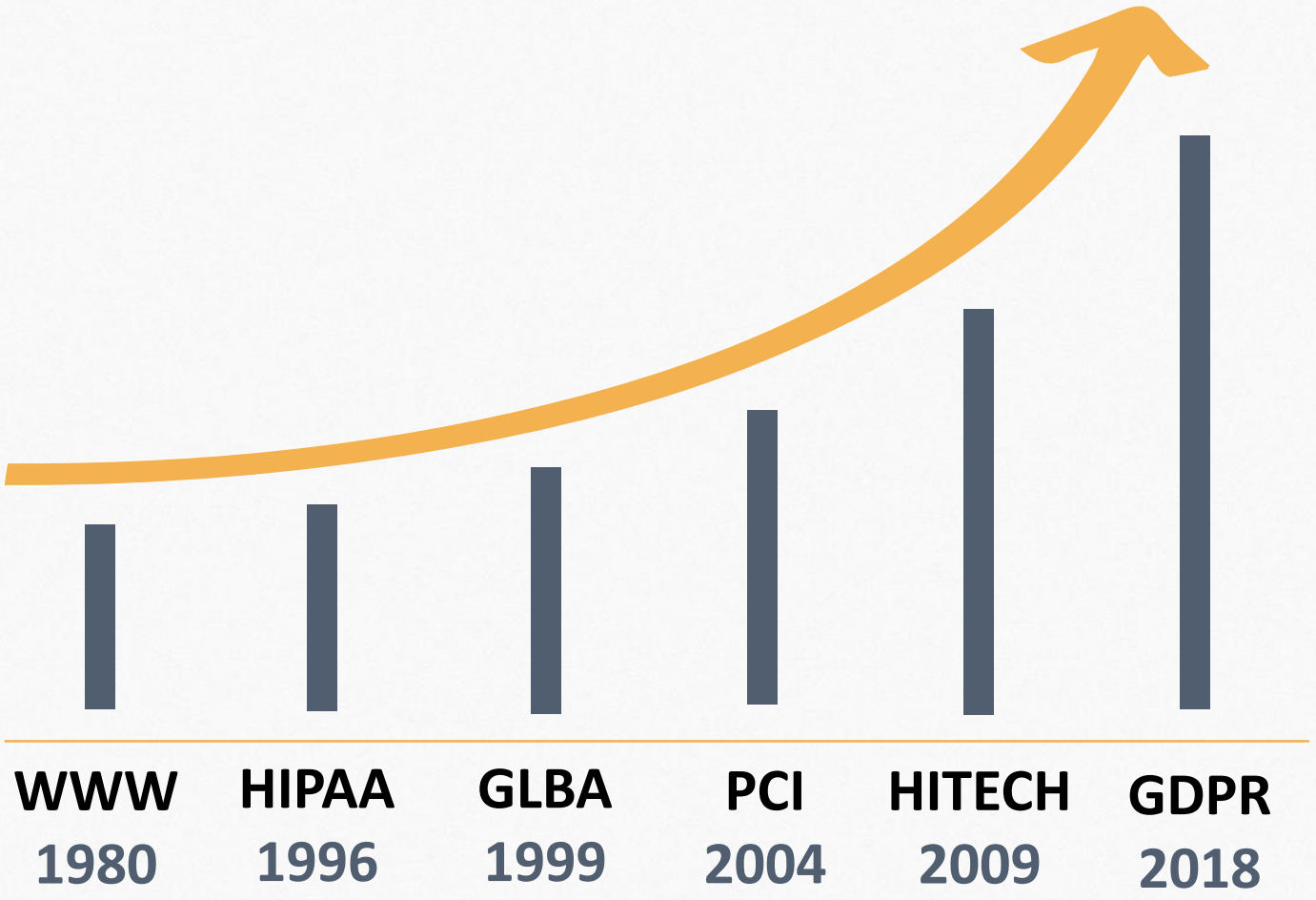
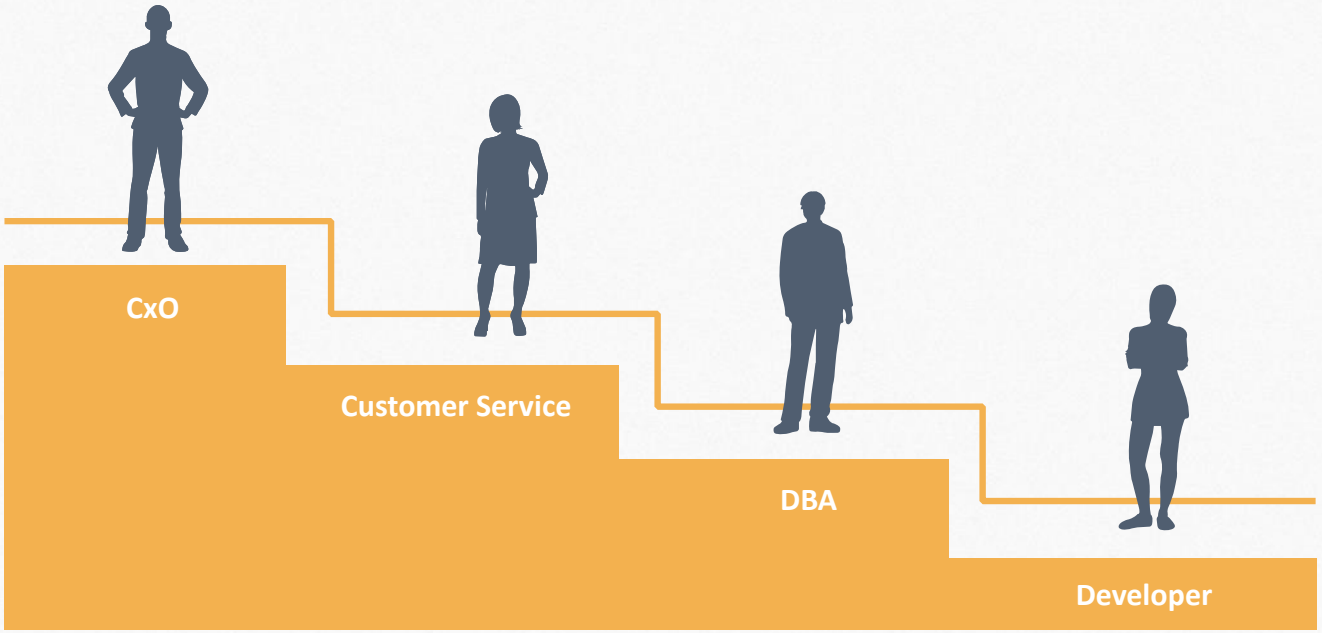
A COMPONENT-BASED DATA MASKING SOLUTION

