

APIs for the Modern World

Addison Hart

Background

- Reverse engineer and data scientist with experience in geospatial, transit and video game APIs
- Created the Public Mario
 Maker 2 API

Super Mario Maker 2 Public API

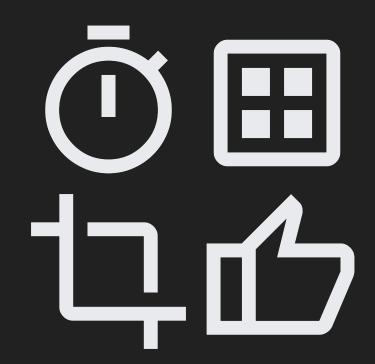
Made by TheGreatRambler with help from Tatiaus, Gible_V, Funnier04, Warspyking, AppleSinger, T04Dw, and Wizulus. This is a hosted instance of MariOver.

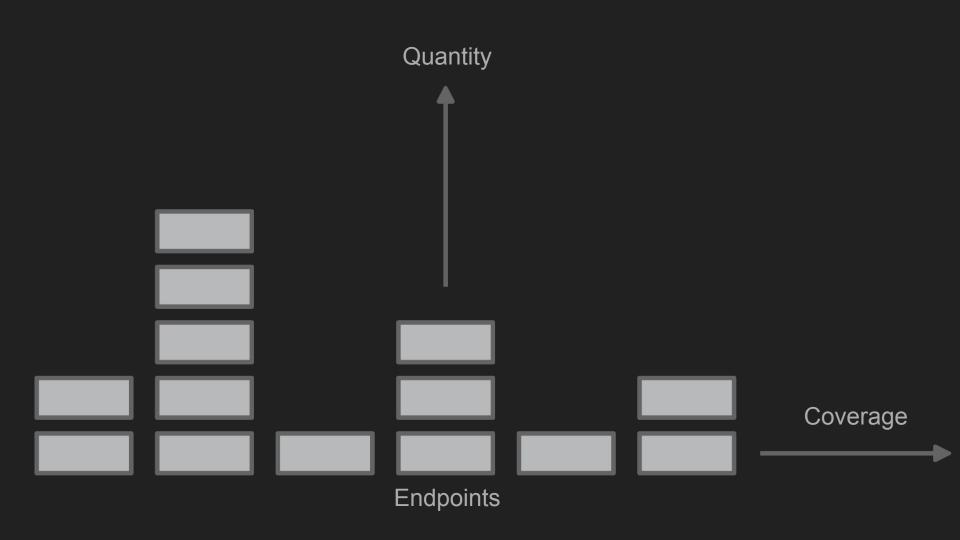
/level_info/{course_id} /user_info/{maker_id} /level_info_multiple/{course_ids} /level_comments/{course id} /level played/{course id} /level_deaths/{course_id} /level_thum /level_ent: /level data/ /level da /ninji_info /ninji_ghosts /get_posted/{ /get_liked/{mak /get_played/{maker_id} /get_first_cleared/{maker_id} /get_world_record/{maker_id} /get_super_worlds /super_world/{super_world_id} /search_endless mode /search_new

/search_popular

What Makes a Great API?

