Privacy-Preserving Analytics in and out of the Clouds

Jon Crowcroft,

http://www.cl.cam.ac.uk/~jac22

This work is funded in part by the EPSRC Databox (EP/N028260/1),
NaaS (EP/K031724/2) and Contrive (EP/N028422/1) and Turing/Maru (EP/N510129/1) projects.

Compute-First Networking

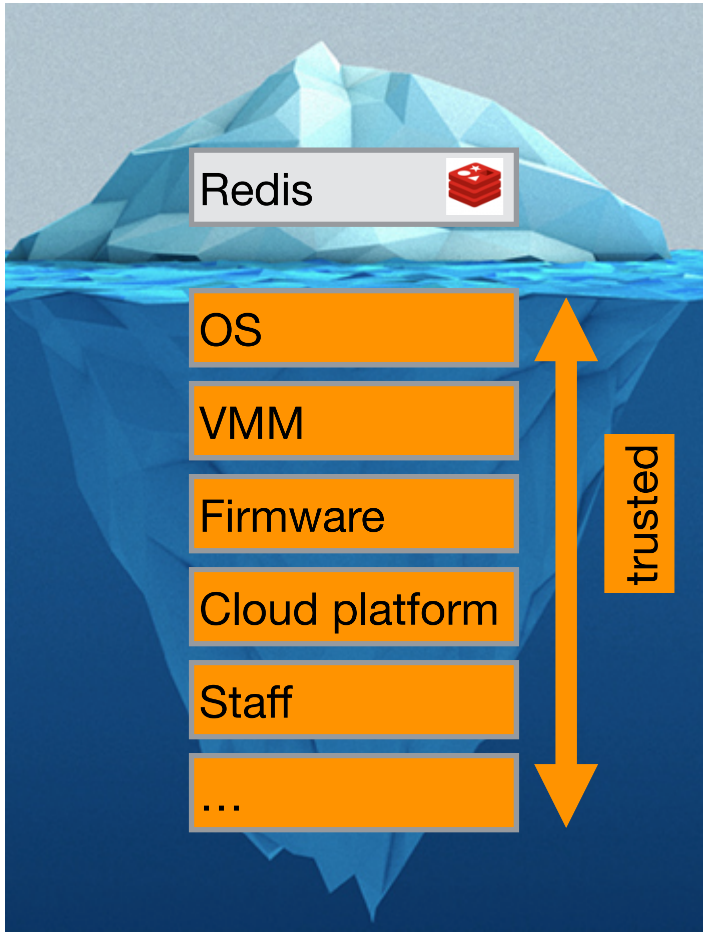
- Should do to IP, what IP did to POTS
- Need to think 40-50 years ahead in s/w terms
- Scale up (10^23 nodes per person)
- Verification or Generative (correct & smaller by design) code
- Invariants (C (max latency), plank’s constant (min size)) matter more…

1 In Cloud. Trust Issues: Provider Perspective

Cloud provider does not trust users

Use virtual machines to isolate users from each other and the host

VMs only provide one way protection

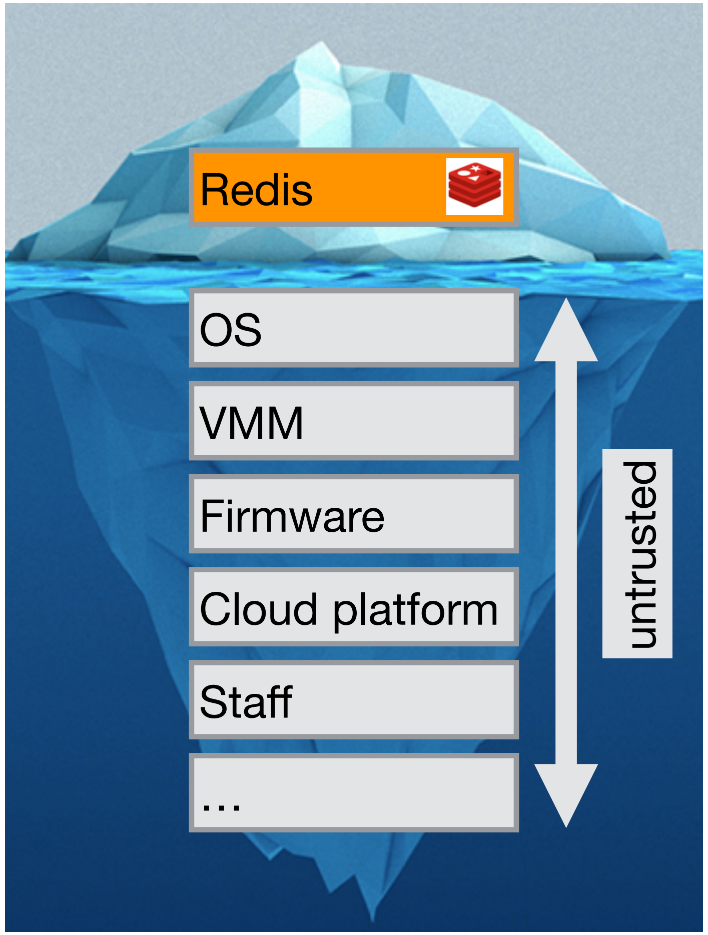


Trust Issues: User Perspective

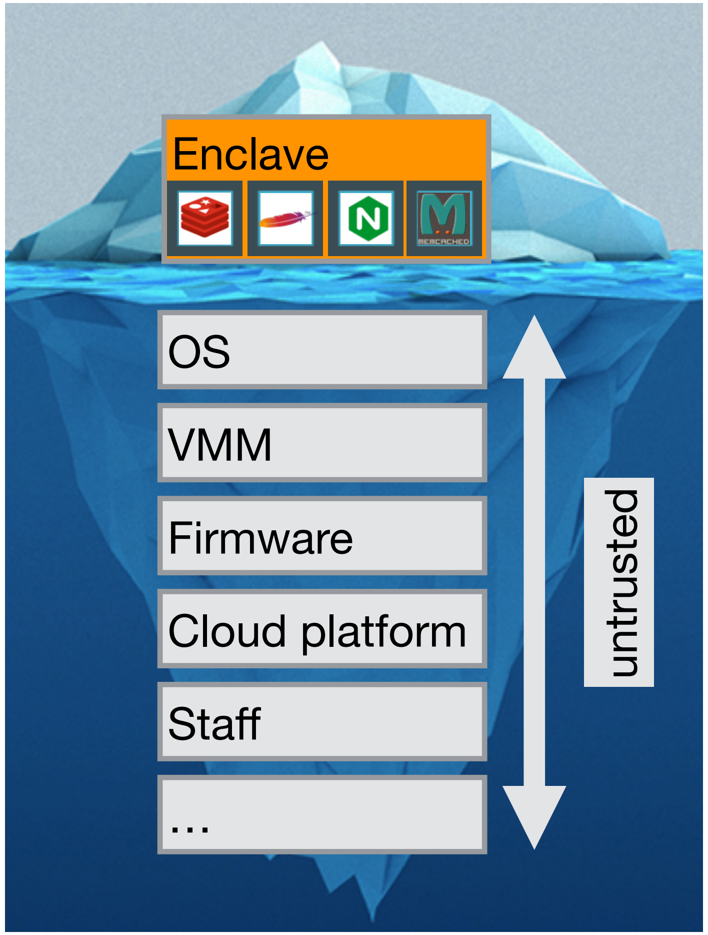
Users trust their applications

Users must implicitly trust cloud provider

Existing applications implicitly assume trusted operating system



Trusted Execution Support with Intel SGX



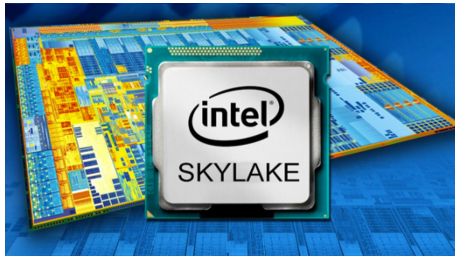
Users create HW-enforced trusted environment (enclave)

Supports unprivileged user code

Protects against strong attacker model

Remote attestation

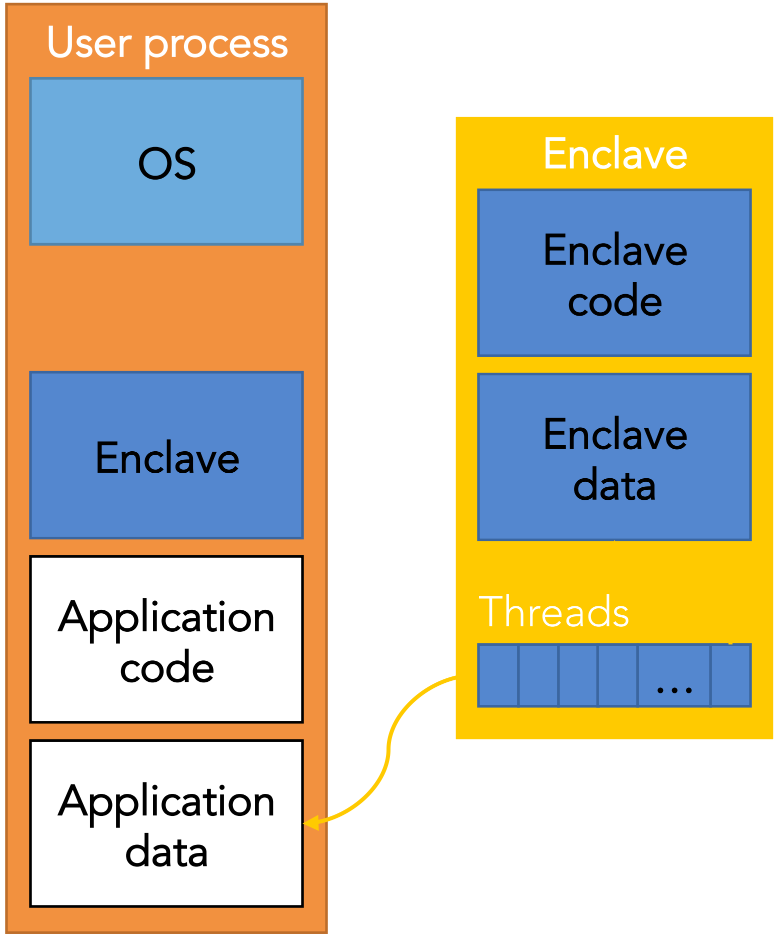
Available on commodity CPUs



Trusted Execution Environments

Trusted execution environment (TEE) in process

- Own code & data
- Controlled entry points
- Provides confidentiality & integrity
- Supports multiple threads
- Full access to application memory



