SUM OF ALL VIDEO GAMES

Our road to make Wikidata the hub of all video game metadata.

Jean-Frédéric Berthelot, @JeanFred
Tracy Hoffmann, @nabatz
Envel Le Hir, @envlh
WikidataCon 2019, 25-26 October, Berlin

Why should we care about video games?

- **Exist since the 40s/50s/60s/70s**
  (it's complicated, check Wikipedia :-þ)
- **Major cultural industry, more than 2.5 billion players worldwide**
- **At least 100K to 200K games published**

WIKI DATA CON
Wikidata & languages 2019

Joi Ito, CC BY 2.0

Current state of video game data (in Wikidata)

- 38,539 video games (Q7889)
- 423 expansion packs (Q209163)
- 294 video game compilations (Q16070115)
- 78 downloadable contents (Q1066707)
- 112 game controllers (Q865422)
- 766 video game consoles (Q8076)
- 206 video game genres (Q659563)
- 287 video game magazines (Q69662460)

WIKI DATA CON
Wikidata & languages 2019

- A *lot* of databases
  - Mobygames, GameFAQs, IGDB, ...
  - Special-interest fan-databases, government agencies, commercial players...
- No common identifier
- Little or no cross-linking between databases
- Different data coverage / specialization
- Conceptual differences

And out of Wikidata

WikiProject Video Games

http://w.wiki/6an

(and subpages)

WikiProject Video Games

Something old...

Properties / Ontology

- Currently 64 properties
  - Mostly for games

Example Queries

Contents [hide]

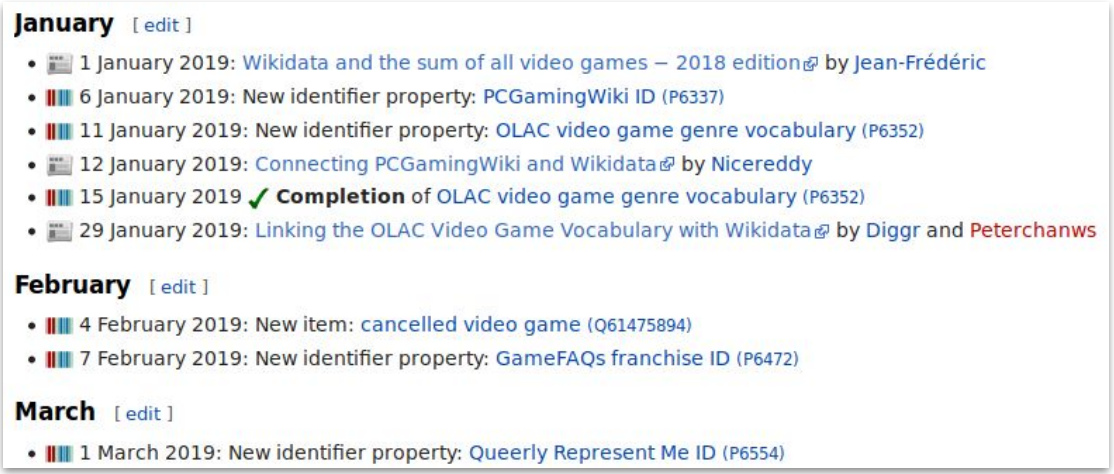
1 Properties for games
- 1.1 Core properties
- 1.2 Main properties
- 1.3 Properties for relationships to other works
- 1.4 Gameplay properties
- 1.5 Staff properties
- 1.6 Characters and setting
- 1.7 Reception properties
  - 1.7.1 Content Rating
- 1.8 Secondary properties

2 Properties for franchise/series
3 Properties for mods
4 Properties for platforms
5 External identifiers

WikiProject Video Games

...something new...