- Latency
- Quantity
- Coverage
- Quality



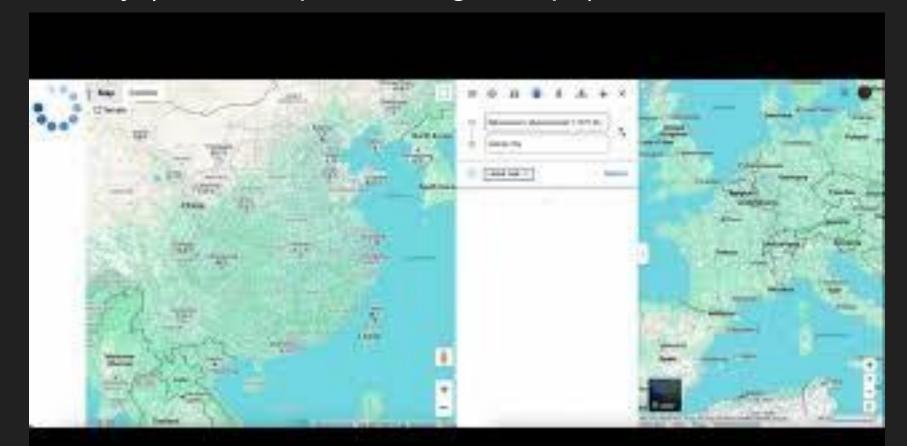


Latency



Serving users fast with good performance

Latency (Baidu Maps vs Google Maps)



Latency (ranking in Mario Maker 2 Ninji)

```
// Obtain player rank
          row := s.Conn.QueryRow(context.Background(),
              SELECT pid_rank FROM (SELECT pid,
              DENSE_RANK() over (ORDER BY time ASC)
                  AS pid_rank FROM ninji_time WHERE
                  data id = $2 AND active IS TRUE) AS _
              WHERE pid = $1 LIMIT 1, current_pid,
              data_id)
          err := row.Scan(&histogram.Rank)
          if err == pgx.ErrNoRows {
              // If have not played rank is 0
              histogram.Rank = 0
          } else if err != nil {
304
              return histogram, err
321
          // Get histogram of values
          // Always starts at 10000, always ends at
          120000, always 110 buckets
          rows, err := s.Conn.Query(context.Background(),
               `SELECT WIDTH_BUCKET(time, 10000, 120000,
              110) FROM ninji_time WHERE data_id = $1
              AND active IS TRUE`, data_id)
          if err != nil {
              return histogram, err
```



Latency (ghosts in Mario Maker 2 Ninji)

```
// Select n times around a time, with an equal
          number on either side
          only active := "AND active IS TRUE"
          if true {
              only active = ""
          rows, err := s.Conn.Query(context.Background(),
              fmt.Sprintf(`(SELECT pid, time, replay_id
              FROM ninji_time WHERE data_id = $1 AND
                  replay_id IS NOT NULL %s AND time > $2
                  ORDER BY ABS($2 - time) ASC LIMIT $3)
              UNION ALL
540
              (SELECT pid, time, replay_id FROM
              ninji_time WHERE data_id = $1 AND
                  replay id IS NOT NULL %s AND time < $2
                  ORDER BY ABS($2 - time) ASC LIMIT $3)
                   `, only active, only active),
              param.DataId, param.Time, param.Count/2)
           if err != nil {
545
546
              return infos, err
          defer rows.Close()
548
```



Latency (world records in Mario Maker 2)

```
func (s *DB) GetPlayerResultWorldRecord(dataId
                                                                  int64) (datastore.Pid, uint32, error) {
                                                                       // Get first person with the best time
                                                                       rows, err := s.Conn.Query(contact)
                                                                       (), "SELECT pid, best
                                                                       WHERE data_id = $1 AN is_world_record = True
                                                                       LIMIT 1". dataId)
                                                                       if err != nil {
                                                                              return datastore.Pid(0), 0, err
func (s *DB) UpdateRecordsAndFirstClears(result PlayResult, player_result
PlayerResult) (bool, bool, error) {
                                                                       defer rows.Close()
    r newWorldRecord, newFirstClear bool
   // Modify world record status of other player if they were beaten
  oldPid, oldTime, err := s.GetPlayerResultWorldRecord(int64(result.
                                                       104
                                                                       if !rows.Next() {
  DataId))
   hadWorldRecord := true
                                                                              // No world record
                                                                             //return datastore.Pid(0), 0xFFFFFFFF, nil
  } else if oldPid == 0 {
      hadWorldRecord = false
                                                                             return datastore.Pid(0), 0, nil
    f hadWorldRecord && wint32(result.Time.Int32) < oldTime -</p>
     newWorldRecord = true
     _, err = s.Conn.Exec(context.Background(), `UPDATE player_result
                                                                       var pid datastore.Pid
     SET is_world_record = False WHERE pid = $1 AND data_id = $2`,
     oldPid, player result.DataId
                                                                        var time pgtype.Int4
                                                                       err = rows.Scan(&pid, &time)
      _, err = s.Conn.Exec(context.Background(), `UPDATE player_result
                                                                       return pid, uint32(time.Int32), err
     SET is_world_record = True WHERE pid = $1 AND data_id = $2',
     player result.Pid, player result.DataId)
  } else if !hadWorldRecord {
     newFirstClear = true
     _, err = s.Conn.Exec(context.Background(), `UPDATE player_result
     SET is first clear = True, is world record = True WHERE pid = $1
     AND data_id = $2', player_result.Pid, player_result.DataId
```

