

Data Classification and Handling: Discovery and Response Prioritization

Darrell Berninger
& Tim Plona

PUBLIC

Introductions:

- Darrell Berninger, CISSP, GSEC, GWAPT, GCIH:
Application Security Architect
 - Previously with: Reed Elsevier and LexisNexis
- Tim Plona, CISSP, CISM, CISA, CRISC, CGEIT:
Business Solution Architect
 - Previously with: Canon Printing, Océ Technologies, Dana Corporation, Bekins Van Lines.

Objectives

- Discuss “The Problem”
- How to engage your business in a focused approach
- How to ask questions to create a data registry
- How to use process to promote a more secure environment
- How to evaluate data risks through content and usage
- How to prioritize data classification efforts
- And how to do all of this on the “cheap”

The Problem Is...

There's usually is a lot of friggin' data!

It's not easy for us to define

- Determine Where – “We have an office in *that* country?”
- Describe What – “Is data supposed to feel squishy?”
- Delineate Who – “Who the heck owns this stuff?”
- Define Controls – “Its sleeping softly on a cloud.”
- Define Process – “Fold tab A into slot Z12!”

*“Hi, my name is Joe and
I am a data-holic...”*

So Why is There a Problem?

People keep making it and technology keeps shoveling it.

- Overwhelming Amounts (Technical term = 'friggin')

Make your storage architect's day by asking how much!

Its consistently inconsistent and can get complicated quickly

- Storage
- Use
- Location
- Language
- Access
- Authorization
- Accountability
- Confidentiality
- Availability
- Format
- Etc., etc., etc...

So what's the Problem?

Data Classification Projects Are Often Overlooked

- Its likely not going to win a “project of the year award”.
- Get's complex with deep back end processes with few sponsors
- Hard to do well with limited funds?
 - Perception that risk reduction means big effort with big dollars.
- Uncovering issues under rocks can overwhelm a security or records management team “*Don't we have enough to do already?*”
- Requests that employees change a process or activity are not appreciated

So what's the Problem?

Policy is Required

- Do policies on the topic already exist or are effective?
 - Are they achievable and measurable?

Risk Models Can Complicate Things:

- Which model is worth the investment for your organization?
- Working with them can require special skills.

Definition Confusion Complicates Things:

- What is data? / What are records? / Paper? / Bits?
- How many levels of classifications will you have?

Defining the Unknown

“Reports that say that something hasn't happened are always interesting to me, because as we know,

- *there are **known knowns**; there are things we know we know.*
- *We also know there are **known unknowns**; that is to say we know there are some things we do not know.*
- *But there are also **unknown unknowns** -- the ones we don't know we don't know. And if one looks throughout the history of our country and other free countries, it is the latter category that tend to be the difficult ones.”*

US Secretary of Defense, referring to the war in Iraq.

Known and Unknown Data Types

- **Known Knowns:** There is some data in any organization that we easily know as being of a nature that require protection:
 - SSN
 - Names and DOB
 - Salaries
 - Medical Records (HIPAA)
- **Known unknowns:** There is some data in any organization that we know as being of a nature that require protection, but we just don't know enough about it:
 - What exactly is Human Resources storing about everyone?
 - What exactly does Mergers and Acquisitions look at as they consider opportunities?
 - What customer data does the account team retain?
 - What customer purchase data does the marketing team hold?
- **Unknown unknowns:** Even if you have been in your organization since day one, you can not know about all the valued data of the organization.

Known and Unknown Data Usage

- **Known Knowns:** Likewise we know there are patterns of data use (transport and storage), such as:
 - E-mail file attachments
 - File Transfers (FTP/HTTP/etc.)
 - Laptops or Mobile Devices
 - Poorly managed network file shares
 - USB Thumb drives and portable hard drives
 - Desk Drawers / File Cabinets
- **Known Unknowns:** We know that there are ways that users move data around without our knowledge – we know it is unknown to us.
 - **Old fashion Sneaker-Net:** Passing media from one person to another. Ubiquitous USB drives make this very possible and very problematic.
 - **Cloud File Sharing:** Passing data from one to another through unknown web based products.
- **Unknown Unknowns:** Are people storing and using data in manners you have never seen before in your organization? How would you know?

A Note About Regulatory Data

Before we move on....

