

Fr-SemCor

Un corpus du français annoté en Supersenses

Lucie Barque

Université Sorbonne Paris Nord & LLF

Avec : M. Candito, B. Crabbé, P. Haas, R. Huyghe, V. Segonne, D. Tribout

Séminaire TALEP
25 mars 2021, Marseille

Annotation sémantique en corpus : deux types d'annotation lexicale

(1) *Quatre électeurs écologistes déposent plainte à Paris avec **constitution**, de partie civile.*

1. Tag : sens lexicaux – utilisation d'un lexique dans lequel les sens des mots sont prédéfinis (ex. Wiktionnaire)

constitution-1 Manière dont une chose est constituée, dont ses éléments combinés forment un tout.

constitution-2 État général du corps humain, complexion.

constitution-3 Loi ou charte fondamentale qui détermine la forme du gouvernement, qui règle les droits politiques des citoyens et qui dicte la politique d'un État ou d'un groupe d'États.

constitution-4 ['en particulier'] Ordonnance, loi, règlement, législation ancienne en matière ecclésiastique.

constitution-5 ['droit'] Le fait de créer, de fonder quelque chose.

☞ Type de corpus annoté utilisé en **WSD** (semi-)supervisé (Navigli 2009)

Annotation sémantique en corpus : deux types d'annotation lexicale

(2) *Quatre électeurs écologistes déposent plainte à Paris avec **constitution**? de partie civile.*

2. Tag : **supersense** – utilisation de classes sémantiques générales, ou supersenses (ex. les 25 *Unique Beginners* de la hiérarchie des noms dans WordNet)

```
...  
noun.act          nouns denoting acts or actions  
noun.animal       nouns denoting animals  
noun.artifact     nouns denoting man-made objects  
noun.attribute    nouns denoting attributes of people and objects  
noun.body         nouns denoting body parts  
noun.cognition    nouns denoting cognitive processes and contents  
noun.communication nouns denoting communicative processes and contents  
noun.event        nouns denoting natural events  
...
```

☞ Type de corpus utilisé en **Supersense tagging** (semi-)supervisé (Ciaramita & Johnson 2003)

Corpus annotés sémantiquement

Corpus	Langue	POS	Annotation	Tagset	Tagged
SemCor 3.0 ¹	anglais	N, V, Adj	manuelle	WordNet synsets (T1)	~ 193K
SemEval-2013 ²	multilingue	N	manuelle	BalbelNet concepts (T1)	~ 3K
AQMAR Corpus ³	arabe	N, V	manuelle	WordNet UBs (T2)	~ 65K
SemDax ⁴	danois	N, V, Adj	manuelle	WordNet UBs (T2)	~ 35K
EuroSense ⁵	multilingue	N, V, Adj	automatique	BabelNet concepts (T1)	~ 123,000K

-
1. Landes *et al.* 1998
 2. Navigli *et al.*, 2013
 3. Schneider *et al.* 2012
 4. Pederson *et al.* 2016
 5. Delli Bovi *et al.* 2017

- **Fr-Semcor** : corpus du français annoté manuellement
 - ▶ À l'aide de **supersenses** adaptés des WordNet *Unique Beginners*
 - ▶ Annotation des **noms communs** uniquement
 - 👉 15 000 occurrences du corpus Sequoia (Candito & Seddah 2012)
- **Objectif**
 - ▶ Contribuer au développement d'un outil d'analyse sémantique du français en fournissant des données pour l'apprentissage semi-supervisé
 - ▶ Désambiguïsation nominale
 - ▶ Désambiguïsation verbale

1. Définition des supersenses

2. Méthode d'annotation

3. Analyse des données annotées

Conclusion

1. Définition des supersenses

Les *Unique Beginners* de WordNet

- Étiquettes issues du projet WordNet (Miller *et al.* 1990, Fellbaum 1998)
 - ▶ 25 UBs pour les noms, 15 pour les verbes, 2 pour les adjectifs
 - ▶ UBs pour les noms : Act, Animal, Artifact, Attribute, Body, Cognition, Communication, Event, Feeling, Food, Group, Location, Motive, Object, Person, Phenomenon, Plant, Possession, Process, Quantity, Relation, Shape, State, Substance, Time, Tops
 - 📖 Pas un jeu d'étiquettes construit pour l'annotation mais pour le travail lexicographique
- Étiquettes traditionnellement utilisées pour la tâche de Supersense tagging (Ciaramita & Johnson 2003)
 - ▶ Les annotations en sens lexicaux de SemCor 3.0 (Landes *et al.* 1998) sont traduites en supersenses