Intel Software Guard Extensions (SGX)

Extension of Instruction Set Architecture (ISA) in recent Intel CPUs

- Skylake (2015), Kaby lake (2016)

Protects confidentiality and integrity of code & data in untrusted environments

- Platform owner considered malicious
- Only CPU chip and isolated region trusted

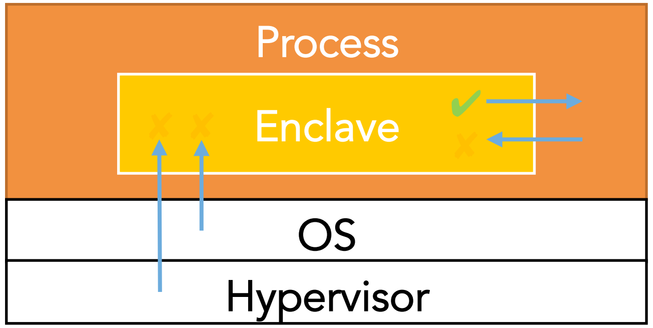
SGX Enclaves

SGX introduces notion of **enclave**

- Isolated memory region for code & data
- New CPU instructions to manipulate enclaves and new enclave execution mode

Enclave memory **encrypted** and **integrity-protected** by hardware

- Memory encryption engine (MEE)
- No plaintext secrets in main memory



Enclave memory can be accessed only by enclave code

- Protection from privileged code (OS, hypervisor)

Application has ability to defend secrets

- Attack surface reduced to just enclaves and CPU
- Compromised software cannot steal application secrets

SGX SDK Code Sample

SGX application: untrusted code

```
char request_buf[BUFFER_SIZE];
char response_buf[BUFFER_SIZE];

int main()
{
  ...
  while(1)
  {
    receive(request_buf);
    ret = EENTER(request_buf, response_buf);
    if (ret < 0)
      fprintf(stderr, "Corrupted message\n");
    else
      send(response_buf);
  }
  ...
}
```

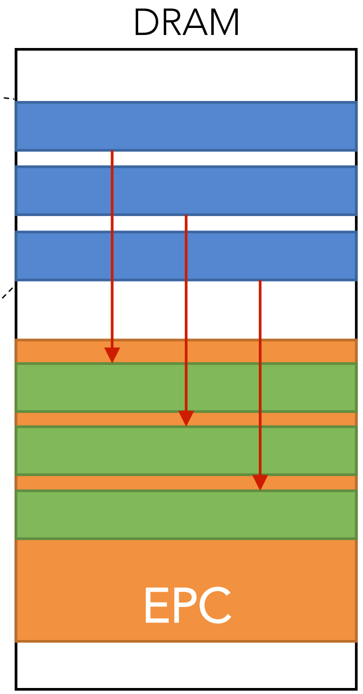
Enclave: trusted code

```
char input_buf[BUFFER_SIZE];
char output_buf[BUFFER_SIZE];

int process_request(char *in, char *out)
{
  copy_msg(in, input_buf);
  if(verify_MAC(input_buf))
  {
    decrypt_msg(input_buf);
    process_msg(input_buf, output_buf);
    encrypt_msg(output_buf);
    copy_msg(output_buf, out);
    EEXIT(0);
  } else
    EEXIT(-1);
}
```

Server:
- Receives encrypted requests
- Processes them in enclave
- Sends encrypted responses

SGX Enclave Construction

```
1 { char input_buf[BUFFER_SIZE];
2 { char output_buf[BUFFER_SIZE];

    int process_request(char *in, char *out)
    {
      copy_msg(in, input_buf);
      if(verify_MAC(input_buf))
      {
        decrypt_msg(input_buf);
3 {     process_msg(input_buf, output_buf);
        encrypt_msg(output_buf);
        copy_msg(output_buf, out);
        EEXIT(0);
      } else
        EEXIT(-1);
    }
```



Enclave populated using special instruction (EADD)

- Contents initially in untrusted memory
- Copied into EPC in 4KB pages

Both data & code copied before execution commences in enclave

SGX Support for Spark

SGX-Spark

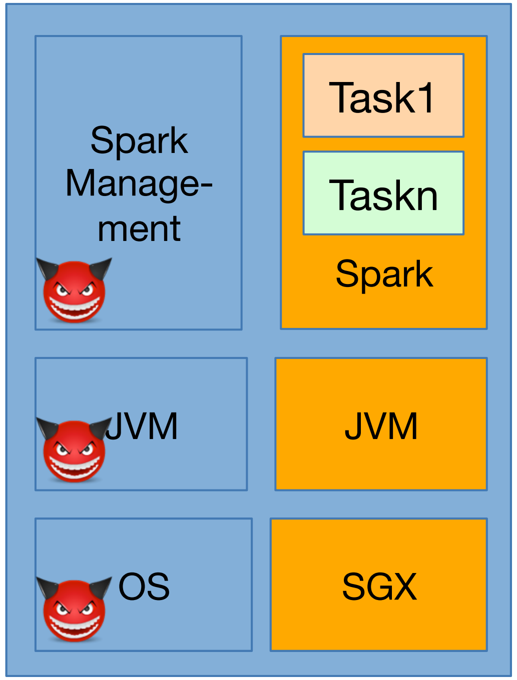
- Protect data processing from infrastructure provider
- Protect confidentiality & integrity of existing jobs
- No modifications for end users
- Acceptable performance overhead

Idea:
Execute only sensitive parts of Spark inside enclave

- Code that accesses/processes sensitive data

Code outside of enclave only accesses encrypted data

- Partition Spark
- Run two collaborating JVMs, inside enclave and outside of enclave



Partitioning Spark

Goal: Move minimal amount of Spark code to enclave

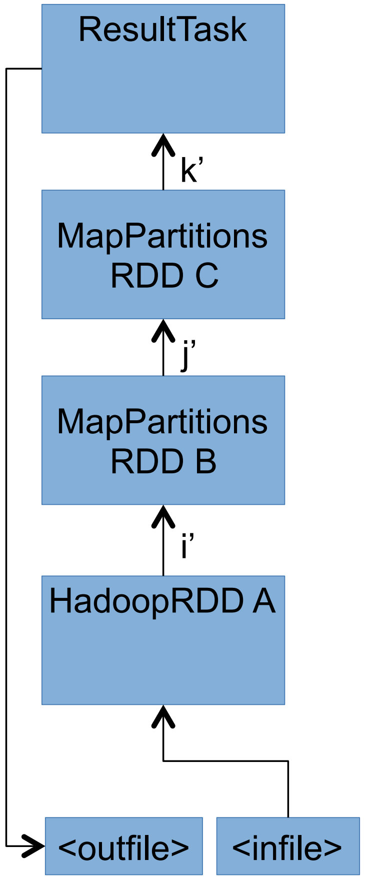
Outside	Enclave
HadoopRDD Provide iterator over input data partition (encrypted)	

`it`
MapPartitionsRDD
Execute user-provided function (f)
(eg `flatMap(line => {line.split(" ")})`
(i) Serialise user-provided function `f`
`f,it`
(ii) Send `f` and `it` to enclave JVM
(iv) Receive result iterator `it_result`
`it2 = it_result`
	(iii) Decrypt input data
	(iv) Compute `f(it) = it_result`
	(v) Encrypt result

ExternalSorter
Execute user-provided reduce function g
`g,it2`
(eg `reduceByKey{case (x, y) => x + y}`)
`it2_result`
	(iii) Decrypt input data
	(iv) Compute `g(it2) = it2_result`
	(v) Encrypt result

