
  - Cognome: massimo 25 caratteri, codifica ASCII
  - Età: 1 byte, numero naturale
  - Telefono: massimo 15 caratteri ????
  - Indirizzo:
    - Via/piazza: ……..
    - Numero:…….. Non basta……
    - Cap:……..
    - Città:…… descrizione o riferimento…….
  - Padre:…… riferimento a ….
  - Madre:…….. Riferimento a …….

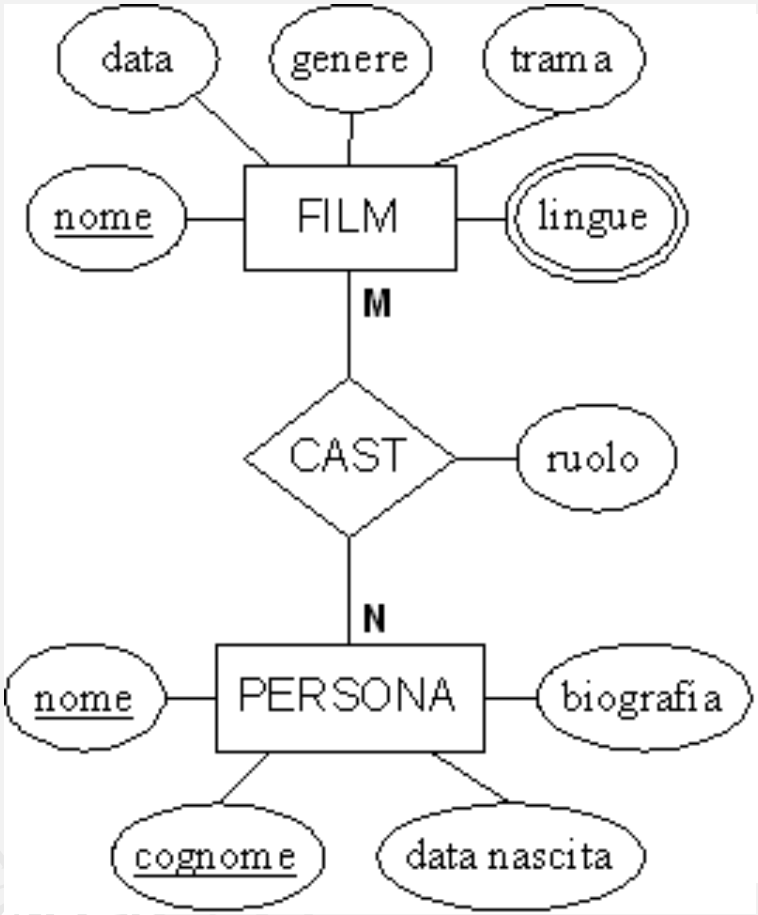
- Semantica e dati ←→ solo dati:
  - Paolo_____|Nesi_____|….|…….|……..|……|

I dati e I formati

- Formati e informazioni
- Codifica
- Informatica ?? Informazione – automatica
- Volume dei dati
  - velocità dei flussi e computabilità:
  - static e real time
- Informazione strutturata
- entità vs relazioni: rappresentazione conoscenza
- tabelle vs reticoli
- XML



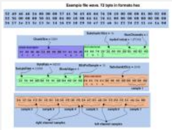
Entità e relazioni

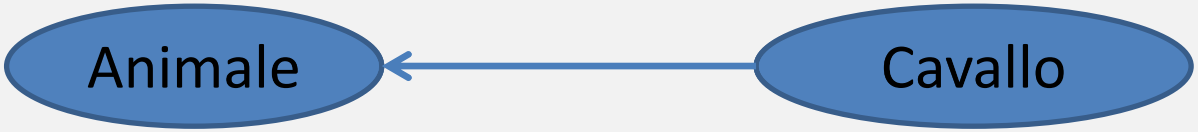


Struttura di classificazione	Simbolo
Entità	Nome entità
Relazione	nome
Attributo semplice	Nome attributo
Attributo composto	Nome attrib. Nome attrib. Nome
Sottoinsieme	
Generalizzazione	

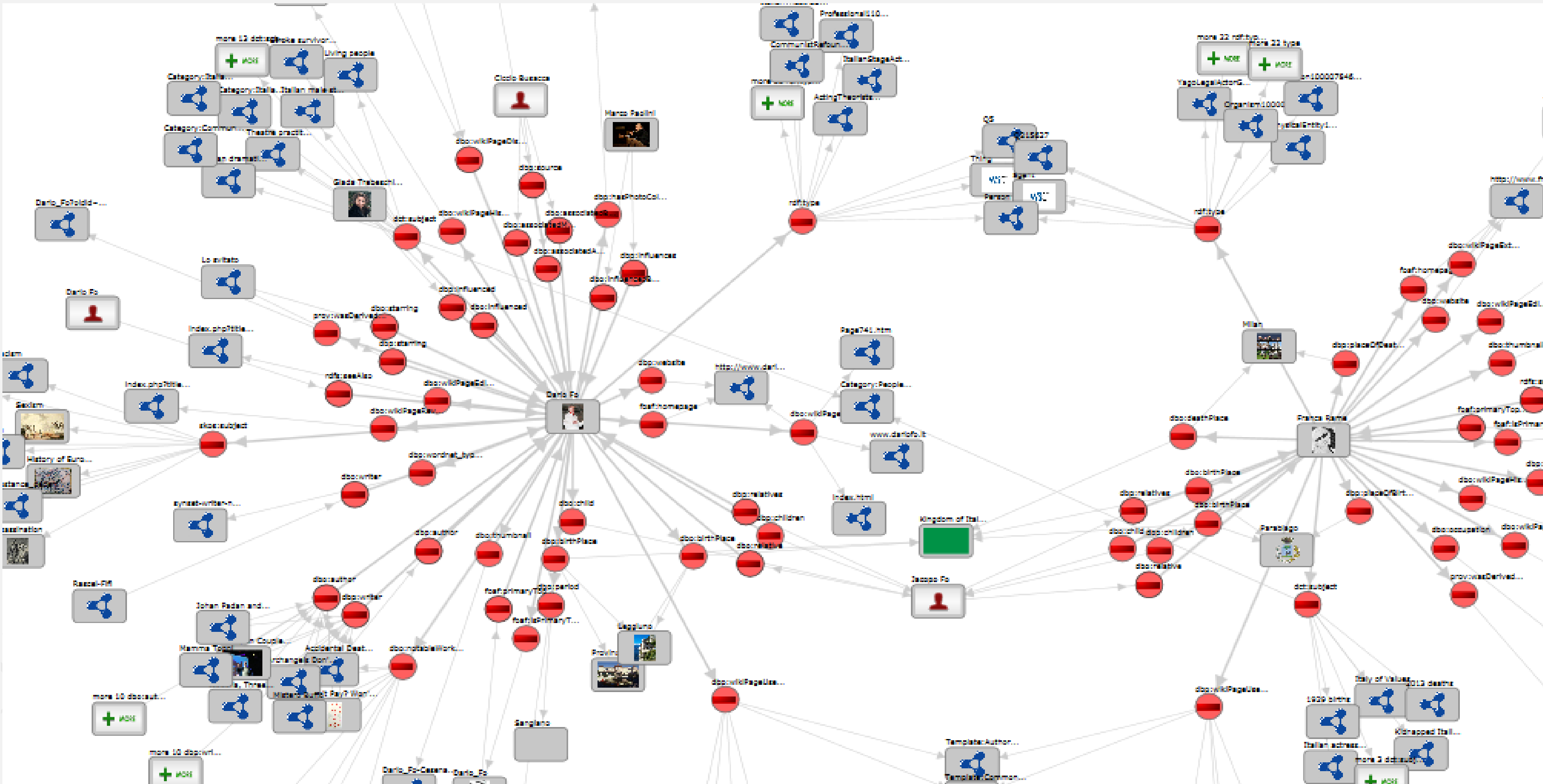
Fig.3 Rappresentazione dei concetti del modello entità-relazione

La semantica, le relazioni

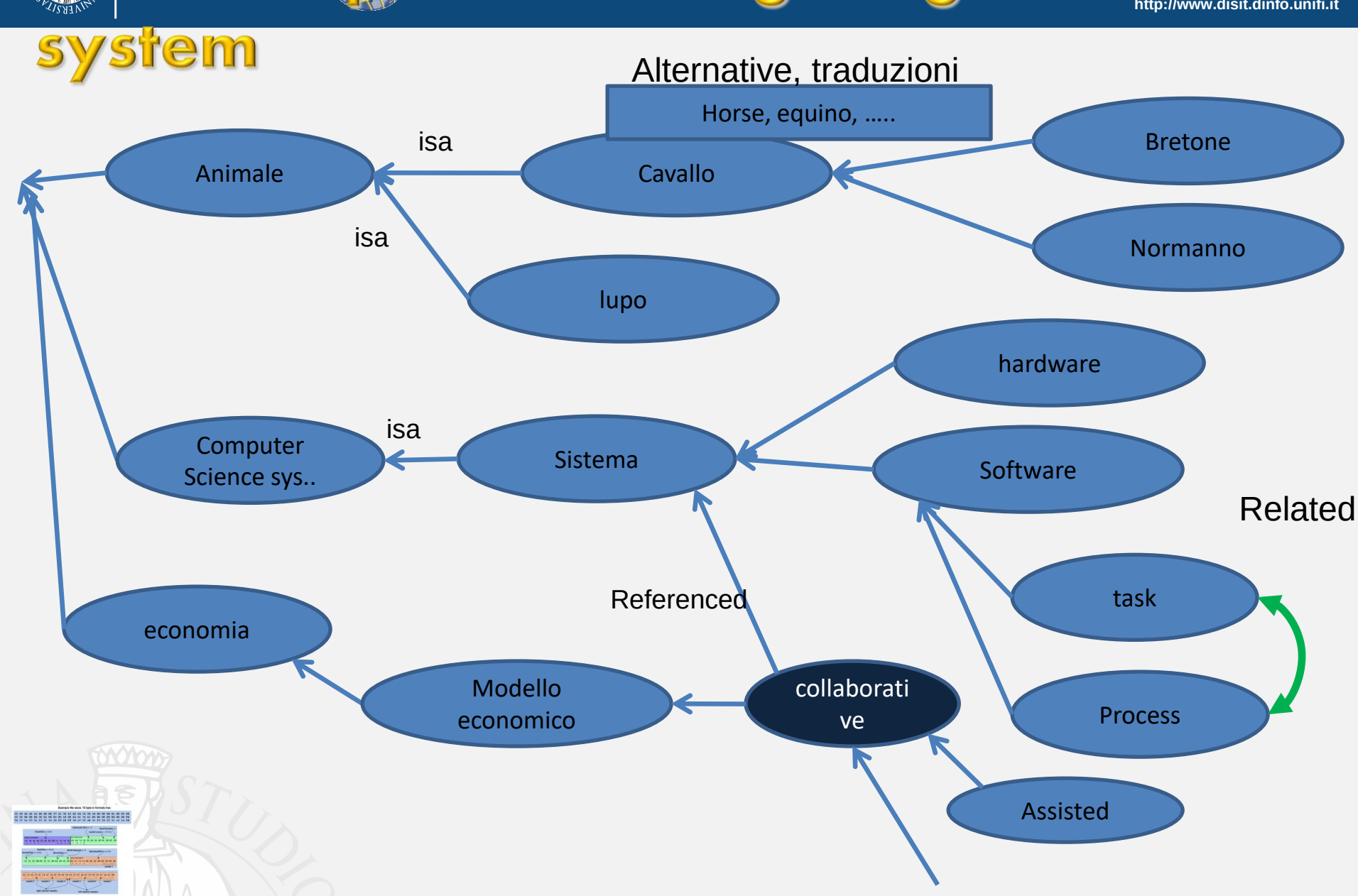




Il cavallo è un animale



SKOS simple knowledge organization system



Tabelle

- Tipi diversi
- Semantica definita
- Le relazioni stanno fra le colonne e sulle colonne

Installation neue PC			
Abteilung	Geplante PC	Installations -termin	Kosten
Einkauf	2 Desktops	August	2.200€
Verkauf	2 Laptops	September	4.600€
Lager	1 Desktop 1 Laptop	Oktober	3.400€

Il concetto di metalinguaggio

- XML è un **metalinguaggio**
  - XML definisce un insieme regole (meta)sintattiche, attraverso le quali è possibile descrivere formalmente un linguaggio di markup, detto “applicazione XML”
  - ogni applicazione XML eredita un insieme di caratteristiche sintattiche comuni
  - ogni applicazione XML a sua volta definisce una sintassi formale particolare
  - ogni applicazione XML è dotata di una semantica specificata in modo non formale

XML: caratteristiche (2)

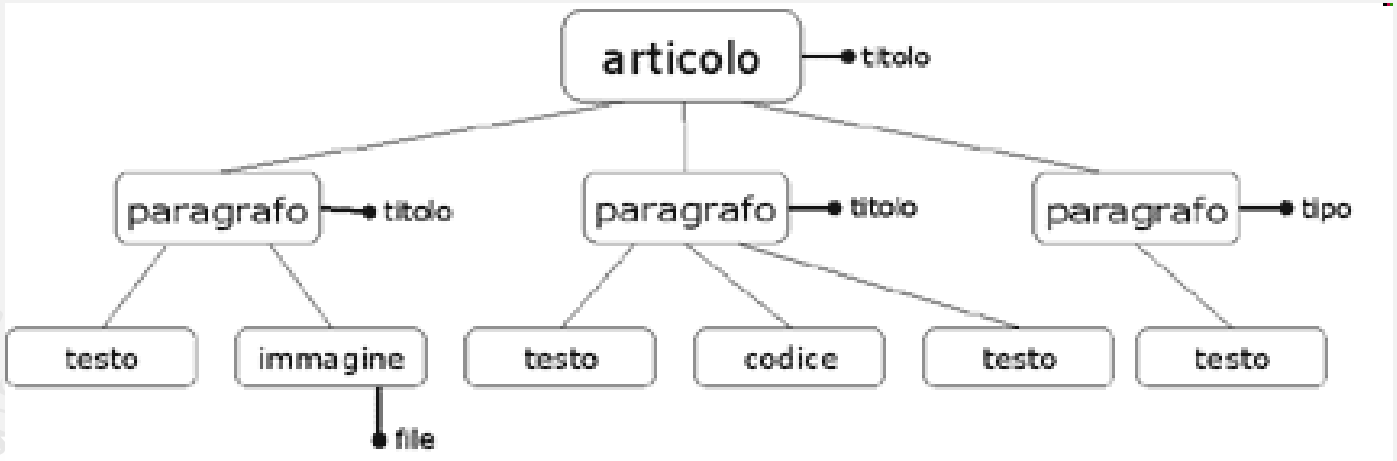
- XML è indipendente dal tipo di piattaforma hardware e software su cui viene utilizzato
- XML permette la rappresentazione di qualsiasi tipo di documento (e di struttura testuale) indipendentemente dalle finalità applicative
- XML è indipendente dai dispositivi di archiviazione e visualizzazione.
- Un Documento XML può essere
  - archiviato su qualsiasi tipo di supporto digitale (attuale e… futuro!)
  - visualizzato su qualsiasi dispositivo di output

XML: caratteristiche (3)

- XML può essere usato per la rappresentazione di dati strutturati (archivi, tabelle, matrici) in alternativa ai formati di database tradizionali
- XML è uno standard di pubblico dominio
- ogni software “conforme XML” è in grado di gestire dati in formato XML
- sono disponibili numerose applicazioni e librerie *open source* per la manipolazione di dati in formato XML basate su diversi linguaggi di programmazione (Java, C, Python, Perl…)
- una applicazione in grado di elaborare dati in formato XML viene definita elaboratore XML

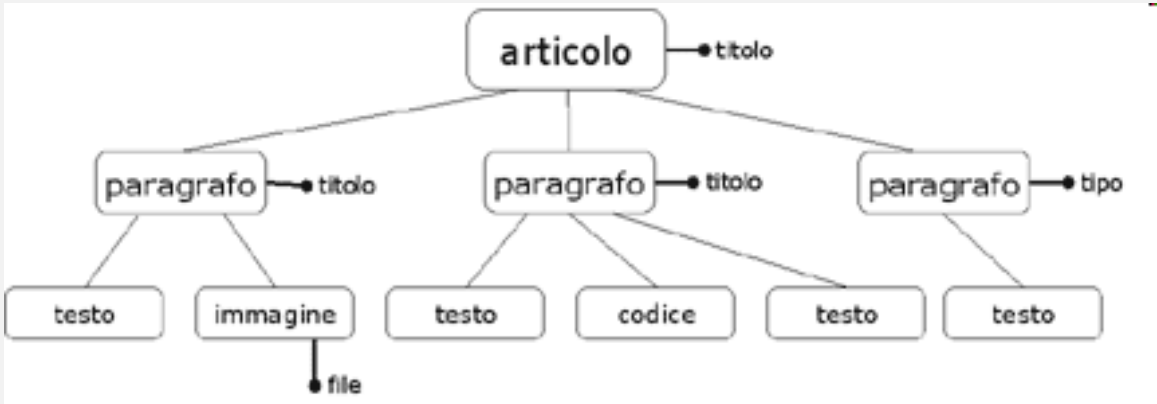
XML: caratteristiche (4)

- XML adotta un formato di file di tipo testuale: sia il mark-up sia il testo sono stringhe di caratteri
- XML si basa sul sistema di codifica dei caratteri ISO 10646/UNICODE
- Un documento XML è “leggibile” da un utente umano senza la mediazione di software specifico
- Concretamente, un documento XML è un file di testo che contiene una serie di tag, attributi e testo secondo regole sintattiche ben definite.



Esempio: articolo

```
<?xml version="1.0" ?>
<articolo titolo="Titolo dell'articolo">
  <paragrafo titolo="Titolo del primo paragrafo">
    <testo>
      Blocco di testo del primo paragrafo
    </testo>
    <immagine file="immagine1.jpg">
    </immagine>
  </paragrafo>
  <paragrafo titolo="Titolo del secondo paragrafo">
    <testo>
      Blocco di testo del secondo paragrafo
    </testo>
    <codice>
      Esempio di codice
    </codice>
    <testo>
      Altro blocco di testo
    </testo>
  </paragrafo>
```



sommario

- *Dati e formati*
- *Comunicazione fra sistemi computerizzati*
- *Modelli di protezione e gestione*
- *Dati vs Metadati*
- *Social network e Smart City*
- *Motori di Ricerca, Crawling e NLP*
- *Data Mining*
- *Data Intelligence*

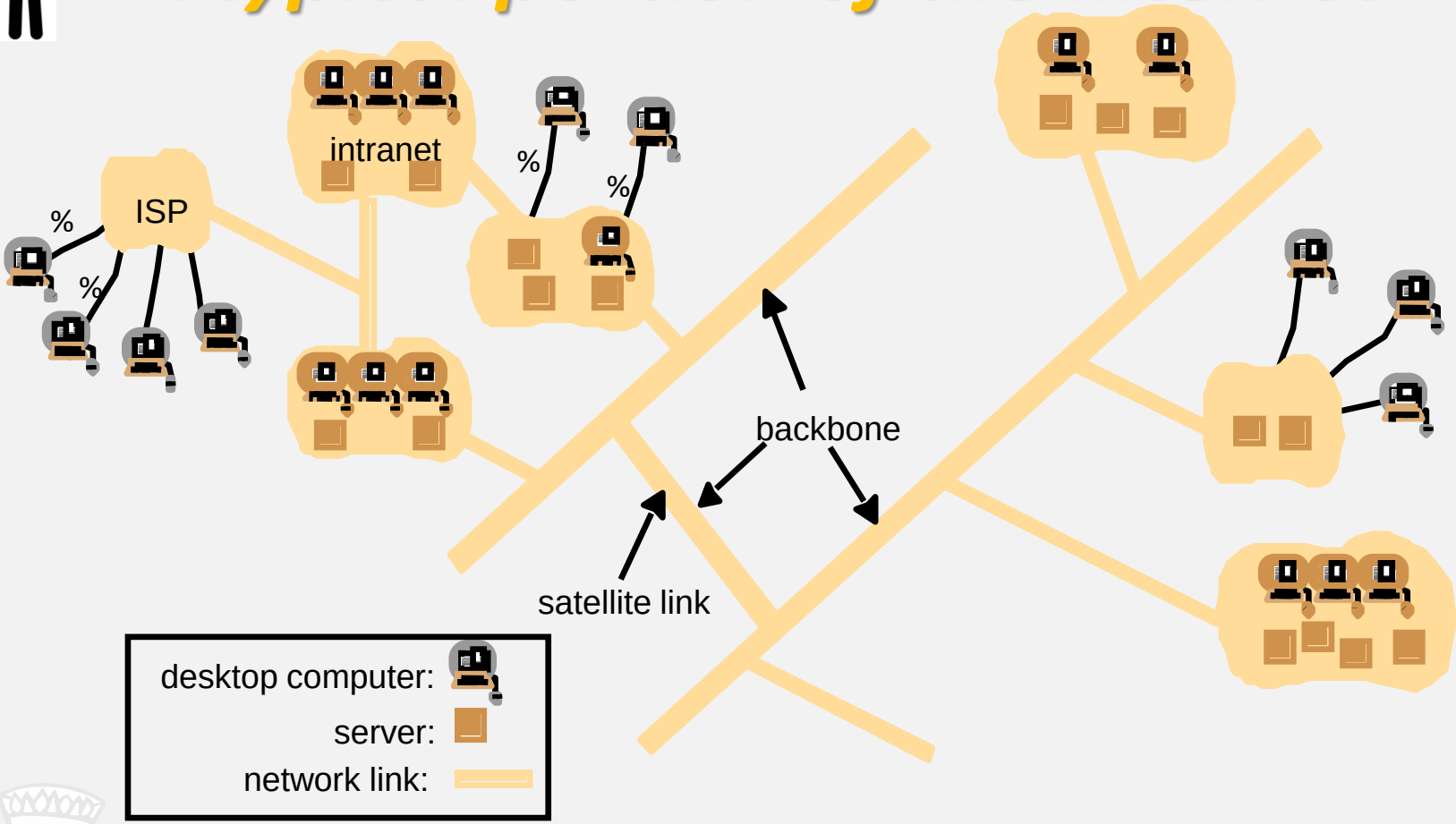
Calcolatori in rete: Networking

- Sistemi Distribuiti
- Connessioni fra calcolatori
- IP: indirizzi e porte
- Stack ISO OSI
- Protocolli
- Connessioni intercontinentali
- Sicurezza sulla rete

Sistemi Distribuiti

- Un Sistema distribuito è composto da componenti / strumenti messi in relazione tramite una rete di computer.
  - Tali componenti comunicano fra di loro tramite messaggi
- Messaggi portano informazioni:
  - Controlli, oppure Dati
- Esempi di sistemi distribuiti sono:
  - Internet, intranet, mobile and ubiquitous computing

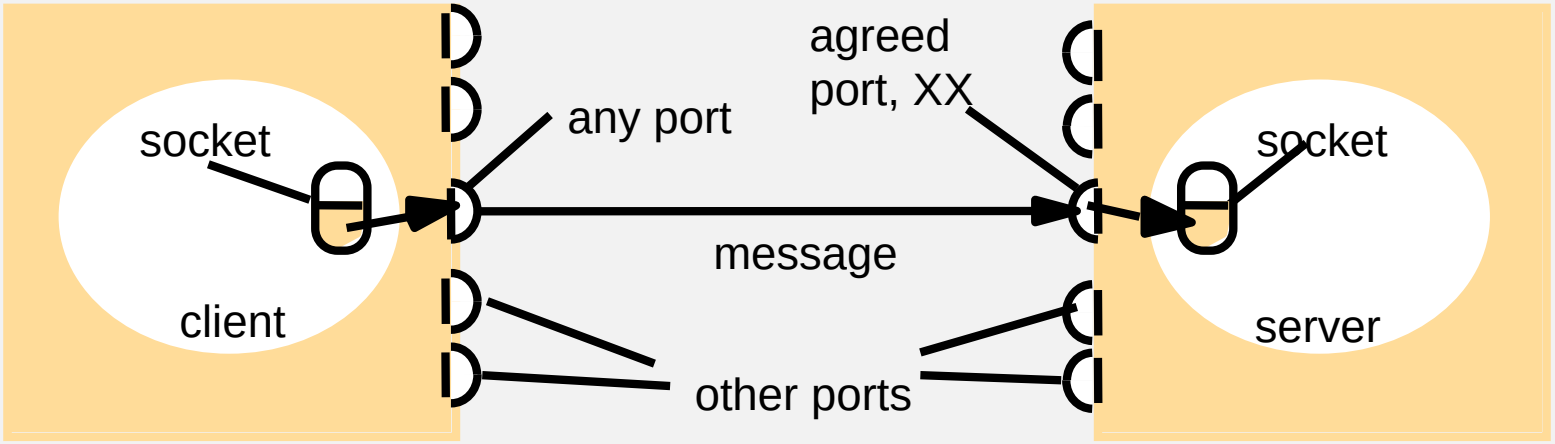
A typical portion of the Internet



Calcolatori in rete: Networking

- Sistemi Distribuiti
- Connessioni fra calcolatori
- IP: indirizzi e porte
- Stack ISO OSI
- Protocolli
- Connessioni intercontinentali
- Sicurezza sulla rete

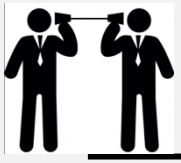
Sockets and ports



Internet address = 138.37.94.248:yy

Internet address = 138.37.88.249

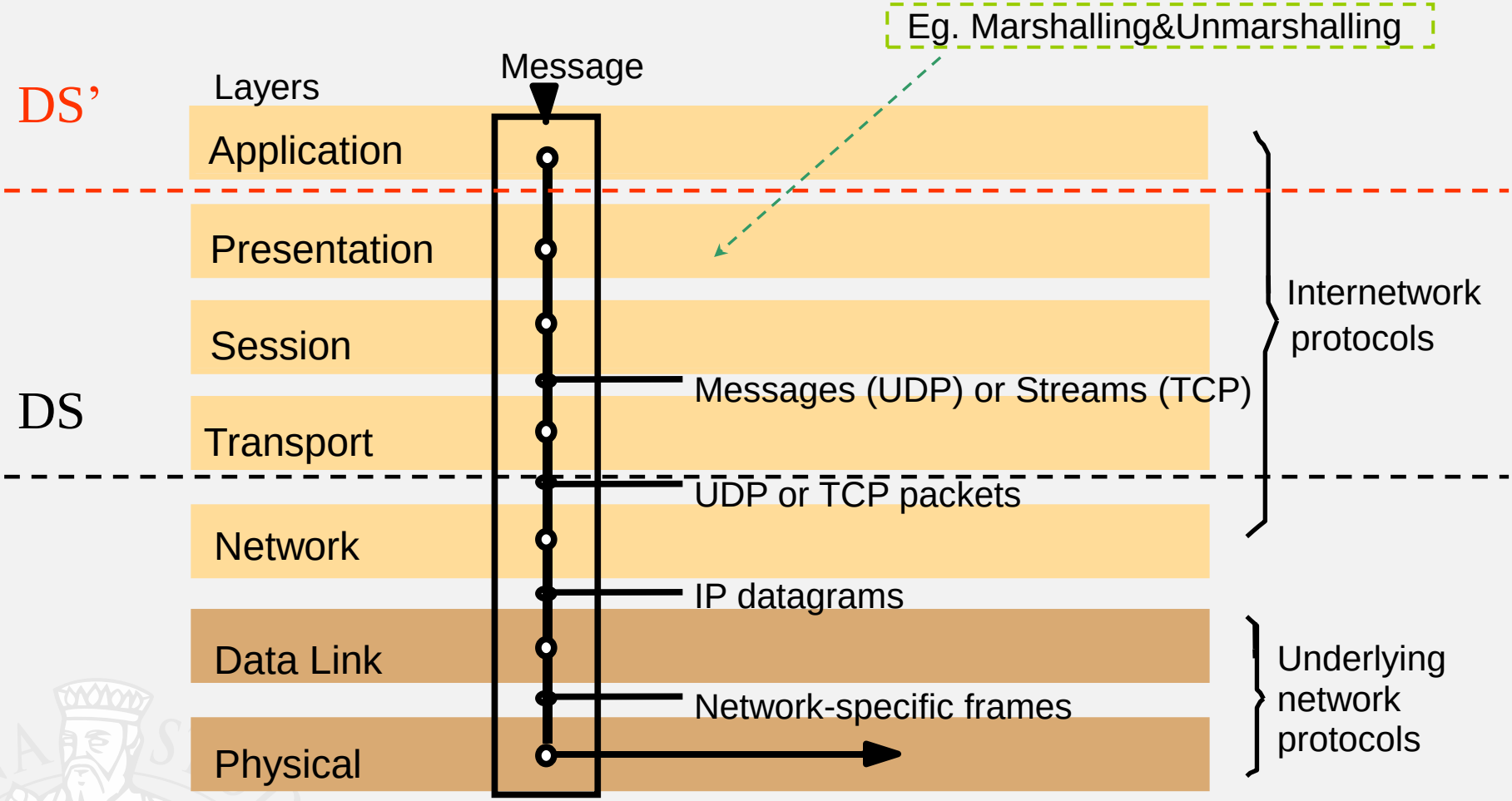
Port=yy

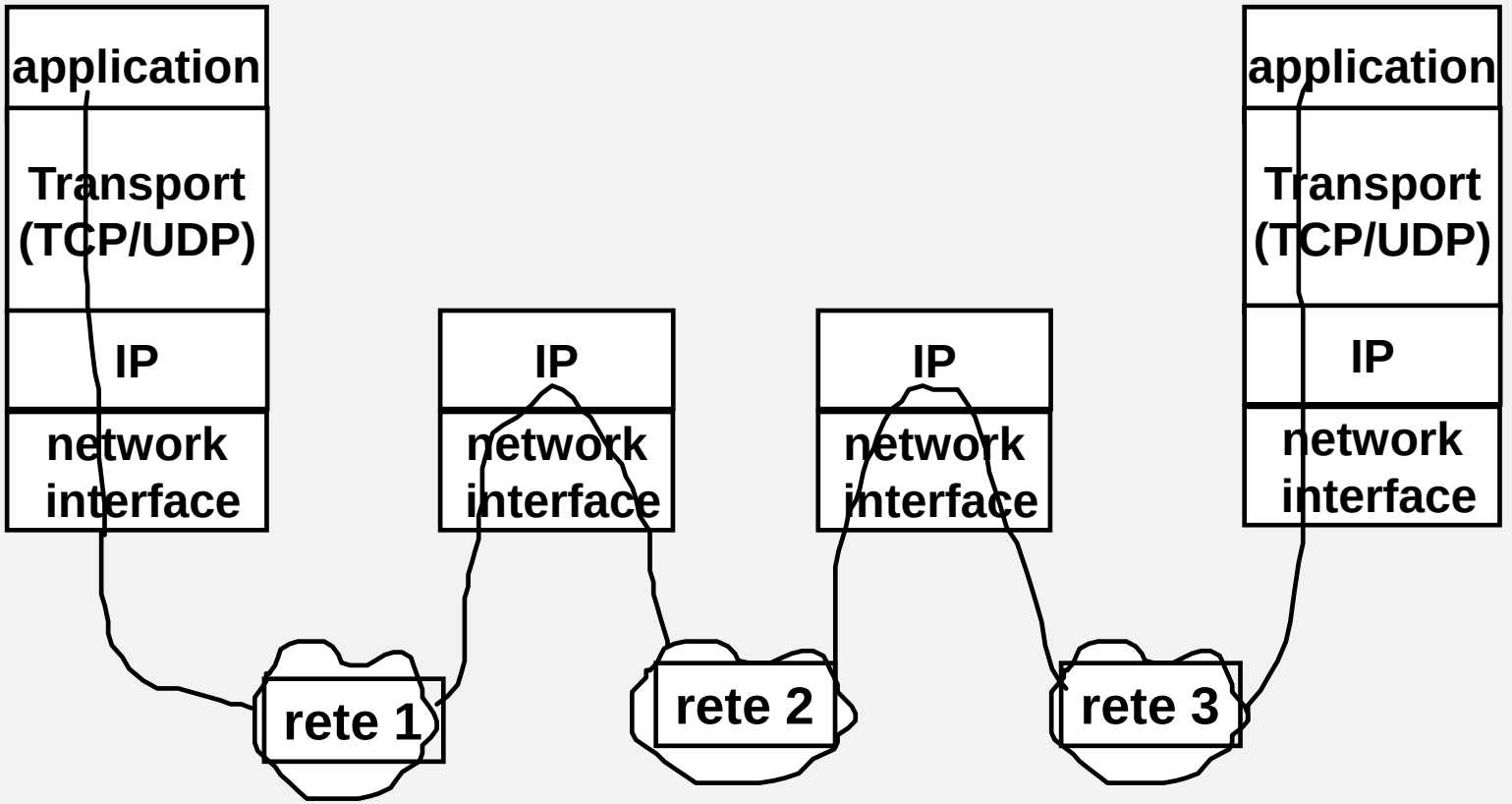


OSI protocol summary

Layer	Description	Examples
Application	Protocols that are designed to meet the communication requirements of specific applications, often defining the interface to a service.	HTTP, FTP, SMTP, CORBA IIOP
Presentation	Protocols at this level transmit data in a network representation that is independent of the representations used in individual computers, which may differ. Encryption is also performed in this layer, if required.	Secure Sockets (SSL),CORBA Data Rep.
Session	At this level reliability and adaptation are performed, such as detection of failures and automatic recovery.	
Transport	This is the lowest level at which messages (rather than packets) are handled. Messages are addressed to communication ports attached to processes, Protocols in this layer may be connection-oriented or connectionless.	TCP, UDP
Network	Transfers data packets between computers in a specific network. In a WAN or an internetwork this involves the generation of a route passing through routers. In a single LAN no routing is required.	IP, ATM virtual circuits
Data link	Responsible for transmission of packets between nodes that are directly connected by a physical link. In a WAN transmission is between pairs of routers or between routers and hosts. In a LAN it is between any pair of hosts.	Ethernet MAC, ATM cell transfer, PPP
Physical	The circuits and hardware that drive the network. It transmits sequences of binary data by analogue signalling, using amplitude or frequency modulation of electrical signals (on cable circuits), light signals (on fibre optic circuits) or other electromagnetic signals (on radio and microwave circuits).	Ethernet base- band signalling, ISDN

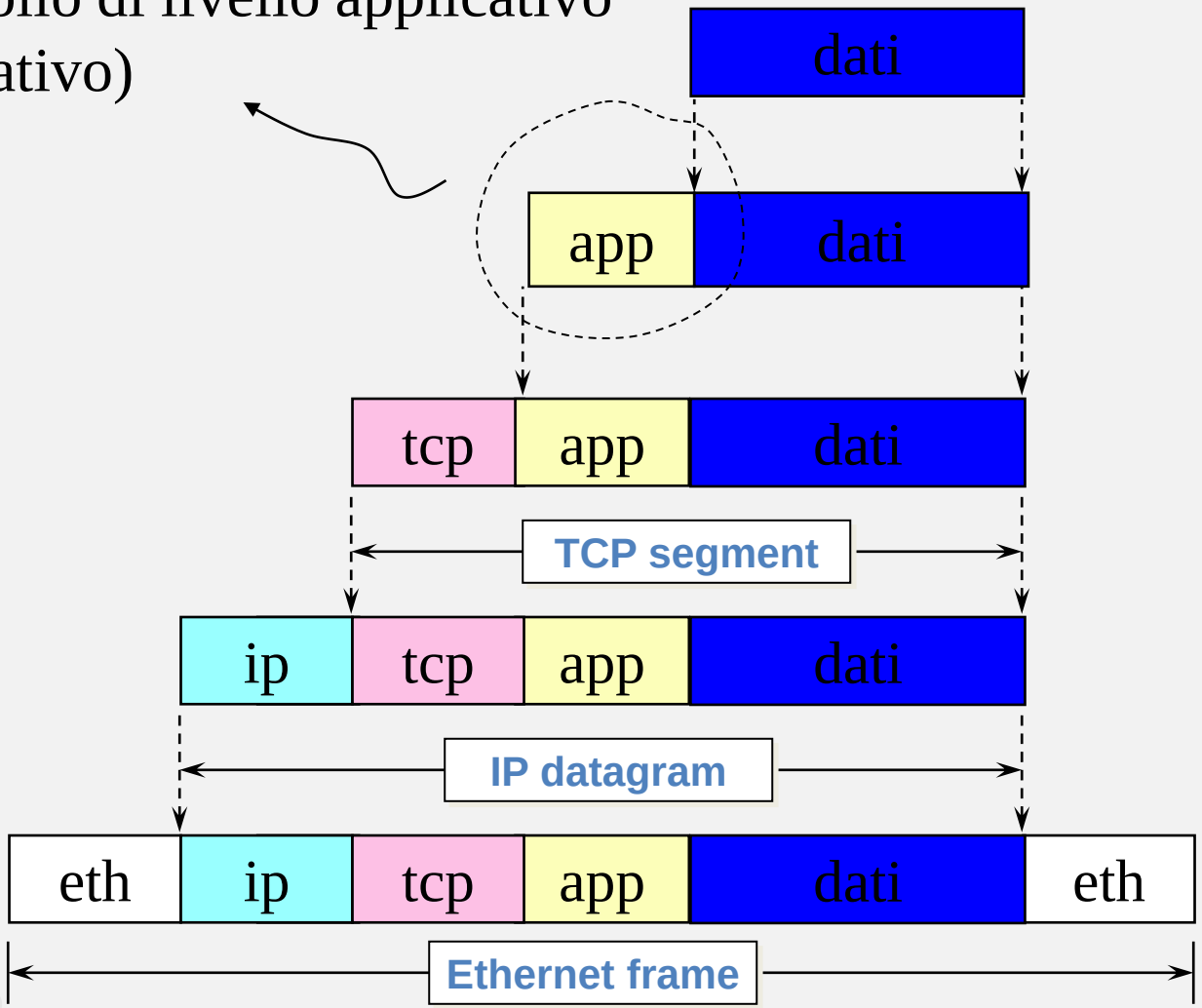
Network Communication



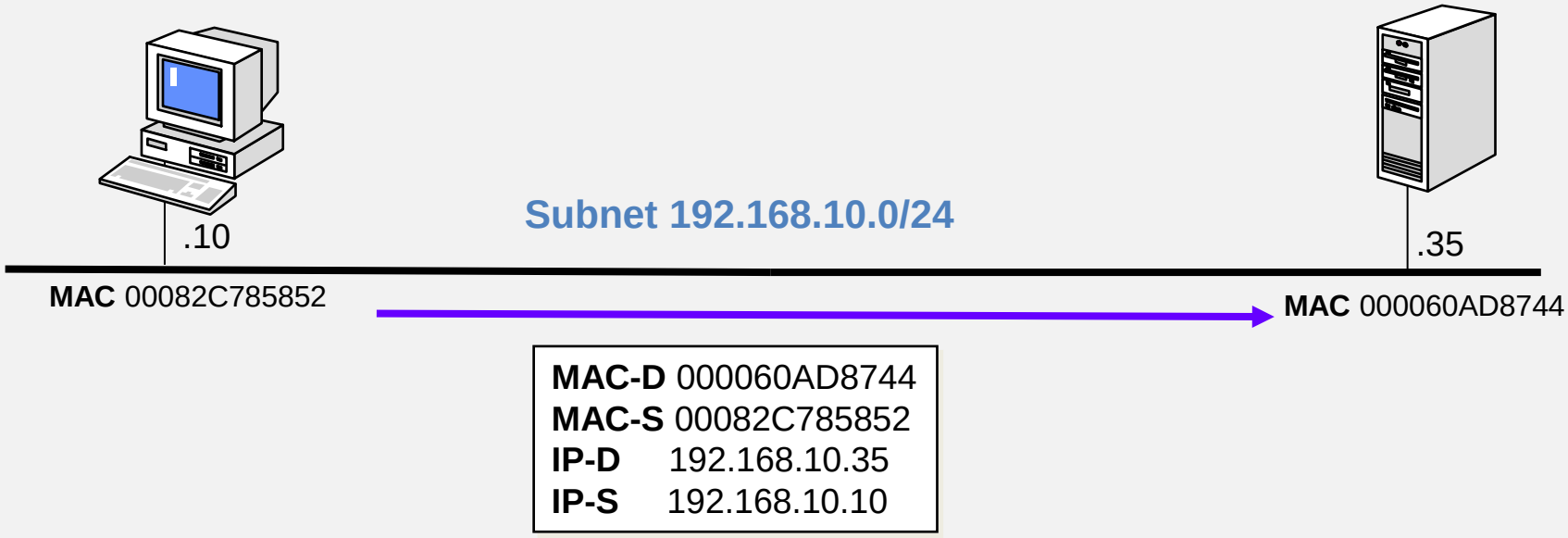


- IP <u>virtualizza omogeneità</u> di gestione delle informazioni tra reti distinte che compongono Internet

protocollo di livello applicativo (facoltativo)

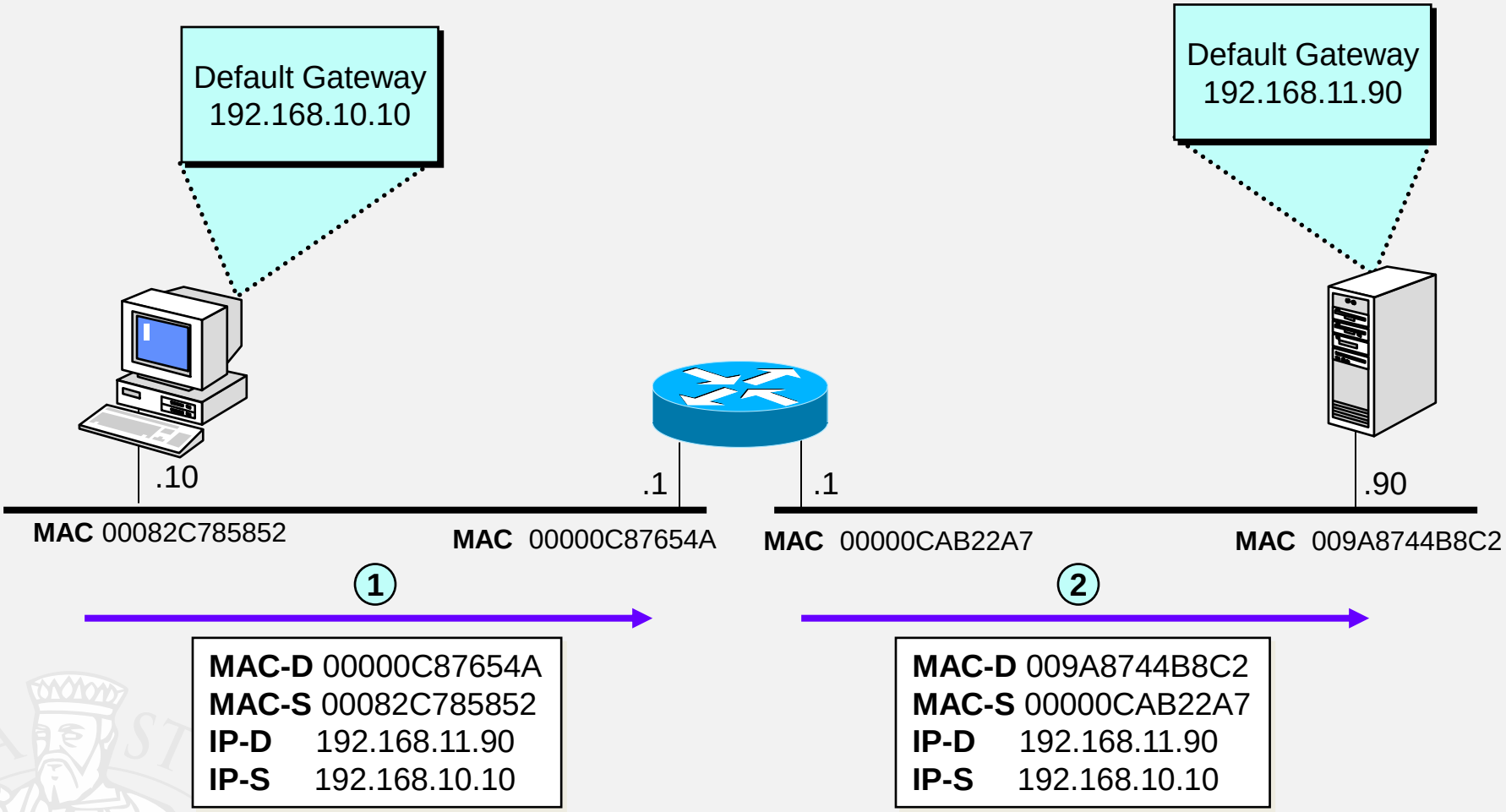


Forwarding diretto: esempio



- il mittente genera un frame contenente l'indirizzo MAC dell'host destinazione

Forwarding indiretto: esempio



Condivisione delle Risorse

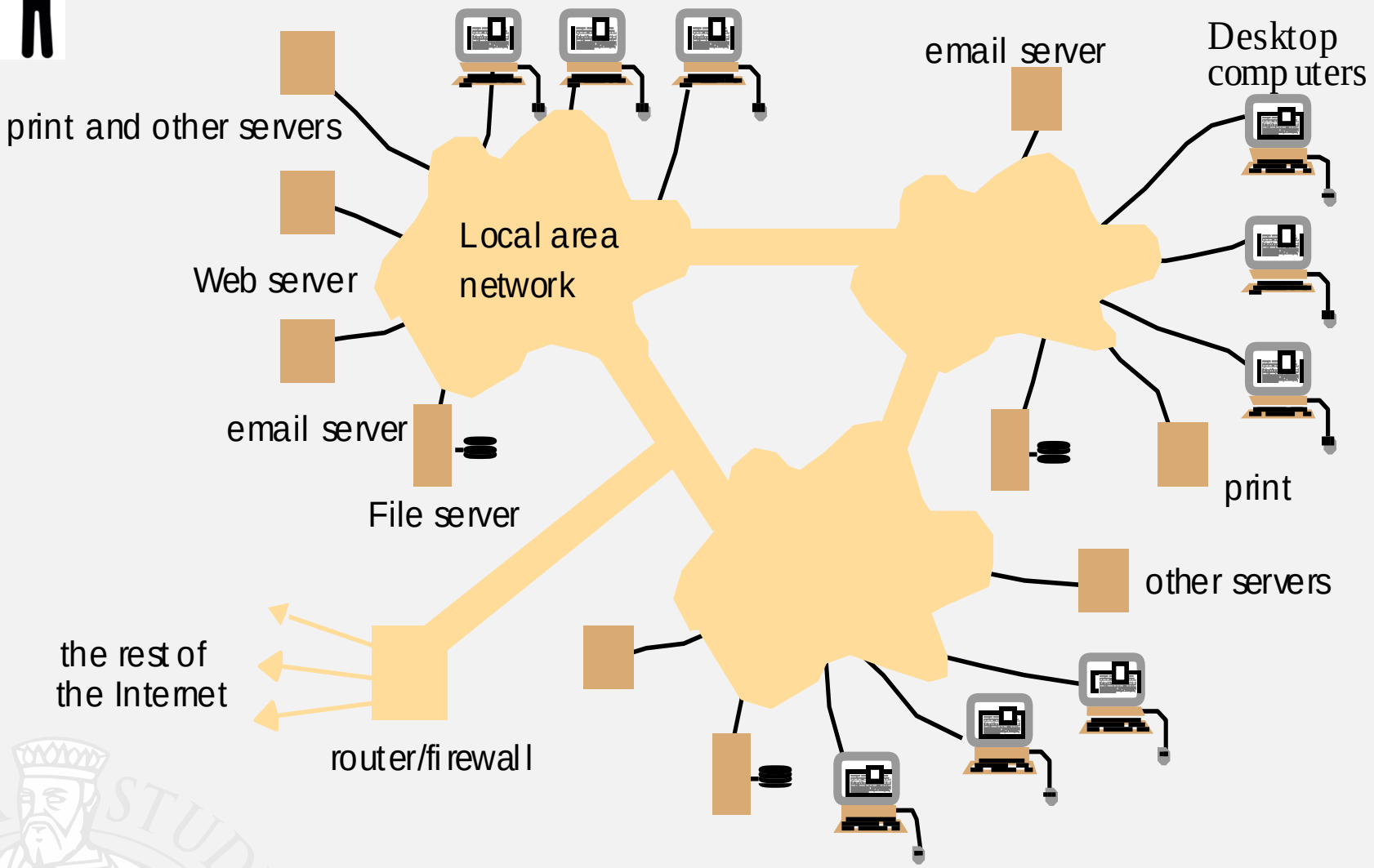
- **Condivisione di risorse hw: stampanti, file, cpu, …**
- **WEB Servers**:
  - Web pages (HTML… XML….XSL), a range of sercives
- **Cooperative Work**
  - Cooperative/collaborative Work, CSCW
  - Configuration Management and development tools, CVS
  - Applicazioni P2P
- **Condivisione di servizi**
  - WEB Services, portali e chiamate REST
  - Remote Procedure Calls, RPC, …RMI
  - Distributed Objects
  - GRID computing, parallel distributed computing
- **Condivizione servizi di calcolo**: cloud computing
  - Massive computing, GRID computing
  - Virtualizzazione, cloud, storage

Sistemi Eterogenei

- **Diversi ??**
  - reti diverse
  - supporti e protocolli diversi
  - computer con hardware diversi
  - sistemi operativi diversi con gli stessi protocolli
  - linguaggi di programmazione diversi per servizi e per la realizzazione di oggetti condivisi e chiamate remote
  - implementazione di servizi diversi
  - implementazioni diverse degli stessi servizi,
  - etc.