- Activity Log
- Tasks list



January [edit]

- 1 January 2019: Wikidata and the sum of all video games – 2018 edition by Jean-Frédéric
- 6 January 2019: New identifier property: PCGamingWiki ID (P6337)
- 11 January 2019: New identifier property: OLAC video game genre vocabulary (P6352)
- 12 January 2019: Connecting PCGamingWiki and Wikidata by Nicereddy
- 15 January 2019 ✓ **Completion** of OLAC video game genre vocabulary (P6352)
- 29 January 2019: Linking the OLAC Video Game Vocabulary with Wikidata by Diggr and Peterchanws

February [edit]

- 4 February 2019: New item: cancelled video game (Q61475894)
- 7 February 2019: New identifier property: GameFAQs franchise ID (P6472)

March [edit]

- 1 March 2019: New identifier property: Queerly Represent Me ID (P6554)

WikiProject Video Games

...something borrowed...

- Statistics , Reports
- External Identifiers dashboard

Powered by Listeria

Property	Subject	Constraints	Wikidata completeness	Mix'n'Match	?usage	?records	Completion (%)
Behind The Voice Actors video game ID (P4965)	Behind The Voice Actors	Constraints	is complete, but new values may exist in the future	2368	3641	3488	104
OLAC video game genre vocabulary (P6352)	OLAC video game vocabulary	Constraints	is complete	2167	68	67	101
GameFAQs platform ID (P6078)	GameFAQs	Constraints	is complete, but new values may exist in the future	1964	129	129	100
IGDB platform ID (P5795)	Internet Game Database	Constraints	eventually complete	1736	137	141	97
old-computers.com ID (P5936)	old-computers.com	Constraints	eventually complete	1807	1176	1246	94
MobyGames platform ID (P5868)	MobyGames	Constraints	eventually complete	1774	224	271	83
NeoGeoSoft ID (P6898)	NeoGeoSoft	Constraints	eventually complete	2655	114	156	73
Sega8bit.com identifier (P4961)	Sega8bit.com	Constraints	eventually complete	1081	235	353	67

WikiProject Video Games

Powered by inteGraality

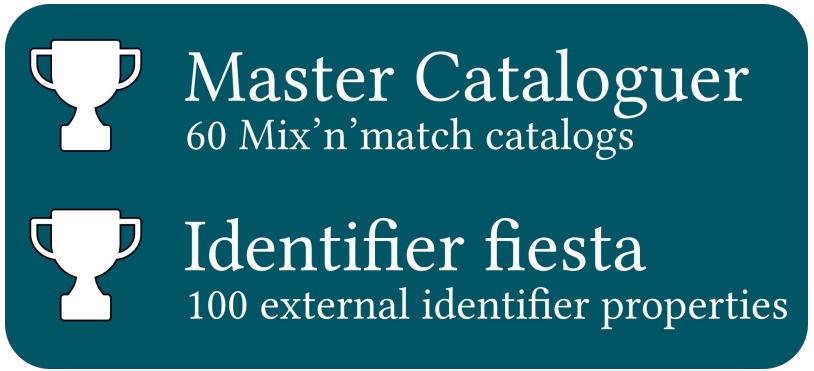
Top groupings (Minimum 20 items)		Top Properties (used at least 10 times per grouping)						
