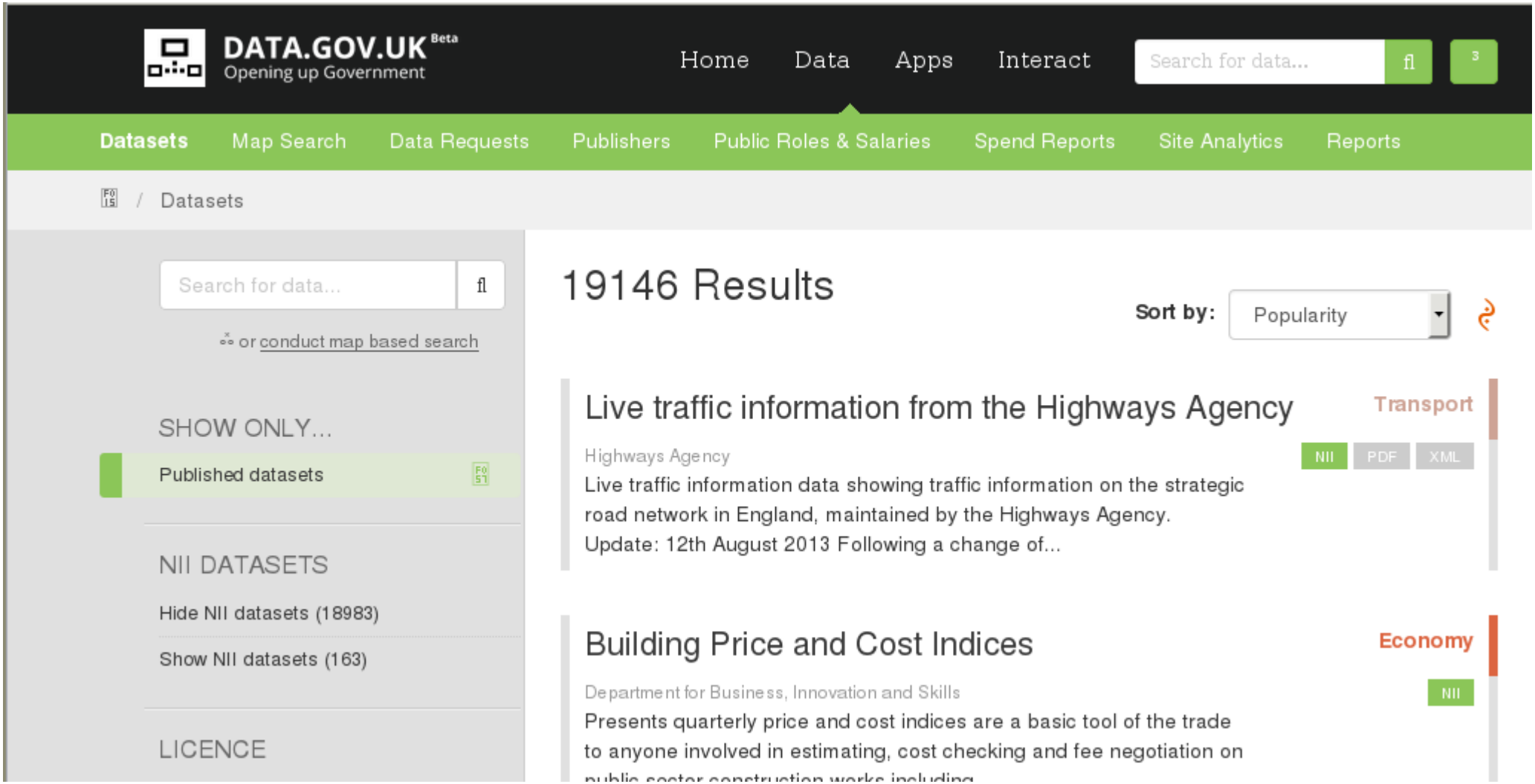
Nexa Center for Internet & Society
Politecnico di Torino

