



TRUST
Over **IP**
FOUNDATION

Semantics WG Weekly Meeting

24 November 2020

 THE **LINUX** FOUNDATION

Antitrust Policy Notice

- › Linux Foundation meetings involve participation by industry competitors, and it is the intention of the Linux Foundation to conduct all of its activities in accordance with applicable antitrust and competition laws. It is therefore extremely important that attendees adhere to meeting agendas, and be aware of, and not participate in, any activities that are prohibited under applicable US state, federal or foreign antitrust and competition laws.
- › Examples of types of actions that are prohibited at Linux Foundation meetings and in connection with Linux Foundation activities are described in the Linux Foundation Antitrust Policy available at <http://www.linuxfoundation.org/antitrust-policy>. If you have questions about these matters, please contact your company counsel, or if you are a member of the Linux Foundation, feel free to contact Andrew Updegrave of the firm of Gesmer Updegrave LLP, which provides legal counsel to the Linux Foundation.

Membership Advisory

- › For the protection of all Members, active participation in working groups, meetings and events is limited to members, including their employees, of the Trust over IP Foundation who have signed the membership documents (including Trust over IP membership agreement as well as relevant working group charters) and thus agreed to the intellectual property rules governing participation.
- › If you or your employer are not a member, we ask that you not participate in meetings by verbal contribution or otherwise take any action beyond observing.

Agenda

- › 1. Welcome (Paul—2.5 mins)
- › 2. Newcomer Introductions (Paul—2.5 mins)
- › 3. Task Force/Focus Group Updates (WG—5 mins)
- › 4. Industry Sector Classification at ToIP (Paul—10 mins)
- › 5. Identity Correlation Bitmap: An object for mitigating against attribute correlation patterns (Paul—35 mins)
- › 6. Logistics and miscellaneous (Paul—5 mins)
 - › a. News from the Operations Team
 - › b. Leadership positions
 - › c. Meeting schedule

Newcomer Introductions

(30 seconds!)

1. Name
2. Location / time zone
3. Affiliation(s)
4. One-sentence summary of your interest in Semantics (or **one particular semantics-related** issue you personally want to see solved)

Task Force/Focus Group Updates

(5 mins)

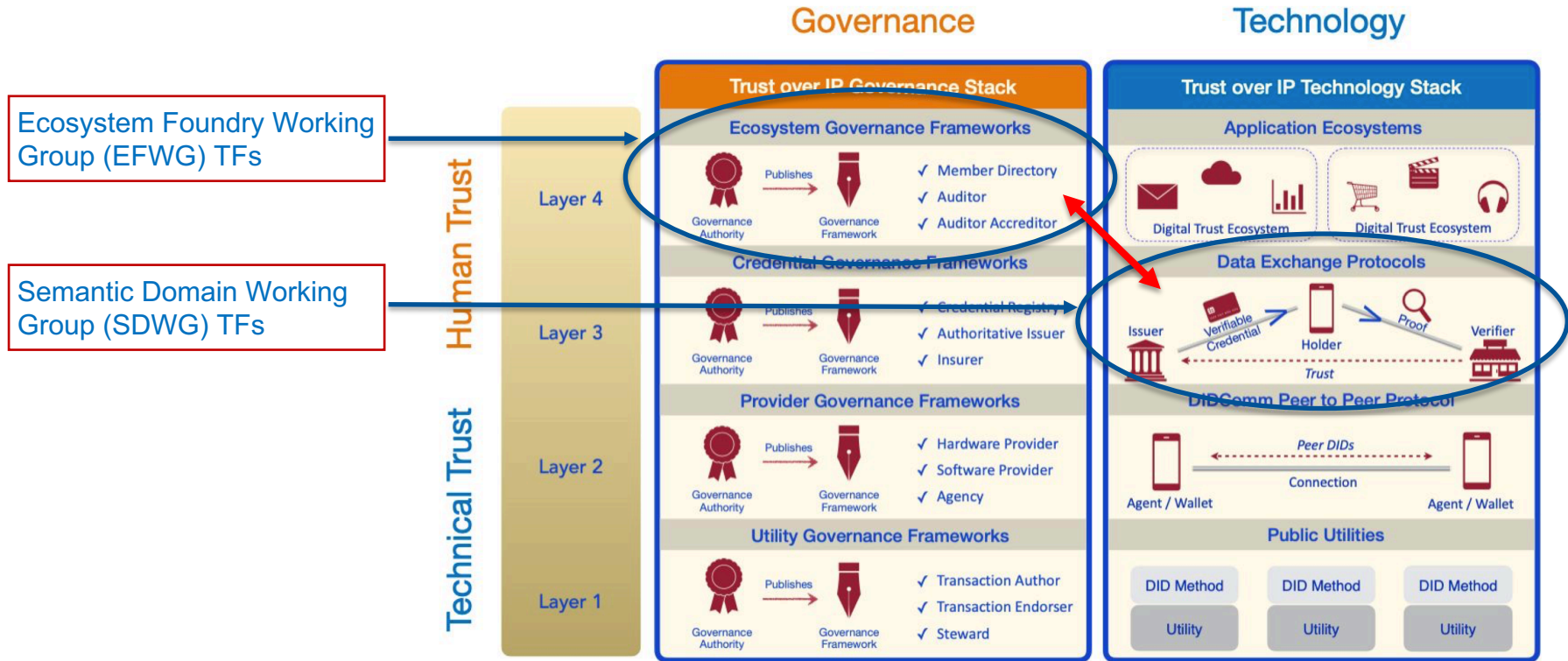
- Imaging TF (Scott)
- Medical Information TF (Scott)
- ✓ OCA-FHIR FG (John/Mukund)
- Notice & Consent TF (Mark)