- Certain data types are going to need extra special treatment
 - HIPAA
 - GLBA/DPPA
 - Credit header data (SPI)
 - FERPA
 - Etc...
- This falls outside the scope of this presentation.

Understand the requirements through your corporate counsel.

- The Complexity Increases
- We are not lawyers, nor do we play ones on TV!

General Questions

To Begin, You Need to Start Asking some Basic Questions:

- Who makes the final decision on how data is handled?
(Ownership)
- Do we already have a policies or standards to work with?
(Awareness)
- What is the risk related to the content of the data?
(qualification)
- How do I build a risk model? (quantification)
- Where do I focus efforts first? (prioritization)

Providing an Approach: Reaching Owners

- Know and Query your Business Owners
 - Determine the best, business culturally acceptable means to get the topic in front of them.
 - Make it real for them:

Confidentiality: *What would happen if your important records were left on a bus stop bench right in front of your office? Or published in a national newspaper?*

Integrity: *What would be the impact if someone changed these records maliciously and you did not catch it?*

Availability: *What would happen if you no longer could access these records?*

Remember, a data owner will always know the value and sensitivity of their data – better than you will

Results of the Approach: Knowledge!

- **Ownership:** Having these conversations, helps the business process owner take ownership for their important records! They buy in!
- **Answers to questions:**
 - Who owns what data?
 - What is the risk appetite of the organization?
 - What are the highest value record sets?
 - What are the highest value business processes?
 - What risky behaviors are present in our organization?
- **End-User Awareness**
 - You are talking to people about data handling. It will pique their interest.
 - You are opening a channel of communication. You now know each other.

Know Your Users and Their Data

- Ask around to seed your list and find out what is important!
 - Talk to company lawyers, privacy officers, compliance officers and internal auditors.
 - Do you have a Records Management group that already knows about retention and securing of certain records?
 - Are there Business Continuity / Disaster Recovery plans that prioritize data sets or business processes?
 - Do you have business analysts or business solution architects that work with business process leaders?
 - Take a quick look at available data stores.

Key Knowledge to Gain: Qualities I

Define the qualities/contents of the data (data types):

- **PII - Personal Identifiable Information** (The stuff that individuals care about not being disclosed)
Examples: Name, date of birth, home address, social security numbers, medical test results, back ground checks, liens and garnishments, criminal records, employment reviews, disciplinary action, etc.
- **CPI - Corporate Proprietary Information** (The stuff that companies care about not being disclosed)
Examples: Company financials, intellectual property, trade secrets, mergers and acquisitions, partner and supplier contracts or agreements, strategic planning, etc.
- **Compliance or Regulated Content** (The stuff everyone, including governments, care about)
Examples: Sarbanes Oxley, HIPAA, EU Privacy Act, EPA, Labor Agreements, GLBA, DPPA, FERPA, etc.

Key Knowledge to Gain: Qualities II

Ask how the data is **created, used and destroyed** (data lifecycle).

- What is the workflow that creates the data?
- What organization or groups share the data?
- Is the data required to be retained? How long?
- How are paper records handled?
- How is access or authorization granted, audited and revoked?

Ask how is the **data is stored** (data at rest) to find out all the places it resides.

- Desktop / laptop / flash drive / portable HDD (are they encrypted?)
- What applications? E-mail Server? Enterprise application, or off the shelf software from Best Buy?
- Servers / network shares / SharePoint
- Cloud (is it contracted? Is it encrypted? How?) There can be huge differences between Google Docs and the business focused side of Azure.

Key Knowledge to Gain: Qualities III

Ask how the **data is transmitted** and where it goes.

- Does the data leave the organization (on purpose)?
- What applications/methodologies are used to transmit the data?
- Is the data transmitted securely?

Ask how the **user values** the data they handle.

- Do they have concerns? Do they have restless nights?
- Do they feel they have the appropriate awareness?
- Do they know who to ask for assistance?
- Do you agree with how the user valued their data?

What to do with what you learned?