- **Middleware to mask heterogeneity**
  - CORBA, Java RMI, J2EE, ……
  - .NET, DCOM….
  - Accesso distribuito a SQL

A typical intranet



Calcolatori in rete: Networking

- Sistemi Distribuiti
- Connessioni fra calcolatori
- IP: indirizzi e porte
- Stack ISO OSI
- Protocolli
- Connessioni intercontinentali
- Sicurezza sulla rete

Internet vs Intranet

- Protocollo TCP/IP
- Servizi di base sono:
  - WWW, World Wide Web: HTTP, XML
  - FTP, File Transfer Protocol
  - Mail, chat, meeting, etc.
  - P2P (vari protocolli), GRID
  - Multimedia distribution: IPTV, VOIP, VOD, webTV, etc.

- Connessioni a Internet
  - Via ISP: Internet Service Provider
  - da Internet a Intranet
  - di Internet sul Backbone ad elevata velocità

- WEB Services and WEB Servers

Layers of technology

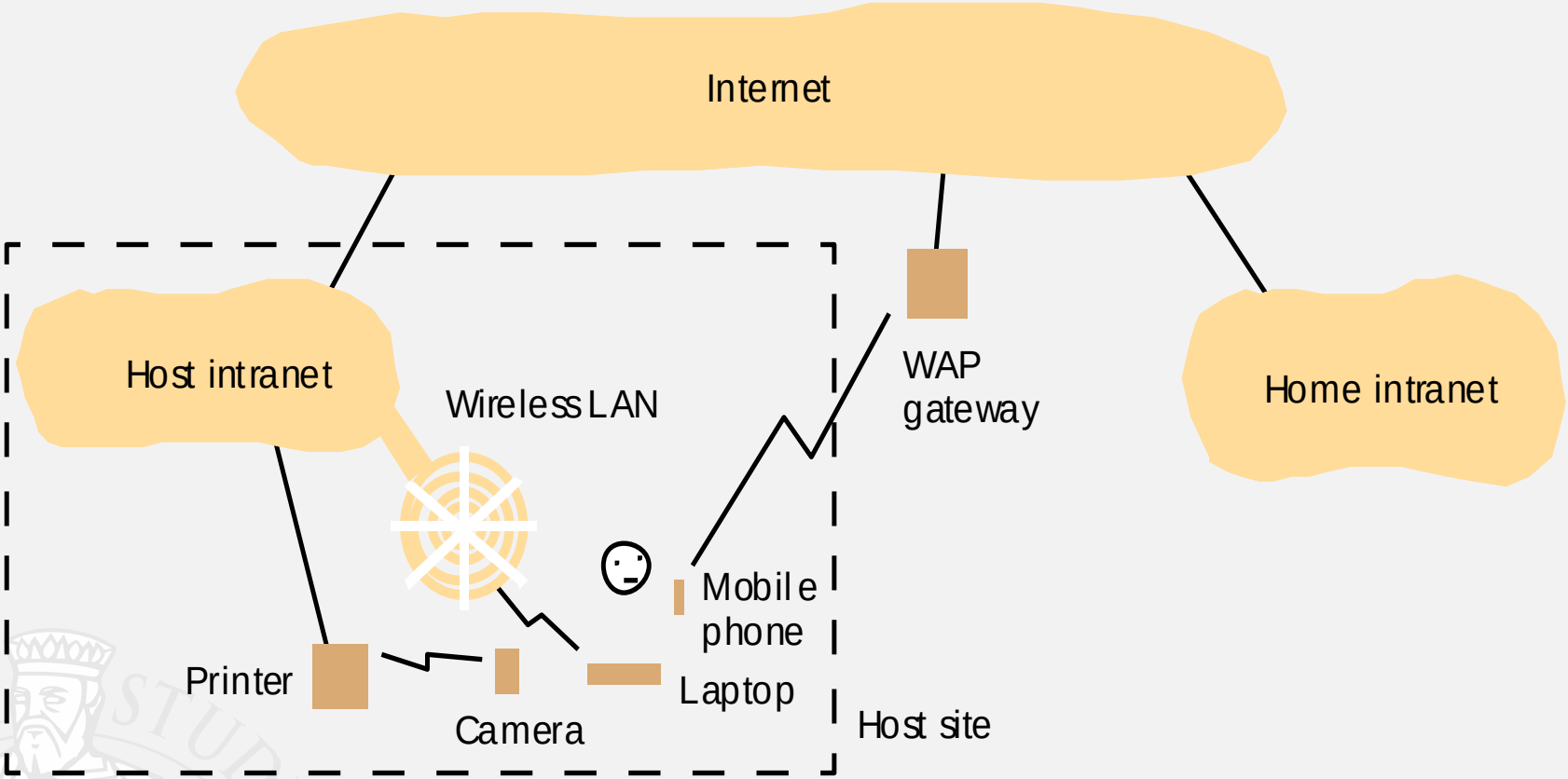
DEVICE	Laptop	PDA	Handset		
NETWORK	WLAN	GSM	GPRS	UMTS	
PROTOCOL	SMS	EMS	MMS	I-mode	WAP
LANGUAGE	WML	XML	HTML		
INTERACTION	One way down: Alert	One way down: Download	Two ways: Near real time browsing	Two ways: Real time browsing	
CONSULTATION MODE	Location based	Non-Location based			
SUPPORT	Text	Image	Video	Software	Audio
APPLICATION	Gaming	News	Financial info	Travel	Edutainment
INDUSTRY PROVIDER	Public inst.	Newspapers	....	Software devel.	

Source: Andersen

Basi Tecnologiche x Sistemi Distribuiti

- **Problemi e tecnologie per la gestione di**
  - **Comunicazioni** fra processi concorrenti e distribuiti
    - prob: di comunicazione, incertezza/latenza, interruzione, etc.
    - soluzioni: protocolli robusti, modelli robusti
  - **Sincronizzazioni** temporali, e.g., far partire azioni in simultanea
    - prob: mancanza di un clock comune assoluto
    - prob: precisione della Sincronizzazione …
    - soluzioni: modelli e metodi
  - **Fault** (fallimenti) in sistemi distribuiti
    - prob: fallimenti Indipendenti/dipendenti, coincidenti/sparsi
    - soluzioni: azioni di Recovering from failure
    - soluzioni: Architetture fault tolerance
- **I sistemi distribuiti sono tipicamente eterogenei**
  - Diversi per: sistema operativo, interfaccia di comunicazione, potenza, CPU, protocolli, posizione, etc.

Portable and handheld devices in a distributed system

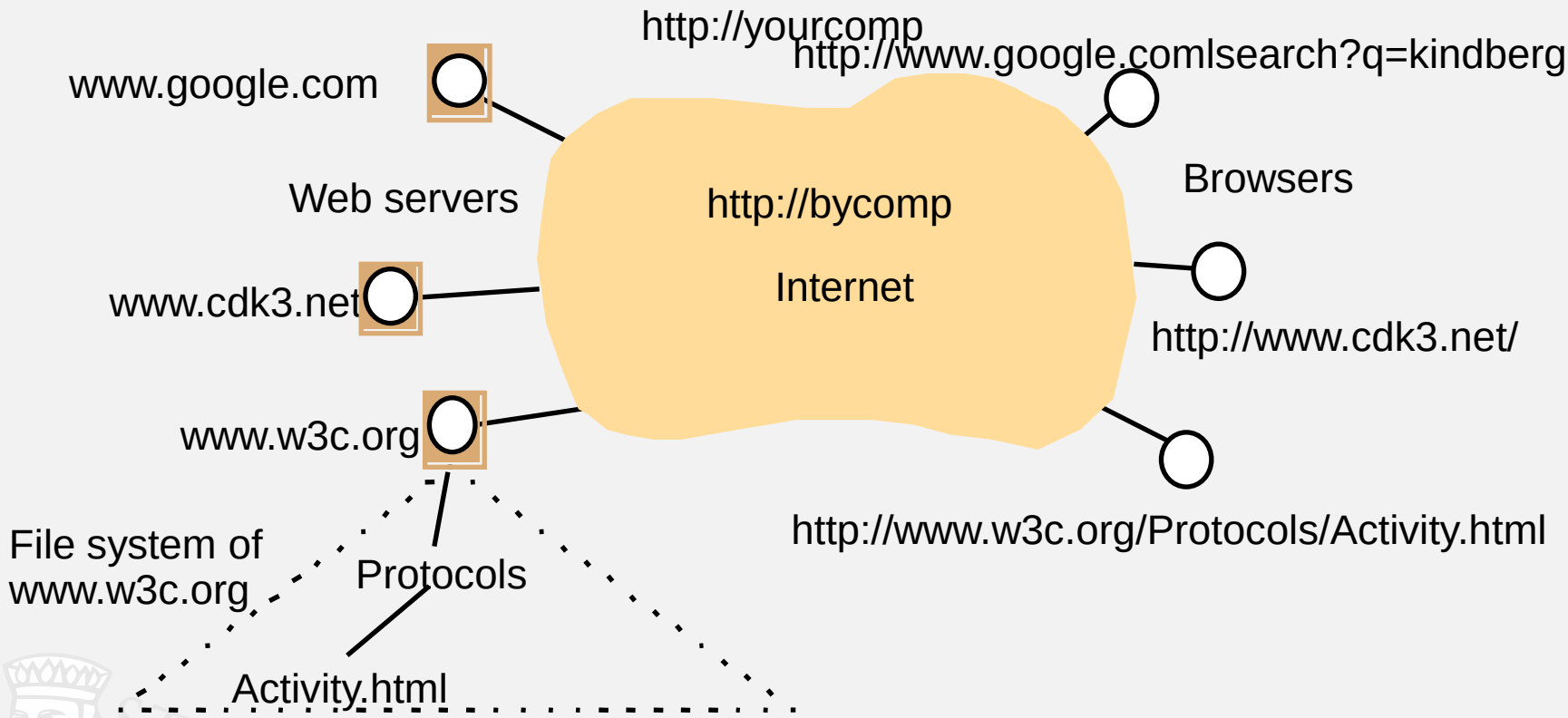


Intranet

- **LAN Services**:
  - Condivisione di File, Stampanti/Printer, …
  - Accesso a Web Services di vario tipo:
    - Accesso a db, ricerche, DNS, etc.
  - Utilizzo di WEB application per la gestione interna dell'azienda che ha realizzato l'Intranet

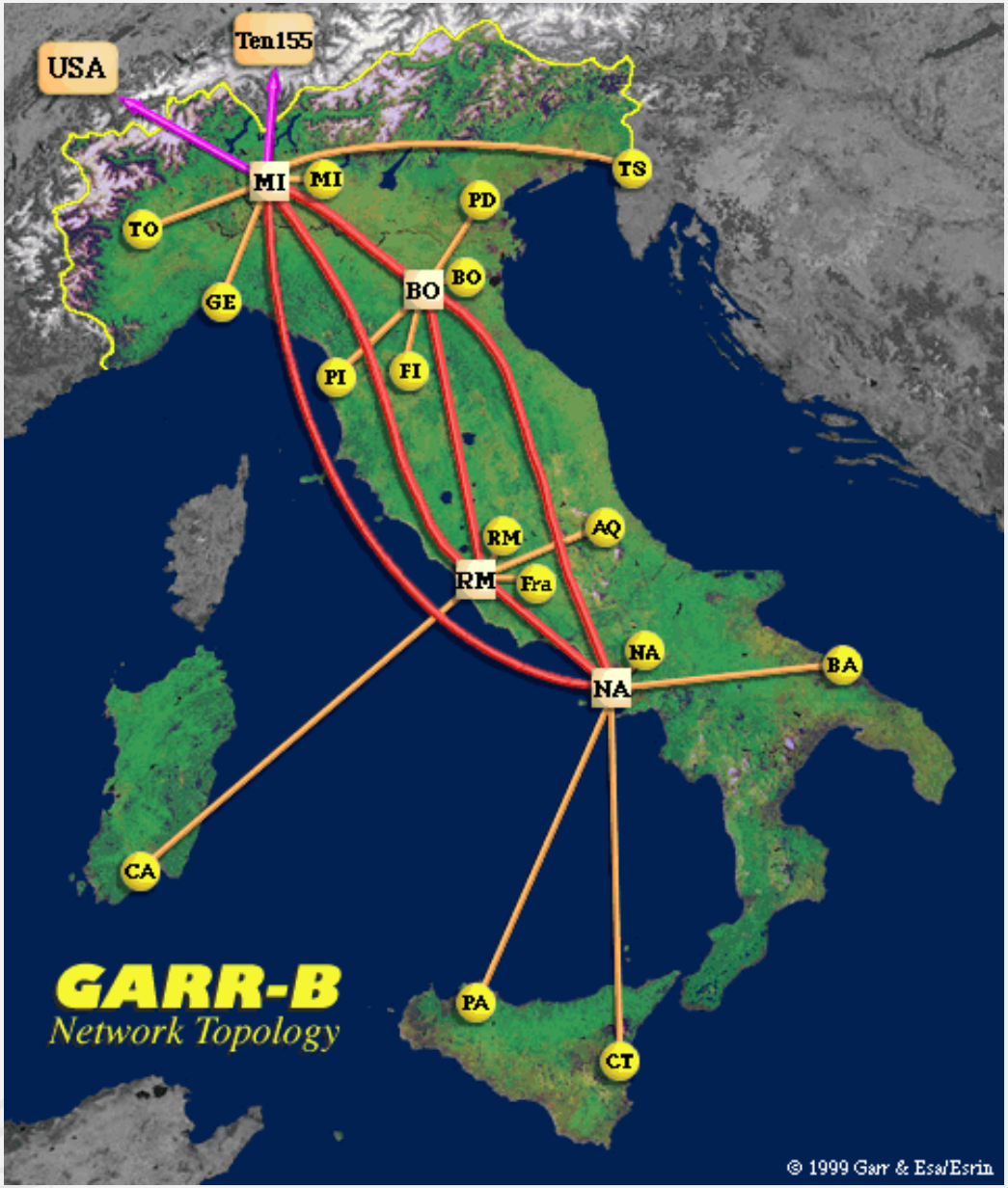
- **Connessioni da Intranet a Internet, gateway**:
  - Problemi di sicurezza
  - Firewall verso/da Internet
  - Connessione via provider: ADSL, mobile net, tc…

Web servers and web browsers



Calcolatori in rete: Networking

- Sistemi Distribuiti
- Connessioni fra calcolatori
- IP: indirizzi e porte
- Stack ISO OSI
- Protocolli
- Connessioni intercontinentali
- Sicurezza sulla rete


USA
Ten155
MI
MI
TS
PD
TO
BO
BO
GE
PI
FI
RM
AQ
RM
Fra
NA
BA
NA
CA
PA
CT
GARR-B
Network Topology
© 1999 Garr & Esa/Esrin

WorldCom's Global UUNET Internet network

For more information see www.uu.net/network/maps

NB: UUNET also has infrastructure within individual countries, which is not shown on this map.

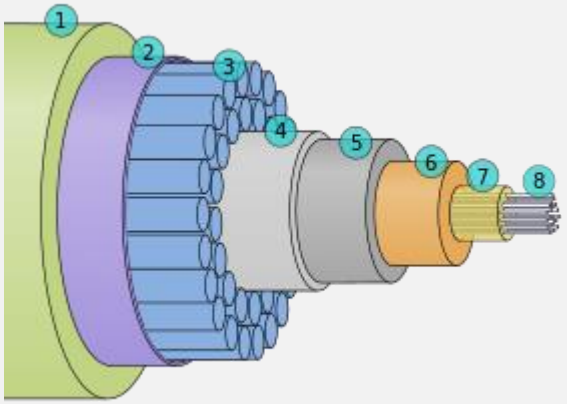
January 2001

WORLDCOM

Stesura cavi sottomarini

Un satellite geostazionario deve essere posto in orbita ad una altezza dalla superficie terrestre di 36 000 km.
Alla velocità della luce le onde radio impiegano circa un quarto di secondo a percorrerlo: tempo di latenza di mezzo secondo (andata e ritorno della risposta).

Un cavo idealmente posato tra Roma e New York ha una lunghezza di 6 600 km e causa un ritardo di soli 5 centesimi di secondo.



- Una sezione di un moderno cavo sottomarino per le telecomunicazioni.
  1 – Polietilene
  2 – nastro in Mylar
  3 – cavi d'acciaio
  4 – Aluminium water barrier
  5 – Policarbonato
  6 – Tubo di rame o d'alluminio
  7 – Vaselina
  8 – fibra ottica

Calcolatori in rete: Networking

- Sistemi Distribuiti
- Connessioni fra calcolatori
- IP: indirizzi e porte
- Stack ISO OSI
- Protocolli
- Connessioni intercontinentali
- Sicurezza sulla rete

Sicurezza

- **Controllo** degli accessi a dati e servizi
  - Consistenza dei dati
  - Firewall (indirizzi IP, protocolli e porte)
  - VPN (Virtual Private Network)
- **Sicurezza**
  - Registrazione e riconoscimento, autenticazione
  - Accesso ai servizi in modo controllato
  - SSO, single sign on

- **Gestione dei Fallimenti**
  - Detecting Failures
  - Masking Failures: resend, raid…
  - Recovering From Failures
  - Fault Tolerance, with Redundancy

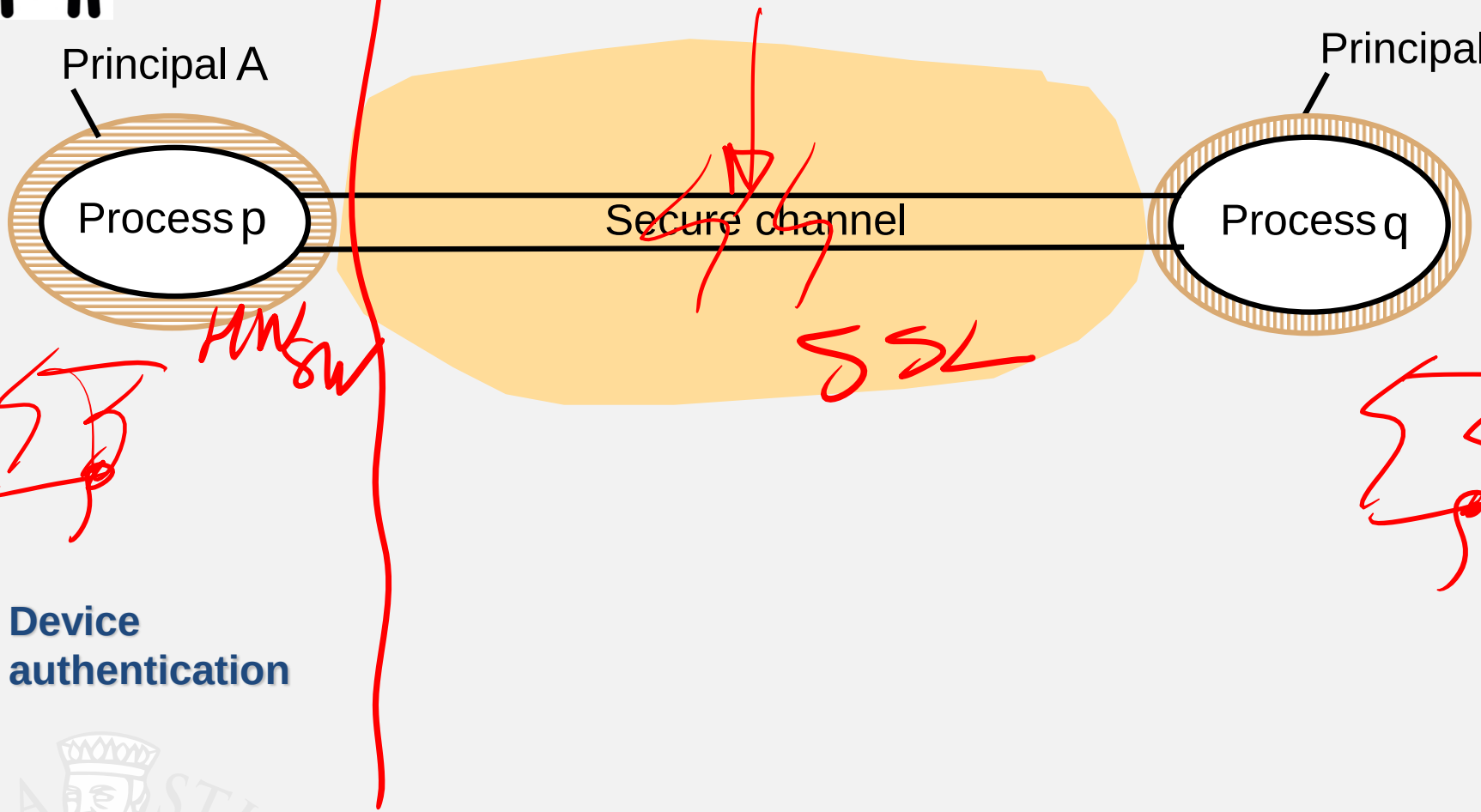
Security Model, controllo accesso

- **Authentication** of users
  - User ID and password, frequently saved as MD5 into database
  - User profiling with several information of the user
  - Production of certificates per la gestione delle chiavi

- **Certification** delle integrità di elementi come
  - devices (HW+SD), SW tools, etc.
  - data e content, dati e contenuti
  - Channel, canale

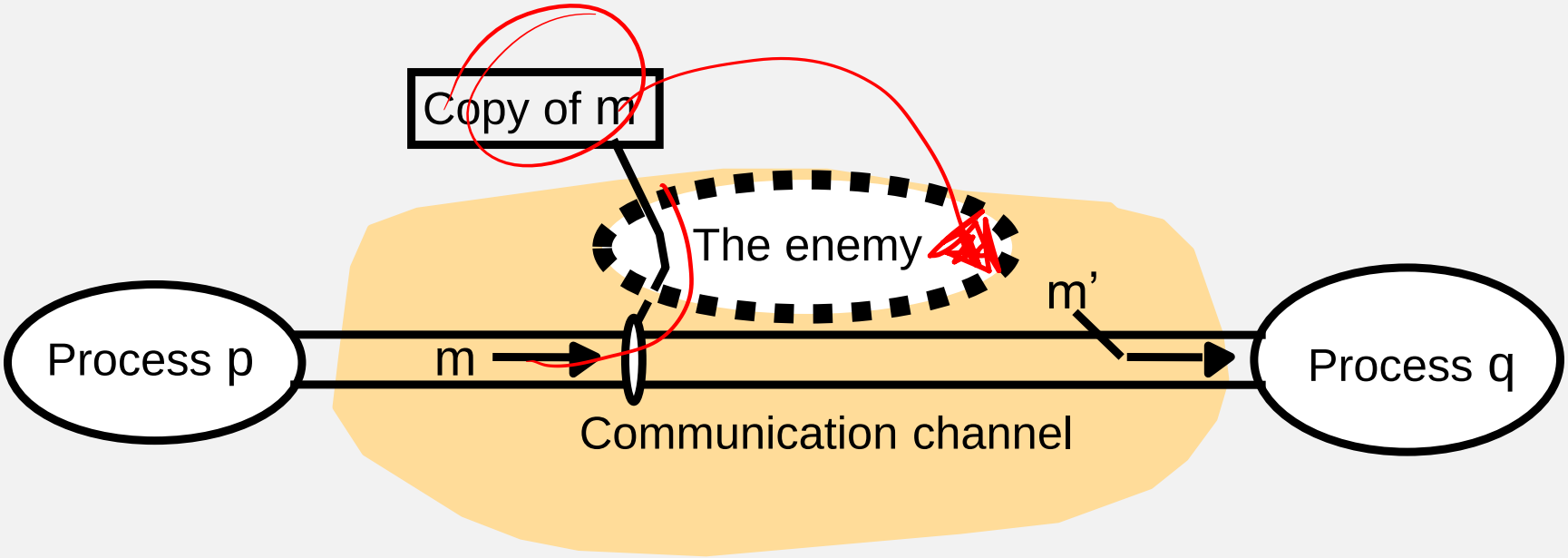
- **Protezione** di canale
  - Per esempio: SSL, HTTPS, SFTP
  - Certificati per la connessione

Secure channels



Device authentication

The Enemy, il nemico



Al momento della conncessione e creazione del canale protetto sia hanno i maggiori rischi

Security Model con diritti (rights)

- **Protezione** dei contenuti, degli oggetti:
  - Objects/components: software tools
  - Data: digital media, audiovisual, etc.

- **Controllo e prevenzione**
  - Controllo accesso all'oggetto:
    - Conditional Access Systems, CAS
  - Controllo dei diritti per l'uso dei servizi degli oggetti
    - Digital Rights Management, DRM

- **Sfruttamento controllato** di ogni loro caratteristica/ feature/ servizio
  - Conto corrente: leggo, muovo soldi, cambio parametri, etc.
  - Processo: start, stop, uso F1, uso F2, etc.
  - Contenuto digitale: play, print, copy, etc.

sommario

- *Dati e formati*
- *Comunicazione fra sistemi computerizzati*
- *Modelli di protezione e gestione*
- *Dati vs Metadati*
- *Social network e Smart City*
- *Motori di Ricerca, Crawling e NLP*
- *Data Mining*
- *Data Intelligence*

Modelli di protezione e gestione

- Copy protection
- I modelli di attacco in rete
- La crittografia, la firma digitale
- Conditional Access System
- Digital Rights Management, DRM
- Licenze per il DRM
- Watermark
- fingerprint

Copy Protection solution

- **Naturally digital items can be freely copied**
  - The copy is a feature of the operating system, file system
  - The operating system cannot be typically controlled
  - Microsoft is going to enforce more control on the Operating System

- **CP Solution**: prevents the Copy of digital content
  - Programs: hardware key
  - Holes in physical domain
  - Special formatting in CDs/DVDs
  - Etc.

Utilizzo della crittografia

- Encryption è il processo che codifica un messaggio in modo da nasconderne il contenuto
- Si basano sull'uso di parametri segreti chiamati *chiavi*
- Si dividono in due classi fondamentali
  - Chiavi segrete condivise *(secret-key)*
  - Coppie di chiavi pubblica/privata *(public-key)*
- Segretezza e integrità
- Autenticazione
- Firma digitale

Modelli di protezione e gestione

- Copy protection
- I modelli di attacco in rete
- La crittografia, la firma digitale
- Conditional Access System
- Digital Rights Management, DRM
- Licenze per il DRM
- Watermark
- fingerprint

Modello di sicurezza

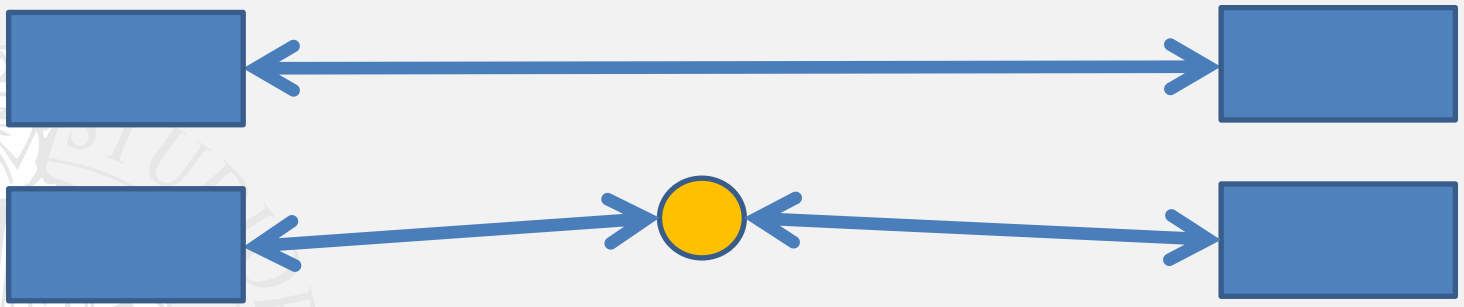
- Elementi chiave:
  - Processi
  - Risorse
  - Interfacce

♣Client
♣Principals
♣Rete

- I **processi** contengono (*encapsulation*) oggetti (linguaggi di programmazione) e altre **risorse** definite dal sistema
- I processi consentono l'accesso ai **client** attraverso le loro **interfacce**
- I **principals** (utenti o altri processi) possono essere autorizzati a operare su di una determinata risorsa
- I processi interagisco attraverso una **rete** condivisa tra molti utenti

Sistema distribuito: esempi di minacce

- In molte architetture di rete è semplice
  - creare un programma che ottenga copie dei messaggi trasmessi sulla rete
  - se i clients non provvedono ad autenticare il server, che un programma possa inserirsi è spacciarsi per il processo server richiesto cosicché il client trasmetta le informazioni ignaro della loro reale destinazione
  - che un programma esegua richieste fraudolente a scapito di un sistema insicuro, a seguito di violazione dei suoi dati



Classificazione delle minacce

- Lo scopo principale della sicurezza è consentire l'accesso alle risorse ed alle informazioni soltanto ai *principal* autorizzati
- Le minacce sono contenute in tre classi:
  - Leakeage
    - Accesso ad informazioni del sistema senza autorizzazione
  - Tampering
    - Modifica non-autorizzata delle informazioni
  - Vandalism
    - Interferenza al corretto funzionamento del sistema senza guadagno da parte di chi la attua

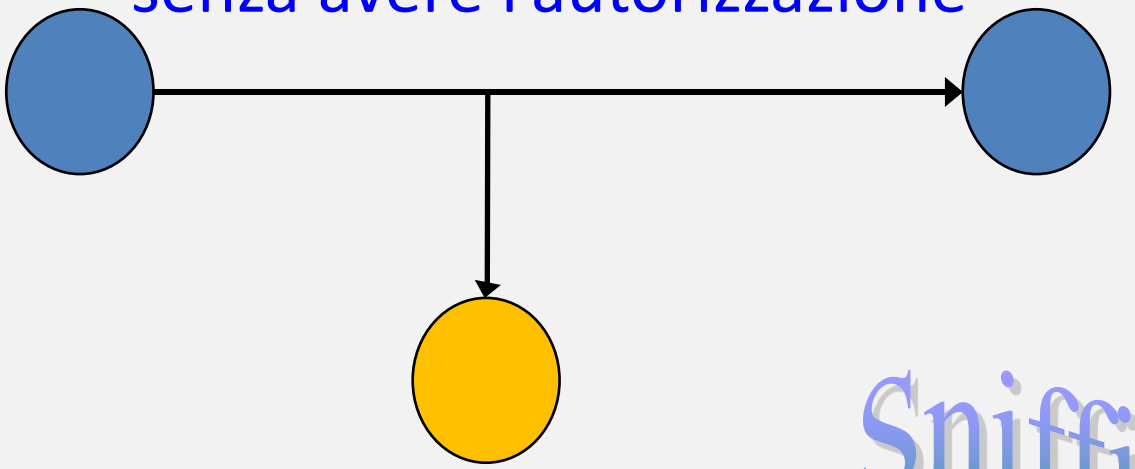
Tipologie di attacco al sistema

- Gli attacchi ad un sistema distribuito contano su
  - dall'ottenere o meno l'accesso a canali di comunicazione esistente
  - dal creare nuovi canali di comunicazione che figurano come autorizzate
- Si distinguono nelle seguenti tipologie:
  - **Eavesdropping** (sniffing)
  - **Masquerading** (spoofing)
  - **Message** tampering
  - **Replayng**
  - **Denial of service, DOS**

Eavesdropping

Ottenere copie dei messaggi

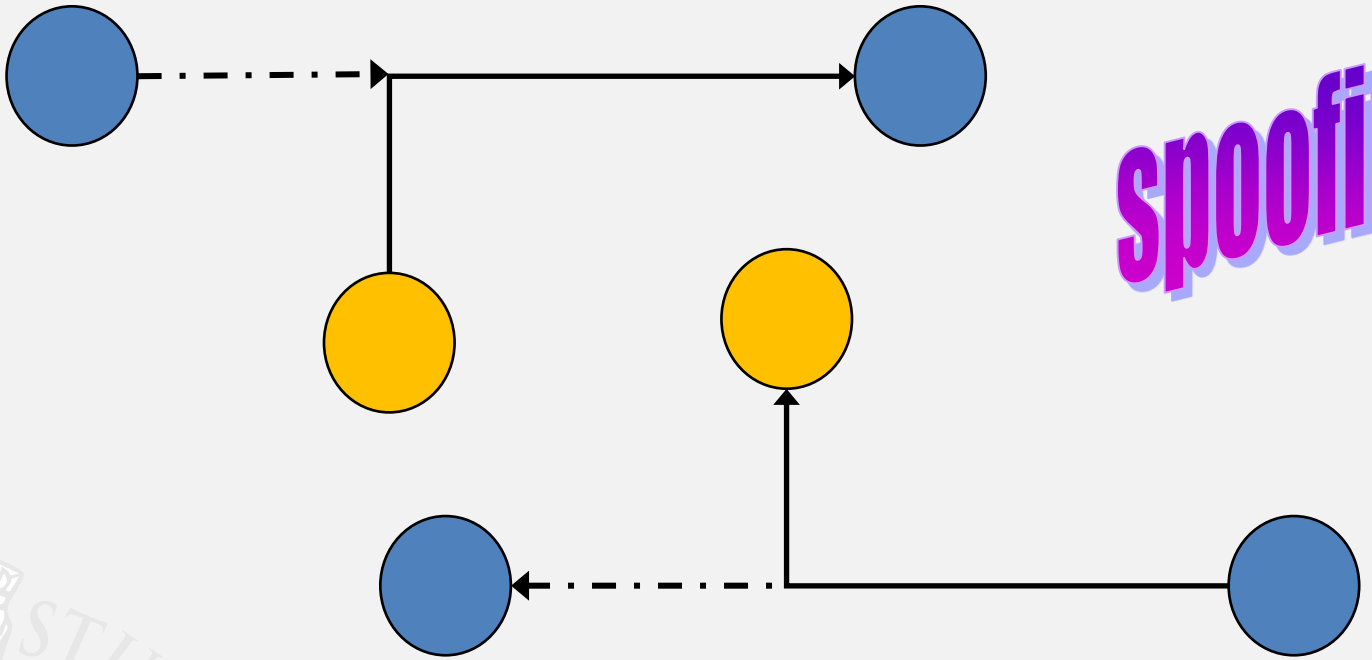
senza avere l'autorizzazione



Sniffing

Masquerading

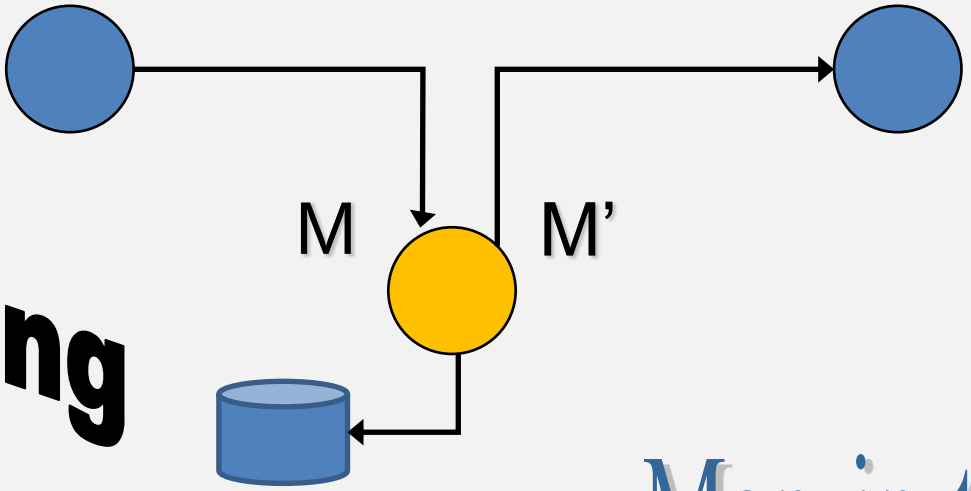
Inviare o ricevere messaggi usando l'identita di altri principals senza la loro autorizzazione



spoofing

Message tampering

Intercettare messaggi ed alterarne il contenuto prima di ritrasmetterli alla destinazione prevista

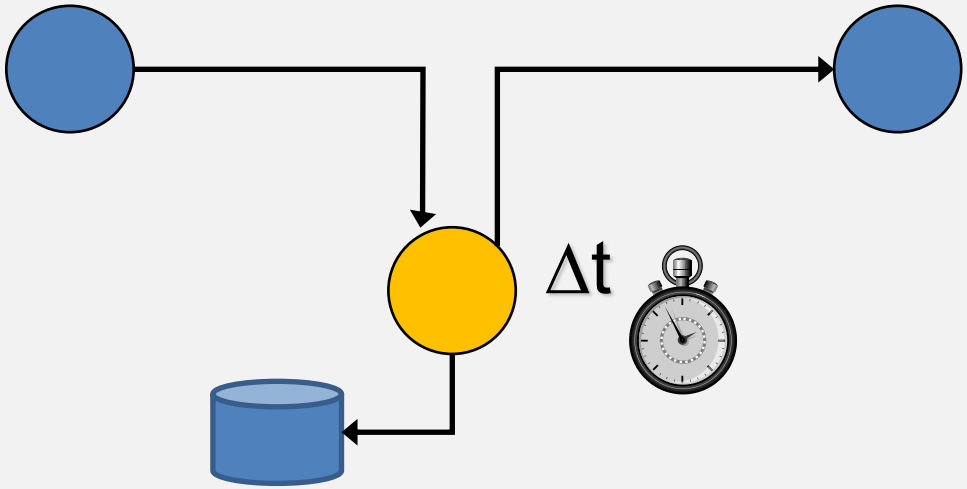


Hijacking

Man-in-the-middle

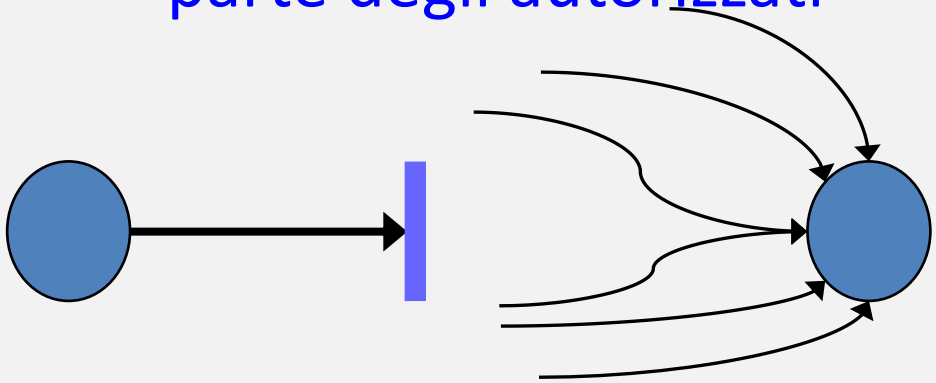
Replaying

Memorizzare messaggi intercettati e inviarli in ritardo rispetto alla loro reale origine



