

# Archiver le web pour les artistes : **Atelier Webrecorder**

**Artexte (webinaire) | 13 décembre 2018**

**Anna Perricci**

Directrice associée, Partenariats stratégiques

Webrecorder à Rhizome

Rhizome au New Museum

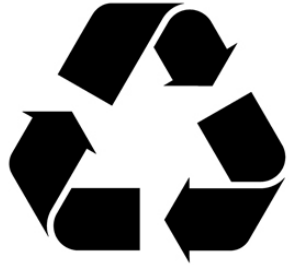
[Anna.Perricci@rhizome.org](mailto:Anna.Perricci@rhizome.org)

# Pendant cette séance, les participants vont :

- Apprendre comment utiliser un logiciel convivial, au code source ouvert et accessible via un navigateur Web pour capturer du contenu Web et rendre des collections accessibles sans frais (Webrecorder.io)
- Obtenir des informations sur les options de téléchargement des archives Web (fichiers WARC) et les logiciels permettant d'accéder aux archives Web (Webrecorder Player)
- Découvrir comment décrire, gérer et partager leurs collections Web en ligne (via Webrecorder.io)



## Web archiving is a multi-step process



SAA  
DAS  
course

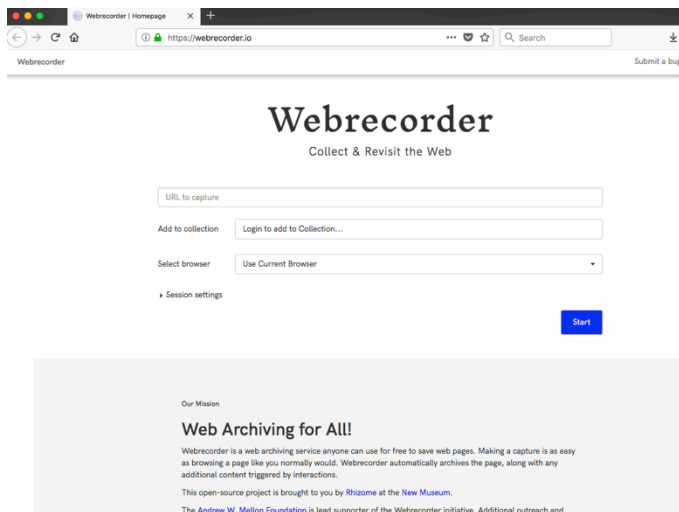
- Collection development and planning
- Selection
- Permissions
- Harvesting
- Description
- Access
- Long-term preservation

Les diapositives relatives aux bases de l'archivage Web sont disponibles à :

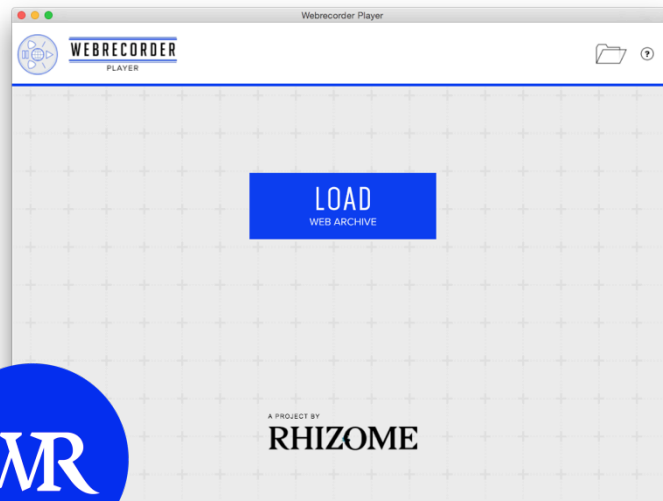
<https://www.slideshare.net/annaperricci/web-archiving-intro-circa-2015>

# À propos de Webrecorder

Créez des captures interactives de haute fidélité de toutes les pages Web que vous parcourez



<http://webrecorder.io>



Webrecorder Player App



# Le projet Webrecorder

- Outils faciles à utiliser et robustes
- Gratuit
- Complètement *open source*
- Utilise des standards ouverts
- Communauté d'utilisateurs grandissante
- Évolue rapidement

