

Mise en œuvre du concept de Linked Data pour la publication de données spatiales environnementales : une application aux cyclones tropicaux

Sylvain OZAINNE

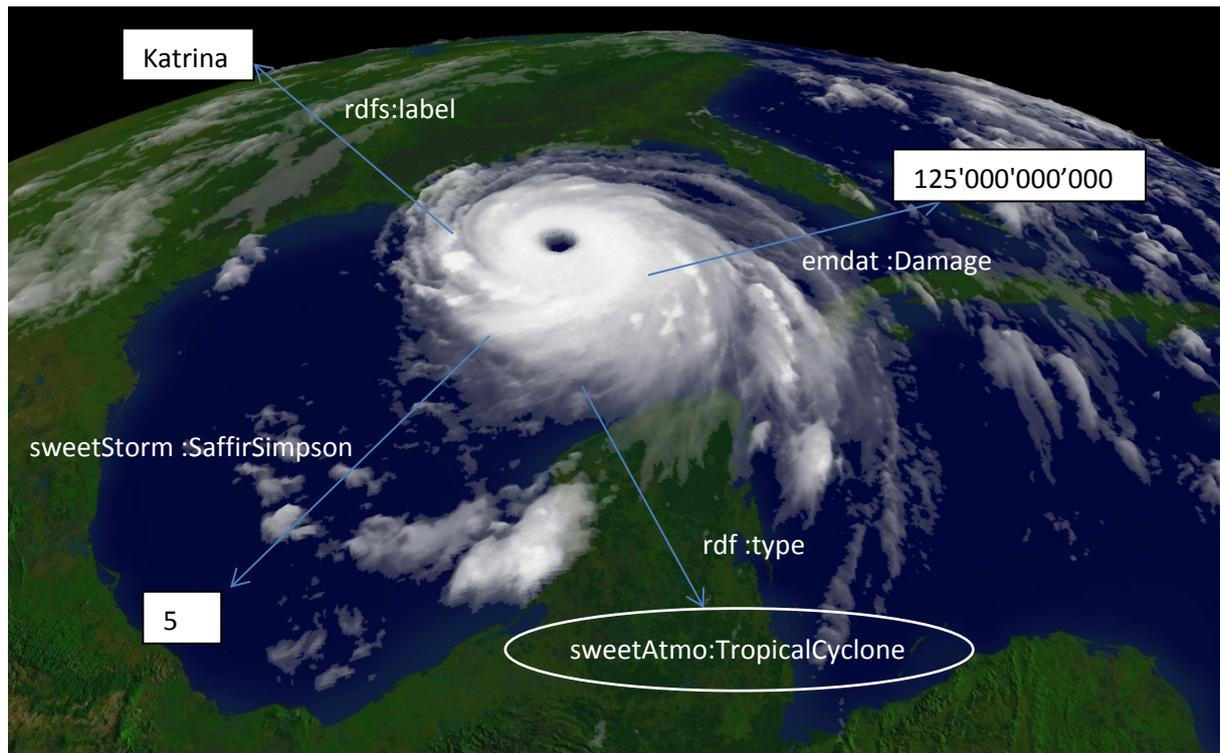


Image NOAA/NESDIS: <http://www.nvvl.noaa.gov/2005/08/28T1515Z>

Certificat complémentaire en géomatique 2015

Université de Genève, Faculté des Sciences de la Société

ISE/GRID-Geneva

Rapport de stage

Responsable de stage: Gregory Giuliani (ISE/GRID-Geneva)

Jury: Gregory Giuliani, Claudine Métral (CUI) et Gilles Falquet (CUI)

Table des matières

Résumé.....	4
Mots-clés	4
1. Problématique et objectifs	5
1.1. Introduction et problématique générale	5
1.2. Objectifs du stage.....	6
2. Linked Data et données géospatiales.....	7
2.1 Le Linked Data et le Web sémantique.....	7
2.2. Le Linked Open Data et le « 5 Star Open Data »	8
2.3. Réaliser le Web sémantique : RDF, vocabulaire, ontologies et syntaxes.....	9
2.3.1. Le standard RDF.....	9
2.3.2. Vocabulaires et ontologies	10
2.3.3. Syntaxes et préfixes.....	12
2.3.4. Publier des données RDF interrogeables : triplestores et endpoints.....	14
2.3.5. Interroger les données RDF : le langage SPARQL.....	14
2.4. Données géospatiales et Linked Data	15
2.4.1. Généralités	15
2.4.2. Standards et modèles de données	16
2.4.3. Le langage GeoSPARQL.....	17
2.4.4. Linked Data et données au format raster	18
2.4.5. Exemples d'applications géospatiales du Linked Data	20
3. Workflow pour la préparation et la publication d'un jeu de données du PREVIEW au format Linked Data	21
3.1. Introduction.....	21
3.2. Données.....	22
3.2.1. PREVIEW	23
3.2.2. EM-DAT.....	23
3.2.3. DBpedia	23
3.3. Choix des outils et des technologies	24
3.4. Préparation des données	25
3.4.1. Simplification des données du PREVIEW.....	25
3.4.2. Enregistrement du shapefile de cyclones PREVIEW au format WKT	27

3.4.3. Aménagement du fichier PREVIEW enregistré en CSV	28
3.4.4. Préparation des données de l'EM-DAT	29
3.5. Préparation du graphe RDF des données du PREVIEW	29
3.5.1. Modélisation des données au format RDF	29
3.5.2. Choix du vocabulaire	30
3.5.3. Ecriture du graphe RDF des données PREVIEW	32
3.6. Préparation du graphe RDF des données EM-DAT.....	39
3.6.1. Modélisation et choix des vocabulaires	39
3.6.2. Ecriture du graphe RDF des données EM-DAT.....	40
3.7. Etablissement des liens entre les graphes PREVIEW et EM-DAT	40
3.8. Préparation d'un graphe PREVIEW lié complet.....	46
3.9. Chargement des graphes PREVIEW et EM-DAT sur Parliament.....	46
4. Démonstrateur	48
4.1. Site web	48
4.2. Applications de visualisation du site web	49
4.2.1. Tableaux de bord.....	49
4.2.2. Formulaire SPARQL.....	51
4.2.3. Exemples de requêtes sur des données spatiales.....	51
4.4. Interrogation des données avec le package R SPARQL	53
5. Bilan et perspectives	55
5.1 Avantages et potentiel du Linked Data	55
5.2 Principaux inconvénients	56
5.3 Conclusion et perspectives de développement	57
Remerciements	58
Références.....	59
Annexes : code du démonstrateur web	62
A. Index.....	62
B. Dashboard PREVIEW	64
C. Dashboard EM-DAT	70
D. Formulaire SPARQL	77
E. Exemples de requêtes	81

Résumé

Le travail réalisé dans le cadre de ce stage à l'ISE/GRID-Geneva s'inscrit dans une problématique générale de partage des données permettant la documentation des catastrophes et l'évaluation des risques. Son objectif était d'évaluer le potentiel du concept de Linked Data pour la publication, la consommation mais aussi l'interconnexion de données à couverture globale, soit d'une part la base de données internationale sur les catastrophes EM-DAT (Emergency Events Database) et d'autre part le PREVIEW Global Risk Data Platform, hébergé au GRID. Ces deux sources de données sont complémentaires mais rassemblent des informations de natures différentes dont l'utilisation conjointe peut s'avérer problématique. Il est notamment compliqué de faire coïncider les chiffres de l'EM-DAT, associés à des entités administratives, avec l'emprise spatiale réelle des événements tels que documentés sur le PREVIEW. Or, l'accès par les décideurs et les chercheurs à des données spatialisées sur les catastrophes passées demeure essentiel pour l'estimation des risques. Il apparaissait ainsi important de tester les capacités du Linked Data, dont le principe général consiste à dépasser la logique des silos de données cloisonnés en partageant et en interconnectant entre elles des données structurées selon des standards du W3C (World Wide Web Consortium).

Dans cette optique, un jeu de données du PREVIEW modélisant l'emprise spatiale des cyclones tropicaux lors des différentes phases de leur évolution sur l'échelle des vents de Saffir-Simpson a été traduit au format Linked Data, en s'appuyant sur le standard RDF (Resource Description Framework) du W3C ainsi que sur des standards de l'OGC (Open Geospatial Consortium). Cette transcription a permis d'interconnecter la représentation spatiale des cyclones du PREVIEW avec les chiffres correspondant de l'EM-DAT ainsi qu'avec d'autres sources de données libres. Une interface Web a ensuite été développée afin de visualiser et d'interroger les données à l'aide du langage de requête SPARQL (SPARQL Protocol and RDF Query Language).

Les résultats de ce travail de stage sont largement positifs, l'utilisation du Linked Data permettant une interconnexion de données différentes ainsi qu'une interopérabilité à l'échelle du Web. Cette solution offre aussi des avantages concernant la mise à jour des informations ainsi que de nombreuses possibilités de réutilisation des données. Les principaux inconvénients constatés sont liés à l'instabilité de certaines ressources Web et à l'immaturité ou à la difficulté de prise en main des outils disponibles, ainsi qu'au temps d'apprentissage nécessaire à la production et à l'utilisation des données à ce format. Globalement, le concept de Linked Data permet d'envisager de nombreuses perspectives de développement pour la publication et la consommation de données géospatiales environnementales.

Mots-clés

Linked Data ; Environnement ; Catastrophes ; Risque ; Cyclones tropicaux ; PREVIEW Global Risk Data Platform ; EM-DAT ; W3C ; OGC ; RDF ; SPARQL

1. Problématique et objectifs

1.1. Introduction et problématique générale

La documentation des catastrophes et l'évaluation des risques nécessitent le recours à des bases de données internationales caractérisées par une couverture globale ainsi que l'utilisation de méthodes comparables pour l'estimation des chiffres associés (Peduzzi et al. 2005). La base de données internationale sur les catastrophes EM-DAT (Emergency Events Database : <http://www.emdat.be/>), élaborée et maintenue par le CRED (Centre for Research on the Epidemiology of Disasters : <http://www.cred.be/>), fournit notamment un accès public online à des résumés de données sur plus de 18000 catastrophes majeures survenues dans le monde depuis 1900. Bien qu'elle permette d'accéder à de nombreuses statistiques sur les pertes humaines et financières passées et offre la possibilité de consulter des résumés sous forme de graphiques ou de cartes, la base de données de l'EM-DAT rassemble des événements et chiffres qui ne sont pas précisément géoréférencés. Or, la connaissance de l'emprise spatiale des événements est indispensable à l'élaboration d'indicateurs importants. Par exemple, une estimation des pertes humaines peut être obtenue en calculant le rapport des personnes tuées au total de la population concernée par l'évènement, cette dernière donnée nécessitant de connaître l'aire exacte touchée par l'aléa (Peduzzi et al. 2005). Cette problématique a notamment motivé la mise au point par le DEWA/GRID Geneva ¹ d'une série de procédures impliquant l'utilisation de SIG et permettant de définir les zones et ainsi les populations affectées par plusieurs types de catastrophes, à savoir les inondations, les tremblements de terre, les cyclones tropicaux et les volcans (Peduzzi et al. 2005).