urn newWorldRecord, newFirstClear, nil

```
func (s *DB) GetPlayerResultWorldRecord(dataId
      int64) (datastore.Pid, uint32, error) {
          // Get first person with the best time
          rows, err := s.Conn.Query(context.Background
          (), "SELECT pid, best time FROM player result
          WHERE data id = $1 ORDER BY best time ASC
          LIMIT 1", dataId)
          if err != nil {
              return datastore.Pid(0), 0, err
          defer rows.Close()
          if !rows.Next() {
104
              // No world record
              return datastore.Pid(0), 0, nil
          var pid datastore.Pid
          var time pgtype.Int4
          err = rows.Scan(&pid, &time)
          return pid, uint32(time.Int32), err
```

Is this overengineered?

Latency (world records in Mario Maker 2)

```
// update user stats cache
func (s *DB) UpdateUserStatsCache(pid datastore.
Pid) error {
    var stats UserStats
    // query from player_result table
    row := s.Conn.QueryRow(context_Background(),
    "SELECT COUNT(pid) _______, SUM(deaths
    SUM(cleared::int SUM(is_world_record::int),
    SUM(is_first_clear. FROM player result
    WHERE pid = $1", pid)
    err := row.Scan(&stats.Plays, &stats.Tries, &
    stats.Deaths, &stats.Clears, &stats.
    WorldRecords, &stats.FirstClears)
    if err != nil {
    _, err = s.Conn.Exec(context.Background(),
        INSERT INTO user_stats (id, pid, plays,
        tries, deaths, clears, world_records,
        first_clears) VALUES (DEFAULT, $1, $2, $3,
        $4, $5, $6, $7)
        ON CONFLICT (pid) DO UPDATE SET plays =
        $2, tries = $3, deaths = $4, clears = $5,
        world records = $6, first clears = $7
     `, pid, stats.Plays, stats.Tries, stats.
    Deaths, stats.Clears, stats.WorldRecords,
    stats.FirstClears)
    return err
```

```
func (s *DB) SearchCoursesWorldRecord(state *user.
      State, query_pid datastore.Pid, param view.
      SearchCoursesParam) ([]datastore.CourseInfo, []
      uint32, bool, error) {
          // Get all world records held by this user
99
100
          rows, err := s.Conn.Query(context.Background
          (), "SELECT data id Till plan result WHERE
          pid = $1 ND is_world_record = True ORD R BY
          time_obtained_DESC_LINIT_ #2_0173E1 $3",
          query_pid, param.Size, param.Offset)
101
          if err != nil {
102
              return []datastore.CourseInfo{}, []uint32
              {}, false, err
103
          defer rows.Close()
104
```

NO! O(1) vs O(n)

Latency (restrict access)

