Open Data & Linked Open Data

Lecture 2 — From Open Data to the Web of Data

Stephane Derrode · Departement MI · Centrale Lyon · 2025-2026

Today's Agenda

Part 1 - Open Data

- Definition & philosophy
- Legal framework
- What qualifies as Open Data?
- Data sources & licences
- Formats & tools
- Use cases

Part 2 - Linked Open Data

- From Web of documents to Web of data
- Resources & representations
- From XML to RDF
- The 4 principles of Linked Data
- LOD star scheme & DBpedia

Both parts feed directly into Lab 1: LOD & SPARQL

Part 1

Open Data

Shown in class

Open Data a la loupe

Source: LiberTIC / Nantes Metropole - A2B Production

Duration: ~4 min

Watch for:

- What types of data are being opened?
- Who benefits, and how?
- What does libre reutilisation mean in practice?

Liberer les donnees publiques pour que chacun puisse les utiliser, les reutiliser et les redistribuer librement.

[Play]

▶️ Search on YouTube: 'Open Data à la loupe LiberTIC'

What is Open Data?

Open data is digital data whose access and use are left free to users. Distributed under an open licence guaranteeing free access and reuse by all, without technical, legal or financial restrictions.

- Open Knowledge Foundation

A movement - a philosophy of access to information

A practice - structured publication of freely usable data

A right - public information as a common good

The Open Knowledge Foundation

- Founded: Cambridge, May 2004
- UK non-profit promoting free culture, open content and open data
- Key initiative: global open data index

Open Research - open scientific data, new collaboration models

Open Culture - open-source tools for cultural heritage

Open Government - public open data, transparency & accountability

🔗 okfn.org

🔗 opendatahandbook.org

Open Government - Why Open Data?

Better territorial knowledge
Geography, public spaces, infrastructure, demographics

Greater public transparency
Budgets, expenditure, subsidies, public contracts

Economic development
Local services, quality improvement, sustainable development

Inter-service efficiency
Better data sharing between public administrations

Source: democratieouverte.org

democratieouverte.org

Open Data - The Legal Framework (France & EU)

1978 French CADA law - right to access administrative documents

2003 EU Directive - access becomes an obligation to publish

2011 French Decree - free reuse of public data established

2016 Loi Republique Numerique - open data by default for public bodies

2016 Energy Transition Law - energy consumption data opened

2018 GDPR - personal data explicitly excluded from open data scope

What Qualifies as Open Data?

Included:
All structured data produced by public or private actors in a public service mission - including geographic data, public services, political exercise, produced by institutions, associations, citizens or academic actors.

Excluded:

- Personal data (GDPR) - any info identifying an individual
- Government deliberation secrecy
- State or public safety secrets
- Third-party intellectual property rights

Key rule: Open Data is not personal data. Anonymisation is required before publication.

Open Data vs Big Data - Key Differences

	Open Data	Big Data
Origin	Public bodies, public service missions	Private actors, individual behaviour
Nature	Structured, public interest, open	Personal, behavioural, private
Examples	Bus schedules, budgets, elections, maps	Location traces, browsing, purchases, IoT
Access	Free, unrestricted	Controlled, monetised
Goal	Transparency, common good	Personalisation, prediction, profit

Where Does Open Data Come From? (1/2)

Collectivites

- DataLab, InfoLab, OpenDataSoft
- 60,000+ datasets on data.gouv.fr
- 2,800+ reuse applications

Etat

- data.gouv.fr
- EtatLab
- OpenData France
- Ministries, national agencies

Citoyens

- Regards Citoyens
- DataPublica
- Data4Citizen
- Civic tech & engaged citizens

Source: data.gouv.fr - figures updated 2024

🔗 data.gouv.fr

Where Does Open Data Come From? (2/2)

Institutes

- INSEE (statistics)
- INPI (industrial property)
- Legifrance (law)
- Air & water quality agencies

International

- Global Open Data Index (OKF)
- World Bank Open Data
- European Data Portal
- UN Data
- NASA Open Data

Geodata

- IGN (French national mapping)
- OpenStreetMap
- Copernicus (EU satellites)
- USGS (US geology)

Open Data Licences

ODbL - Open Database Licence

Open Knowledge Foundation

- Free reuse incl. commercial
- Attribution required
- Integrity obligation
- Share-alike: derivatives must use same licence