Les résultats de ces recherches sont notamment disponibles via la Global Risk Data Platform (PREVIEW : <http://preview.grid.unep.ch/>) du GRID, dont le but est de mettre à disposition des données spatiales relatives aux risques globaux liés à différents aléas naturels tels que cyclones tropicaux et ondes de tempête associées, sécheresses, tremblements de terre, feux de brousse, inondations, glissements de terrain, tsunamis et éruptions volcaniques (Giuliani & Peduzzi 2011). Les données du PREVIEW sont accessibles par l'intermédiaire d'interfaces de téléchargement (en format shapefile, raster ou GML, par exemple) mais aussi sous la forme de services Web respectant les standards de l'Open Geospatial Consortium ou OGC (<http://www.opengeospatial.org/>) tels que WMS, WFS, WCS, KML et GeoRSS. Un extracteur de données permet également de sélectionner des sous-ensembles de données par pays ou par région. Il est donc possible, par exemple, d'intégrer directement des données sur le tracé des cyclones des 40 dernières années dans des infrastructures de données spatiales ou dans un simple SIG par l'intermédiaire d'un service WMS.

Depuis quelques années, un nombre croissant d'informations géospatiales ouvertes, fournies notamment par des gouvernements, est également accessible online grâce au développement du Web des données ou Linked Data. Si les concepts généraux de Web sémantique et de Web des données ont été posés dès les années 1990 par le fondateur du World Wide Web Tim Berners-Lee (voir chapitre 3.1), leur mise en œuvre et leur utilisation sont toujours en cours de développement et il était intéressant de pouvoir évaluer leur potentiel pour le partage de données spatiales environnementales.

¹ Le DEWA/GRID-Geneva fait partie d'un groupe global de centres d'informations de l'UNEP (United Nations Environment Program), le Global Resource Information Database (GRID). LE GRID-Geneva fonctionne en étroite collaboration avec l'Université de Genève.

Dans cette optique, il a été planifié de tester dans le cadre de ce stage les possibilités offertes par le Linked Data en utilisant des données spatiales du PREVIEW, plus précisément un jeu de données modélisant l'emprise spatiale et la force des vents des cyclones tropicaux (<http://preview.grid.unep.ch/index.php?preview=data&events=cyclones&evcat=1&lang=eng>). Ces données occupent en effet une place importante dans la recherche précédemment citée et dont le but était d'apporter une meilleure dimension spatiale aux données accessibles via l'EM-DAT (Peduzzi et al. 2005). La mise à disposition des décideurs de données sur les cyclones tropicaux passés est en outre particulièrement importante. Bien que les liens entre activité humaine, réchauffement global et croissance de la fréquence ou de la puissance des cyclones aient été sujets à discussion depuis une dizaine d'années (Pielke et al. 2005 ; Knutson et al. 2010 ; Geophysical Fluid Dynamics Society 2015), le nombre et la proportion d'ouragans atteignant des catégories 4 ou 5 sur l'échelle de Saffir-Simpson, ainsi que leur potentiel de destruction aurait montré une tendance à l'augmentation entre 1970 et 2005 (Webster et al. 2005 ; Emmanuel 2005). Bien que des projections récentes montrent que la fréquence des cyclones tropicaux diminuera ou restera stable pendant le 21^{ème} siècle, il est aussi probable que leur intensité moyenne augmente (Peduzzi et al. 2012). Sachant que le risque de mortalité lié aux cyclones est dépendant de leur intensité, du nombre de personnes exposées, du niveau de pauvreté et de la gouvernance, les prévisions de croissance démographique et d'augmentation de l'intensité des cyclones pour les décennies à venir sont susceptibles d'accroître de façon importante le nombre de personnes exposées et donc le risque de catastrophe (Peduzzi et al. 2012). L'accès par les décideurs aux données sur les impacts passés afin d'estimer les risques futurs est donc crucial. Il est également important que les chercheurs travaillant notamment à la modélisation des cyclones puissent accéder facilement aux mêmes données.

Ce travail de stage mené au GRID-Geneva entre juin et novembre 2015 avait ainsi pour but de tester le potentiel d'un nouvel outil de partage des données du PREVIEW, reposant sur le concept de Linked Data.

1.2. Objectifs du stage

Les principaux objectifs du travail ont été initialement définis comme suit :

- Effectuer une revue du concept de Linked Data pour les données géospatiales et statistiques ainsi que des technologies permettant sa réalisation ;
- Evaluer les possibilités de mise en œuvre du concept de Linked Data en utilisant les données de l'EDE et du PREVIEW ;
- Sélectionner des technologies applicables à la problématique et développer des services Linked Data pour l'EDE et le PREVIEW, cette étape impliquant le développement d'un démonstrateur/application.

Lors de discussions menées au début de ce travail avec Gregory Giuliani, responsable du stage, il a été décidé de tester l'établissement de liens entre données du PREVIEW et données de l'EM-DAT plutôt que celles de l'EDE, afin d'associer entités spatiales représentant les cyclones tropicaux et données statistiques sur les catastrophes survenues depuis 1970.

Une présentation générale du concept de Linked Data ainsi qu'un aperçu des possibilités de son application aux données géospatiales sont consultables dans le chapitre 2.

Les choix méthodologiques et technologiques ainsi qu'une description étape par étape du workflow sont rassemblés dans le chapitre 3.

Une présentation du démonstrateur et plusieurs exemples de l'utilisation des applications développées sont disponibles dans le chapitre 4.

Les principaux résultats issus de ce travail ainsi que des perspectives de développement sont discutés dans le chapitre 5.

2. Linked Data et données géospatiales

2.1 Le Linked Data et le Web sémantique

Le Web sémantique est un concept que l'on doit à l'inventeur du World Wide Web (W3), Tim Berners-Lee. Il est mis en œuvre par le World Wide Web Consortium (W3C)². Bien que les premières recommandations concernant le Web sémantique remontent à la fin des années 1990, c'est un article paru en 2001 dans le Scientific American qui pose véritablement les bases du concept (Berners-Lee et al. 2001). Cette référence est très souvent relatée comme étant le point de départ de la diffusion de l'idée d'un ajout de sémantique aux ressources du web (Campbell et MacNeill 2010 ; Verborgh 2014), notamment grâce à la définition suivante : « The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation. » (Berners-Lee et al. 2001).

En complément au Web des documents, le Web sémantique a ainsi pour but de construire un ensemble de technologies soutenant le Web des données (« Web of data », largement plus connu sous le nom de « Linked Data »), c'est-à-dire constitué de données telles que dates, titres, numéros, propriétés diverses, etc. La différence entre Web des documents et Web des données peut être expliquée par l'analogie suivante : le Web actuel est une plateforme pour la distribution de présentations, tandis que le Web sémantique est une plateforme pour la distribution de connaissances (Tauberer 2006). Les connaissances disponibles sur le Web sous forme de documents sont en effet inaccessibles pour les ordinateurs. Par exemple, une page Wikipedia peut contenir une grande quantité d'information pour un lecteur humain, tandis que le contenu réel, c'est-à-dire la connaissance devant être transmise par les fichiers, ne signifie rien pour un ordinateur (Tauberer 2006).

Une définition très simple du Linked Data est donnée dans un autre article très largement cité par les utilisateurs de cette technologie : « In summary, Linked Data is simply about using the Web to create typed links between data from different sources » (Bizer et al. 2009).

Concrètement, le terme de Linked Data renvoie à un ensemble de recommandations et de pratiques exemplaires pour la publication et la connexion entre elles de données structurées sur le Web en utilisant les standards définis par le W3C (Wood et al. 2014), l'idée générale étant de sortir de la logique des silos de données cloisonnées. L'ensemble des informations publiées sous cette forme constitue le Web sémantique ou le Web des données, soit un espace global de données (Bauer & Kaltenböck 2011 ; Heath & Bizer 2011 ; Roda et al. 2013).

² Le W3C est une communauté internationale qui œuvre au développement de standards ouverts destinés à conduire le World Wide Web à son potentiel maximum tout en assurant sa croissance à long terme (<http://www.w3.org/>). Fondé par Tim Berners-Lee en 1994, le W3C a pour mission de promouvoir la compatibilité des technologies du W3, telles que le HTML et le CSS.

Tim Berners-Lee (2006) a proposé un ensemble de quatre règles fondamentales pour la publication de données sur le Web, de façon à ce que toutes les données publiées deviennent des éléments d'un unique espace de données global (Berners-Lee 2006 ; Bizer et al. 2009). Ces principes sont les suivants :

1. Utiliser des **URIs** pour nommer les ressources;
2. Utiliser des **URIs http** afin que l'on puisse rechercher et accéder à ces ressources ;
3. Lorsqu'une **URI** est consultée, elle doit fournir une description de la ressource en utilisant les standards **RDF** et **SPARQL** (voir 2.3) ;
4. Inclure des **liens** vers d'autres **URIs**, de manière à ce que de nouvelles ressources puissent être découvertes et consultées.

2.2. Le Linked Open Data et le « 5 Star Open Data »

L'ensemble des données publiées et librement accessibles sur le Web est couramment nommé Linked Open Data (LOD), ce concept associant le concept de Linked Data à une publication des données en «Open License ». La quantité de données librement accessibles via le Web a connu un important essor dans la dernière décennie³.

En 2010, Tim Berners-Lee a proposé un système d'évaluation à cinq étoiles de la qualité des LOD (5 star Open Data : <http://5stardata.info/en/>):

★ Disponible sur le Web sous n'importe quel format avec une licence Open Data ;

★★ Disponible sous forme de données structurées exploitable par une machine (par exemple au format Excel plutôt qu'un tableau en format PDF) ;

★★★ Comme le point précédent, mais en utilisant un format non-propriétaire (par exemple CSV à la place d'Excel) ;

★★★★ Tous les points précédents, en utilisant les standards du W3C (RDF et SPARQL) pour identifier les objets ;

★★★★★ Tous les points précédents, en ajoutant des liens vers des données extérieures afin de fournir un contexte à ses propres données.

³ On peut trouver un aperçu de l'évolution du Web des données sur la page Wikipédia sur le Linked Open Data (https://fr.wikipedia.org/wiki/Linked_open_data) ainsi que via le « Linked Open Data cloud diagram » (<http://lod-cloud.net/>)

2.3. Réaliser le Web sémantique : RDF, vocabulaire, ontologies et syntaxes

2.3.1. Le standard RDF

La publication de données sur le Web sous forme de Linked Data, impliquant une interconnexion avec d'autres sources de données, repose sur l'utilisation d'un modèle de données de type graphe, utilisant comme format le standard RDF (Resource Description Framework : <http://www.w3.org/RDF/>) du W3C (W3C 1999 ; Shadbolt et al. 2006). Le standard RDF modélise les données sous la forme de triplets regroupant trois types d'objets (W3C 1999) :

-des ressources, ou sujet, que l'on veut décrire. Il peut s'agir d'une page web entière, d'une partie d'une page web ou, à l'inverse, d'un site Web entier. Une ressource peut également être un objet n'étant pas directement accessible via le Web, comme par exemple un livre imprimé. Les ressources sont toujours nommées par des URI (Uniform Resource Identifier : <http://www.w3.org/wiki/URI>), dont l'extensibilité permet l'introduction d'identifiants pour n'importe quelle entité imaginable ;

-des propriétés, ou prédicats, soit une caractéristique, un attribut ou une relation utilisée pour décrire une ressource ;

-des déclarations. Une ressource spécifique et une propriété nommée plus la valeur de cette propriété constituent une déclaration RDF (« RDF statement »). Les trois parties d'une déclaration sont communément nommées respectivement le *sujet*, le *prédicat* et l'*objet*.

Le modèle RDF est un modèle de graphe, constitué d'un ensemble de nœuds et d'arcs reliant des paires de nœuds. Un graphe est composé d'un ensemble de triplets RDF, chaque triplet étant constitué d'un nœud de type *sujet* (le nœud de départ, soit la ressource), d'un *prédicat*, qui correspond à une propriété de la ressource décrite et établit une relation avec un autre nœud, et d'un nœud de type *objet*, qui correspond à la valeur de la propriété et constitue le nœud d'arrivée.