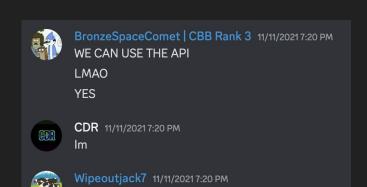
- Sometimes clever engineering is not enough
- Rate limiting or request throttling may be enough, as long as your users are aware of such measures and not trying to actively break them

lock = asyncio.Semaphore(5)

```
await check_tokens()
async with lock:
    async with backend.connect(s, HOST,
    PORT) as be:
        async with be.login(str(user_id),
        auth info=auth info) as client:
            store = datastore.
            DataStoreClientSMM2(client)
            print("Want course info for "
            + course id)
            course info json = await
            obtain course info(course id,
            store, noCaching)
            if invalid level
            (course info json):
                return ORJSONResponse
                (status code=400,
                content=course info json)
            return ORJSONResponse
            (content=course info json)
```

Latency (benefits)

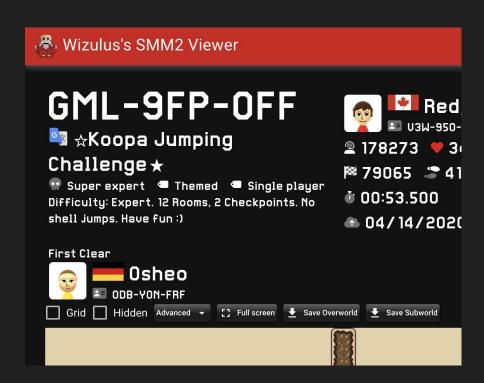
- Live data feeds enable real time reactions to events
- Tools like Google Takeout are not suitable for real time reactions





Latency (benefits)

 Have special exemptions on rate limiting or request throttling for specific trusted users with smart caching policies



Quantity



Providing as much data as possible over each endpoint

Quantity

 Mario Maker 2 only serves the last 1000 comments of a level

```
"courses": [

"name": "Celestial Sprint Speedrun [20s]",
"description": "Never stop running! Trust
the fake traps! And most importantly,
enjoy!",
"uploaded_pretty": "29-6-2019 21:08:13",
"uploaded": 1561842493,
"data_id": 3456358,
"and the second second
```

```
Returns the comments of the specified course

[ {course_id}: The 9 digit code used to identify a course | 7N1MVBWKF
```

https://tgrcode.com/mm2/level_comments/7N1MVBWKF

```
unk17: 0,
 poster: [Object]
},
 comment_id: '20241016150033345909_81c7a1db10a5a50f_34bd66',
 posted_pretty: '16-10-2024 15:00:33',
 posted: 1729090833,
 clear required: false,
 reaction_image_id: 5,
 reaction_image_id_name: 'Wow!',
 type_name: 'Reaction Image',
 type: 2,
 has beaten: false,
 x: 2148,
 y: 128,
 reaction face: 0,
 reaction_face_name: 'Normal',
 unk8: 0,
 unk10: 0,
 unk12: false,
 unk14: '',
 unk17: 0,
 poster: [Object]
... 883 more items
```

Quantity (pagination)

- Enable pagination to reduce average payload size while still allowing access to all data
- Up to clients how much they value quantity (in both time and money)



Quantity (options for advanced users)

Single request option for normal use



Bulk request option for advanced use

Accessing Bulk Whois Data

If your request is approved, you will be able to access the Bulk Whois data by performing these steps:

- 1. Log in to ARIN Online.
- 2. Select Downloads & Services from the navigation menu.
- 3. In the Bulk Whois section, choose Bulk Whois Data. The next window provides a list of files and formats in which the data is available. You have the option to download a file of all Whois object types (POCs, Orgs, NETs, and ASNs) or to selectively download a particular Whois object type in compressed archive (.zip), text (.txt), or extended markup language (.xml) format.

