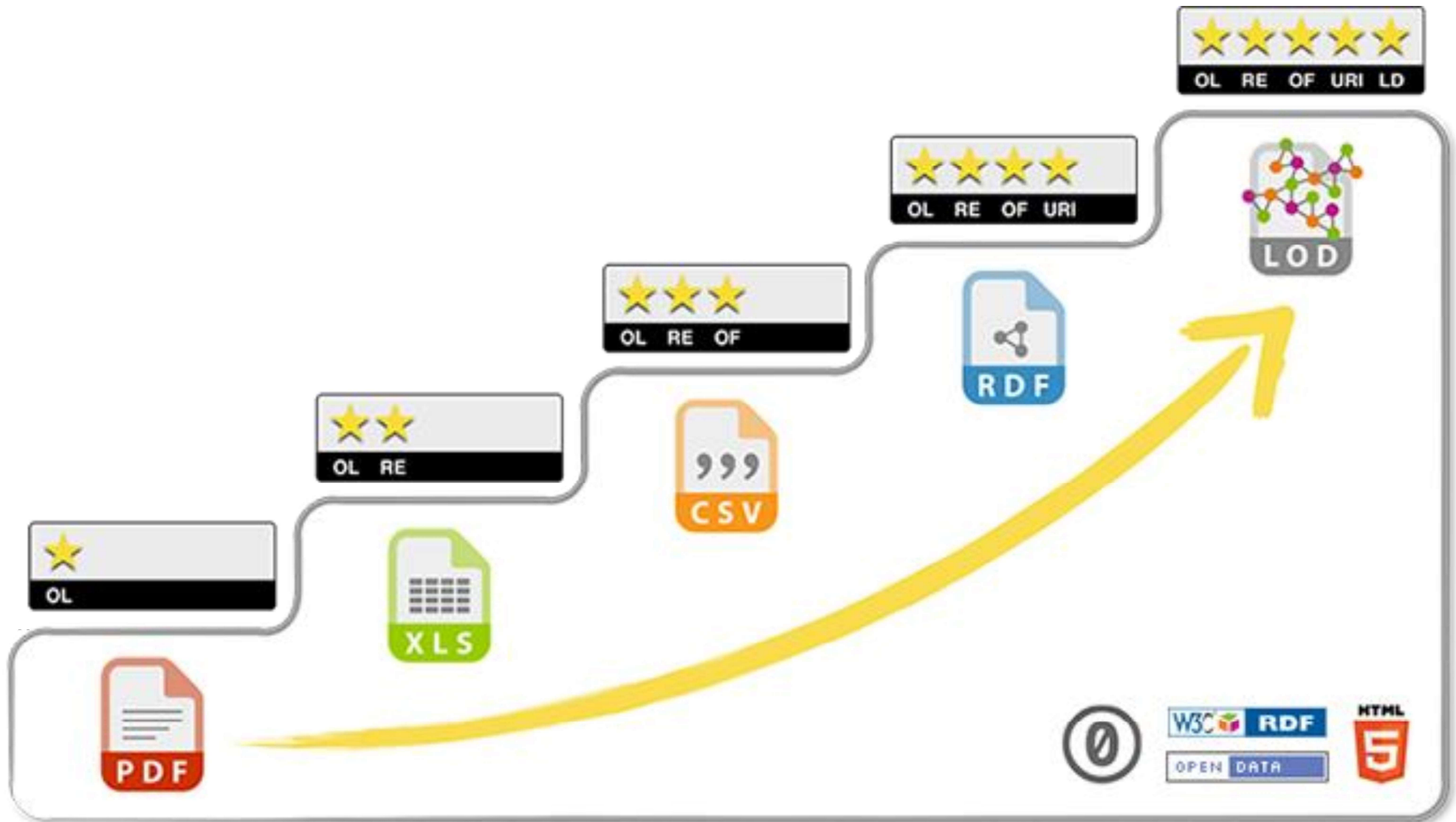


5 ★ Open Data



Source: <http://5stardata.info/>

- ★ Publiez vos données sur le Web avec une licence ouverte (peu importe le format)
- ★★ Publiez vos données dans un format structuré (par exemple Excel au lieu d'une image tirée d'un scanner)
- ★★★ Utilisez un format ouvert (par exemple CSV au lieu d'Excel)
- ★★★★ Utilisez des URIs pour chaque objet, afin que n'importe qui puisse cibler vos données
- ★★★★★ Liez vos données avec d'autres données pour enrichir le contexte



Publiez vos données sur le Web avec une licence ouverte
(peu importe le format)

Les utilisateurs peuvent

- Voir/imprimer les données
- Sauvegarder les données
- Réutiliser les données
- Redistribuer les données

Pour les fournisseurs

- La publication est facile;
- Les données sont disponibles.