Smart Disclosure
e
government as a platform

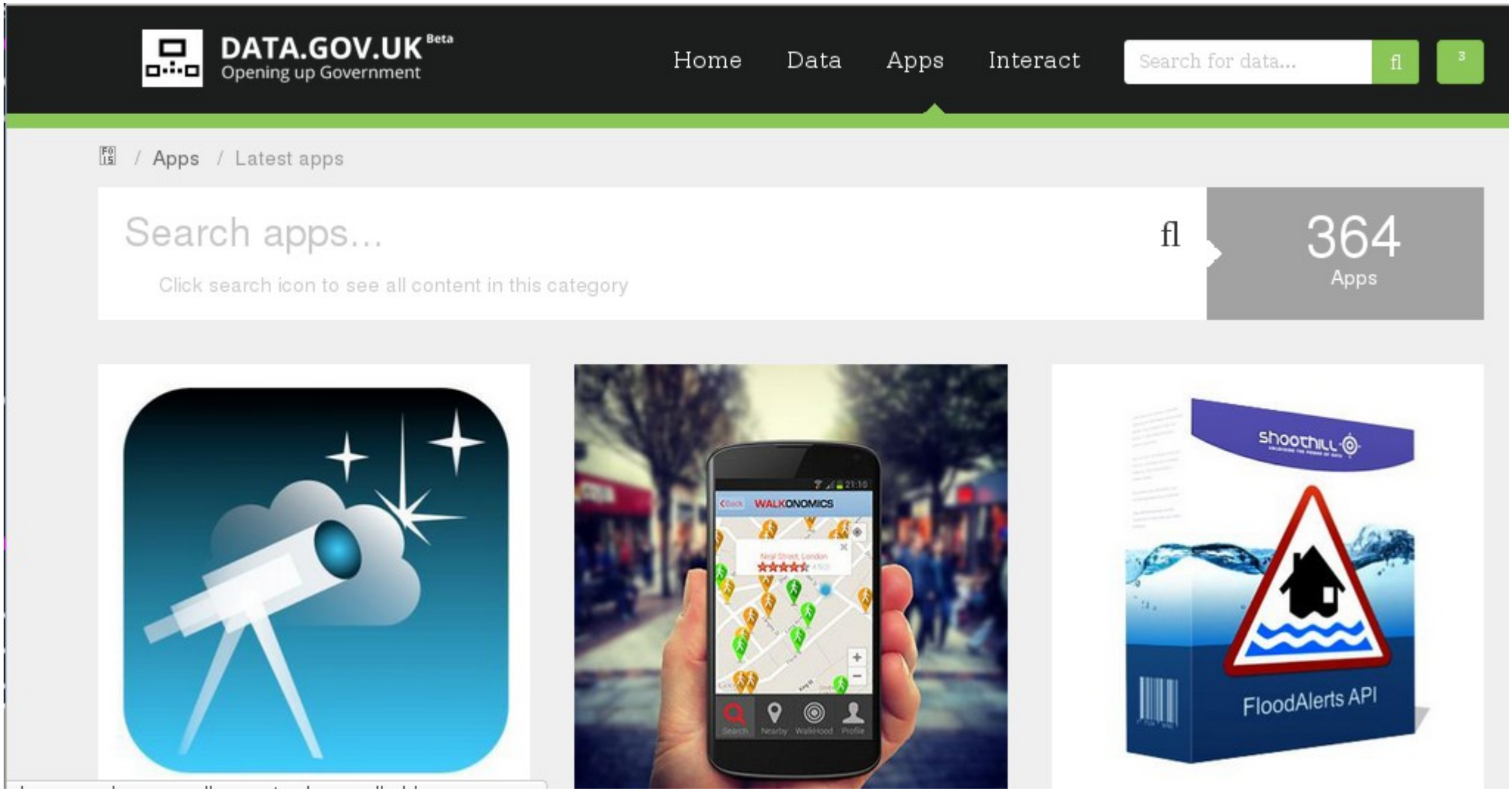
Giuseppe Ciaccio
DIBRIS, Università di Genova