Denial of service, DOS

Saturare un canale di comunicazione o altre risorse con messaggi ripetuti in modo da negarne l'accesso da parte degli autorizzati



Pericoli potenziali ed effettivi

- Tutti questi sono pericoli soltanto in teoria, ma gli attacchi che possono andare a buon fine dipendono dal sistema di sicurezza
- Attacchi con successo contano sul fatto di individuare imperfezioni del sistema di sicurezza (*holes*)

- Negli attuali sistemi in uso sono comuni e ben evidenti

Minacce dal mobile code

- Si definisce *mobile code* un programma che viene caricato da un server remoto e viene eseguito in locale
  - Plugin, activeX, applet, ... javascript
- In questo contesto le risorse all'interno del sistema locale possono subire un attacco dal mobile code
- JVM da ad ogni applicazione mobile code un suo ambiente predefinito ed un security manager determina quali risorse sono disponibili
  - Le classi scaricate in memoria diversa dalle classi locali
  - Il bytecode viene controllato prima di essere eseguito
- Molti browser impediscono alle applet JAVA di accedere ai file locali, alle stampanti o alle socket del sistema

  - Si veda per esempio la definitiva cancellazioni dell'uso di Flash player nei browser

Information leakage

- Il problema della riservatezza delle informazioni non riguarda soltanto il contenuto dei messaggi scambiati

- Anche *osservare* che in un canale sorgente-destinazione il flusso di dati è rilevante può essere un informazione importante

Transazioni elettroniche sicure 1

- E-mail
  - Anche il protocollo dedicato allo scambio di posta non prevedeva originariamente un supporto alla sicurezza, ma la crittografia è adesso comune a molti applicativi

- Acquisto di beni e servizi
  - E-commerce prevede il selezionamento dei beni da acquistare ed il pagamento attraverso il Web

Transazioni elettroniche sicure 2

- Transazioni Bancarie
  - Le banche elettroniche offrono virtualmente i tipici servizi presenti allo sportello di una banca comune (estratto conto, bonifico bancario, domiciliazione utenze)

- Micro-transazioni
  - Altri servizi posso essere forniti dal Web tipo supporto alla *comunicazione vocale* o alla *videoconferenza*, pagabili a tempo tipicamente con importi bassi da non giustificare la sicurezza prevista per altre transazioni

Requisiti di sicurezza per le transazioni

- Autenticare il venditore al compratore, cosicché egli sia sicuro di essere in contatto con il server del venditore che gli interessa
- Tenere nascoste (ed inalterate) informazioni importanti di pagamento in modo che non cadano in mani sbagliate (i.e. carta di credito)
- Se i beni sono fruibili tramite download assicurare che il contenuto sia consegnato al compratore senza alterazioni e senza accesso da parte di altri

- In aggiunta a questi requisiti può essere necessario autenticare l'identità del client per fornirgli i diritti previsti all'accesso (e-banking)

non-ripudio

Progetto di sicurezza

- Progettare un sistema senza punti deboli è simile a scrivere un programma senza *bugs*
- La validazione formale è l'unica possibilità di garantire completa sicurezza
- La prova di validità è articolata in due fasi:
  - Si crea una lista di minacce possibili al sistema
  - Si mostra come ognuna di essi è gestita con successo dal sistema
- La dimostrazione può avere un aspetto informale anche se si predilige un approccio di tipo formale

Progettare un sistema di sicurezza è un esercizio nel bilanciare i costi in relazione alle minacce

Considerazioni worst-case

Le interfacce dei processi server sono necessariamente aperte

- Gli indirizzi degli host possono subire *spoofing*
- Una chiave segreta generata è sicura al momento della sua generazione, ma la sua segretezza diminuisce con il tempo
- Gli algoritmi sono disponibili ai responsabili di sicurezza come agli attacker
- Gli attacker dispongono spesso di grandi risorse di calcolo
- La base di fiducia (hardware e software) è spesso la causa delle debolezze

Tecnologie per la sicurezza

- Utilizzo della crittografia

- Certificati
- Controllo di accesso
- Credenziali
- Firewall

Modelli di protezione e gestione

- Copy protection
- I modelli di attacco in rete
- La crittografia, la firma digitale
- Conditional Access System
- Digital Rights Management, DRM
- Licenze per il DRM
- Watermark
- fingerprint

Utilizzo della crittografia

- Encryption è il processo che codifica un messaggio in modo da nasconderne il contenuto
- Si basano sull'uso di parametri segreti chiamati *chiavi*
- Si dividono in due classi fondamentali
  - Chiavi segrete condivise *(secret-key)*
  - Coppie di chiavi pubblica/privata *(public-key)*
- Segretezza e integrità
- Autenticazione
- Firma digitale

Algortimi di crittografia

Un messaggio si dice criptato quando il mittente applica alcune regole per trasformare il testo originale (*plaintext*) in un altro testo (*ciphertext*)

$$E(K_1, M) = \{M\}_K$$

- Il ricevente deve conoscere la trasformazione inversa per ritrasformare il *ciphertext* nel messaggio originale

$$D(K_2, \{M\}_K) = M$$

$K_1 = K_2$ **simmetrico**

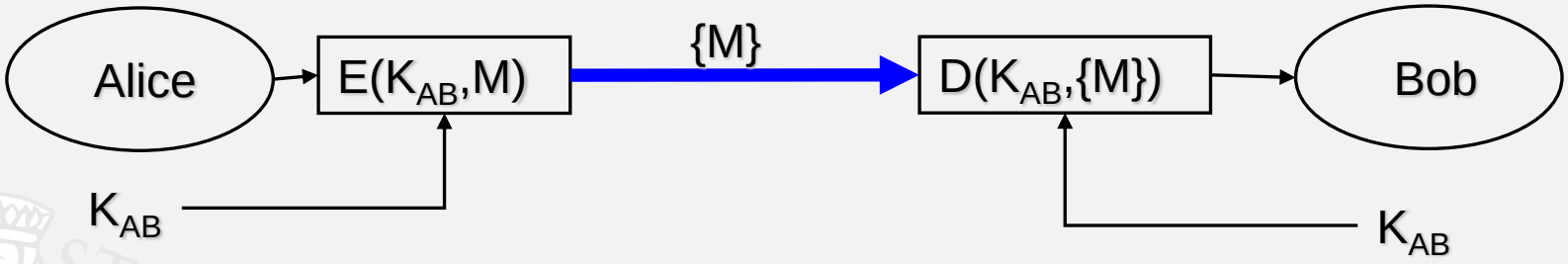
$K_1 != K_2$ **asimmetrico**

Scenario 1: secret communication

Alice vuole inviare alcune informazioni segretamente a Bob

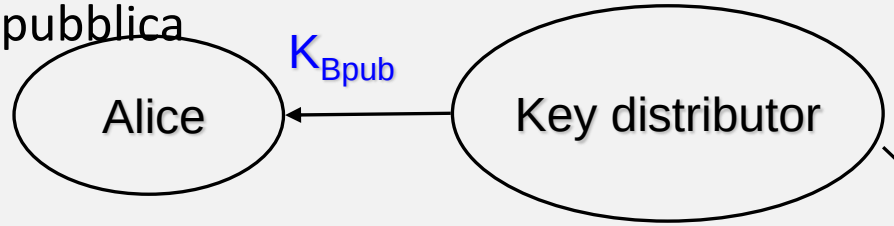
$\{M\}=E(K_{AB},M)$

- Alice e Bob conoscono entrambi la chiave segreta K_{AB}
- La comunicazione è segreta finchè K_{AB} non è compromessa

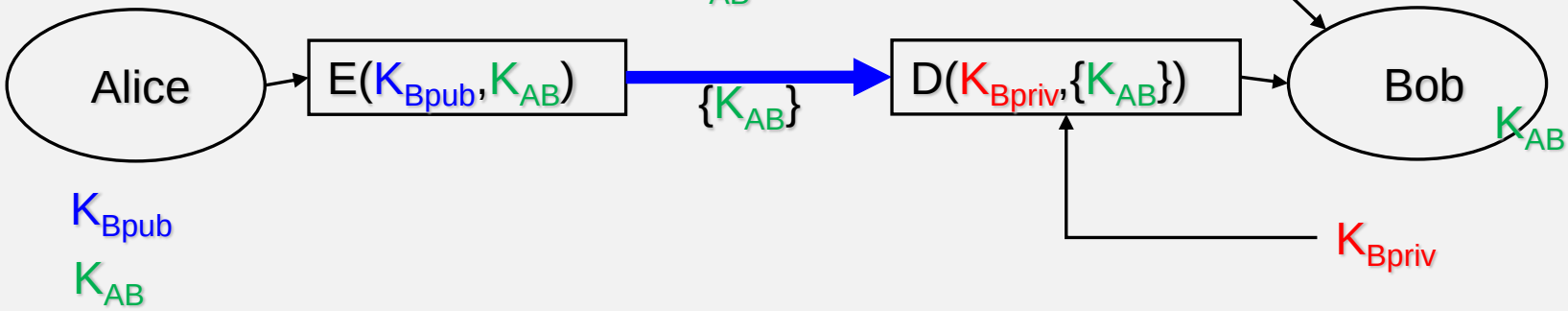


Scenario 3: authenticated with public-key

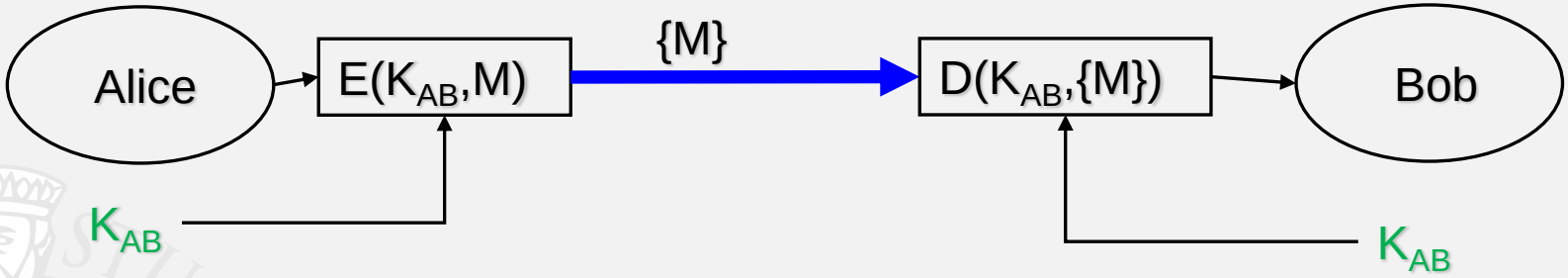
- Richiesta chiave pubblica



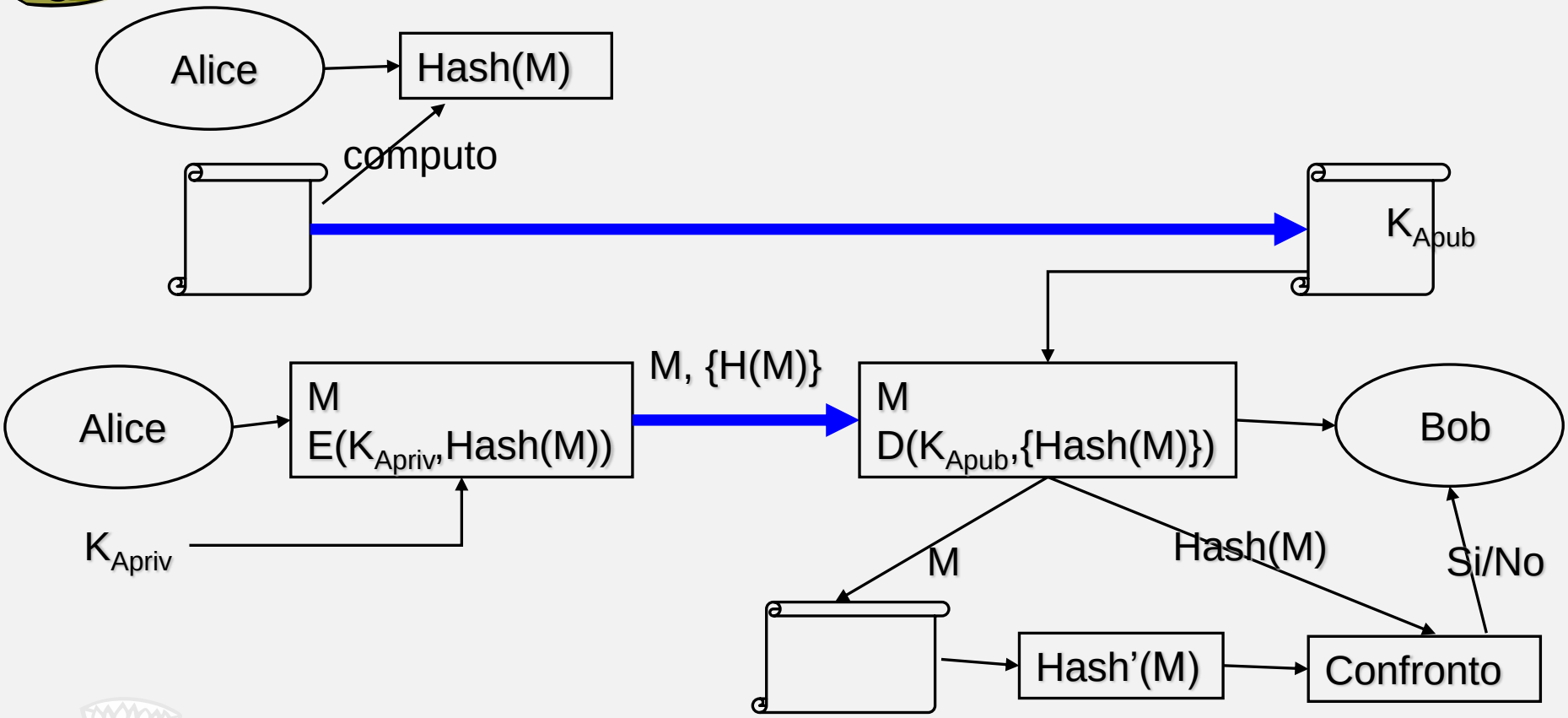
- Determinazione chiave di sessione K_{AB}



- Cumunicazione sicura



Scenario 4: digital signature



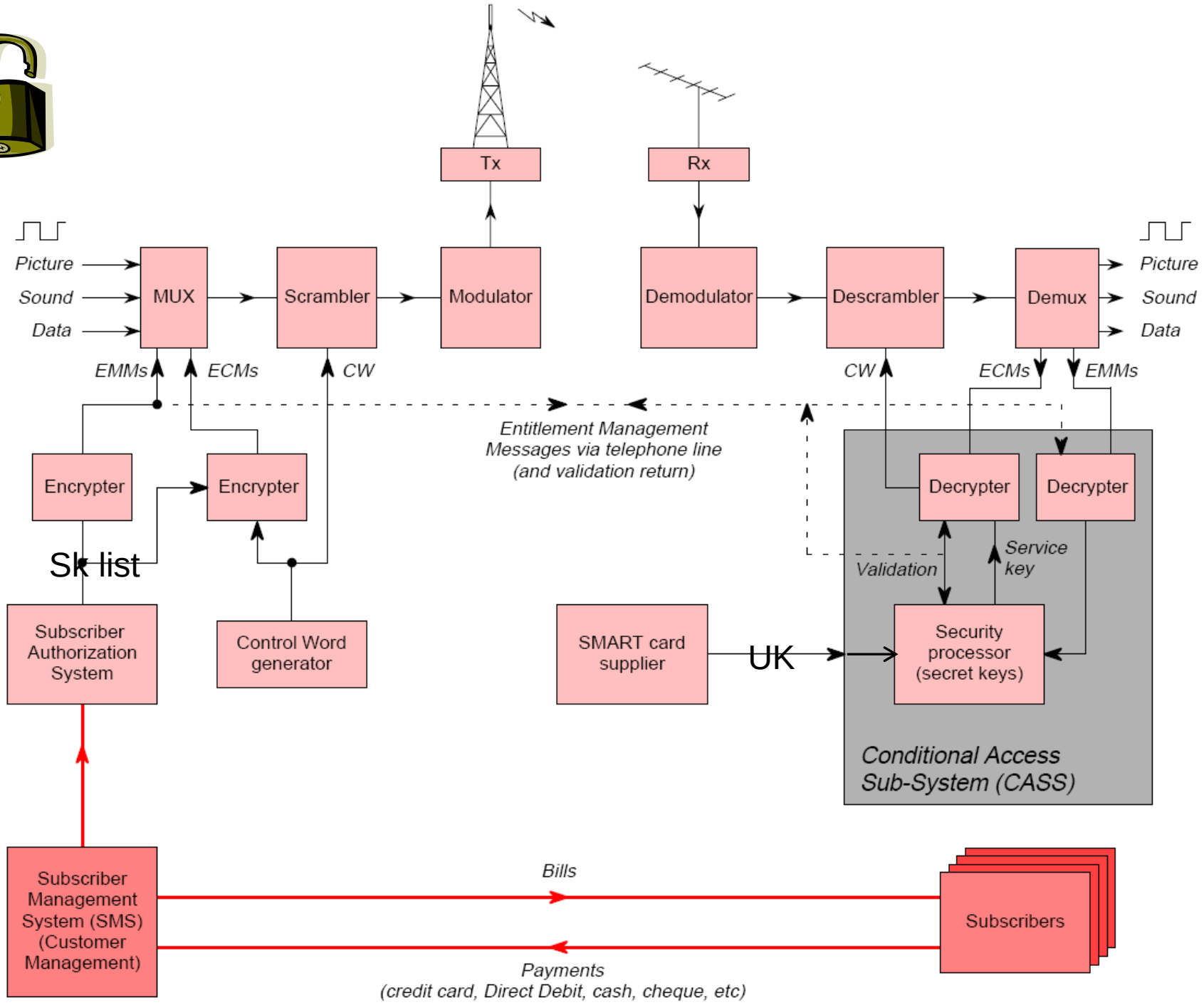
Modelli di protezione e gestione

- Copy protection
- I modelli di attacco in rete
- La crittografia, la firma digitale
- Conditional Access System
- Digital Rights Management, DRM
- Licenze per il DRM
- Watermark
- fingerprint

CAS: Conditional Access Systems

- **Systems that controls the access to the content**
  - Typically used on streaming towards STB/Decoders
  - Copy is assumed not possible since the content is not stored locally and neither accessible to the final user.

- **For PC:**
  - Partially suitable for open platforms such as PC
    - On PC: SSL, HTTPS, etc.
  - Temporary storage of smaller content on the disk, may be encrypted
- **For STB:**
  - The most interesting and diffuse solution


Tx
Rx
Picture
Sound
Data
MUX
Scrambler
Modulator
Demodulator
Descrambler
Demux
Picture
Sound
Data
EMMs
ECMs
CW
CW
ECMs
EMMs
Entitlement Management
Messages via telephone line
(and validation return)
Encrypter
Encrypter
Decrypter
Decrypter
Sk list
Validation
Service
key
Subscriber
Authorization
System
Control Word
generator
SMART card
supplier
UK
Security
processor
(secret keys)
Conditional Access
Sub-System (CASS)
Subscriber
Management
System (SMS)
(Customer
Management)
Bills
Subscribers
Payments
(credit card, Direct Debit, cash, cheque, etc)

Modelli di protezione e gestione

- Copy protection
- I modelli di attacco in rete
- La crittografia, la firma digitale
- Conditional Access System
- Digital Rights Management, DRM
- Licenze per il DRM
- Watermark
- fingerprint

Rights Management

- **DRM**: Digital Rights Management
  - general term many times abused, confused, ...

- Management of Digital Rights
  - Limited to the management of rights of digital content ? → NO!!!
- **Digital Management of Rights → YES!!!**
  - More correct and reasonable
  - Management of both rights for original *works* and related *manifestations*, digital *resources*, etc.
  - in many solutions DRM is not intended in this way

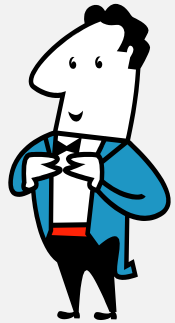
Aim of …… Digital Rights Management

- To allow exploiting digital content functionalities (rights) in a controlled manner
  - To who has been **authenticated/certified**
  - To do what (are the rights) is defined in a formal **license**
  - **Verifying/Control/Supervise** if the above conditions and others are respected
  - By using technologies to **protect content** (e.g., encryption, fingerprint, watermark, etc.)

- Cons:
  - Registration of users (in some case can be relaxed)
  - Authentic. of users and/or tools/terminal/devices
  - Control of users
- *It has to be supported by a set of additional technical solutions*

Simple protection with Key sending

Selling Server



License Production

License distribution and player/device verification and supervision

Ask for the License

Content deprotection and rights exploitation

Content Packaging, Protection

content

Distribution Server

Get the content

Production and distribution

Player device

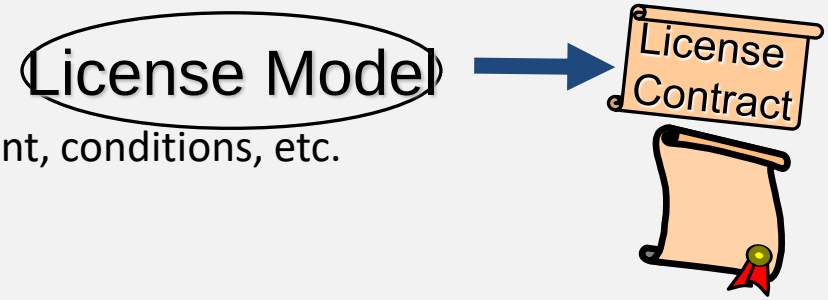
Content Elements of the package

- **Metadata:**
  - Identification information, unique ID, distributor ID, etc.
  - Classification information also for indexing: Dublin core, etc.
  - Semantic Descriptors, MPEG-7, for indexing, etc.
  - References to Owner, to distributor, etc.
  - Etc.

- **Digital Resources:**
  - Any digital information: images, doc, txt, video, game, application, file, audio, etc.
  - Hierarchy of digital resources

- **Protection Information:**
  - What has to be done to access at a given information/resource
  - Tools used, their parameters, etc.

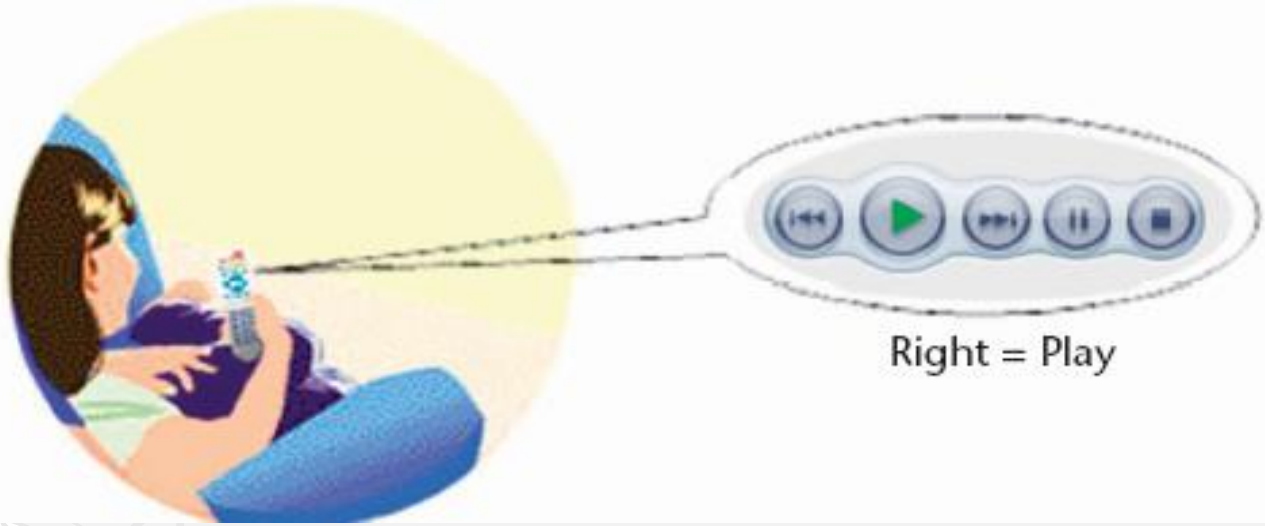
- **License:**
  - Which rights are provided, who is the recipient, conditions, etc.



An example of statement

Condition = November 2003

Resource = Ocean Wilds



Right = Play

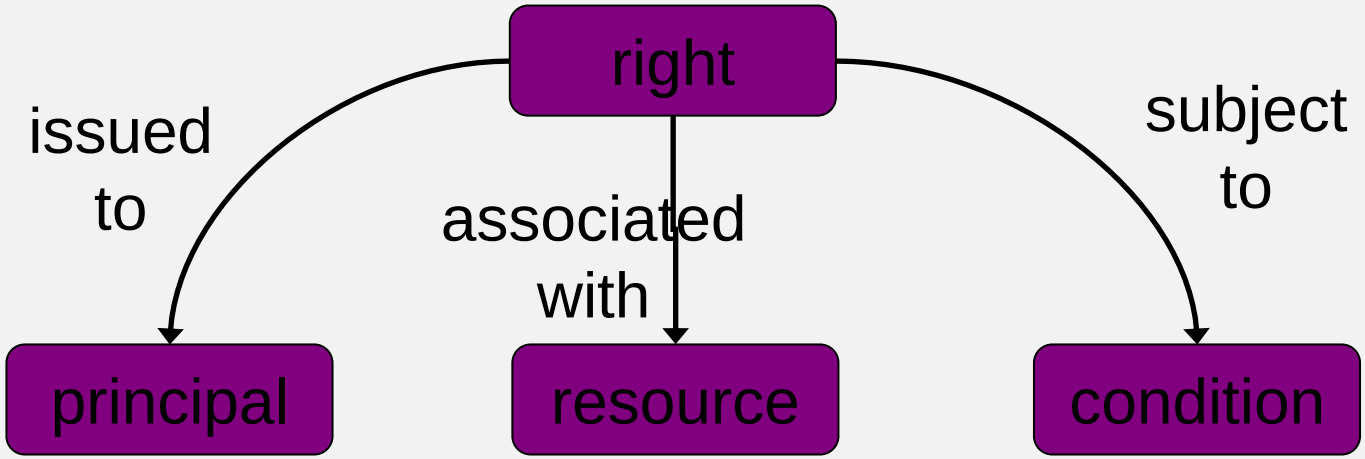
- Rosy can Play 3 times the Ocean Wilds in November 2003.

MPEG-21 — REL, Rights Expression Language

ISO/IEC JTC1/SC29 WG11 MPEG MOVING PICTURE EXPERTS GROUP

- **REL is** a machine-readable language, XML
  - to declare rights and permissions
  - uses terms defined in the Rights Data Dictionary, RDD
- **REL allows to define licenses** that give specific permissions to Users to perform certain actions on certain resources, given that certain conditions are met
  - Grants can also allow Users to delegate authority to others
- **Systems and device have to**
  - parse and validate the REL formalizations
  - check permissions before any further action is done
- **REL licenses** are wrapped into MPEG-21 Digital Items when the object is governed
- **MPEG-21 DID parser** is responsible for discovering and identifying where to gather licenses

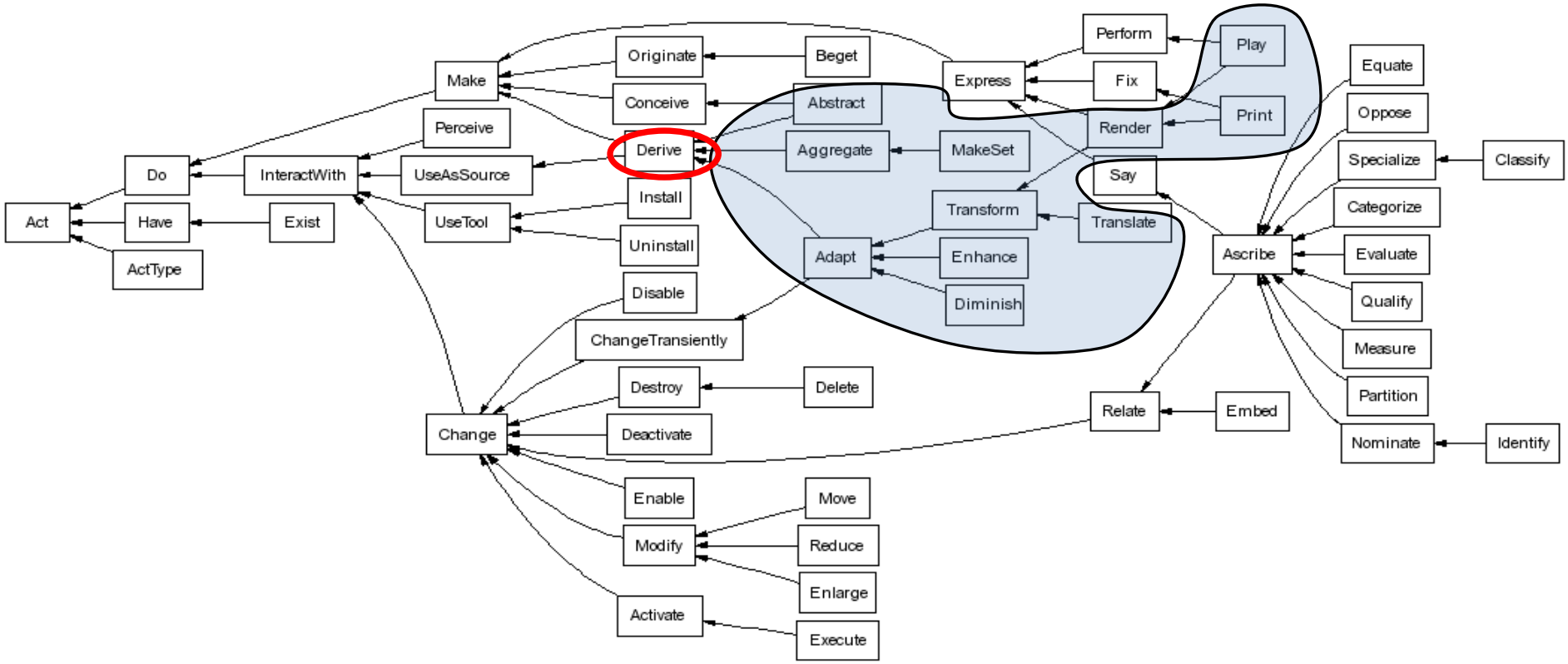
REL data model



- REL grant formalization consists of
  - principal to whom grant is issued
  - rights the grant specifies
  - resource to which right in grant applies
  - condition to be met before grant can be exercised

RDDOnto (Garcia, Delgado, 2007)

- Example: *Act* hierarchy.



Licences

- Relazioni fra rights!
  - Relazioni di Implicazione
- Relazioni fra licenze
  - Licenza di emissione di licenze per certi diritti
- License di Dominio
- License con diritti condizionati:
  - tipologia di utenti, location, tempo, etc…
- …

Creative Commons, CC

- Nel commercio elettronico e' necessario includere anche alcune slide relative alle licenze CC.
- Fino ad ora abbiamo visto tecnologia al servizio della protezione della proprieta' intellettuale
- CC mette a disposizione degli strumenti, dei formalismi legali che possono essere o meno adottati da chi pubblica i propri contenuti
- Questi sono license formalizzate:
  - **Struttura della Licenza CC:**
    - Legal Code: testo legale che la descrive
    - Commons Deed: short description della licenza
    - Digital Code: metadati da associare

    (see S. Aliprandi, 2005, 2006…)

http://www.copyleft-italia.it/cc/brochureCCv2.pdf

Creative Commons, CC

- ***Le licenze CC***
  - Nate in USA
  - Sono state adattate alla legislazione nazionale di diversi stati e anche a livello europeo
    - specialmente per contenuti generati dagli utenti e in ambito culturale
    - Questo permette in un certo qual modo di avere una trascodifica fra le questioni legali nazionali e quelle di altre nazioni, ma solo per certe questioni.
  - Licenze CC ipotizzano uno share, copyleft
- E' stato fatta una codifica delle licenze CC in MPEG-21 REL
  - Non e' vero l'opposto, tutte le licenze che si possono formalizzare in MPEG-21 non hanno una controparte in CC

Ogni Licenza chiede che il Licensiatario

- Ottenga il tuo permesso per fare una qualsiasi delle cose che
  - hai scelto di limitare con la licenza,
  - for example: limitare gli usi commerciali o quelli di opera derivata
- Mantenga l'indicazione di diritto di autore intatta su tutte le copie del tuo lavoro
- Faccia un link alla tua licenza dalle copie dell'opera
- Non alteri i termini della licenza
- Non usi mezzi tecnologici per impedire ad altri licenziatari di esercitare uno qualsiasi degli usi consentiti dalla legge

Ogni licenza CC permette le seguenti azioni a patto che:

i licenziatari rispettino le condizioni della licenza CC assegnata:

- Possono
  - Copiare l'opera;
  - Distribuire l'opera;
  - Comunicare al pubblico, rappresentare, esegure, recitare o esporre l'opera in pubblico, ivi inclusa la trasmissione audio digitale dell'opera;
  - Cambiare il formato dell'opera.

Alcuni marker CC

- **Liberta' per l'utente**

  - Sei libero di distribuire, comunicare, rappresentare, eseguire, recitare o esporre l'opera in pubblico, ivi inclusa la trasmissione audio digitale dell'opera;

  - Sei libero di modificare questa opera

- **Condizioni di uso**

  - Devi riconoscere la paternita' di questa opera
  - Per esempio citando e riportanto un link alla sorgente

Licenze Creative Commons

- Offrono 6 diverse articolazioni
  - per artisti, giornalisti, docenti, istituzioni e, in genere, creatori che desiderino **condividere in maniera ampia** le proprie opere secondo il modello **"alcuni diritti riservati"**.

- **Altre condizioni d'uso, il detentore dei diritti puo'**

  - non autorizzare a priori **usi prevalentemente commerciali** dell'opera (opzione *Non commerciale*, acronimo inglese: *NC*)

  - non autorizzare la creazione di **opere derivate** (*Non opere derivate*, acronimo: *ND*); no extract, no aggregate, ..

  - Imporre di rilasciarle **con la stessa licenza dell'opera originaria** (*Condividi allo stesso modo*, acronimo: *SA*, da "Share-Alike").

Le combinazioni di queste scelte generano 6 licenze CC, disponibili anche in versione italiana, come descritto in seguito!

2 of the 6 CC licenses, 1/3

- **Attribution Non-commercial No Derivatives (by-nc-nd)**

  - The most restrictive of our six main licenses, allowing redistribution.
  - This license is often called the “free advertising” license
  - *it allows others to download your works and share them with others as long as they mention you and link back to you*,
  - they can’t change them in any way or use them commercially

- **Attribution Non-commercial Share Alike (by-nc-sa)**

  - *Let others remix, tweak, and build upon your work non-commercially, as long as they credit you and license their new creations under the identical terms.*
  - Others can download and redistribute your work just like the by-nc-nd license, but they can also translate, make remixes, and produce new stories based on your work.
  - All new work based on yours will carry the same license, so any derivatives will also be non-commercial in nature.

2 of the 6 CC licenses, 2/3

- **Attribution Non-commercial (by-nc)**
  - Let others remix, tweak, and build upon your work non-commercially, and their new works must also acknowledge you and be non-commercial,
  - they don't have to license their derivative works on the same terms.

- **Attribution No Derivatives (by-nd)**
  - allows for redistribution, commercial and non-commercial,
  - as long as it is passed along unchanged and in whole, with credit to you

2 of the 6 CC licenses, 3/3

- **Attribution Share Alike (by-sa)**
  - *lets others remix, tweak, and build upon your work even for commercial reasons, as long as they credit you and license their new creations under the identical terms.*
  - This license is often compared to open source software licenses.
  - All new works based on yours will carry the same license, so any derivatives will also allow commercial use.

- **Attribution (by)**
  - *lets others distribute, remix, tweak, and build upon your work, even commercially, as long as they credit you for the original creation.*
  - This is the most accommodating of licenses offered, in terms of what others can do with your works licensed under Attribution.

Part 1a: sistemi di protezione

- Distribution models
- Terminologies
- Business Models & Value Chain
- Copy protection
- Conditional Access Systems
- Digital Rights Management
- Content Modeling and Packaging
- Licensing and content distribution
- Creative Commons Licensing
- Composition of Licenses, data aggregation

Licenses and More..

- **Licenses**: MPEG-21, ODRL, XACML, Xrml, etc..
  - Suitable for media, Unsuitable for data
- **Data Licenses**: CC, ODC, OGL, IODL
  - Mainly open data and declinations
  - **Permissions**: derivative, commercialize, derivative...
  - **Restrictions/dutie**s: attribution, notice, ...
- **Getting Composing Data Set → Licences Composition is needed**
  - **See www.disit.org/6877 extension**
- Formal models to **grant rights**
- Techniques for **right enforcement/verification**
  - **Almost missing on RDF stores**

Composing Licenses

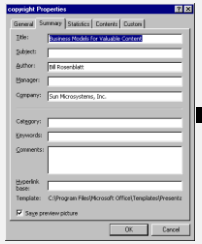
DISIT extension

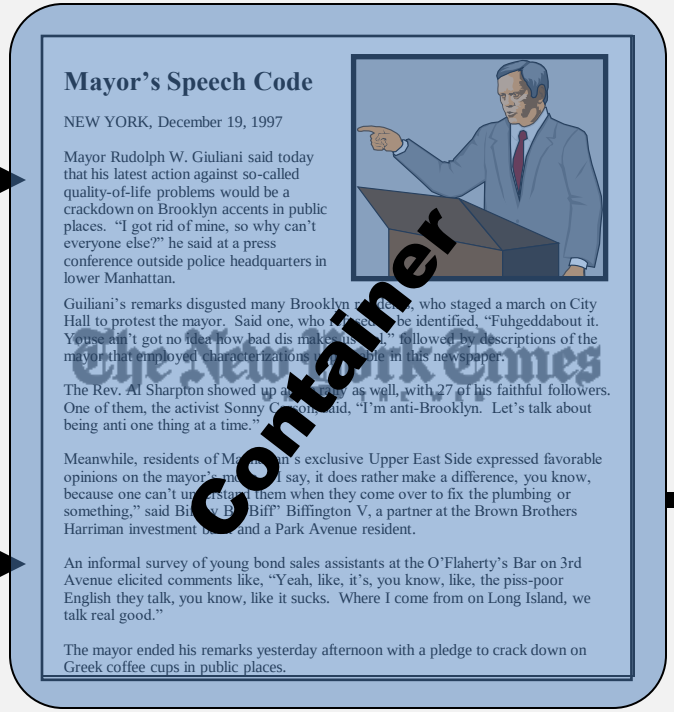
First License \ Second License	CC0	CC-PDM	CC-BY-ND	CC-BY-NC-ND	CC-BY	CC-BY-SA	CC-BY-NC	CC-BY-NC-SA	ODC-PDDL	ODC-BY	ODC-ODbL	OGL 2.0	OS OpenData
CC0	No restrictions	No restrictions	-	-	CC-BY	CC-BY-SA	CC-BY-NC	CC-BY-NC-SA	No restrictions	ODC-BY	ODC-ODbL	OGL 2.0	OS OpenData
CC-PDM	No restrictions	No restrictions	-	-	CC-BY	CC-BY-SA	CC-BY-NC	CC-BY-NC-SA	No restrictions	ODC-BY	ODC-ODbL	OGL 2.0	OS OpenData
CC-BY-ND	-	-	-	-	-	-	-	-	-	-	-	-	-
CC-BY-NC-ND	-	-	-	-	-	-	-	-	-	-	-	-	-
CC-BY	CC-BY	CC-BY	-	-	CC-BY	CC-BY-SA	CC-BY-NC	CC-BY-NC-SA	CC-BY	CC-BY	ODC-ODbL	CC-BY	OS OpenData
CC-BY-SA	CC-BY-SA	CC-BY-SA	-	-	CC-BY-SA	CC-BY-SA	-	-	CC-BY-SA	CC-BY-SA	ODC-ODbL	CC-BY-SA	CC-BY-SA
CC-BY-NC	CC-BY-NC	CC-BY-NC	-	-	CC-BY-NC	-	CC-BY-NC	CC-BY-NC-SA	CC-BY-NC	CC-BY-NC	-	CC-BY-NC	--
CC-BY-NC-SA	CC-BY-NC-SA	CC-BY-NC-SA	-	-	CC-BY-NC-SA	-	CC-BY-NC-SA	CC-BY-NC-SA	CC-BY-NC-SA	CC-BY-NC-SA	-	CC-BY-NC-SA	--
ODC-PDDL	No restrictions	No restrictions	-	-	CC-BY	CC-BY-SA	CC-BY-NC	CC-BY-NC-SA	No restrictions	ODC-BY	ODC-ODbL	OGL 2.0	OS OpenData
ODC-BY	ODC-BY	ODC-BY	-	-	ODC-BY	CC-BY-SA	CC-BY-NC	CC-BY-NC-SA	ODC-BY	ODC-BY	ODC-ODbL	ODC-BY	OS OpenData
ODC-ODbL	ODC-ODbL	ODC-ODbL	-	-	ODC-ODbL	ODC-ODbL	-	--	ODC-ODbL	ODC-ODbL	ODC-ODbL	ODC-ODbL	ODC-ODbL
OGL 2.0	OGL 2.0	OGL 2.0	-	-	CC-BY	CC-BY-SA	CC-BY-NC	CC-BY-NC-SA	OGL 2.0	ODC-BY	ODC-ODbL	OGL 2.0	OS OpenData
OS OpenData	OS OpenData	OS OpenData	-	-	OS OpenData	CC-BY-SA	--	--	OS OpenData	OS OpenData	ODC-ODbL	OS OpenData	OS OpenData

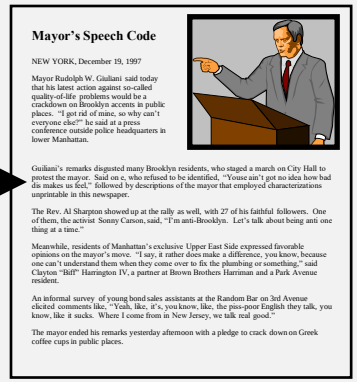
Modelli di protezione e gestione

- Copy protection
- I modelli di attacco in rete
- La crittografia, la firma digitale
- Conditional Access System
- Digital Rights Management, DRM
- Licenze per il DRM
- Watermark
- fingerprint

Watermark: Persistent Association







- Watermark
- Encryption

Watermark

What is the watermark (also called steganographic)

- a technology to embed an information in the content: image, video, text, audio, etc

- Which information is watermarked:
  - Object ID
  - Owner ID
  - Distributor ID
  - Eventual coding of the license (governed object)
  - Etc.
- Once read it can be used
  - to hide IDs to demonstrate the ownership of the content
  - To hide a sort of license

Watermark features

- Transparency: visible, invisible
- Robustness: tolerance to attacks
  - Adaptation, DA-AD
- Capacity: amount of information embedded
- Blindness: reference to the source image Hidden or visible
- Removable or not:
  - when it is separable from the digital resource obtaining the original digital resource
- Single of multiple:
  - when more than one Watermark is present
- Readable
  - by all or only by the owner: when there is not need to have a special key/parameters to read it
  - with an absolute certainty or with some statistical confidence
  - To be estimated during streaming
- Etc.