More restrictive

Licence Ouverte (Etalab)

French government - Etalab

- Free reuse incl. commercial
- Attribution required
- Integrity obligation
- No share-alike requirement

More permissive

ODbL more common for geographic data - Licence Ouverte for administrative data

Open Data File Formats

To be Open Data, data must be:

- In electronic format
- Accessible online
- Structured
- In an open (preferably non-proprietary) format

Common formats by type:

Alphanumeric:

CSV, ODS, XML, JSON (open) | XLS, DOC (proprietary)

Geographic:

GeoJSON, KML, Shapefile (open) | DWG, DXF (proprietary)

Domain-specific:

GTFS (transport), GPX (GPS tracks)

A PDF file is NOT Open Data - Adobe-dependent, not truly open

Tools for Exploiting Open Data

Alphanumeric data:

Office suites: LibreOffice, Google Sheets, Excel

Programming: Python (pandas, requests), R

OpenRefine - data exploration & cleaning

Geographic data:

QGIS - full GIS suite, open-source (used in Lab 3)

uMap / GeoServer - web map creation

Commercial: ArcGIS, MapInfo, Google Maps Platform

Web APIs: data.gouv.fr API - Overpass API (OSM) - OpenDataSoft - Socrata

Use Case: Open Data in Urban Mobility

What open data enables:

- Real-time bus & tram positions
- Multi-modal journey planners
- Cycling route optimisation
- Parking availability (real-time)
- Air quality mapping
- Examples: Citymapper - OpenWeatherMap - TCL Lyon

Velo'v Lyon - bike share data open since 2013

GTFS feeds - any developer can build a transit app

Handimap - accessibility mapping from open geodata

Shown in class

How we found the worst place to park in New York City

Source: Ben Wellington - TEDxNewYork

Duration: ~12 min

Watch for:

- How a simple question led to a surprising discovery
- What open data made this analysis possible
- How the city reacted to the findings

Municipal agencies have access to a wealth of data... but sometimes they don't know what to do with it.

[Play]

▶ Watch on TED: Ben Wellington — NY Parking

Videos to Watch - Open Data

The following videos are in French - recommended for viewing outside of class

Title	Author	Duration
L'Open Data a la Loupe (extended)	Yann Bresson	~15 min
Quel est l'impact de l'Open Data ?	Charles Ruelle, 2014	~12 min
Open Data explained in a nutshell	Simpleshow Foundation	~3 min
L'Open Data, Avenir des Big Data	Jean-Marc Lazar, TEDxUTCompiegne	~15 min

Yann Bresson

Charles Ruelle

Simpleshow Open Data

Jean-Marc Lazar TEDx

Quick Check - Open Data

Which of the following files qualifies as Open Data?

A - A PDF scan of a city budget report

B - A CSV file of bus schedules on data.gouv.fr (correct answer)

C - An Excel file of citizen names and addresses

D - A proprietary database of commercial transactions

Answer: B - Open, structured, machine-readable, freely reusable, no personal data

Part 2

Linked Open Data

Shown in class

The Next Web of Open, Linked Data

Source: Tim Berners-Lee - TED2009

Duration: ~16 min

Watch for:

- Why is simply publishing data not enough?
- What does 'linked' add to 'open data'?
- The 'Raw Data Now!' argument

Linked data - I want you to make it. I want you to demand it.

[Play]

Watch on TED: Tim Berners-Lee — The Next Web

From the Web of Documents to the Web of Data

Today's Web (of documents)

- Decentralised - HTTP
- Interconnected - URLs
- Interoperable - HTML, CSS, JavaScript
- Designed for humans to read
- Links connect pages

->

The Web of Data

- Decentralised - HTTP
- Interconnected - IRIs
- Interoperable - RDF, SPARQL
- Designed for machines to read AND reason
- Links connect data entities

Key question: how do we make data on the Web as interconnected as pages on the Web?

Resources & Representations

What is a Web resource?