28° Nexa Lunch Seminar, 25 febbraio 2015

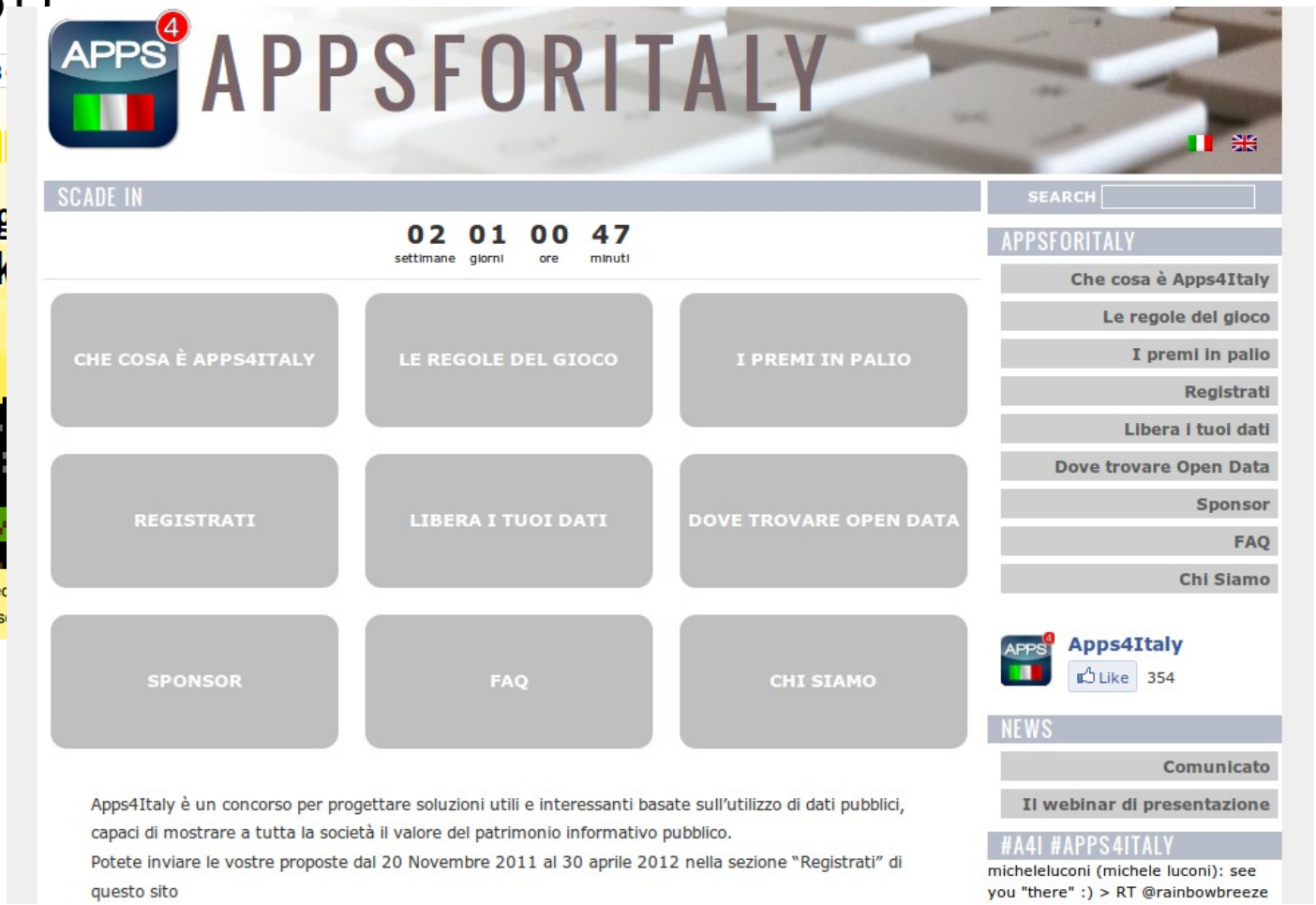
data.gov.uk


DATA.GOV.UK Beta
Opening up Government
Home
Data
Apps
Interact
Search for data...
Datasets
Map Search
Data Requests
Publishers
Public Roles & Salaries
Spend Reports
Site Analytics
Reports
/ Datasets
Search for data...
or conduct map based search
SHOW ONLY...
Published datasets
NII DATASETS
Hide NII datasets (18983)
Show NII datasets (163)
LICENCE
19146 Results
Sort by:
Popularity
Live traffic information from the Highways Agency
Transport
Highways Agency
NII
PDF
XML
Live traffic information data showing traffic information on the strategic road network in England, maintained by the Highways Agency.
Update: 12th August 2013 Following a change of...
Building Price and Cost Indices
Economy
Department for Business, Innovation and Skills
NII
Presents quarterly price and cost indices are a basic tool of the trade to anyone involved in estimating, cost checking and fee negotiation on

data.gov.uk


DATA.GOV.UK Beta
Opening up Government
Home
Data
Apps
Interact
Search for data...
/ Apps / Latest apps
Search apps...
Click search icon to see all content in this category
364
Apps
WALKONOMICS
shoothill
FloodAlerts API