Usage of Watermark

- **Content Producers/Distributor**
  - watermark the content (images, audio, video, etc.)
- **Content integrators and distributors**
  - are informed and may add one more watermark with their code or reference
- **End users**
  - are not aware about that, if it is undetectable is easy
- **The terminal**
  - may or may not be capable to read it

Fingerprint and descriptors

What is the Fingerprint

- It is an ID-code estimated on the digital content or resource that present in practical an high probability to be unique for that content with respect to other similar content
- To make the recognition of the digital content possible
  - Indexing into the database

Fingerprint as a high level content descriptor

- Resources
  - Audio: Rhythm, tonality, duration, genre, etc.
  - Video: number of scenes, description of the scene, etc.
  - Text: main keywords, summary, topics, etc.
- Collected as MPEG-7 descriptors
- Vectors of those features, etc.
- Independent on the resolution, format, etc.
- May be Computationally intensive
- Etc.

Fingerprint Features

Features:

- Never included with the content if its aim is the usage for content protection
- Included in the content (package) only if it is used as content descriptor
- Robust to adaptation processing: Scaling: time, space, color, etc.
- Short and concise
- Repeatable
- Light to be estimated
  - estimable during streaming, on the basis of a short duration of the content streaming
- Robust to eventual watermark addition
- Etc.

- **Typically more computational intensive** with respect to WM:
  - The WM code is read/extracted from the content
  - The FP code has to be estimated from the content

sommario

- *Dati e formati*
- *Comunicazione fra sistemi computerizzati*
- *Modelli di protezione e gestione*
- *Dati vs Metadati*
- *Social network e Smart City*
- *Motori di Ricerca, Crawling e NLP*
- *Data Mining*
- *Data Intelligence*

Dati vs Metadati

- Dati, open data
- Dati privati, dati pubblici
- Linked Open Data
- Metadati e indicizzazione
- Dati e smart city
- User Profile

i dati: una classificazione

- **Dati pubblici**:
  - web, broadcasting (radio, tv), etc.
- **Dati istituzionali e/o privati: archivi, etc.**
  - Tipicamente accessibili solo per service provider: banche, uffici GOV, poste, etc.
- **Dati privati per servizi privati: archivi, etc.**
  - Se noti possono essere accessibili per questioni di sicurezza nazionale
  - Operatori: energia, telecom, …
- *Alcuni di questi dati sono difficilmente accessibili in modo sistematico*

Privati Statici

- Codice fiscale
- Foto non condivise
- Aspetti legali
- Cartella clinica
- ..

Pubblici statici (open data)

statistiche: incidenti, censimenti, votazioni

- Statistiche accessi alla ZTL
- Strutture pubbliche UNIFI

- Movimenti personali non pubblicati
- Relazioni personali non pubblicate

posizione dei punti di interesse

- Musei
- Strutture della città
- Servizi attivi

- comportamenti
- social media
- contributi
- consumi

- Traffico personale
- Posizione mezzi,
- Meteo
- Parcheggi
- Posizione taxi
- Code ai musei
- Posizione CarSharing
- ...

- Info traffico
- video camere
- Info Meto
- Info Ambiente
- Terremoti
- Parcheggi

- Stato accessi alla ZTL
- Stato dei servizi

Privati Tempo reale

Pubblici Tempo reale (open data)

Relatore **Paolo Nesi**
Ordinario di Sistemi di elaborazione delle informazioni

i dati

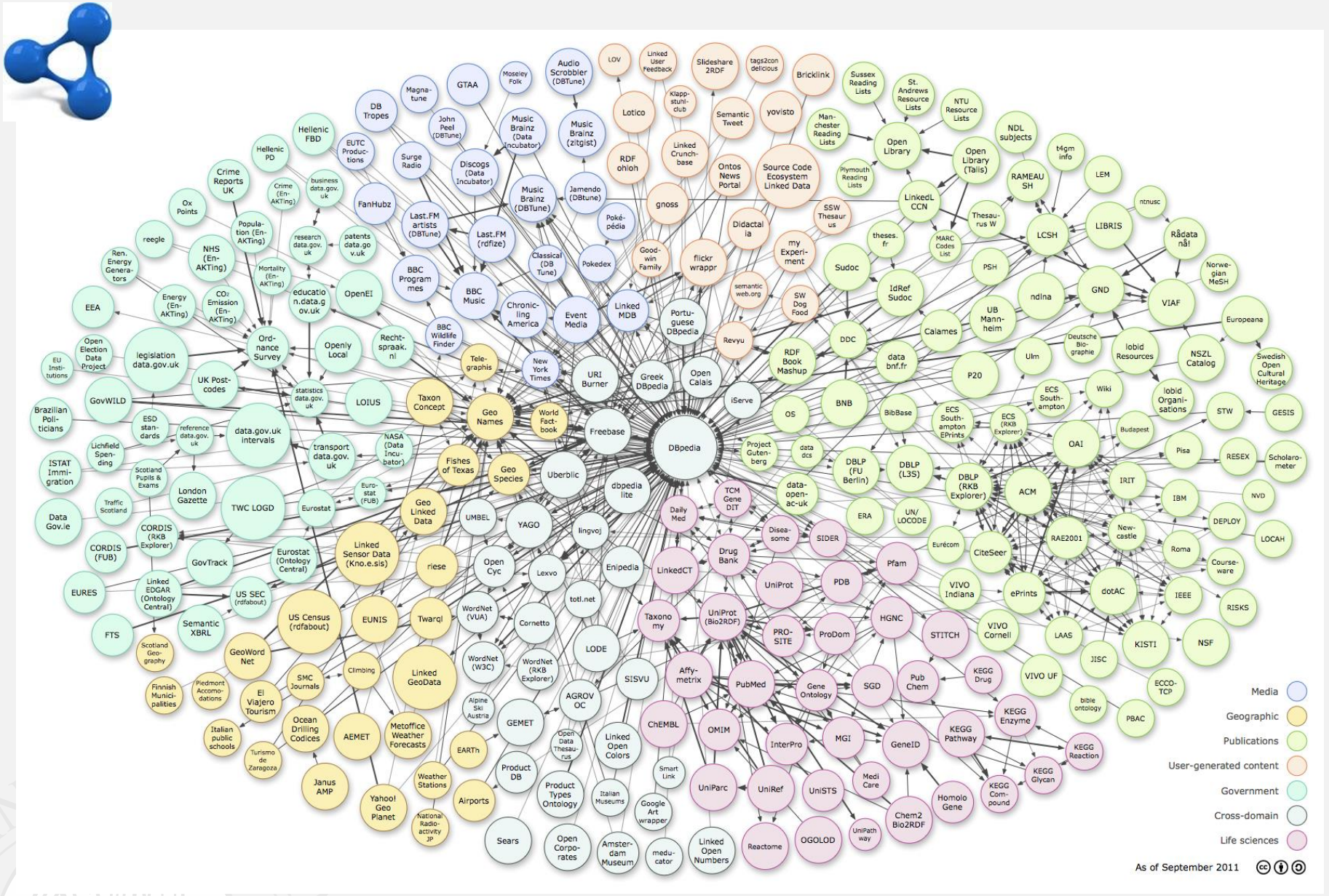
- **Dati Privati e Semantica**:
  - **A. Trasmissioni: streaming, progressive download, P2P**
    - Generati da un o più persone, o riprodotte, etc.
    - *Conversazioni audio, video, telefonate, chat, etc.*
  - **B. Dati personali generati:** prodotti dalla singola persona
    - Generati o modificati dalla persona intenzionalmente,..
    - *email, documenti, fogli di calcolo, immagini, video, agende, etc*.
  - **C. Profili utenti indotti**: caratterizzano la singola persona
    - Generati dal comportamento e dalle azioni della persona in vari servizi privati e GOV, anche in modo non intenzionale
    - *Descrittori del comportamento, relazioni, scelte, spese, conoscenze, log chiamate, etc.*

Dati vs Metadati

- Dati, open data
- Dati privati, dati pubblici
- Linked Open Data
- Metadati e indicizzazione
- Dati e smart city
- User Profile

Linked Data

DISIT Lab, Distributed Data Intelligence and Technologies
Distributed Systems and Internet Technologies
Department of Information Engineering (DINFO)
http://www.disit.dinfo.unifi.it



Linked Open Graph: Linked Open Data

Linked Open Graph

Select a SPARQL endpoint:
OSIM UNIFI Competences (by DISIT)
Examples:
- Paolo Nesi
- Dip. Ingegneria dell\'Informazione

Choose a class:
Search for keyword
keyword:
università
uri: /ww.dsi.unifi.it/CMSAteneoCompetence#University Request
Multiple endpoint search

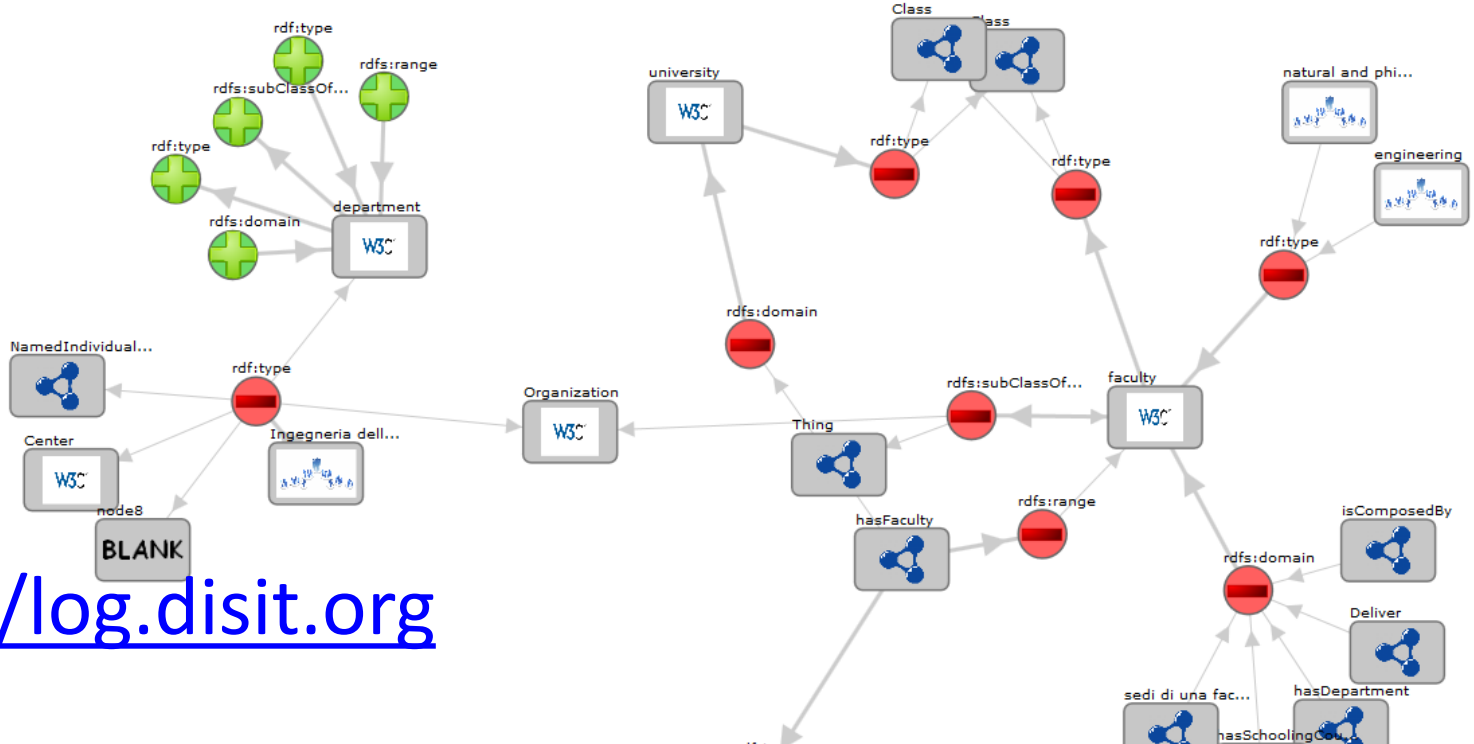
Your data
sparql endpoint: (optional)
http://...
uri: http://... Request
Multiple endpoint search

Status
Requests:
università [OSIM UNIFI Competences (b
Dip. Ingegneria dell\\\'Informazione [OSIM (b
Remove Clear

Type of relations
Select all | Deselect all | Invert | Hide all inverse
- foaf:depiction
- owl:sameAs
- rdf:type
- rdfs:domain
- rdfs:range
- rdfs:seeAlso
- rdfs:subClassOf

Linked Open Graph

Shown:34
Entities:20
Relations:14



http://log.disit.org

Dati vs Metadati

- Dati, open data
- Dati privati, dati pubblici
- Linked Open Data
- Metadati e indicizzazione
- Dati e smart city
- User Profile

Metadati

- Sono descrittori di dati
- Vi sono svariati standard di descrizione
  - DC, METS, MPEG-21, MPEG-7, EDM, ..FRBR, …
- Descrivono vari aspetti, in varie lingue,
- Possono essere specifici per il contesto, etc.
- Oggi sono spesso formalizzati in XML e codifcati in ASCII.

Indexing

Media Types	DC (ML)	Technical	Performing Arts	Full Text	Tax, Group (ML)	Comments, Tags (ML)	Votes
# of Index Fields*	468	10	23	13	26	13	1
Cross Media: html, MPEG-21, animations, etc.	Y_n	Y	Y	Y	Y_n	Y_m	Y_n
Info text: blog, web pages, events, forum, comments	T	N	N	N	N	Y_m	N
Document: pdf, doc, ePub	Y_n	Y	Y	Y	Y_n	Y_m	Y
Audio, video, image	Y_n	Y	Y	N	Y_n	Y_m	Y_n
Aggregations: play lists, collections, courses, etc.	Y_n	Y	Y	Y/N	Y_n	Y_m	Y_n

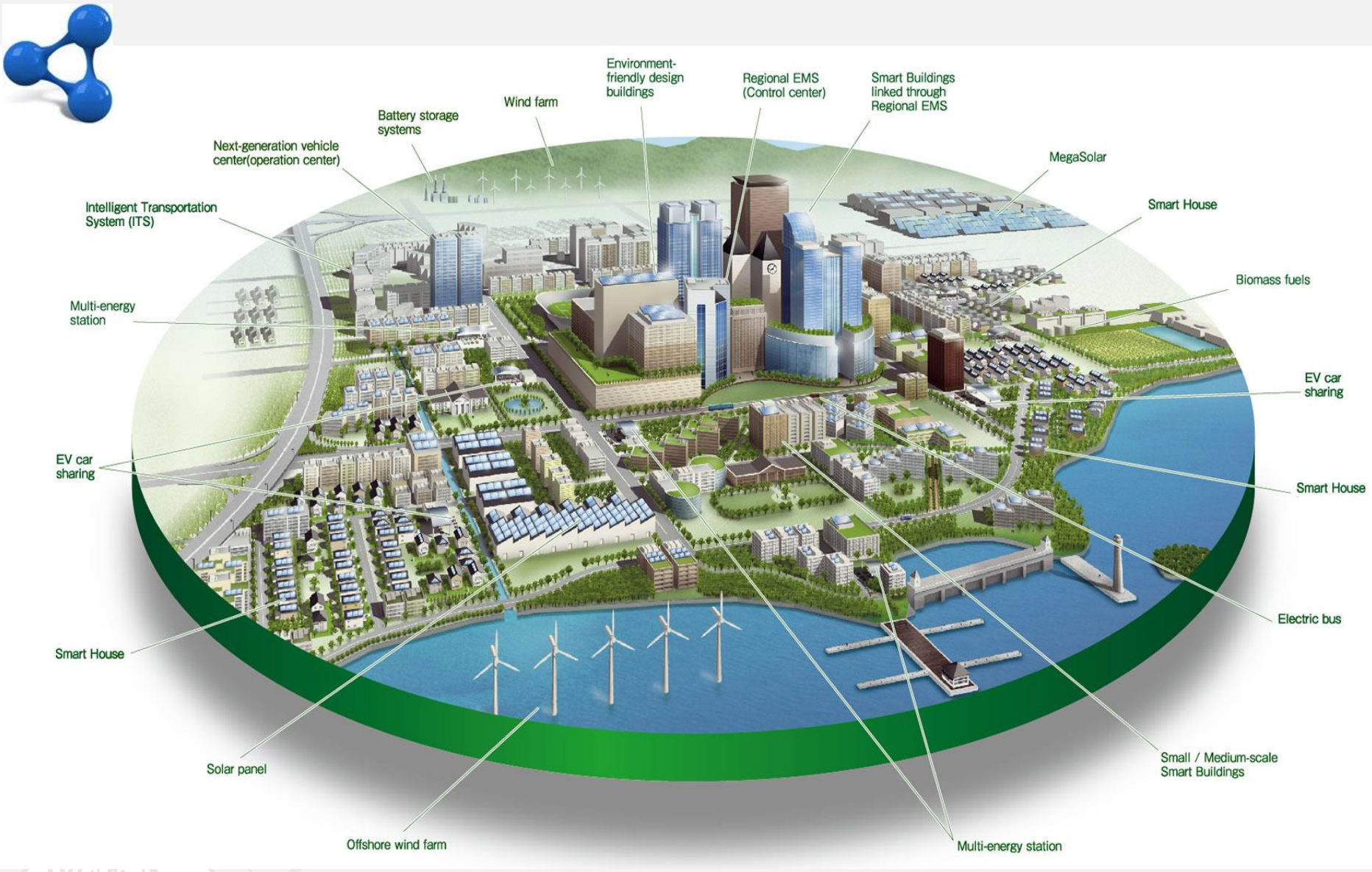
* = (# of Fields per Metadata type) × (# of Languages)
ML: Multilingual; DC: Dublin Core; Tax: Taxonomy

Modello di ricerca vs algebra

- **Esempio:**
  - **query1 AND query2** will match all the items containing both **query1** and **query2**.
  - **query1 OR query2** will match all the items containing **query1** or **query2** or both.
  - **query1 [AND] NOT query2** will match all the items containing **query1** and *NOT* **query2**.
  - **+query1 -query2** will match all the items containing **query1** but not **query2**.
- Modelli indicizzati Faceted: e.g.: www.eclap.eu
- Modelli

Dati vs Metadati

- Dati, open data
- Dati privati, dati pubblici
- Linked Open Data
- Metadati e indicizzazione
- Dati e smart city
- User Profile


Environment-friendly design buildings
Regional EMS (Control center)
Smart Buildings linked through Regional EMS
Wind farm
Battery storage systems
Next-generation vehicle center(operation center)
MegaSolar
Intelligent Transportation System (ITS)
Smart House
Biomass fuels
Multi-energy station
EV car sharing
EV car sharing
Smart House
Electric bus
Smart House
Small / Medium-scale Smart Buildings
Solar panel
Offshore wind farm
Multi-energy station

Sustainability of the Growth

- To be **planned and managed** with respect to increment of population and their needs
  - increment of efficiency:
    - compensation of the increments of costs
  - Increment of quality of life:
    - compensation of the decrement of quality of life
  - provisioning of new services:
    - compensation of the inadequacy of services
  - Decision support for strategic aspects
    - Corrections, prediction, new services, etc.
- **Towards citizens**
  - Informing citizens on the new adaptations, making them aware about that
  - Forming citizens to adopt virtuous behaviour in the usage of services and resources

Smartness, smart city needs 6 features

- Smart Health
- Smart Education
- Smart Mobility
- Smart Energy
- Smart Governmental
  - Smart economy
  - Smart people
  - Smart environment
  - Smart living
- Smart Telecommunication

Smart health

(can be regarded as smart governmental)

- Online accessing to health services:
  - booking and paying
  - selecting doctor
  - access to EPR (Electronic Patient Record)
- **Monitoring** services and users for,
  - learn people behavior, create collective profiles
  - personalized health
  - Inform citizens to the risks of their habits
  - Improve efficiency of services
  - redistribute workload, thus reducing the peak of consumption

Smart Education

(can be regarded as smart governmental)

- Diffusion of ICT into the schools:
  - LIM, PAD, internet connection, tables, ..
- Primary and secondary schools → university → industry & services
- **Monitoring** the students and quality of service,
  - learn student behavior, create collective profiles,
  - personalized education
- suggesting behavior to
  - Informing the families
  - moderate the peak of consumption
  - increase the competence in specific needed sectors, etc.
  - Increase formation impact and benefits

Smart Mobility

- Public transportation:
  - bus, railway, taxi, metro, etc.,
- Public transport for services:
  - garbage collection, ambulances,
- Private transportation:
  - cars, delivering material, etc.
- New solutions (public and/or private):
  - electric cars, car sharing, car pooling, bike sharing, bicycle paths
- Online:
  - ticketing, monitoring travel, infomobility, access to RTZ, parking, etc.

Smart Mobility and urbanization

- **Monitoring** the city status,
  - learn city behavior on mobility
  - learn people behavior
  - create collective profiles
  - tracking people flows
- **Providing Info/service**
  - personalized
  - **Info** about city status to
  - help moving people and material
  - education on mobility,
  - moderate the peak of consumption
- **Reasoning to**
  - make services sustainable
  - make services accessible
  - Increase the quality of service

Smart Energy

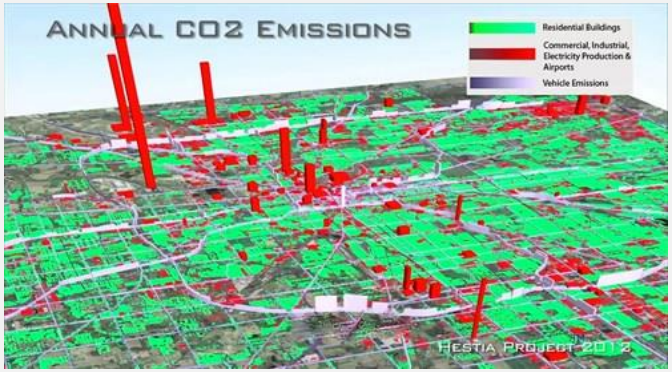
- **Smart building:**
  - saving and optimizing energy consumption, district heating
  - renewable energy: photovoltaic, wind energy, solar energy, hydropower, etc.
- **Smart lighting**:
  - turning on/off on the basis of the real needs
- **Energy points for electric**: c
  - ars, bikes, scooters,
- **Monitoring** consumption, learn people/city behavior on energy consumption, learn people behavior, create collective profiles
- Suggesting consumers
  - different behavior for consumption: different time to use the washing machine
- Suggesting administrations
  - restructuring to reduce the global consumption,
  - moderate the peak of consumption

Smart Governmental Services

- Service toward citizens:
  - on-line services:
    - register, certification, civil services,
      taxes, use of soil, …
  - Payments and banking:
    - taxes, schools, accesses
  - Garbage collection:
    - regular and exceptional
  - Quality of air:
    - monitoring pollution
  - Water control:
    - monitoring water quality, water dispersion, river status

Smart Governmental Services

- **Service toward citizens**:
  - **Cultural Heritage**: ticketing on museums,
  - **Tourism**: ticketing, visiting, planning, booking (hotel and restaurants, etc.)
  - **social networking:** getting service feedbacks, monitoring
- **Social sustainability of services**:
  - crowd services
- **Social recovering** of infrastructure,
  - New services, exploiting infrastructures
- **Monitoring** consumption and exploitation of services, learn people behavior, create collective profiles
  - Discovering problems of services,
  - Finding collective solutions and new needs…



Telecommunication, broadband

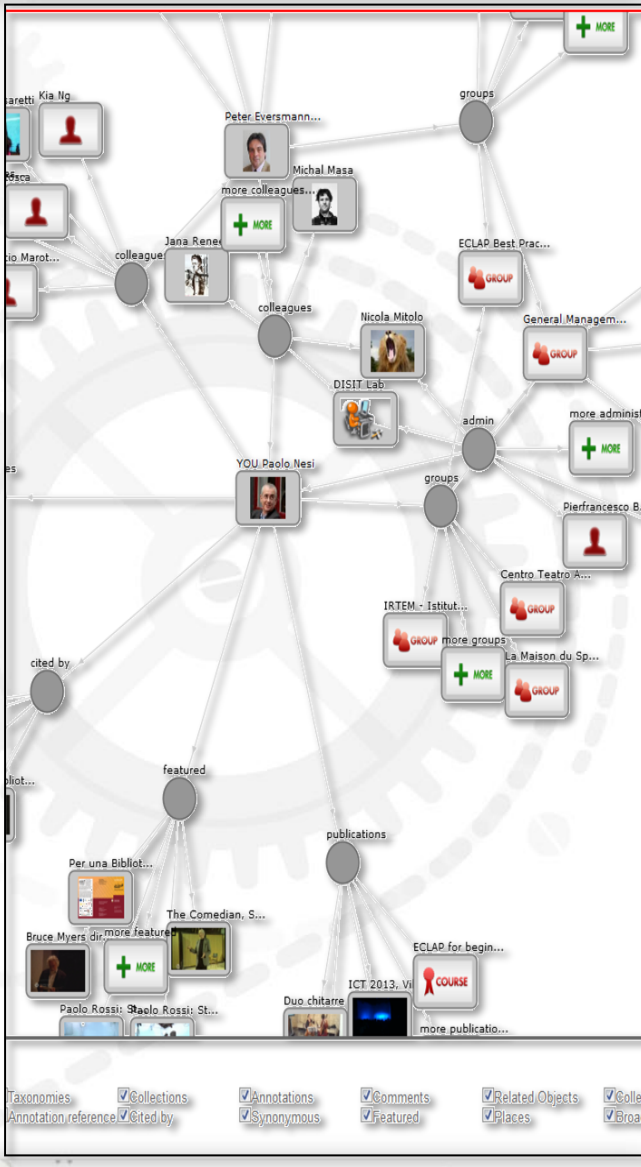
- **Fixed Connectivity**:
  - ADSL or more, fiber,
- **Mobile Connectivity**:
  - Public wifi, Services on WiFi, HSPDA, LTE
- **Monitoring** communication infrastructure
- Providing information and formation on:
  - how to exploit the communication infrastructure
  - Exploiting the communication for the other services,
  - moderate the peak of consumption

Dati vs Metadati

- Dati, open data
- Dati privati, dati pubblici
- Linked Open Data
- Metadati e indicizzazione
- Dati e smart city
- User Profile

I profili degli utenti

- ***Gli utenti possono:***
  - fornire informazioni preziose sulla citta come «***sensori intelligenti»*** per tenere sotto controllo il livello dei servizi della città e/o nuove necessità
  - **essere profilati per ricevere dei servizi personalizzati, benefici diretti**
- Informazioni anonime:
  - *velocità degli spostamenti: auto a piedi, code e flussi cittadini, temperature, meteo*
  - *Uso dei servizi*
- private in consenso informato, statistiche e attuali:
  - *Azioni e dati personali*
  - *Relazioni con altre persone*
  - *Movimenti puntuali*



User Profile

- **Static user profile aspects**
  - generically provided during registration
    - frequently not so much detailed in generic Social Networks,
    - users prefer to avoid filling in ‘useless’ forms and/or to provide false data.
  - In small thematic and business oriented Social Networks the information is much more reliable.
  - Dependent on the Social Network objectives

- **Dynamic user profile aspects**
  - generate on the basis of the user’s activities performed on the SN elements, such as the actions performed on
    - content, other users, on groups, on chat, etc.
  - estimated/inferred by assessment/analysis

Static Aspects of User profile

- **Static information collected during registration++**
  - Name, surname,
  - Nationality and languages (multiple)
  - Genre, age, etc..other personal info,.
  - Instruction/School, work, family structure, etc.
  - Personal photo
  - Jobs: several different jobs with periods, etc..
  - Competences: several skills
  - Economical data: range, etc.
  - Explicit Preferred content:
    - topics, genre, period, area, etc.
  - Subscribed (slow dynamic):
    - lists, groups, ..

Dynamic Aspects of User profile

- **dynamic information collected on the basis of the activities**:
  - votes and comments/annotations on:
    - contents, forums, web pages, etc.;
  - downloads and play/view/executions of content, web pages, etc.;
  - uploads and publishing of user provided content;
  - marked content as preferred/favorite;
  - recommend content/groups or users to other users;
  - chat with other users, publishing forum topic on groups;
  - queries performed on the portal, etc.;
  - create a topic in a forum or contribute to a discussion;
  - relationships/connections with other users or groups;
  - Etc.

Dynamic aspects of user actions

- Statements written on blogs and micro blogs
  - Short Comments in a context
  - Recurrent user and statement
  - The same statement on more than one blog (pushed by pushers)
  - Dates and time, successive blog posts
- Statements on comments
- Assessment of:
  - Market trends, market vigilance
  - Pharmavigilance

Esempio di Profilo utenti

- *Informazioni statiche:*
  - *Informazioni generali:*
    - *nome, cognome, sesso,*
    - *foto, data di nascita,*
    - *descrizione personale,*
    - *località di provenienza (ISO 3166),*
      - *Nazione*
      - *Suddivisione*
      - *Provincia*
    - *lingue parlate (ISO 369)*
  - *Informazioni di contatto:*
    - *lista di contatti di instant messaging*
  - *Scuola e Lavoro:*
    - *scelta del livello scolastico,*
    - *nome della scuola,*
    - *tipo di lavoro,*
    - *nome del posto di lavoro*
  - *Interessi:*
    - *Vettore contenente la lista di valori del campo Type degli oggetti scelti dall'utente*
- *Informazioni dinamiche:*
  - *Lista di oggetti preferiti*
  - *Lista di amici*
  - *Lista gruppi*
  - *Voti positivi ad oggetti*
  - *Commenti ad oggetti*
  - *Blog post*
  - *…*
  - *…*
  - *Informazioni sulle preferenze sulla base delle visualizzazioni degli oggetti*
    - *Format*
    - *Type*
    - *Taxonomy*

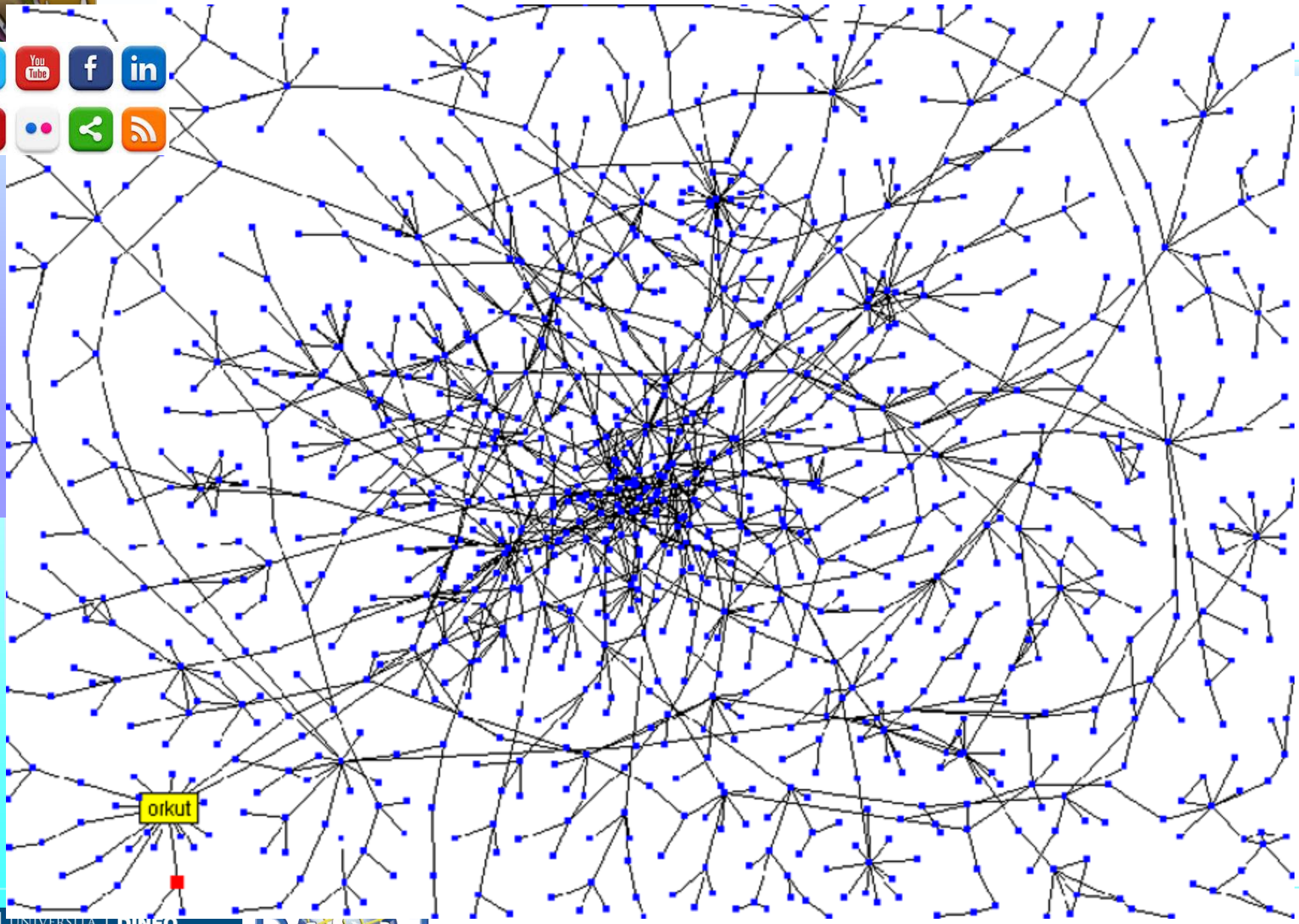
sommario

- *Dati e formati*
- *Comunicazione fra sistemi computerizzati*
- *Modelli di protezione e gestione*
- *Dati vs Metadati*
- *Social network e Smart City*
- *Motori di Ricerca, Crawling e NLP*
- *Data Mining*
- *Data Intelligence*

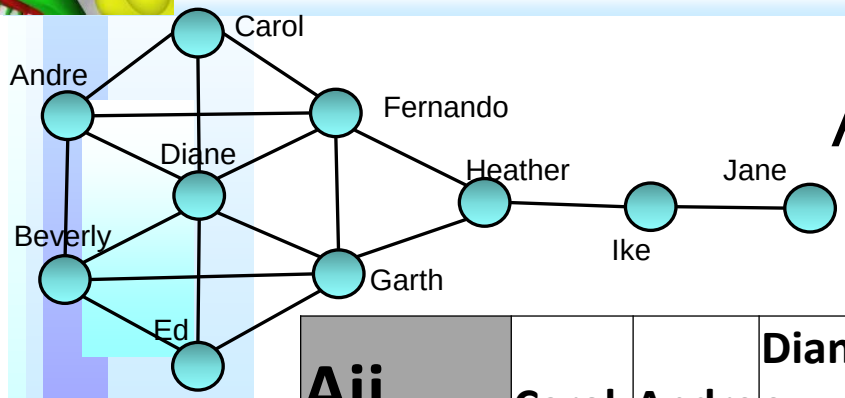
Social Network e Smart City

- Relazioni fra utenti
  - Si veda http://osim.disit.org per UNIFI
- Social network analysis
- Social Network
- Social graphs
- Social network interoperability
- Smart city model

Stanford Social Web



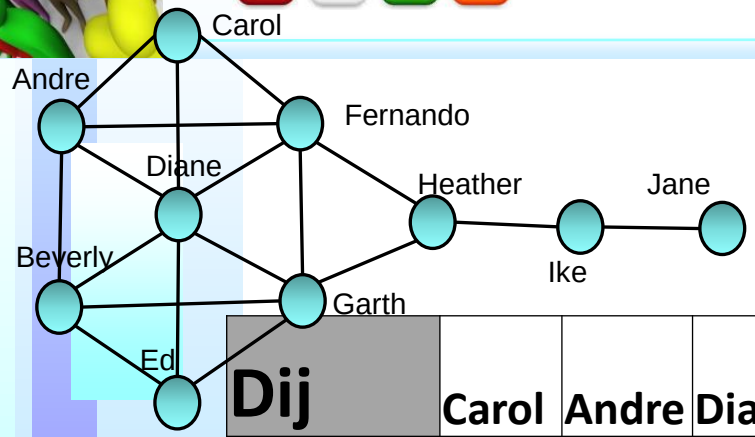
Matrix of connections



A[i][j]: matrix of connections

Aij	Carol	Andre	Diane	Fernando	Beverly	Ed	Garth	Heather	Ike	Jane	NC
Carol	0	1	1	1	0	0	0	0	0	0	3
Andre	1	0	1	1	1	0	0	0	0	0	4
Diane	1	1	0	1	1	1	1	0	0	0	6
Fernando	1	1	1	0	0	0	1	1	0	0	5
Beverly	0	1	1	0	0	1	1	0	0	0	4
Ed	0	0	1	0	1	0	1	0	0	0	3
Garth	0	0	1	1	1	1	0	1	0	0	5
Heather	0	0	0	1	0	0	1	0	1	0	3
Ike	0	0	0	0	0	0	0	1	0	1	2
Jane	0	0	0	0	0	0	0	0	1	0	1
nc	3	4	6	5	4	3	5	3	2	1	36

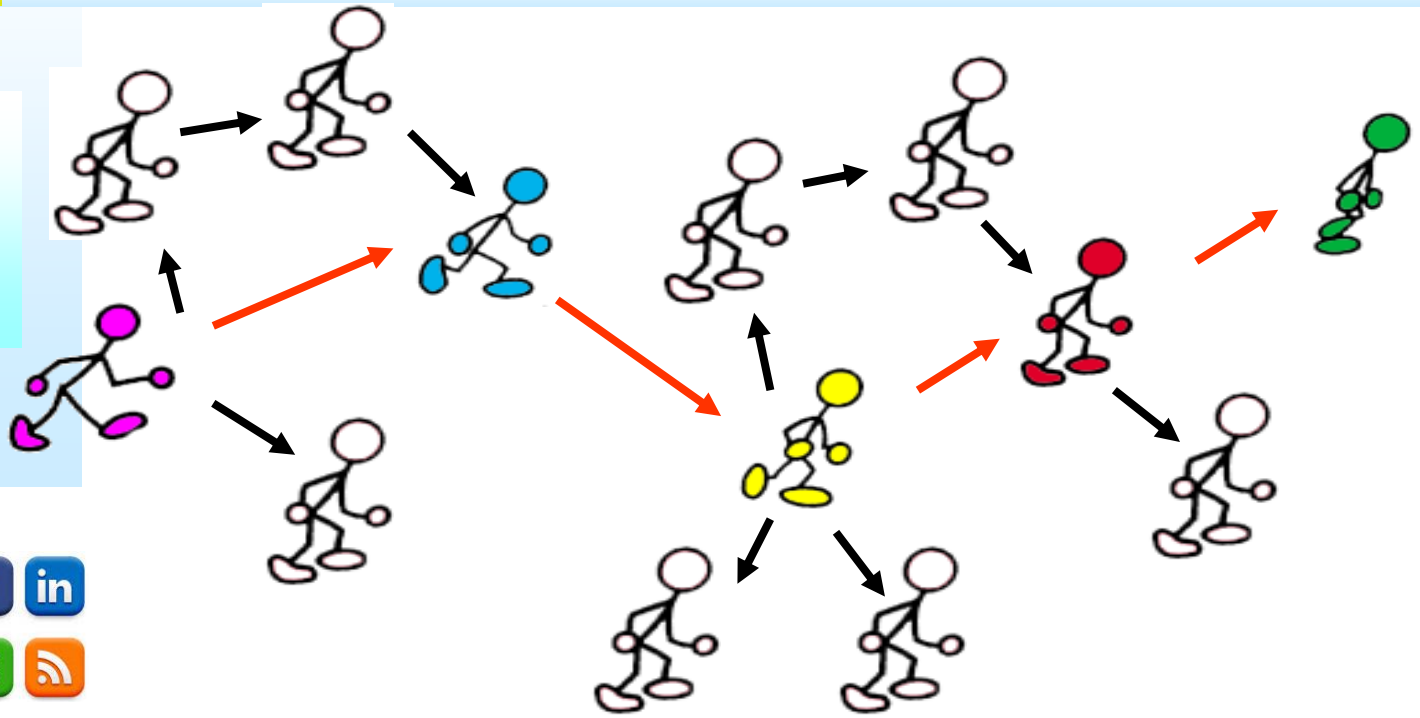
Matrix of distances



D[i][j]: matrix of distances
N*(N-1)/2 elements

Dij	Carol	Andre	Diane	Fernando	Beverly	Ed	Garth	Heather	Ike	Jane	
Carol	0	1	1	1	2	2	2	2	3	4	18
Andre		0	1	1	1	2	2	2	3	4	16
Diane			0	1	1	1	1	2	3	4	13
Fernando				0	2	2	1	1	2	3	11
Beverly					0	1	1	2	3	4	11
Ed						0	1	2	3	4	10
Garth							0	1	2	3	6
Heather								0	1	2	3
Ike									0	1	1
Jane										0	0
											89