Industry Sector Classification at ToIP (10 mins)

Update by: P.Knowles

<https://docs.google.com/document/d/1EOSfkcAtMYyT4rlzpq-VWV1mQ90So9U9/edit#>
 THE **LINUX** FOUNDATION

Direct relationship between the stacks



Every Ecosystem Governance Framework defined at Layer 4 of the Governance Stack will have a direct relationship with an associated Data Exchange Protocol at Layer 3 of the Technology Stack

Industry Sector Classification

Option 1

GICS: Global Industry Classification Standard

"The Industry Standard"



The **GICS** indices is an industry taxonomy for use by the global financial community as a basis to assign companies to a sub-industry, and to an industry, industry group, and sector, by its principal business activity.

- 11 Sectors



Ground-zero

- 24 Industry Groups

- 69 Industries

- 158 Sub-Industries

Industry Sector Classification

GICS: Global Industry Classification Standard

GICS codes were developed by MSCI, a leading provider of research-based, investment decision support tools for investors globally and Standard & Poor's, an American financial services company.

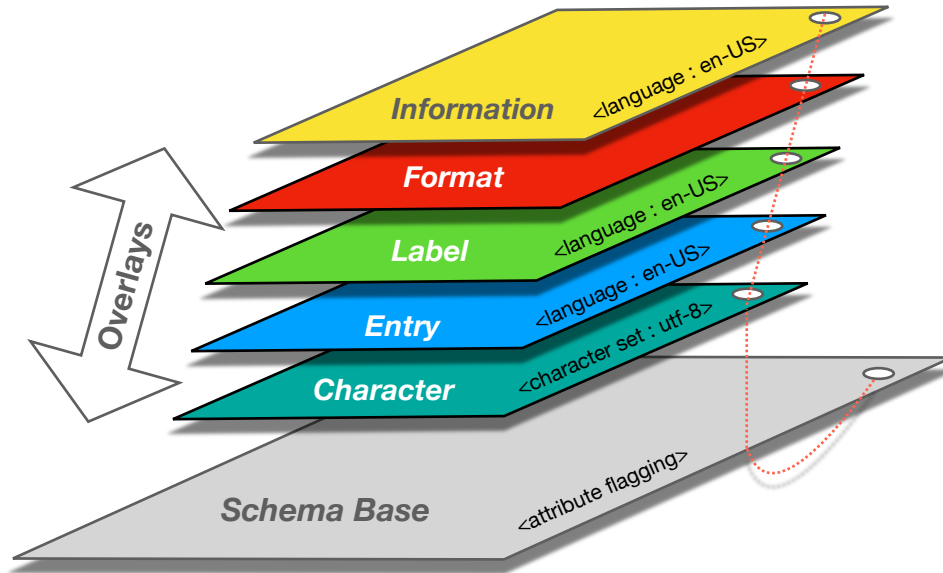
The aim of GICS is to enhance investment research and asset management processes for financial professionals worldwide. The methodology used is now widely accepted in the financial and investment community and has led to efficiencies and transparencies throughout investment processes.

- 11 Sectors (2-digits)
- 24 Industry Groups (4-digits)
- 69 Industries (6-digits)
- 158 Sub-Industries (8-digits)

Industry Sector Classification

“Demographics” schema example

“classification”: “GICS:35202010”



GICS = 35202010

Sector code:

35 - Health Care

Industry group code:

3520 - Pharmaceuticals, Biotechnology & Life Sciences

Industry code:

352020 - Pharmaceuticals

Sub-industry code:

35202010 - Pharmaceuticals

Description:

Companies engaged in the research, development or production of pharmaceuticals. Includes veterinary drugs.