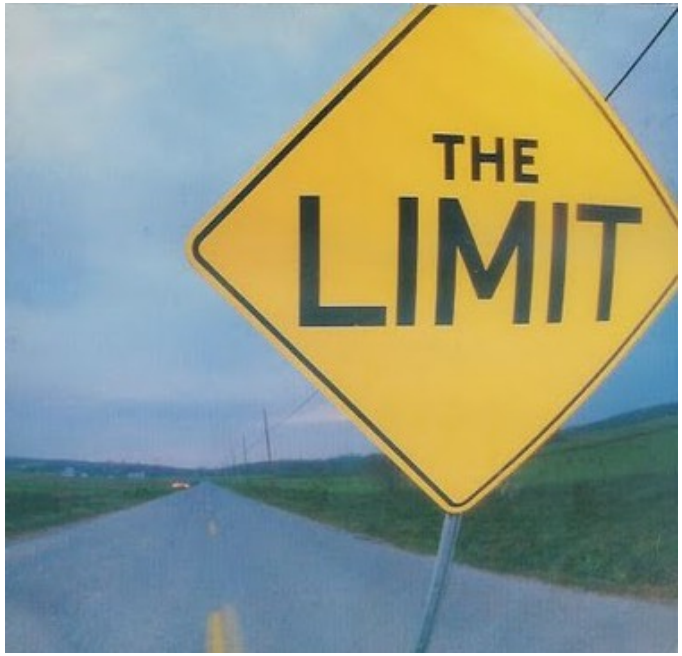
concorsi a premi



il limite

Open Data attuali:

- (di solito) **statici**
- **aggregati**
- **anonimizzati**
- **con semantica ambigua**



Osservazione:

la **assenza di dati individuali** delle persone, la **ambiguità semantica** e la **natura statica** dei dataset **limitano l'uso** degli Open Data

e dunque?

I dati devono essere “freschi”:
abbandonare l'idea dei dataset scaricabili

Sorgenti di dati multiple ma
semantica condivisa

Servizi utili ai singoli individui:
usare informazioni personali online

ma i servizi online ci sono già...

- E’ vero, ma **chi gestisce i nostri dati ha il monopolio dei servizi** e decide cosa offrire e in che modo. Prendere o lasciare.
- Per pratiche complesse dobbiamo usare parecchi servizi indipendenti e non interoperabili. Lungo lavoro “manuale” tramite browser. Possibili errori.
- Per automatizzare queste pratiche **serve un approccio *machine-to-machine*** (m2m), non *machine-to-human*
- **La massima flessibilità si ha offrendo i dati**, sui quali costruire o “lasciar costruire” i servizi più svariati: *government as a platform*

“government as a platform”

Un sogno visionario?

*“[...] the **federal government** must reimagine its role as an **information provider**. Rather than struggling...to design sites that meet each end-user need, it **should focus on creating a simple, reliable and publicly accessible infrastructure that “exposes” the underlying data.** Private actors, either nonprofit or commercial, are better suited to deliver government information to citizens and can constantly create and reshape the tools individuals use to find and leverage public data.[...]”*

Tim O'Reilly, 2010

smart disclosure

- **L'accesso online ai dati** in modalità m2m si può accoppiare a tecnologie di **autorizzazione online**
- **Gli individui possono autorizzare applicazioni di terze parti ad accedere ai propri dati personali** a prescindere da chi gestisce materialmente questi dati (PA, banche, etc.)

smart disclosure

- **L'accesso online ai dati** in modalità m2m si può accoppiare a tecnologie di **autorizzazione online**
- **Gli individui possono autorizzare applicazioni di terze parti ad accedere ai propri dati personali** a prescindere da chi gestisce materialmente questi dati (PA, banche, etc.)
- Dati non più “sequestrati” dai gestori
- Gli individui ridiventano proprietari dei dati personali

smart disclosure

- **L'accesso online ai dati** in modalità m2m si può accoppiare a tecnologie di **autorizzazione online**
- **Gli individui possono autorizzare applicazioni di terze parti ad accedere ai propri dati personali** a prescindere da chi gestisce materialmente questi dati (PA, banche, etc.)
- Dati non più “sequestrati” dai gestori
- Gli individui ridiventano proprietari dei dati personali

Si chiama **Smart Disclosure**

(rilascio intelligente)

esempio: Green Button

- Smart disclosure made in USA:
  http://en.openei.org/wiki/Green_Button
- Campo di applicazione:
  fornitura di energia al dettaglio (*utilities*)
- Dati individuali di consumo (raccolti dalle *utilities*) a disposizione di terzi, previa autorizzazione dell'individuo
- Solito concorso a premi per sviluppatori di app:
  http://appsforenergy.challengepost.com/submissions
- Ancora nessuna *killer app*...?

autorizzazione all'accesso online

- Accesso ai **dati personali**: problemi di **privacy**
- Accesso in **scrittura**: ulteriori problemi di **sicurezza/integrità**