- Make it usable – get it into a database
- Add value to your knowledge.
- How much risk is there in a DOB + Name + SSN compared to an employee bank routing and account number.
- Prioritize your risk:
 - Content Risk
 - Personal Identifiable Information risk
 - Corporate Proprietary Information Risk
 - Usage Risk
 - Storage Risk
 - Transmission Risk

PII Content Risk Considerations

Personal Identifiable Information	
Bank Information	1
Credit Report	3
General Individual Info (name, address, DOB together)	20
ID or Log In name	1
Individual Medical Information	40
Individual Names	1
Individual's Payment Card Information	5
Individual's PIN	1
Individual's SSN or FEIN's	20
Individual's Salary / Benefit	20

- What is most important to your company?
- Where is the greatest risk if confidentiality, integrity or availability is lost?
- What government regulations automatically set a risk score for you? (i.e. HIPAA, Consumer Protection Act.)
- Does the PII risk score change if it is an employee's information or customer's information?
- Does your company's business vertical change the risk scores? Does a bank score these differently than a mining company?

CPI Content Risk Considerations

Corporate Proprietary Information	
5.1 Company Strategy	20
5.1 Company Finances	3
5.1 CPI Intellectual Property	40
5.1 Company Credit Information	3
5.1 Company Bank Info	3
5.1 Company Merger, Acquisition and Divesture Information	40

- What is most important to your company?
- Where is the greatest risk if confidentiality, integrity or availability is lost?
- What government regulations automatically set a risk score for you? (SOx, GLBA, etc.)
- Do some scores change when they are pre or post release? (i.e. Financials)

Storage Risk Considerations

Storage Risk Scores	
E-mail	5
Enterprise Vault	1
Employee desktop	3
Desktop (encrypted)	1
Notebook	10
Notebook (encrypted)	3
Mobile Device	6
Network Share	3
SharePoint	1
Other Database	5
SAP	1
Paper	3
External HDD	9
External HDD (encrypted)	3
Flash Drive	12
Flash Drive (encrypted)	3
CD/DVD	3
Personal Cloud	6
Enterprise Cloud solution	1
Scanned images	3

- Where does this data come to rest?
- Is the resting location secure? How? Encrypted? Managed access? On prem / off Prem?
- Does storage in one location automatically imply storage in other locations? (i.e E-mail leads to desktops, laptops, smart phones etc.)

Transport Risk Considerations

Transport Risks	
Email	8
Voltage (encrypted mail)	1
Encrypted Zip	1
FTP	15
SFTP	2
BizTalk (automated file transfer)	1
Hightail (Enterprise managed Secure FTP)	1
HTTP	15
HTTPS	5
VPN	1
Wire	1
Fax	1

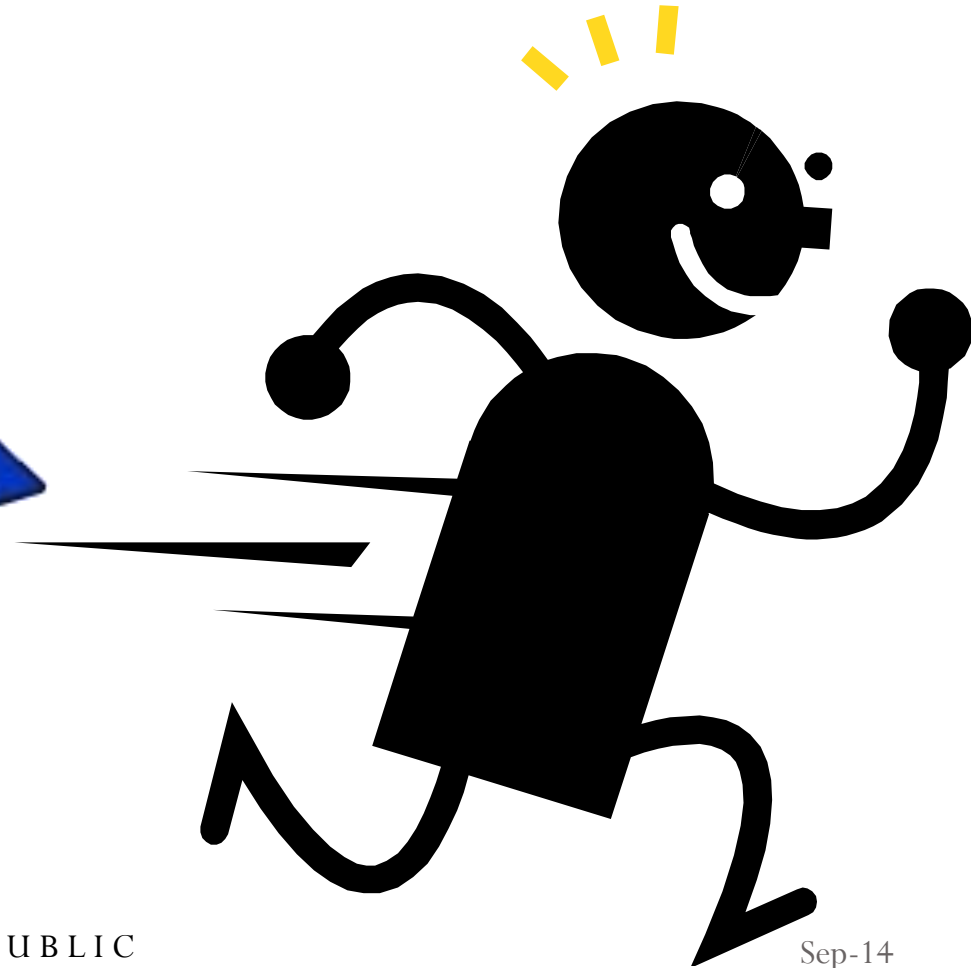
- How do users in a real day to day environment move data around?
- What do the users find to be an easy and quick method to move things around? It often is not the most secure method.