Industry Sector Classification

- Schema base “classification” meta attribute

“Demographics” schema example

```
{
  "@context": "https://odca.tech/v1",
  "name": "Demographics-DM",
  "type": "spec/schema_base/1.0",
  "description": "The DM (Demographics) domain includes a set of essential standard variables that",
  "classification": "GICS:35202010",
  "daisy_chain": "",
  "issued_by": "",
  "attributes": {
    "STUDYID": "Text",
    "DOMAIN": "Text",
    "USUBJID": "Text",
    "AGE": "Number",
    "SEX": "Text",
    "RACE": "Text"
  },
  "attr_blinding": [
    "STUDYID",
    "USUBJID"
  ]
}
```

GICS = 35202010

Sector code:
35 - Health Care

Industry group code:
3520 - Pharmaceuticals, Biotechnology & Life Sciences

Industry code:
352020 - Pharmaceuticals

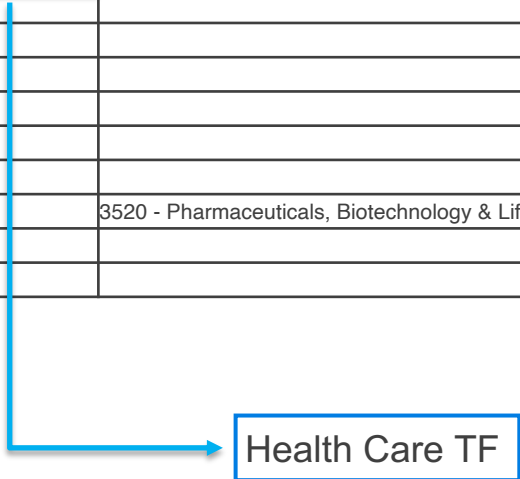
Sub-industry code:
35202010 - Pharmaceuticals

Description:
Companies engaged in the research, development or production of pharmaceuticals. Includes veterinary drugs.

Vertical focus

Global Industry Classification Standard

Sector	Industry Group	Industry	Sub-Industry
35 - Health Care	3510 - Health Care Equipment & Services	351010 - Health Care Equipment & Supplies	35101010 - Health Care Equipment
			35101020 - Health Care Supplies
		351020 - Health Care Providers & Services	35102010 - Health Care Distributors
			35102015 - Health Care Services
			35102020 - Health Care Facilities
			35102030 - Managed Health Care
		351030 - Health Care Technology	35103010 - Health Care Technology
	3520 - Pharmaceuticals, Biotechnology & Life Sciences	352010 - Biotechnology	35201010 - Biotechnology
		352020 - Pharmaceuticals	35202010 - Pharmaceuticals
		352030 - Life Sciences Tools & Services	35203010 - Life Sciences Tools & Services



Vertical focus

- Ecosystem Foundry WG TF: **Patient Identity**

Sector	Industry Group	Industry	Sub-Industry
35 - Health Care	3510 - Health Care Equipment & Services	351010 - Health Care Equipment & Supplies	35101010 - Health Care Equipment
			35101020 - Health Care Supplies
		351020 - Health Care Providers & Services	35102010 - Health Care Distributors
			35102015 - Health Care Services
			35102020 - Health Care Facilities
			35102030 - Managed Health Care
	3520 - Pharmaceuticals, Biotechnology & Life Sciences	351030 - Health Care Technology	35103010 - Health Care Technology
			352010 - Biotechnology
		352020 - Pharmaceuticals	35201010 - Biotechnology
			35202010 - Pharmaceuticals
	352030 - Life Sciences Tools & Services	35203010 - Life Sciences Tools & Services	

GICS: Health Care (35)
SIC: Health Services (80)

Vertical focus

- Ecosystem Foundry WG TF: **Human Trafficking Ecosystem**

Major Group	Industry	Extended SIC 6-Digit
83 - Social Services	8399 - Social Services, Not Elsewhere Classified	839901 - Drug Abuse & Addiction Info & Treatment
		839902 - Alcoholism Information & Treatment Ctrs
		839903 - Abortion Alternatives Organizations
		839904 - Child Abuse Information & Treatment Ctrs
		839905 - Disability Services
		839906 - Gambling Abuse/addiction Info/treatment
		839907 - Fund Raising Counselors & Organizations
		839908 - Human Services Organizations
		839909 - Handicapped Services & Organizations
		839910 - Smokers Information & Treatment Centers
		839911 - Medical Management Service
		839912 - Suicide Prevention Service
		839913 - Indian Reservations & Tribes
		839914 - Community Action Agencies
		839915 - Gay & Lesbian Organizations
		839916 - Breastfeeding Supplies & Information
		839917 - Crime Prevention Programs
		839918 - Volunteer Workers Placement Service
		839919 - Charitable Institutions
		839921 - Addiction Treatment Centers
		839922 - Background Screening
		839924 - Dependency Information & Help Centres
		839925 - Memorial Societies
		839929 - Epilepsy Educational Referral/sprt Services
		839930 - Tax Advocacy
		839998 - Non-Profit Organizations
		839999 - Social Services Nec