- **OAuth 2.0**
  tecnologia standard per l'autorizzazione online
- Google e Facebook (e molti altri)
  la usano da tempo

un ecosistema dei dati personali

Web API per accesso dati + autorizzazione online =

Smart Disclosure

separazione tra **gestore** e **proprietario** dei dati,

dati personali diventano **open** pur rimanendo **“non pubblici”**

servizi online personalizzati e dunque **più utili,**

libertà di realizzare nuovi servizi,

“government as a platform”

OAuth 2.0

Idea semplice:

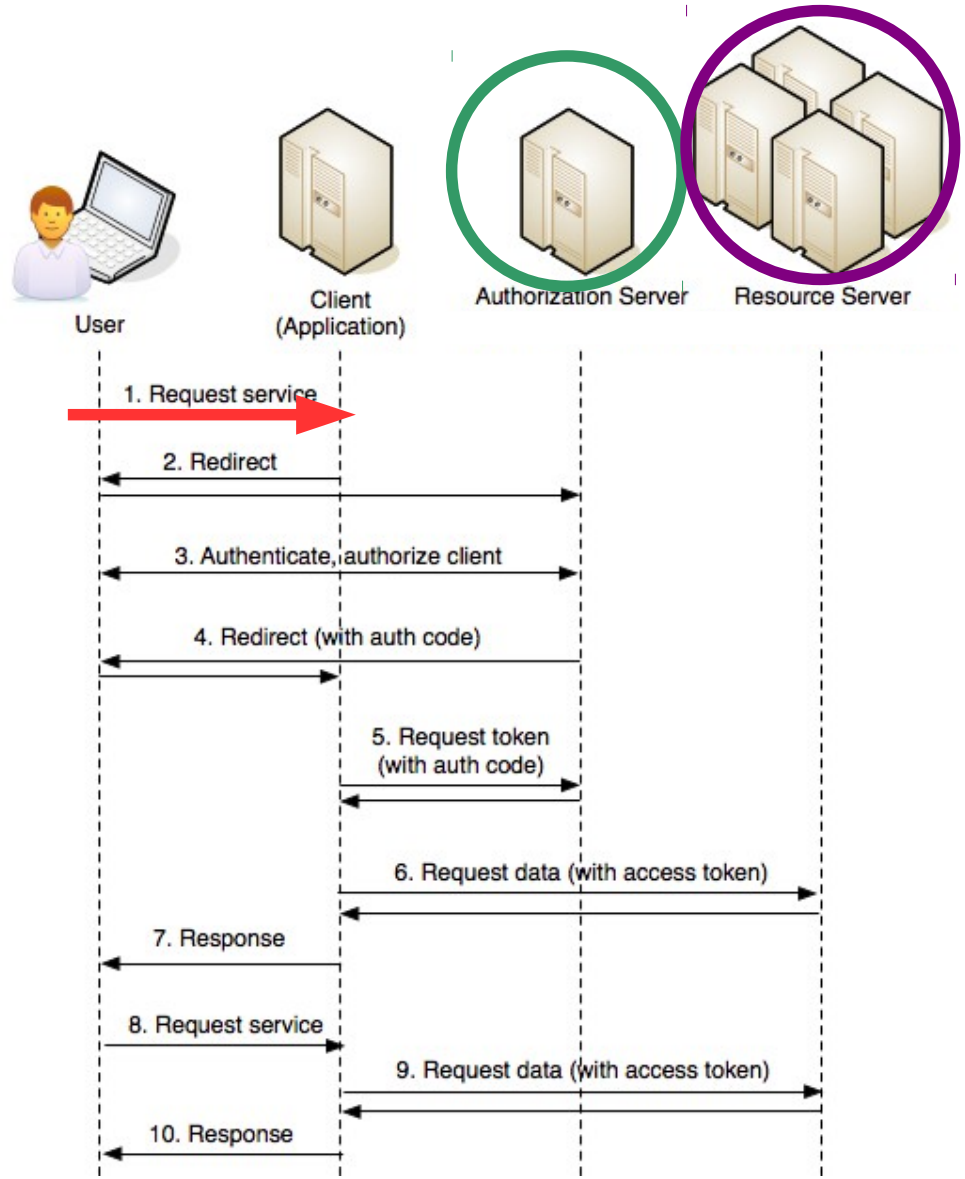
1. **Ottieni un token di accesso** da un **Authorization Server**

2. **Spendi il token di accesso** per ottenere i dati localizzati presso uno o più **Resource Servers** anche indipendenti tra loro (ma che riconoscono autorità a chi ha emesso il token)