A titre d'exemple simple, le graphe de données de la fig. 1 décrit un T-shirt de couleur blanche :

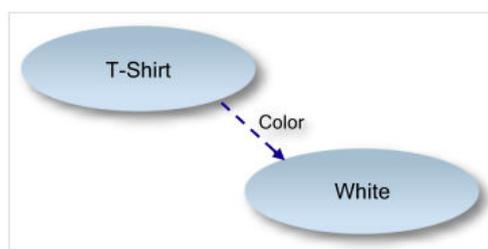


Fig. 1. Exemple simple de graphe de données. Source : <http://www.linkeddatatools.com/introducing-rdf>

Dans cet exemple, le sujet est le T-shirt, le prédicat ou propriété est la couleur et l'objet (ou valeur) de la propriété couleur est Blanc.

L'objet d'une déclaration RDF (c'est-à-dire la valeur de la propriété d'une ressource) peut être lui-même soit une autre URI, soit une valeur constante que l'on appelle un « littéral » et pouvant par exemple être constitué d'une chaîne de caractères. Graphiquement, on représente les sujets ou objets qui sont des URI entourés d'un ovale, les objets de type « littéral » entourés d'un rectangle, tandis que les prédicats sont simplement représentés par l'URI correspondante (voir fig. 2).

Le graphe d'exemple de la fig. 2 représente le groupe de déclarations suivant : « Il y a une ressource identifiée par l'URI <http://ua.unige.ch/fr/personne/sylvainozainne/> qui permet d'identifier une personne dont le nom est Sylvain Ozainne, dont l'adresse e-mail est sylvain.ozainne@unige.ch et dont le titre est Dr. ». Ce graphe est donc composé de quatre triplets ayant tous comme sujet une seule et même URI.

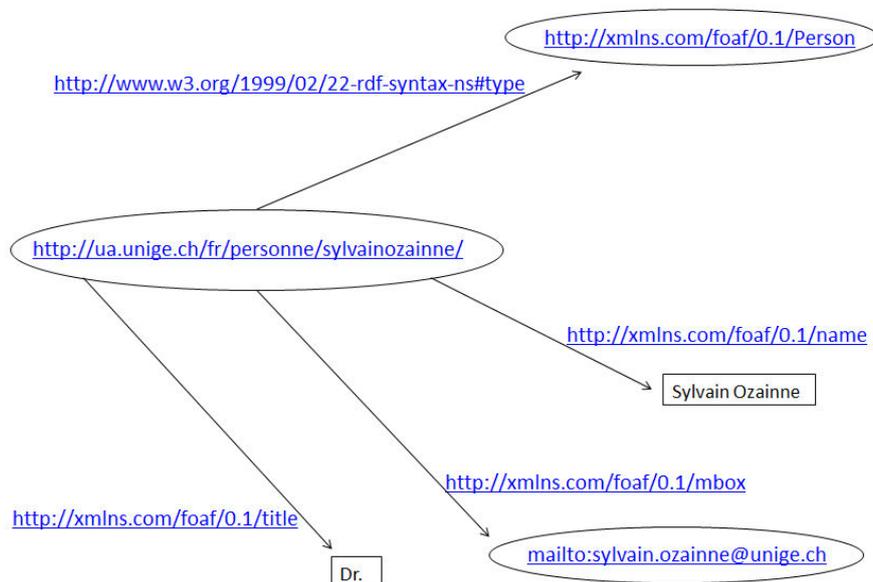


Fig. 2. Exemple de représentation de graphe RDF

Sans lui appliquer de syntaxe précise, la description complète du graphe de la fig. 2 sous forme de triplets s'écrit de la façon suivante :

```
<http://ua.unige.ch/fr/personne/sylvainozainne/> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://xmlns.com/foaf/0.1/Person>

<http://ua.unige.ch/fr/personne/sylvainozainne/> <http://xmlns.com/foaf/0.1/name> « Sylvain Ozainne »

<http://ua.unige.ch/fr/personne/sylvainozainne/> <http://xmlns.com/foaf/0.1/mailbox>
<mailto:sylvain.ozainne@unige.ch>

<http://ua.unige.ch/fr/personne/sylvainozainne/> <http://xmlns.com/foaf/0.1/title> « Dr. »
```

Les données RDF peuvent aussi être exprimées sous forme de « N-Quads » (<http://www.w3.org/TR/n-quads/>). Comme les triplets, les quads sont composés de trois termes représentant un sujet, un prédicat et un objet, ainsi que du nom du graphe auquel ils appartiennent.

2.3.2. Vocabulaires et ontologies

Pour décrire les différentes propriétés de la ressource, le graphe de la fig. 2 fait appel à deux vocabulaires différents. Pour décrire la première propriété de la ressource, c'est-à-dire son type, le graphe utilise notamment le vocabulaire de modélisation des données propre au RDF (<http://www.w3.org/1999/02/22-rdf-syntax-ns#>). Le vocabulaire RDF regroupe plusieurs prédicats et classes, notamment le prédicat <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> (abrégé

rdf:type ou encore plus simplement a ; voir l'utilisation des préfixes *infra*). Incontournable et très largement utilisé, il permet d'établir qu'une ressource est une instance d'une certaine classe (voir exemple fig. 1). Le RDF Schema (RDFS 1.1. : <http://www.w3.org/TR/rdf-schema/>) est une extension du RDF permettant de décrire plus précisément des taxonomies de classes et de propriétés et ainsi de définir une hiérarchie des ressources. Dans le graphe exemple de la fig. 2, les propriétés désignant le nom, l'adresse mail et le titre associés à la ressource font appel à l'ontologie FOAF (Friend of a Friend : <http://www.foaf-project.org/>), également largement usitée et dont le vocabulaire est spécifié à l'URL <http://xmlns.com/foaf/spec/>.

C'est l'utilisation de vocabulaires et d'ontologies qui permet d'appliquer une dimension sémantique aux graphes. L'ajout d'une sémantique permet alors de connecter les données automatiquement à travers plusieurs domaines de connaissances différents mais dont certaines informations pourraient être partagées. La fig. 3 illustre deux bases de données conçues indépendamment l'une de l'autre mais dont certains champs pourraient être mis en commun. La base des Oscars ne permet par exemple que d'accéder au nom et à la date de naissance des acteurs récompensés, alors que des informations beaucoup plus détaillées sur les mêmes personnes sont disponibles dans la base des biographies d'acteurs.



Fig.3. Exemple de bases de données indépendantes dont certaines données pourraient être partagées.

Source : <http://www.linkeddatatools.com/semantic-modeling>

Faire correspondre les informations entre ces bases de données conçues de façon indépendantes nécessiterait un certain investissement en temps et en moyens humains et financiers. Une fois réalisé, l'échange de données se limiterait en outre à ces deux bases de données et ne permettrait pas directement d'autre ajout d'informations extérieures. L'intervention humaine demeure dans ce cas de figure indispensable pour comprendre la signification des données et pour définir un format commun permettant aux deux bases de collaborer (<http://www.linkeddatatools.com/semantic-modeling>). L'utilisation du standard RDF et d'une modélisation sémantique constituent une solution plus simple. Le recours à un vocabulaire (soit une collection de termes possédant une signification bien définie et cohérente d'un contexte à l'autre) et une ontologie (qui permet de définir des relations contextuelles pour un vocabulaire défini) consistent les fondements de la définition d'un domaine commun. En reprenant le même exemple des Oscars, il faudrait que les deux sites adoptent un vocabulaire standard dont chaque terme à la même signification dans les deux ensembles de données et donc utilisent la même ontologie de base (ou un vocabulaire commun) pour exprimer la signification des données qu'ils mettent à disposition (fig. 4).

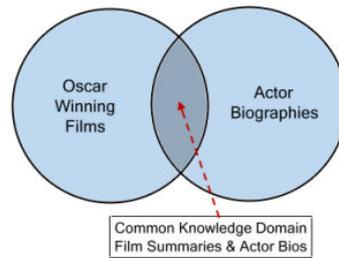


Fig. 4. Même exemple que la fig. 3 mais avec une modélisation sémantique. L'utilisation d'un vocabulaire commun permet aux deux sites de s'interroger mutuellement en utilisant les mêmes termes. Source : <http://www.linkeddatatools.com/semantic-modeling>

Le Web Ontology Language (OWL : <http://www.w3.org/2001/sw/wiki/OWL>), une autre extension aux vocabulaires RDF et RDFS, est une syntaxe formelle permettant d'écrire des ontologies web.

De nombreux vocabulaires standards ou ontologies formelles réunissant des termes généraux ou associés à des domaines de connaissances spécifiques sont mis à disposition par diverses organisations. Les ontologies Dublin Core Metadata Initiative (<http://dublincore.org/>), Friend Of A Friend (<http://xmlns.com/foaf/spec/>), Simple Knowledge Organization System (skos : <http://www.w3.org/2004/02/skos/>) ou encore OpenCyc (<http://www.opencyc.org/>), largement utilisées, permettent de décrire des entités et des relations générales (personnes, adresses, etc.). Les ontologies SWEET (Semantic Web for Earth and Environmental Terminology : <https://sweet.jpl.nasa.gov/>) proposées par la NASA rassemblent 6000 concepts liés à l'environnement et constituent un bon exemple d'ontologie plus spécifique. S'il est souvent nécessaire d'écrire une nouvelle ontologie pour publier un jeu de données, il est important de vérifier d'abord si les termes que l'on désire utiliser n'existent pas déjà, puis de choisir le vocabulaire le mieux adapté et si possible le plus usité, par exemple en effectuant une recherche sur le site LOV (Linked Open Vocabularies : <http://lov.okfn.org/dataset/lov>).

Le RDF Data Cube Vocabulary (<https://dvcs.w3.org/hg/gld/raw-file/default/data-cube/index.html>) est quant à lui conçu pour publier des données à dimensions multiples, essentiellement des statistiques, et son modèle est compatible avec le standard SDMX (https://sdmx.org/?page_id=5008), prévu pour faciliter l'échange de données statistiques et de métadonnées. Il comprend notamment un concept de « tranches » (slices), qui permet de regrouper des sous-ensembles à l'intérieur d'un jeu de données.

2.3.3. Syntaxes et préfixes

Un graphe RDF peut être écrit à l'aide de plusieurs syntaxes, notamment le RDF/XML (<http://www.w3.org/TR/rdf-syntax-grammar/>), qui est une recommandation du W3C et permet d'exprimer un ensemble de triplets en utilisant le langage XML (voir aussi http://www.w3schools.com/xml/xml_rdf.asp).

Le graphe de la fig. 2 s'écrit en syntaxe RDF/XML de la façon suivante⁴ :

```
<?xml version="1.0" encoding="utf-8" ?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:foaf="http://xmlns.com/foaf/0.1/">

  <foaf:Person rdf:about="http://ua.unige.ch/fr/personne/sylvainozainne/">
    <foaf:name>Sylvain Ozainne</foaf:name>
    <foaf:title>Dr.</foaf:title>
    <foaf:mailbox rdf:resource="mailto:sylvain.ozainne@unige.ch"/>
  </foaf:Person>
</rdf:RDF>
```

Il existe d'autres syntaxes RDF recommandées par le W3C, comme Notation3 (N3 : <http://www.w3.org/TeamSubmission/n3/>) ou la syntaxe Turtle (<http://www.w3.org/TR/turtle/>), plus légère que le RDF/XML et très largement utilisée. Il existe également une syntaxe JSON-LD (<http://json-ld.org/>), utilisant le format JSON.