*A project by*

**RHIZOME**

*with generous support from*

THE  
ANDREW W.  
**MELLON**  
FOUNDATION

# L'Équipe Webrecorder



**Dragan Espenschied**  
Directeur de la  
conservation, Rhizome



**Ilya Kreymer**  
Directeur software



**Mark Beasley**  
Développeur senior,  
spécialiste du Front-End



**Pat Shiu**  
Directrice associée,  
Design



**Anna Perricci**  
Directrice associée,  
Partenariats stratégiques



**John Berlin**  
Développeur senior,  
spécialiste du Back-  
End

# Archivage Web haute fidélité

- Fidélité = similitude entre l'original et la capture
- Capture de toute page Web chargée dans le navigateur
- Archive le contenu web interactif (disponible uniquement après saisie du code d'utilisateur)
- Même système pour la capture et la navigation (navigateur Web)



Webrecorder crée des archives Web de haute  
fidélité **comprenant** des  
éléments complexes dont le contenu  
interactif, **qui ne sont pas normalement pris**  
**en charge** par les moissonneurs Web.

# Moissonner à l'échelle humaine

- La collecte est réalisée attentivement par une personne via le navigateur Web, une page à la fois.
- Il est possible d'importer et de bonifier des collections créées par les robots moissonneurs.
- N'importe qui peut utiliser Webrecorder: l'archivage Web pour tous!

**La récompense d'une capture minutieuse est une représentation fidèle de l'original**

# Création de comptes : optionnel

- Il n'est pas nécessaire d'avoir un compte pour effectuer des captures de contenu Web
- Les utilisateurs peuvent **télécharger les captures** immédiatement après le moissonnage (en fichier WARC) et les **sauvegarder localement**
- Pour avoir un **accès continu** aux contenus en ligne et pouvoir **bonifier une collection**, il est nécessaire de se **connecter** à un compte gratuit

# Capter / parcourir

- Webrecorder.io est utilisé pour réaliser des **captures** interactives de pages Web que les utilisateurs voient lors de la collecte, mais ce n'est pas un logiciel d'enregistrement sur écran capable de lire des enregistrements comme une vidéo.
- **Parcourir** signifie que vous pouvez accéder au contenu capturé dans l'archive Web et le parcourir de manière interactive, comme sur le Web en direct.

# Parcourir une archive liée

- Chaque collection est une unité distincte; vous ne pouvez donc naviguer que dans le contenu d'une seule collection à la fois.
- Cela donne un contrôle sur le contenu conservé, bien que les limites de la collection peuvent parfois être rapidement identifiées.

# Options d'accès et de partage

- Les collections et les listes créées par un utilisateur peuvent être **maintenues privées ou rendues publiques** via Webrecorder.io
- Les collections publiques et les listes publiques peuvent être consultées par tous
- Des contrôles d'accès plus détaillés sont envisagés actuellement
- Des fonctionnalités d'édition partagées seront probablement disponibles dans le futur

# Pourquoi est-il difficile de conserver des sites Web en constante évolution (i.e. les nouvelles)?

- Les sites d'informations repoussent les limites de la technologie actuelle
- Contenu interactif!
- Beaucoup de contenu audio/vidéo et de contenu publicitaire
- Les robots traditionnels ne capturent pas Javascript
- Évolution très rapide

# Applications mobiles?

- Webrecorder ne convient pas aux applications mobiles, qui sont souvent des navigateurs Web personnalisés pour un ou plusieurs sites Web particuliers.
- Les sites Web riches en médias, comme les « applications d'actualité », peuvent être capturés par Webrecorder. Par contre, l'archivage à échelle humaine signifie aussi que la capture peut prendre beaucoup de temps.



# Qu'en est-il des médias sociaux?

- Webrecorder peut capturer le contenu de sites de médias sociaux et fonctionne particulièrement bien avec Instagram et Twitter.
- Certains sites Web proposent un contenu **personnalisé pour chaque utilisateur**.
- Webrecorder peut capturer le contenu **que vous voyez lorsque vous êtes connecté à un profil de réseau social**.
  - *Webrecorder.io est conçu pour ne pas capturer ni conserver vos identifiants de connexion aux médias sociaux.*

# Patching

Webrecorder / annaperricci / Steph Davidson's Excellent Design Work

Submit a bug

Browsing

Current Browser

https://www.nytimes.com/2015/05/03/magazine/what-happened-when-i-joked-about-the-president-of-ecuador.f

7/4

Section Navigator

Item lists

[Collection Index](#)

Steph Davidson's Excellent Design Work (94)

The Automation of Sectarianism: .