1. Définition des supersenses

Les *Unique Beginners* de WordNet

- Définitions sommaires

```
...  
noun.act          nouns denoting acts or actions  
noun.animal       nouns denoting animals  
noun.artifact     nouns denoting man-made objects  
noun.attribute    nouns denoting attributes of people and objects  
noun.body         nouns denoting body parts  
noun.cognition    nouns denoting cognitive processes and contents  
noun.communication nouns denoting communicative processes and contents  
noun.event        nouns denoting natural events  
...
```


1. Définition des supersenses

Les *Unique Beginners* de WordNet

- Définitions sommaires

```
...  
noun.act          nouns denoting acts or actions  
noun.animal       nouns denoting animals  
noun.artifact     nouns denoting man-made objects  
noun.attribute    nouns denoting attributes of people and objects  
noun.body         nouns denoting body parts  
noun.cognition    nouns denoting cognitive processes and contents  
noun.communication nouns denoting communicative processes and contents  
noun.event        nouns denoting natural events  
...
```

- Classes hétérogènes

- ▶ En termes ontologiques

Person, *Act* vs *Cognition* (ex. *logic*, *adeptness*, *night_vision*), *Communication*
(ex. *transmission*, *recto*, *group discussion*)

- ▶ En termes de taille

Person (11 087), *Act* (6650) vs *Shape* (341), *Motive* (42)

1. Définition des supersenses

Les Unique Beginners de WordNet

- Définitions sommaires

```
...  
noun.act          nouns denoting acts or actions  
noun.animal       nouns denoting animals  
noun.artifact     nouns denoting man-made objects  
noun.attribute    nouns denoting attributes of people and objects  
noun.body         nouns denoting body parts  
noun.cognition    nouns denoting cognitive processes and contents  
noun.communication nouns denoting communicative processes and contents  
noun.event        nouns denoting natural events  
...
```

- Classes hétérogènes

- ▶ En termes ontologiques

Person, *Act* vs *Cognition* (ex. *logic*, *adeptness*, *night_vision*), *Communication* (ex. *transmission*, *recto*, *group discussion*)

- ▶ En termes de taille

Person (11 087), *Act* (6650) vs *Shape* (341), *Motive* (42)

- Classes complémentaires

Ex. Bien que des N étiquetés *Food* correspondent également à des substances, ils ne sont pas étiquetés *Substance*

1. Définition des Supersenses

Adaptation des *Unique Beginners* pour Fr-SemCor

Supersenses Fr-SemCor
Act, Event, Process
Attribute, State, Feeling
Animal, Person
Body, Object, Plant
Cognition, Communication
Part , Quantity, Group
Artifact, Food, Institution , , Phenomenon Location, Motive, Possession, Shape, Substance, Time, Relation

- 1 Suppression de 4 WN UBs (Process, Location, Motive, Shape)
- 2 Création de 2 supersenses (Part et Institution)
- 3 Rédéfinition des contours de certaines classes

1. Définition des Supersenses

Adaptation des *Unique Beginners* pour Fr-SemCor

1 Suppression de 4 WN UBs

Process

Processus agentif (ex. *anglicisation*) → Act

Processus non agentif (ex. *biosynthèse*) → Event

Location

Lieu naturel (ex. *plage, campagne, terrain*) → Object

Lieu construit (ex. *pont, salon, ville*) → Artifact

N de localisation interne (ex. *coin, bord*) → Part

Motive, Shape

1. Définition des Supersenses

Adaptation des *Unique Beginners* pour Fr-SemCor

1 Suppression de 4 WN UBs

Process

Processus agentif (ex. *anglicisation*) → Act

Processus non agentif (ex. *biosynthèse*) → Event

Location

Lieu naturel (ex. *plage, campagne, terrain*) → Object

Lieu construit (ex. *pont, salon, ville*) → Artifact

N de localisation interne (ex. *coin, bord*) → Part

Motive, Shape

2 Ajout de 2 supersenses

Institution

Organisations, collectivités, etc (ex. *gendarmerie, syndicat, ville*)