WITH INTERNET ADVANCE, DATA BREACHES, INCREASE AND REGULATIONS TAKE HOLD



INTERNAL THREATS – WHO BREACHES PRIVACY?

Rare	Selling PII on the "black market"	Selling PII for Gain	Selling PII, sabotage
CxO	Customer Service	DBA	Developer



Accidental Disclosures
Partner/Researcher

WHAT DO GDPR/GLBA/HIPAA SAY ABOUT INTERNAL THREATS?

- *the Privacy Rule*
- *the Transactions and Code Sets Rule*
- *the Security Rule*
- *….*
- *the Unique Identifiers Rule*
- *the Enforcement Rule*
- *The Pseudonymization Rules*
- *….*

STATIC DATA MASKING/ANONYMIZATION

… When using or disclosing protected health information or when requesting protected health information from another covered entity, a covered entity must make reasonable efforts to limit protected health information to the minimum necessary to accomplish the intended purpose of the use, disclosure, or request.

TOKENIZATION/PSEUDONYMIZATION

…Tokenization involves removing or obscuring direct identifiers and, in some cases, certain indirect identifiers that could combine to reveal a person's identity. These data points are then held in a separate database that could be linked to the de-identified database through the use of a key, such as a random identification number or some other pseudonym.

DYNAMIC DATA MASKING (Section 164.308)

…Information access management's implementation specifications: Implement policies and procedures for granting access to electronic, protected health information, for example, through access to a workstation, transaction, program, process, or other mechanism.

LACK OF CLEAR GUIDANCE RESULTS IN CHAOS. LET'S BRINGING AN ORDER TO CHAOS.