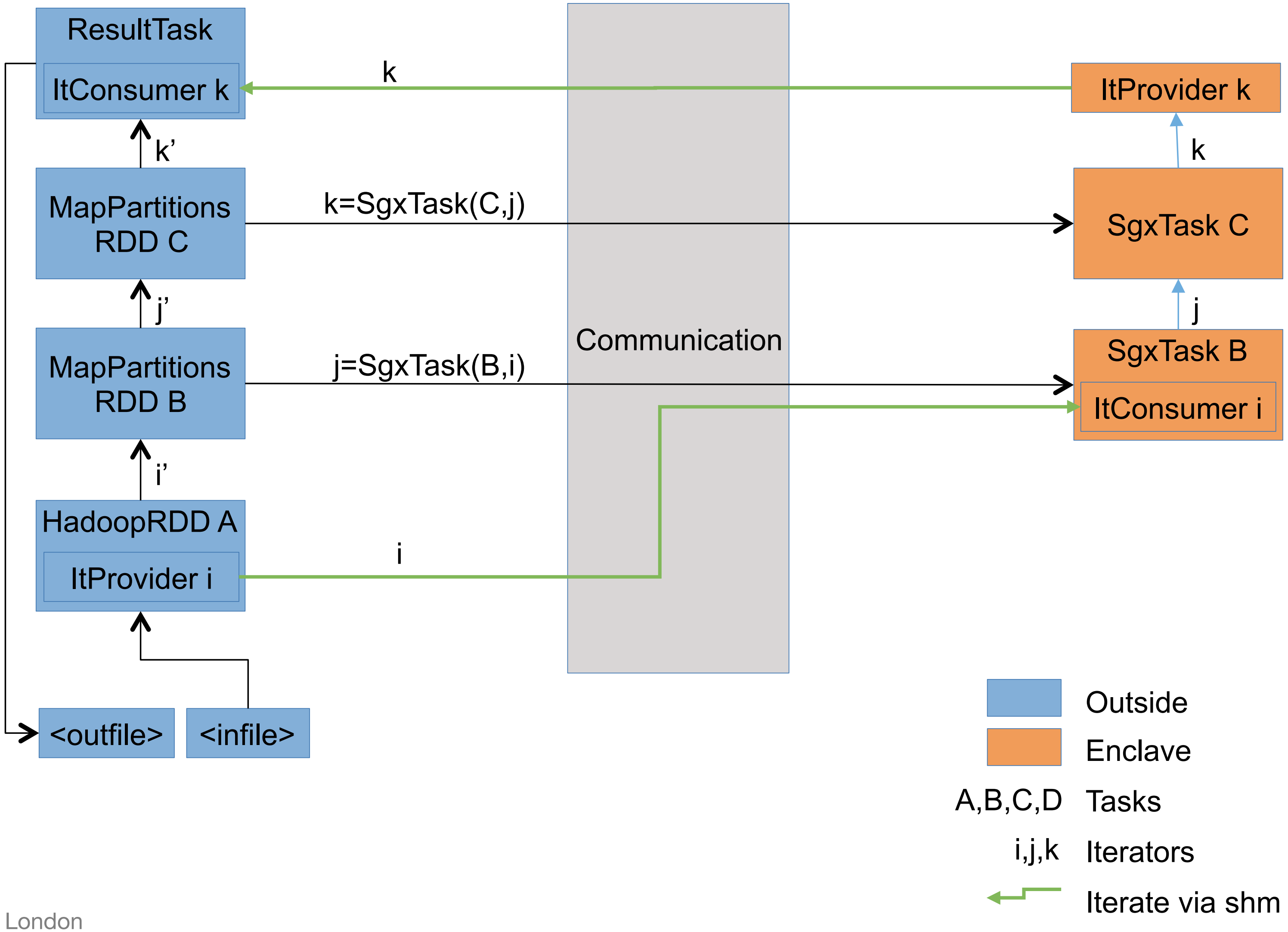
ResultTask Output results	

Partitioning Spark



🟦	Outside
🟧	Enclave
A,B,C,D	Tasks
i,j,k	Iterators
←	Iterate via shm

Partitioning Spark



2. Alternatives (as well as)

- MPC
  - Hard to comprehend
  - High latency
- Homomorphic Cryptography
  - Easier to understand
  - Lower throughput
- Differential Privacy or K-Anon or Fuzzing
  - Easier still to understand
  - Limited number of duty cycles

Hub-of-All-Things

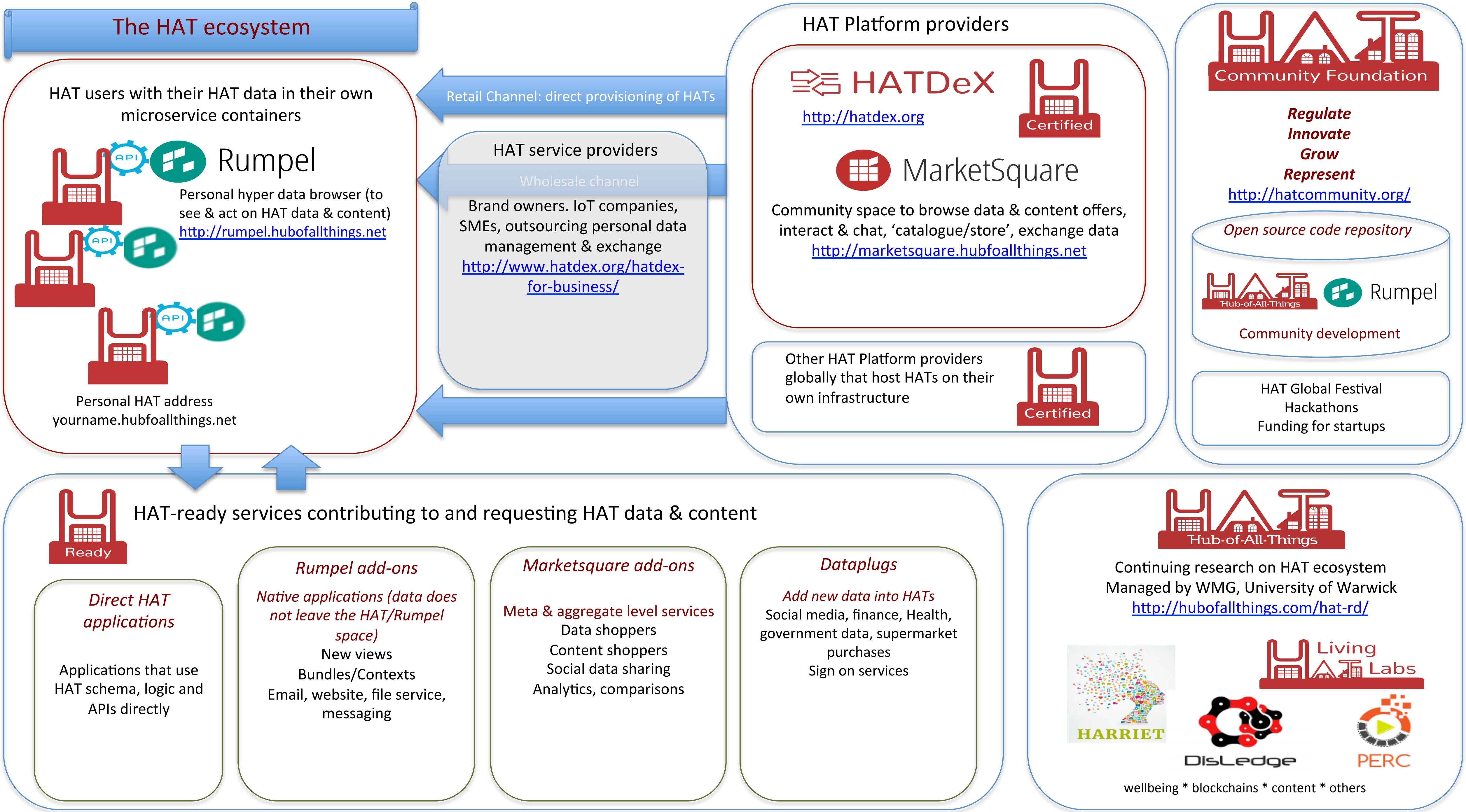
3. The HAT ecosystem – Not a Cloud == Databox in the Wild

To evolve and emerge a digital service ecosystem of organisations and people where individuals are able to acquire, use, control and exchange their own data for their own good and the good of society

Ownership Model:

http://wrap.warwick.ac.uk/108357/

The HAT ecosystem



HAT users with their HAT data in their own microservice containers

Rumpel

Personal hyper data browser (to see & act on HAT data & content)
http://rumpel.hubofallthings.net

Personal HAT address
yourname.hubfoallthings.net

Retail Channel: direct provisioning of HATs

HAT service providers

Wholesale channel

Brand owners. IoT companies, SMEs, outsourcing personal data management & exchange
http://www.hatdex.org/hatdex-for-business/

HAT Platform providers

HATDeX
http://hatdex.org

MarketSquare

Community space to browse data & content offers, interact & chat, 'catalogue/store', exchange data
http://marketsquare.hubfoallthings.net

Other HAT Platform providers globally that host HATs on their own infrastructure

HAT Community Foundation

Regulate
Innovate
Grow
Represent
http://hatcommunity.org/

Open source code repository

Community development

HAT Global Festival
Hackathons
Funding for startups

HAT-ready services contributing to and requesting HAT data & content

Direct HAT applications

Applications that use HAT schema, logic and APIs directly

Rumpel add-ons

Native applications (data does not leave the HAT/Rumpel space)
New views
Bundles/Contexts
Email, website, file service, messaging

Marketsquare add-ons

Meta & aggregate level services
Data shoppers
Content shoppers
Social data sharing
Analytics, comparisons

Dataplugs

Add new data into HATs
Social media, finance, Health, government data, supermarket purchases
Sign on services

HAT Hub-of-All-Things

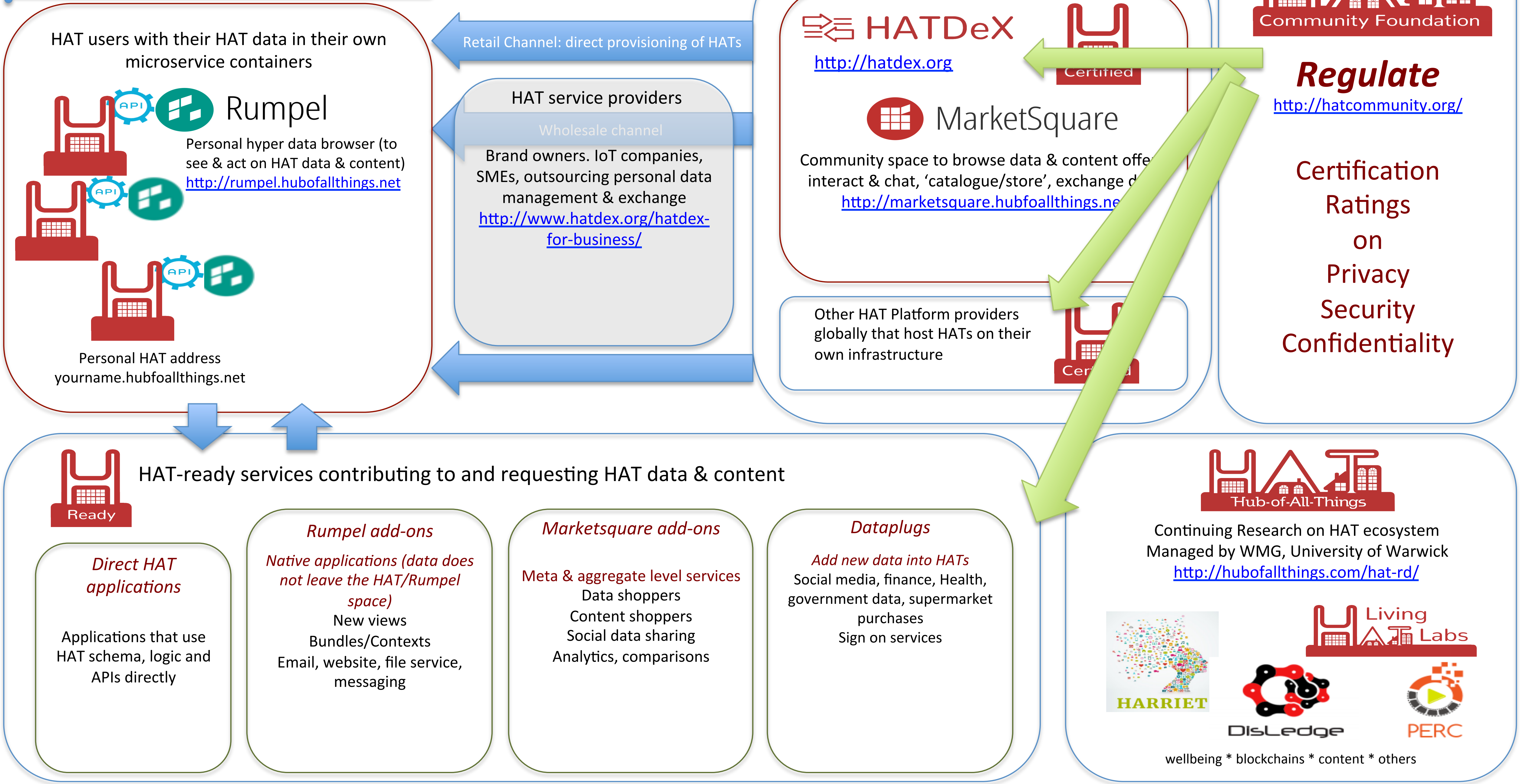
Continuing research on HAT ecosystem
Managed by WMG, University of Warwick
http://hubofallthings.com/hat-rd/