Name	Count	genre (P136)	developer (P178)	publisher (P123)	country of origin (P495)	publication date (P577)	game mode (P404)	distribution format (P437)
Microsoft Windows (Q1406)	12686	73.88% (9372)	58.31% (7397)	55.14% (6995)	43.73% (5548)	79.77% (10119)	67.28% (8535)	55.6% (7054)
PlayStation 2 (Q10680)	2909	76.07% (2213)	69.71% (2028)	70.13% (2040)	64.52% (1877)	81.27% (2364)	66.76% (1942)	41.08% (1195)
macOS (Q14116)	2565	81.68% (2095)	70.8% (1816)	61.4% (1575)	44.02% (1129)	89.75% (2302)	82.5% (2116)	79.03% (2027)
iOS (Q48493)	2459	73.65% (1811)	58.19% (1431)	49.61% (1220)	50.79% (1249)	84.34% (2074)	61.53% (1513)	40.71% (1001)
Commodore 64 (Q99775)	2097	65.43% (1372)	41.34% (867)	50.64% (1062)	50.69% (1063)	89.65% (1880)	49.69% (1042)	22.84% (479)
Disk Operating System (Q170434)	1979	88.48% (1751)	70.04% (1386)	72.46% (1434)	76.96% (1523)	94.54% (1871)	80.75% (1598)	47.75% (945)
PlayStation 3 (Q10683)	1934	87.59% (1694)	84.75% (1639)	82.26% (1591)	65.41% (1265)	90.95% (1759)	81.08% (1568)	69.03% (1335)
Xbox 360 (Q48263)	1742	90.3% (1573)	89.38% (1557)	88.92% (1549)	65.67% (1144)	95.41% (1662)	88.35% (1539)	68.54% (1194)
Nintendo DS (Q170323)	1638	72.41% (1186)	64.41% (1055)	68.13% (1116)	57.08% (935)	79.18% (1297)	63.74% (1044)	29.67% (486)
Linux (Q388)	1507	78.3% (1180)	59.99% (904)	51.09% (770)	35.5% (535)	86.99% (1311)	76.64% (1155)	77.57% (1169)
Android (Q94)	1452	78.58% (1141)	62.95% (914)	55.3% (803)	47.52% (690)	85.95% (1248)	67.01% (973)	45.8% (665)
PlayStation 4 (Q5014725)	1405	84.7% (1190)	74.38% (1045)	71.17% (1000)	59.43% (835)	85.55% (1202)	74.59% (1048)	67.9% (954)
Super Nintendo Entertainment System (Q183259)	1369	86.05% (1178)	69.47% (951)	73.78% (1010)	65.08% (891)	86.12% (1179)	78.45% (1074)	29.95% (410)
Amiga (Q100047)	1363	76.23% (1039)	58.25% (794)	62.51% (852)	60.9% (830)	95.08% (1296)	63.46% (865)	32.94% (449)
ZX Spectrum (Q23882)	1352	77.22% (1044)	50.89% (688)	65.09% (880)	66.42% (898)	85.58% (1157)	64.28% (869)	37.87% (512)
PlayStation (Q10677)	1280	86.88% (1112)	76.8% (983)	80.08% (1025)	76.88% (984)	94.06% (1204)	79.22% (1014)	51.8% (663)