Geographical Risk Considerations

Geography Risks	
Stored in Corporate Data center	1
Stored in remote office data “closet”	8
Stored in second country	12
Stored in tertiary country with lax controls	36
Stored in partner’s data center	12
Stored in contracted vendor’s tier I data center	2
Stored in MIS Director’s garage	30

- Does the location of the data center where the data is at rest matter?
- Do the laws of the geographic location of the data storage location matter?
- Political stability?
- Ethics culture?

From theory and process to analysis



Now connect the dots – Content Risk

- Human Resources has an Excel spreadsheet that has the following contents:
 - Names of all employees
 - Date of Birth
 - Home Address
 - Social Security numbers
 - Current salary with new salary and bonus information.
 - Employee bank routing numbers

Personal Identifiable Information	
Bank Information	1
Credit Report	3
General Individual Info (name, address, DOB together)	20
ID or Log In name	1
Individual Medical Information	40
Individual Names	1
Individual's Payment Card Information	5
Individual's PIN	1
Individual's SSN or FEIN's	20
Individual's Salary / Benefit	20

PII Content Risk

Score = 62

Now Connect the Dots: Storage Risk

- The Excel file is created on an HR Analyst's unencrypted laptop, shared with various managers via e-mail. It is stored on a corporate network share and carried to various meetings to discuss on an unencrypted USB stick.

Storage Risk Scores	
E-mail	5
Enterprise Vault	1
Employee desktop	3
Desktop (encrypted)	1
Notebook	10
Notebook (encrypted)	3
Mobile Device	6
Network Share	3
SharePoint	1
Other Data Base	5
SAP	1
Paper	3
External HDD	9
External HDD (encrypted)	3
Flash Drive	12
Flash Drive (encrypted)	3
CD/DVD	3
Personal Cloud	6
Enterprise Cloud solution	1
Scanned images	3

Storage Risk
Score = 30

Now Connect the Dots: Transport Risk

- The Excel file is created on an HR Analyst's unencrypted laptop, shared with various managers via e-mail. It is stored on a corporate network share and carried to various meetings to discuss on an unencrypted USB stick.

Transport Risks	
Email	8
Voltage (encrypted mail)	1
Encrypted Zip	1
FTP	15
SFTP	2
BizTalk (automated file transfer)	1
Hightail (Enterprise managed Secure FTP)	1
HTTP	15
HTTPS	5
VPN	1
Wire	1
Fax	1

Transport Risk
Score = 8

Now Connect the dots: Plotting

- **Content risk score** = 61
- **Usage Risk Score** = 38
 - Storage risk score = 30
 - Transport risk score = 8
- **TOTAL RISK Score** = 99

HR Raises
Record



Usage Risk Score

Now Connect the dots: 2 Choices

Usage Risk Score

- **Content risk score** = 61
- **Usage Risk Score** = 38
 - Storage risk score = 30
 - Transport risk score = 8
- **TOTAL RISK Score** = 99

Lower the Content Risk Score:
eliminate data fields that may not be necessary, mask certain fields

HR Raises Record

Lower the Usage Risk Score:
Provide shared, secured, managed, SharePoint storage. Provide encrypted USB if required.