wellbeing * blockchains * content * others

HAT Community Foundation

A members organisation (charity) startup for the Regulation, Innovation, Growth & Representation in the HAT personal data ecosystem

The HAT ecosystem



HAT users with their HAT data in their own microservice containers

Rumpel

Personal hyper data browser (to see & act on HAT data & content)
http://rumpel.hubofallthings.net

Personal HAT address
yourname.hubfoallthings.net

Retail Channel: direct provisioning of HATs

HAT service providers

Wholesale channel

Brand owners. IoT companies, SMEs, outsourcing personal data management & exchange
http://www.hatdex.org/hatdex-for-business/

HAT Platform providers

HATDeX
http://hatdex.org

MarketSquare

Community space to browse data & content offe interact & chat, 'catalogue/store', exchange d
http://marketsquare.hubfoallthings.ne

Other HAT Platform providers globally that host HATs on their own infrastructure

Community Foundation

Regulate

http://hatcommunity.org/

Certification
Ratings
on
Privacy
Security
Confidentiality

HAT-ready services contributing to and requesting HAT data & content

Direct HAT applications

Applications that use HAT schema, logic and APIs directly

Rumpel add-ons

Native applications (data does not leave the HAT/Rumpel space)
New views
Bundles/Contexts
Email, website, file service, messaging

Marketsquare add-ons

Meta & aggregate level services
Data shoppers
Content shoppers
Social data sharing
Analytics, comparisons

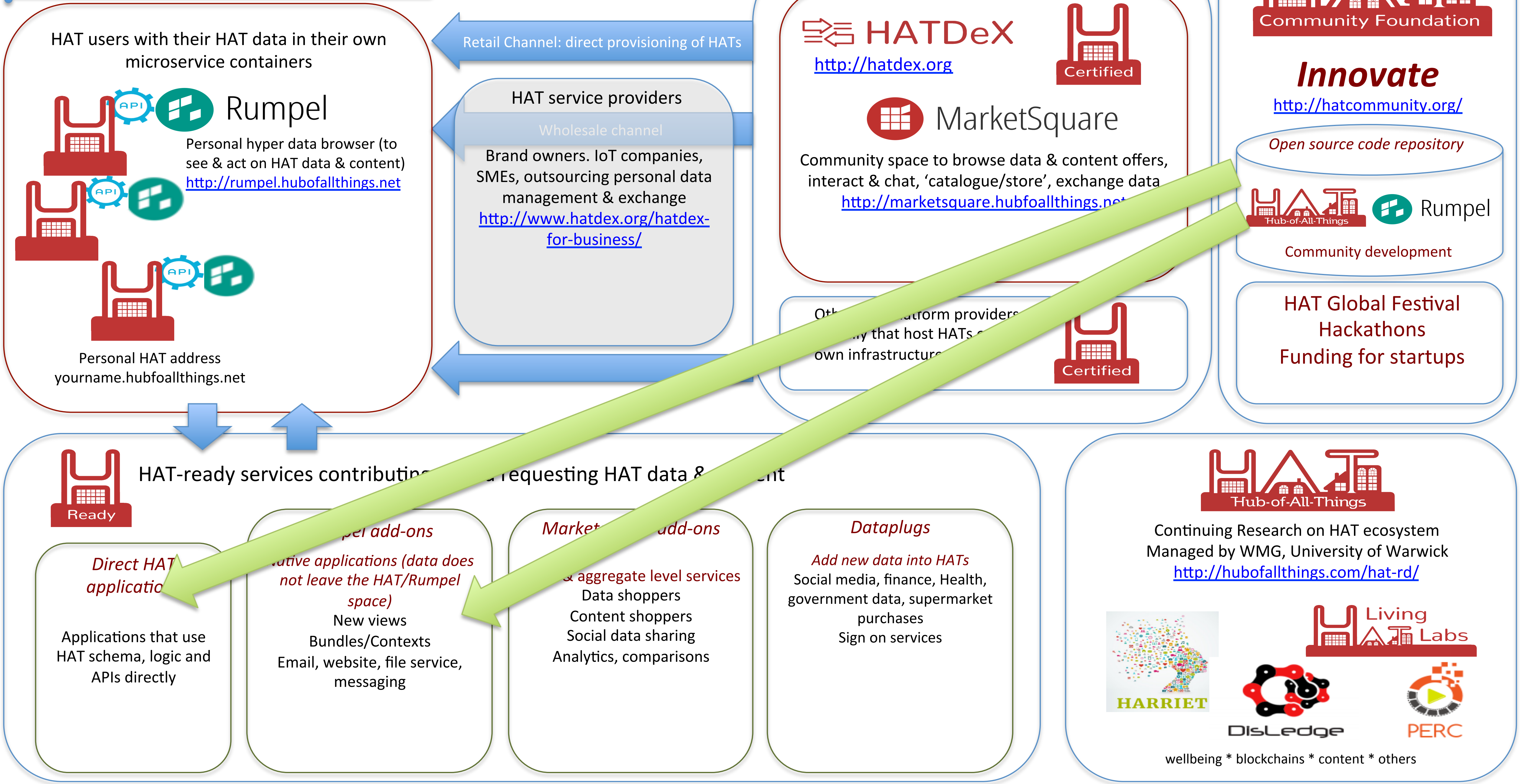
Dataplugs

Add new data into HATs
Social media, finance, Health, government data, supermarket purchases
Sign on services

Continuing Research on HAT ecosystem
Managed by WMG, University of Warwick
http://hubofallthings.com/hat-rd/

wellbeing * blockchains * content * others

The HAT ecosystem



HAT users with their HAT data in their own microservice containers

Rumpel

Personal hyper data browser (to see & act on HAT data & content)
http://rumpel.hubofallthings.net

Personal HAT address
yourname.hubfoallthings.net

Retail Channel: direct provisioning of HATs

HAT service providers

Wholesale channel

Brand owners. IoT companies, SMEs, outsourcing personal data management & exchange
http://www.hatdex.org/hatdex-for-business/

HAT Platform providers

HATDeX
http://hatdex.org

Certified

MarketSquare

Community space to browse data & content offers, interact & chat, 'catalogue/store', exchange data
http://marketsquare.hubfoallthings.ne

Oth... ...atform providers
...y that host HATs ...
own infrastructur...

Certified

Community Foundation

Innovate

http://hatcommunity.org/

Open source code repository

Hub-of-All-Things

Rumpel

Community development

HAT Global Festival
Hackathons
Funding for startups

Ready

HAT-ready services contributin... ...requesting HAT data &... ...nt

Direct HAT applicatio...

Applications that use HAT schema, logic and APIs directly

...pel add-ons

...ative applications (data does not leave the HAT/Rumpel space)

New views
Bundles/Contexts
Email, website, file service, messaging

Market... ...add-ons

...& aggregate level services

Data shoppers
Content shoppers
Social data sharing
Analytics, comparisons

Dataplugs

Add new data into HATs

Social media, finance, Health, government data, supermarket purchases
Sign on services

Hub-of-All-Things

Continuing Research on HAT ecosystem
Managed by WMG, University of Warwick
http://hubofallthings.com/hat-rd/