Any entity identifiable by an IRI: a weather bulletin, a coffee order, a person, a concept...

http://meteo.example.com/lyon -> Lyon weather

http://meteo.example.com/ici -> location-aware weather

http://commerce.example.com/order/192837 -> a specific order

Key insight:
A resource is not a file. It is an active object:

- Content can change over time
- Content varies by context
- You can interact with it (read, update, delete)

A resource is always accessed through representations (JSON, XML, HTML, RDF) - never directly

From HTML to XML to RDF

HTML (1991)

- Structure for display
- Tree model
- Weak semantics
- Good for humans

->

XML (1998)

- Structured data exchange
- Tree model (XML Infoset)
- Extensible semantics
- Better for machines

->

RDF (1999, rev. 2014)

- Web data model
- Graph model
- Universal semantics
- Designed for machine reasoning

W3C recommendation: RDF 1.1 (2014) - foundation of the Web of Linked Data

The LOD Star Scheme - 5 Levels of Openness

1 star - Data available online in any format (e.g. PDF, image)

2 stars - Machine-readable structured data (e.g. Excel spreadsheet)

3 stars - Non-proprietary open format (e.g. CSV instead of Excel)

4 stars - W3C open standards used: RDF + SPARQL URIs

5 stars - All above + linked to other open data on the Web (LOD)

Defined by Tim Berners-Lee - Most French public data is at 3 stars - DBpedia is at 5 stars

The 4 Principles of Linked Data

- Tim Berners-Lee, 2006

1 Use IRIs to name things
Every entity (person, place, concept) gets a unique web identifier

2 Use HTTP IRIs
So that anyone can look up the IRI and get a representation back

3 Use standards when serving data
Provide information using RDF and SPARQL

4 Include links to other IRIs
So that users can discover more related data on the Web

🔗 w3.org/DesignIssues/LinkedData.html — Tim Berners-Lee

Source: http://www.w3.org/DesignIssues/LinkedData.html

RDF - From Tables to Graphs

Classic table:

Name | City | Born
Ada | London | 1815

->

As a graph:

Ada -> born in -> London
Ada -> born in year -> 1815
(subject -> predicate -> object)

->

With IRIs:

dbpedia:Ada_Lovelace
-> dbo:birthPlace
-> dbpedia:London

RDF triple: subject (IRI) - predicate (IRI) - object (IRI or literal value)

DBpedia - A Flagship LOD Project

- Launched by Chris Bizer, 2007
- Goal: extract Wikipedia infoboxes and expose as RDF
- 2024: 6M+ things described, 125M+ RDF triples
- Includes: persons, places, organisations, creative works, species...
- Linked to: Wikidata, GeoNames, MusicBrainz, 1000+ LOD datasets

Why it matters:

DBpedia turns the human-readable Wikipedia into a machine-readable knowledge graph - queryable via SPARQL.

Try: dbpedia.org/resource/Ada_Lovelace

🔗 **dbpedia.org**

🔗 dbpedia.org/resource/Ada_Lovelace (example)

The LOD Cloud

- Maps all published LOD datasets and their interconnections
- First version: 2007 - 12 datasets
- Today: 1,300+ datasets, billions of triples
- Domains: government, life sciences, geography, media, publications
- -> lod-cloud.net (live, interactive)

Key hubs:

- DBpedia - encyclopedic knowledge
- Wikidata - structured facts
- GeoNames - 11M geographic features
- LinkedGeoData - OpenStreetMap as RDF
- EuroStat - EU statistical data

Principle 4 in action: data linking creates an emergent, self-growing knowledge web

🔗 lod-cloud.net

Quick Check - Linked Open Data

According to Tim Berners-Lee's 4 principles, what distinguishes Linked Data from simply published Open Data?

A - It must be in PDF format

B - It must be free of charge

C - Entities use HTTP IRIs and link to other datasets (correct answer)

D - It must be published by a government body

Answer: C - Principles 2 & 4: HTTP IRIs for lookup + links to other data for discovery

Key Takeaways & What's Next

Summary:

- Open Data = free, structured, reusable public data
- LOD = Open Data + universal identifiers + links
- The LOD cloud: 1,300+ interconnected datasets
- RDF triples: the foundation of the Web of Data

Prepare for Lab 1 (LOD & SPARQL):

- Review the 4 principles of Linked Data
- Explore dbpedia.org before the session

Resources:

- opendatahandbook.org
- data.gouv.fr
- lod-cloud.net
- dbpedia.org

🔗 opendatahandbook.org

🔗 data.gouv.fr

🔗 lod-cloud.net

🔗 dbpedia.org

Next lecture:

Cloud Computing - Actors & Strategies

Yann Fornier - Oct. 20

🔗 Course page

🔗 Pedagogie3 platform

Questions? -> stephane.derrode@ec-lyon.fr