Averaged shortest path from one person to another



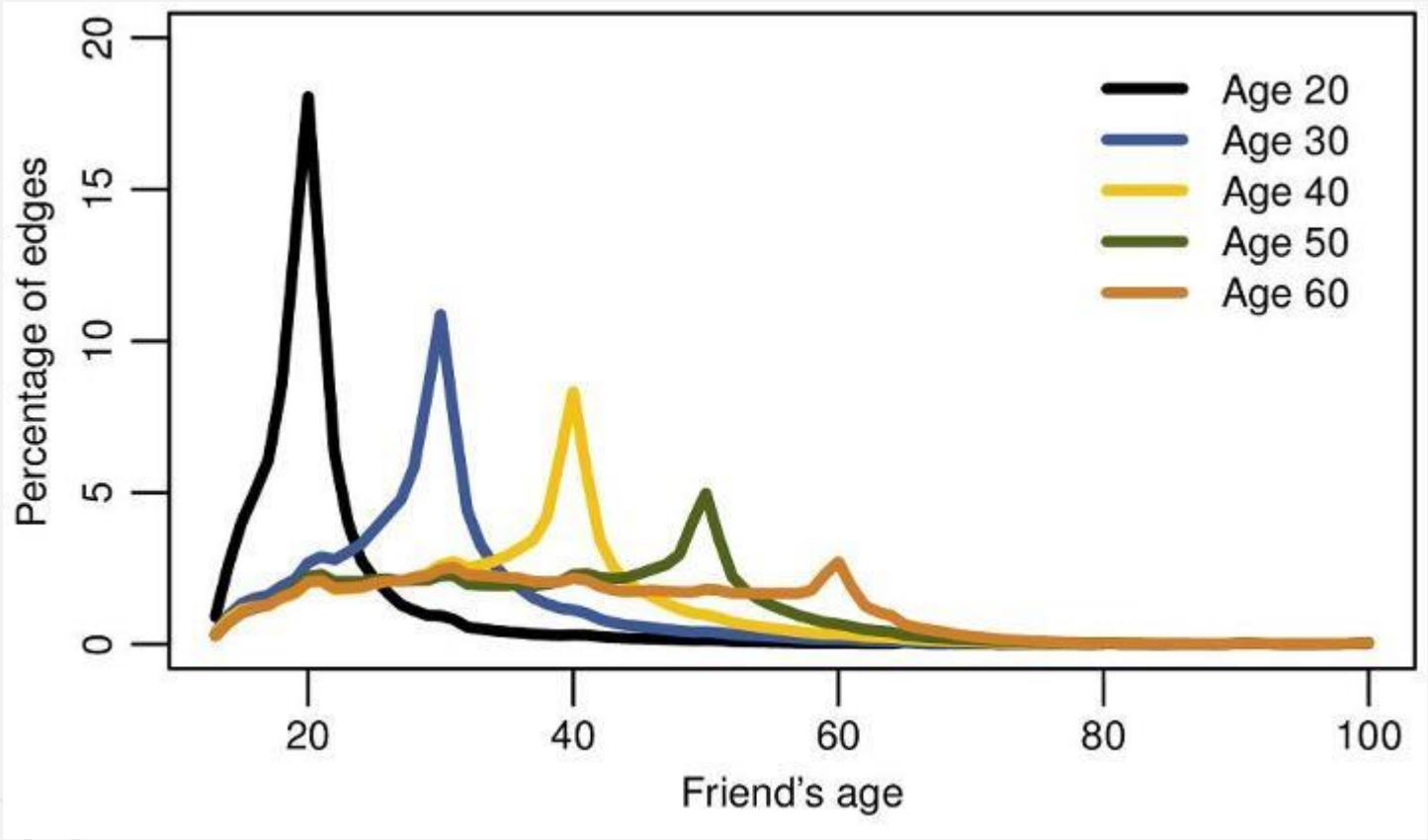
MIT: 6.4 hops

Stanford: 9.2 hops

Our example: 1.97 hops

Sum of shortest paths: 89
10 Nodes
45 possible connections

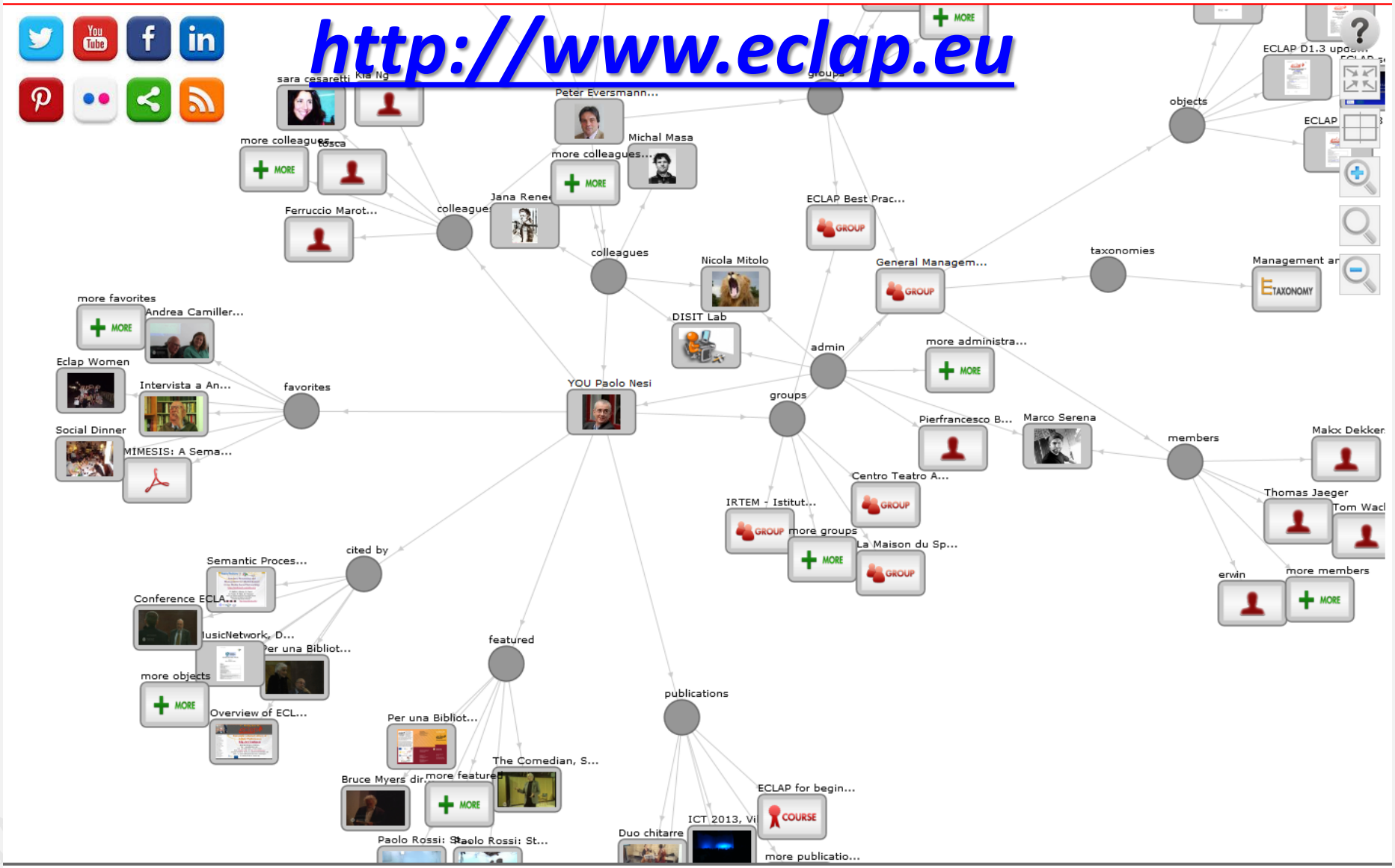
Distribuzione dei colleghi (facebook)



Social Network e Smart City

- Relazioni fra utenti
  - Si veda http://osim.disit.org per UNIFI
- Social network analysis
- Social Network
- Social graphs
- Social network interoperability
- Smart city model

http://www.eclap.eu



tions Select all Deselect all

er
- Groups
- Administrators
- Cited Names
- Favorite
- Writer
- Taxonomies
- Annotation reference
- Collections
- Cited by
- Annotations
- Synonymous
- Comments
- Featured
- Related Objects
- Places
- Colleagues
- Broader
- Publications
- Narrower
- User's favorites
- Formed

Social Network e Smart City

- Relazioni fra utenti
  - Si veda http://osim.disit.org per UNIFI
- Social network analysis
- Social Network
- Social graphs
- Social network interoperability
- Smart city model

Interoperability among Social Networks

- **SN may be interoperable** with other portals and SN
- **Allowing**:
  - **posting** comments and contributions via the so called ***Social Icon*** interface
  - **importing** user registration/profile and info or directly with some SSO
  - **exporting** SN content in other portals, for example via some API.
  - **hosting** SN players into other WEB portal pages, via some HTML segment to be copied
  - **hosting** widgets/applications into the WEB pages of the Social Network, via some programming model

Interoperability for Users

- Interchange of user profiles
  - **OpenID**: user identity standard, to allow user profile and credential interoperability among portals, is SSO method
  - **OAuth**: delegation protocol for accessing to credentials, user authentication
  - **OpenSocial** (by **Google** with MySpace): exchange of user profile.
    - Many big Social Networks have joined the OpenSocial API movement, including hi5, LinkedIn, Netlog, Ning, Plaxo, Orkut, Friendster, Salesforce, Yahoo, Ning, SixApart, XING, etc.
  - **Facebook** Connect is in competition with OpenSocial
  - Other technologies:
    - XUP of W3C, XMPP, FOAF, XFN, etc..

Obiettivi: Twitter Vigilance

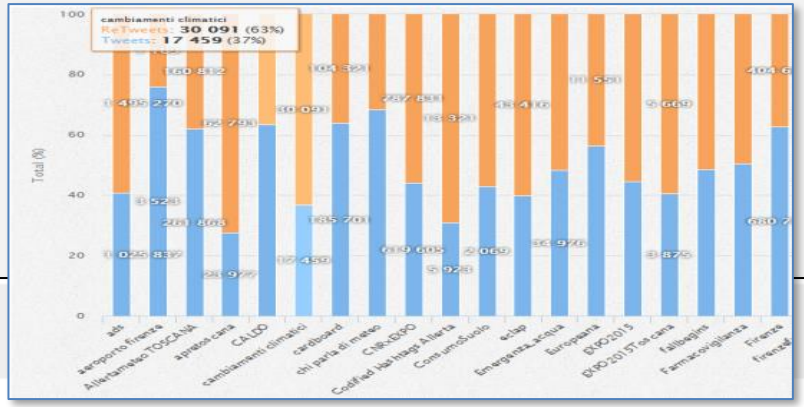
- **Monitorare** Canali Twitter con
  - **Elevata affidabilità** e precisione, seguendo eventi lenti, veloci ed esplosivi
  - **Gestione multiutente** per canali multipli: pubblici e privati
  - **Data analytics e sentiment analysis** in modo assistito e diretto

- ***Un Canale di Twitter Vigilance:*** un insieme di ricerche attive e adattive su social media Twitter.
  - Ogni ricerca può essere semplice o complessa
  - Ogni utente qualificato può gestire più canali e ricerche
- **Attivo**
  - da Aprile 2015 per il collezionamento dati,
  - da settembre per il calcolo automatico big data analytics,
  - da ottobre per la Sentiment Analysis automatizzata.
- **Gestisce** con disinvoltura
  - Anche 580.000 tweet per giorno
  - Oltre 30 Milioni di Tweet

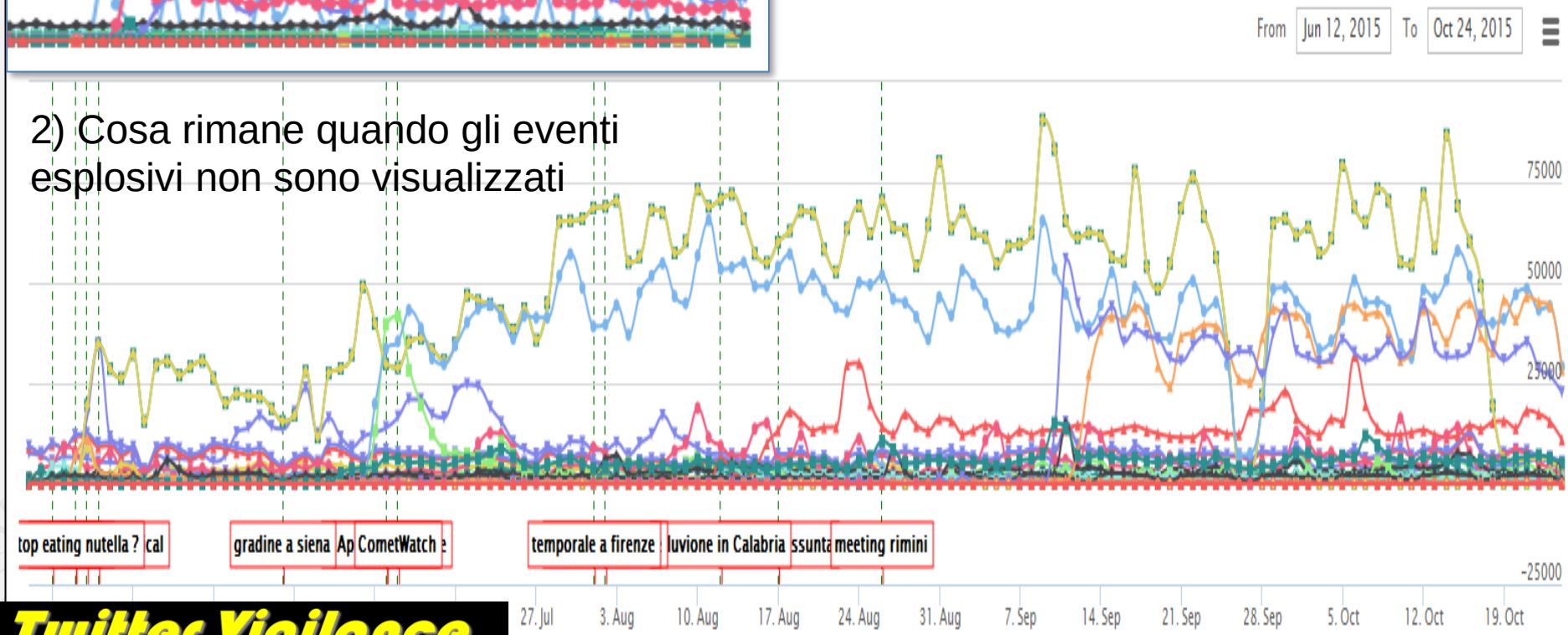
Twitter Vigilance

Canali Attivi in studio (protetti)

1) Visione completa con svariati eventi esplosivi

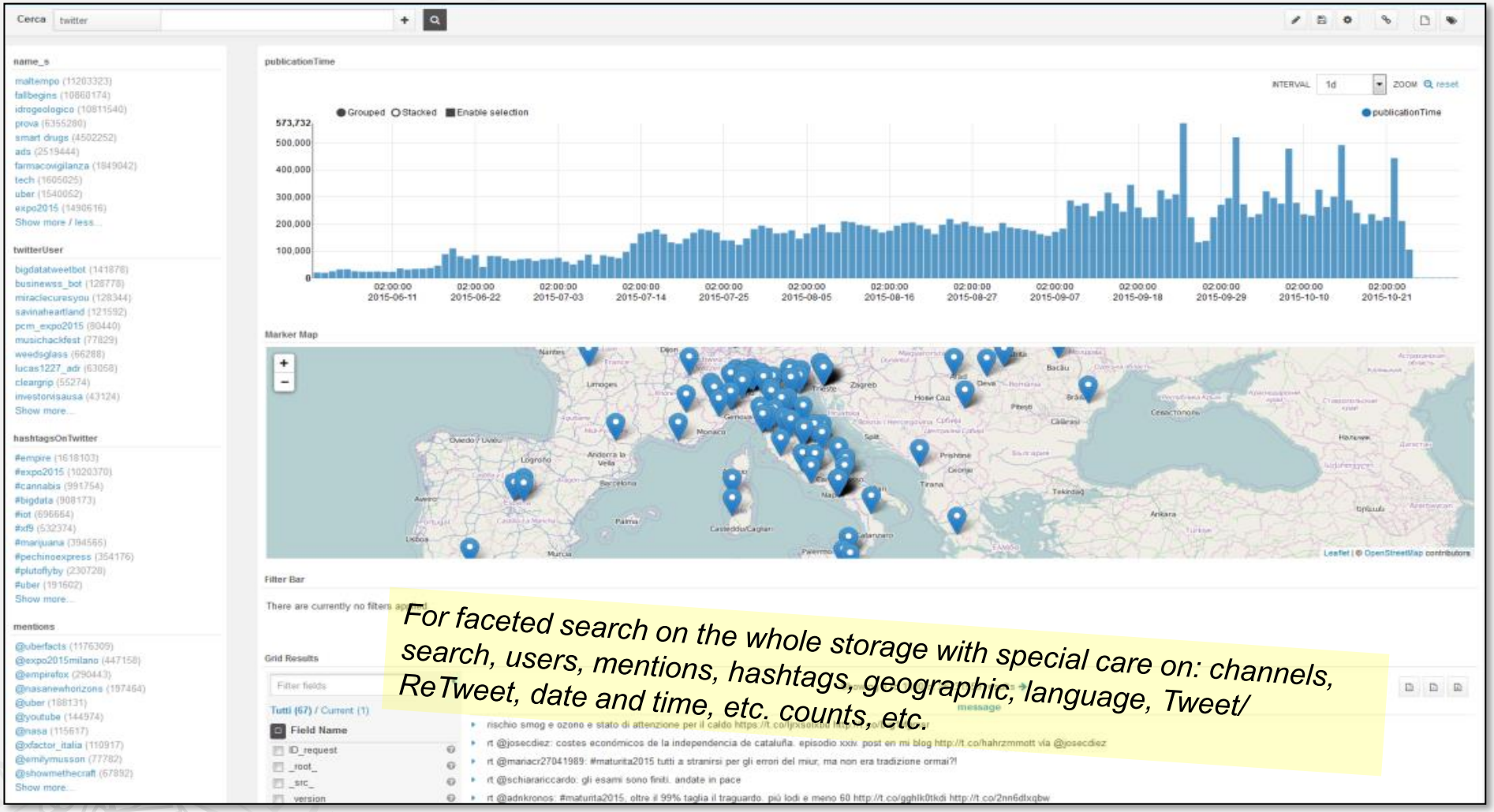


2) Cosa rimane quando gli eventi esplosivi non sono visualizzati



Twitter Vigilance

Ricerca full text e Faceted



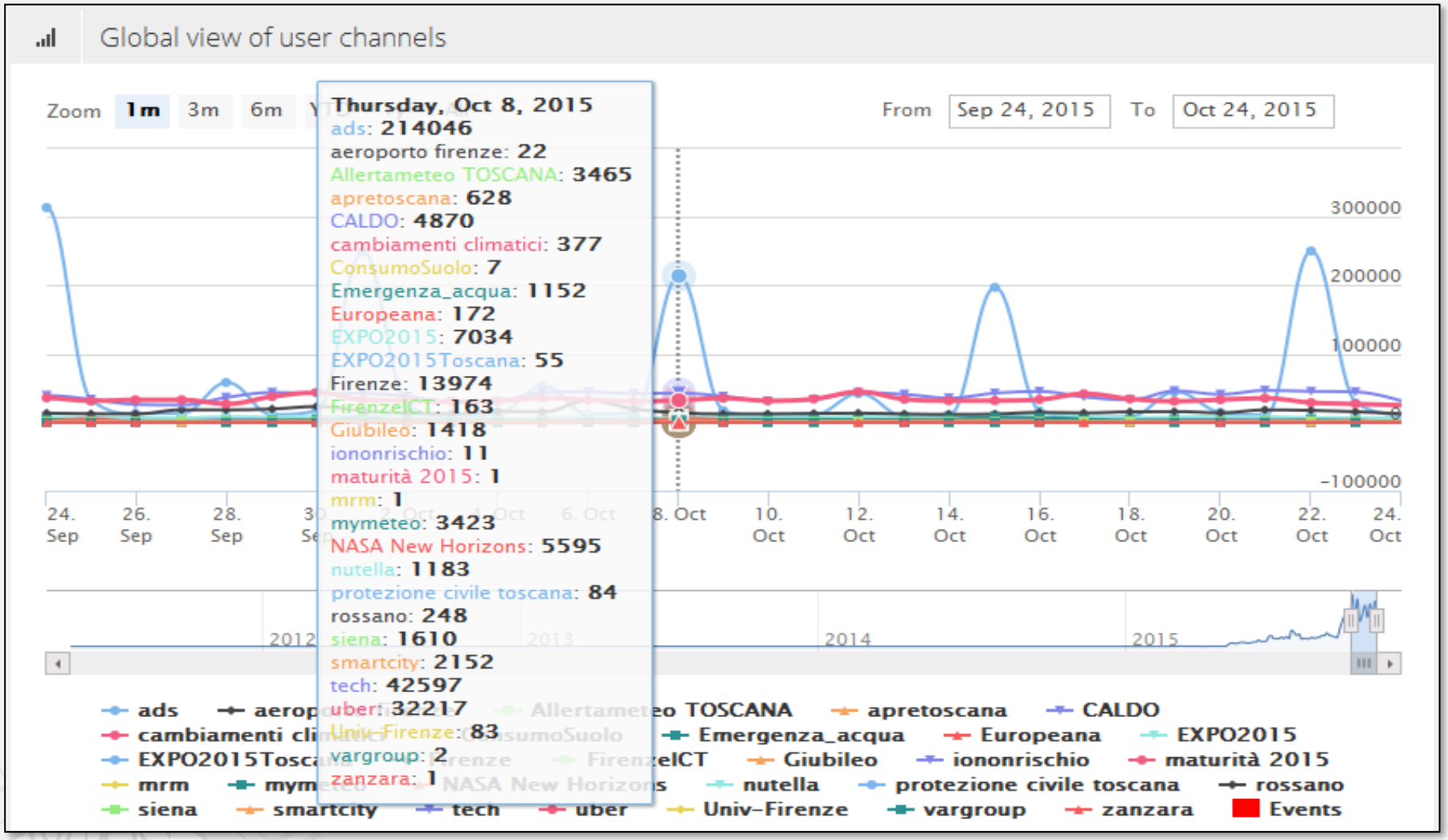
For faceted search on the whole storage with special care on: channels, search, users, mentions, hashtags, geographic, language, Tweet/ ReTweet, date and time, etc. counts, etc.

Twitter Vigilance

Alcuni Canali di TwitterVigilance

- **http://www.disit.org/tv canali pubblicati**
  - **Esempi di Canali:**
    - EXPO 2015, CNR EXPO2015,
    - Firenze, ApreToscana, maturità,
    - ConsumoSuolo, meteo, allerta meteo toscana, protezione civile,
    - farmaco vigilanza, smart drug,
    - ECLAP e Europeana, advertising TV, laudatesi, etc.
- **Aree tematiche:** meteo, ambiente, advertising, eventi pubblici, farmacovigilanza, smart city, politica, etc.

Canali Pubblici al Nov. 2015

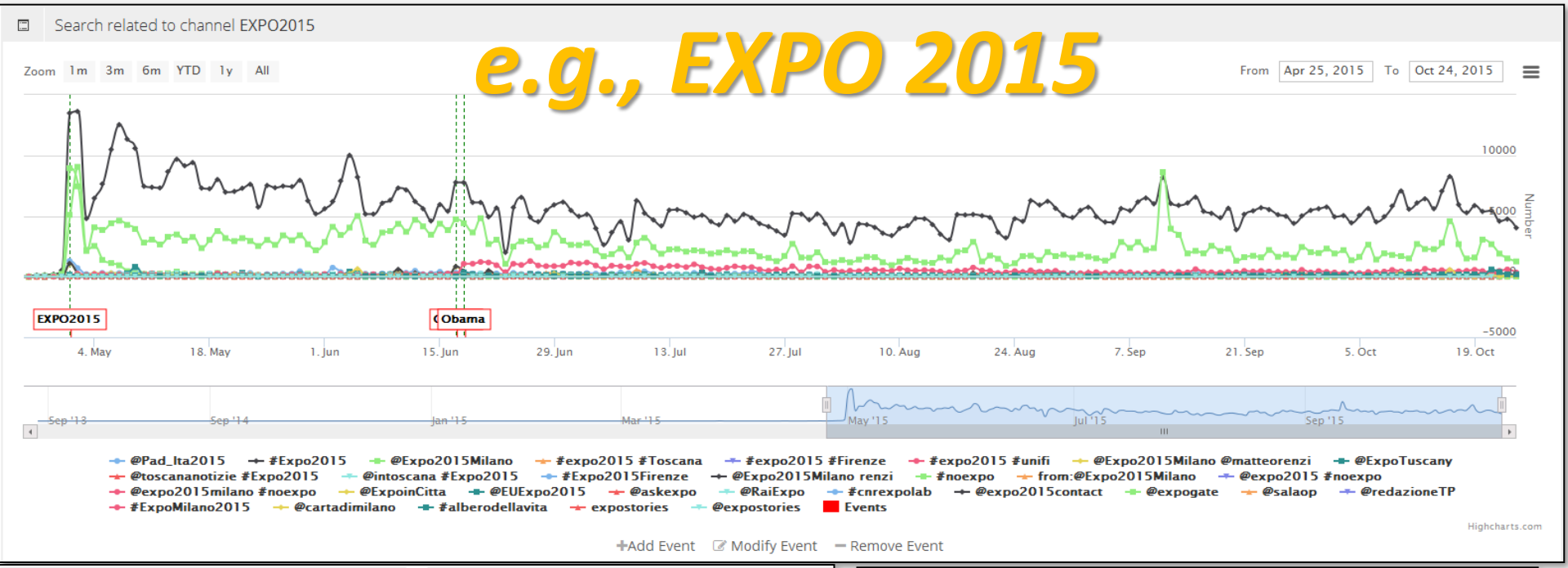


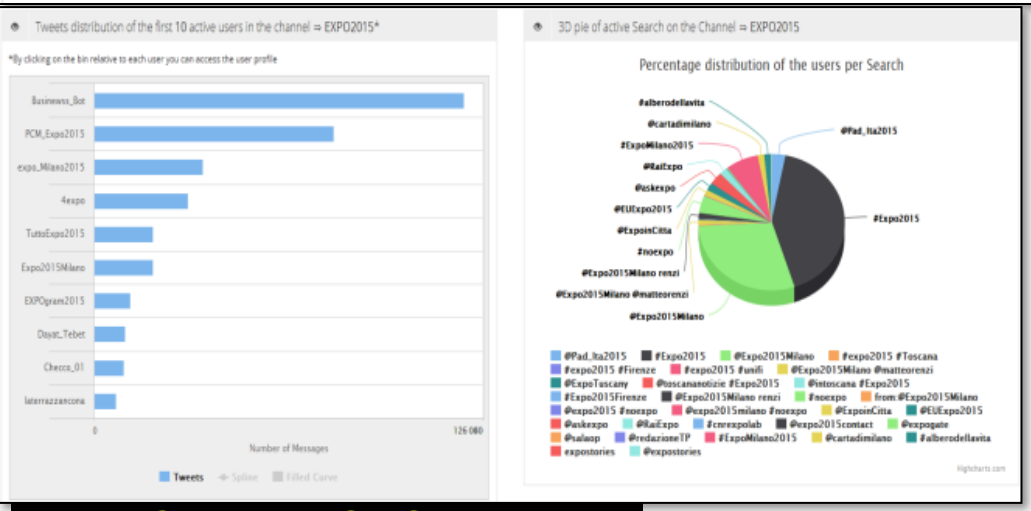
Twitter Vigilance

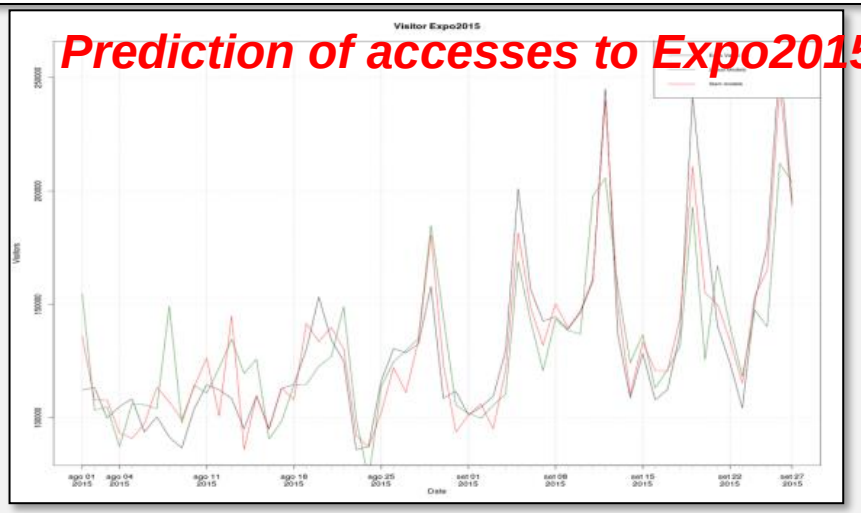
Twitter Vigilance: le analisi

- **Analisi e caratterizzazione della comunicazione**
- **Percezione sociale, eventi pubblici, naturali..**
- **Scoprire, identificare e calcolare**
  - **Nascita / crescita di nuove occorrenze** in tempo reale: eventi, fatti, meteo, condizioni critiche, etc.
    - ***Supporto alla decisioni, ridurre i tempi di reazione, valutare la percezione, ridurre i costi, incrementare la resilienza come capacità di reagire***
  - **Chi influenza** la comunicazione, le comunità e come: i pusher, gli attori, i *follower*, le sorgenti, etc.
  - **Predizione su eventi** periodici, per esempio presenze ad eventi, presenze sui canali televisivi, etc.
  - **Misure indirette** basate sulla popolazione: rischio sicurezza, degrado, neve, grandine, vento, etc.

UNIVERSITÀ DEGLI STUDI FIRENZE
DISIT
DISIT Lab, Distributed Data Intelligence and Technologies
Distributed Systems and Internet Technologies
Department of Information Engineering (DINFO)
http://www.disit.dinfo.unifi.it







Twitter Vigilance

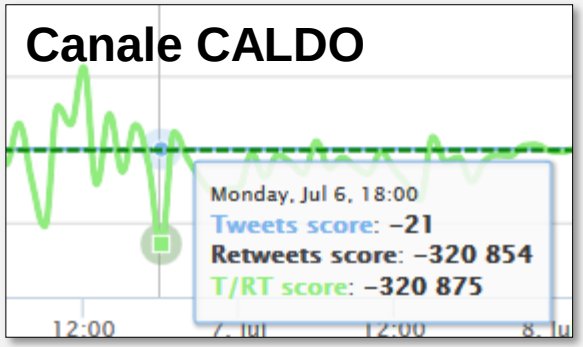
Twitter Vigilance: sentiment analysis

- **Controlla e analizza in automatico** il *livello di apprezzamento e/o dissenso* per:
  - prodotti, servizi, promozioni, cambiamenti in città,
  - persone, azioni politiche, eventi,
  - programmi TV, attori, cantanti (surrogato di "auditel")
- Permette di effettuare
  - Valutazioni di **andamento** a breve e lungo termine
  - Valutazioni **comparative** a breve e lungo termine
  - **Predizioni** in certe situazioni
  - **Identificazioni in quasi real time** della nascita di eventi esplosivi, situazioni critiche, etc., uso di utenti come sensori diffusi

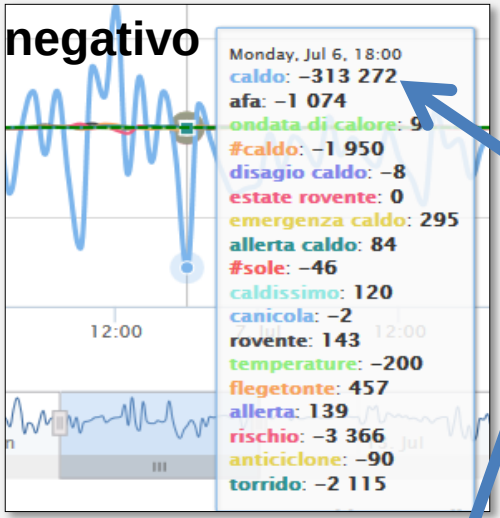
UNIVERSITÀ DEGLI STUDI FIRENZE
DISIT
DISIT Lab, Distributed Data Intelligence and Technologies
Distributed Systems and Internet Technologies
Department of Information Engineering (DINFO)
http://www.disit.dinfo.unifi.it

Sentiment Analysis

Canale CALDO

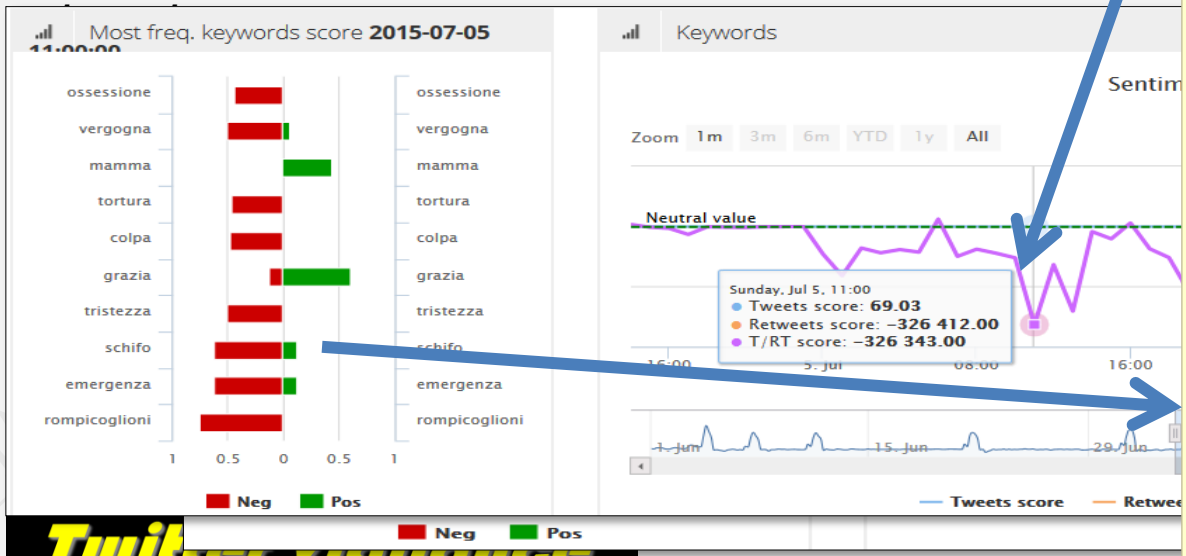


SA sul canale CALDO: la ricerca "caldo" ha dato un sentiment negativo



Un sentiment marcatamente negativo: -313272
Anche su key multiple

Pochi Tweet negativi hanno provocato un enorme numero di



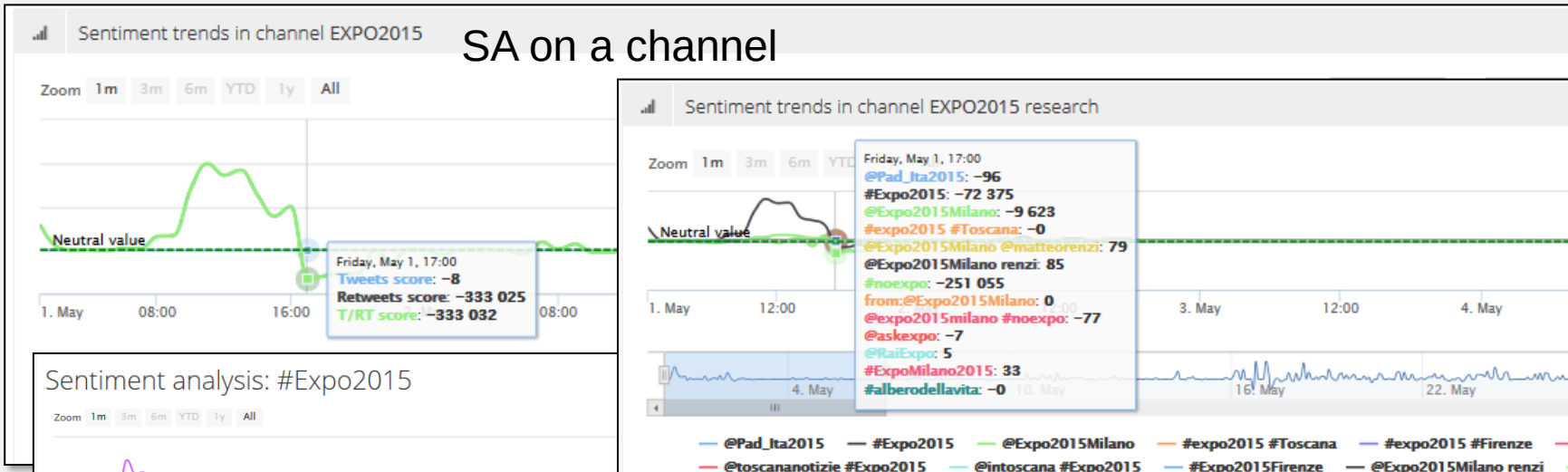
- **emergenza** caldo, una rete per i più deboli http://t.co/x6e4dkwqk8
- continua l'**emergenza** #caldo. domani fino a 43 gradi percepiti in #piemonte http://t.co/gvuyafeq0o
- cerolini su gestione **emergenza** caldo a pescara http://t.co/zecmr7t4zv #**emergenza**caldo #pescara
- **emergenza** caldo, in arrivo i 38 gradi nella cittò metropolitana di milano. consulta il bollettino emesso: http://t.co/n5oocelibv

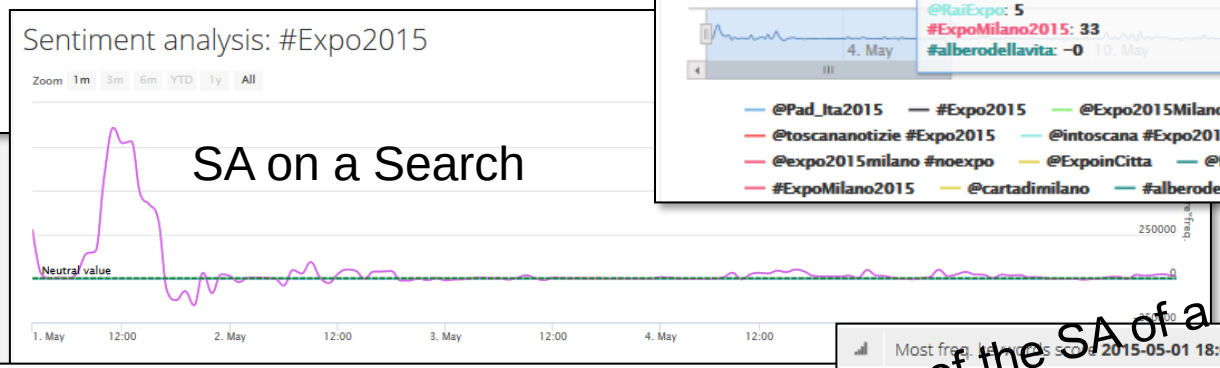
Twitter Vigilance

UNIVERSITÀ DEGLI STUDI FIRENZE
DISIT Lab, Distributed Data Intelligence and Technologies
Distributed Systems and Internet Technologies
Department of Information Engineering (DINFO)
http://www.disit.dinfo.unifi.it

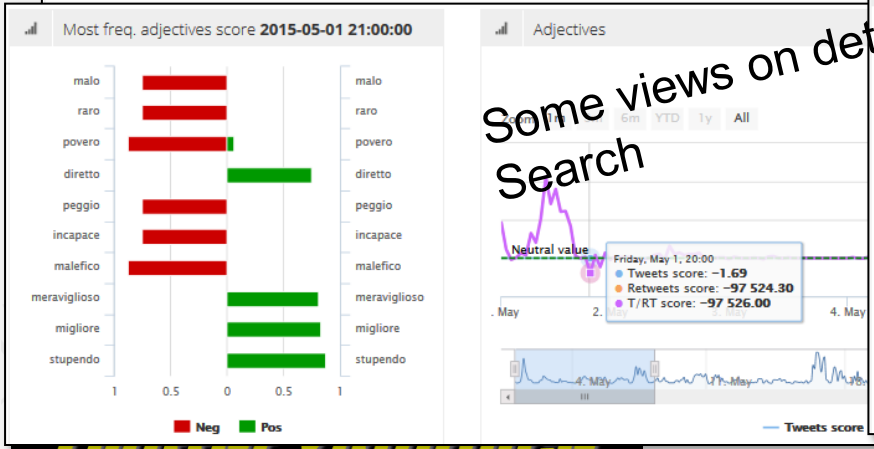
Sentiment Analysis

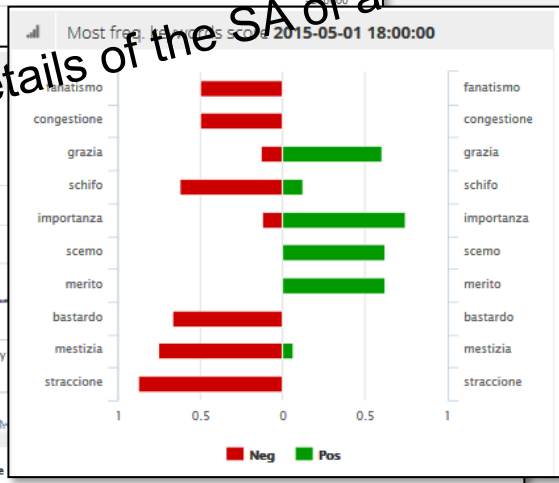
SA on a channel

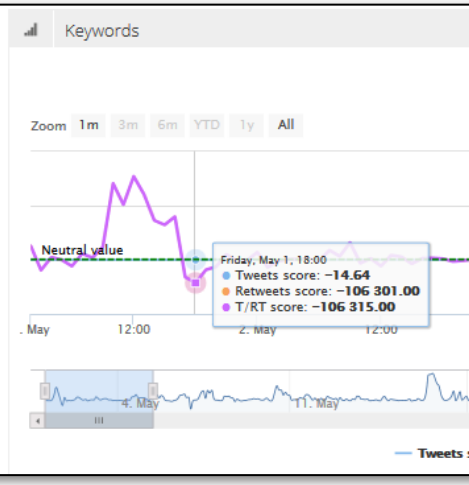




Some views on details of the SA of a Search







Twitter Vigilance

Social Network e Smart City

- Relazioni fra utenti
  - Si veda http://osim.disit.org per UNIFI
- Social network analysis
- Social Network
- Social graphs
- Social network interoperability
- Smart city model

Sfide: richieste e deduzioni

Pub. Amm.

Miglioramento dei servizi

vendita di servizi, miglioramento, vendita di dati

Operatori: Mobilità, energia, Salute,..

profilazioni

Commercio

Miglioramenti e predizioni sui servizi

Turismo Cultura

Accesso ai dati per le PMI, PA

Sistema Smart City

User profiling
Profili Collettivi
Segmentazione

Servizi & Suggerimenti
Trasporto, Mobilità, Commercio, Turismo Cultura, salute,..

Personal Time Assistant biglietti, suggerimenti, informazioni, aiuto, etc.

Comportamenti degli utenti

Dati: pubblici e privati, statici e in tempo reale

Privati: comportamenti, social media, contributi, consumi
Pubblici: mobilità, traffico, video camere, ambiente, acqua, statistiche, accesso alla ZTL, servizi, musei, punti di interesse, …

Relatore Paolo Nesi
Ordinario di Sistemi di elaborazione delle informazioni

Scenari che da fantascientifici diventano reali…

Grazie a infrastrutture che..