Living HAT Labs

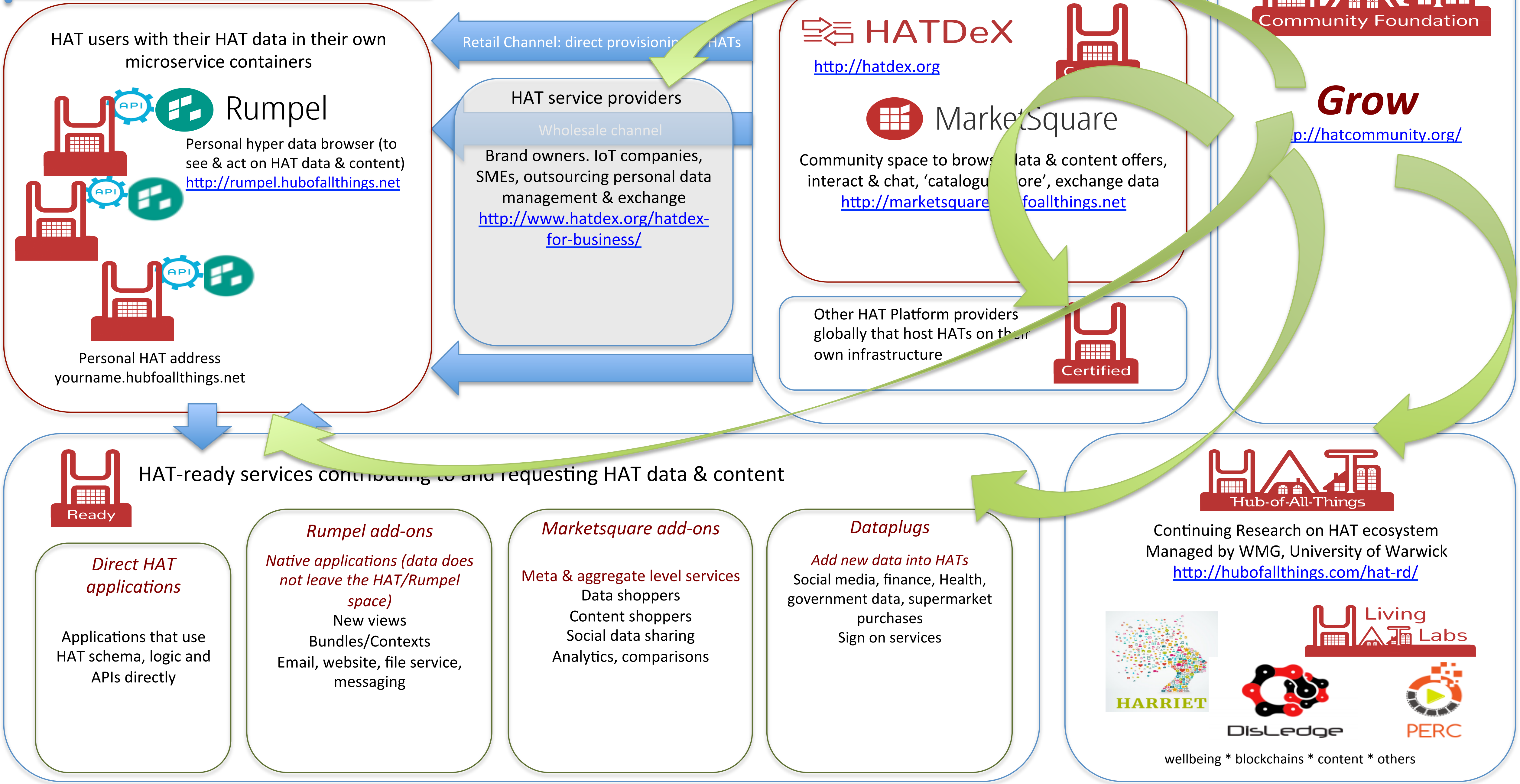
HARRIET

DisLedge

PERC

wellbeing * blockchains * content * others

The HAT ecosystem



HAT users with their HAT data in their own microservice containers

Rumpel

Personal hyper data browser (to see & act on HAT data & content)
http://rumpel.hubofallthings.net

Personal HAT address
yourname.hubfoallthings.net

Retail Channel: direct provisionin HATs

HAT service providers

Wholesale channel

Brand owners. IoT companies, SMEs, outsourcing personal data management & exchange
http://www.hatdex.org/hatdex-for-business/

HAT Platform providers

HATDeX
http://hatdex.org

MarketSquare

Community space to brows lata & content offers, interact & chat, 'catalogu ore', exchange data
http://marketsquare foallthings.net

Other HAT Platform providers globally that host HATs on their own infrastructure

Certified

Community Foundation

Grow
p://hatcommunity.org/

Ready

HAT-ready services contributing to and requesting HAT data & content

Direct HAT applications

Applications that use HAT schema, logic and APIs directly

Rumpel add-ons

Native applications (data does not leave the HAT/Rumpel space)
New views
Bundles/Contexts
Email, website, file service, messaging

Marketsquare add-ons

Meta & aggregate level services
Data shoppers
Content shoppers
Social data sharing
Analytics, comparisons

Dataplugs

Add new data into HATs
Social media, finance, Health, government data, supermarket purchases
Sign on services

Hub-of-All-Things

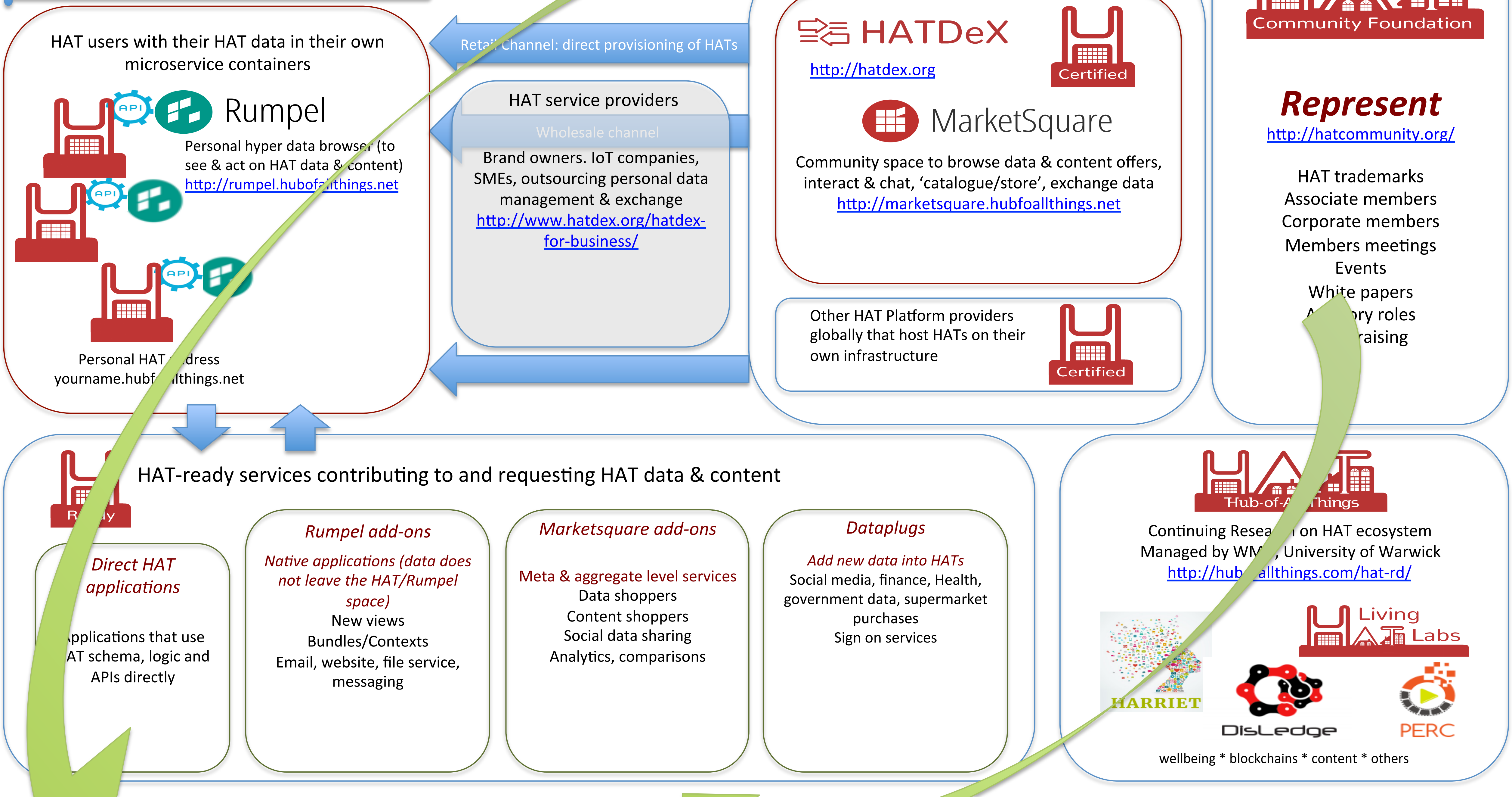
Continuing Research on HAT ecosystem
Managed by WMG, University of Warwick
http://hubofallthings.com/hat-rd/

Living HAT Labs

HARRIET
DisLedge
PERC

wellbeing * blockchains * content * others

The HAT ecosystem



HAT users with their HAT data in their own microservice containers

Rumpel

Personal hyper data browser (to see & act on HAT data & content)
http://rumpel.hubofallthings.net

Personal HAT address
yourname.hubofallthings.net

Retail Channel: direct provisioning of HATs

HAT service providers

Wholesale channel

Brand owners. IoT companies, SMEs, outsourcing personal data management & exchange
http://www.hatdex.org/hatdex-for-business/

HAT Platform providers

HATDeX
http://hatdex.org

Certified

MarketSquare

Community space to browse data & content offers, interact & chat, 'catalogue/store', exchange data
http://marketsquare.hubfoallthings.net

Other HAT Platform providers globally that host HATs on their own infrastructure

Certified

Community Foundation

Represent
http://hatcommunity.org/

HAT trademarks
Associate members
Corporate members
Members meetings
Events
White papers
Advisory roles
Fundraising

HAT-ready services contributing to and requesting HAT data & content

Relay

Direct HAT applications

Applications that use HAT schema, logic and APIs directly

Rumpel add-ons

Native applications (data does not leave the HAT/Rumpel space)
New views
Bundles/Contexts
Email, website, file service, messaging

Marketsquare add-ons

Meta & aggregate level services
Data shoppers
Content shoppers
Social data sharing
Analytics, comparisons

Dataplugs

Add new data into HATs
Social media, finance, Health, government data, supermarket purchases
Sign on services

Hub-of-All-Things

Continuing Research on HAT ecosystem
Managed by WMG, University of Warwick
http://hubofallthings.com/hat-rd/