Supersense introduit dans le projet SemDax (Pederson et al. 2016)

Part (ex. *coin, bord*)

1. Définition des Supersenses

Adaptation des *Unique Beginners* pour Fr-SemCor

3. Redéfinition de l'extension de certaines classes

Cognition

Cognition dans WordNet inclut les actes et les contenus cognitifs

Cognition dans FR-SemCor inclut seulement les contenus cognitifs, les actes cognitifs sont dans **Act**

Communication

Communication dans WordNet inclut les actes et les contenus communicationnels, les médiums de communication et les objets linguistiques

Communication dans FR-SemCor inclut seulement les objets linguistiques et médium de communication, les actes de communication sont dans **Act**, les contenus communiqués dans **Cognition**

1. Définition des Supersenses

Adaptation des *Unique Beginners* pour Fr-SemCor

- Guide d'annotation (<https://github.com/FrSemCor/FrSemCor/blob/master/guideAnno-FR-SemCor.pdf>)
 - ▶ Définition des classes, avec explicitation des répartitions entre classes

Animal

Définition : être animé non humain

- ⊘ les bactéries et les virus (micro-organismes), classés dans Object
- ⊘ la viande/chaire de l'animal, classés dans Food

NotaBene 1 : les N comme *bébé*, *jeune*, etc. sont étiquetés Animal s'ils ont dans la phrase un référent de type animal

NotaBene 2 : les animaux dans les histoires sont annotés Animal même s'ils ont des comportements humains

Exemples :

- *A la grande époque, le chien_{Animal} vedette passait son temps à se castagner avec son compère, le chat_{Animal} Hercule.*
- *La destruction de l'habitat naturel de cette espèce_{Group & Animal} endémique de l'île-continent a causé une hécatombe : 53 % des marsupiaux_{Animal} ont disparu dans le Queensland.*

1. Définition des Supersenses

Adaptation des *Unique Beginners* pour Fr-SemCor

- Guide d'annotation (<https://github.com/FrSemCor/FrSemCor/blob/master/guideAnno-FR-SemCor.pdf>)

▶ Tests linguistiques indicatifs

Attribute

Définition : propriété constitutive d'une entité (cf +/- ILP) ou d'une situation ou propriété en soi (considérées de manière abstraite, ex. Aimer la beauté)

∈ les noms de propriétés mesurables (ex. *taille*), y compris lorsque ceux-ci sont employés métonymiquement pour désigner la mesure elle-même (ex. *Indiquez-moi votre taille*).

Tests indicatifs : *Pierre/ce mur est d'un grand N, Pierre a fait preuve de N/d'un N exp*

Exemples :

- *Ils s'éloignent lentement en direction de la passe du Lido, pour retrouver l'immensité_{Attribute} de la mer Adriatique.*
- *La descente du canal de la Giudecca restera autorisée, mais seulement aux bateaux de taille_{Attribute} plus modeste.*
- *Guy Hosneld était une personne très connue dans la commune et aux alentours, pour son grand dévouement_{Attribute} aux autres et sa participation à la vie associative du village.*

1. Définition des Supersenses

Supersenses du projet FR-SemCor

1 Classes complémentaires

Entités concrètes :

animée : **Animal, Person**

naturelle : **Body, Plant, Object**

fabriquée : **Artifact**

naturelle ou fabriquée : **Food, Substance**

Situations dynamiques : **Act, Event**

Situations statives ou propriétés : **Attribute, State, Feeling**

Autres : **Institution, Cognition, Communication, Relation, Possession, Phenomenon, Time**

2 Classes relationnelles

Groupes ou parties (de qqch) : **Group, Part, Quantity**

1. Définition des supersenses

2. Méthode d'annotation

3. Analyse des données annotées

Conclusion

2. Méthode d'annotation

Opérateurs pour la construction de Supersenses complexes

- L'opérateur / permet de rendre compte de cas d'ambiguïté contextuelle (3-a) ou de cas d'indétermination (3-b), propres aux noms à facettes
- (3)
- a. *L'introducteur artériel est retiré deux heures après l'arrêt^{Act/Event} de la perfusion.*
 - b. *Le visiteur pourra aussi découvrir un exposé^{Act/Cognition} sur l'histoire et le fonctionnement des planètes.*