- ***Raccolgono dati e statistiche su***
  - Ambiente & energia
  - Trasporti & mobilità
  - Commercio & Turismo
  - Servizi al cittadino
  - Comportamento e stato della popolazione nel rispetto della privacy, anonymity
- ***Producono analisi, previsioni e deduzioni*** su base
  - Statistica, analitica, logica…
  - sporadiche e/o in tempo reale

possiamo dire che

- *Gli utenti dovrebbero consumare la loro energia quando le industrie non lo fanno…*
- *Le auto elettriche dovrebbero essere ricaricate vicino alla generazione di energia*
- *Ora: vi sono 34 posti liberi in Piazza Stazione*
- *Ora: Il #4 arriva alla fermata in 3 minuti*

Relatore Paolo Nesi
Ordinario di Sistemi di elaborazione delle informazioni

Firenze

43.7693, 11.2560

Tue 10 Nov @ 12:53:21

ServiceMap, SmartDS

Wifi operativi 21

10.7% 89.3%

SmartDS: via santa marta 3 21

58.8% 21.3% 20.1%

Informazioni disponibili 21

Servizi al Cittadino: 23452

Eventi giornalieri: 29

Open Data disponibili: 143

Servizi duplic. 20

19.4%

Accessi ai servizi singoli (web e mobile) 22

Bus attivi 20

29

Ataf RT 20

64.1%

Parcheggi liberi 20

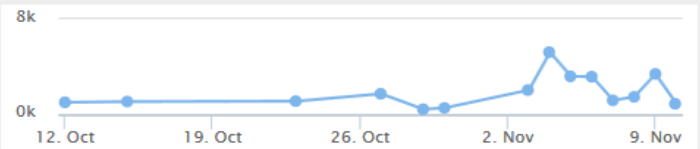
0%

Twitter trends/citazioni 21

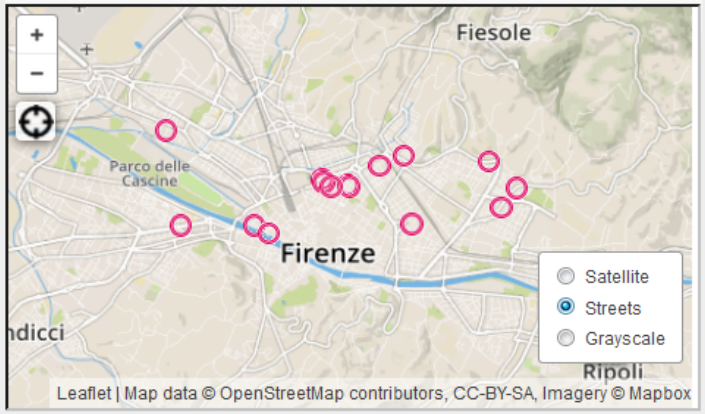
Principali Twitter Trends: #firenze #fiorentina #pontevecchio

Citazioni: @comunefi #fiorentina #pontevecchio

Rilievi wifi 22



Posizione in tempo reale degli autobus ATAF



Previsioni RT 20

98.9%

Previsioni meteo del comune di FIRENZE 299961

Tuesday	Wednesday	Thursday	Friday	Saturday
bit cloudy	mist	bleary	bit cloudy	cloudy
13°C / 19°C	9°C / 20°C	8°C / 20°C	/	/

Latest Update: 2015-11-10T08:35:00+01:00

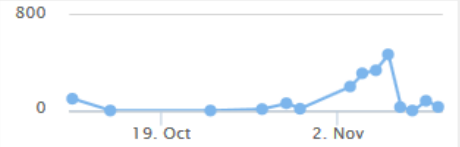
Parcheggi RT 20

66.0%

Smart City Engine 21

CPU	Mem	Job
0.1%	Totale	Eseguiti
	24034 MB	954

Query Servicemap 22



Stato corse ATAF 21

Linee: 6 17 4

in Orario	in Anticipo	in Ritardo
60.6%	17.1%	22.4%

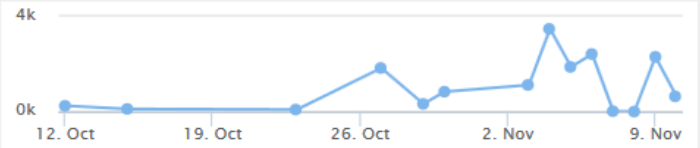
Sensori RT 20

99.3%

Eventi del giorno 22

- L'arte di Francesco - Capolavori d'arte e terre d'Asia dal XIII al XV secolo
- "The Medici Dynasty Show"
- TOSCANA '900
- Carlo Dolci 1616- 1687

Query Km4city API 22



twitter 2264 per giorno retweet 1378 per giorno Twitter Vigilance

http://www.disit.dinfo.unifi.it

Dati singoli ai dati aggregati ….

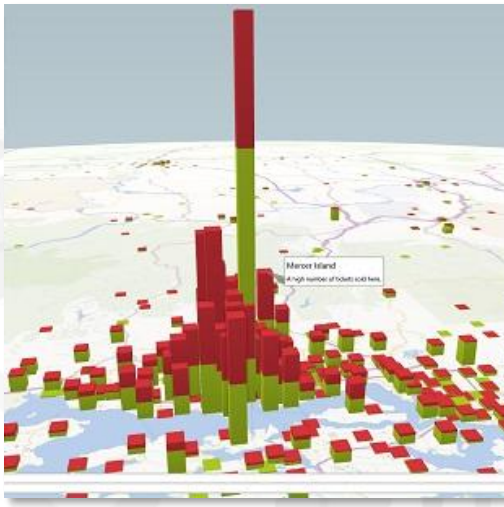
- Sistemi di **raccolta dati che devono essere integrati a livello semantico**
  - milioni di milioni di dati complessi arrivano ogni giorno alle centrali per essere analizzati: **Open Data, Real Time Data, Linked Data**

Sensori ed attuatori, sistemi di comunicazione, kit su veicoli

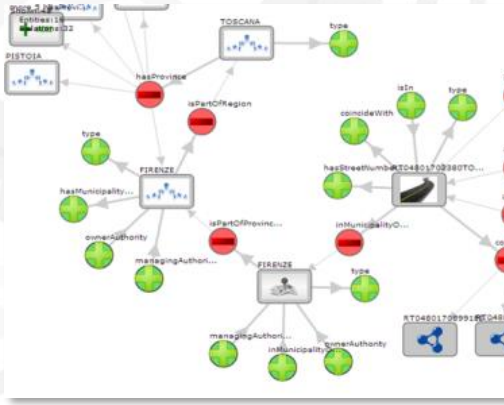
  - OD, sensori, social network, blog, etc.

Da dati aggregati alle deduzioni….

- Soluzioni di **intelligence per l'analisi dei dati**, per produrre in automatico:
  - deduzioni, correlazioni, implicazioni….
  - Supporto alle decisioni per le Pub. Ammin.
  - suggerimenti/raccomandazioni agli utenti anche in base ai loro profili (per esempio: medicina personalizzata), planning; alle aziende
- Soluzioni di **analisi per la comprensione di dati complessi**
  - fraseggi delle persone sulle social network, i commenti riguardo ai servizi della PA, le richieste di miglioramento dei servizi… (Natural Language Processing)
  - Comprensione di andamenti complessi da misurazioni puntuali (Data Mining, Knowledge Mining)



Microsoft geoflow



http://Log.disit.org

Infrastruttura della Smart City



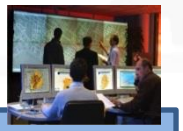


Gestore del traffico
Sensori
Energia
Ferrovie
Operatori Telefonia
……
…….
Social Media

centrali

Acquisizione dati
Aggregazione dati
Dati in Real Time
Social Data trends

Acquisizione dati e comprensione



Motore Smart City
Elaborazione
Real Time Computing

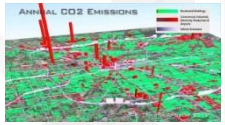
Ragionamento e deduzione

Servizi Profilati
Servizi Profilati
InFormazione dei cittadini
Data / info visualizzaz.
Data / info esport
Suggerimenti e allarmi

Azione e visualizzazione

Applicazioni

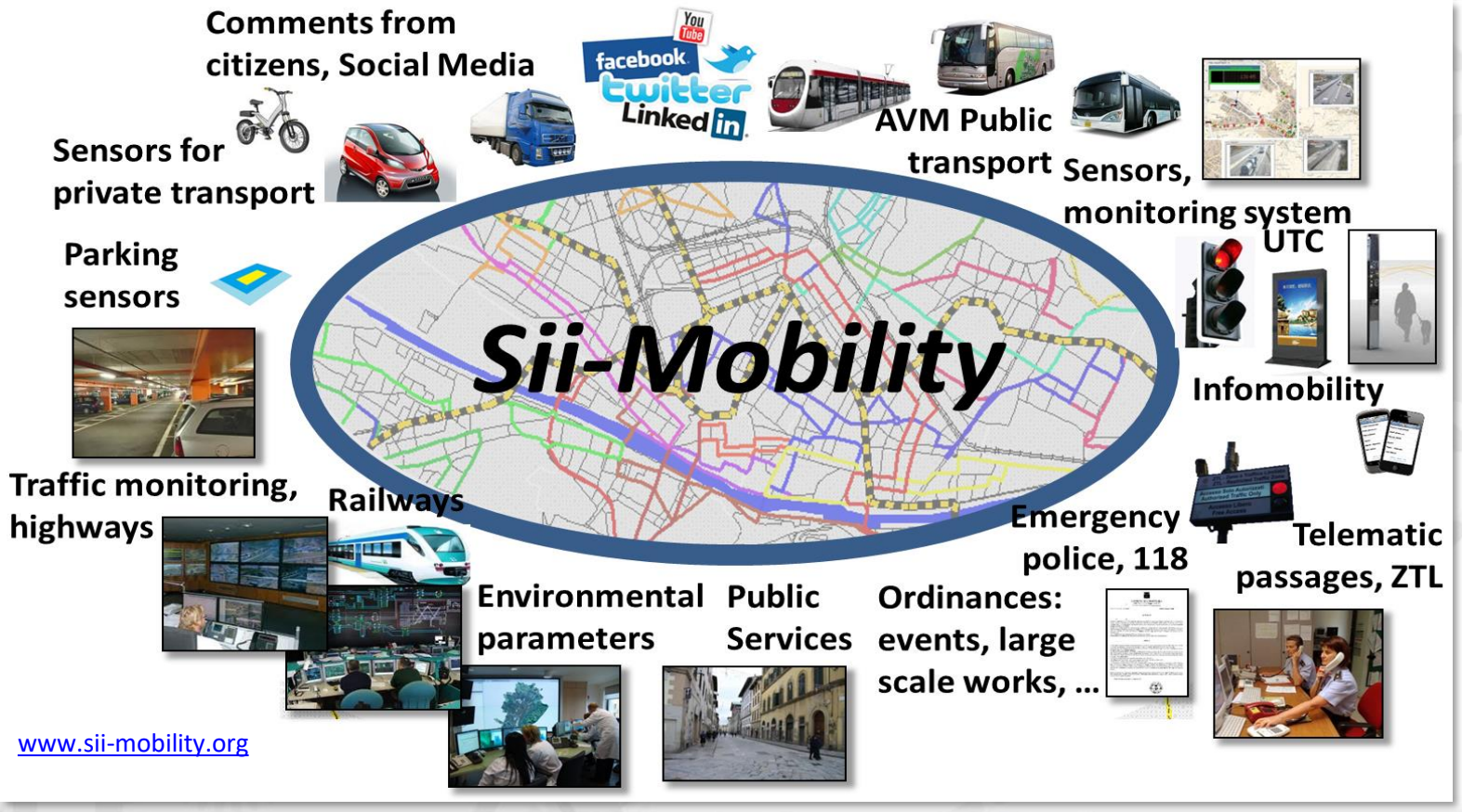





interconnessione


Relatore Paolo Nesi
Ordinario di Sistemi di elaborazione delle informazioni

- La sfida va verso ***l'integrazione di grosse moli dati non omogenei*** per produrre ***deduzioni più ampie e precise***
  - Dalle ***infrastrutture di monitoraggio e controllo***: energia, ambiente, salute, traffico, taxi, etc.



Sii-Mobility

- **servizi personalizzati**, connessi alla mobilità nella città
- Piattaforma di **partecipazione** e **sensibilizzazione**
- integrazione di **metodi di pagamento** e di identificazione
- gestione delle aree a traffico controllato
  - **dinamica dei confini**
  - **politiche di accesso**
- **interoperabilità** ed integrazione dei sistemi di gestione
- **scambio dati fra PA e privati**





Relatore Paolo Nesi
Ordinario di Sistemi di elaborazione delle informazioni


PROVINCIA DI PISTOIA
COMUNE DI PISTOIA
Provincia di Prato
COMUNE DI PRATO
COMUNE DI FIRENZE
PROVINCIA DI FIRENZE
Area Metropolitana
Firenze-Prato-Pistoia
Pontremoli
Fivizzano
Aulla
Carrara
Massa
Marina di Massa
Forte dei Marmi
Pietrasanta
Viareggio
Lucca
Pistoia
Prato
Montecatini Terme
Pescia
Pisa
Empoli
Firenze
Livorno
Cecina
Volterra
San Gimignano
Siena
Arezzo
Sansepolcro
Castiglioncello
Rosignano Marittimo
Castagneto Carducci
Massa Marittima
Piombino
Portoferraio
Marciana Marina
Porto Azzurro
Punta Ala
Follonica
Grosseto
Castiglione della Pescaia
Area Arezzo-Siena
Autostrade
SS Fi-Pi-Li
SS Fi-Si
Ferrovie (primarie)
Aree
Comune di Pisa
REGIONE TOSCANA
Comune di Perugia
COMUNE DI NAPOLI
Sii-Mobility

Km4City La Sfida

- **Grandi moli dai dati:** Open Data, Linked Data, Real Time data, sensori, social media Twitter, Wi-Fi, etc.
(**big data: velocity, variety, volume, veracity, …**)
  - Molti di questi dati sono **semanticamente non interoperabili**
  - Le città necessitano di creare **sistemi di supervisione e controllo** di più alto livello con strumenti open e non vincolati a tecnologie proprietarie.
  - Sono necessarie **competenze integrate** su un ampia gamma di problematiche: ICT, energia, protezione dati, mobilità, rischio, resilienza, supporto alle decisioni, etc.
- **Modello unificante**: la sola soluzione per fornire dei servizi integrati a Pubbliche Amministrazioni, ai cittadini, agli Operatori (mobilità, energia, acqua, telecom, etc.)
  - Complessità di avere ha una visione completa sui dati e sui modelli degli altri

Km4City: Aggregatore Dati

- **Fornire più informazioni di quelle che riceve in ingresso**:
  - Sfrutta soluzioni big data e di intelligenza artificiale: learning, self correction, reasoning, data mining per connessioni con altri dati
  - dati aggregati come un servizio via API
    - Fornisce servizi integrati
    - Supporto alla decisioni, suggerimenti, dashboard
    - →Valutazione del rischio, resilienza…

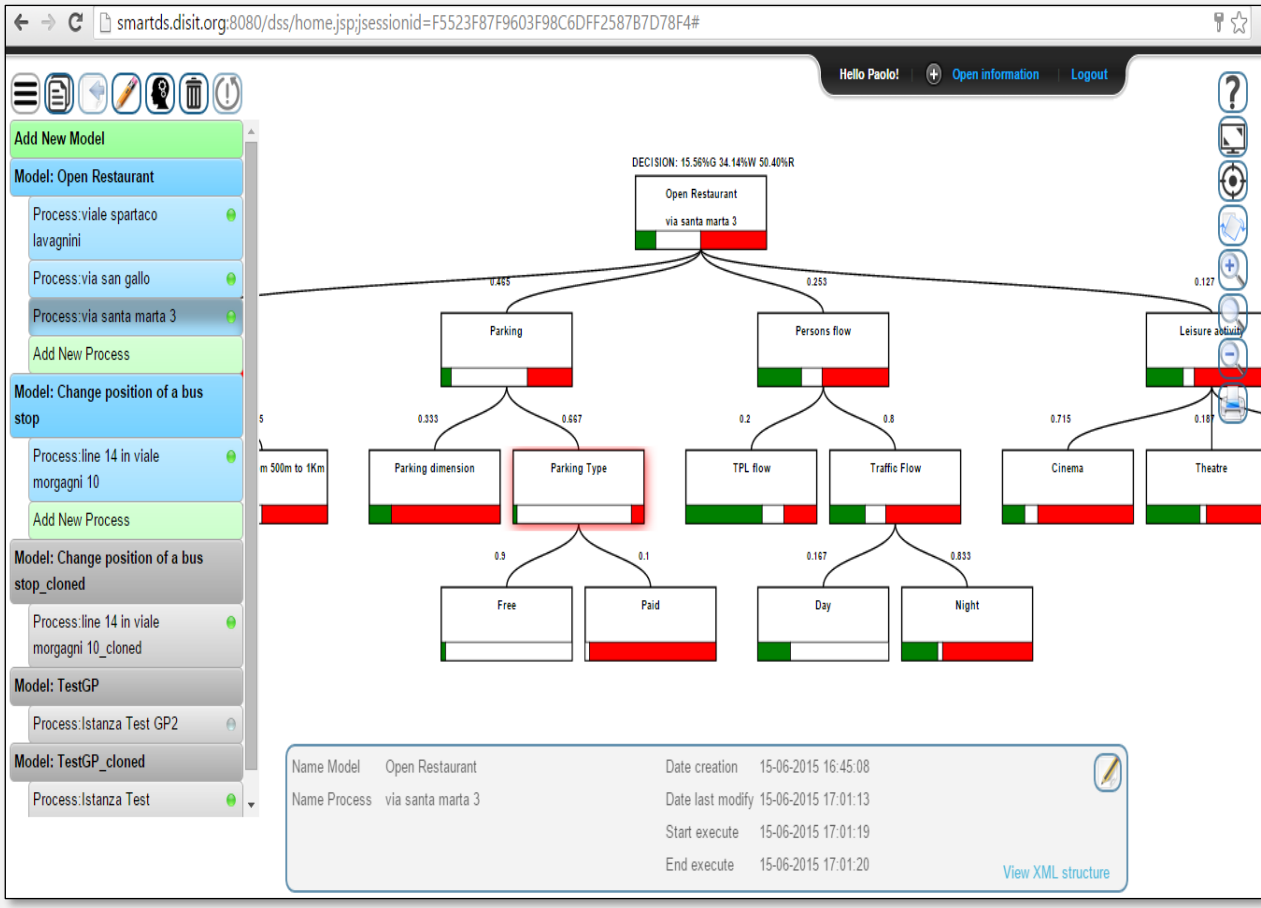
DISIT Lab, Distributed Data Intelligence and Technologies
Distributed Systems and Internet Technologies
Department of Information Engineering (DINFO)
http://www.disit.dinfo.unifi.it

Smart City Decision Support

- ***Smart Decision Support System*** *basato su System Thinking avanzato*

Integra

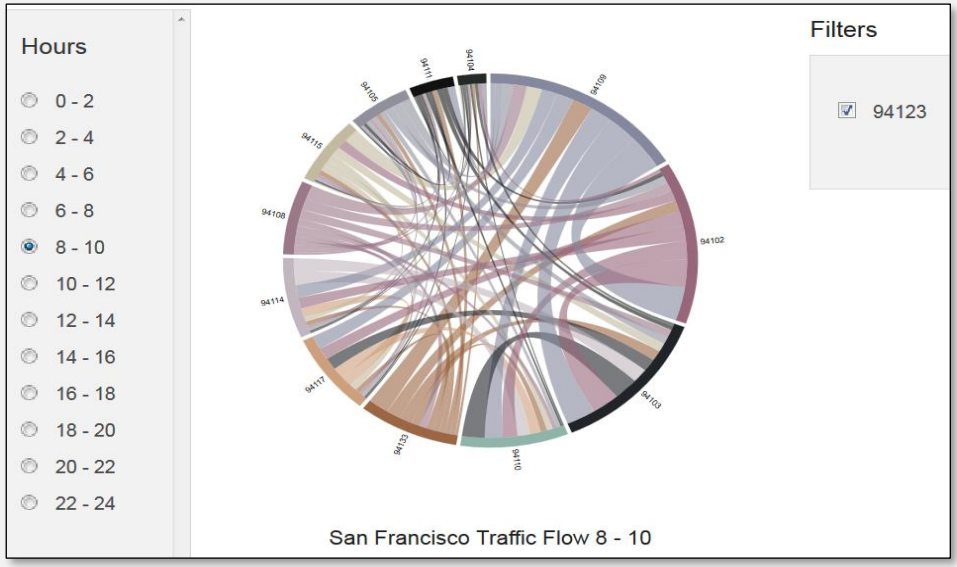
- Modello matematico per la propagazione della conoscenza di esperti e dei gruppi di lavoro
- Modello di lavoro collaborativo per il Supporto alle decisioni
- Processi decisionali misti che combinano esperti, social media, valori dai dati base, valori statistici, sondaggi, workshop, etc.
- Produzione e ricalcolo in tempo reale, produzioni di allarmi e alert
- Connessione a modelli computazionali, Twitter Vigilance, Km4City Data store, City DashBoard
- …

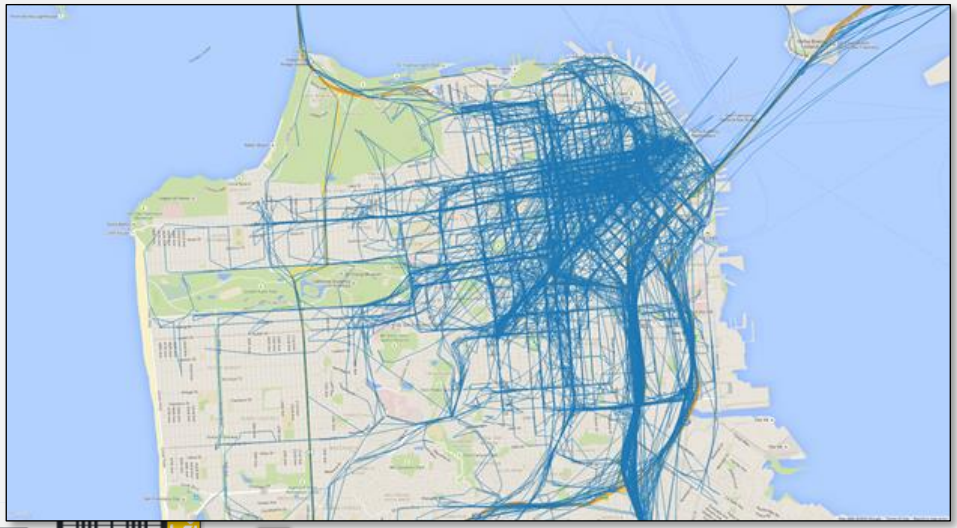


Traffic and People Flow Assessment

- **Mappe di Origine Destinazione**
  - Sensors, vehicle toolkit, mobile App, Wi-Fi Access Points, etc.
  - Calcolo delle matrici e contestualizzazione dei dati
- **Valutazione dei Flussi di Persone e veicoli**
  - Miglioramento dei servizi
  - Adattamento in tempo reale dei servizi,
  - Calcolo di percorsi ottimi
  - Predizione di condizioni critiche
  - Identificazione precoce di condizioni critiche
  - Supporto alla decisioni, connessione con SmartDS e Twitter
  - Incremento della Resilienza
  - Analisi sui servizi connessi alla mobilità in città
  - Analisi dei flussi turistici

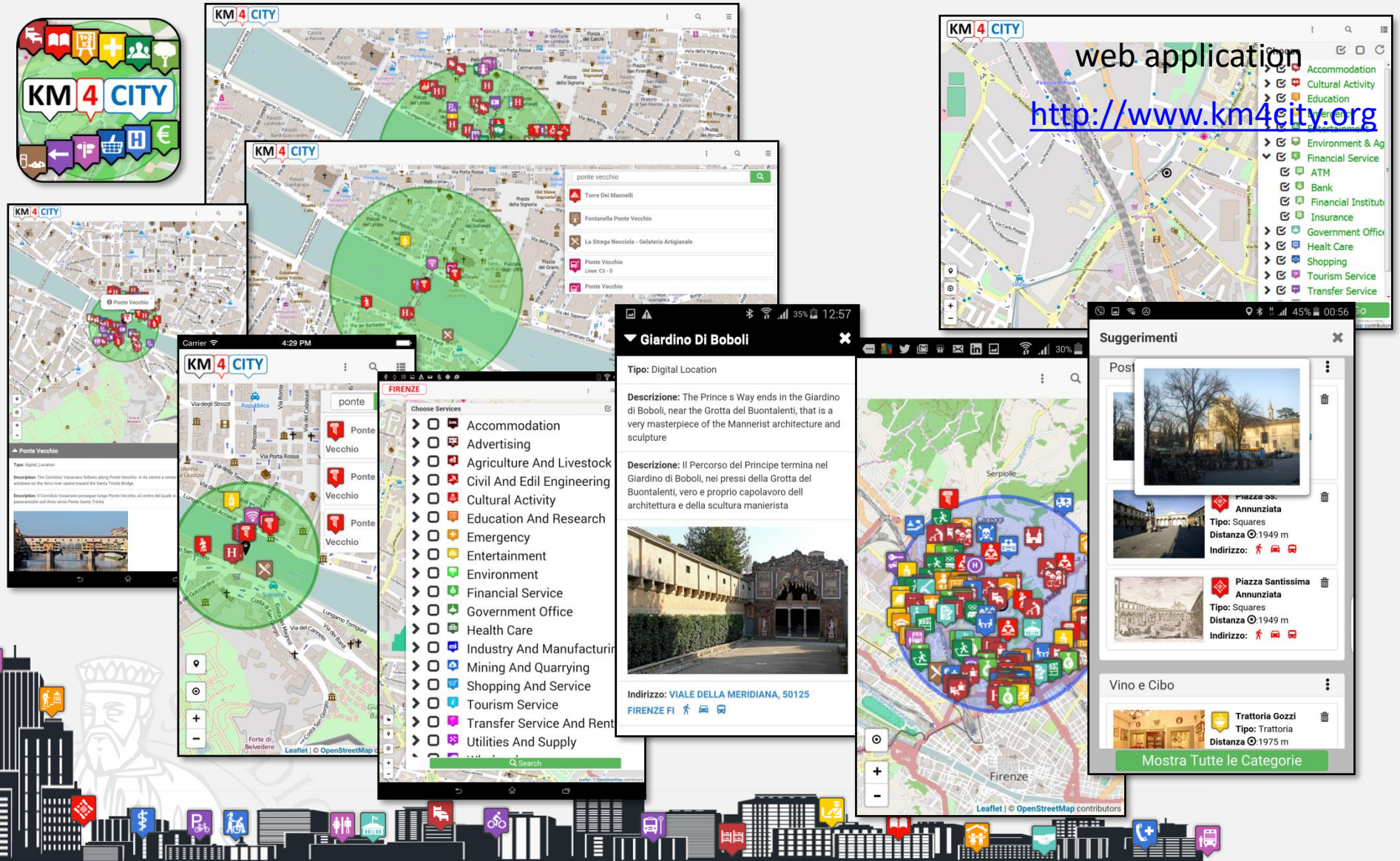
http://www.disit.org/6694





UNIVERSITÀ DEGLI STUDI FIRENZE
DISIT
DISIT Lab, Distributed Data Intelligence and Technologies
Distributed Systems and Internet Technologies
Department of Information Engineering (DINFO)
http://www.disit.dinfo.unifi.it

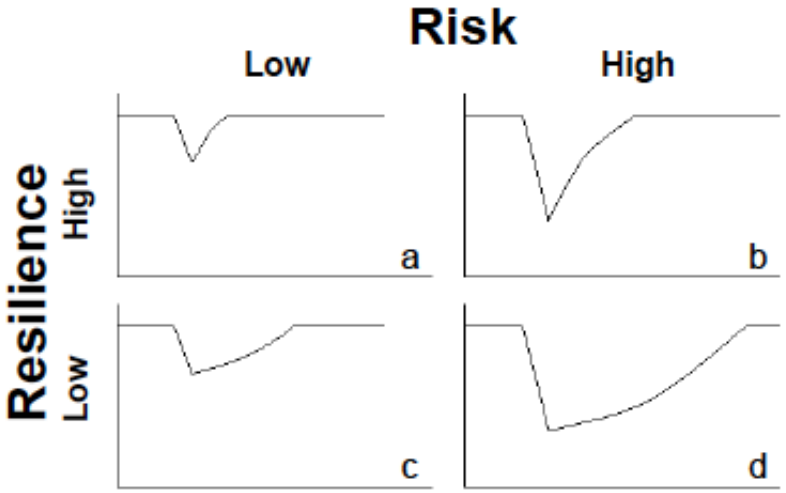
Km4CityMobile App: all stores



Co-funded by the European Union under H2020 DRS' 07-2014

Resilienza

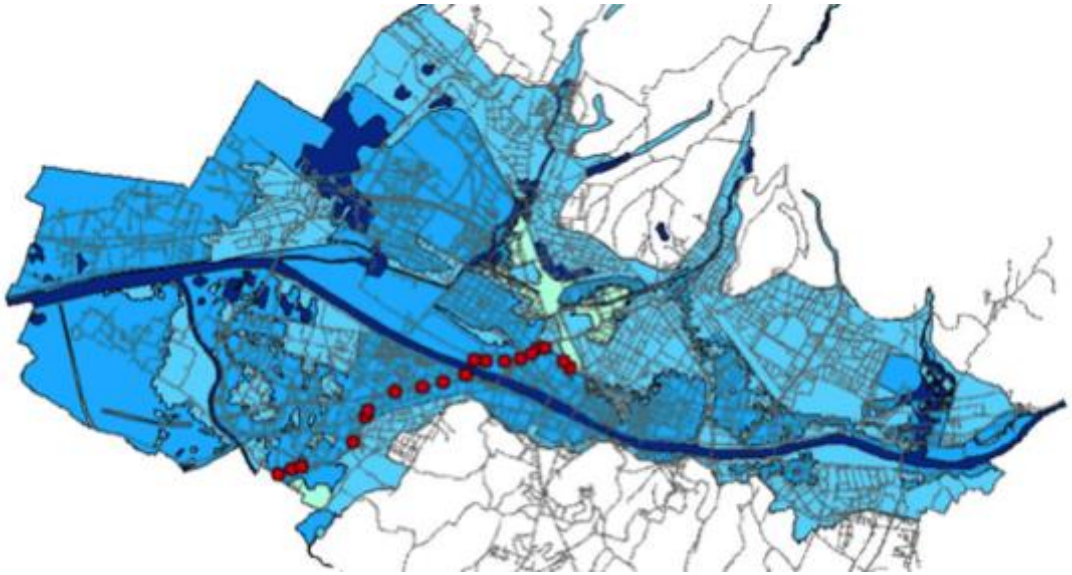
la *capacità di un sistema, una comunità o una società esposti a catastrofi di* ***resistere***, ***assorbire***, ***adattarsi*** *e* ***riprendersi*** *dagli effetti di una catastrofe in maniera efficiente e tempestiva, attraverso la protezione e il ripristino delle sue strutture e funzioni essenziali*.[UNISDR 2009]



Paolo Nesi, Emanuele Bellini, UNIFI
Smart City Meeting,
Firenze, 13 Novembre

Co-funded by the European Union under H2020 DRS' 07-2014

Firenze Pilota



Città di FIRENZE

Oltre il 70% delle infrastrutture cittadine sono a rischio idrogeologico

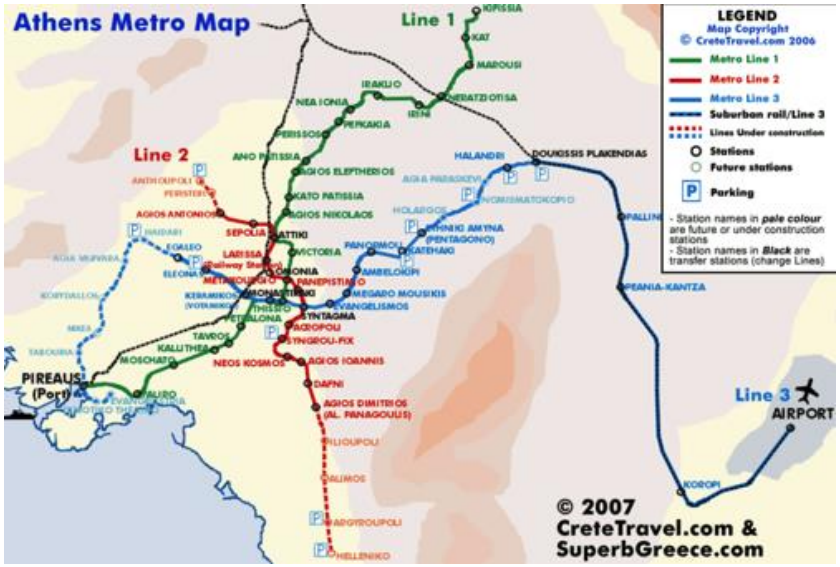
Definire modelli per analizzare il traffico per automatico e semi-automatico riconoscimento di eccezionali (flood) situazioni e relative azioni correttive attraverso l'applicazione delle ERMG definite nel progetto

Eseguire vari scenari di eventi di piena, al fine di valutare la resilienza del sistema di trasporto urbano e le interdipendenze attraverso l'integrazione del sistema CRAMSS con il sistema di gestione del traffico reale a Firenze

Paolo Nesi, Emanuele Bellini, UNIFI
Smart City Meeting,
Firenze, 13 Novembre

Co-funded by the European Union under H2020 DRS' 07-2014

Attiko METPO pilota



ATTIKO METPO A.E.

98Km linee
65 stazioni
1M passeggeri al giorno

- Esaminare le prestazioni delle regole e procedure di funzionamento della metropolitana esistenti per la gestione di eventi critici
- Eseguire vari scenari di minacce per valutare la resilienza del sistema di trasporti urbani di Atene attraverso l'impiego delle ERMG definite nel progetto
- Misurare l'impatto dei sentimenti umani negativi (ad esempio, la paura) sulla capacità di tenuta del sistema di trasporto urbano nel periodo post-attentato tramite un sondaggio preferenza

sommario

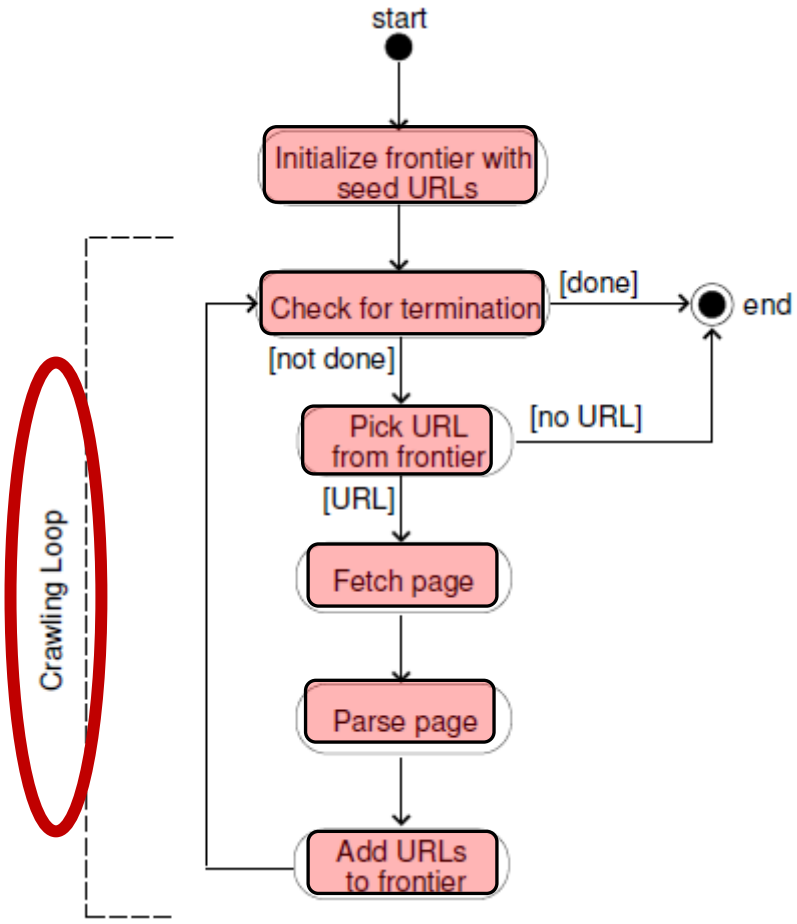
- *Dati e formati*
- *Comunicazione fra sistemi computerizzati*
- *Modelli di protezione e gestione*
- *Dati vs Metadati*
- *Social network e Smart City*
- *Motori di Ricerca, Crawling e NLP*
- *Data Mining*
- *Data Intelligence*

Motori di Ricerca e Crawling

- Cosa fa‘ un motore di ricerca
- Crawling
- Come viene fatto
- Indexing
- NLP: Natural language Processing

Web Crawler - Generalità

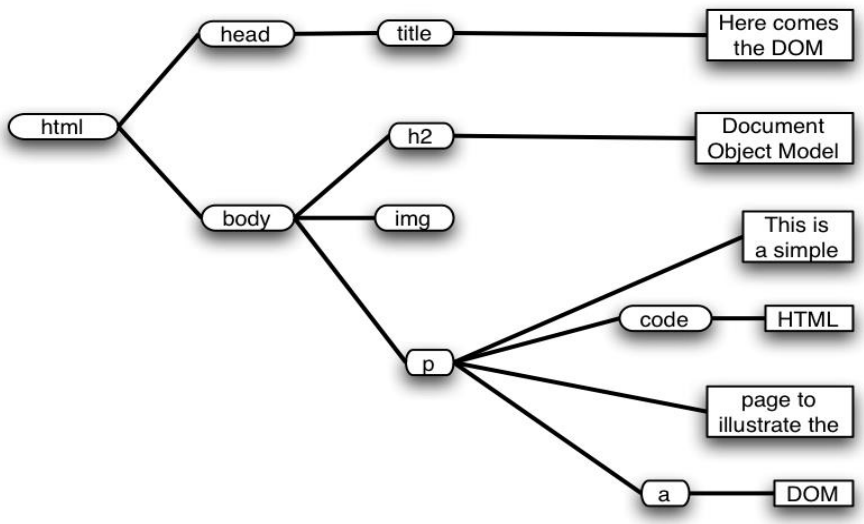
- Un web crawler è un programma per estrarre automaticamente l'informazione contenuta nel web
- E' utilizzato per creare una copia locale di tutte le pagine visitate per una loro successiva elaborazione (estrazione di informazioni, indicizzazione, ecc..)
- Il web è visto come un grafo: i nodi sono le pagine web e gli archi sono gli hyperlinks



Considerazioni implementative: Parsing

- I documenti HTML hanno una struttura ad albero - DOM (Document Object Model)
- Spesso i documenti HTML non rispettano gli standard di sintassi
- Occorre trattare le entità HTML e gli unicode nei testi
- Vi sono molti formati diversi di files:
  - Flash, SVG, RSS, AJAX…

```
<html>
  <head>
    <title>Here comes the DOM</title>
  </head>
  <body>
    <h2>Document Object Model</h2>
    <img align="right" alt="dom pict" src="dom.png">
    <p>
      This is a simple
      <code>HTML</code>
      page to illustrate the
      <a href="http://www.w3.org/DOM/">DOM</a>
    </p>
  </body>
</html>
```



Motori di Ricerca e Crawling

- Cosa fa‘ un motore di ricerca
- Crawling
- Come viene fatto
- Indexing
- NLP: Natural language Processing

Indexing

- **Indexing & Search system**
  - Based on Apache Solr
- **Multilingual aspects**
  - Translate the metadata or translate the query?.. both
    - metadata translation
    - Query translation
- **Indexing schema**
  - Dublin Core + DCTerms (multi language)
  - Performing Arts
  - Technical (provider, content type, GPS, IPR, duration, quality, …)
  - Groups associations (multi language)
  - Taxonomy associations (multi language)
  - Comments & multi language tags
  - FullText of the textual digital resources

Indexing

Media Types	DC (ML)	Technical	Performing Arts	Full Text	Tax, Group (ML)	Comments, Tags (ML)	Votes
# of Index Fields*	468	10	23	13	26	13	1
Cross Media: html, MPEG-21, animations, etc.	Y_n	Y	Y	Y	Y_n	Y_m	Y_n
Info text: blog, web pages, events, forum, comments	T	N	N	N	N	Y_m	N
Document: pdf, doc, ePub	Y_n	Y	Y	Y	Y_n	Y_m	Y
Audio, video, image	Y_n	Y	Y	N	Y_n	Y_m	Y_n
Aggregations: play lists, collections, courses, etc.	Y_n	Y	Y	Y/N	Y_n	Y_m	Y_n

* = (# of Fields per Metadata type) × (# of Languages)
ML: Multilingual; DC: Dublin Core; Tax: Taxonomy

Technical

N° accesses 28

Format video

Type video

Duration 00:12:48.0

Video quality available LD MD

Available platforms PC, iPhone/iPad, Android, Windows Phone 7, Windows Mobile 6.5

Upload date Sun, 2012-01-15 04:08

Group
Centro Teatro Ateneo, University of Rome La Sapienza, UNIROMA, Italia

Published by CTA-UNIROMA

Upload user marcomaci

Original filename ITUR1CTAVGA09074.mov

Workflow type Europeana

Content-url link to this content

QR code

axoid urn:axmedis:00000:obj:03b7454d-6b7f-464a-9050-1d42e98cff88

Classification

IPR information

Technical

Classification

Type Videosintesi [illegible] di prove dello spettacolo, a cura di Valentina Valentini

Language Ita

Source BVU

isPartOf Vittorio Gassman

Original metadata language it

IPR information

Rights
Ferruccio Marotti, Centro Teatro Ateneo – "Sapienza" Università di Roma, Eredi Gassman

IPR owner page link to page

Europeana rights Europeana: Unknown copyright status

Public No

PC permission

	Public	Group	Educ.	Trusted
Download HD PC	No	No	No	**Yes**
Play HD PC	**Yes**	Yes	Yes	**Yes**
Download LD/MD PC	No	No	No	**Yes**
Embed	No	No	No	**Yes**
Play LD/MD PC	Yes	Yes	Yes	**Yes**

Mobile permission

	Public	Group	Educ.	Trusted
Download mobile browser	No	No	No	**Yes**
Play mobile browser	Yes	Yes	Yes	**Yes**
Download mobile app	No	No	No	**Yes**
Play mobile app	Yes	Yes	Yes	**Yes**

Performing arts metadata

Location

Performing arts metadata

First performance date 1983-07-[illegible]

Performance place Teatro Ateneo

Performance city Roma

Performance country Italia

Performing Arts Group Compagnia del Teatro Manzoni diretta da Vittorio Gassman, con la collaborazione del 46° Maggio Musicale Fiorentino, della Bottega Teatrale di Firenze e dell'Estate Teatrale Veronese

Performers and crew Riprese: A. Muschietti; Tecnico audio e video: S. Casaluci; Interpreti e personaggi: Alessandro Esposito (Duncan, re di Scozia, un portiere), Danilo De Girolamo (Malcom), Roberto Medina (Donalbain), Vittorio Gassman (Macbeth), Carlo Montagna (Banquo), Luciano Virgilio (Macduff), Gian Franco Baroni (Lennox), Stefano De Sando (Ross), Luca Lazzareschi (Angus), Sergio Basile (Siward, conte di Northumberland, un capitano, un sicario), Lorenzo Gioielli (figlio di Siward, Fleance, figlio di Banquo), Alessandro Nisivoccia (un medico), Sergio Meogrossi (primo sicario), Federico Grassi (un sicario), Annamaria Guarnieri (Lady Macbeth), Giovanna Carcasci (Lady Macduff), Regina Senatore, Gabriella Chiari, Francesca Tardella, Franco Concilio, Franco Felici, Federico Grassi, Luca Lazzareschi, Stefano Molinari, Guido Paternesi (streghe). Scene e costumi: Paolo Tommasi; Musiche: Gianandrea Gazzola; Regia: Vittorio Gassman

Genre Teatro

Historical period XX Secolo

Location

Motori di Ricerca e Crawling

- Cosa fa‘ un motore di ricerca
- Crawling
- Come viene fatto
- Indexing
- NLP: Natural language Processing

Natural Language Processing NLP (1)

- **Scenario / Requisiti**
  - Dotare l'IA delle abilità linguistica proprie dell'essere umano
  - Comprensione e generazione del testo
  - Contesto multi-language: differenti regole e strutture a seconda della lingua

- **Applicazioni**
  - Generalizzazione delle query nei motori di ricerca
    - *"Chi si occupa di sistemi distribuiti nell'Università di Firenze ?"*
  - Supporto automatizzato per Help-Desk
  - Tutoring assistito (e-tutoring, e-teaching…)
  - Summarization: creare compendi da una collezione eterogenea di documenti
  - Machine translation: tradurre testi in lingue diverse