Problème

Les données restent **prisonnières** du document. Récupérer les données pour les réutiliser peut en particulier s'avérer difficile.



Publiez vos données dans un format structuré
(par exemple Excel au lieu d'une image tirée d'un scanner)

Les utilisateurs peuvent

- Analyser les données avec un logiciel propriétaire
- Exporter les données dans un format structuré

Pour les fournisseurs

- La publication est toujours aussi facile

Problème

L'analyse de données nécessite l'usage
d'un logiciel propriétaire



Utilisez un format ouvert
(par exemple CSV au lieu d'Excel)

Les utilisateurs peuvent

- Manipuler les données sans avoir recours à un logiciel propriétaire

Pour les fournisseurs

- La publication est encore facile
- Des convertisseurs de données peuvent avoir à être utilisés

Problème

Les données sont encore isolées.
Elles sont accessibles **via le Web**,
mais elles ne font pas encore
partie du Web.



Utilisez des URIs pour chaque objet, afin que n'importe qui puisse cibler vos données

Les utilisateurs peuvent

- Lier vers les données depuis une page Web
- Enregistrer un signet
- Combiner les données avec d'autres données sans problème
- Utiliser des outils tiers pour analyser une partie des données... à condition de comprendre la structure d'un « graphe » RDF

Pour les fournisseurs

- La publication plus fine permet d'optimiser l'accès aux données (caching, load balancing, API)
- La publication est plus exigeante, il faut créer des URIs et penser à la représentation des données

Problème

Les données font bien partie du Web, mais peuvent toujours être isolées



Liez vos données avec d'autres données pour enrichir le contexte

Les utilisateurs peuvent

- Découvrir des données connexes
- Comprendre le schéma de données plus facilement

Pour les fournisseurs

- Les données peuvent être découvertes
- Les données font partie d'un tout, leur valeur ajoutée augmente en conséquence