USA	EU
DATA MASKING / ANONYMIZATION	ANONYMIZATION
HIPAA	GDPR PREFERENCE
HITECH	
GLBA	
TOKENIZATION/PSEUDONIMIZATION	PSEUDONYMIZATION
PCI/DSS	GDPR – recital 26, 29, article 5, article 6(4)
NEWEST ACT (1/14/2018)	
OPEN GOVERNMENT DATA ACT https://www.datacoalition.org/press-releases/president-signs-government-wide-open-data-bill/	ANONYMIZATION REQUIRES ROLLUPS (Patent pending)

PSEDONYNIZATION ENACTMENT CONFUSION

REQUIRES A PROOF OF SUFFICIENCY –YOU NEED A RISK ASSESSMENT PROCEDURE

REGULATION NECESSITATES DEFINING AND FINDING ALL THE SENSITIVE DATA.

HUSH HUSH

WHAT IS SENSITIVE DATA?

SENSITIVE DATA MODELS: PII

The term “PII,” as defined in OMB Memorandum M-07-1616 refers to information that can be used to **distinguish or trace an individual’s identity**, either alone or when combined with other personal or identifying information that is linked or linkable to a specific individual.
US General Services Administration

Personally Identifiable Information is a sensitive and critical organizational resource.


Credit Card Numbers


Names


Social Security Numbers


DOBs

HOW DO WE FIND SENSITIVE DATA?

DATA CLASSIFICATION – INDUSTRY WAY , PII AND BEYOND, LACK OF ATTACKS METRICS

"CLASS": FINANCIAL DATA

"CLASS": HEALTHCARE DATA

"CLASS": EDUCATIONAL DATA

"SAFE HARBOR" - ACADEMIC AND HITECH WAY, PII/PHI, K-ANONYMITY AND OTHER METRICS

LATANYA SWEENEY'S HIPAA MODEL ("Safe Harbor", "Expert Determination")

WITH RISK ASSESSMENT ELEMENTS AND ANONYMITY ATTACKS METRICS

GDPR - VERY VAGUE, BASED ON ANONYMIZATION CONCEPT, PII, K-ANONYMITY METRICS

PSEUDONYMIZATION: DIRECT IDENTIFIERS + SEPARATION CONCEPT

ANONYMIZATION: DIRECT + INDIRECT IDENTIFIERS

WHAT DO WE THINK?

YOU NEED TO CLASSIFY DATA BASED ON TYPE OF ATTACKS

CREATE MODELS OF RISK ASSESSMENT

FINANCIAL –

data needed for the application for bank account

HEALTHCARE CLAIM –

data needed to get money for the claim

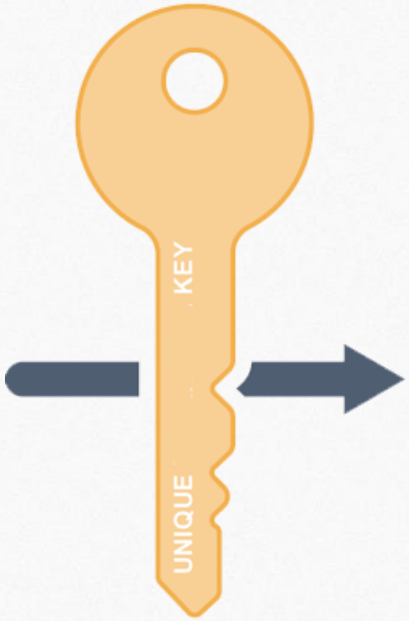
MIX OF INDUSTRY AND ACADEMIA, DATA COMES FROM REAL ATTACKS AND EXTENDS INTO CLASSES.

SATISFIES THE STRONGEST HITECH AND GDPRA REQUIREMENTS INCLUDING RISK DETECTION AND FOLLOWING ASSESSEMENT OF THE SOLUTION.

WHAT ARE DIRECT IDENTIFIERS PER GDPR?