Ambiguità dei linguaggi naturali (1)

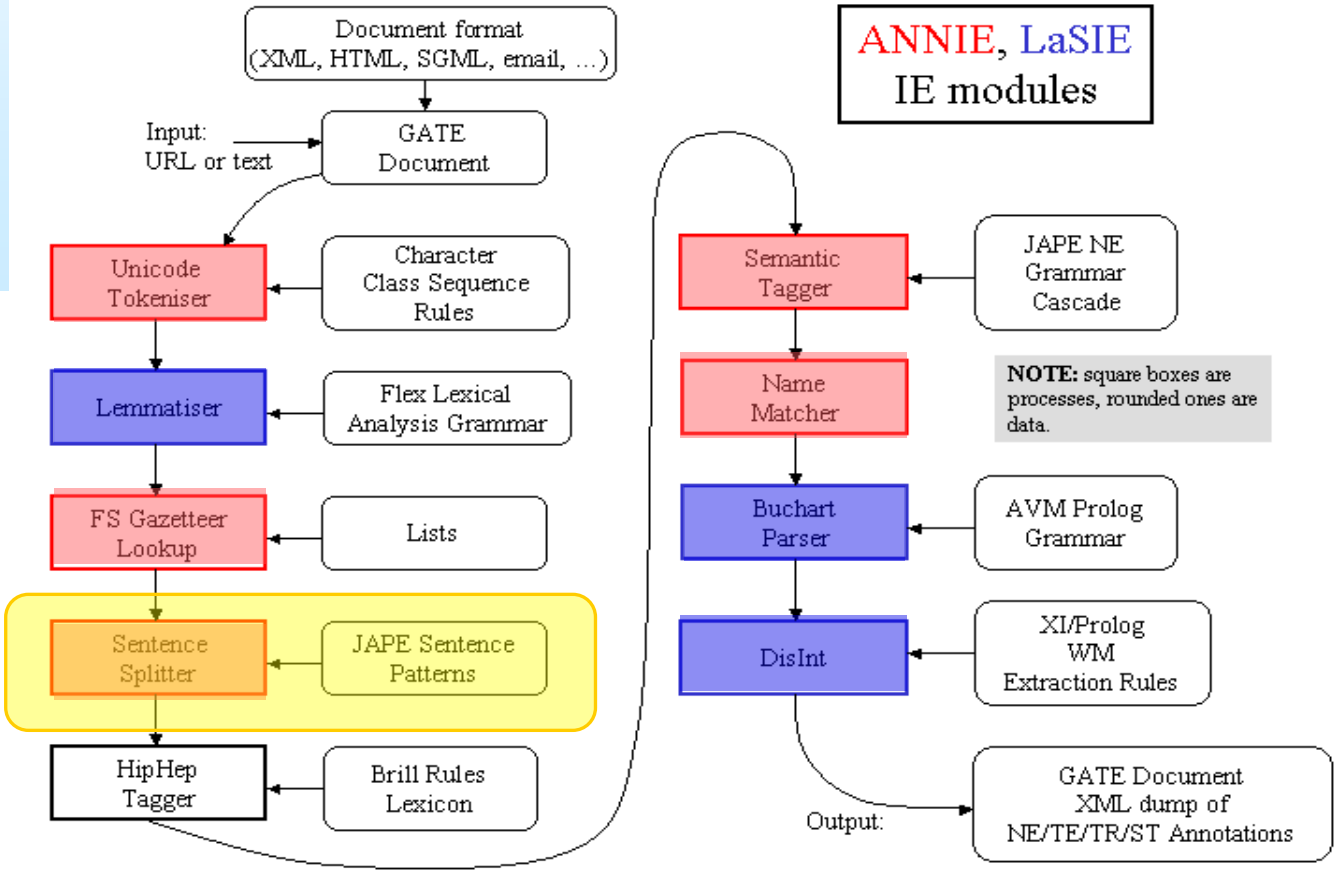
- **Scenario / Requisiti**
  - I linguaggi sono purtroppo ambigui.
  - Le ambiguità si possono avere a 4 livelli:
    - ✓ *Ambiguità lessicale*: *«attacco» (verbo, sostantivo)*
    - ✓ *Ambiguità strutturale*: *«Ieri ho visto l'uomo col telescopio»*
      «Una vecchia legge la regola»
    - ✓ *Ambiguità semantica*: *«acuto» (persona intelligente, tipo di suono)*
    - ✓ *Ambiguità pragmatica:* *«se Buffon non gioca contro la Spagna, l'Italia perderà»*

      L'intensione comunicativa viene recepita diversamente dagli interlocutori:
      - *interpretazione emotiva: l'assenza di Buffon è psicologicamente fondamentale per i tifosi*
      - *Interpretazione referenziale: l'Italia senza Buffon è più debole*

Ciò rende il processo di elaborazione del linguaggio naturale molto complicato

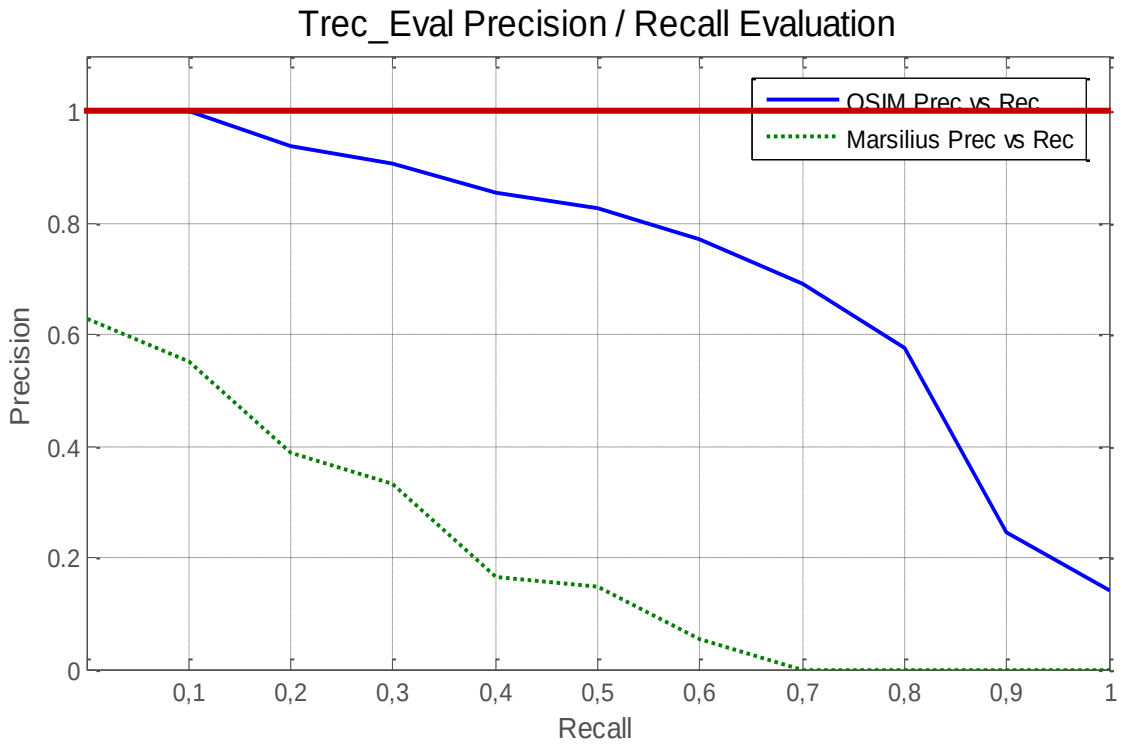
NLP Tools: GATE (1)

- **GATE – *General Architecture for Text Engineering***
  - Supporta documenti plain text, XML, RTF, HTML, SGML



Validazione (2)

- Set di query su un sottoinsieme di 4 dipartimenti
- Profondità dei risultati fissata N = 20
- Curva *Precision – Recall* ottenuta con il software standard *Trec_Eval*



Ottimo ideale

sommario

- *Dati e formati*
- *Comunicazione fra sistemi computerizzati*
- *Modelli di protezione e gestione*
- *Dati vs Metadati*
- *Social network e Smart City*
- *Motori di Ricerca, Crawling e NLP*
- *Data Mining*
- *Data Intelligence*

Data Mining

- Il processo di data mining consiste nel estrarre significato dai dati.
- Sono strumenti e processi di data mining (molti li abbiamo già visti):
  - Data harvesting and reconciliation
  - Semantic modeling and knowledge modeling
  - Web crawling
  - Blog crawling e vigilance
  - Sentiment analysis
  - Keyword extraction, KP extraction
  - NLP
  - Business intelligence

Data mining

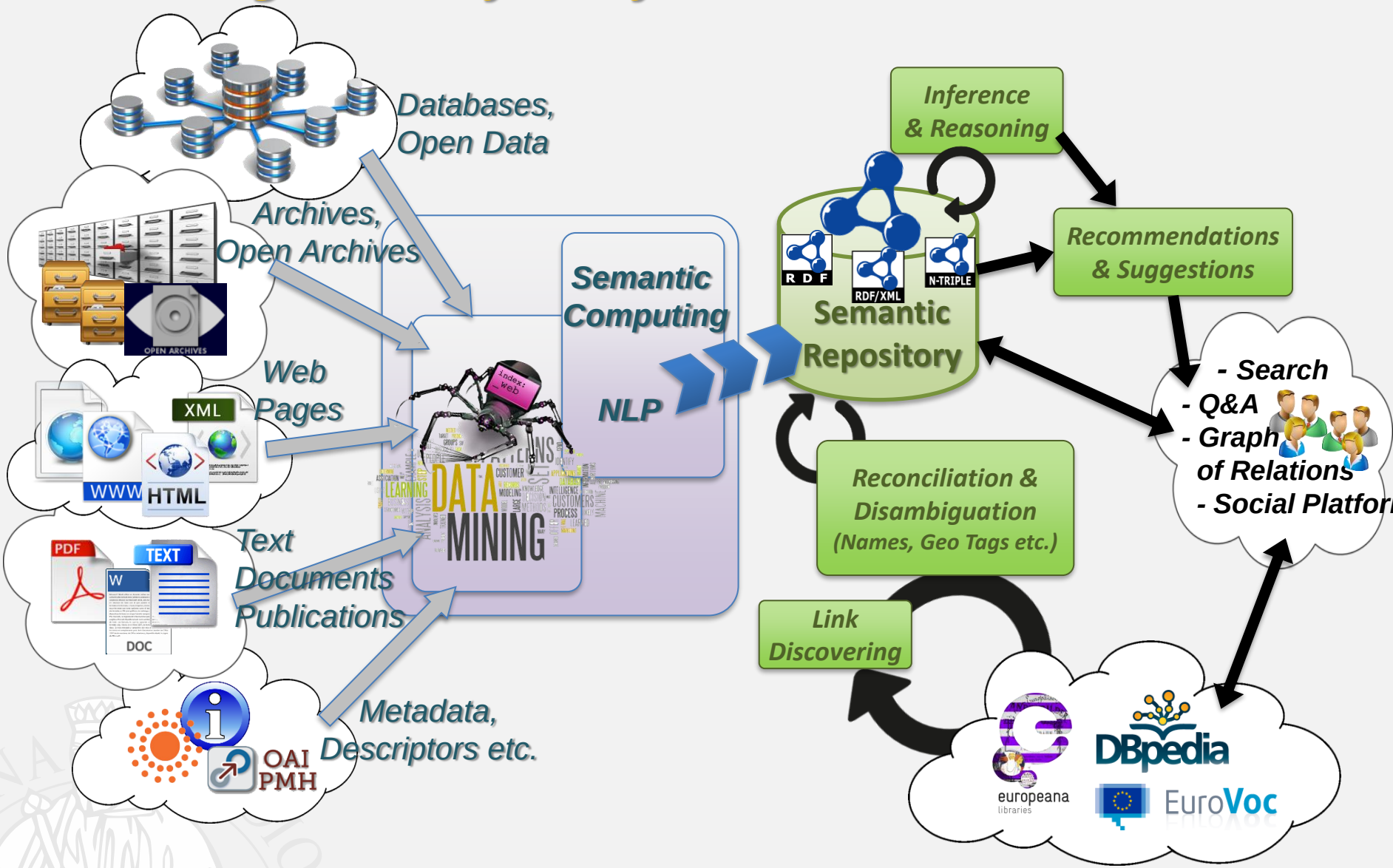
- Data base technology
- Knowledge modeling
- Information science
- Semantic computing
- Machine learning
- Pattern matching
- Visualization
- Statistics
- NLP
- …..

- Data acquisition
- Data integration and aggregation
- Data selection
- Data cleaning/filtering
- Data transformation
- Knowledge representation
- …..

Data integration

- Data warehousing provides the enterprise with a memory
- Data Mining provide the Enterprise with the basis for intelligence

Knowledge Work-flow: from Sources to Final Users



Localization via web crawling

- Using the **Ge(**o)**Lo(**cator) framework:
  - Mining, retrieving and geolocalizing web-domains associated to companies in Tuscany (thanks to a Distribute Web Crawler based on Apache Nutch + Hadoop)
  - Extraction of geographical information based on a hybrid approach (thanks to Open Source **GATE** Framework + using external gazetteers)
  - Validation in 2 steps: Evaluation of Complete Address Array Extraction, Evaluation of Geographic Coordinate Extraction
- New services found, can be transformed into RDF triples and added to the repository!

RDF Store Enrichment, for service Localization via web crawling

TABLE I. COMPARISON TABLE REPORTING EVALUATION DETAILS FOR EVALUATION TASK 1: ADDRESS ARRAY EXTRACTION, AND TASK 2: GEOGRAPHIC COORDINATES EXTRACTION (EMPLOYING BOTH THE SMART CITY REPOSITORY AND THE GOOGLE GEOCODING API).

Evaluation Tasks	TP	FP	FN	TN	Precision	Recall	F-Measure
1) Address Array Extraction	74.5%	7.8%	5.9%	11.8%	90.5%	92.7%	91.6%
2a) Geographic Coordinates Extraction (Smart City Semantic Repository)	57.8%	4.7%	29.5%	8.0%	92.5%	66.2%	77.1%
2b) Geographic Coordinates Extraction (Google Geocoding)	48.9%	31.1%	11.1%	8.9%	61.1%	81.5%	69.8%

- ***Precision*** rate for geographic coordinates extraction (employing the Smart City Semantic Repository) has increased, with respect to the value obtained in the evaluation of address array extraction.
- Slightly decreasing ***TN*** rate for Test (2a) with respect to Test (1): exploiting the extraction of high level features (such as building names) allows the system to obtain correct coordinates even for domains with incomplete Address Array.
- ***Recall*** rate for Test (2a) significantly decrease with respect to Test (1). This is due mainly to the noise generated by the supplementary logic and the extended semantic queries required to obtain the geographical coordinates.
- Higher ***Recall*** rate achieved when using the Google Geocoding APIs: Google Repository is by far larger than DISIT Smart City RDF datastore, so that it is able to index a huge amount of resources, even if this can affect the precision rate.

RDF Store Enrichment, VIP names identification

DISIT Lab, Distributed Data Intelligence and Technologies
Distributed Systems and Internet Technologies
Department of Information Engineering (DINFO)
http://www.disit.dinfo.unifi.it

- Searching RDF and/or MySQL stores looking for VIP names (citations) into strings:
  - *Via Leonardo Da Vinci*
  - *Piazza Lorenzo il Magnifico*
  - *Palazzo Medici Riccardi*
  - *Etc.*
- The idea is to link those entities with LD/LOD information as dbPedia information and

Uso del Data Mining

- Il Data Mining trova le relazioni di base (Modeling and representation) che in seguito sono utilizzate per algoritmi di:
  - Business intelligence
  - DSS: decision support system
  - Knowledge models and deduction queries
  - …

Data Mining Techniques

- Matchmaking
  - Matching demand and offer
- Classification and prediction
  - Recommendation, guessing new behaviors, fault prediction
- Cluster analysis:
  - Classification and recommendation
  - Market segmentation, collective intelligence, collective profiling
- Outlier analysis:
  - Fraud detection
- Etc.

sommario

- *Dati e formati*
- *Comunicazione fra sistemi computerizzati*
- *Modelli di protezione e gestione*
- *Dati vs Metadati*
- *Social network e Smart City*
- *Motori di Ricerca, Crawling e NLP*
- *Data Mining*
- *Data Intelligence*

I Dati e Metadati

- **Metadati associati ai dati** li contestualizzano, e.g.:
  - *Autore, data/time, località, user name, numero telefonico, etc.*
  - *Nomi delle persone coinvolte in una chat*
- I Metadati **accessibili per un servizio**
  - Possono permettere il congiungimento di tracce su servizi diversi.
  - Non è detto che siano riconciliabili con quelli di un altro, e.g.:
    - *Email vs numero telefonico, residenza vs user name su skype, etc.*

I Dati e vie di comunicazione

- **Dati e flussi** (real time, streaming e non) in relazione a server o persone singole, come da metadati
  1. **entranti ed uscenti dal paese (dati anche criptati o in chiaro)**
     - In US e altri paesi, questi dati vengono acquisiti, memorizzati ed analizzati per questioni di sicurezza nazionale.
  2. **interni alla nazione (da e vero server nel paese)**.
     - Questo accesso e registrazione dati è tipicamente visto come intercettazione, invasione della privacy, ed è giustificata a termini di legge.

Accesso Dati e Metadati

- Questi transitano o sono memorizzati in modo stabile o temporaneo su apparati e su server tipicamente
  - posizionati all'estero o nel nostro paese,
    - in cloud tipicamente
  - accessibili solo per:
    - Service provider (nazionali e/o internazionali)
    - gestori di rete a livello nazionale
    - fornitori di servizi

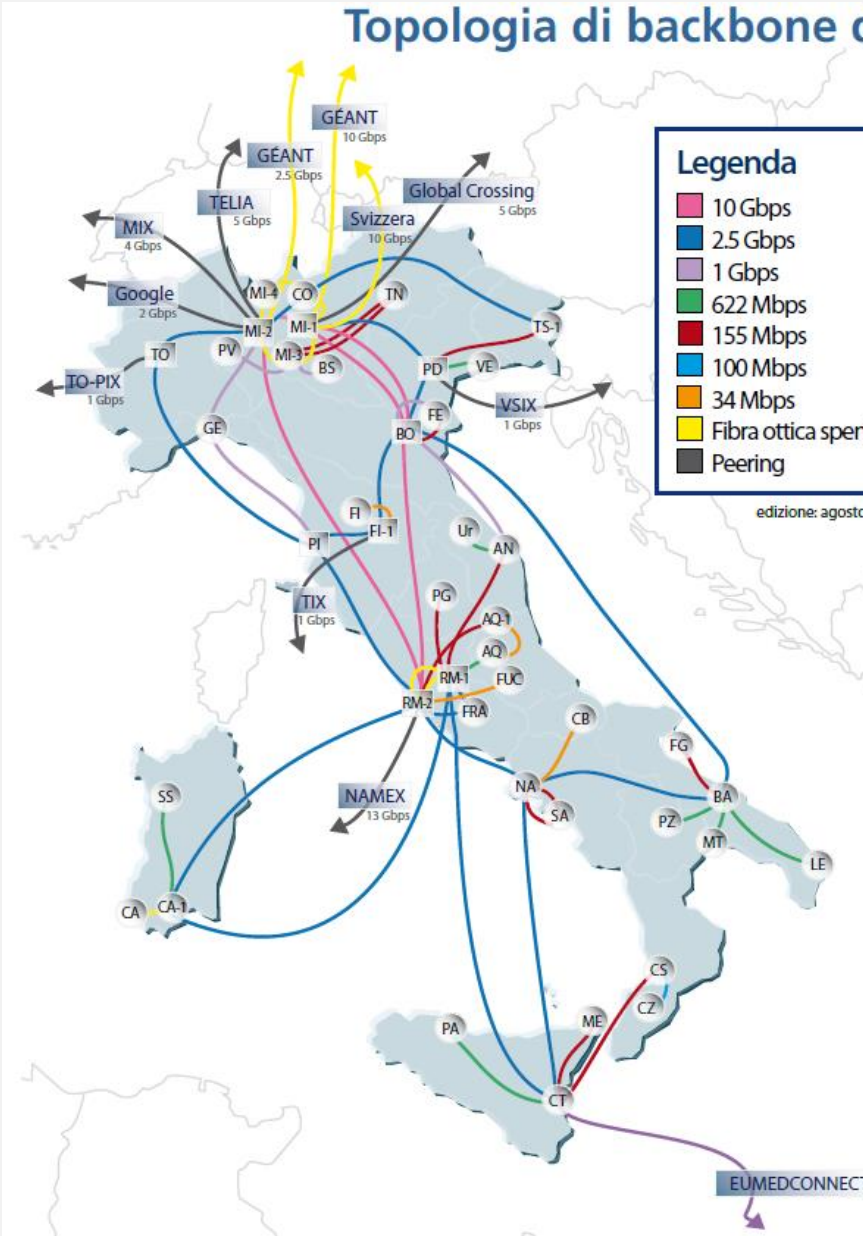
stream entranti ed uscenti dal paese

- **Accesso al flusso dati**:
  - I flussi entranti ed uscenti dal paese possono includere dati del tipo A, B, e C, anche codificati in modo criptato
  - Reti nazionali:
    - Rete pubblica: GARR
    - Rete privata tramite svariati operatori
  - Eventuali connessioni satellitari dirette

- L'accesso agli stream può fornire un dato complessivo più completo ma molto più costoso da ricostruire.
  - *Gli stream sono in tempo reale, e le informazioni di contesto sono complesse da estrarre ed ottenere*
- Per certe tipologie di dato è meno costoso estrarre i metadati ed il contesto dal server che eroga il servizio invece che a livello di stream:
  - per esempio: *è più semplice accedere ad email, chat, docs, etc. tramite il loro server, che dallo stream.*

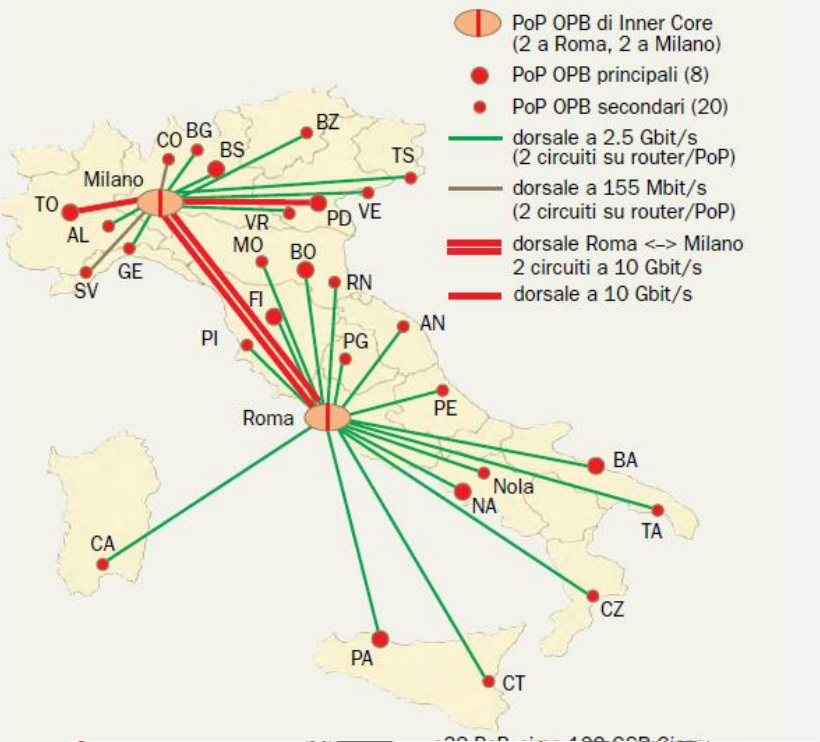
Connessioni Nazionali, GARR

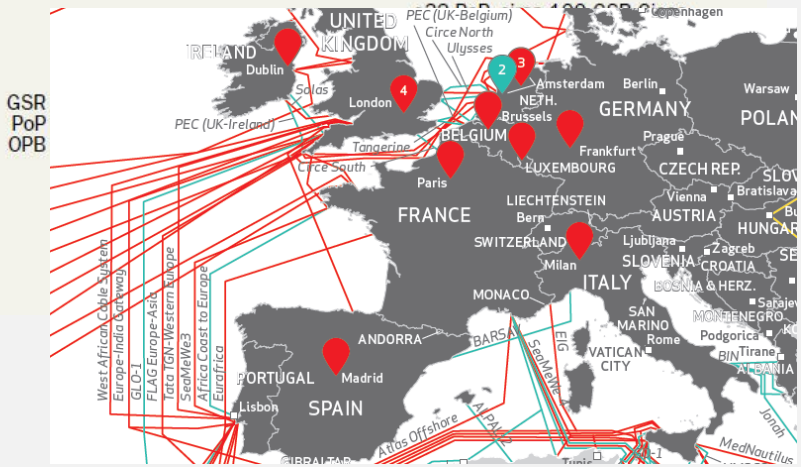
- pieno supporto ad applicazioni innovative quali griglie, telemedicina, e-learning, multimedia, fisica delle alte energie, radioastronomia, osservazione della Terra, supercalcolo ed offre servizi di rete avanzati quali VPN (Virtual Private Network), Multicast e IPv6.
- integrante del sistema mondiale delle reti della ricerca e collabora con le principali organizzazioni che operano nel campo del networking quali DANTE, TERENA, Internet2, IETF. È interconnessa con le altre reti della ricerca europee e mondiali, tramite collegamenti a 10 Gbit/s con la rete GÉANT, e con il resto dell'Internet commerciale con multipli collegamenti a 2.5 Gbit/s e a 10 Gbit/s.



Connessioni Nazionali, Private

- ISP provider, operatori hanno connessioni internazionali verso altre reti. Per esempio:
  - Telecom da Roma e Milano
- A livello nazionale dovrebbe essere possibile censire i flussi in/out di questi operatori nazionali
- Solo questi operatori possono accedere agli stream in entrata ed uscita dal paese





Data Mining in breve

- **insieme di tecnologie**: decodifica, integrazione (riconciliazione), comprensione, deduzione….
- **Decodificare** il dato:
  - dal formato trasmissivo (anche se criptato) ad uno comprensibile
- **Estrazione della semantica**
  - da formati a descrizioni semantiche, e.g.:
    - video → *descrizioni delle scene, riconoscimento facce, riconoscimento azioni ed oggetti, …*
    - audio → *testo, rumori, …*
    - Blog → *contesto e significato, …*

Comprensione del Dato

- **mettere in relazione ed integrare:**
  - i dati vs contesto (persone, data, luogo, etc.), le informazioni di contesto possono essere
    - nel contenuto: *citazioni di nomi, date, etc.*
    - informazioni in possesso del service provider, e pertanto vanno dedotte dal contesto sui server di servizio:
      - *chi ha aperto la connessione, chi ha telefonato, dove era, etc.*
      - *Nicola ha detto a Fra45 che «……..», Nicola e Fra 45 sono Mario Rossi, e Carlo Bianchi*
  - i dati elementari + metadati provenienti da server e stream diversi fra di loro per
    - identificare correlazioni, cause ←→effetti, predizioni, evoluzioni temporali, comportamenti tipici, etc.

Esempio NSA

- Sembra che vi sia un accordo fra il sistema di intelligence US/NSA e alcuni service provider per accedere a stream / dati + metadati di contesto:
  - Per esempio: *Apple, Google, Facebook, Microsoft, Yahoo, AOL, dropbox, etc.*
  - Questo non significa che NSA a un accesso diretto ai server
- **I service provider hanno accesso a dati diversi in funzione del loro business**, per esempio casi come per esempio:
  - *ISP provider (AOL): stream voip, contatti, …..*
  - *FaceBook-SN: profili, comportamenti, amicizie, contatti, posizioni gps, preferenze di vario tipo, …..*
  - *MS-game: contatti, date e time presenze/posizioni*
  - *MS-email: contatti, email,….*
  - *Dropbox-filesharing: file, contatti, …….*

Accedere ai Flussi e Dati

- **I Service Provider che agiscono su:**
  - **Flussi e connessioni** (*carrier, ISP provider, op. telefonia*..):
    - Non è detto che per i loro fini effettuino la registrazione /memorizzazione di tutti i dati del flusso che gestiscono
    - Devono avere apparati e programmi appositi per memorizzare tali dati e rendere possibile la ricerca mirata e contestualizzata.
      - Sono possibili anche registrazioni mirate per contesto, keyword, etc.
      - Invio di dati compressi per ogni ricerca
  - **Servizi e Dati**: SN, email, img, etc. → tipicamente su cloud
    - Tipicamente, memorizzano tali dati perché sono funzionali a fornire il loro servizio e suggerimenti/raccomandazioni
    - Devono avere apparati e programmi appositi per rendere possibile la ricerca mirata e contestualizzata.
      - Invio di dati compressi per ogni ricerca

Dati e Cloud

- **I Cloud dei Service Provider hanno**
  - storage e capacità computazionale estesa
- **Perché farlo nei loro cloud!:**
  - I flussi e i contesti sono complessi
  - Costi elevati di replicazione e dello storage
  - Costi elevati di banda / trasmissione del dato
  - Costi elevati di filtraggio del dato alla sorgente per ridurre la memorizzazione e i costi in generale, ma anche per mirare la ricerca a certe condizioni e casi, …
  - Aspetti legali relativi al trattamento del dato.
    - I service provider hanno accordi con i loro utenti che gli autorizzano alla memorizzazione dei loro dati personali (per un certo lasso di tempo e per certi fini) ma non hanno il permesso di condividerli, etc…..

Completezza dei dati dei Service Provider

- **Non è detto** che i dati associati all'erogazione di servizi dei Service Provider (come SN, email, chat, File sharing, etc.)
  - includano i dati di contesto necessari per le operazioni di integrazione e di intelligence, e.g.:
    - Per una SN non è indispensabile memorizzare ore minuti e secondi del momento in cui si è marcato come preferito un contenuto, o che altro.
    - alcune SN tengono conto solo del numero delle azioni e non delle singole azioni
  - siamo memorizzati per un lasso di tempo sufficiente, e.g.:
    - i testi delle chat dopo un certo lasso di tempo vengono cancellati
  - non siano già stati anonimizzati per garantire l'utente finale
- **Non è detto** che i service provider abbiano i dati con lo stesso dettaglio e qualità per tutti i loro utenti, e che siano quelli necessari per gli scopi di intelligence:
  - I fornitori di energia, acqua, gas, etc., non fanno monitoraggio dei consumi di tutti i loro utenti con un campionamento fine
  - I dati delle telecamere per la sicurezza non registrano tutto, etc.

Completezza dei dati dei Service Provider

- **Al fine di garantire la presenza e l'accessibilità di tali informazioni importanti, sono necessari accordi** con gli operatori per definire:
  - cosa è necessario accedere/conservare
  - il formato delle richieste (target) e di cosa deve essere prodotto come risultato verso i servizi di integrazione dell'intelligence

Integrazione del Dato e Ragionamento

- La **visione di insieme è può essere importante e fondamentale per fare delle deduzioni** con integrazione/riconciliazione di dati che provengono
  - da service provider:
    - diversi, e.g., *telefonata e email, registrazione account Apple, etc.*
    - in condizioni diverse*: nazionali ed internazionali*
  - da archivi istituzionali, dati pubblici in rete (web), open data, anagrafe, etc.
- **Questa operazione** in US viene svolta dall'FBI DITU, che
  - Integra e re-invia questi dati a FBI, CIA, etc.

Target vs Integrazione

- La definizione del Target è importante come l'integrazione del dato
  - Target può includere: lista dei service provider, canali di trasmissione, keyword, contesti, etc.
- Sulla base del Target, i dati ricevuti vanno passati ad una profonda analisi:
  - *aggregazione, comprensione, valutazione di completezza e consistenza*
  - Problematiche relative a:
    - Tecniche di data mining, intelligenza artificiale, interventi umani, Big Data, ….
    - Mole e complessità del dato, costi, tempi di elaborazione, competenze

Service provider e Copertura

- I Service Provider, anche se integrati fra di loro, **non hanno una copertura** al 100% sui dati nazionali anche solo di dati/flussi entranti/uscenti
  - Nazionali
  - Internazionali (posizione geografica dei loro server)
- **Accordi internazionali** di scambio e/o accesso ai dati dei server locali (anche localizzati in altre nazioni) possono permettere di completare un quadro di intelligence.
  - Si veda accordi presunti fra NSA e Service Provider in altre nazioni e/o altre nazioni
  - In Europa si potrebbe ipotizzare la necessità di avere degli accordi bilaterali multipli per definire protocolli di scambio e di accesso a informazioni.

Considerazioni generali

- Complessità ed incertezze dovute al:
  - processo di acquisizione, integrazione e comprensione dei dati ha dei margini di incertezza
    - mancanza di completezza e coerenza dei dati (certi dati sono sporadici)
    - Incertezze nei processi di data mining, che possono essere ridotte per integrazione, si una qualità e certezza sul dato variabile
  - difficile accessibilità dei dati, specialmente se nazionali
  - costi per la memorizzazione e l'elaborazione
    - dimensioni immense e crescenti dei dati che transitano
    - velocità e deperibilità del dato sui server
  - ….
- I Target sono dei desiderata più che delle *query di ricerca:*
  - Per questo motivo, gli investimenti per aumentare l'affidabilità del processo di intelligence sono molto elevati.
  - Spesso gli investimenti hanno coinvolto direttamente i Service Provider, per farli conformare a direttive sul trattamento dei dati e per predisporre soluzioni di accesso on demand

Quadro di Riferimento (CINI)

- Evoluzione della minaccia cyber
  - I professionisti nel crimine come servizio
  - Comportamenti e vittime: pervasività e vulnerabilità
  - Modus operandi e servizi diversificati e personalizzati
- Dinamiche del Cyberspace e Governance di Internet
- Multipolarità e Cyber-war
  - Canali, cloud, strutture intermedie, …
- Politica digitale e sicurezza informatica in Italia

Le raccomandazioni

- **Strategia, pianificazione e controllo**
- **Sicurezza come investimento**
- **Cyber Security Center - un'alleanza nazionale tra accademia, pubblico e privato**
- **Razionalizzazione del patrimonio informativo della Pubblica Amministrazione**
- **Formazione**
- **Certificazioni, Best Practices e Framework di Sicurezza Nazionale**

UNIVERSITÀ DEGLI STUDI FIRENZE
DINFO DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE
DISIT DISTRIBUTED SYSTEMS AND INTERNET TECHNOLOGIES LAB

Hacker vs cybersecurity: Quanto è sicuro lo spazio digitale delle nostre informazioni?

http://www.disit.org
Paolo Nesi, paolo.nesi@unifi.it
Tel: 335-5668674

Le nostre informazioni

- Profilo personale vs profilo collettivo
- Dati anonimi vs dati non anonimi

- Diretti: **spostamenti** (assicurazioni, telefoni,), **azioni**…
- Indirette: video, wifi touch, BT touch, RFID, etc.

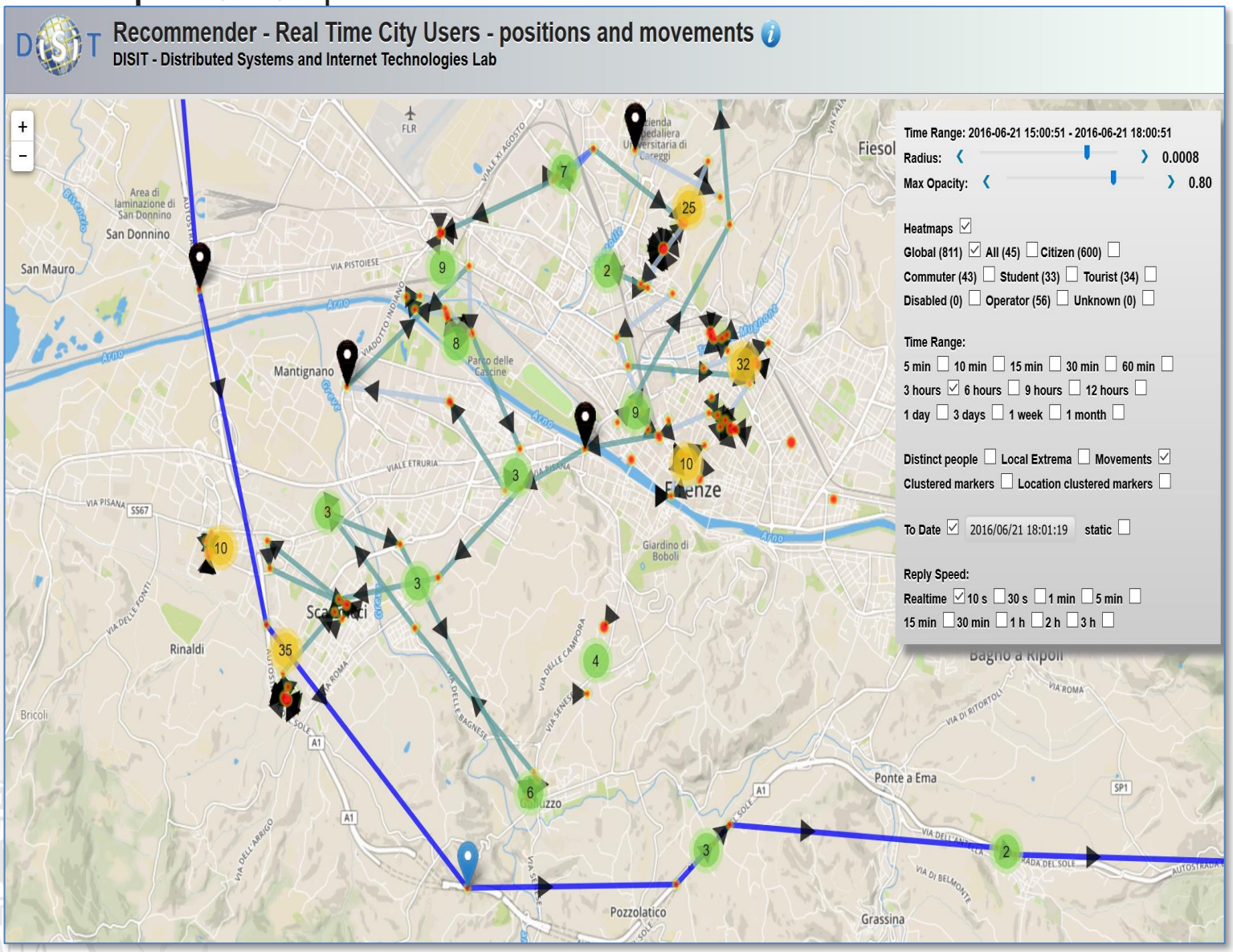
- **Comunicazioni**: telefoniche, email, sms, chat, ..
  - Ancora movimenti, ma anche preferenze, ..
- **Transazioni**: bancarie, commerciali, ...
  - Ancora movimenti, ma anche preferenze, ..
- Etc.