<https://marcowenjones.wordpress.com/2016/06/2>

Saudi Women Free After 73 Days

<https://www.nytimes.com/2015/02/13/world/mid>

Loujain Alhathloul

<http://www.loujainhathloul.com/>

Pro-Government Twitter Bots Try

<https://www.wired.com/2015/08/pro-government>

**What Happened When I Joked At**

<https://www.nytimes.com/2015/05/03/magazine/>

Packrat: Seven Years of a South A



## Resource not Found

Try to patch

Still missing? Report a bug.

The url <http://radioambulante.org/audio/correa-vs-crudo> was not found in the archive.

# Exemples de collections

1. <https://webrecorder.io/demo/narrative-archives>
2. <https://webrecorder.io/imamuseum>
3. <https://webrecorder.io/ClarkArtLibrary>
4. <https://webrecorder.io/ivypluslibraries/state-elections-web-archive>
5. <https://webrecorder.io/sup/enchanting-the-desert>
6. <https://webrecorder.io/wrsc/snowfall---ny-times>
7. <https://webrecorder.io/demo/narrative-archives/list/flotus-x-td4w/b9/20170104093624/http://www.td4wbutton.com/>

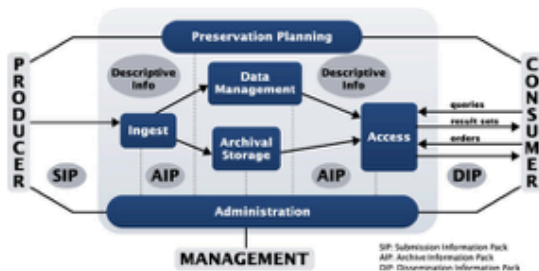
# Navigateurs préconfigurés (« *remote* »)

- Utilisation de navigateurs préconfigurés pour capturer et parcourir le contenu Web qui pourrait ne pas être pris en charge par les futurs navigateurs Web

ex. : Flash

- L'accès à un site Web via un navigateur préconfiguré garantit une plus grande fidélité à l'aspect et à l'expérience utilisateur d'origine.
- Les navigateurs utilisent le mode proxy HTTP = une fidélité encore plus grande

## Interactive Production : Preservation of Experience



**SIP** : WARC file grab by Webrecorder (Rhizome)

**AIP** : Archived using Atempo Digital Archive with all rules to ensure long term access

**DIP** : WARC with live index and playback with **Webrecorder** (Rhizome) in NFB infrastructure

Development Collaboration between NFB and Rhizome

# La description, la gestion et le partage

# Narrative Archives

We've reached the twilight hours of Barack Obama's presidency.

A typical transition period requires many handovers, but this year, one of these is an altogether new endeavor for a White House: turning the sitting President's social media presence over to the new office holder. These Twitter, Instagram, Vine, Facebook, etc. accounts will be wiped and exported, some to archival accounts—for instance, [Obama's tweets will live on at @POTUS44](#)—and some to a public data archive.

On the occasion of this first social media transition, the White House issued a public call to action, asking Americans to try their hand at breathing life into the huge amount of data that is President Obama's eight years of social media presence. We were thrilled to respond. (With others, too!) The data archive, vast and disparate as it is, is largely unintelligible; what we proposed was using our built in-house [Webrecorder](#) tool to give it some narrative shape, some

## Lists in this Collection

### 📖 FLOTUS x TD4W

Lil Jon and Michelle Obama's inadvertent collaboration on the First Lady's now-infamous "turnip for what" Vine.

[View list \(14 Bookmarks\) »](#)

### 📖 Thanks Obama

Exploring the satirization of conservative disappointment with Obama, mapping the *Thanks Obama* meme's progression from its earnest right-wing origins all the way to its self-aware usage by Obama himself.

[View list \(1 Bookmark\) »](#)

### 📖 LoveWins

# Steph Davidson's Excellent Design Work

This collection contains pieces published on Bloomberg.