2. Méthode d'annotation

Opérateurs pour la construction de Supersenses complexes

- L'opérateur / permet de rendre compte de cas d'ambiguïté contextuelle (3-a) ou de cas d'indétermination (3-b), propres aux noms à facettes
 - (3) a. *L'introducteur artériel est retiré deux heures après l'arrêt^{Act/Event} de la perfusion.*
 - b. *Le visiteur pourra aussi découvrir un exposé^{Act/Cognition} sur l'histoire et le fonctionnement des planètes.*
- L'opérateur x permet d'annoter les noms dont le sens agrège un élément fonctionnel/relationnel et une classe ontologique
 - (4) a. *Guy était arrivé ici tout jeune avec sa famille^{Group x Person}.*
 - b. *Dimanche en milieu d'après-midi, quelques enfants s'adonnaient à des glissades alors que l'eau, cachée, recouvrait encore tout le secteur^{Part x Object}.*

2. Méthode d'annotation

Méthode d'annotation guidée pas la référence en contexte

- Les noms à référents hétérogènes sont annotés selon leur référent en contexte
 - (5) a. *un membre***Person** *de la famille*
 - b. *un membre***Institution** *de l'ONU*
 - c. *un membre***Animal** *de la meute*

2. Méthode d'annotation

Méthode d'annotation guidée pas la référence en contexte

- Les noms à référents hétérogènes sont annotés selon leur référent en contexte
 - (5)
 - a. *un membre**Person* *de la famille*
 - b. *un membre**Institution* *de l'ONU*
 - c. *un membre**Animal* *de la meute*
 - Les arguments implicites des noms à sens fonctionnel sont rétablis dans l'annotation (6-b)
 - (6)
 - a. *Les évènements indésirables sont listés par classes**Group* *de systèmes d'organes dans le tableau 1.*
 - b. *Effet de classe**Group x Substance* : *anomalie de la fonction rénale*

2. Méthode d'annotation

Méthode d'annotation guidée pas la référence en contexte

- Les noms à référents hétérogènes sont annotés selon leur référent en contexte
 - (5) a. *un membre**Person* *de la famille*
 - b. *un membre**Institution* *de l'ONU*
 - c. *un membre**Animal* *de la meute*
- Les arguments implicites des noms à sens fonctionnel sont rétablis dans l'annotation (6-b)
 - (6) a. *Les évènements indésirables sont listés par classes**Group* *de systèmes d'organes dans le tableau 1.*
 - b. *Effet de classe**Group x Substance* : *anomalie de la fonction rénale*
- Les noms en apposition sont annotés selon le nom caractérisé (phénomène marginal)
 - (7) a. *le roi**Possession* *dollar*
 - b. *Je suis souvent sceptique vis-à-vis des exigences très sévères, par exemple en matière d'environnement, qui sont imposées aux pays candidats**Institution*.

2. Méthode d'annotation

- Annotation des noms communs du corpus Sequoia

Nombre de lemmes nominaux	2 974
Nombre de tokens nominaux	15 197
Nombre de tokens dans des unités polylexicales	3 689

- Annotation faite dans l'environnement WebAnno (Eckart de Castilho et al., 2016) par trois annotatrices non expertes et par les membres de l'équipe lexicographique du projet
 - Pas de pré-annotation automatique
 - Pas de joker possible en cas d'occurrences difficiles à annoter
- Chaque token a été annoté en double aveugle et adjudiqué

1. Définition des supersenses
 2. Méthode d'annotation
 3. Analyse des données annotées
- Conclusion

3. Analyse des données annotées

Distribution des Supersenses dans le corpus Sequoia

Supersense	Nb tokens	Nb types
Act	2079	613
Person	1853	396
Time	1052	108
Cognition	906	252
Institution	735	155
Event	712	149
Substance	711	129
Quantity	687	93
State	621	160
Attribute	565	184
Artifact	530	236
Possession	366	71
Act/Cognition	314	97
Group×Person	222	54
Artifact/Cognition	213	51
Body	203	54
Object	159	56
Feeling	81	46
Part	67	28
Phenomenon	60	18

3. Analyse des données annotées

Au niveau lexical : env. 15% des lemmes nominaux ont reçu au moins 2 Supersenses distincts