A titre d'exemple, le graphe de la fig. 2 s'écrit en syntaxe Turtle de la façon suivante :

```
@prefix foaf: <http://xmlns.com/foaf/0.1/>.

<http://ua.unige.ch/fr/personne/sylvainozainne/>
  foaf:name "Sylvain Ozainne" ;
  a foaf:Person ;
  foaf:title "Dr." ;
  foaf:mailbox <mailto:sylvain.ozainne@unige.ch> .
```

Les syntaxes utilisent le concept de préfixe, qui permet d'abrégier des URIs longues et répétitives en déclarant un nom court. Dans l'exemple de la figure 2, plusieurs prédicats de l'ontologie foaf sont utilisés et la déclaration, en syntaxe Turtle, d'un préfixe @prefix foaf: <http://xmlns.com/foaf/0.1/> au début du document permet de simplifier le graphe RDF, puisqu'il suffira par exemple d'écrire le prédicat <http://xmlns.com/foaf/0.1/name> simplement sous la forme foaf :name. On peut attribuer n'importe quel nom à un préfixe, mais les ontologies les plus courantes sont le plus souvent abrégées de la même façon, comme par exemple rdfs pour le RDF Schema ou dc pour les termes du Dublin Core Metadata Initiative. Il n'y a toutefois pas de véritable standardisation et il convient par exemple de se méfier de l'emploi du préfixe geo, qui peut être utilisé pour renvoyer à des termes d'ontologies bien distinctes. Un outil très utile pour vérifier quelle sont les utilisations de divers préfixes est disponible à l'URL <http://prefix.cc/>.

Il existe aussi plusieurs outils permettant de valider les graphes RDF que l'on produit, par exemple :

<http://validator.linkeddata.org/vapour>
<http://graphite.ecs.soton.ac.uk/checker/>
<http://oops.linkeddata.es/>

⁴ Les traductions du graphe d'une syntaxe à l'autre ont été réalisées avec <http://www.easyrdf.org/converter>

2.3.4. Publier des données RDF interrogeables : triplestores et endpoints

Une façon efficace de publier des données sous forme de Linked Data est de les charger dans un entrepôt RDF, souvent nommés RDF stores, repositories ou triplestores. Les applications Web d'entrepôt RDF font majoritairement appel aux technologies de [Sesame](#), [Apache Jena](#) ou [Virtuoso](#). Le chargement de données RDF dans des entrepôts peut aussi permettre l'inférence de nouveaux triplets à partir de ceux existants. Une fois les données publiées dans un entrepôt, elles doivent idéalement être accessibles par un endpoint SPARQL (voir 2.3.5), soit une interface web permettant d'interroger les données RDF du triplestore. Une liste de endpoints SPARQL publiquement accessibles est disponible sur le wiki du W3C : <http://www.w3.org/wiki/SparqlEndpoints>.

A titre d'exemple, il est possible d'interroger, via un endpoint SPARQL, les métadonnées de la plateforme française ISIDORE (<http://www.rechercheisidore.fr/sqe/>) qui offre accès à des données sur la recherche en sciences humaines et sociales. Il est également possible d'accéder via un endpoint SPARQL aux données de LinkedGeoData (<http://linkedgeodata.org/sparql/>), qui est une version en Linked Data des données d'Open Street Maps. Les ressources de DBpedia (<http://dbpedia.org/sparql/>), une transcription en Linked Data des entrées de l'encyclopédie Wikipédia, sont également interrogeables par le même biais.

Ces exemples sont des endpoints spécifiques, c'est-à-dire qu'ils ont destinés à interroger un jeu de données précis, mais on peut aussi avoir recours à des endpoints génériques. Par exemple, le General purpose processor de SPARQLer (<http://sparql.org/sparql.html>) permet d'interroger n'importe quel jeu de données en précisant le nom du graphe que l'on veut explorer.

2.3.5. Interroger les données RDF : le langage SPARQL

Le langage de requête SPARQL (SPARQL Protocol and RDF Query Language : <http://www.w3.org/TR/sparql11-query/>) est le pendant Linked Data du SQL destiné à interroger les données structurées en format RDF. La version 1.0 est devenue un standard du W3C en 2008, tandis que la version 1.1., intégrant d'importantes fonctionnalités supplémentaires, est devenue un standard en 2013. Il permet d'explorer un jeu de données en interrogeant des relations inconnues mais aussi d'ajouter, de modifier ou de supprimer des données. SPARQL permet aussi de construire de nouveaux graphes en utilisant des requêtes CONSTRUCT. Un excellent tutoriel est disponible à l'URL <http://www.cambridgesemantics.com/semantic-university/sparql-by-example>.

Une requête SPARQL se compose des principaux éléments suivants :

- Une déclaration des préfixes utilisés (clause PREFIX) ;
- Une clause de résultat, qui précise quelles données doivent être renvoyées par la requête (clause SELECT) ;
- La définition du jeu de données à interroger (clause FROM) ;
- Le schéma de requête (« Query pattern »), qui précise quels sont les éléments d'un graphe que l'on veut interroger (clause WHERE) ;
- Des modificateurs de requête, permettant d'ordonner ou de filtrer les résultats (par exemple ORDER BY, FILTER,...)

L'exemple de requête suivant renvoie les pages web de toutes les personnes que connaît Tim Berners-Lee, à partir du graphe-cible <http://dig.csail.mit.edu/2008/webdav/timbl/foaf.rdf> (fig. 5.):

```

PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX card: <http://www.w3.org/People/Berners-Lee/card#>
SELECT ?homepage
FROM <http://dig.csail.mit.edu/2008/webdav/timbl/foaf.rdf>
WHERE {
  card:i foaf:knows ?known .
  ?known foaf:homepage ?homepage .
}

```



Fig. 5. Exemple de requête avec illustration de son schéma (« Query pattern »).

Source : <http://www.cambridgesemantics.com/semantic-university/sparql-by-example>

Dans cet exemple, l'utilisation de la variable ?known à la fois en tant qu'objet d'un triplet et en tant que sujet du triplet suivant, permet de suivre plusieurs liens du graphe interrogé.

Ressource utile, le Query demo de sparqler (<http://www.sparql.org/query.html>) rassemble sur une même page une interface permettant de tester des requêtes SELECT et une interface pour réaliser des requêtes CONSTRUCT, ces dernières permettant de produire un nouveau graphe à partir des données interrogées.

2.4. Données géospatiales et Linked Data

2.4.1. Généralités

Depuis quelques années, on constate un intérêt croissant pour les données géospatiales sous forme de Linked Data et leur quantité sur le Web sémantique augmente en conséquence (Nikolaou et al. 2013). En dehors de l'initiative gouvernementale pionnière de l'Ordnance Survey de Grande-Bretagne (<http://data.ordnancesurvey.co.uk/>), la mise à disposition d'importants jeux de données ouvertes au format Linked Data, tels que LinkedGeoData (<http://linkedgeodata.org>), la version RDF d'Open Street Map, ou GeoLinked Data (<http://geo.linkeddata.es>), doivent être signalées. Des applications permettant d'exploiter cette abondance de données ont également vu le jour ces dernières années, notamment l'outil SexTant (<http://www.earthobservatory.eu/SexTant>) développé dans le cadre du projet européen TELEIOS (Nikolaou et al. 2013 : <http://www.earthobservatory.eu/>). Ainsi que le soulignent les auteurs du projet européen GeoKnow (<http://geoknow.eu/Welcome.html>), bien que le partage, l'intégration et la réutilisation de données spatiales aient fait de nombreux progrès, notamment grâce aux efforts de standardisation de l'OGC, les technologies liées au Linked Data permettant l'interopérabilité, l'exploitation et la visualisation de données géospatiales n'émergent que lentement (Lehmann et al. 2015). Le projet GeoKnow, dont l'objectif est de proposer un ensemble d'outils permettant d'enrichir le Web des données en données géospatiales, travaille entre autres à une méthodologie destinée à établir des ponts entre la directive européenne INSPIRE (<http://inspire.ec.europa.eu/>) et le Web sémantique.

2.4.2. Standards et modèles de données

Plusieurs approches d'établissement de standards ou vocabulaires spécifiques aux données géospatiales au format RDF ont été développées ces dernières années (Athanasidou et al. 2013), le plus simple étant le vocabulaire basique du W3C GeoPos84, un vocabulaire permettant de représenter des coordonnées en longitude et latitude selon le système de coordonnées standard WGS84, tel que le montre l'exemple suivant en RDF/XML :

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:geo="http://www.w3.org/2003/01/geo/wgs84_pos#">
  <geo:Point>
    <geo:lat>55.701</geo:lat>
    <geo:long>12.552</geo:long>
  </geo:Point>
</rdf:RDF>
```

Il existe d'autres possibilités, comme par exemple l'utilisation de vocabulaire GeoRSS (http://www.georss.org/rdf_rss1.html) dans l'écriture de triplets en RDF/XML.

C'est surtout l'important développement du standard de l'OGC GeoSPARQL (<http://www.opengeospatial.org/standards/geosparql>) qui a récemment ouvert de larges possibilités pour la représentation et l'interrogation de données géospatiales en format RDF. Ce standard, déjà assez largement utilisé et offrant de nombreuses possibilités, a été exploité pour structurer et interroger les jeux de données produits en format RDF dans le cadre de ce stage. Il fait donc l'objet d'une description détaillée dans le chapitre suivant (2.4.3.).

Il est important de signaler ici la technologie Strabon (<http://www.strabon.di.uoa.gr/about>), un entrepôt RDF permettant notamment d'interroger de façon spatiale et temporelle des données grâce au format stRDF et au langage stSPARQL (http://www.strabon.di.uoa.gr/files/stSPARQL_tutorial.pdf), deux extensions des standards du W3C RDF et SPARQL proposées dans le cadre du projet FP7 ICT SensorGrid4Env (<http://www.sensorgrid4env.eu/>), lui-même destiné à faciliter le développement d'applications SSW (Semantic Sensor Web) pour la gestion de l'environnement. Le modèle de données stRDF est notamment fondé sur l'utilisation du standard de l'OGC Well Known Text (WKT : <http://www.opengeospatial.org/standards/wkt-crs>; voir 2.4.3) pour représenter les données géospatiales. Strabon permet aussi d'interroger des données géospatiales à l'aide d'un sous-ensemble de fonctions du standard GeoSPARQL, l'implémentation de l'ensemble de ses composantes étant en cours de réalisation.

D'autres modèles de données géospatiales ont été développés dans le cadre de divers projets ou initiatives, comme par exemple celle de [GeoLinkedData](#), une initiative ouverte de [l'Ontology Engineering Group](#) basé à l'Université Polytechnique de Madrid.

2.4.3. Le langage GeoSPARQL

Le standard de l'OGC GeoSPARQL (OGC 2012 ; Battle & Kolas 2012) est une extension spatiale du langage de requête SPARQL, dont les principaux éléments sont un vocabulaire RDF/OWL permettant de représenter l'information spatiale.

Modélisation et représentation d'objets géométriques avec GeoSPARQL

L'ontologie GeoSPARQL est composée de 3 classes principales :

- geo:Feature : Un objet qui peut avoir une localisation spatiale (un bâtiment, un monument, etc.) ;
- geo:Geometry : une représentation d'une localisation spatiale ;
- geo:SpatialObject: Une superclasse englobant les Features et les Geometries.

Les relations entre ces classes sont représentées sur la fig.6:

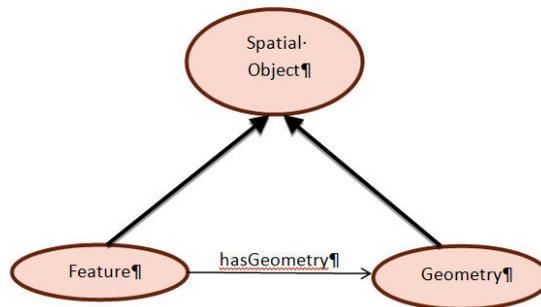


Fig. 6 : les trois classes de GeoSPARQL. Source : Kolas et Battle 2012, GeoSPARQL User Guide.

La propriété geo:hasGeometry permet de lier des objets à leur géométrie (leur position). En séparant les entités de leur localisation, GeoSPARQL permet ainsi de lier de multiples géométries à un objet en fonction du but recherché (un objet peut être représenté aussi bien par un point central que par un polygone, par exemple).

La propriété geo:asWKT permet de lier la géométrie à un littéral WKT. Le WKT (« Well Known text » : <http://www.opengeospatial.org/standards/wkt-crs>) est un standard de l'OGC permettant une représentation compacte d'objets géométriques lisible à la fois par les humains et par les machines, sous forme de chaîne de caractères. Par exemple, un point s'écrira au format WKT de la façon suivante :