+ New Session



Collection Cover

Public

Pages (94)



LISTS (4 Public)

EDIT +

Title list for pieces in collection (6)



What is Code? (1)



How to Spend Your Bonus (1)



A Global Guide to State-Sponsored Troll... (1)



## Pages

INDEX	TIMESTAMP	PAGE TITLE	URL
1	7/27/2018, 1:19:24 PM	<a href="#">The Automation of Sectarianism: Are Twitt</a>	<a href="https://marcowenjones.wordp">https://marcowenjones.wordp</a>
2	7/27/2018, 1:13:49 PM	<a href="#">Saudi Women Free After 73 Days in Jail fo</a>	<a href="https://www.nytimes.com/201">https://www.nytimes.com/201</a>
3	7/27/2018, 1:13:39 PM	<a href="#">Loujain Alhathloul</a>	<a href="http://www.loujainhathloul.co">http://www.loujainhathloul.co</a>
4	7/27/2018, 1:11:33 PM	<a href="#">Pro-Government Twitter Bots Try to Hush</a>	<a href="https://www.wired.com/2015">https://www.wired.com/2015</a>
5	7/27/2018, 1:09:26 PM	<a href="#">What Happened When I Joked About the I</a>	<a href="https://www.nytimes.com/201">https://www.nytimes.com/201</a>
6	7/27/2018, 1:09:20 PM	<a href="#">Packrat: Seven Years of a South American</a>	<a href="https://citizenlab.ca/2015/12">https://citizenlab.ca/2015/12</a>
7	7/27/2018, 1:08:07 PM	<a href="#">Ecuador Transparente - Godwin</a>	<a href="https://ecuadortransparente.o">https://ecuadortransparente.o</a>
8	7/27/2018, 1:07:06 PM	<a href="#">Amazon.com: Swati Chaturvedi: Books</a>	<a href="https://www.amazon.com/s/re">https://www.amazon.com/s/re</a>
9	7/27/2018, 1:06:36 PM	<a href="#">I am a Troll: Inside the Secret World of the</a>	<a href="https://www.amazon.com/Tro">https://www.amazon.com/Tro</a>
10	7/27/2018, 1:05:47 PM	<a href="#">Leaked Documents Show That Ethiopia's R</a>	<a href="https://advox.globalvoices.org">https://advox.globalvoices.org</a>
11	7/27/2018, 1:05:26 PM	<a href="#">Amazon.com: Customer reviews: I am a Tr</a>	<a href="https://www.amazon.com/Tro">https://www.amazon.com/Tro</a>
12	7/27/2018, 1:04:08 PM	<a href="#">I am a Troll: Inside the Secret World of the</a>	<a href="https://www.amazon.com/Tro">https://www.amazon.com/Tro</a>
13	7/27/2018, 1:02:44 PM	<a href="#">Automated sectarianism and pro-Saudi prc</a>	<a href="https://exposingtheinvisible.or">https://exposingtheinvisible.or</a>



+ New Session ...

Collection Cover Public

Pages (94) lock icon

LISTS (4 Public) EDIT +

- Title list for pieces in collection (6) ●
- What is Code? (1) ●
- How to Spend Your Bonus (1) ●
- A Global Guide to State-Sponsored Troll... (1) ●

2 7/27/2018, 1:13:49 PM Saudi Women Free After 73 Days in Jail to

3 7/27/2018, 1:13:39 PM Loujain Alhathloul

4 7/27/2018

5 7/27/2018

6 7/27/2018

7 7/27/2018

8 7/27/2018

9 7/27/2018

10 7/27/2018, 1:05:47 PM Leaked Documents Show That Ethiopia's R

11 7/27/2018, 1:05:26 PM Amazon.com: Customer reviews: I am a Tr

12 7/27/2018, 1:04:08 PM I am a Troll: Inside the Secret World of the

x New list| +

- Title list for pieces in collection
- What is Code?
- How to Spend Your Bonus
- A Global Guide to State-Sponsored Trolling

Done

# Steph Davidson's Excellent Design Work

This collection contains pieces published on Bloomberg.com

+ New Session



Collection Cover

Public

Pages (94)

LISTS (4 Public)

EDIT +

Title list for pieces in collection (6)

What is Code? (1)

How to Spend Your Bonus (1)

A Global Guide to State-Sponsored Troll... (1)

INDEX

TIMESTAMP

PAGE TITLE

URL

7/27/2018 1:19:24 PM

The Automation of Sectarianism: Are Twitt

https://marcove

## Edit Collection Details

Collection Name:

Steph Davidson's Excellent Design Work

Collection Description:

**B**

*I*

U

~~S~~

**M+**

This collection contains pieces published on Bloomberg.com that were illustrated by Steph Davidson. This is not a comprehensive collection of her work but these examples are quite excellent!