Activities

Alignment

- Property creation (Identifiers ❤️)
- Mix 'n' match-ing
- More advanced matching & bulk import
    - Examples: Mixer ID, PCGamingWiki, MediaArt/GameFAQ import

Only 9,6K items without identifiers!
(down from 15K 5 months ago)



Data Integration

The project relies on several tools to import data into Wikidata:

- Wikidata website
- Tools dedicated to Wikidata: Mix'n'match and its gadget, QuickStatements, ...
- General tools: OpenRefine, Dataiku DSS, ...

→ Importance of an healthy ecosystem of tools.

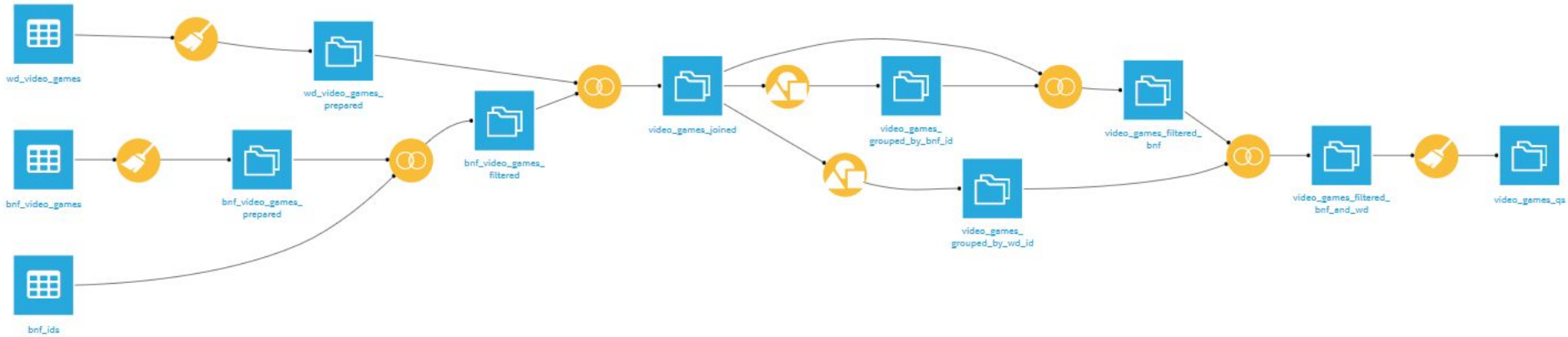
Some imports:

- Connecting PCGamingWiki and Wikidata
- Linking the OLAC Video Game Vocabulary with Wikidata

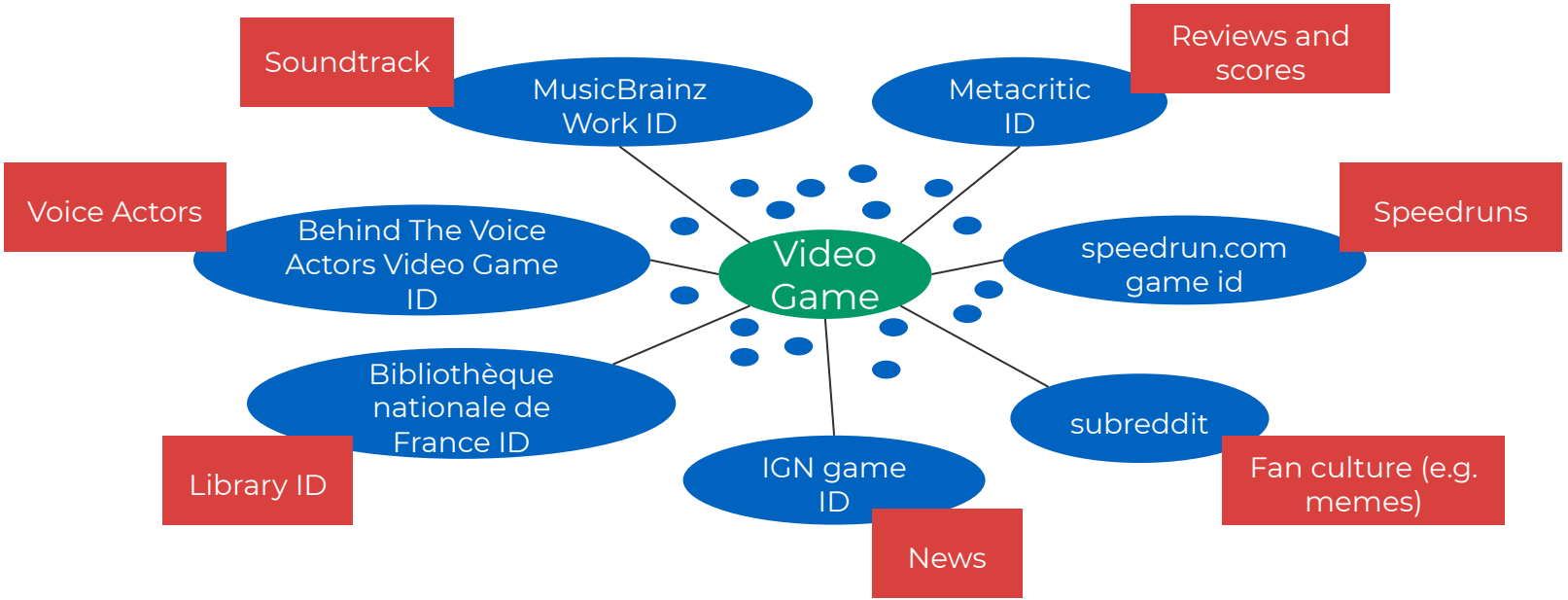
Case Study

Matching BnF and Wikidata video games using Dataiku DSS

- Project to align the catalog of National Library of France (BnF) about video games (~4000 entries) with Wikidata
- Development of a Dataiku project
- Development of a Dataiku plugin to facilitate SPARQL querying
- Contact established with BnF



THE Linking Hub for Video Games



Data Modeling challenges

- At the moment: (more or less) one Wikidata item per video game
- Challenging for identifiers
  - e.g. Identifier points to a specific edition or release of a game
  - Specific qualifiers needed
- Editions/versions/releases, hardware, relation with other items, ...
- refer to several other cultural artifacts, actors and subjects
  - leads to several (sometimes complex) relationships with other Wikidata items or with external identifiers
- In discussion: one item per video game, release, ...
- Academic literature: FRBR considered not precise enough, several models proposed

Video Game Research

- Video games have become a subject of interest in research in recent years
- Relative new field in academia
- Little (reliable) research data
- No authority for video game data
- Wikidata to the rescue!