```
"POINT(-77.03524 38.889468)"^^geo:wktLiteral.
```

Le littéral est écrit entre guillemets puis typé, c'est-à-dire que l'on indique à sa suite qu'il s'agit d'un littéral WKT, le préfixe geo: abrégant le prédicat <http://www.opengis.net/ont/geosparql#>.

Le format WKT, conçu au départ pour représenter des systèmes de coordonnées, permet de représenter sous forme de littéraux non seulement des points mais aussi des lignes, polygones et multipolygones.

L'exemple suivant (Kolas et Battle 2012) montre la modélisation complète d'un objet et de sa géométrie en WKT avec GeoSPARQL :

```

ex:Monument1 a ex:Monument;
  rdfs:label "Washington Monument";
  geo:hasGeometry ex:Point1.
ex:Point1 a geo:Point;
  geo:asWKT "POINT(-77.03524 38.889468)"^^geo:wktLiteral.

```

Les fonctions de GeoSPARQL

GeoSPARQL comprend aussi un ensemble de fonctions d'interrogation spatiale, permettant par exemple de retourner les points situés dans un certain rayon à partir d'un autre point (buffer) ou les points situés à l'intersection de deux géométries (intersection). La *Geometry Topology Extension* regroupe un ensemble de fonctions de requêtes topologiques applicables à des géométries de type littéral, qui permettent ainsi d'effectuer des requêtes sur des géométries écrites au format WKT, comme par exemple les fonctions `sfIntersects` ou `sfWithin` (OGC 2012).

L'exemple de requête suivant (Kolas et Battle 2012) demande : « Quels monuments sont-ils compris dans quels parcs ? » :

```

SELECT ?f
WHERE {
  ex:Park1 geo:hasGeometry ?g1 .
  ?f a ex:Monument;
    geo:hasGeometry ?g2 .
  ?g2 geo:sfWithin ?g1 .
}

```

Il est aussi possible d'utiliser les fonctions de topologie dans la clause `FILTER` d'une requête, comme dans l'exemple suivant, qui permet de sélectionner des objets possédant une géométrie au format WKT et localisés à l'intérieur d'un polygone décrit directement dans le filtre de la requête :