Cancel

Save

## Steph Davidson's Excellent Design Work

This collection contains pieces published on Bloom

+ New Session ...

Collection Cover

Public

Pages (94)

LISTS (4 Public)

EDIT +

- Title list for pieces in collection (6)
- What is Code? (1)
- How to Spend Your Bonus (1)
- A Global Guide to State-Sponsored Troll... (1)

### Metadata

BOOKMARK 1 OF 1

Paul Ford: What Is Code? | Bloomberg

### DESCRIPTION

**B** *I* U ~~S~~ 

Engaging and educational!

## What is Code?

Show Description

REM...	INDEX	TIMESTAMP	BOOKMARK TITLE	URL	CAPTURE B
	1	10/30/2017, 2:36:04 PM	<a href="#">Paul Ford: What Is Code?   Bloomberg</a>	<a href="https://www.bloomberg.com/graphics/2015-p">https://www.bloomberg.com/graphics/2015-p</a>	 Firefox v49

 COLLECTION by [annaperricci](#)

## Steph Davidson's Excellent Design Work


This collection contains pieces published on Bloomberg

+ New Session

...

Collection Cover

 Public

 Pages (94)




LISTS (2 Public)

EDIT +

 Title list for pieces in collection (6)



 What is Code? (1)



 How to Spend Your Bonus (1)

PRIVATE 

 A Global Guide to State-Sponsored Troll... (1)



## Steph Davidson's Excellent Design Work

This collection contains pieces published on Bloomberg

+ New Session



New Session

Cover

Manage Sessions

Upload To Collection

Download Collection

Delete Collection

Tour New Features

Help

Collection Cover

Pages (95)

LISTS (4 Public)

Title list for pie

What is Code?

How to Spend

A Global Guide

## Pages

INDEX	TIMESTAMP	PAGE TITLE
1	9/18/2018, 6:22:01 PM	<a href="#">Correa vs Crudo - Radio Amb</a>
2	7/27/2018, 1:19:24 PM	<a href="#">The Automation of Sectarianis</a>
3	7/27/2018, 1:13:49 PM	<a href="#">Saudi Women Free After 73 D</a>
4	7/27/2018, 1:13:39 PM	<a href="#">Loujain Alhathloul</a>
5	7/27/2018, 1:11:33 PM	<a href="#">Pro-Government Twitter Bots</a>
6	7/27/2018, 1:09:26 PM	<a href="#">What Happened When I Joker</a>
7	7/27/2018, 1:09:20 PM	<a href="#">Packrat: Seven Years of a Sout</a>
8	7/27/2018, 1:08:07 PM	<a href="#">Ecuador Transparente - Godw</a>
9	7/27/2018, 1:07:06 PM	<a href="#">Amazon.com: Swati Chaturver</a>
10	7/27/2018, 1:06:36 PM	<a href="#">I am a Troll: Inside the Secret</a>
11	7/27/2018, 1:05:47 PM	<a href="#">Leaked Documents Show That</a>
12	7/27/2018, 1:05:26 PM	<a href="#">Amazon.com: Customer review</a>
13	7/27/2018, 1:04:08 PM	<a href="#">I am a Troll: Inside the Secret</a>

## MANAGE SESSIONS

Remove content from collection by **deleting sessions**. Add content by **uploading warcs**.


Please note that it is not possible to delete individual pages.

### OVERVIEW

## Steph Davidson's Excellent Design Work

Created on 10/30/2017, 6:35:47 PM

26 sessions over approximately **10months 23days 3hrs 47mins** containing **95 pages**. Total capture time is **2months 13days 34mins** and total data size is **501.91 MB**.