UNIQUE DATA

- Social security number
- **(123-45-6789)**
- Passport number
- **(C00001234)**
- Driver's license
- **(123-456-789)**
- Account Numbers
- Credit card
- **(4234-5678-9123-4567)**
- phones, faxes, VINs, IPs
- **NON-UNIQUE DATA**
- names, DOB



SDM: MASKED DATA

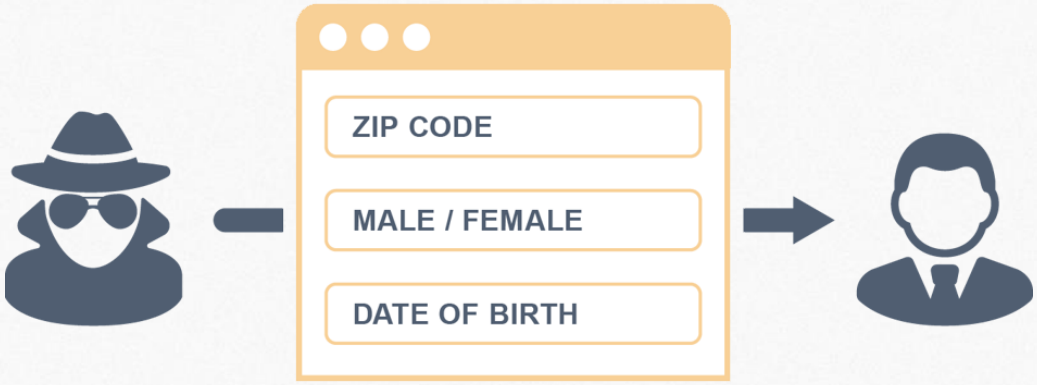
- **987-65-4321**
- **A00009876**
- **4276-5432-1987-6543**
- **654-987-321**

PSEUDONYMIZATION

- **Key1, ID**
- **Key2, ID**
- **Key3, ID**

INDIRECT IDENTIFIERS - ANONYMIZATION

PUBLIC DATA SETS: K-ANONYMITY, LATANYA SWEENEY



Statistics:

- **Zip code 27615: 45,090 people (Raleigh, NC)**
- **Over 65 → 5,190**
- **Female over 65 → 2,595**
- **Female over 65 with birthday on April 3 → 7**
- **DOB: April 3, 1946 → 1**
- **87% of people in the US can be re-identified with Gender, Zip & DOB**

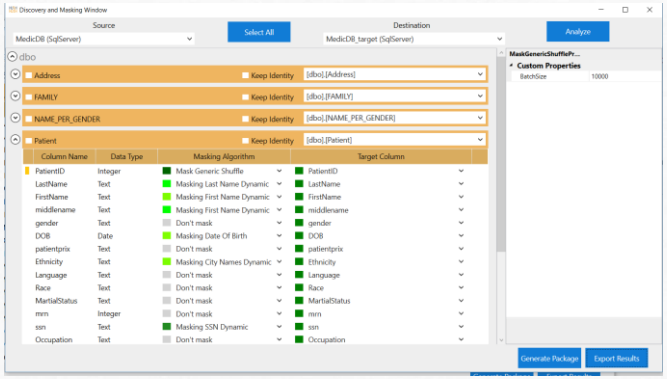
DATA MASKING PROCESS

Mask Personally Identifiable Information:

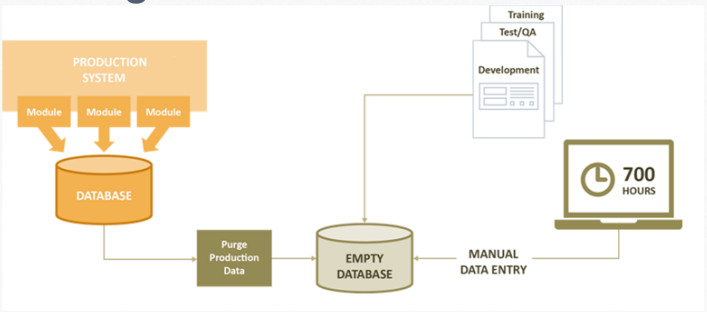
1. DISCOVER SENSITIVE DATA
2. FIGURE OUT THE TYPES OF ATTACKS
3. DECIDE “PSEUDO-” VS “ANO-”
4. DE_IDENTIFY: COMMON METHODS + REMOVE THE OUTLIERS
5. ASSESS THE RISKS OF RE-IDENTIFICATION