GICS: -
SIC: Social Services (83)

Vertical focus

- Ecosystem Foundry WG TF: [Enterprise – Identity and Access Management](#)

Sector	Industry Group	Industry	Sub-Industry
45 - Information Technology	4510 - Software & Services	451020 - IT Services	45102010 - IT Consulting & Other Services
			45102020 - Data Processing & Outsourced Services
			45102030 - Internet Services & Infrastructure
		451030 - Software	45103010 - Application Software
			45103020 - Systems Software

GICS: Information Technology (45)
SIC: Information Technology Services (737109)

Vertical focus

- Ecosystem Foundry WG TF: COVID-19 Credentials Initiative Governance Framework

Sector	Industry Group	Industry	Sub-Industry	
35 - Health Care	3510 - Health Care Equipment & Services	351010 - Health Care Equipment & Supplies	35101010 - Health Care Equipment	
			35101020 - Health Care Supplies	
		351020 - Health Care Providers & Services	35102010 - Health Care Distributors	
			35102015 - Health Care Services	
			35102020 - Health Care Facilities	
	3520 - Pharmaceuticals, Biotechnology & Life Sciences	351030 - Health Care Technology	35102030 - Managed Health Care	
			35103010 - Health Care Technology	
		352010 - Biotechnology	35201010 - Biotechnology	
			352020 - Pharmaceuticals	35202010 - Pharmaceuticals
				352030 - Life Sciences Tools & Services

GICS: Health Care (35)
SIC: Health Services (80)

Vertical focus

- Ecosystem Foundry WG TF: [Sovrin Ecosystem Governance Framework](#)

Sector	Industry Group	Industry	Sub-Industry
45 - Information Technology	4510 - Software & Services	451020 - IT Services	45102010 - IT Consulting & Other Services
			45102020 - Data Processing & Outsourced Services
			45102030 - Internet Services & Infrastructure

GICS: Information Technology (45)
SIC: Information Technology Services (737109)

Vertical focus

- Ecosystem Foundry WG TF: [Internet of Education \(IoE\) Ecosystem](#)

Sector	Industry Group	Industry	Sub-Industry
25 - Consumer Discretionary	2530 - Consumer Services	253020 - Diversified Consumer Services	25302010 - Education Services
			25302020 - Specialized Consumer Services

GICS: Education Services (25302010)
SIC: Educational Services (82)

Identity Correlation Bitmap: An object for mitigating against attribute correlation patterns (35 mins)

Presented by: P.Knowles

<https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-122.pdf>

What is Personally Identifiable Information (PII)?

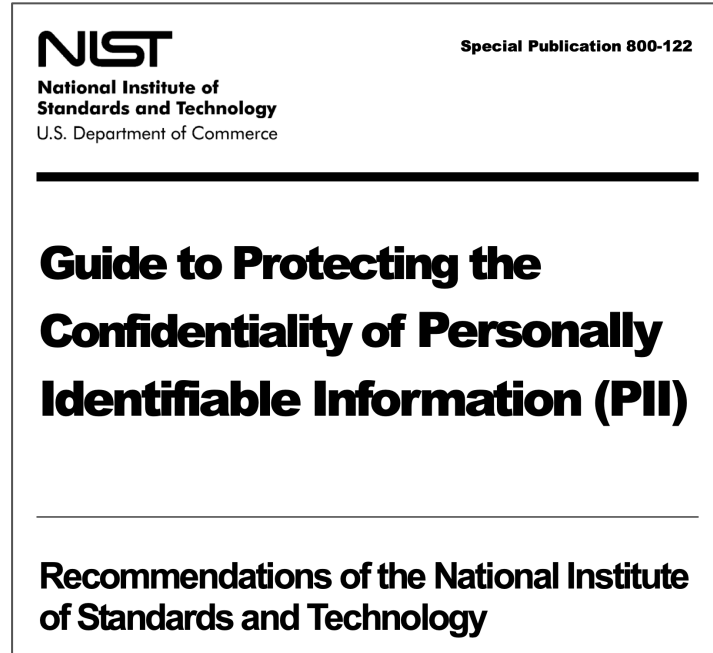
Personally identifiable information (PII) is any data that can be used to identify a specific individual. Social Security numbers, mailing or email address, and phone numbers have most commonly been considered PII, but technology has expanded the scope of PII considerably. It can include an IP address, login IDs, social media posts, or digital images. Geolocation, biometric, and behavioral data can also be classified as PII.