Differenti flussi (“protocolli”) a seconda del tipo di app (desktop, web, mobile)

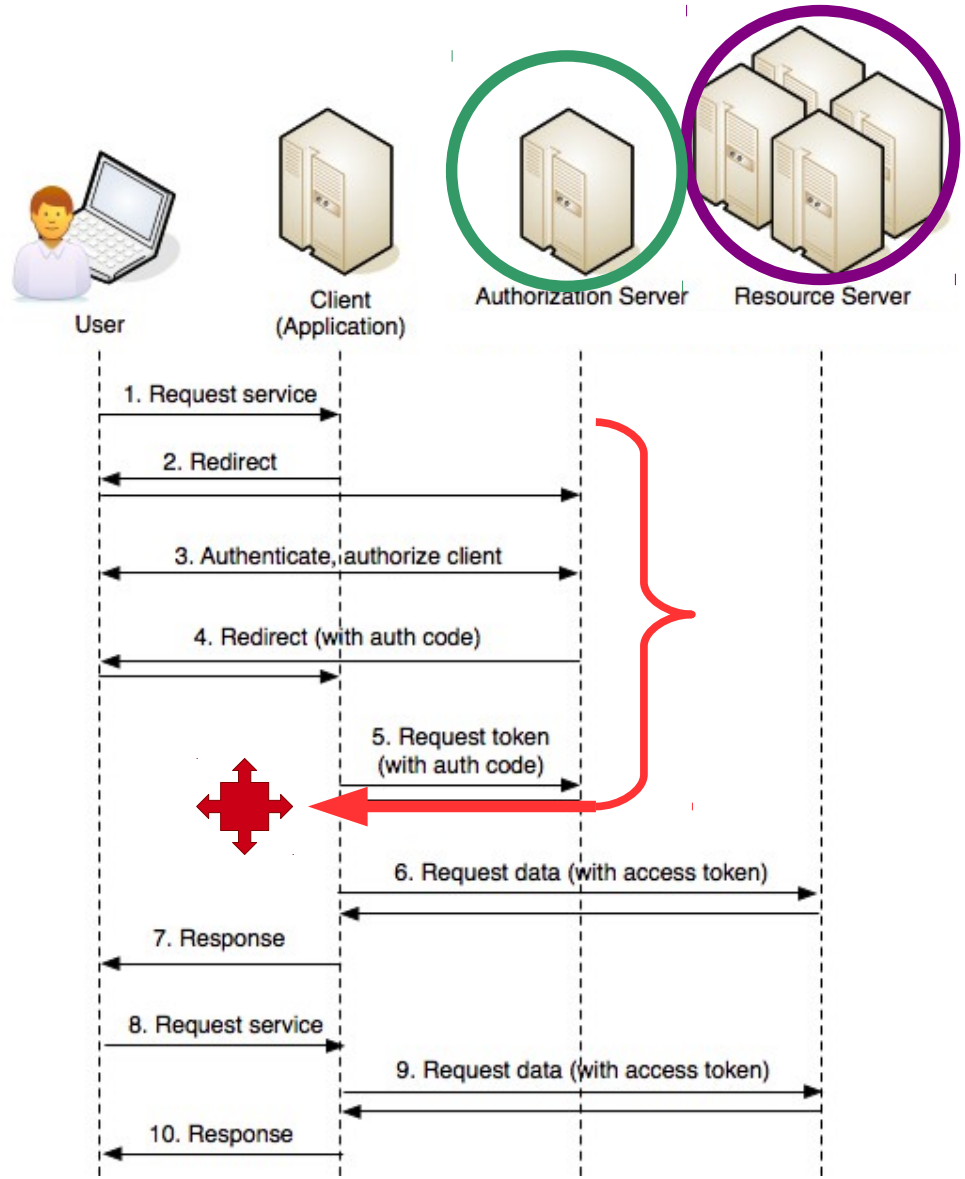
OAuth 2.0

Web app flow



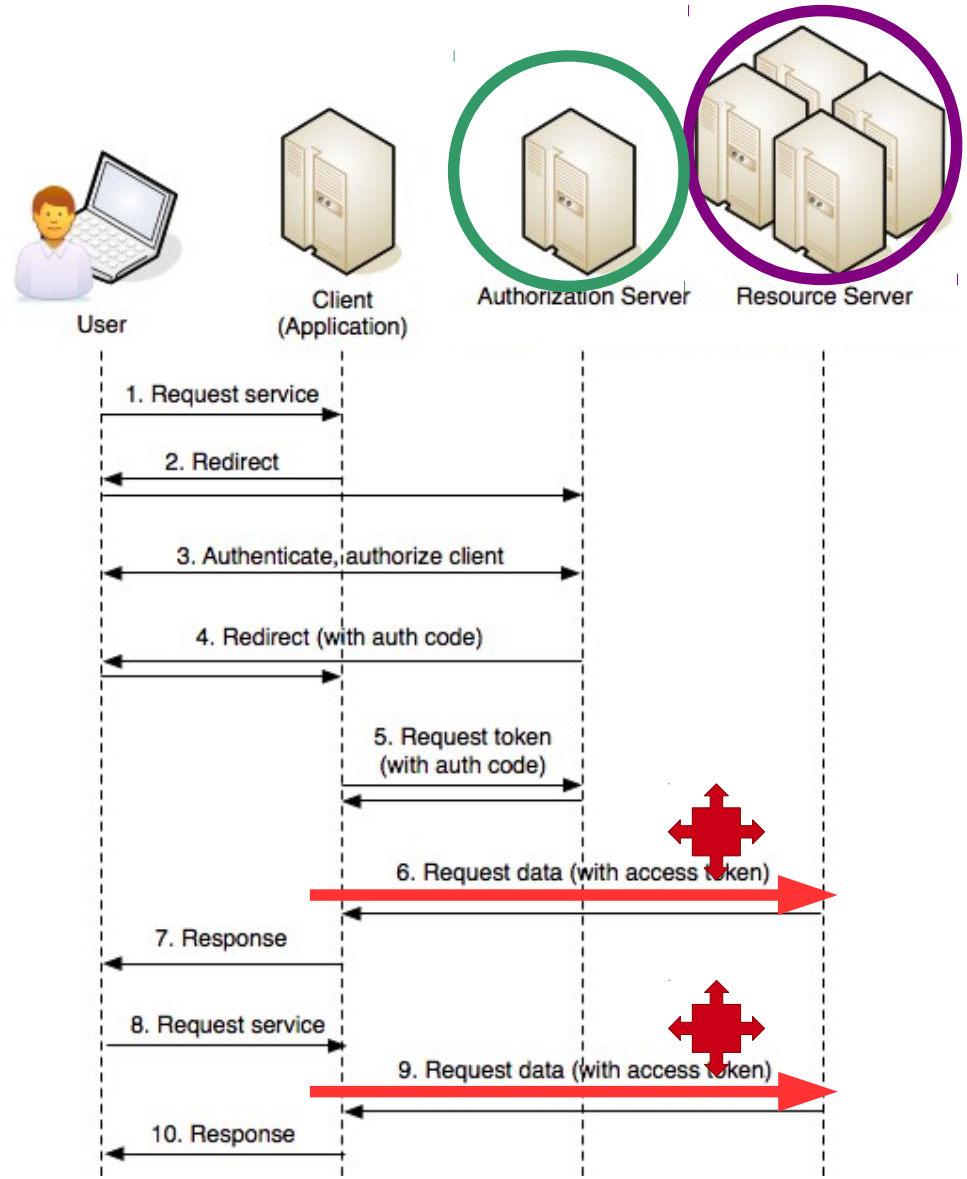
OAuth 2.0

Web app flow



OAuth 2.0

Web app flow



OAuth 2.0 token

1.4. Access Token

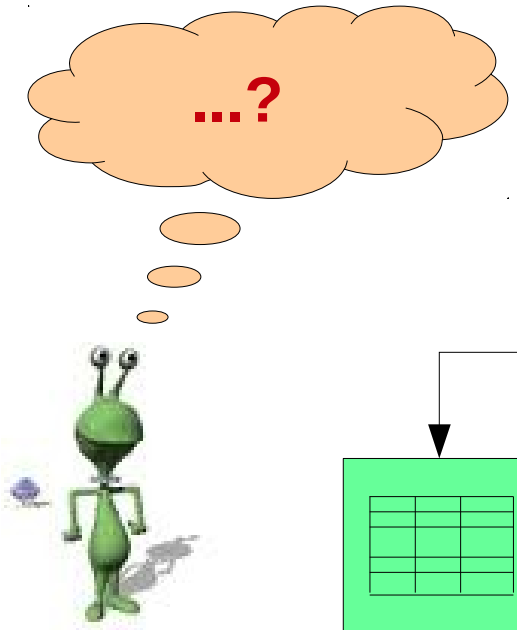
Access tokens are credentials used to access protected resources. An access token is a string representing an authorization issued to the client. The string is usually opaque to the client. Tokens represent specific scopes and durations of access, granted by the resource owner, and enforced by the resource server and authorization server.