WIKI DATA CON
Wikidata & languages 2019

diggr.link

Video Game Research

- Creating and using links
- Curation of data about video games and video game companies
- Alignment with OLAC Video Game Genre Vocabulary
- Using Wikidata as Work Authority for Video Games (Fukuda, 2019)

WIKI DATA CON
Wikidata & languages 2019

diggr.link

Video Game Research

LEMONGRAB

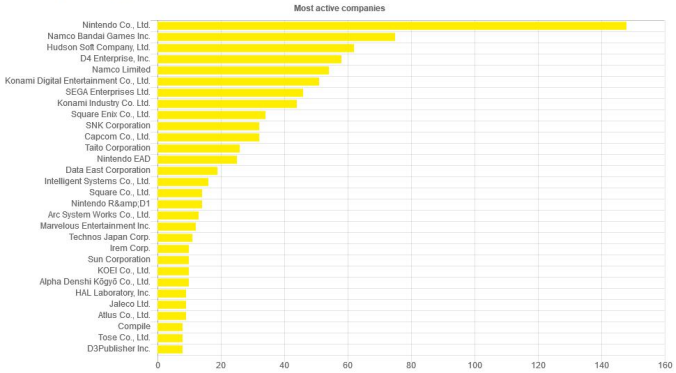
Platforms

- Nintendo Wii

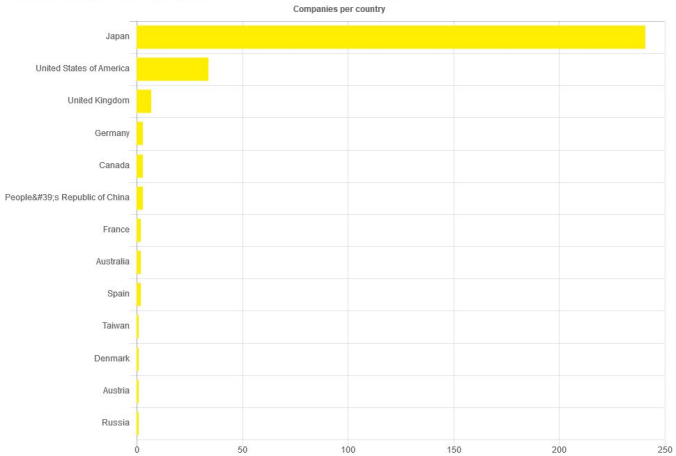
Countries

- Japan

Number of companies: 341

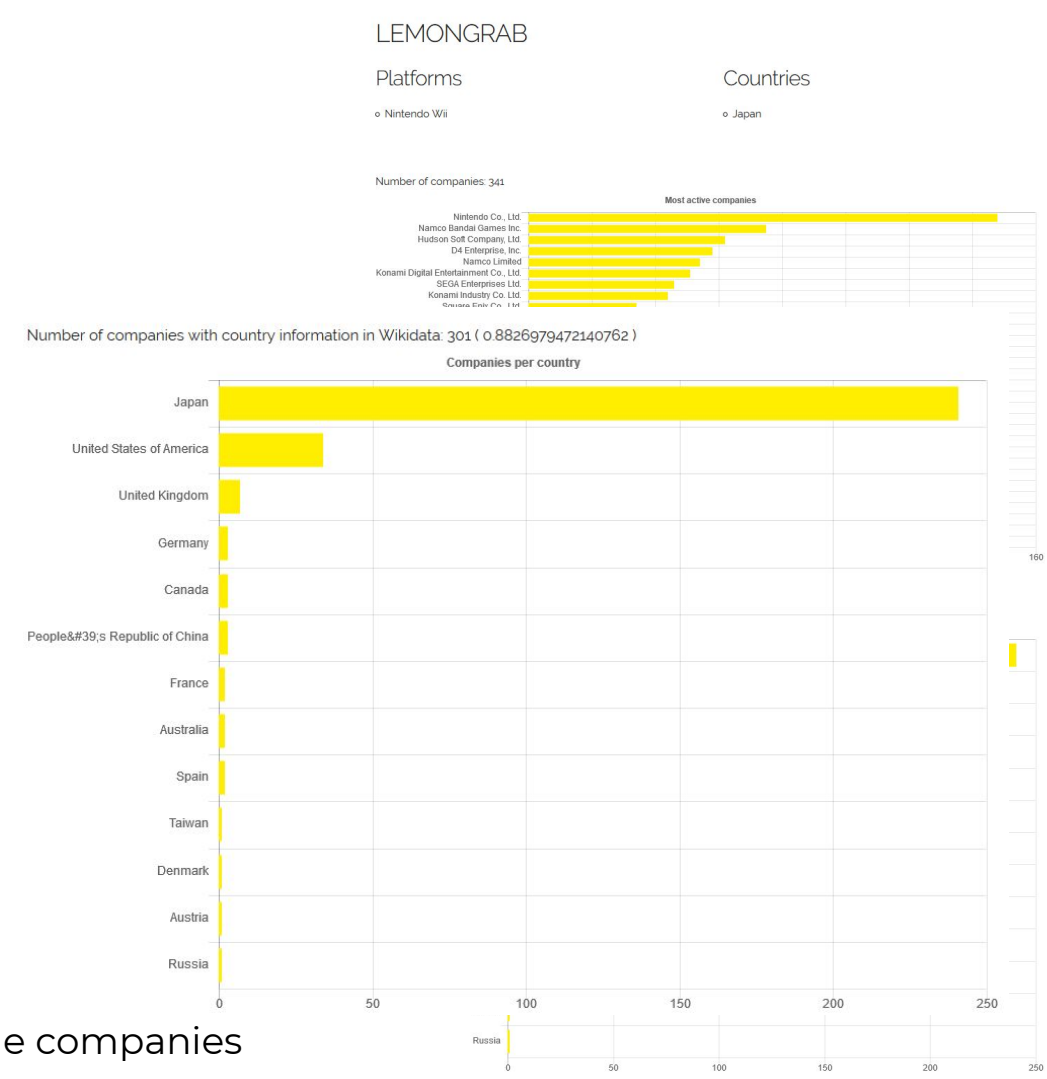


Number of companies with country information in Wikidata: 301 (0.8826979472140762)