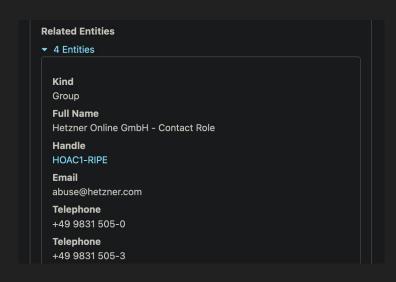
If you want to download a file that contains a combination of Whois object types, you can enter the download URL directly into your browser window, using the following format:

https://accountws.arin.net/public/rest/downloads/ bulkwhois/XXX+YYY.zip

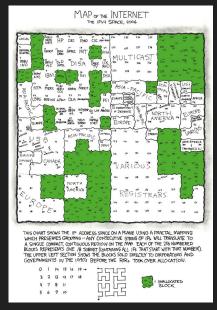
where XXX and YYY can be any combination of:

Quantity (options for advanced users)

This use case...



Is different from this use case



But you want to support both

Quantity (provide cross references to other data)

- Provide cross references to additional data if the average user does not use it
- Most users access this endpoint for times, not for replay files

```
"ghosts": [
       "time": 32518,
       "replay file": {
           "url": "https://dd04qv7ro5r2l.
           cloudfront.net/10.
           ngs_lp1_22306d00_datastore/ds/1/
           relation_data/event_course_ghost/
           event-course-ghost 982 13ef99468f2d4233
            _b9b71a_20191207232345036295",
           "size": 2434,
           "filename":
           "event-course-ghost_982_13ef99468f2d423
           3_b9b71a_20191207232345036295"
        "region": 0.
       "region_name": "Asia",
       "code": "TQ61RQJHF",
       "pid": 1436535334483345971,
       "name": "あさひさし",
       "country": "JP",
```

Quantity (progressive loading)

Some APIs can benefit from serving in chunks in different resolutions

 Google Maps serves different zoom values (resolution) and different offsets into an equirectangular image

Once again, up to clients how much they value quantity (in both time and

money)



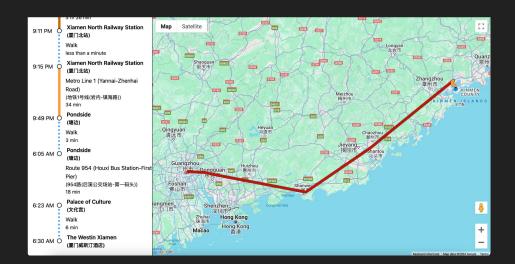
Coverage

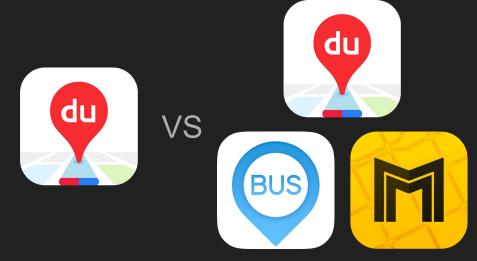


Providing to users as much data as you have

Coverage (expose all)

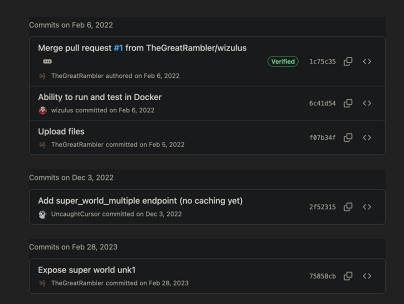
- Baidu Maps (website version) does not expose bus or metro schedules, had to RE 2 additional apps
- App had access to additional data that it did not expose to users





Coverage (accommodate new use cases)

- Listen to the community for new endpoint requests
- Try to design API such that new endpoints are improvements on old ones, you could still access that data earlier



Quality (interoperability)