```

SELECT ?f
WHERE {
  ?f geo:hasGeometry ?g .
  ?g geo:asWKT ?gWKT .
  FILTER (geof:sfWithin(?gWKT,
    "POLYGON ((-77.2 38.8, -77 38.8, -77 39, -77.2 39.9,
              -77.2 38.8))"^^geo:wktLiteral))
}

```

2.4.4. Linked Data et données au format raster

Les chapitres 2.4.1. à 2.4.3 montrent qu'il est relativement aisé de transformer en Linked Data des entités géospatiales représentées par des données vectorielles, qu'elles soient écrites sous forme de littéraux WKT comme au chapitre 2.4.3. ou sous d'autres formes, par exemple en utilisant le standard d'encodage GML (Geographic Markup Language : <http://www.opengeospatial.org/standards/gml>), un autre standard de l'OGC. Il est en effet simple d'attribuer à chaque objet vectoriel (un point, une ligne ou un polygone) une URI qui pourra être liée à d'autres URI par des prédicats.

En revanche, la production de Linked Data à partir de données raster s'avère plus compliquée. Dans un raster, dont la vocation est de représenter des champs pouvant correspondre à des catégories ou à des valeurs continues, l'équivalent d'une entité spatiale est en effet un ensemble de pixels répondant à une certaine définition contextuelle. Concrètement, il peut s'agir soit d'objets du monde réel, comme une route ou une rivière, représentés par des ensembles de pixels connectés et

correspondant à une même catégorie, soit à une surface de pixels comprises dans une certaine plage de valeurs d'un champ numérique continu, comme par exemple les bornes altitudinales de l'habitat des marmottes (Scharrenbach et al. 2012). Pour produire du Linked Raster Data, on peut donc recourir soit à des entités explicites (« explicit features »), c'est-à-dire des objets géographiques rasterisés, soit à des descriptions d'entités (« feature descriptions ») (Scharrenbach et al. 2012 ; Schüpfer 2014). L'approche « explicit feature » nécessite une méthode d'extraction de l'entité que l'on veut représenter. Par exemple, une montagne peut être représentée par un ensemble de pixels auquel une URI sera attribuée, cette méthode impliquant toutefois une difficulté quant à la définition des limites exactes de l'objet dans le raster : comment déterminer où commence et où se termine une montagne dans un raster d'élévation ? (Schüpfer 2014).

Un exemple de ce type d'approche est une représentation du Meteor Crater en Arizona à partir de données raster, qui soulève l'importance de la connexion entre un objet ontologique et une entité géographique du monde réel (Usery & Varanka 2012). Cette méthode nécessite d'identifier une entité dans le raster et de choisir un pixel ou un ensemble de pixels représentant son empreinte géométrique de base. Un seul pixel suffirait pour identifier une entité pouvant être traitée comme un point à la résolution du raster (comme par exemple une source), alors qu'un groupe de pixels contigus ou non-contigus est nécessaire pour représenter des entités s'étendant sur des aires importantes comme le cratère cité dans l'exemple (Usery & Varanka 2012). Les définitions, attributs et relations, y compris spatiales, de l'entité sont ensuite identifiées, permettant de définir une « entité ontologique » qui peut être liée à l'empreinte géométrique sur le raster précédemment établie. Cette géométrie raster est structurée à l'aide du standard GML Coverage, en utilisant notamment l'élément gml:Grid, renvoyant à un rectangle à limite minimum (« Minimum Bounding Rectangle » ou MBR) (Usery & Varanka 2012).

Une autre possibilité consiste à identifier un objet par la position des cellules correspondantes dans le raster et de ne stocker que les métadonnées du raster et la position de l'objet dans un fichier RDF (Scharrenbach et al. 2012 ; Schüpfer 2014).

L'approche « features descriptions » consiste à décrire une opération de géotraitement paramétrée qui calcule un résultat en fonction d'informations contextuelles supplémentaires. Pour pouvoir interroger des données raster avec le langage SPARQL, il faut d'abord lire les métadonnées du fichier raster et les rendre accessibles sous forme de graphe RDF. F. Schüpfer (2014) a notamment proposé une méthode permettant de lire et de convertir en triplets RDF les métadonnées d'un raster, mais aussi d'effectuer à l'aide de plusieurs fonctions des opérations spatiales sur les données avec le langage SPARQL, en faisant appel à plusieurs bibliothèques Python (GDAL, RdfLib).

La conversion et l'utilisation de données raster constituant une approche plus complexe et un sujet de recherche à part entière, les travaux réalisés dans le cadre de ce stage se sont uniquement intéressés à la traduction et à l'exploitation au format RDF de données vectorielles. Aborder les données raster du PREVIEW selon le concept de Linked Data serait une étape de développement ultérieure souhaitable.

2.4.5. Exemples d'applications géospatiales du Linked Data

Plusieurs projets ont récemment proposé divers outils ou applications permettant d'exploiter le concept de Linked Data dans le cadre d'une approche géospatiale :

-Le projet [Sextant](#), dont le site a fait l'objet d'une récente et importante mise à jour, est une plateforme dédiée à l'exploration et à la visualisation de données géospatiales au format Linked Data. Sextant offre l'accès à de nombreuses ressources et outils et permet de produire des cartes thématiques en combinant données géospatiales et temporelles. Développé dans le cadre du projet européen TELEIOS, Sextant utilise la technologie de [Strabon](#), qui est un entrepôt RDF spatio-temporel permettant d'interroger des données au format stRDF avec le langage stSPARQL (voir 2.4.2) ;

-Faisant partie des nombreux outils du projet GeoKnow (<http://geoknow.eu/Welcome.html>), Facete2 (<https://github.com/GeoKnow/Facete2>) est une application web de recherche facettée et de visualisation de données géospatiales RDF fonctionnant avec le standard SPARQL ;

-FAGI-gis (<https://www.w3.org/community/geosemweb/2015/10/05/fagi-gis-fusing-geospatial-rdf-data/>) est un autre outil issu du projet GeoKnow, qui permet de fusionner des entités liées issues de deux sources de données distinctes, ces dernières pouvant être des fichiers ou des endpoints SPARQL ;

-[map4rdf](http://oeg-dev.dia.fi.upm.es/map4rdf/) (<http://oeg-dev.dia.fi.upm.es/map4rdf/>) est un autre exemple d'application web de recherche facettée permettant d'explorer et de cartographier des données géospatiales au format RDF. Elle permet de visualiser divers types de géométries en utilisant le modèle de données de GeoLinkedData (<http://geo.linkeddata.es/web/guest/home;jsessionid=7F2708A83AD3E233BCE2ADC6BC88B64A>).

-GeoQuery (Wiegand et al. 2014) est une proposition d'interface de requêtes sur des données spatiales utilisant le standard GeoSPARQL ;

Plusieurs initiatives mettent également à disposition d'importantes quantités de données géospatiales :

-le projet communautaire le plus visible est sans doute celui de LinkedGeoData (<http://linkedgeodata.org/About>), dont le but est d'enrichir la dimension spatiale du Web sémantique en mettant à disposition les données du projet OpenStreetMap au format RDF. Les données de LinkedGeoData sont notamment accessibles via un endpoint SPARQL Virtuoso : <http://linkedgeodata.org/sparql>;

-le projet Geonames (<http://www.geonames.org/>), qui permet de télécharger sous divers formats plus de huit millions de noms d'endroits, possède sa propre ontologie et une version RDF de ses ressources via plusieurs accès. Les données RDF sont ainsi téléchargeables en un seul document rassemblant 150 millions de triplets. L'accès au RDF des ressources individuelles par l'interface cartographique Mother earth semble en revanche ne pas fonctionner et aucun accès via un endpoint SPARQL n'est malheureusement proposé sur le site.

3. Workflow pour la préparation et la publication d'un jeu de données du PREVIEW au format Linked Data

3.1. Introduction

Conformément aux objectifs de départ et à l'issue de la revue des outils disponibles, un schéma méthodologique (fig.7) comprenant les principales étapes suivantes a été défini :

-traduction au format RDF d'un jeu de données spatiales du GRID au format shapefile, consistant en un modèle d'emprise spatiale des cyclones tropicaux. Cette étape inclut la transformation des polygones issus du shapefile en objets géométriques au format WKT ;

-traduction au format RDF d'un jeu de données issu de la base de données des catastrophes de l'EM-DAT ;

-interconnexion des données PREVIEW avec celles de l'EM-DAT ainsi qu'avec une source de données extérieure, en l'occurrence le projet DBpedia ;

-publication et partage des données liées via un triplestore et un endpoint SPARQL ;

-élaboration d'un démonstrateur sous la forme d'une application web, permettant de présenter et d'interroger ces données, y compris par l'intermédiaire de requêtes spatiales.

Les différentes applications utilisées pour la préparation et la publication des données ainsi que le site et les applications web du démonstrateur ont été installées sur une machine Debian 8 Jessie hébergée au GRID.

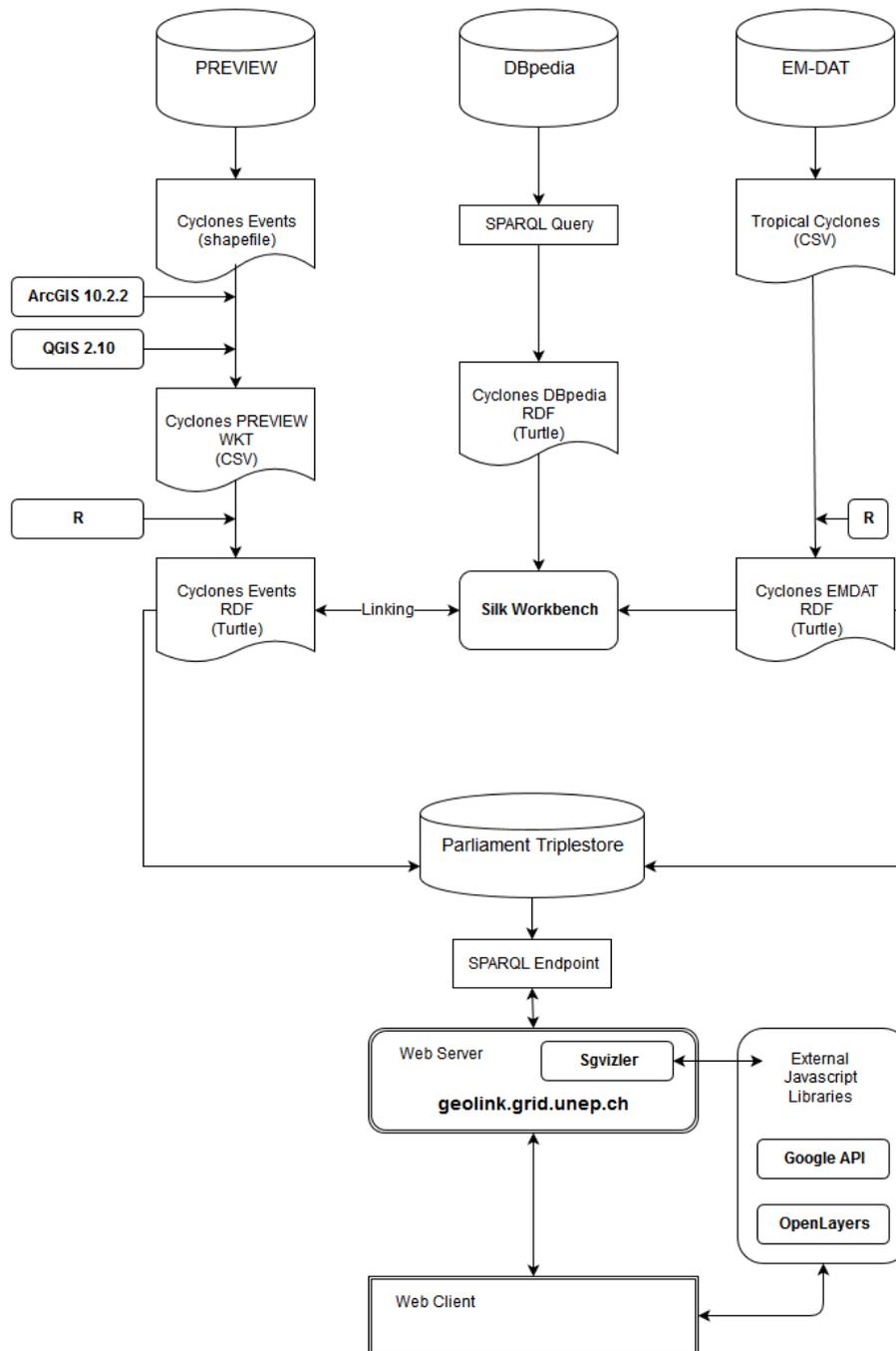


Fig. 7. Schéma général du workflow de préparation et de publication des données.

3.2. Données

Deux principales sources de données relatives aux cyclones tropicaux ont été utilisées pendant ce travail, soit le PREVIEW et la base de données des catastrophes EM-DAT. Des données issues du projet DBpedia ont également été exploitées afin de tester les possibilités de liens entre jeux de données différents.

3.2.1. PREVIEW

Le jeu de données Cyclones-Events a été téléchargé depuis le PREVIEW en format shapefile:

<http://preview.grid.unep.ch/index.php?preview=data&events=cyclones&evcat=1&lang=eng>

Il contient une modélisation sous forme de buffers représentant les surfaces touchées par les différents événements, les buffers étant obtenus à partir des trajets des cyclones issus de la base de données Ibtracs (<https://www.ncdc.noaa.gov/ibtracs/>) et d'une formule prenant en compte la pression au centre des cyclones, la vitesse des vents (échelle de Saffir-Simpson : https://fr.wikipedia.org/wiki/%C3%89chelle_de_Saffir-Simpson) et d'autres variables. Le modèle a été élaboré à partir du modèle de Holland (1980), lui-même adapté d'une approche originale de Schloemer (1954) et adapté par l'UNEP/GRID (Giulani & Peduzzi 2011). Ce jeu de données est complexe puisque chaque événement consiste en autant de polygones que de catégories atteintes sur l'échelle de Saffir-Simpson, numérotées de 0 à 5, 0 correspondant à un événement resté au stade de tempête tropicale. Chaque polygone de cyclone/catégorie est lui-même redécoupé en fonction des polygones de frontières de pays touchés par l'évènement.