- Cas de polysémie ou d'homonymie, ex. INTÉRÊT

- (8) a. *L'enquête avait mis en évidence l'intérêt^{Feeling} des adultes pour ces activités.*
b. *Le budget de la France ne peut rembourser 16 milliard de francs plus intérêts^{Possession}*

- Noms à facettes, ex. TRAITEMENT

- (9) a. *La dose d'Angiox et la durée du traitement^{Act} dépendent des motifs de son utilisation.*
b. *Angiox s'est avéré aussi efficace que les traitements^{Act / Substance} auxquels il a été comparé.*

- Noms à référents hétérogènes, ex. MEMBRE cf. (5)

- Noms incluant une composante fonctionnelle, ex. CLASSE cf. (6)

3. Analyse des données annotées

Accord inter-annotatrices

	Full set	Coarse set
Exact match	κ : 0.649 ⁶ ag : 0.683	κ : 0.700 ag : 0.742
Partial match	κ : 0.734 ag : 0.763	κ : 0.790 ag : 0.822

Full set : 88 Supersenses distincts (24 simples + 64 complexes)

Coarse set : 25 Supersenses distincts (6 simples + 19 complexes)

GroupxPerson vs Group : Exact match ✗ / Partial match ✓

Act/Substance vs Artifact/Substance : Exact match ✗ / Partial match ✓

6. Accord similaire à celui rapporté dans (Schneider et al. 2012), calculé uniquement sur les annotations jugées non problématiques et sur un nombre de classes moindre

3. Analyse des données annotées

Analyse de l'accord

- Grande variabilité au sein des classes
Person (94.3) Substance (83.3) Artifact (72.1) Act (70.9)
State (50.7), Cognition (47.5), Attribute (43.6) Event (37.5)
- Principales zones de désaccords
 - Entre classes sémantiquement proches
 - Act et Event
 - ☞ Difficulté à déterminer si une situation dynamique est agentive
 - Attribute et State
 - ☞ Difficulté à déterminer si une situation stative est épisodique
 - Entre classes concernées par les N à facettes (ex. *constat*, *versement*)
Annotations alternatives les + fréquentes pour Cognition : Act et Artifact
Annotations alternatives les + fréquentes pour Possession : Act (et Cog)

1. Définition des supersenses
2. Méthode d'annotation
3. Analyse des données annotées

Conclusion

- Données sémantiques gold pour les noms du français
 - ▶ Jeu d'étiquettes défini avec soin
 - ▶ Double annotation en aveugle et adjudication de l'ensemble des données annotées (~ 600 heures)
- Utilisées comme données d'évaluation d'une méthode de supersense tagging non supervisée (Aloui *et al* 2020)
- Evaluation de méthodes supervisées et semi-supervisées pour la classification sémantique
 - des noms du français
 - des verbes, via leur arguments nominaux

Mise à disposition du corpus

- Nouvelle couche d'annotation du Deep-Sequoia (Candito *et al.* 2014)
<https://deep-sequoia.inria.fr/>
- Disponible sous Licence LGPL-LR

```
# sent_id = annodis.er_00033
# text = Le mobilier est subventionné à 50% par le conseil général.
1 Le le DET _ Definite=Def|Gender=Masc|Number=Sing|PronType=Art 2 det _ _ * *
2 mobilier mobilier NOUN _ Gender=Masc|Number=Sing 4 nsubj:pass _ _ * GroupxArtifact
3 est être AUX _ Mood=Ind|Number=Sing|Person=3|Tense=Pres|VerbForm=Fin 4 aux:pass _ _ * *
4 subventionné subventionner VERB _ Gender=Masc|Number=Sing|Tense=Past|VerbForm=Part|Voice=Pass 0 root _ _ * *
5 à à ADP _ _ 7 case _ _ * *
6 50 50 NUM _ NumType=Card 7 nummod _ _ * *
7 % % NOUN _ Gender=Masc 4 obl:mod _ _ * Quantity
8 par par ADP _ _ 10 case _ _ * *
9 le le DET _ Definite=Def|Gender=Masc|Number=Sing|PronType=Art 10 det _ _ * *
10 conseil conseil NOUN _ Gender=Masc|Number=Sing 4 obl:agent _ _ 1:|MWE|LEX 1:Institution
11 général général ADJ _ Gender=Masc|Number=Sing 10 amod _ _ 1 1
12 . . PUNCT _ _ 4 punct _ _ * *
```