- Expose GUIDs that are compatible with other APIs
- Data like names and coordinates do not always correspond with other APIs

```
"kind": 0,
"last_station": "\u7b2c\u4e00\u7801\u5934
"name": "954\u8def(\u540e\u6eaa\u516c\u4e
"st_uid": "1340124536317357055",
"startTime": "06:05",
"station_num": 4,
"time": 471,
"timetable": "06:05-21:00",
"timetable_ext": "",
```

```
"start_station": "后溪公交场站",
"end_time": "2100",
"lineId": "0592-954-1",
"final_station": "第一码头站",
"start_time": "0605",
"buses": [

"syncTime": 1,
"pRate": -1,
"travelTime": 1441
```

uno.csv

- 1 XMML01<,>ML<,>Line 1<,>1号线<,>1號綫<,>
- 2 XMML02<,>ML<,>Line 2<,>2号线<,>2號綫<,>
- 3 XMML03<,>ML<,>Line 3<,>3号线<,>3號綫<,>
- 4 XMMW01A<,>MW<,>For Zhenhai Road
- 5 XMMW01B<,>MW<,>For Yannei<,>岩口
- 6 XMMW02A<,>MW<,>For Wuyuanwan<,

Examples of Good GUIDs

- ISBN
 - Google Books
 - Open Book
 - Library of Congress
- UPC (Universal Product Code)
 - Walmart
 - o Go-UPC
- ISIN / ticker symbol
 - o Bloomberg
 - Robinhood

Oftentimes the biggest API should become the GUID



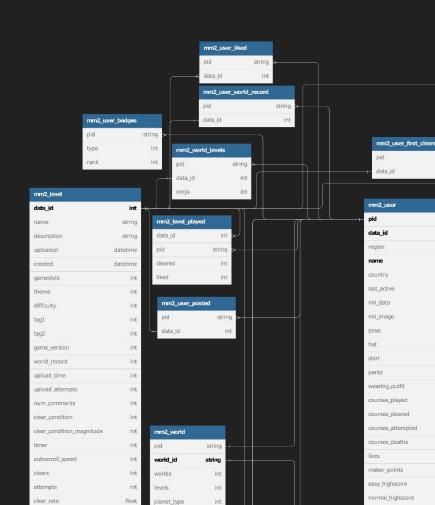
Quality



Providing a public service that inspires experimentation

Quality (self documenting)

- Give fields representing the same data type the same name
- Ensure every endpoint taking in the same data type accepts the same domain for that data type



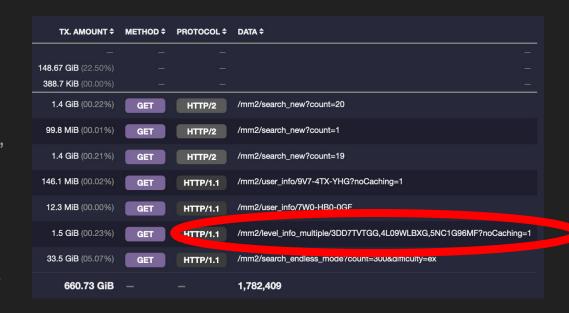
Quality (right data format for the job)