Lowering Risk:

- **Prior to Project:** The Excel file is created on an HR Analyst's unencrypted laptop, shared with various managers via e-mail. It is stored on a corporate network share and carried to various meetings to discuss on an unencrypted USB stick.
- **After Project:** The Excel file is created on an HR Analyst's encrypted laptop, shared with various managers via SharePoint Link. SSN, home address and DOB have been removed. It is stored on SharePoint and carried to various meetings to discuss on an encrypted USB stick.



Total Risk
Score = 99



New Total
Risk Score
= ??

Now connect the dots – New Content Risk

- Human Resources has an Excel spreadsheet that has the following contents:
 - Names of all employees
 - ~~Date of Birth~~
 - ~~Home Address~~
 - ~~Social Security numbers~~
 - Current salary with new salary and bonus information.
 - Employee bank routing numbers

Personal Identifiable Information	
Bank Information	1
Credit Report	3
General Individual Info (name, address, DOB together)	20
ID or Log In name	1
Individual Medical Information	40
Individual Names	1
Individual's Payment Card Information	5
Individual's PIN	1
Individual's SSN or FEIN's	20
Individual's Salary / Benefit	20

PII Content Risk
 Score was = 62
 Now = 22

Now Connect the Dots: Storage Risk

- The Excel file is created on an HR Analyst's encrypted laptop, shared with various managers SharePoint Link. It is stored on SharePoint and carried to various meetings to discuss on an encrypted USB stick.

Storage Risk Scores	
E-mail	5
Enterprise Vault	1
Employee desktop	3
Desktop (encrypted)	1
Notebook	10
Notebook (encrypted)	3
Mobile Device	6
Network Share	3
SharePoint	1
Other Data Base	5
SAP	1
Paper	3
External HDD	9
External HDD (encrypted)	3
Flash Drive	12
Flash Drive (encrypted)	3
CD/DVD	3
Personal Cloud	6
Enterprise Cloud solution	1
Scanned images	3

Storage Risk
Score was = 30
Score now = 7

Now Connect the Dots: Transport Risk

- The Excel file is created on an HR Analyst's encrypted laptop, shared with various managers SharePoint Link. It is stored on SharePoint and carried to various meetings to discuss on an encrypted USB stick.

Transport Risks	
Email	8
Voltage (encrypted mail)	1
Encrypted Zip	1
FTP	15
SFTP	2
BizTalk (automated file transfer)	1
Hightail (Enterprise managed Secure FTP)	1
HTTP	15
HTTPS	5
VPN	1
Wire	1
Fax	1



Transport Risk
Score = 1

Changing the risk

Usage Risk Score

- **Content risk score** = 61
- **Usage Risk Score** = 38
 - Storage risk score = 30
 - Transport risk score = 8
- **ORIGINAL TOTAL RISK Score** = 99

- **Content risk score** = 22
- **Usage Risk Score** = 9
 - Storage risk score = 8
 - Transport risk score = 1
- **ORIGINAL TOTAL RISK Score** = 31

Removed SSN, DOB and Home Address

HR Raises Record



Removed transport by e-mail. Removed storage of Network Share, added SharePoint and Encrypted USB drive.

Four Possible Movements

USAGE RISK

You can only change the
Content Risk

You can only change the
Usage Risk

You can change the
Content and Usage Risk

You cant change what is in
the record or how it is
used.

Measuring the current risk

Usage Risk Score

Each data point now has a score:

- Either a combined score of Content risk + Usage risk
- The organization can now have an average score
- HR risk score = 109
- Finance risk score = 89
- M & A risk score = 89
- Sales risk score = 68
- Recruiting risk score = 59
- **AVERAGE Risk Score = 82.9**