UNIVERSITÀ DEGLI STUDI FIRENZE

DINFO DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

DISIT DISTRIBUTED SYSTEMS AND INTERNET TECHNOLOGIES LAB

Traiettorie personali

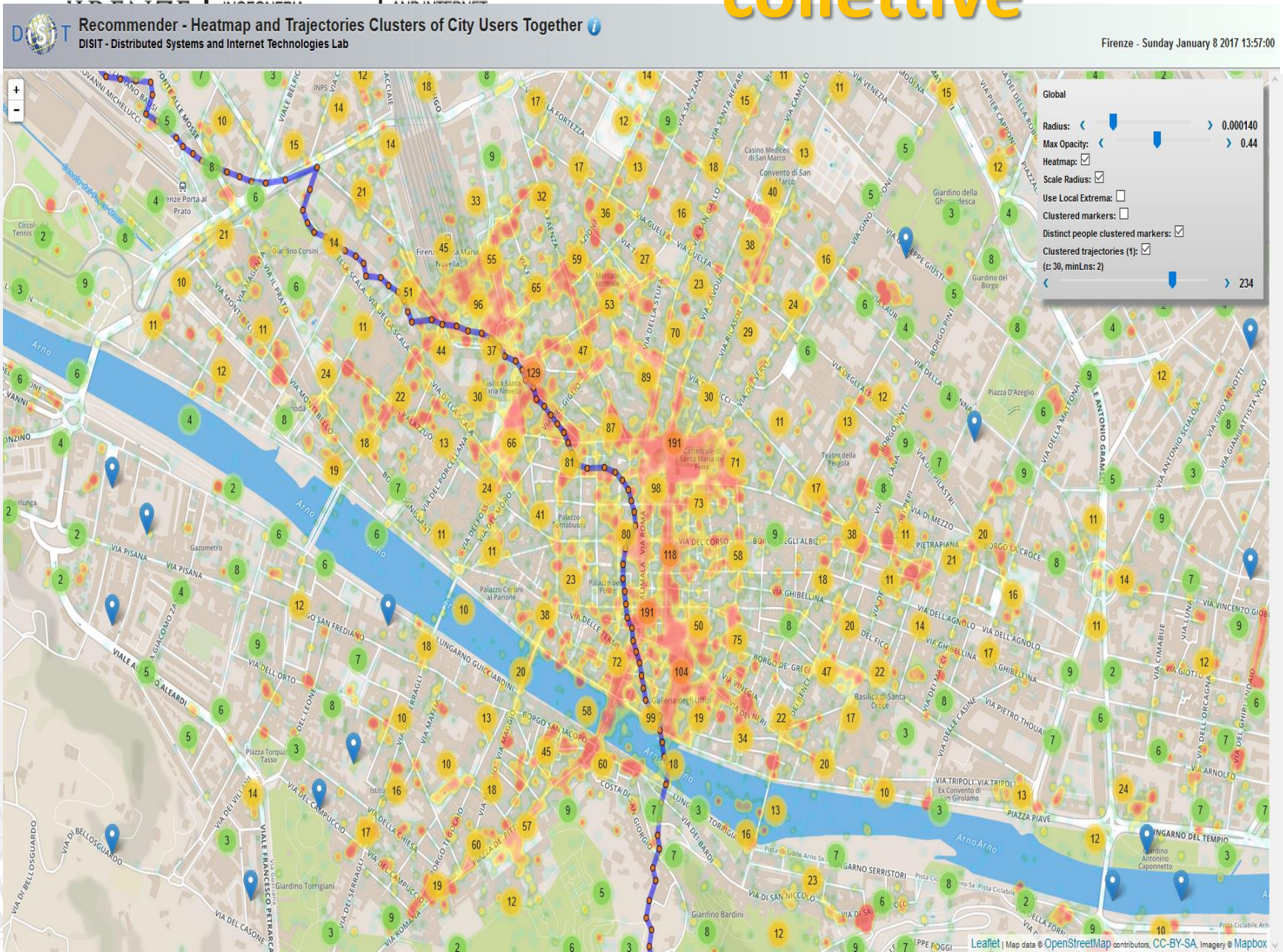


UNIVERSITÀ DEGLI STUDI

DINFO DIPARTIMENTO DI

DISIT DISTRIBUTED SYSTEMS

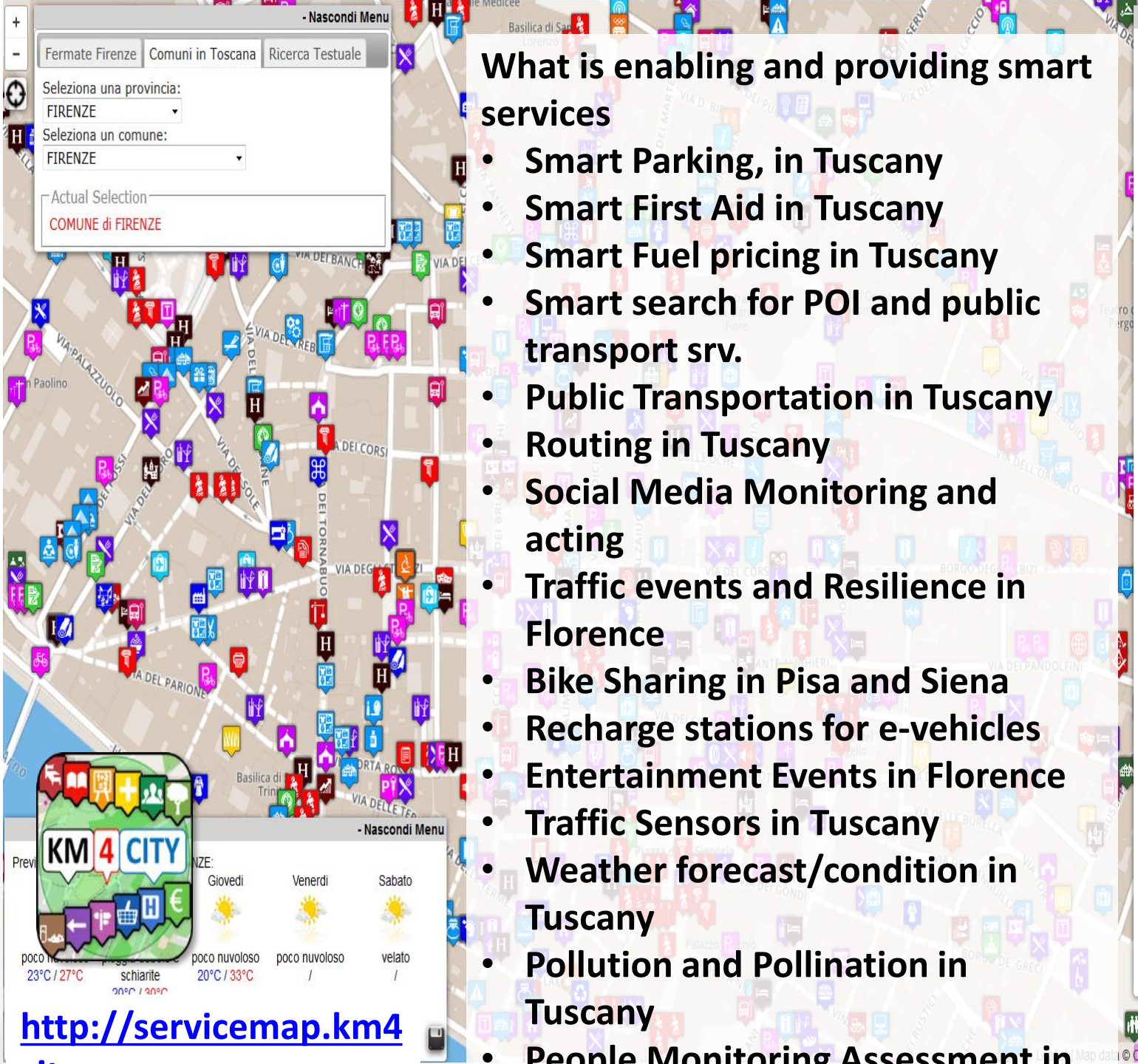
Traiettorie collettive



Profilo personale vs collettivo

- Benefici e ritorni (profilo collettivo vs personale)
  - Essere informati
  - Avere dei suggerimenti personalizzati
  - ..
  - Medicina personalizzata
  - ..
- → Termini d'uso
- → GDPR (General Data Protection Regulation)

Apprendere il modello collettivo per produrre suggerimenti, aiuto nei profili personali

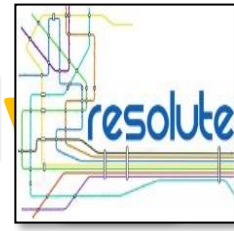


What is enabling and providing smart services

- Smart Parking, in Tuscany
- Smart First Aid in Tuscany
- Smart Fuel pricing in Tuscany
- Smart search for POI and public transport srv.
- Public Transportation in Tuscany
- Routing in Tuscany
- Social Media Monitoring and acting
- Traffic events and Resilience in Florence
- Bike Sharing in Pisa and Siena
- Recharge stations for e-vehicles
- Entertainment Events in Florence
- Traffic Sensors in Tuscany
- Weather forecast/condition in Tuscany
- Pollution and Pollination in Tuscany
- People Monitoring Assessment in

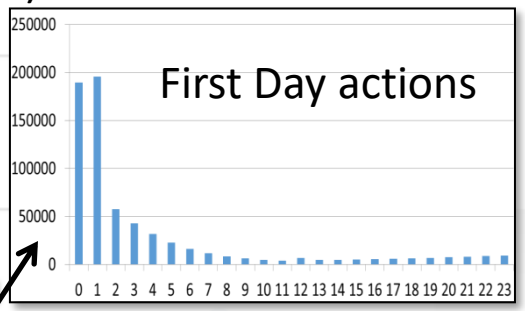
http://servicemap.km4

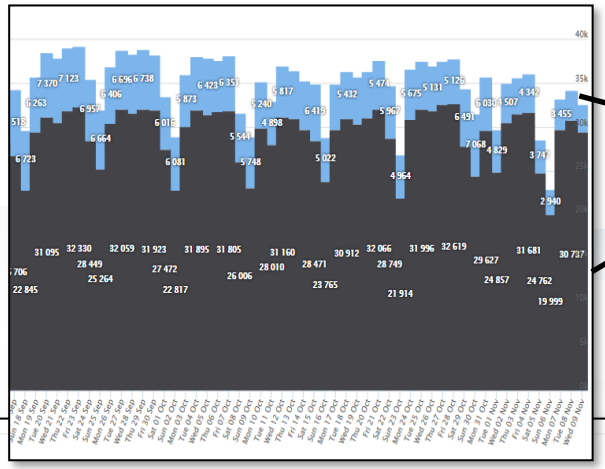
User Behavior Anal



Distinct APs: 343
Distinct APs (last 24 hours): 311
Distinct Users (last 180 days): 1102098
Distinct Excursionists (last 180 days, < 24 h): 687025

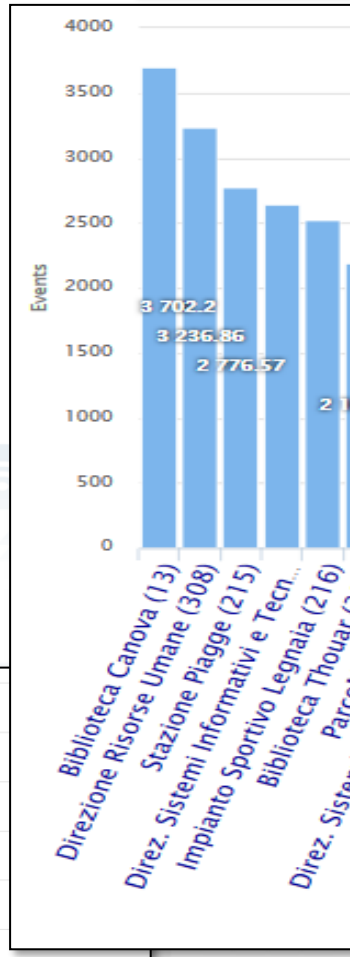
Excursionists



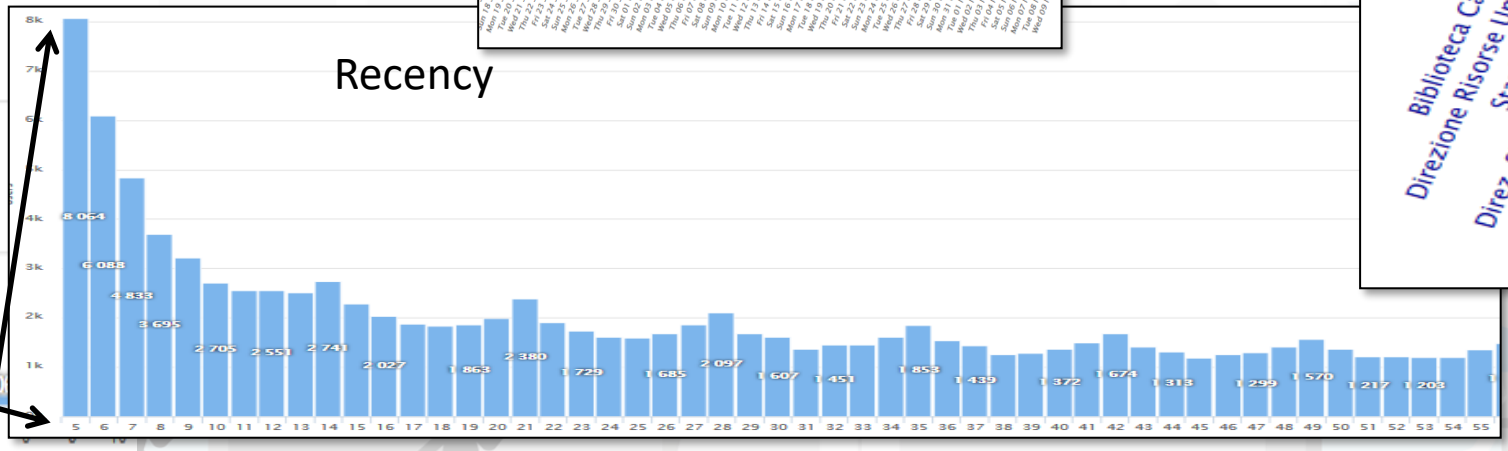


New City Users VS Returning

Where



Recency



UNIVERSITÀ DEGLI STUDI FIRENZE

DINFO DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

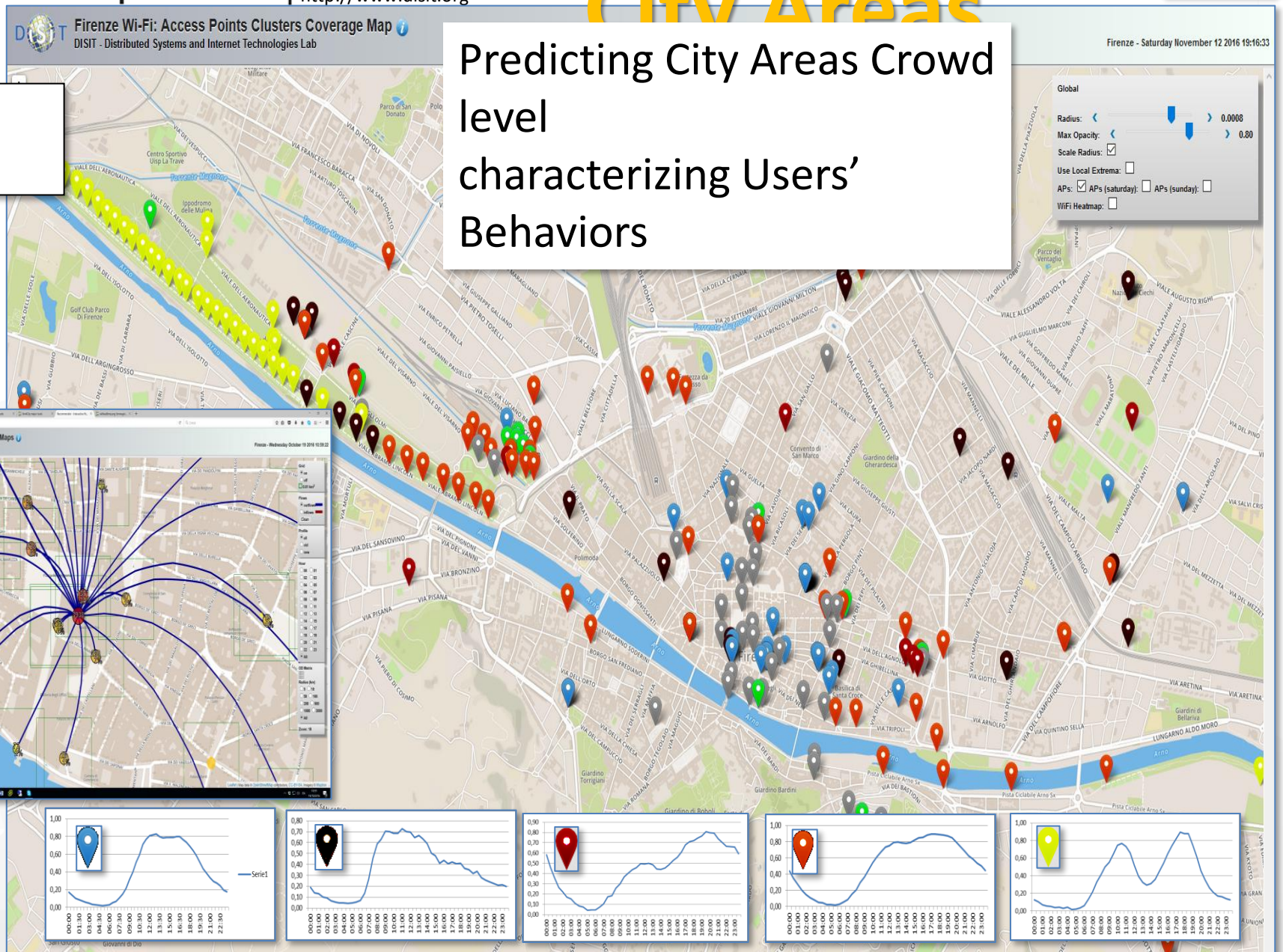
DISIT DISTRIBUTED SYSTEMS AND INTERNET TECHNOLOGIES LAB
http://www.disit.org

Characterizing City Areas

resolute

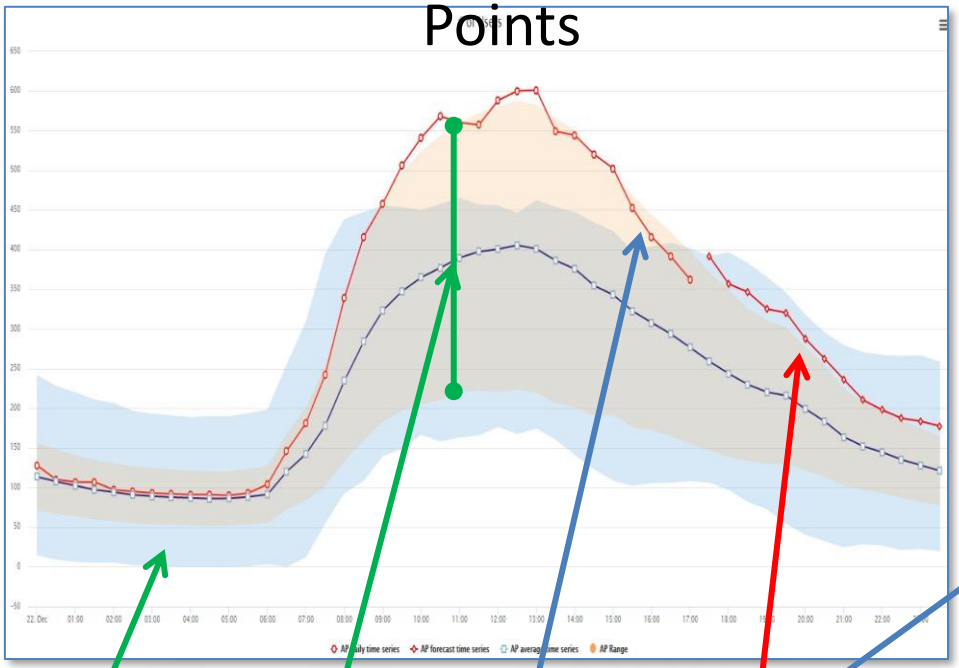
Wi-Fi based

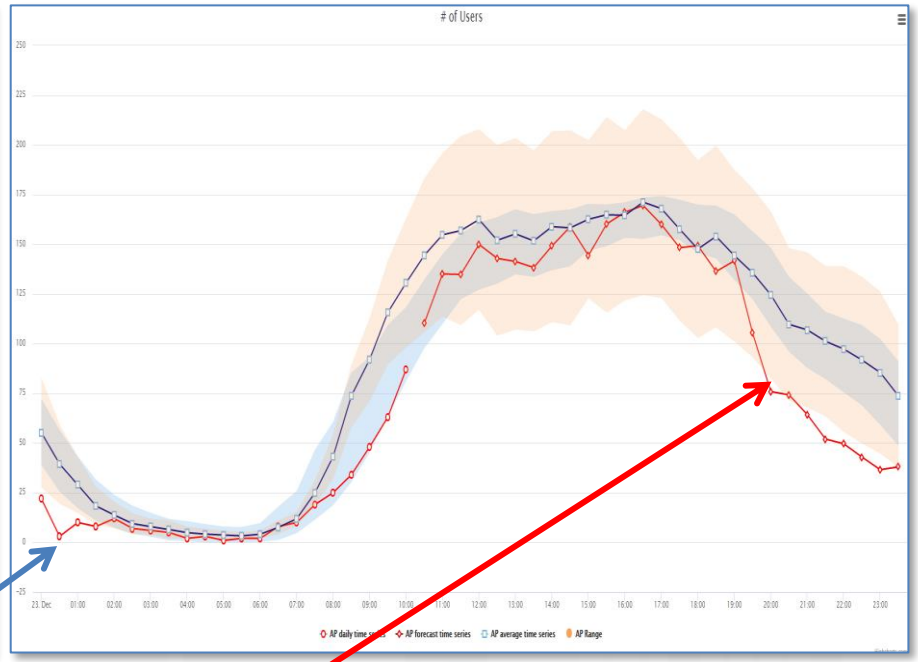
Predicting City Areas Crowd level characterizing Users' Behaviors



Prediction and identification of anomalies

Guessing number of users of Wi-Fi Access Points





Cluster confidence

AP average and confidence

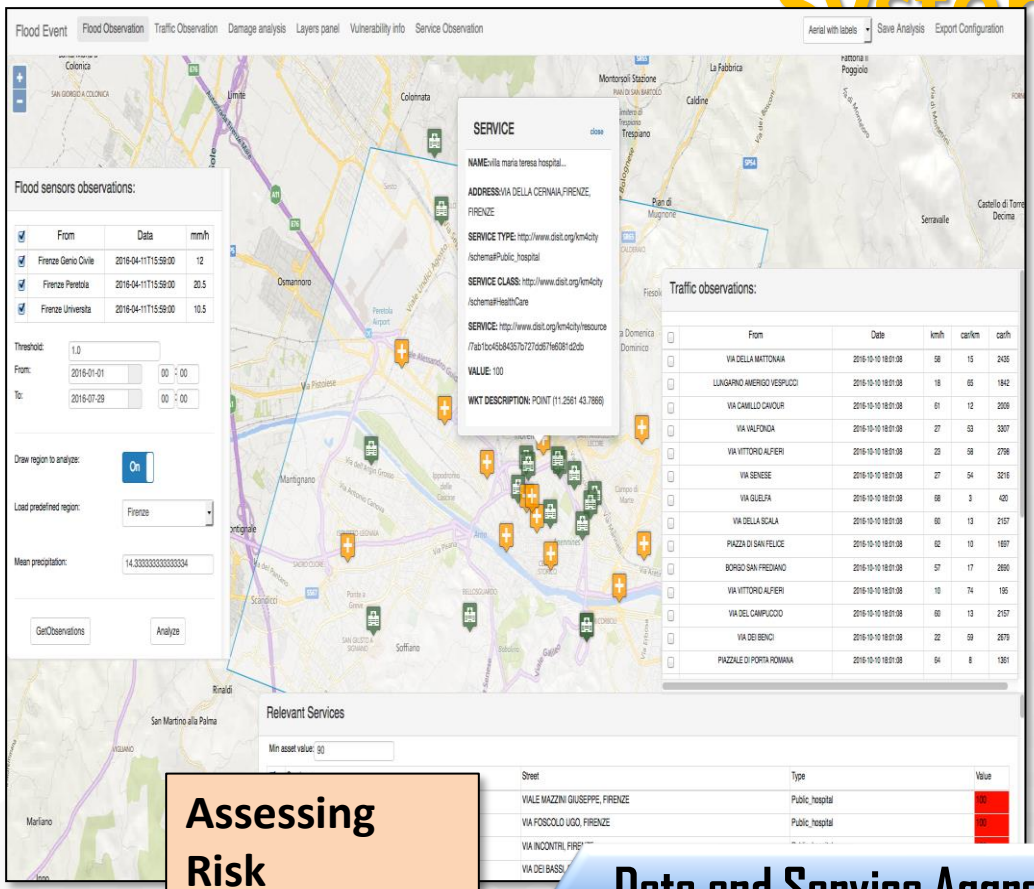
Actual AP trend for today

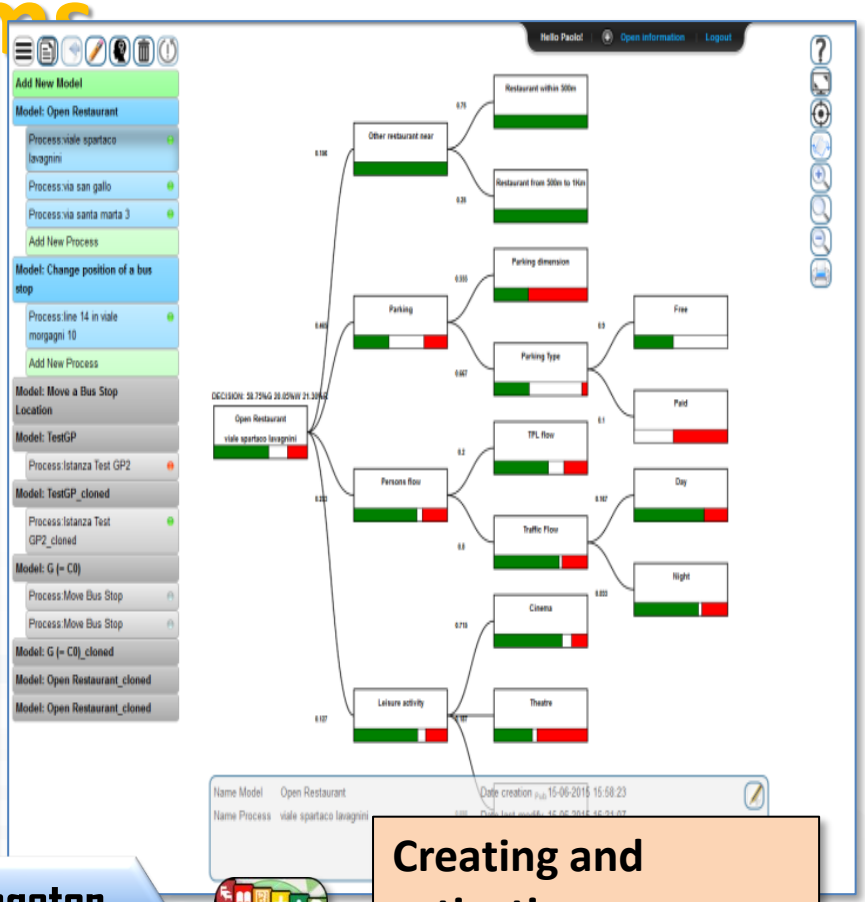
AP prediction for the next time slot in the day on the basis of past

UNIVERSITÀ DEGLI STUDI FIRENZE
DINFO DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE
DISIT DISTRIBUTED SYSTEMS AND INTERNET TECHNOLOGIES LAB

Risk Assessment

Decision Support Systems





Assessing Risk

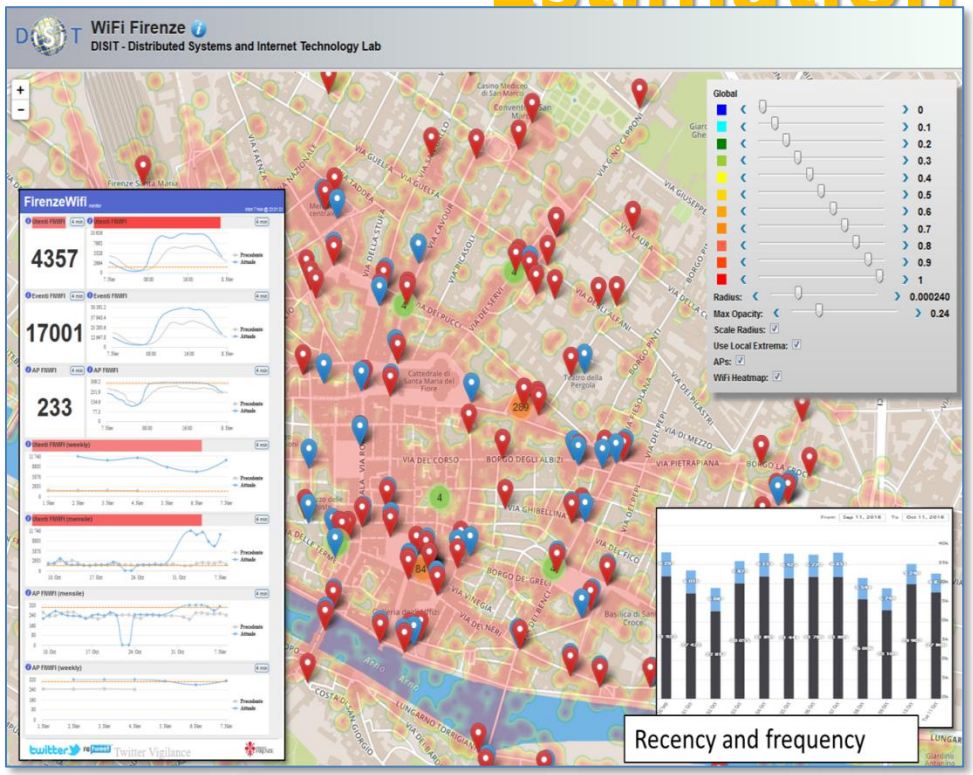
- hydraulic
- Seismic

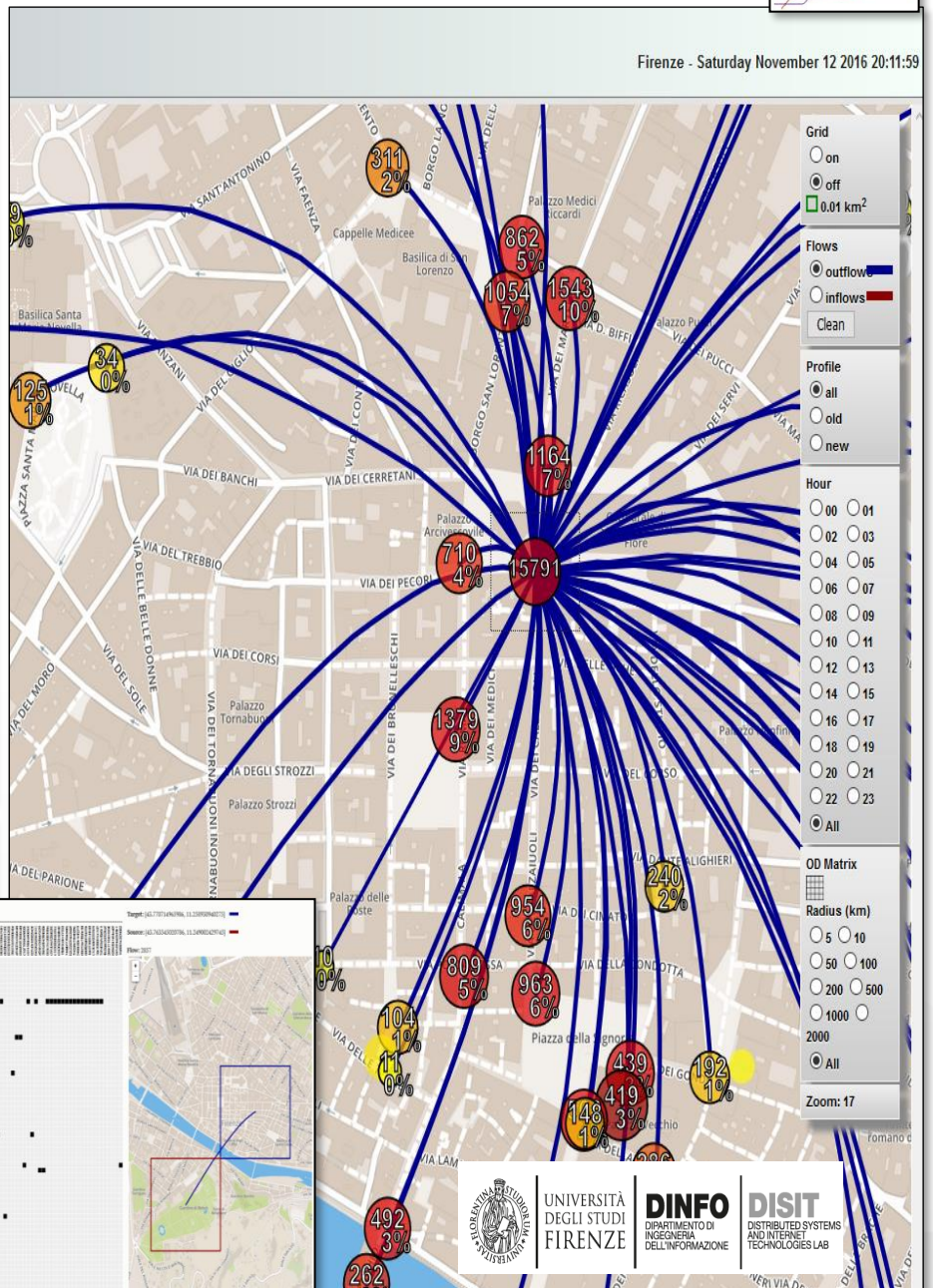
Data and Service Aggregator

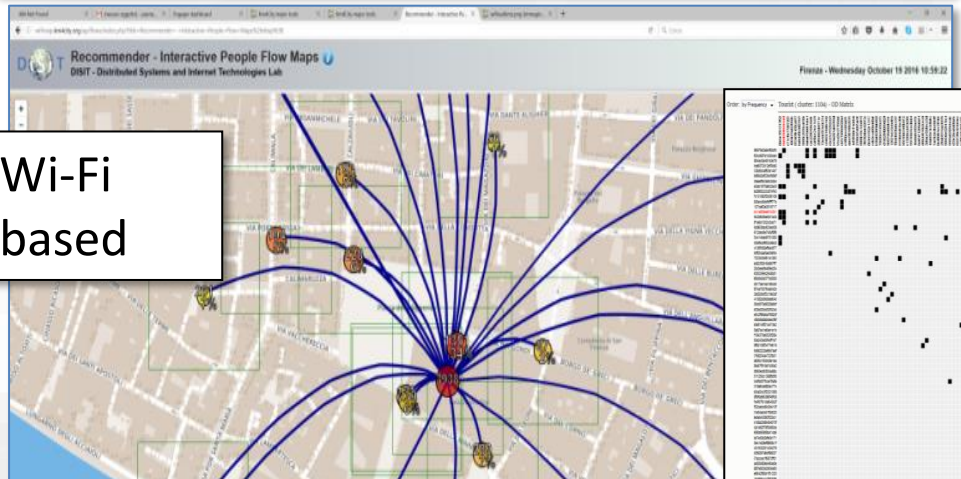
Creating and activating

- Decision Support models

Origin Destination Matrix Estimation







Wi-Fi based

UNIVERSITÀ DEGLI STUDI FIRENZE
DINFO DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE
DISIT DISTRIBUTED SYSTEMS AND INTERNET TECHNOLOGIES LAB

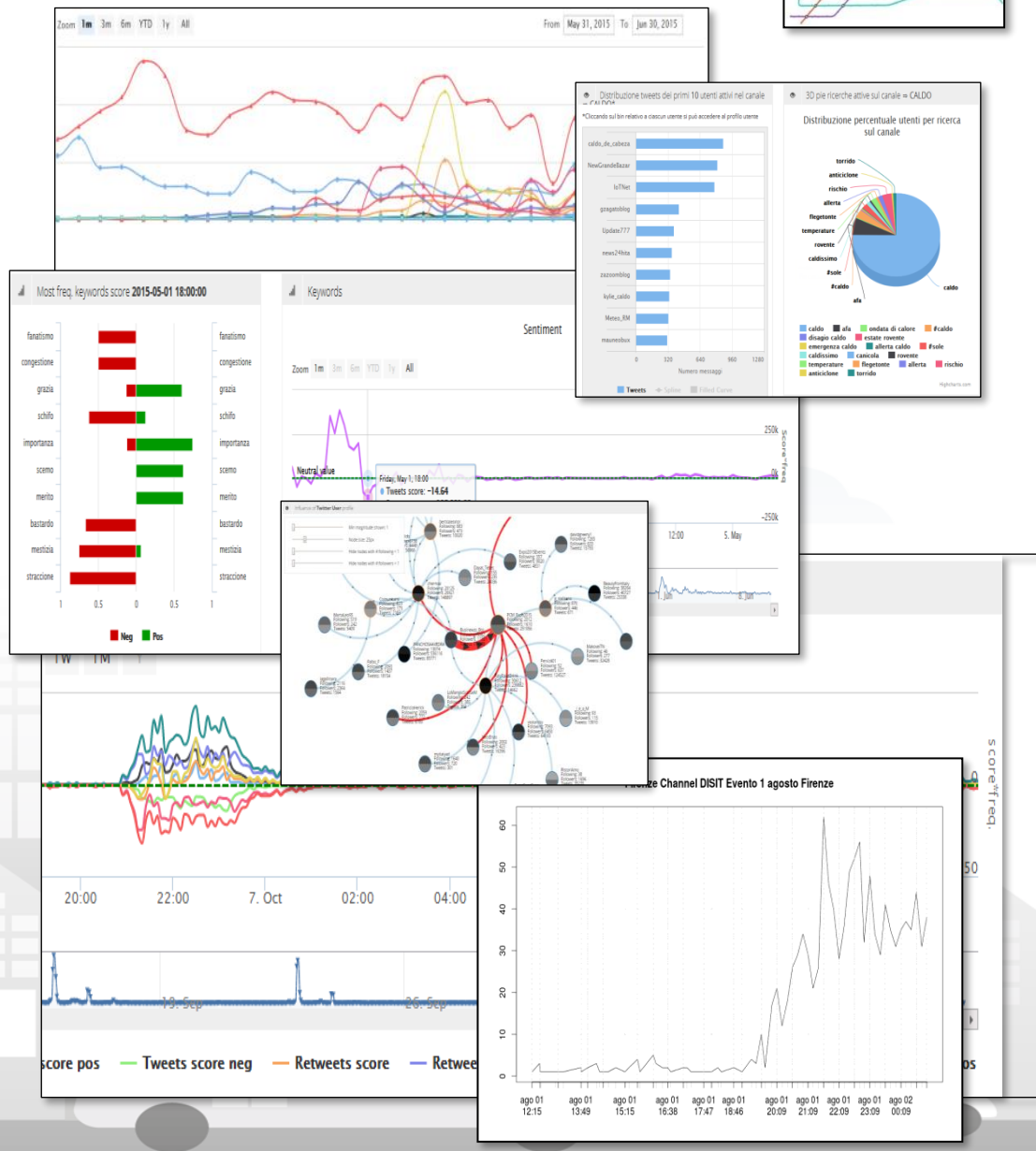
Le nostre informazioni

- **Transazioni**: bancarie, commerciali, ...
  - Ancora movimenti, ancora in rete, ma anche preferenze, ..
  - semplici da comprendere
    - Propensione all'acquisto, capacità di spesa, etc.
    - Analisi dei comportamenti anomali.
- **Comunicazioni**: telefoniche, social media, email, sms, chat, ..
  - movimenti, preferenze, azioni, ..
  - Analisi del testo in linguaggio naturale
  - ..
- Etc.

UNIVERSITÀ DEGLI STUDI FIRENZE | DINFO DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE | DISIT DISTRIBUTED SYSTEMS AND INTERNET TECHNOLOGIES LAB
http://www.disit.org

Twitter Vigilance

- http://www.disit.org/tv
- http://www.disit.org/rttv
- Citizens as sensors to
  - Assess sentiment on services, events, …
  - Response of consumers wrt…
  - Early detection of critical conditions
  - Information channel
  - Opinion leaders
  - Communities
  - Formation
  - Predicting volume of visitors for tuning the services



Twitter Vigilance

UNIVERSITÀ DEGLI STUDI FIRENZE

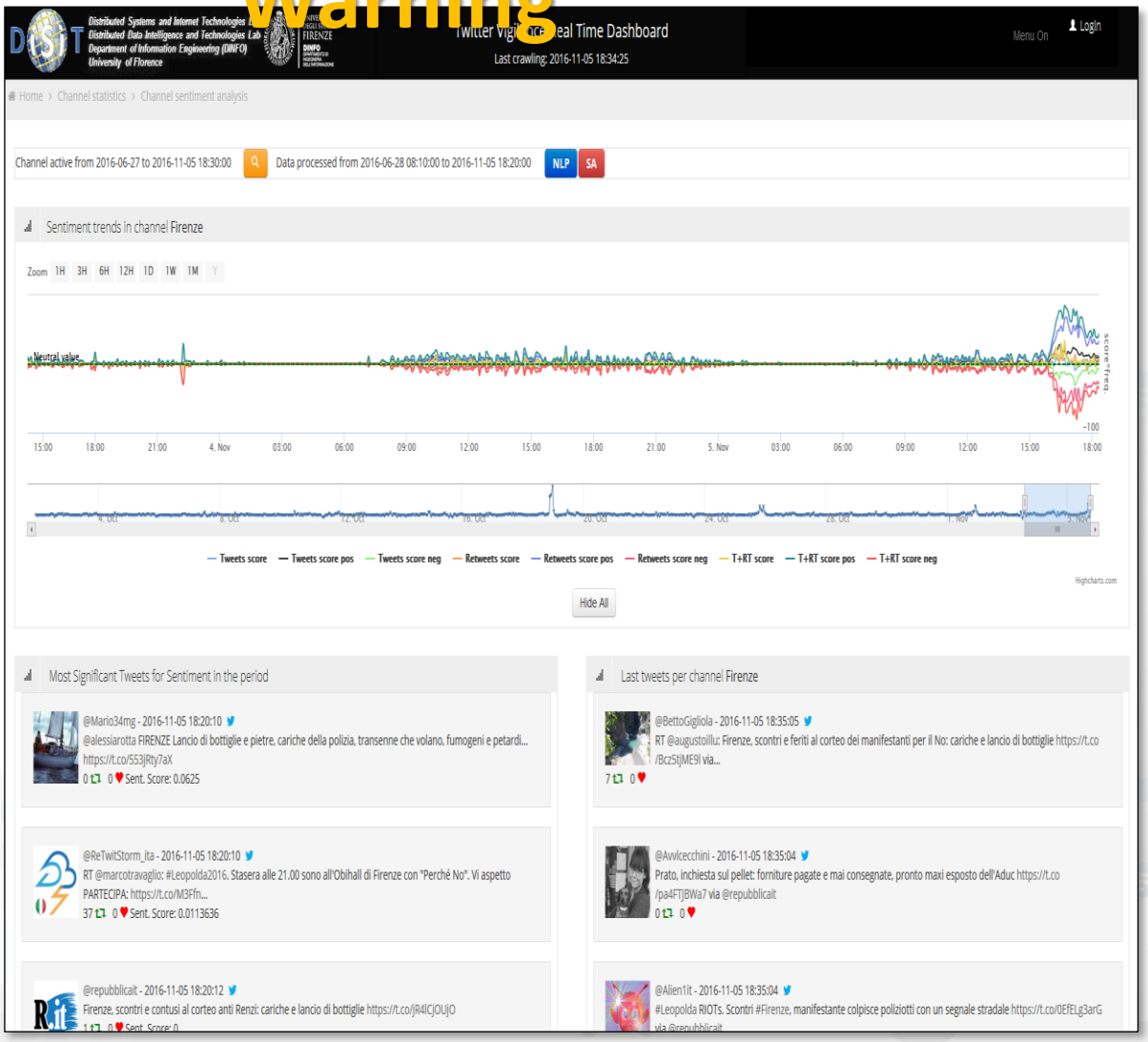
DINFO DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

DISIT DISTRIBUTED SYSTEMS AND INTERNET TECHNOLOGIES LAB

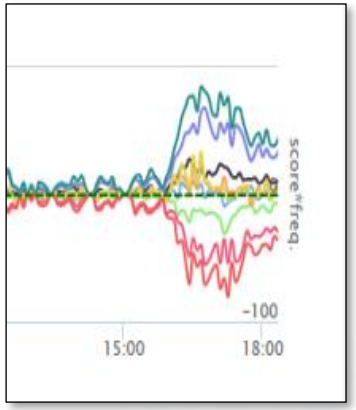
http://www.disit.org

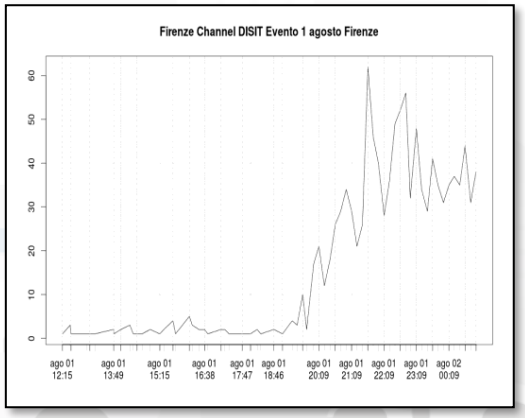
Twitter Vigilance

Real Time Twitter Vigilance, Early Warning



Sentiment Analysis





Gli attacchi ai dati personali

- **Keylogger**: processo che registra cosa digitate
- **Fake AP**: un AP finto che registra il vostro accesso
- **Phishing**: Fake WEB page di banca
- **Sniffing**: su vari media (rame, fibra, etc..)
  - Man in the middle
- **Hijacking**: acquisizione della vostra identita' nel digtale

- **Molti di questi attacchi** arrivano via: email, USB, pagine WEB, Applicazini, giochi, ..

La sicurezza delle info

- **La sicurezza dovrebbe garantirla il provider, fornitore del servizio**
  - Protocolli: HTTPS, LDAP (SSO),
  - Certo non è cosa facile lo dimostrano i furti di password effettuati con successo a carico di grossi providers
- **GDPR (General Data Protection Regulation)**
  - Diritto di sapere quali tipi di dati personali sono memorizzati
  - Diritto di concedere i propri dati o meno per tipo di dato
  - Diritto di vedere cancellati i propri dati (vs security nazionale)
  - …

Le tecnologie di base

- **Soluzioni di rete**: *Firewall*, antivirus, anti *rootkit*, etc.
- Soluzioni corporate per monitoraggio e il *Data Analytic* per identificare
  - pattern/firme di aggressori, se note, …
  - comportamenti anomali:
    - acquisti fatti in luoghi distanti, movimenti bruschi, errori nella password, movimenti strani sui conti, acquisti non usuali, volume acquisiti, ..
    - perdite di potenza sui cavi in rame o fibra
    - …

Data Intelligence

Corso di perfezionamento post laurea
«Intelligence e sicurezza nazionale»

Prof. Paolo Nesi
DISIT Lab
Distributed Data Intelligence and Technologies Lab
Distributed Systems and Internet Technologies Lab
Dipartimento di Ingegneria dell'Informazione
Università degli Studi di Firenze
Via S. Marta 3, 50139, Firenze, Italia
tel: +39-055-2758515, fax: +39-055-2758570
http://www.disit.dinfo.unifi.it
paolo.nesi@unifi.it