- Sending the wrong data format can incur additional performance penalties or impact developer experience
- Use well accepted standards when possible
 - JSON
 - Protobuffer
- When the data format isn't right, a reverse engineer like me will make a new one!

```
"isonr": {
   "data":
       "encryptResult": "xJE7SF7oqbMpNHvQ9Nu1cQCpwIhcVb5N
       +BI2s4kvPS5e4oKuNFGLweos+kPVMhYjwybFVa/aaeXF10YKltriwz+CJGcREESy
       +vsc/
       Ftee01ULCzBttrpGmwm7QMR10iqHZnH0TxfXTRPDc9qZV6A7nu5rvsCu0lMQqirRpCs
       nfS7X6P5wX6X4WFCwUaJP4sfnex7iL8Qe0iu7qKUGk3ub5tiKA1m1xpQHdFSQru1rrY
       +ydQGURDVaygK0nMHtYZVPnmDxVeKQt03awUuWbKnND6KtY0Goa0IFcoCw2ZyZzo1Ey
       N1eAjQcKLaRF0
       +4BURRy0prTx90UGMmDAsJhwU9FhzFCPIQBDxnMcnrTeQpUU43aHUjq4EWzG
       +j4I454UzhTaqXCrmR0AbPBZ6+DHIW7fzSTb8iIydJAMUj6/7t0fSdsmSEjFdzraR
       +Vh3vHF93W5EwGY+4yITCdP2uXw3gf7t07TCNzB3vyf0r2he8SPBp8TojExGk56h/
       0ebexLl0XAy/01E/ShcYPo0x8a1/
       0E4WpzUDlKkFv2y5pPjXkpNcaK9ZOaxBqHpA3Iv8hrYsuneM5Snpvu1Ity7kNhN/qT/
       Af6T0NugDBuBftihT80wCL9tYTc1fgLpeqFiT5/CE3MwF/
       78lSZYeKQnu1q7k460DUF2F0aXJdoHAtzVrkdVoDUDRoiziFUqie/
       Lgon1clZ3hyTp0tRhWU6+RoMTp/
       GqJwxXhcd5D8xb9m35PRlyh8sTyjGnkeHRSrHOtiPdVEclkGWU9yV52riBMDjhU8Von
       MpGT2do9ifCWnLco1MliZgVFQveLkd0bcuhRD54amsuf3dW5Ws9BH0cL9zQQJ8gGfcE
       RuXR0F0slH0qzEVZs9BbTphgXfFfgaFuNBN89E6VJ
       +p0dh9D1qytA8EpfpADwZqR9rvGP+5MeMAOW2nTBIqmpFUSCsl/
```

Quality (use logging to understand user behaviors)

- One of the most popular requests is level info for 3 seemingly random levels with no caching
- Upon further inspection, and asking the community, this user is using the "first cleared" field of these levels as a replacement for a "user_info_multiple" endpoint
- These 3 players are the top 3 in endless easy, a competitive gamemode leaderboard



Quality (example code)

Provide example code for both kinds of users

Python example constructing raw request

Go example using a client library

Quality (documentation)

- Document both the domain (kinds of inputs) and range (kinds of outputs) of your endpoints
- Link to external documentation when using a standardized format, like GTFS or DZI (Deep Zoom Image)

```
GameStyles = {
    0: "SMB1",
    1: "SMB3",
    2: "SMW".
    3: "NSMBU",
    4: "SM3DW"
Difficulties = {
    0: "Easy",
    1: "Normal",
   2: "Expert",
    3: "Super expert"
CourseThemes = {
    0: "Overworld",
    1: "Underground",
    2: "Castle",
    3: "Airship",
    4: "Underwater",
   5: "Ghost house",
    6: "Snow",
    7: "Desert",
    8: "Sky",
    9: "Forest"
```

```
TagNames = {
   0: "None",
   1: "Standard",
   2: "Puzzle solving",
    3: "Speedrun",
    4: "Autoscroll",
    5: "Auto mario",
   6: "Short and sweet",
    7: "Multiplayer versus",
   8: "Themed",
   9: "Music",
    10: "Art",
   11: "Technical",
   12: "Shooter".
   13: "Boss battle",
   14: "Single player",
   15: "Link"
Regions = {
   0: "Asia",
   1: "Americas",
    2: "Europe",
    3: "Other"
```

Quality (compatibility)

- Have an established plan to preserve compatibility
 - Endpoint prefixes
 - Backwards compatibility
- Advertise changes
- Choose the right UIDs at the start

○ 2024 update

The endpoint https://www.google.com/maps/preview/photo has been turned into https://www.google.com/maps/rpc/photo/listentityphotos. That change is reflected in this blopost where applicable.

⇔ Motivation

There are two websites which have always led to hours of unplanned exploration for me: Wikipedia and Google Maps.