(10) Le mobilier_{Groupx:Artifact} est subventionné à 50%_{Quantity} par le conseil_{Institution} général

Références 1

- Candito M. and Seddah D. (2012) Le corpus Sequoia : annotation syntaxique et exploitation pour l'adaptation d'analyseur par pont lexical. *Actes de TALN'2012*, Grenoble, France
- Candito M., Perrier G., Guillaume B., Ribeyre C., Fort K., Seddah D. and de la Clergerie E. (2014) Deep Syntax Annotation of the Sequoia French Treebank. *Proceedings of LREC 2014*, Reykjavic, Iceland.
- Ciaramita M. and M. Johnson (2003) Supersense Tagging of Unknown Nouns in WordNet. *Proceedings of EMNLP-2003*.
- Cruse, D. (1995). Polysemy and Related Phenomena from a Cognitive Linguistic Viewpoint. In : P. St-Dizier and É. Viegas (dir.) : *Computational Lexical Semantics*. Cambridge (G.-B.) : Cambridge University Press, 33-49.
- Delli Bovi C., J. Camacho-Collados, A. Raganato, and R. Navigli. (2017) Eurosense : Automatic harvesting of multilingual sense annotations from parallel text. In *Proceedings of ACL*.
- Djemaa, M., Candito, M., Muller P. and Vieu L. (2016) Corpus annotation within the French Framenet : methodology and results, Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC), Portorož, Slovenia, 2016.
- Eckart de Castilho, R., Mújdricza-Maydt, É., Yimam, S.M., Hartmann, S., Gurevych, I., Frank, A. and Biemann, C. (2016) A Web-based Tool for the Integrated Annotation of Semantic and Syntactic Structures. *Proceedings of the LT4DH workshop at COLING 2016*, Osaka, Japan
- Fellbaum C. (1998) *WordNet : An Electronic Lexical Database*. Cambridge, MA : MIT Press
- Landes, S., Leacock, C., and Teng, R. I. (1998). Building semantic concordances. In Fellbaum, C. (Ed.), *WordNet : An Electronic Lexical Database*, pp. 199–216. MIT Press.
- Miller, R. Beckwith, C. D. Fellbaum, D. Gross, K. Miller (1990) WordNet : An online lexical database. *International Journal of Lexicography* 3(4). 235–244.

Références 2

- Navigli R. (2006) Meaningful clustering of senses helps boost word sense disambiguation performance. In *Proceedings of the 44th Annual Meeting of the Association for Computational Linguistics joint with the 21st International Conference on Computational Linguistics (COLING-ACL, Sydney, Australia)*. 105–112.
- Navigli R., Litkowski, K.C., and Hargraves, O. (2007) Semeval-2007 task 07 : Coarse-grained English all-words task. In *Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval, Prague, Czech Republic)*. 30–35.
- Palmer M., Dang, H., and Fellbaum, C. (2007) Making fine-grained and coarse-grained sense distinctions, both manually and automatically. *J. Nat. Lang. Eng.* 13, 2, 137–163
- Pedersen B. S, A. Braasch, A. Johannsen, H. Mart´nez Alonso, S. Nimb, S. Olsen, A. Søgaard, and N. Sørensen (2016) The semdax corpus–sense annotations with scalable sense inventories. In *LREC*.
- Petrolito T and F. Bond (2014). A survey of WordNet Annotated Corpora, In *Proceedings of the Seventh Global WordNet Conference, Tartu, Estonia*, pp. 236-243. Tartu, Estonia.
- Snyder B. and M. Palmer (2004) The English all-words task. In *Proceedings of the 3rd International Workshop on the Evaluation of Systems for the Semantic Analysis of Text (Senseval-3, Barcelona, Spain)*. 41–43.
- Schneider N., B. Mohit, K. Oflazer and N. Smith (2012) Coarse lexical semantic annotation with supersenses : an Arabic case study. *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*. Jeju Island, Korea.
- Weischedel, Ralph, et al. *OntoNotes Release 5.0 LDC2013T19*. Web Download. Philadelphia : Linguistic Data Consortium, 2013.