This broad definition of PII creates security and privacy challenges, especially when specific and stringent safeguards for it are spelled out in regulations such as the European Union's (EU's) General Data Protection Regulation (GDPR).

Ref.:

<https://www.csoonline.com/article/3215864/how-to-protect-personally-identifiable-information-pii-under-gdpr.html>

NIST 800-122: Guide to Protecting the Confidentiality of Personally Identifiable Information (PII)



Blinding Identity Taxonomy (BIT)



- Names (incl. First Names, Last Names, Full Names, Entity Names)
- Physical Addresses
- E-mail Addresses
- Telephone Numbers
- Postal Codes
- Personal Software Application Handles (e.g. Skype, Slack, Hyperledger Chat, etc.)
- Profile Pages
- Passport Numbers
- Social Security Numbers
- National Insurance Numbers
- Driving License Numbers
- Vehicle Registration Numbers
- Bank Account Numbers
- Credit (or Debit) Card Numbers
- Personal Identification Numbers (PIN)
- Private Keys / Master Keys
- Symmetric Keys
- Public Keys
- Link Secrets
- Employee Identifiers
- Account Identifiers
- Governmental Identifiers
- Membership Identifiers (e.g. Trade Union Membership, etc.)
- Institutional Identifiers (e.g. Private Health Care Identifiers, etc.)
- Case Identifiers (e.g. Case ID Numbers, Benefit Plan Participation Identifiers, etc.)
- User Identifiers (e.g. User IDs, Logins, etc.)
- Passwords
- Signatures
- Digital Certificates
- Photos
- Videos
- Images
- Vocal Sound Bites
- Dates and timestamps (e.g. Date of Birth, transaction dates, etc.)*
- Genetic Identifiers (incl. chromosomal, deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) data)
- Biometric Identifiers (incl. voiceprints, iris scans, facial imaging and dactyloscopic (fingerprint) data)
- Internet Protocol (IP) Addresses
- Media Access Control (MAC) Addresses
- Service Set Identifiers (SSID) (incl. local WiFi SSIDs)
- Bluetooth Device Addresses (BD_ADDR)
- Locational Information (incl. Global Positioning System (GPS), 3 word address, etc.)
- Cookie Browser Identifiers
- Radio Frequency Identifiers
- IoT Identifiers (incl. smart meter data)
- International Mobile Equipment Identity (IMEI)
- International Mobile Subscriber Identity (IMSI)
- Social media interactive elements, posts and comments (incl. likes, emojis and polling results)
- Free-Form Text Fields / Unstructured Data**

* Note: Not all captured dates will reveal identity but some will so, if in doubt, encrypt.

** Defn.: Text which does not have a given structure, nor which is entered in any specific format. Note: All free-form text fields should be encrypted.

Blinding attributes in a schema base

```
{
  "@context": "https://oca.tech/v1",
  "name": "Demographics",
  "type": "spec/schema_base/1.0",
  "description": "Personal demographic data",
  "classification": "GICS:35202010",
  "issued_by": "did:example:ebfeb1f712ebc6f1c276e12ec21",
  "attributes": {
    "email": "Text",
    "firstname": "Text",
    "lastname": "Text",
    "salutation": "Text",
    "birthdate": "Date",
    "gender": "Text"
  },
  "attr_blinding": [
    "email",
    "firstname",
    "lastname",
    "birthdate"
  ]
}
```



The BIT is a taxonomy of data fields to be blinded for the purpose of removing identity data from a dataset.

Mitigation against privacy attacks

Existing security mechanisms focusing on confidentiality and integrity cannot preserve privacy effectively. For instance, while data is protected over encrypted communication, external attackers still learn query location and data location from eavesdropping. Combining types of unintentionally disclosed information, the attacker could further infer the privacy of different stakeholders through attribute-correlation attacks and inference attacks.

Attribute-correlation attacks

Attribute-Correlation Attack: The Predicates of an XML query describe conditions that often carry sensitive and private data (e.g., name, SSN, credit card number, etc.) If an attacker intercepts a query with multiple predicates or composite predicate expressions, the attacker can correlate the attributes in the predicates to infer sensitive information about the data owner. This is known as the attribute-correlation attack.