Living HAT Labs

HARRIET

DisLedge

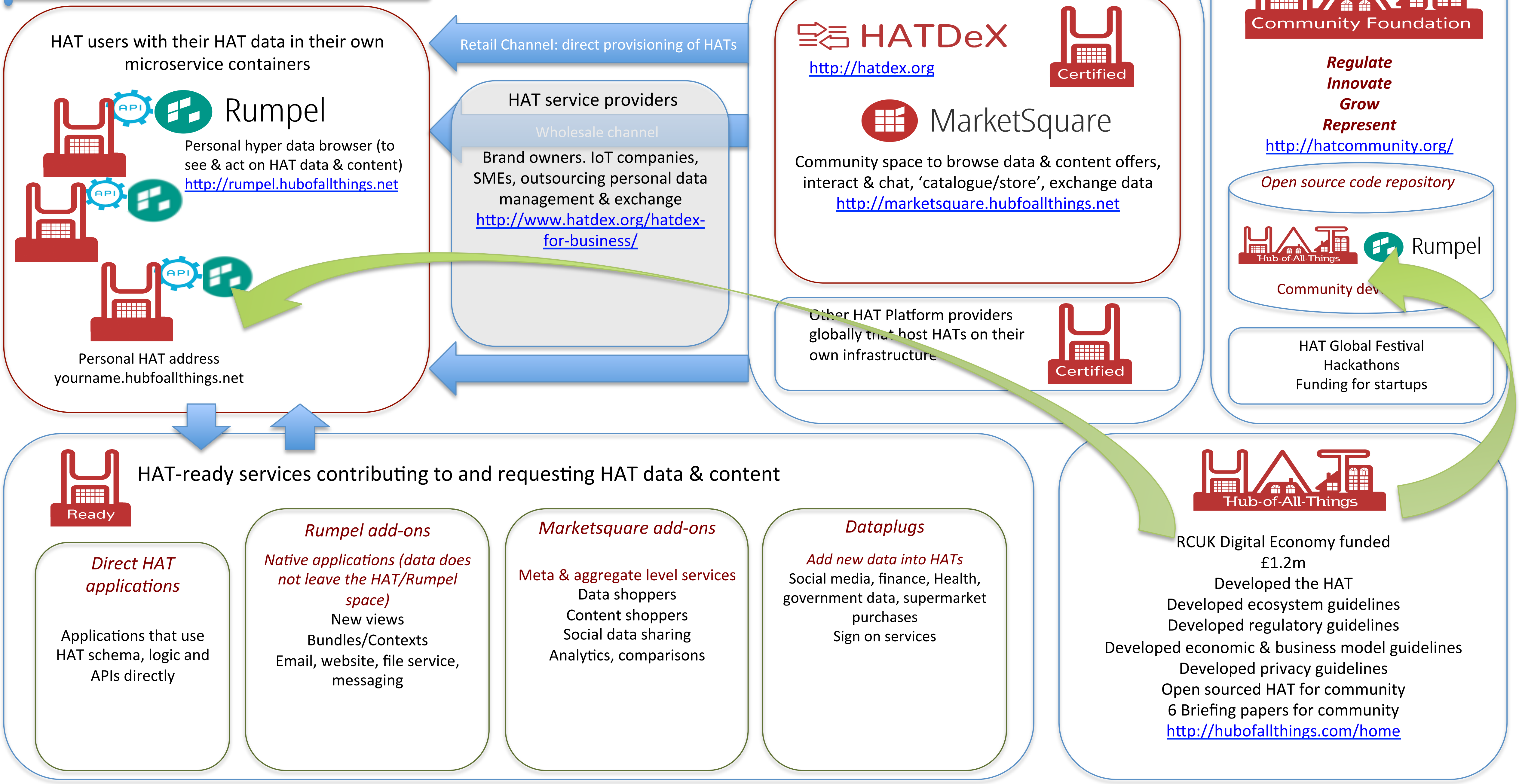
PERC

wellbeing * blockchains * content * others

HAT Research

Increasing knowledge in the HAT personal data ecosystem

The HAT ecosystem



HAT users with their HAT data in their own microservice containers

Rumpel

Personal hyper data browser (to see & act on HAT data & content)
http://rumpel.hubofallthings.net

Personal HAT address
yourname.hubfoallthings.net

Retail Channel: direct provisioning of HATs

HAT service providers

Wholesale channel

Brand owners. IoT companies, SMEs, outsourcing personal data management & exchange
http://www.hatdex.org/hatdex-for-business/

HAT Platform providers

HATDeX
http://hatdex.org

Certified

MarketSquare

Community space to browse data & content offers, interact & chat, 'catalogue/store', exchange data
http://marketsquare.hubfoallthings.net

Other HAT Platform providers globally that host HATs on their own infrastructure

Certified

HAT Community Foundation

Regulate
Innovate
Grow
Represent
http://hatcommunity.org/

Open source code repository

HAT Hub-of-All-Things

Rumpel

Community dev

HAT Global Festival
Hackathons
Funding for startups

Ready

HAT-ready services contributing to and requesting HAT data & content

Direct HAT applications

Applications that use HAT schema, logic and APIs directly

Rumpel add-ons

Native applications (data does not leave the HAT/Rumpel space)
New views
Bundles/Contexts
Email, website, file service, messaging

Marketsquare add-ons

Meta & aggregate level services
Data shoppers
Content shoppers
Social data sharing
Analytics, comparisons

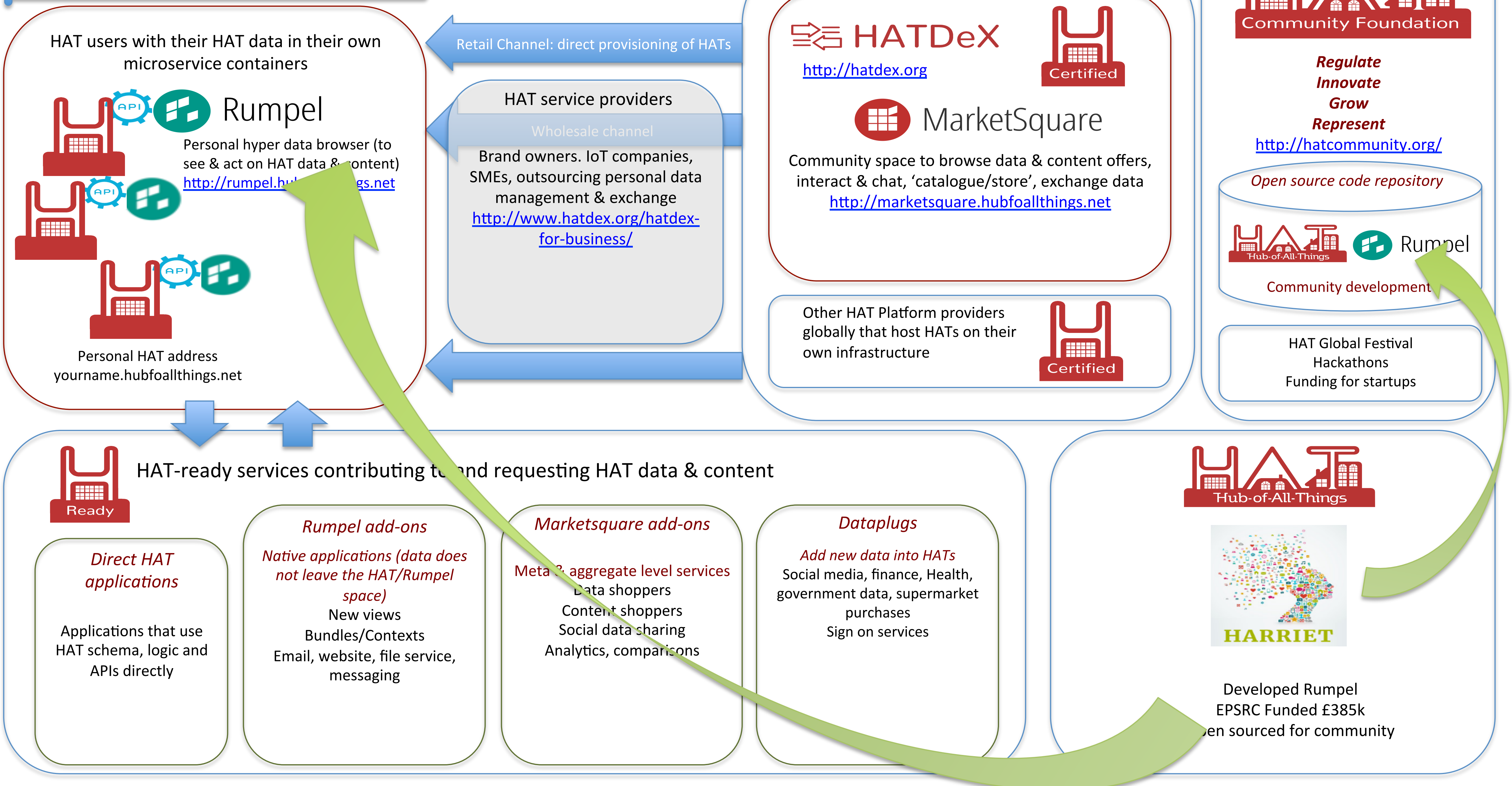
Dataplugs

Add new data into HATs
Social media, finance, Health, government data, supermarket purchases
Sign on services

HAT Hub-of-All-Things

RCUK Digital Economy funded
£1.2m
Developed the HAT
Developed ecosystem guidelines
Developed regulatory guidelines
Developed economic & business model guidelines
Developed privacy guidelines
Open sourced HAT for community
6 Briefing papers for community
http://hubofallthings.com/home

The HAT ecosystem



HAT users with their HAT data in their own microservice containers

Rumpel

Personal hyper data browser (to see & act on HAT data & content)

http://rumpel.hu…gs.net

Personal HAT address
yourname.hubfoallthings.net

Retail Channel: direct provisioning of HATs

HAT service providers

Wholesale channel

Brand owners. IoT companies, SMEs, outsourcing personal data management & exchange
http://www.hatdex.org/hatdex-for-business/

HAT Platform providers

HATDeX
http://hatdex.org

Certified

MarketSquare

Community space to browse data & content offers, interact & chat, 'catalogue/store', exchange data
http://marketsquare.hubfoallthings.net