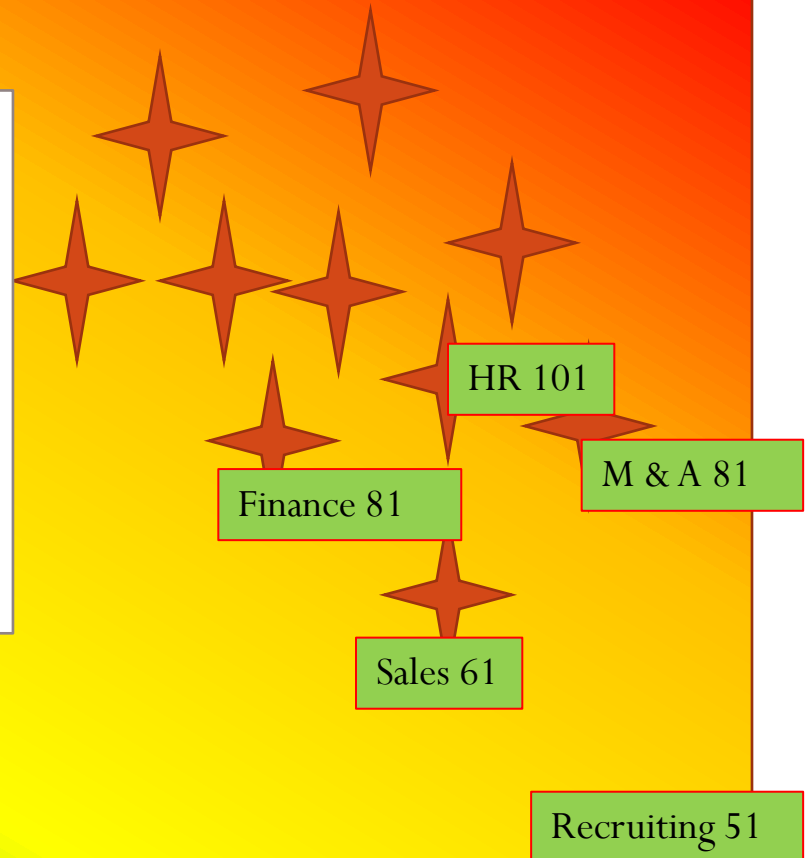


Changing the risk (en masse)

Usage Risk Score

Adding new technology can change the risk score of many records at once:

- Encrypted hard drives, USB thumb drives and encrypted portable HDD
- Move traditional file shares to SharePoint (send links not files)
- Move to more secure data center
- NEW AVERAGE RISK SCORE = 74.8




Prioritizing your efforts by content:

Group	Record	PII RISK	CPI RISK	Total Content Risk Score	Storage Risk Score	Transport Risk Score	Total Usage Risk Score	Aggregate Risk Score
Legal	Pending Litigation files	91	109	200	64	42	106	306
HR	Employee Investigation Files	42	109	151	24	24	48	199
R & D	Pending Product Development Projects	22	129	151	27	11	38	189
Sales	Forecasts	22	120	142	11	2	13	155
Admin.	DR / BC Plans	12	74	86	34	25	59	145
Facilities	Facility Plans	6	43	49	22	25	47	96



Prioritizing your efforts by Group:

Group	AVG PII RISK	AVG CPI RISK	AVG Content Risk Score	AVG Storage Risk Score	AVG Transport Risk Score	AVG Usage Risk Score	AVG Aggregate Risk Score
Legal	85	109	194	64	42	106	300
R & D	15	125	140	45	30	75	215
HR	109	45	154	20	21	41	195
Sales	12	120	142	11	2	13	155
Admin.	6	66	72	34	25	59	131
Facilities	4	34	38	22	25	47	95



Doing this with limited \$\$\$

- Soft Costs:
 - Employee time
 - Analyst time
 - Contacts within departments
 - Remediation teams
 - Database
 - Excel works just fine
 - Process Improvements (i.e. handling, storing etc.)

*Although,
these are also
possible
outcomes*

- Potential Real Costs (not cheap):
 - Remediation tooling (i.e. encrypted devices)
 - Formalize risk management tooling (i.e. Archer)
 - Employee Awareness Program
 - Dedicated FTE for larger Orgs

Final Notes:

- What this is not:
 - This is not an audit, it will not satisfy any formal inquiries
 - This is not a full assessment
 - This is not a tool dependent, technical review
- What this is:
 - This is a means to get your handle on your situation
 - This helps prioritize your efforts:
 - to dig deeper
 - to prioritize assessments and audits
 - to remediate issues.
 - A means to quantify your situation through a risk management strategy.
 - This can be maintained as an ongoing process with the right resources.

Thanks for your attendance !

Darrell Berninger

Darrell.beringer@gmail.com

Tim Plona

Tim.plona@plimtuna.com