 Delete Entire Collection

 Download Collection as WARC

 Upload WARC to Collection

## Sessions

Expand All

Patch	1 Pages	1min	3.66 MB
<div data-bbox="241 709 285 813"><b>18</b> Sep TUE 2018</div> <div data-bbox="353 709 714 753">Session Notes Add notes about this session. Visible only to you.</div> <div data-bbox="546 769 602 791">EDIT</div> <div data-bbox="227 857 299 879">Delete</div> <div data-bbox="218 906 309 928">Download</div>	<div data-bbox="801 709 923 726">Session Pages (1)</div> <div data-bbox="817 742 1081 758">Correa vs Crudo - Radio Ambulante</div> <div data-bbox="1342 742 1719 758"><a href="http://radioambulante.org/audio/correa-vs-crudo2">http://radioambulante.org/audio/correa-vs-crudo2</a></div>		
<div data-bbox="218 988 285 1010"> </div> Patch	1 Pages	2mins	2.49 MB

COLLECTION by [annaperricci](#)

## \_ Steph Davidson's Excellent Design Work

This collection contains pieces published on Bloom:

+ New Session



Collection Cover

New Session

Pages (95)

Cover

LISTS (4 Public)

Manage Sessions

Title list for pie

Upload To Collection

What is Code?

Download Collection

How to Spend

Delete Collection

A Global Guide

Tour New Features

Help

## Pages

INDEX	TIMESTAMP	PAGE TITLE
1	9/18/2018, 6:22:01 PM	<a href="#">Correa vs Crudi</a>
2	7/27/2018, 1:19:24 PM	<a href="#">The Automator</a>
3	7/27/2018, 1:13:49 PM	<a href="#">Saudi Women F</a>
4	7/27/2018, 1:13:39 PM	<a href="#">Loujain Alathlk</a>
5	7/27/2018, 1:11:33 PM	<a href="#">Pro-Governmer</a>
6	7/27/2018, 1:09:26 PM	<a href="#">What Happene</a>
7	7/27/2018, 1:09:20 PM	<a href="#">Packrat: Seven</a>
8	7/27/2018, 1:08:07 PM	<a href="#">Ecuador Transp</a>
9	7/27/2018, 1:07:06 PM	<a href="#">Amazon.com: S</a>
10	7/27/2018, 1:06:36 PM	<a href="#">I am a Troll: Ins</a>
11	7/27/2018, 1:05:47 PM	<a href="#">Leaked Docume</a>
12	7/27/2018, 1:05:26 PM	<a href="#">Amazon.com: C</a>
13	7/27/2018, 1:04:08 PM	<a href="#">I am a Troll: Ins</a>
14	7/27/2018, 1:02:44 PM	<a href="#">Automated cont</a>

# Webrecorder Player

- Application de bureau pour OSX, Windows et Linux
- Application conviviale pour parcourir n'importe quelle archive Web (enregistrée au format WARC standard)
- Vous pouvez parcourir les archives Web hors ligne, aucune connexion Internet n'est requise!



# Téléchargez Webrecorder Player

<https://github.com/webrecorder/webrecorderplayer-electron>

# Gérance

- Les fichiers stockés en format WARC (par exemple, à court ou à moyen terme) peuvent être traités comme n'importe quel format de fichier normalisé et largement adopté
- Comme toujours, toute migration ou dépendance peut présenter des risques
- Les logiciels qui peuvent ouvrir et rendre navigables les fichiers de format WARC seront importants à considérer pour planifier l'accès à long terme
- Des flux de travail pour intégrer les fichiers de format WARC dans les dépôts d'archives sont en cours de développement (y compris comment décrire/assigner des métadonnées à ces fichiers)

# Quelles sont les prochaines étapes pour Webrecorder?

- Des services basés sur des abonnements, notamment pour obtenir plus d'espace de stockage et des outils de collecte automatisés, sont à venir en 2019
- Mettre en œuvre un plan à plusieurs facettes pour assurer la pérennité de Webrecorder visant à générer des revenus tout en maintenant une version *open source* entièrement fonctionnelle et gratuite
- Plus d'options pour gérer et partager les collections

Conclusion,  
discussion,  
questions/réponses,  
atelier?

N'hésitez pas à nous contacter :  
Anna.Perricci@rhizome.org  
Support@Webrecorder.io

**Vous pouvez également  
contacter Artexte :  
eartexte@artexte.ca**

# Webrecorder

*Thank you*

*Merci !*

*A project by*

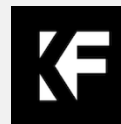
**RHIZOME**

*with generous support from*

THE  
ANDREW W.

**MELLON**  
FOUNDATION

*additional outreach support*



**KNIGHT  
FOUNDATION**