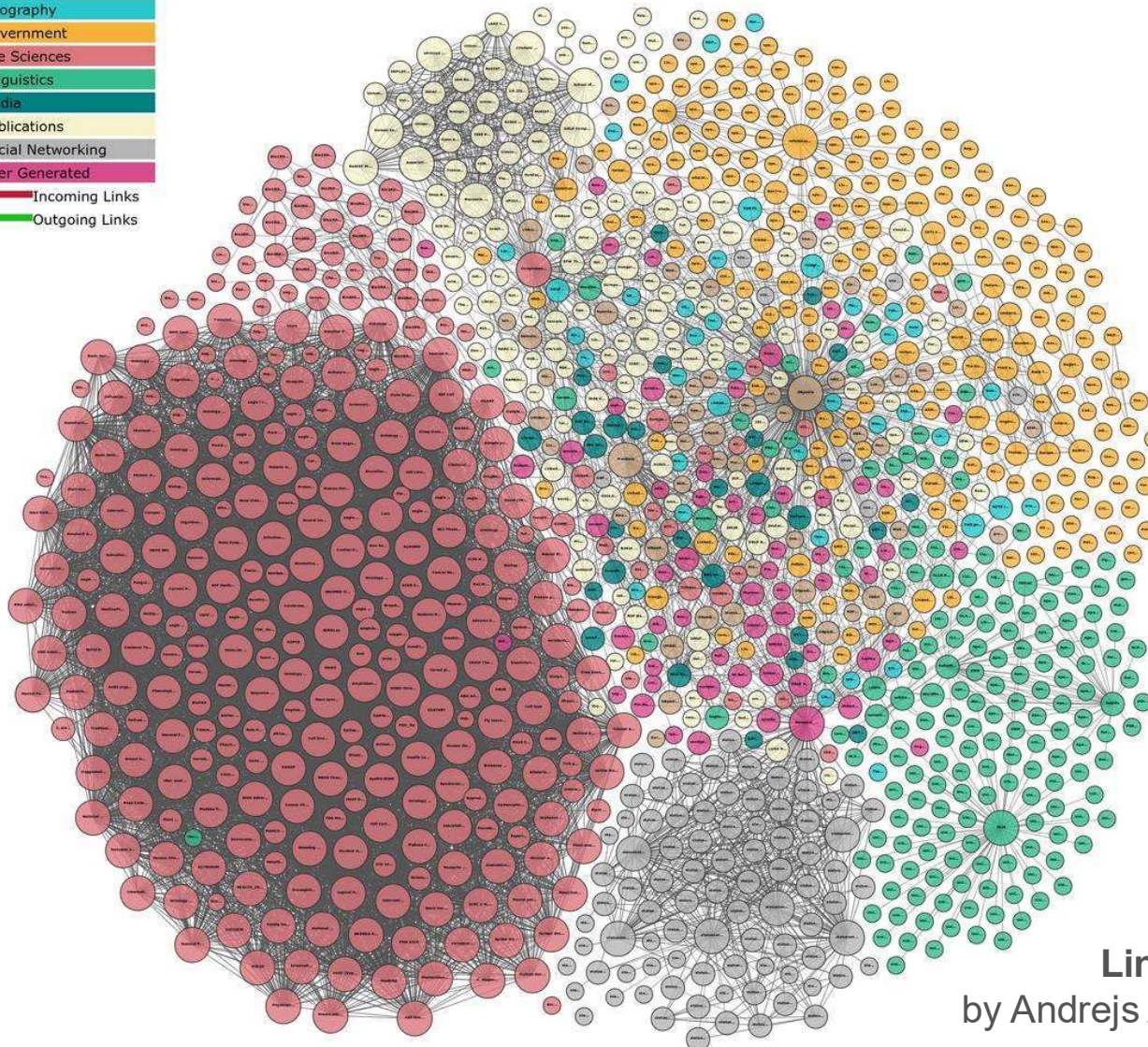
Problèmes

Les données sont désormais liées dans une trame globale.
Comment distinguer le faux du vrai?
Comment gérer les liens brisés?
Quels vocabulaires utiliser?