USING HUSH-HUSH PRODUCTS FOR DIFFERENT SCENARIOS

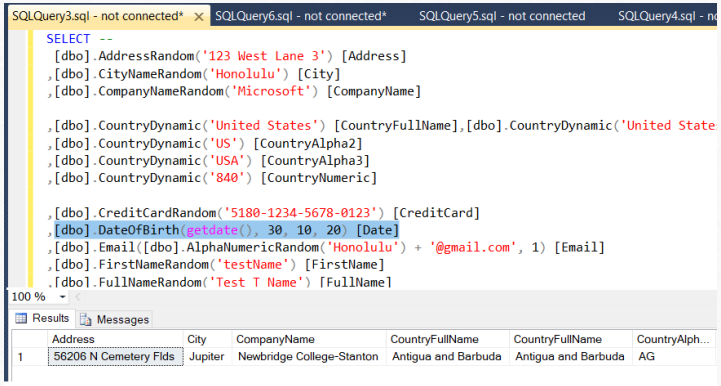
SENSITIVE DATA DISCOVERY/ RISK ASSESSMENT



DEVOPS (SSIS) – populating the environments, sharing



IN-LINE SQL FUNCTIONS – privacy in sql,ease of use



PRIVACY IN DESIGN (API) – public reports, sharing

HUSHHUSH DISCOVERY SOLUTIONS

SENSITIVE DATA DISCOVERY AND PACKAGE GENERATION TOOL

- Scanning both data and metadata
- Categorizing data into PII classes
- Suggesting algorithms with the gradual indication of levels of confidence
- Mapping source and destination schemas based on the best match
- Creating Audit reports
- Generating masking workflows that embed in to Microsoft's SSIS (upload tool)

- Also, hl7/x12 beta file processing tool, text capabilities are POC based on Stanford parser.

HUSHHUSH MASKNG SOLUTIONS

VARIETY OF ALGORITHMS

Unique Elements:

- Entity and format specific (SSN, SIN, PHONE, CREDIT CARD #, IP, etc.)
- Mask-a-Key (patent-pending)

Non Unique Elements : Entity Specific, Generic

Statistical Challenge –**Rollups (Patent Pending),** Number and Date variance

Consistently maintaining integrity across enterprise and random

COMPONENT FORM

API,

SSIS,

CLR,

POSTGRESQL (beta)