Other HAT Platform providers globally that host HATs on their own infrastructure

Certified

HAT Community Foundation

Regulate
Innovate
Grow
Represent
http://hatcommunity.org/

Open source code repository

Hub-of-All-Things Rumpel

Community development

HAT Global Festival
Hackathons
Funding for startups

Ready

HAT-ready services contributing to and requesting HAT data & content

Direct HAT applications

Applications that use HAT schema, logic and APIs directly

Rumpel add-ons

Native applications (data does not leave the HAT/Rumpel space)
New views
Bundles/Contexts
Email, website, file service, messaging

Marketsquare add-ons

Meta & aggregate level services
Data shoppers
Content shoppers
Social data sharing
Analytics, comparisons

Dataplugs

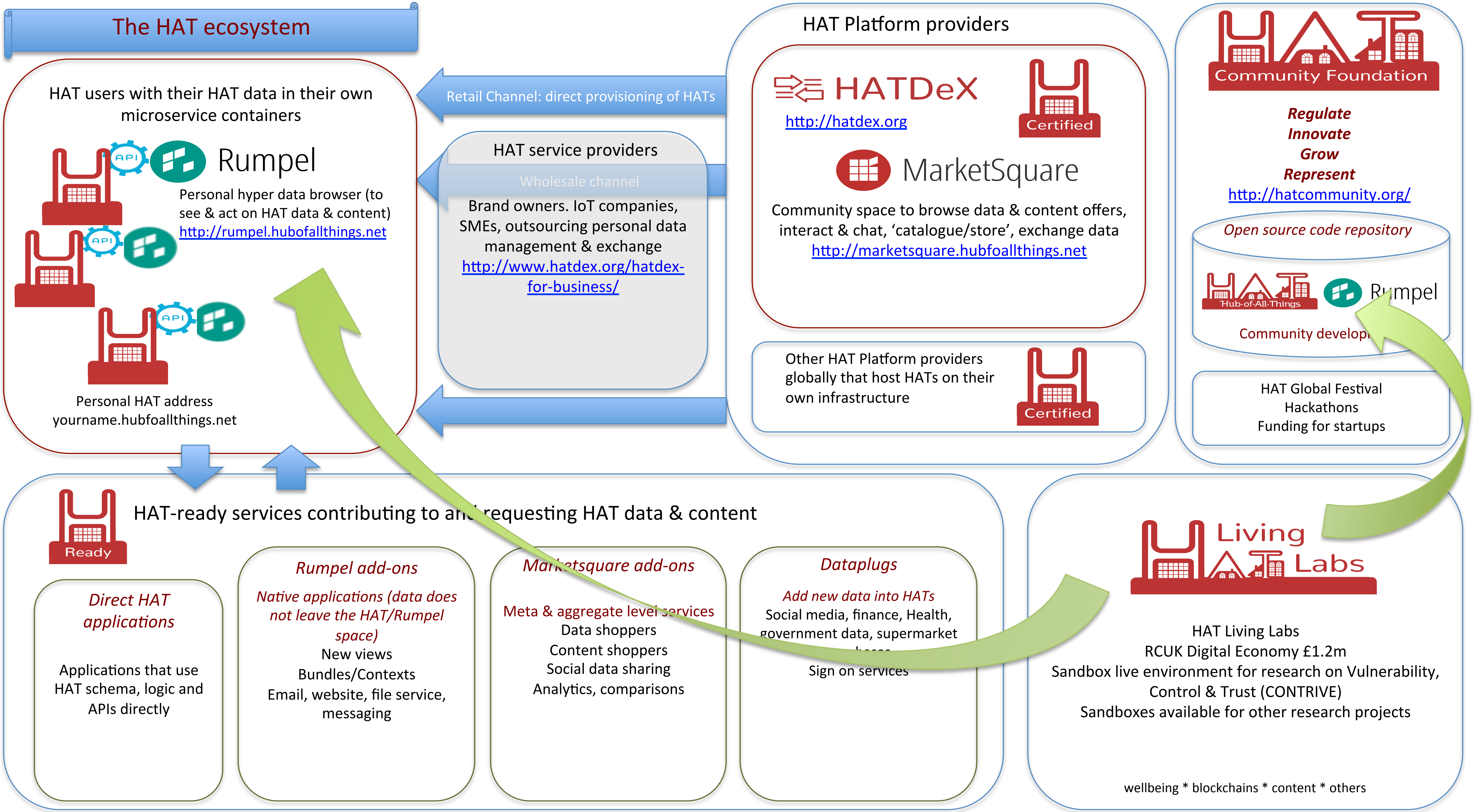
Add new data into HATs
Social media, finance, Health, government data, supermarket purchases
Sign on services

HAT Hub-of-All-Things

HARRIET

Developed Rumpel
EPSRC Funded £385k
…en sourced for community

The HAT ecosystem



HAT users with their HAT data in their own microservice containers

Rumpel

Personal hyper data browser (to see & act on HAT data & content)
http://rumpel.hubofallthings.net

Personal HAT address
yourname.hubfoallthings.net

Retail Channel: direct provisioning of HATs

HAT service providers

Wholesale channel

Brand owners. IoT companies, SMEs, outsourcing personal data management & exchange
http://www.hatdex.org/hatdex-for-business/

HAT Platform providers

HATDeX
http://hatdex.org

MarketSquare

Community space to browse data & content offers, interact & chat, 'catalogue/store', exchange data
http://marketsquare.hubfoallthings.net

Other HAT Platform providers globally that host HATs on their own infrastructure

HAT Community Foundation

Regulate
Innovate
Grow
Represent
http://hatcommunity.org/

Open source code repository

Community develop

HAT Global Festival
Hackathons
Funding for startups

HAT-ready services contributing to and requesting HAT data & content

Direct HAT applications

Applications that use HAT schema, logic and APIs directly

Rumpel add-ons

Native applications (data does not leave the HAT/Rumpel space)
New views
Bundles/Contexts
Email, website, file service, messaging

Marketsquare add-ons

Meta & aggregate level services
Data shoppers
Content shoppers
Social data sharing
Analytics, comparisons

Dataplugs

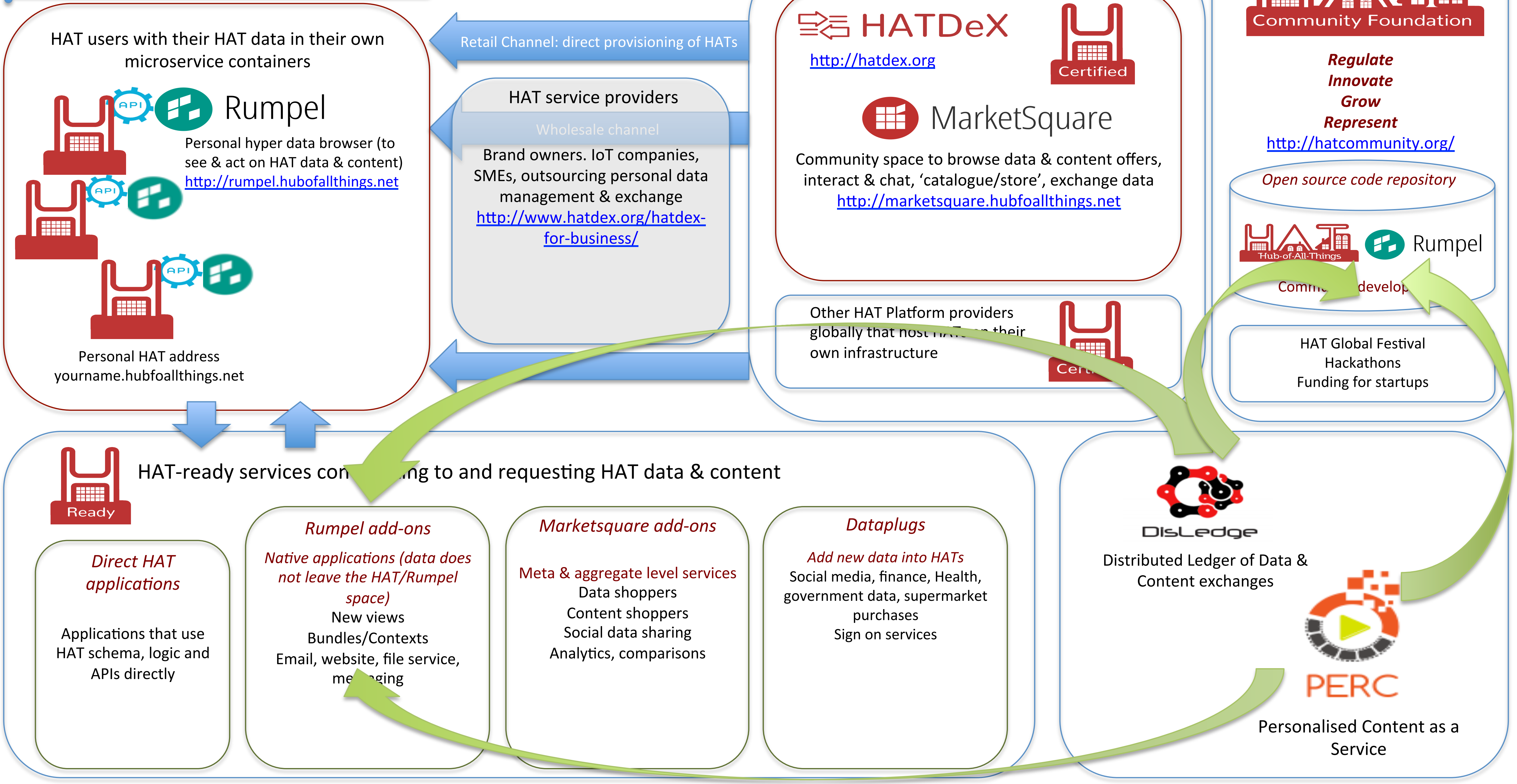
Add new data into HATs
Social media, finance, Health, government data, supermarket purchases
Sign on services

HAT Living Labs

HAT Living Labs
RCUK Digital Economy £1.2m
Sandbox live environment for research on Vulnerability, Control & Trust (CONTRIVE)
Sandboxes available for other research projects

wellbeing * blockchains * content * others

The HAT ecosystem



HAT users with their HAT data in their own microservice containers

Rumpel

Personal hyper data browser (to see & act on HAT data & content)
http://rumpel.hubofallthings.net

Personal HAT address
yourname.hubfoallthings.net

Retail Channel: direct provisioning of HATs

HAT service providers

Wholesale channel

Brand owners. IoT companies, SMEs, outsourcing personal data management & exchange
http://www.hatdex.org/hatdex-for-business/

HAT Platform providers

HATDeX
http://hatdex.org

Certified

MarketSquare

Community space to browse data & content offers, interact & chat, 'catalogue/store', exchange data
http://marketsquare.hubfoallthings.net

Other HAT Platform providers
globally that host HAT... on their
own infrastructure

Cer...