5 ★ Open Data

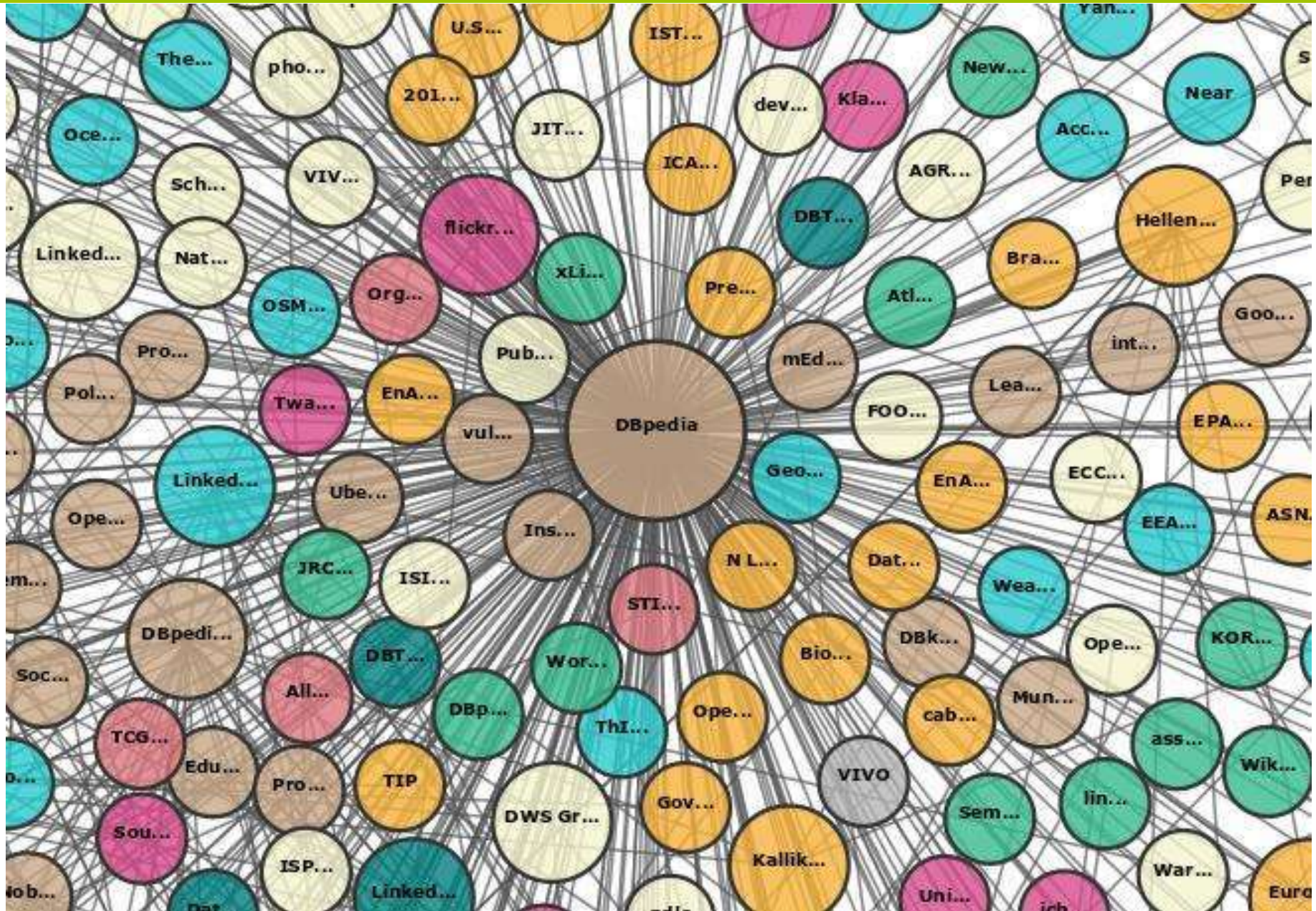
Legend

Cross Domain
Geography
Government
Life Sciences
Linguistics
Media
Publications
Social Networking
User Generated
Incoming Links
Outgoing Links



Linking Open Data cloud diagram 2017
by Andrejs Abele, John P. McCrae, Paul Buitelaar,
Anja Jentsch and Richard Cyganiak.
<http://lod-cloud.net/>

5 ★ Open Data



`http://{base}/{type}[/{domain}][/{sector}][/{resource}]`

- L'utilisation d'**URIs** est à la base du Web
- L'utilisation d'adresses **HTTP** permet de chercher la ressource
- Les identifiants doivent être **pérennes** et **logiques**, la recherche d'informations en sera grandement facilitée, et les liens entre les données renforcés et stables.
- Prévoir une gestion des **versions** de chaque document!

7. Bonnes pratiques pour rendre les données exploitables

**Formats de fichiers:
Privilégier les formats
ouverts et « non visuels »**

ouverts et « non visuels »

- **Format ouvert:**
Exploitable sans logiciel propriétaire,
traitement automatisable
- **Format « non visuel »:**
Données destinées à être traitées par une
machine, pas à être directement
visualisées par les humains



CSV, JSON, XML, HTML, RDF, GeoTIFF,
GeoJSON, ShapeFile, etc.



PDF, DOC, XLS, PNG, GIF, JPEG, etc.

- **Format propriétaire:**
Nécessite Excel
Traitement automatique difficile
- **Format visuel:**
Ex: 37 façons de cacher les données dans
une feuille Excel
http://john.raffensperger.org/ArtOfTheSpreadsheet/Chapter09_ShowAllTheInformation.html

Attention: plusieurs CSV possibles

- Plusieurs définitions du format CSV
- Délimiteur: « , » ou « ; » ou tabulation
- Séparateur décimal: « , » ou « . »
- Problème des textes sur plusieurs lignes
- ...

- Encodé en UTF-8
- Utilise « CRLF » pour chaque fin de ligne
- Utilise « , » comme délimiteur
- Utilise « . » comme séparateur décimal
- Possède la définition des champs en en-tête (1^{ère} ligne)

Certains formats sont dérivés de formats de fichiers plus bruts (comme JSON, XML, CSV, RDF), par exemple:

- **GTFS:**
Ensemble de fichiers texte dans un ZIP avec des noms de fichiers et des structures imposées
- **OCDS (Open Contracting Data Standard):**
Format de données basé sur JSON qui impose la structure et un certain nombre de contraintes sur les données.

- **Données tabulaires**
Ex: CSV
- **Données hiérarchiques**
Ex: JSON, XML
- **Données géographiques**
Ex: GeoJSON, ShapeFile, GeoTIFF
- **Données de transport**
Ex: GTFS
- **Documents**
Ex: HTML, TXT