APIs Done Right | Github

- Latency: Webhook informs your website immediately upon some action
- Quantity: Pagination allows retrieval of all available data, like list of all stargazers
- Coverage: Covers a ton of endpoints, from repositories to available emojis
- Quality: Example code for cURL and JS

```
"data": {
  "repository": {
    "stargazers": {
      "edges": [
          "node": {
            "login": "friedkeenan"
          "starredAt": "2020-11-10T20:20:59Z"
        },
          "node": {
            "login": "noahc3"
          "starredAt": "2020-11-11T01:21:31Z"
          "node": {
            "login": "XorTroll"
          "starredAt": "2020-11-11T20:52:36Z"
          "node": {
            "login": "Pysis868"
```

APIs Done Right | Hacker News

- Latency: Low latency and small payloads
- Quantity: Provides public, incrementing, ID you can use to walk backwards
- Coverage: Advertised as "a dump of our in-memory data structures", has been used to create a custom client
- Quality: Prefixed endpoints to preserve compatibility

| | | hckr news an unofficial alternative <u>hacker news</u> interface. | |
|----------|--------|--|--|
| | | auto-refresh disabled | |
| comments | points | | |
| 2 | 4 | Rocket Cars in Cleveland (www.wsj.com) | |
| 7 | 11 | Astronomers Found Something Cold and Wet Near Uranus (g | |
| 7 | 38 | The history of Monokai (monokai.pro) | |
| 11 | 51 | A comparison of Rust's borrow checker to the one in C# (em- | |
| 20 | 20 | Typing fast is not so important (morgan.zoemp.be) | |
| 12 | 28 | ebras Trains Llama Models to Leap over GPUs (www.nextpl | |
| 9 | 19 | FDA Permits Marketing of Digital Game to Improve Attention | |
| 49 | 32 | Testing for gender differences in Python programming style | |

Expose All the Data!

| Vorld Re | Corus | | | ₩ Worldwide |
|----------|----------------------|------------|---------------------|-------------|
| Cup | Track | Time | Player | Nation |
| | Mario Kart Stadium | 1'34"042 安 | Rai-Oh | • |
| | Water Park | 1'38"816 | jacob | |
| | Sweet Sweet Canyon | 1'47"589 | Shira | • |
| | Thwomp Ruins | 1'45"499 | jacob | |
| | Mario Circuit | 1'42"913 | Shira | • |
| | Toad Harbor | 1'53"947 | jacob | |
| | Twisted Mansion | 1'53"646 | m ^[alt] | |
| | Shy Guy Falls | 1'51"596 | jacob | |
| | Sunshine Airport | 1'54"623 | Alberto | 100 |
| | Dolphin Shoals | 1'56"984 | Darren• | |
| | Electrodrome | 1'54"706 | Fi ^[alt] | • |
| | Mount Wario | 1'39"450 | jacob | |
| | Cloudtop Cruise | 1'56"353 😝 | ROA◆どらさ | • |
| | Bone-Dry Dunes | 1'46"527 | jacob | |
| | Bowser's Castle | 1'57"758 | jacob | |
| | Rainbow Road | 1'59"170 | jacob | |
| | Wii Moo Moo Meadows | 1'20"283 | Kyle Wade | |
| | GBA Mario Circuit | 1'20"357 | | |
| | DS Cheep Cheep Beach | 1'42"852 | さーもんほうさく | • |



Credits

- https://tgrcode.com/
 - https://tgrcode.com/posts/mario_maker_2_api
 - https://tgrcode.com/posts/mario_maker_2_datasets
 - https://tgrcode.com/posts/reverse engineering google streetview
- https://smm2.wizul.us/
- https://search.arin.net/rdap/
- https://www.arin.net/reference/research/bulkwhois/
- https://xkcd.com/195/
- https://developers.google.com/maps/documentation/elevation/start
- https://docs.github.com/en/rest
- https://github.com/HackerNews/API
 - https://hckrnews.com/
- https://www.mkdatabase.com/
- https://www.verizon.com/coverage-map/