Example:

Mr.Ami is sent to ER at California Hospital. Doctor Sham queries for her medical records through a medicare IBS. Since Ami has the symptom of cancer, the query contains two predicates: [pName=Ami], and [symptom=cancer]. Any malicious broker that has helped routing the query could guess Ami has leukemia by correlating the two predicates in the query. Unfortunately, query content including sensitive predicates cannot be simply encrypted since such information is necessary for content-based query routing. Therefore, we are facing a paradox of the requirement for content-based brokering and the risk of attribute-correlation attacks.

Ref.:

<https://www.ijert.org/privacy-preserving-and-information-security-forensics-brokering>

Inference attacks

Inference Attack: More severe privacy leaks occur when an attacker obtains more than one type of sensitive information and learns explicit or implicit knowledge about the stakeholders through association. By implicit, we mean the attacker infers the fact by guessing. For example, an attacker can guess the identity of a requestor from her query location (e.g., IP address). Meanwhile, the identity of the data owner could be explicitly learned from query content (e.g., name or Credit card details). Attackers can also obtain publicly-available information to help inference. For example, if an attacker identifies that a data server is located at a leukemia research center, they can tag the queries as leukemia-related.

Identity Correlation Bitmap : Preparing a Schema Base

```
{
  "@context": "https://oca.tech/v1",
  "name": "Demographics",
  "type": "spec/schema_base/1.0",
  "description": "Personal demographic data",
  "classification": "GICS:35202010",
  "issued_by": "did:example:ebfeb1f712ebc6f1c276e12ec21",
  "attributes": {
    "email": "Text",
    "firstname": "Text",
    "lastname": "Text",
    "salutation": "Text",
    "birthdate": "Date",
    "gender": "Text"
  },
  "attr_blinding": [
    "email",
    "firstname",
    "lastname",
    "birthdate"
  ]
}
```

Schema base attributes would need to hold an attrib number for the bitmap to work well.

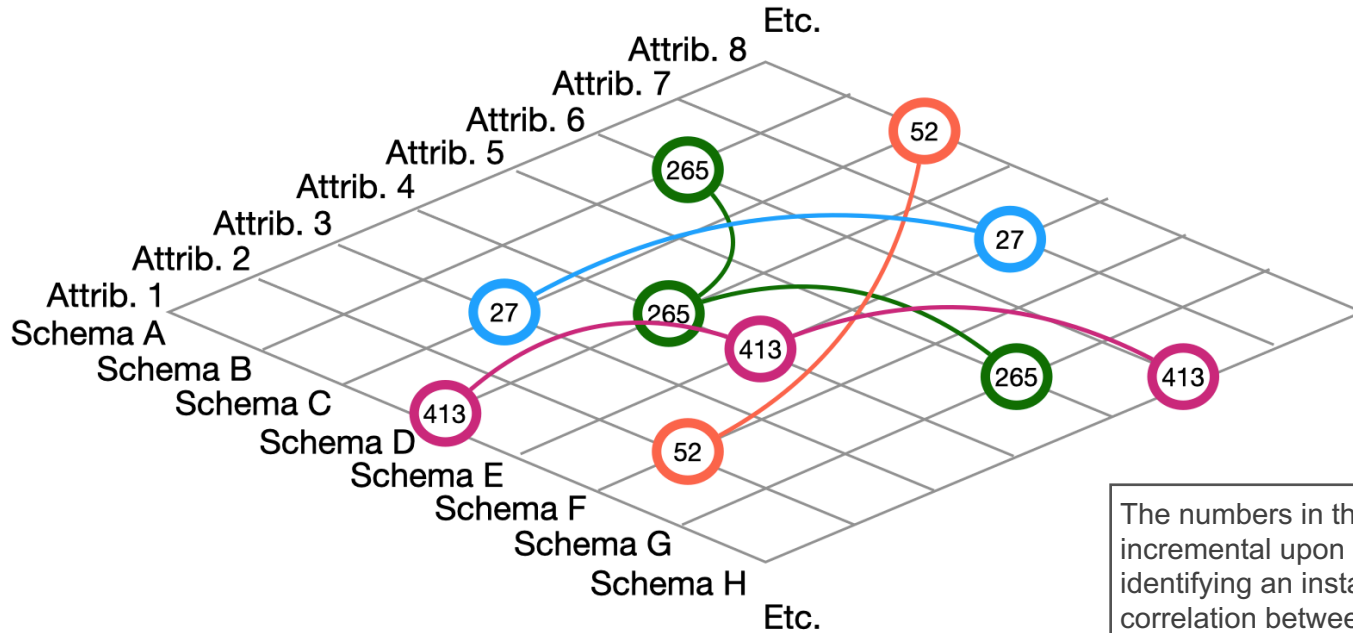
e.g., in this example ...