3.2.2. EM-DAT

Un fichier Excel contenant des données sur l'ensemble des catastrophes de type « Tempête » a été obtenu auprès de l'Emergency Events Database (EM-DAT : <http://www.emdat.be/>). Un sous-ensemble de ce fichier, limité au sous-type de catastrophe « Tropical Cyclone » a ensuite été préparé. Ce fichier contient des données caractérisant les impacts des cyclones, soit le nombre de personnes touchées (c'est-à-dire de personnes nécessitant une aide immédiate en besoins élémentaires de survie tels que nourriture, eau, abri, sanitaires ou assistance médicale), le nombre de sans-abris, de blessés et le nombre total de morts, ce dernier correspondant au total des morts et des disparus, ainsi que la somme totale des dégâts en milliers de dollars. Dans le cadre de ce travail, les principaux champs suivants ont été utilisés :

-Total deaths: somme du nombre de morts et de disparus ;

-Injured: personnes souffrant de blessures physiques, de trauma ou d'une maladie résultant directement de la catastrophe et nécessitant une assistance médicale urgente ;

-Homeless: Nombre de personnes dont le logement est détruit ou lourdement endommagé et ont donc besoin d'un abri suite à un évènement ;

-Affected: Personnes nécessitant une assistance immédiate pendant une période d'urgence, c'est-à-dire ayant besoin de moyens de survie élémentaires tels que nourriture, eau, abri, sanitaires et assistance médicale immédiate ;

-Total affected: Somme des catégories Injured, Homeless et Affected;

-Estimated Damage: Le montant des dommages aux propriétés, cultures et bétail. Les dommages estimés sont donnés en milliers de dollars US ('000). Pour chaque catastrophe, le chiffre correspond à la valeur des dommages au moment de l'évènement, c'est-à-dire les valeurs réelles pour l'année de l'évènement.

3.2.3. DBpedia

Un jeu de données supplémentaire a été récupéré via une requête sur le endpoint SPARQL de DBpedia (<http://wiki.dbpedia.org/>). Ce jeu, rassemblant l'ensemble des villes de plus de 10'000 habitants disponibles sur DBpedia, était destiné à réaliser des requêtes spatiales à l'aide de fonctions du standard GeoSPARQL (voir 4.2.3).

3.3. Choix des outils et des technologies

Les outils et technologies suivantes ont été utilisés pour la préparation, la publication et l'exploitation des données au format RDF (fig. 7) :

-Des outils d'ArcGis 10.2.2. et QGIS 2.10 Pisa ont été exploités pour la simplification des données du shapefile Cyclones Events du PREVIEW (voir 3.4.1) ;

-La plateforme Datalift (<http://datalift.org/>) a largement été sollicitée lors de l'élaboration de ce travail. Issue d'un projet de l'ANR (Agence Nationale de la Recherche, France) lancé en 2010, la plateforme Datalift rassemble de nombreux outils permettant de publier et de lier des jeux de données au format Linked Data à partir de sources diverses (CSV, XML, GML,...). Elle offre aussi un outil permettant de convertir des shapefiles au format en RDF en traduisant les entités sous forme de littéral WKT, ainsi qu'une interface de traduction de vocabulaire d'une ontologie locale vers des vocabulaires existants. Datalift intègre également un endpoint SPARQL permettant d'interroger les données RDF stockées localement ou publiées, ainsi que des interfaces offrant la possibilité de lier des données avec un module Silk ou encore de visualiser des données publiées.

Dans le cadre de ce travail, la plateforme Datalift a été sollicitée surtout dans une optique d'apprentissage de l'écriture, de la publication et de l'interrogation de données au format RDF. Lors de l'élaboration des jeux de données définitifs, une approche plus simple comprenant l'utilisation d'outils développés pendant le stage a été préférée afin de maîtriser complètement toutes les étapes de préparation et de publication des données (voir 3.4 à 3.6). La procédure de publication d'un graphe RDF à partir de données CSV avec Datalift est toutefois décrite au chapitre 3.5.3.

Datalift présentait en outre un inconvénient majeur par rapport aux objectifs de ce travail, puisqu'il ne supporte pas le standard GeoSPARQL et n'offre donc pas la possibilité d'effectuer des requêtes spatiale sur les données publiées. C'est pour cette raison que les graphes RDF définitifs ont été publiés en utilisant la technologie de Parliament (voir 3.9) ;

-La traduction des données PREVIEW et EM-DAT au format RDF a ainsi fait recours à de simples scripts écrits avec R. R a également été sollicité pour la visualisation de données publiées (voir 4.4) ;

-L'outil Silk Workbench (<http://silk-framework.com/>) a été utilisé pour lier les jeux de données du PREVIEW, de l'EM-DAT et de DBpedia. Cet outil offre notamment une interface d'établissement de liens permettant d'utiliser diverses méthode de modification, d'agrégation et de comparaison des données chargées sous forme de graphes RDF (voir 3.7).

-Un entrepôt RDF Parliament (<http://parliament.semwebcentral.org/>) a été installé pour la publication et l'interrogation des graphes. Parliament est un triplestore complet qui permet d'entreposer et de modifier des données. Outre les standards RDF, RDFS, OWL et SPARQL, Parliament est compatible avec le standard OGC GeoSPARQL et permet ainsi de sélectionner des données spatiales en utilisant les fonctions de topologie décrites en 2.4.3. Parliament a été préféré à la solution du triplestore Strabon, ce dernier nécessitant la maîtrise de standards supplémentaires (stRDF et stSPARQL) et ne permettant pas encore l'utilisation de toutes les fonctionnalités de GeoSPARQL.

-La bibliothèque Javascript Sgvizler (Skjæveland 2012) a été utilisée pour élaborer les applications web du démonstrateur (voir 4.2.). Sgvizler offre une collection d'outils permettant de visualiser sous

forme de tableaux, de graphiques mais aussi de cartes les résultats de requêtes SPARQL. Son utilisation était donc parfaitement adaptée aux objectifs de ce travail.

La combinaison de Parliament et de Sgvizler pour stocker, interroger et visualiser les données spatiales dans le cadre de ce stage a été retenue car elle permettait une approche didactique des principaux concepts et standards liés au Linked Data ainsi qu'une grande souplesse dans l'élaboration du démonstrateur. Les solutions d'application présentées dans le chapitre 2.4.5 telles que Sextant ou Facete2, plus élaborées mais pas forcément faciles d'accès et surtout beaucoup moins souples compte tenu des objectifs fixés, n'ont donc pas été retenues.

3.4. Préparation des données

3.4.1. Simplification des données du PREVIEW

Le fichier shapefile Cyclones-Events du PREVIEW a d'abord fait l'objet d'une simplification. En effet, ce jeu de données est complexe puisque chaque cyclone est associé à autant de polygones que de catégories atteintes sur l'échelle de Saffir-Simpson et de pays touchés par chacun de ces polygones de cyclones/catégories. Le shapefile de limites administratives utilisé par le PREVIEW pour découper les empreintes de cyclones étant très précis, certains évènements touchant des côtes aux contours irréguliers peuvent être associés à des polygones possédant un nombre de sommets très élevés. Les îles peuvent également produire de nombreux polygones pour une seule catégorie de vents d'un évènement donné (fig. 8).

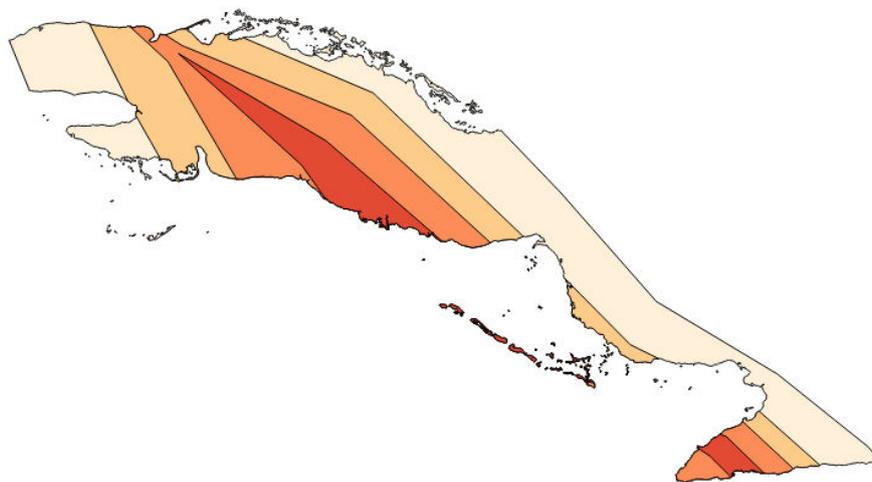


Fig. 8. Exemple d'évènement lié à des polygones complexes dans le jeu de données Cyclones-Events du PREVIEW : polygones représentant le cyclone Dennis (2005) lorsqu'il était classé comme ouragan avec des vitesses de vents classées de 1 à 4 (rouge foncé) sur l'échelle de Saffir-Simpson et qu'il a touché Cuba.

Le jeu de données devant être traduit sous forme de littéral WKT, il était important de réduire le nombre de sommets et donc de simplifier le contour des polygones. L'objectif de ce travail étant également de tester des fonctions de géotraitement du langage GEOSPARQL, il fallait aussi revenir à des polygones non découpés selon les pays touchés et représentant donc l'emprise totale de chaque évènement/catégorie de Saffir-Simpson.

Pour réaliser cette étape, l'ensemble du jeu de données a fait l'objet d'une opération de regroupement des polygones avec l'outil Dissolve d'ArcGis, en utilisant un champ (DISSID) concaténant le numéro d'évènement (ev_id) et la catégorie de Saffir-Simpson (ss_cat). Lors de cette opération, seuls les champs year (stat=min), other_name (stat=first) et serial_num (le numéro de série l'abrégé ; stat=first) ont été conservés via le calcul de statistiques (min et first) (fig. 9).

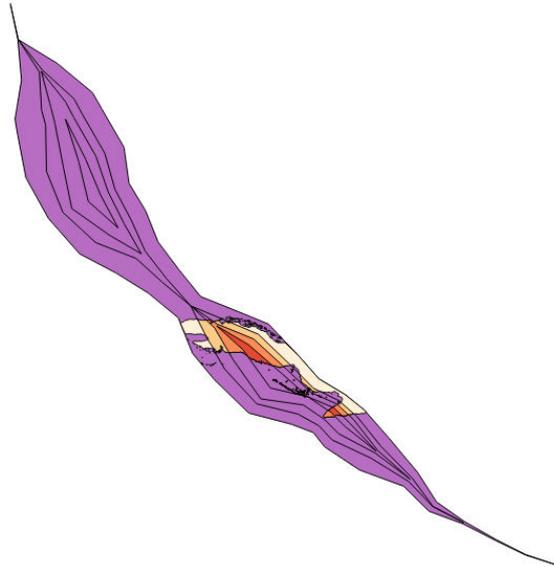


Fig. 9. En violet, exemple de résultat du géotraitement « Dissolve » sur le shapefile Cyclone-Event : emprise totale des polygones représentant l'ouragan Dennis (2005) atteignant les catégories 1 à 4 sur l'échelle de Saffir-Simpson. On voit ainsi toute l'ampleur de l'évènement et non pas seulement son impact sur Cuba tel que montré sur la figure 8.

Une jointure avec le fichier d'origine a ensuite été réalisée sur le champ concaténé DISSID utilisé pour le géotraitement afin de récupérer les champs écartés lors du dissolve. Seule l'information sur les pays touchés (iso3) est encore manquante à l'issue de cette étape (voir *infra*).

Une opération supplémentaire de simplification des polygones a ensuite consisté à réduire le nombre de leurs sommets afin de produire des géométries WKT moins volumineuses, en utilisant l'outil « Simplifier la géométrie » de QGIS (Vecteur->Outils de géométrie), avec une tolérance de 0.001 permettant de ne pas obtenir, à l'inverse, des polygones aux contours trop anguleux.