PERFORMANCE

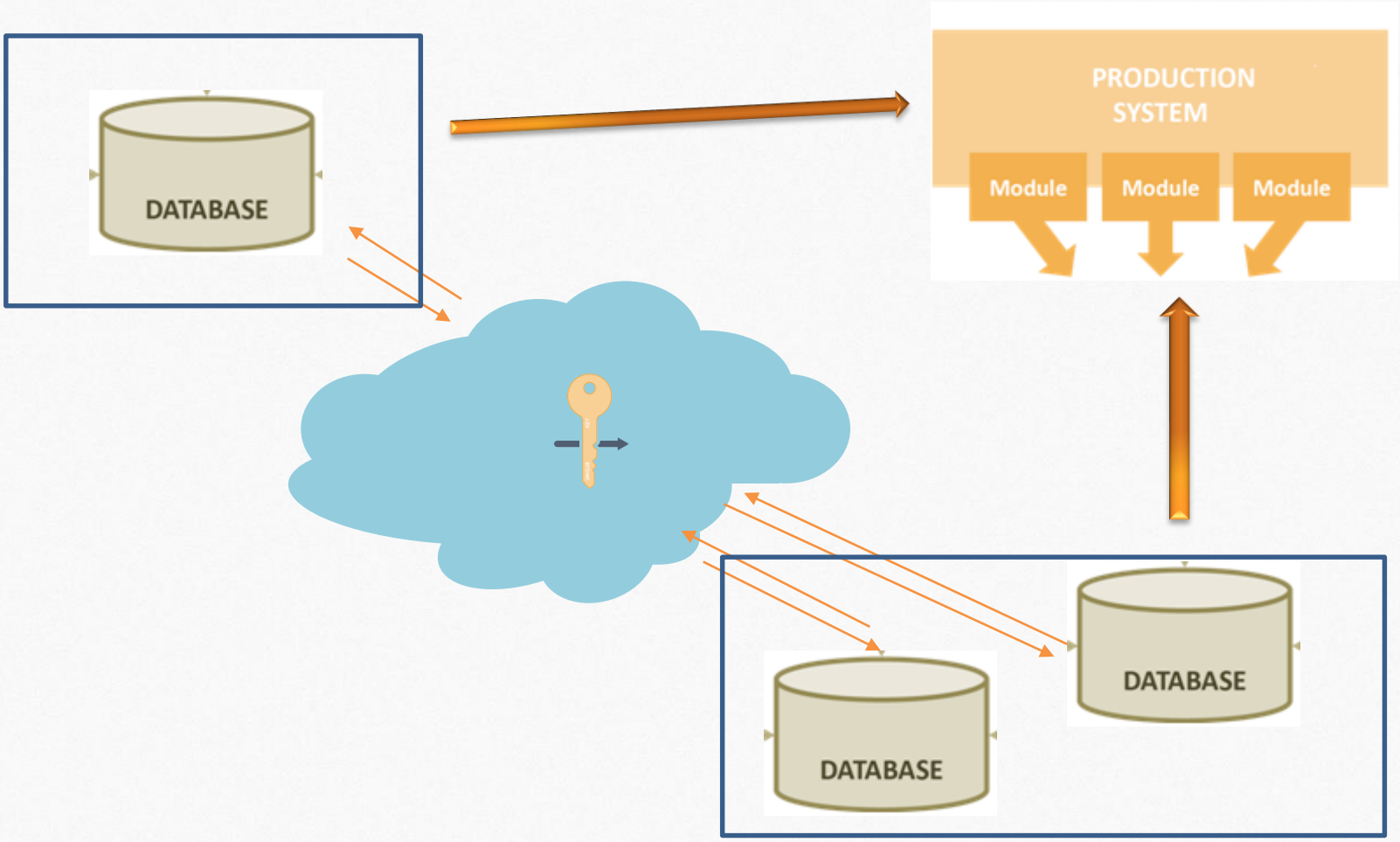
Benchmarked components

High performance

VIRTUAL MACHINE IMAGES IN AZURE

- Components in SSIS
- SDDT and Components

APPENDIX 1
PSEUDONIMIZATION/TOKENIZATION AS PRESCRIBED BY GDPR



APPENDIX 2
TYPICAL EXAMPLE OF A PUBLIC REPORT POSSIBLE WITH ROLLUPS

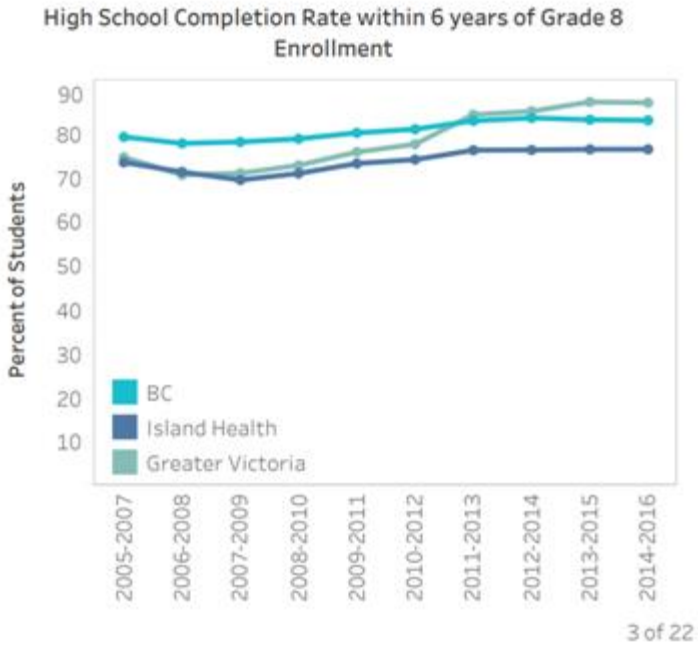
Education

High school completion rates have been increasing over the last ten years and are higher than the rates for Island Health and BC. Also, according to the 2016 Census, a higher proportion of the Greater Victoria LHA adult population have completed post-secondary education.

Population Aged 25 to 64 with Post-Secondary Certificate, Diploma or Degree (%)

2016

BC	Island Health	Greater Victoria
63.9	63.7	70.3



3 of 22