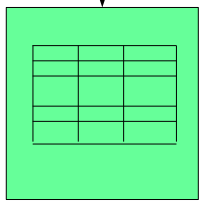
The token may denote an identifier used to retrieve the authorization information, or self-contain the authorization information in a verifiable manner (i.e. a token string consisting of some data and a signature). Additional authentication credentials, which are beyond the scope of this specification, may be required in order for the client to use a token.

The access token provides an abstraction layer, replacing different authorization constructs (e.g. username and password) with a single token understood by the resource server. This abstraction enables

http://tools.ietf.org/html/draft-ietf-oauth-v2-25

dichiarazione redditi – la killer app?





web app
per mod. 730

(public)
Authorization Server

Resource Servers

Anagrafe

Fatturazione elettronica

Agenzia delle Entrate

Servizio Sanitario (Regione)

Mutui (Banche)

fatto @ dibris.unige.it

- ***OAuthwo:* a free OAuth 2.0 authorization server**

  scritto in PHP usando Zend Framework.

  Autore: Antonio Pastorino @ dibris, assegno di ricerca cofinanziato da CCIAA Genova

- ***mod730***: una web app che, con *OAuthwo* e alcuni Resource Servers artificiali, dimostra una dichiarazione mod. 730 automatizzata.

  Tesi di laurea, autore: Artur Tolstenco @ dibris

- https://github.com/giuseppe.ciaccio/oauthwo
  https://github.com/giuseppe.ciaccio/mod730
  https://dione.disi.unige.it

in progress

- **Proprietari multipli**
- **Autorizzazione asincrona**
- **Delega a terzi dei poteri di autorizzazione**
- **Diritti di modifica/creazione/cancellazione**
- **Individui vs. ruoli**
- **Semantica delle risorse (dati)**

proprietari multipli

- Es: fattura rilasciata per una prestazione
- Nel DB di un Resource Server, ogni record dovrebbe indicare chi sono i proprietari del dato rappresentato
- Notificare tutti i comproprietari se uno di essi autorizza accesso
- Problema: nei DB legacy non sono indicati i proprietari dei dati (e non c’è un modo standard per desumerli)

autorizzazione asincrona

- Il proprietario dovrebbe poter **autorizzare in anticipo**, in modo che la parte autorizzata poi proceda all’accesso senza convocare il proprietario
- Es.: dich. redditi dal commercialista
- Sembra facile da risolvere (proprietario autorizza ma si tiene il token di autorizzazione, e lo trasmette poi alla parte autorizzata che lo usa quando vuole)

delega a terzi dei poteri di autorizzazione

- Es.: familiare fiscalmente a carico oppure persona interdetta o incapace di intendere
- “X delega Y ad autorizzare accessi ai dati di Z” + periodo temporale + eventuali limitazioni + altro?
- X può coincidere con Z ma anche no (es.: X = magistrato, Z = persona interdetta o incapace)
- Serve un livello ulteriore, con un token aggiuntivo
- Problema: trovare soluzione che resista a legislatori irrazionali.

diritti di modifica/creazione/cancellazione

- Es.: tutti i dati mantenuti dalle PA; il cittadino e' proprietario ma non crea, non modifica, non cancella
- Associare a ciascun proprietario un **insieme di diritti** che esso esercita sul dato
- CRUD + diritto di autorizzare + diritto di delegare (+ altro?)
- Idealmente tale insieme varia da un dato all'altro, anche per uno stesso proprietario...

individui vs. ruoli

- Es.: magistrato, funzionario di polizia, medico di P.S.
- Se accedono ai miei dati, lo fanno per il loro ruolo
- Impossibile elencare tra i proprietari di ogni dato tutti gli individui che hanno ruolo di magistrato...
- Servono “credenziali di ruolo” individuali.
- Asserzioni SAML, *attribute provider.*

semantica delle risorse

- Enti diversi possono gestire dati analoghi ma con nomi/struttura differente.
- Es.: dati sanitari, anagrafi varie...
- Serve uno strato semantico uniforme, se vogliamo scrivere app che funzionino ovunque.
- Resource Server registrano certi loro insiemi di dati presso l’Authorization Server inviando metadati annotati in base a una ontologia condivisa.
- Oppure: resource server si dotano di endpoint SPARQL e mappano il modello interno dei dati sulla ontologia condivisa (più flessibile ma laborioso).