"email" = Attrib. 1
"firstname" = Attrib. 2
"lastname" = Attrib. 3
"salutation" = Attrib. 4
"birthdate" = Attrib. 5
"gender" = Attrib. 6

The schema DRI could be used along the other axis.

Building an Identity Correlation Bitmap

The creation of a dynamic bitmap that constantly evolves each time an assessor identifies a correlation risk between attributes that could potentially unblind the identity of a governing entity. The numbers in the circles are incremental upon assessors identifying an instance of correlation.



The numbers in the circles are incremental upon assessors identifying an instance of correlation between two or more datasets. It is not a record of how many "Data Subjects" are affected.

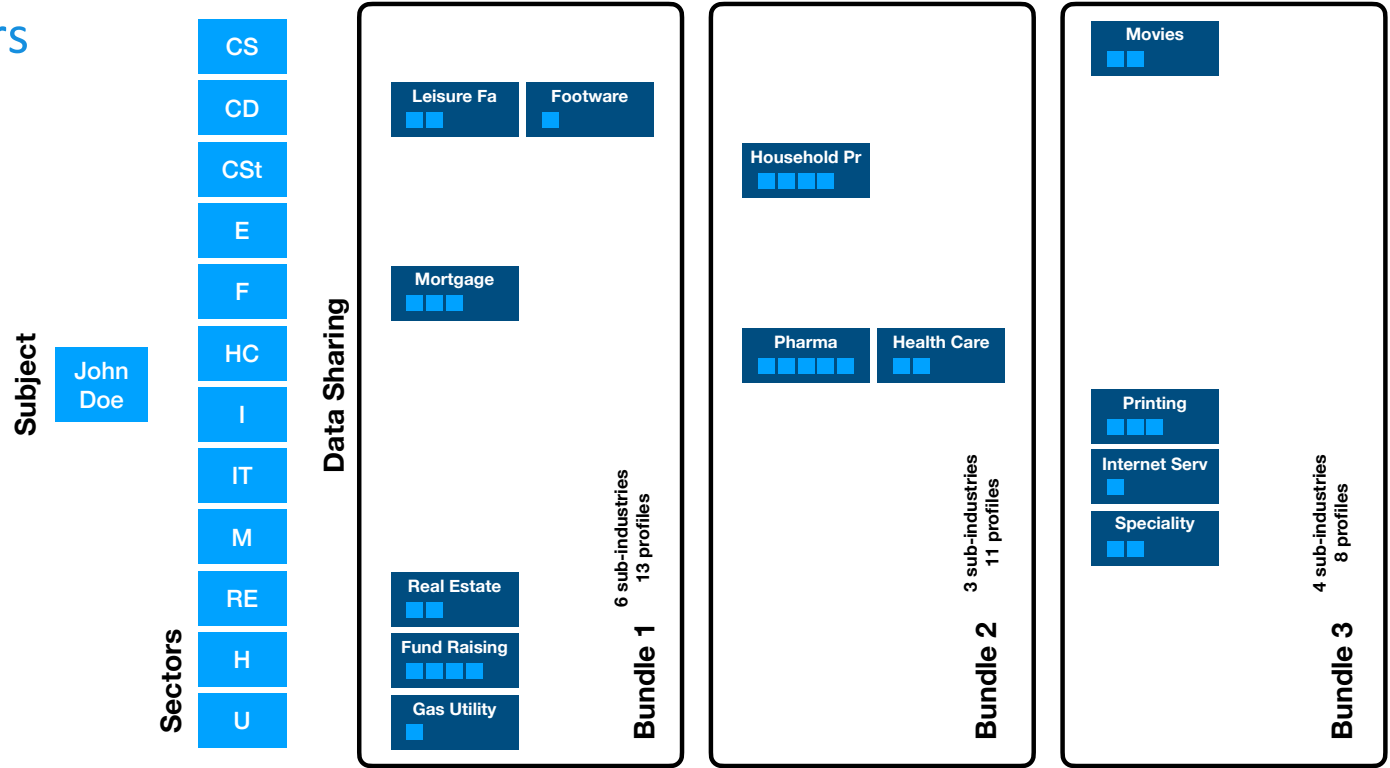
Correlation coefficients

When information of a dataset are analysed, whose origin or “feed” may be a database, information of raw files, logs, spreadsheet data, etc. one of the most powerful tools for drawing conclusions is to carry out correlations.