HAT Community Foundation

Regulate
Innovate
Grow
Represent
http://hatcommunity.org/

Open source code repository

Hub-of-All-Things

Rumpel

Comm... develop...

HAT Global Festival
Hackathons
Funding for startups

HAT-ready services con... ng to and requesting HAT data & content

Ready

Direct HAT applications

Applications that use HAT schema, logic and APIs directly

Rumpel add-ons

Native applications (data does not leave the HAT/Rumpel space)
New views
Bundles/Contexts
Email, website, file service, me... ging

Marketsquare add-ons

Meta & aggregate level services
Data shoppers
Content shoppers
Social data sharing
Analytics, comparisons

Dataplugs

Add new data into HATs
Social media, finance, Health, government data, supermarket purchases
Sign on services

DisLedge

Distributed Ledger of Data & Content exchanges

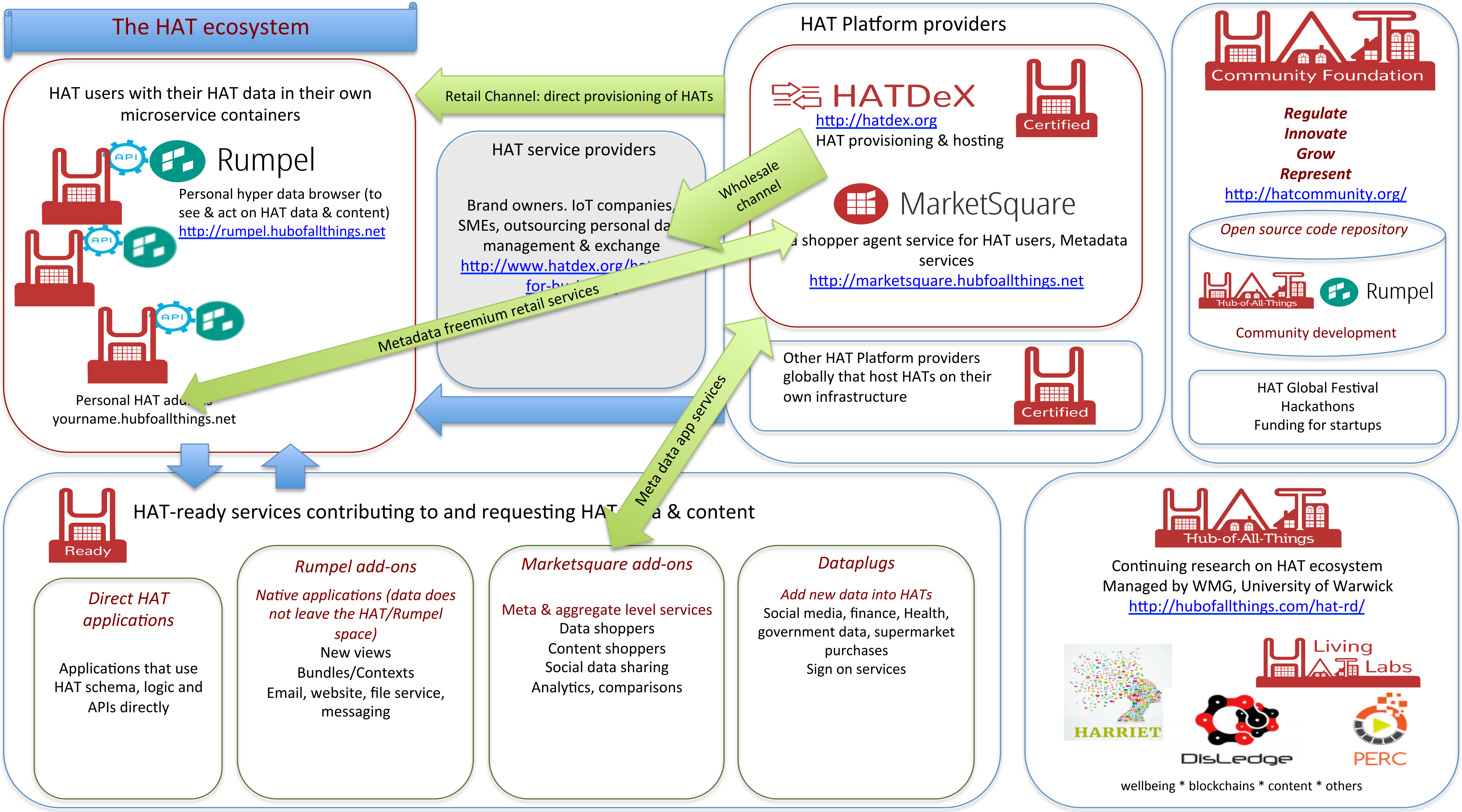
PERC

Personalised Content as a Service

HAT Data Exchange

A commercial startup managing and operating the HAT personal data ecosystem

The HAT ecosystem



HAT users with their HAT data in their own microservice containers

Rumpel

Personal hyper data browser (to see & act on HAT data & content)
http://rumpel.hubofallthings.net

Personal HAT add...s
yourname.hubfoallthings.net

Retail Channel: direct provisioning of HATs

HAT service providers

Brand owners. IoT companies, SMEs, outsourcing personal da... management & exchange
http://www.hatdex.org/h... for-bu...

Wholesale channel

Metadata freemium retail services

Meta data app services

HAT Platform providers

HATDeX
http://hatdex.org
HAT provisioning & hosting

Certified

MarketSquare

...a shopper agent service for HAT users, Metadata services
http://marketsquare.hubfoallthings.net

Other HAT Platform providers globally that host HATs on their own infrastructure

Certified

HAT Community Foundation

Regulate
Innovate
Grow
Represent
http://hatcommunity.org/

Open source code repository

Hub-of-All-Things

Rumpel

Community development

HAT Global Festival
Hackathons
Funding for startups

Ready

HAT-ready services contributing to and requesting HAT ...a & content

Direct HAT applications

Applications that use HAT schema, logic and APIs directly

Rumpel add-ons

Native applications (data does not leave the HAT/Rumpel space)
New views
Bundles/Contexts
Email, website, file service, messaging

Marketsquare add-ons

Meta & aggregate level services
Data shoppers
Content shoppers
Social data sharing
Analytics, comparisons

Dataplugs

Add new data into HATs
Social media, finance, Health, government data, supermarket purchases
Sign on services

HAT Hub-of-All-Things

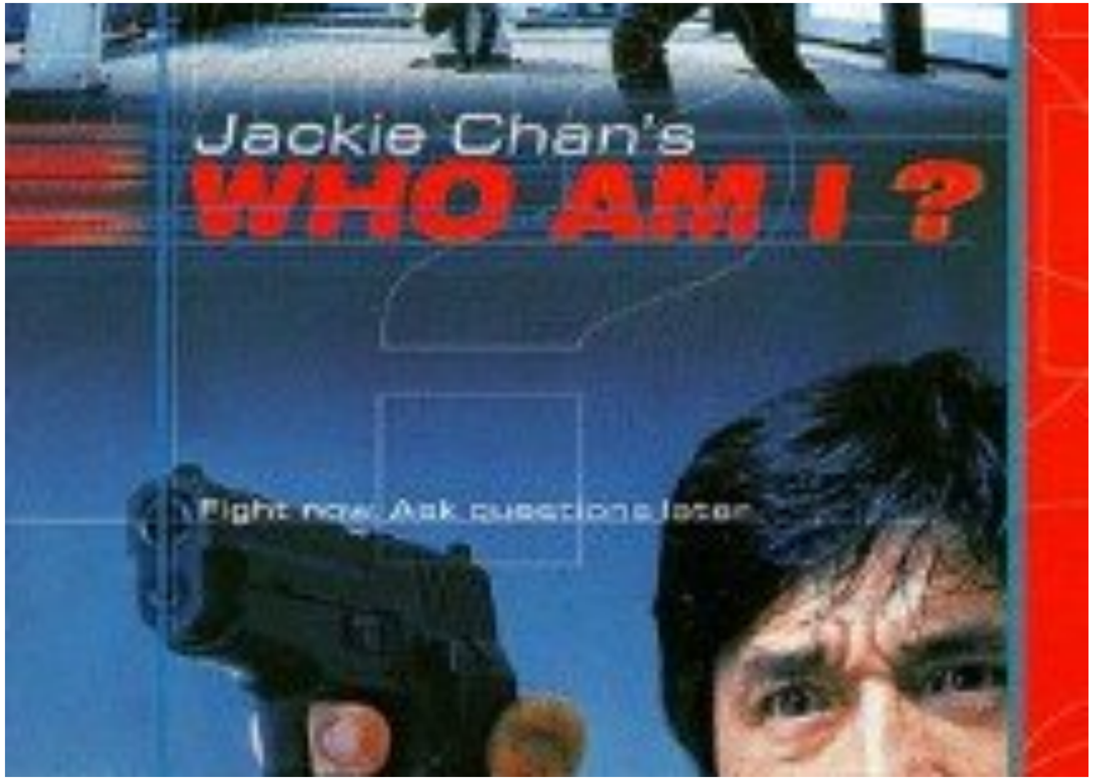
Continuing research on HAT ecosystem
Managed by WMG, University of Warwick
http://hubofallthings.com/hat-rd/

wellbeing * blockchains * content * others

Who Am I? & lets not speculate further ☺

Thanks to EPSRC/databox

- &Liang Wang, et al, Cambridge



Thanks to Turing/Maru

- &Peter Pietzuch, Imperial