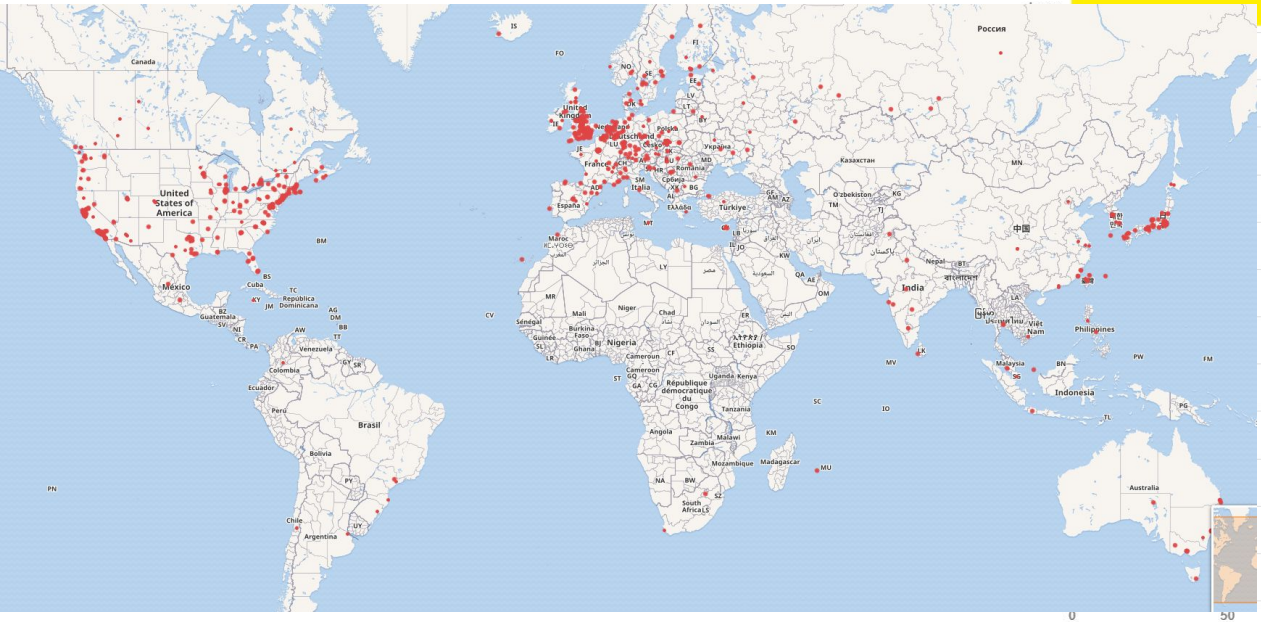
Locations of video game companies

Video Game Research



Locations of video game companies

Video Game Research



Locations of video game companies

LEMONGRAB

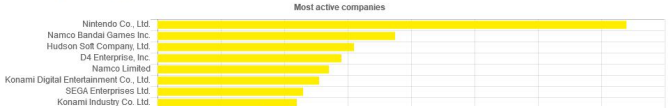
Platforms

- Nintendo Wii

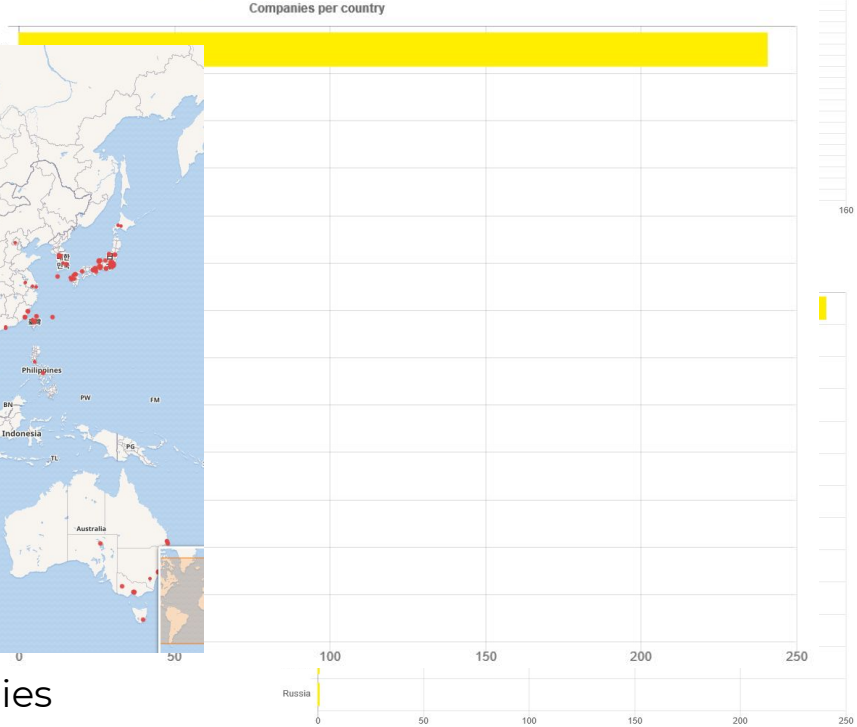
Countries

- Japan

Number of companies: 341



Number of companies with country information in Wikidata: 301 (0.8826979472140762)



What do we want to be?

- The one-stop shop for researchers
- Being/Staying useful to Wikipedias
- Enhancing other databases
  - Visual Novel Database
  - Internationale Computerspielesammlung
- Powering third-party applications
  - Example: https://vglist.co/

Thank you SPARQL!

But our data is in another database!