It is a statistical-based, and thus, mathematics-based information analysis technique. It consists of analysing the relationship between at least two variables, e.g. two fields of a database or of a log or raw data. The result will display the strength and direction of the relationship.

Correlation coefficients are used in statistics to measure how strong a relationship is between two variables. There are several types of correlation coefficient, but the most popular is *Pearson's*. *Pearson's correlation* (also called *Pearson's R*) is a correlation coefficient commonly used in linear regression. If you're starting out in statistics, you'll probably learn about *Pearson's R* first. In fact, when anyone refers to “**the**” correlation coefficient, they are usually talking about *Pearson's*.

Linking identifiers



A visual on what a linking identifier needs to achieve. In the diagram, there is one subject ('John Doe') with three consented data bundles. Each bundle includes a number of profiles. For each bundle, a linking identifier is needed as a thread to knit the profiles together. One linking identifier per consented bundle.

Logistics and miscellaneous (5 mins)

<https://wiki.trustoverip.org/display/HOME/2020-11-24+Weekly+Meeting>

News from the Operations Team

Nick Hayfack

(Semantics WG representative on the ToIP Operations Team)

The purpose of the Operations Team is to create a small group of ToIP members who will share information on the workplans of our WGs, help ensure that draft deliverables are advancing as intended through the stages of the ToIP workflow, resolve any bottlenecks that arise around decision-making/approvals and discuss issues such as (for example) introducing firmer parameters for the creation of Taskforces under all WGs.

The Operations Team will not be directing or otherwise interfering with the development of content and deliverables in the WGs themselves.

Leadership positions

- › Semantics WG Chair
 - › **Paul Knowles** (Human Colossus Foundation)
- › Semantics WG Vice-chair
 - › **John Wunderlich** (JLINC Labs)
- › Operations Team Group Representative
 - › **Nick Nayfack** (Team Ikigai)
- › We can periodically rotate chairs as needed
- › Volunteer via the meeting page at ...
 - › <https://wiki.trustoverip.org/display/HOME/2020-11-24+Weekly+Meeting>

Meeting schedule

- › **Notice & Consent TF** bi-weekly meeting
 - › Thursday, **November 26th** @ 08:30 US PT / 17.30 CET
 - › Zoom link: <https://zoom.us/j/92346573961?pwd=RmZHNNQxS2Iya3NCMHZTVXYra3Rrdz09>
- › **Semantics Domain WG** weekly meeting
 - › Tuesday, **December 1st** @ 09:00 US PT / 18.00 CET
 - › Zoom link: <https://zoom.us/j/93406719136?pwd=SUozZHBQM0N5TUhYMhJqL0ZQM3I3Zz>
- › **OCA-FHIR FG** bi-weekly meeting
 - › Thursday, **December 3rd** @ 08:00 US PT / 17.00 CET
 - › Zoom link: <https://zoom.us/j/93406719136?pwd=SUozZHBQM0N5TUhYMhJqL0ZQM3I3Zz09>



Closing Q & A

Legal Notices

The Linux Foundation, The Linux Foundation logos, and other marks that may be used herein are owned by The Linux Foundation or its affiliated entities, and are subject to The Linux Foundation's Trademark Usage Policy at <https://www.linuxfoundation.org/trademark-usage>, as may be modified from time to time.

Linux is a registered trademark of Linus Torvalds. Please see the Linux Mark Institute's trademark usage page at <https://lmi.linuxfoundation.org> for details regarding use of this trademark.

Some marks that may be used herein are owned by projects operating as separately incorporated entities managed by The Linux Foundation, and have their own trademarks, policies and usage guidelines.

TWITTER, TWEET, RETWEET and the Twitter logo are trademarks of Twitter, Inc. or its affiliates.

Facebook and the "f" logo are trademarks of Facebook or its affiliates.

LinkedIn, the LinkedIn logo, the IN logo and InMail are registered trademarks or trademarks of LinkedIn Corporation and its affiliates in the United States and/or other countries.

YouTube and the YouTube icon are trademarks of YouTube or its affiliates.

All other trademarks are the property of their respective owners. Use of such marks herein does not represent affiliation with or authorization, sponsorship or approval by such owners unless otherwise expressly specified.

The Linux Foundation is subject to other policies, including without limitation its Privacy Policy at <https://www.linuxfoundation.org/privacy> and its Antitrust Policy at <https://www.linuxfoundation.org/antitrust-policy>, each as may be modified from time to time. More information about The Linux Foundation's policies is available at <https://www.linuxfoundation.org>.

Please email legal@linuxfoundation.org with any questions about The Linux Foundation's policies or the notices set forth on this slide.