Enfin, la récupération du nom des pays touchés par les polygones de cyclones/catégories a été réalisée en effectuant une jointure spatiale entre le shapefile des cyclones et un shapefile de limites administratives nommé TM WORLD BORDERS, téléchargé sur le site Thematicmapping.org (http://thematicmapping.org/downloads/world_borders.php). Le but était d'obtenir un polygone par catégorie et par pays touché pour chaque évènement avec un champ contenant le nom du pays concerné. Cette opération a été réalisée avec l'outil Spatial join d'Arcgis (toolbox Analysis) en utilisant l'option JOIN_ONE_TO_MANY (fig. 10) afin qu'en cas de contact entre un cyclone et plusieurs pays, un polygone soit créé par cyclone/catégorie et par pays (le fichier résultant possède par exemple trois polygones différents représentant l'ouragan Katrina de 2005 en catégorie 0, l'évènement ayant touché les Bahamas, Cuba et les USA).

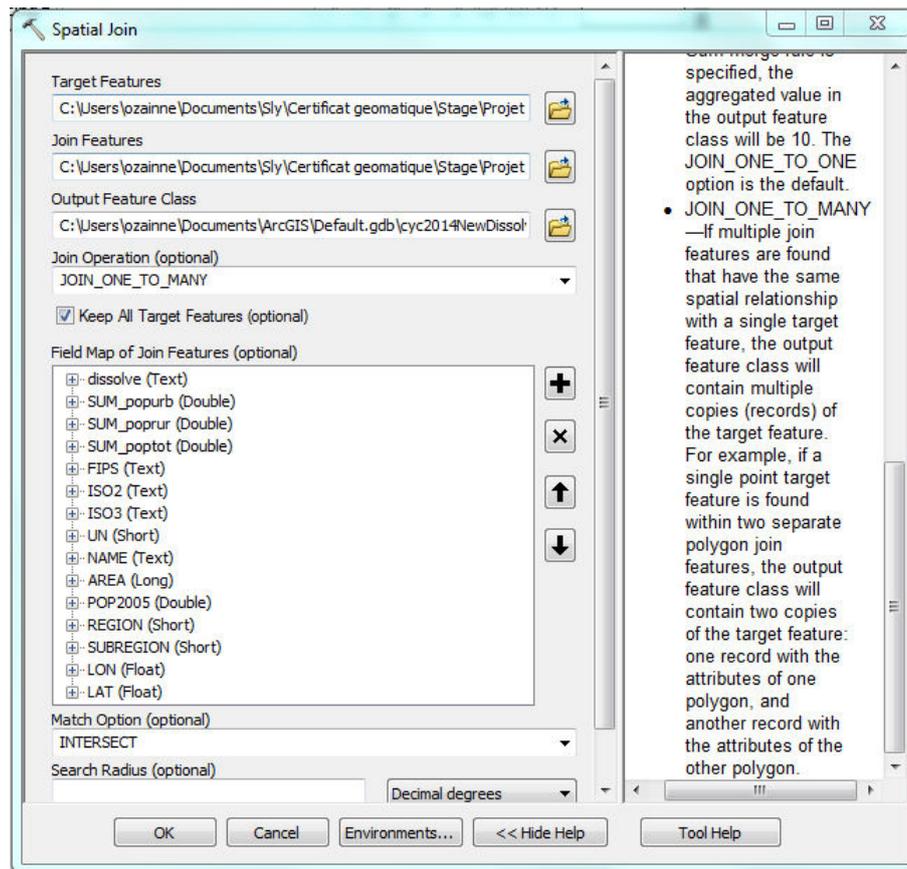


Fig. 10. Attribution des noms de pays à chaque polygone avec l’outil Spatial Join d’Arc Toolbox.

3.4.2. Enregistrement du shapefile de cyclones PREVIEW au format WKT

Le passage au format WKT a été réalisé avec QGIS, qui permet cette opération grâce à un simple « Enregistrer sous ». Il suffit d’enregistrer la couche choisie en tant que document CSV, en précisant le format de la géométrie désirée l’option GEOMETRY, soit AS_WKT, comme indiqué sur la figure 11 :

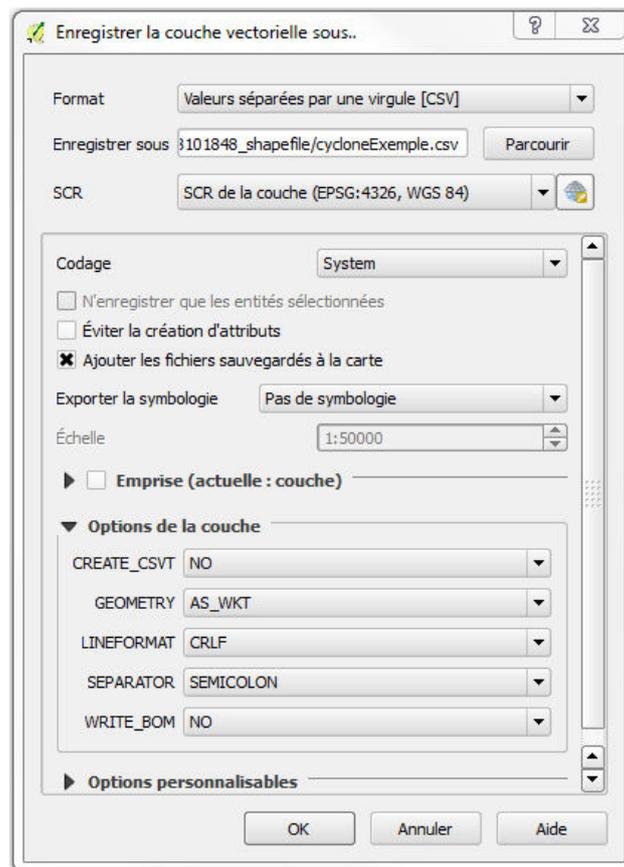


Fig. 11. Enregistrement du fichier shapefile simplifié en format WKT avec QGIS (clic droit sur la couche ->Enregistrer sous)

Il est important de relever ici que Datalift dispose d'un outil de traduction de shapefiles en graphes RDF, traduisant des points, lignes ou polygones également au format WKT. Cet outil n'a toutefois pas été employé, les graphes étant produits selon une modélisation complexe décrivant chaque sommet des polygones non seulement sous forme de littéraux WKT mais aussi en coordonnées utilisant une ontologie de l'IGN. Cette modélisation aboutissant à un graphe volumineux et peu aisé à interroger, un workflow plus artisanal a été conçu, impliquant la traduction des shapefiles en WKT avec QGIS (voir aussi la méthode d'écriture des graphes en 3.5.3).

3.4.3. Aménagement du fichier PREVIEW enregistré en CSV

Le fichier shapefile des cyclones converti en WKT et enregistré comme simple fichier CSV a ensuite fait l'objet de plusieurs aménagements dans l'optique de faciliter son passage au standard RDF. Les enregistrements ont notamment été triés en ordre croissant par année, par nom d'évènement, par nom de pays et par numéro de catégorie de Saffir-Simpson. Les noms d'évènements ont été simplifiés, le champ other_name issu de la base de données lbracs contenant parfois des valeurs complexes et répétitives. Un champ StartMonth a également été ajouté dans l'optique de faciliter les liens avec les données de l'EM-DAT. L'objectif de ce travail étant de tester l'établissement de liens sémantiques entre le jeu de données spatiales du PREVIEW et les statistiques de l'EM-DAT, aucune procédure plus complexe permettant d'établir une correspondance parfaite entre les deux jeux de données à partir de leur date n'a été élaborée. En revanche, une uniformisation des noms des évènements entre les deux jeux de données a été réalisée lorsqu'elle était possible (voir 3.7). Les enregistrements du champ Pays pour lesquels aucune valeur n'était disponible ont été systématiquement renommés avec NoData.

3.4.4. Préparation des données de l'EM-DAT

Comme pour le fichier PREVIEW passé au format WKT et enregistré en CSV, les enregistrements de l'EM-DAT ont fait l'objet d'un tri croissant dans l'ordre année, nom de l'évènement, nom de pays. Comme décrit précédemment, les noms d'évènements ont été uniformisés avec ceux du PREVIEW lorsque c'était possible, afin de faciliter les futurs liens entre les deux jeux de données.

3.5. Préparation du graphe RDF des données du PREVIEW

3.5.1. Modélisation des données au format RDF

Une première étape a consisté à définir la modélisation du graphe RDF des données PREVIEW, impliquant une définition des URI de base associées à chacune de ses ressources.

Une modélisation simple a été définie, chaque ressource correspondant aux buffers simplifiés issus du jeu de données Cyclones Events du PREVIEW. Chaque buffer est rattaché à un évènement via le numéro de série de la base de données IbTracs, permettant une identification précise des évènements et de pouvoir envisager une modélisation plus complexe ultérieurement.

Dans le cadre de ce travail, la modélisation s'est donc bornée à reproduire une version simplifiée du shapefile de départ issu du PREVIEW correspondant à un ensemble de polygones. Des ressources représentant chaque évènement unique (donc chaque numéro de série IbTracs) n'ont pas été créées.

Chaque ressource (correspondant à chaque buffer) est identifiée par l'URI de base <http://geolink.grid.unep.ch/tropCyc#> et le numéro de buffer auquel elle correspond, par exemple <http://geolink.grid.unep.ch/tropCyc#Buffer1>. Chaque ressource est définie par différentes propriétés (prédicats) que sont l'année (year), le nom du cyclone (label), etc., qui sont des champs issus du shapefile d'origine. En l'état, l'URI de chaque ressource renvoie à une information simple, donnée par un commentaire général généré automatiquement lors de l'écriture des fichiers RDF et stocké dans un tableau HTML constituant une liste des ressources disponibles, chaque ressource (ou ligne du tableau) étant directement accessible grâce au # précédent chaque identifiant dans les URIs. Chaque URI est donc déréférencée en offrant une description de la ressource lisible par l'humain, sous forme de tableau HTML. Une étape ultérieure serait de déréférencer les URI à un format lisible par les machines, en publiant par exemple l'ensemble des ressources en format RDF/XML.

Chaque ressource correspond également à une géométrie selon le standard GEOSPARQL de l'OGC, les géométries étant elles-mêmes de nouvelles ressources identifiées de la façon suivante : <http://geolink.grid.unep.ch/tropCyc#MULTIPOL1>

Chacun de ces objets est lui-même caractérisé par un ensemble de coordonnées écrit sous forme de littéral au format WKT. On respecte ainsi la modélisation de données spatiales avec GeoSPARQL telle que présentée dans le chapitre 2.4.3. Il est indispensable que les polygones écrits sous forme de littéraux WKT soient typés, en indiquant après chaque chaîne de caractères qu'elle correspond à un littéral WKT respectant le standard GEOSPARQL :

^^<http://www.opengis.net/ont/geosparql#wktLiteral>.

La modélisation du graphe PREVIEW est résumée par la fig. 12 :

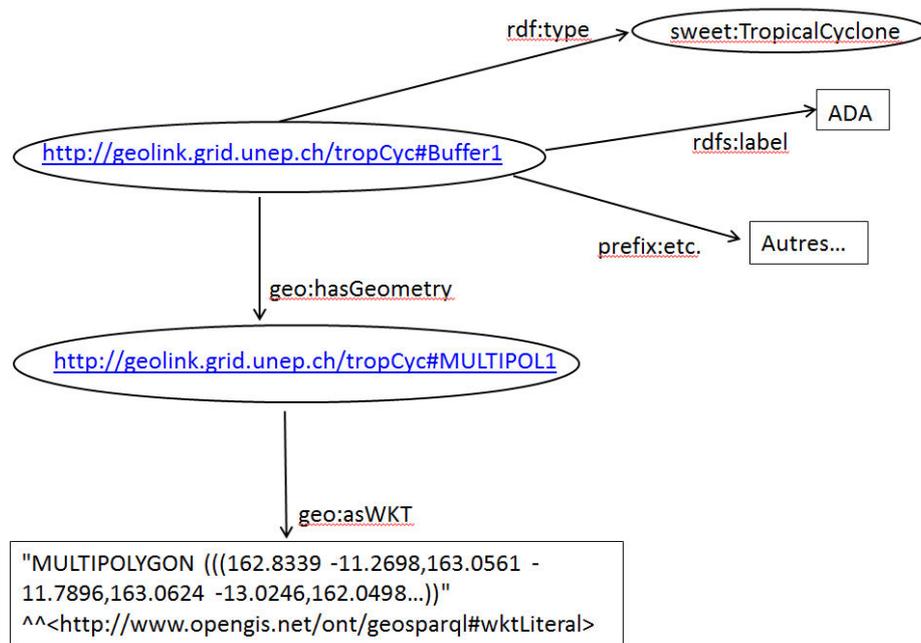


Fig. 12. Modélisation du graphe RDF du jeu de données des cyclones du PREVIEW. Le préfixe « etc. » et le littéral « Autres... » servent à résumer la présence dans le graphe de plusieurs autres propriétés et valeurs de propriétés correspondant aux différents champs issus du PREVIEW.

3.5.2. Choix du vocabulaire

Le vocabulaire choisi pour décrire les différentes propriétés de chaque ressource utilise le plus possible des ontologies existantes. Des prédicats internes écrits directement à partir des titres de champs du fichier CSV ont également été utilisés pour simplifier la production des fichiers RDF.

RDF, RDF Schema, OWL

Le prédicat <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> (que l'on peut abrégé `a`) a été utilisé pour désigner le type de chaque ressource. Le prédicat <http://www.w3.org/2000/01/rdf-schema#comment> du `rdf-schema` désigne pour sa part le commentaire automatiquement généré pour chaque ressource lors de l'écriture du fichier RDF (voir chapitre 3.5.3). Ce commentaire est également utilisé pour décrire rapidement les URIs correspondant à chaque ressource (voir par exemple <http://geolink.grid.unep.ch/tropCyc#Buffer4563>).

Le prédicat <http://www.w3.org/2000/01/rdf-schema#label> a été utilisé pour décrire le nom de l'évènement associé à chaque buffer.

Enfin, le prédicat <http://www.w3.org/2002/07/owl#sameAs> a servi à lier chaque ressource du jeu de données PREVIEW aux ressources décrivant le même évènement dans le jeu de données de l'EM-DAT.

Ontologies SWEET (NASA)

Les ontologies SWEET (Semantic Web for Earth and Environmental Terminology) regroupent plus de 6000 concepts. Deux ontologies ont été utilisées, d'une part pour associer le type « TropicalCyclone » à chaque ressource (<http://sweet.jpl.nasa.gov/2.3/phenAtmoPressure.owl#TropicalCyclone>), d'autre part pour désigner la catégorie de l'échelle de Saffir-Simpson à laquelle est associée chaque ressource (<http://sweet.jpl.nasa.gov/2.3/stateStorm.owl#SaffirSimpsonScale>).

FAO

L'ontologie géopolitique de la FAO a été utilisée pour les prédicats désignant le code ISO3 correspondant à chaque pays (<http://aims.fao.org/aos/geopolitical.owl#codeISO3>) ainsi que le nom de chaque pays (<http://aims.fao.org/aos/geopolitical.owl#nameListEN>).

GEOSPARQL

Le prédicat du standard GEOSPARQL <http://www.opengis.net/ont/geosparql#hasGeometry> a servi à assigner des géométries à chaque ressource, le prédicat <http://www.opengis.net/ont/geosparql#asWKT> permettant à son tour de décrire chaque géométrie par un littéral WKT.

DBpedia

Le prédicat de DBpedia <http://dbpedia.org/property/year> a été utilisé pour définir l'année associée à chaque ressource.

Prédicats « internes »

Pour décrire les autres propriétés, des prédicats internes au jeu de données ont été générés automatiquement à partir des noms de colonnes du document CSV contenant l'ensemble des ressources et leurs géométries. Les prédicats suivants ont ainsi été créés pour décrire le numéro de série IbTracs, le mois de début de chaque évènement ainsi que leurs dates précises de début et de fin :

<http://geolink.grid.unep.ch/tropCyc#ibtracsSerial>
<http://geolink.grid.unep.ch/tropCyc#StartMonth>
<http://geolink.grid.unep.ch/tropCyc#StartDate>
<http://geolink.grid.unep.ch/tropCyc#EndDate>

Une solution complète consisterait à créer et à publier une véritable ontologie définissant des classes et des prédicats permettant de décrire l'ensemble des champs utilisés par PREVIEW, comme par exemple pour le numéro de série IbTracs :

```
<owl:Ontology rdf:about="http://geolink.grid.unep.ch/previewOnt.xml#">
  <rdfs:comment>The GEOLINK ontology 0.1</rdfs:comment>
  <rdfs:label>Some properties for the PREVIEW Linked Dataset test
sample</rdfs:label>

  <rdf:Description rdf:about="#IbtracsSerial">
    <rdfs:comment>The IbTracs Serial Number property, as a unique
    identifier for distinct tropical cyclones events</rdfs:comment>
  </rdf:Description>

</owl:Ontology>
```

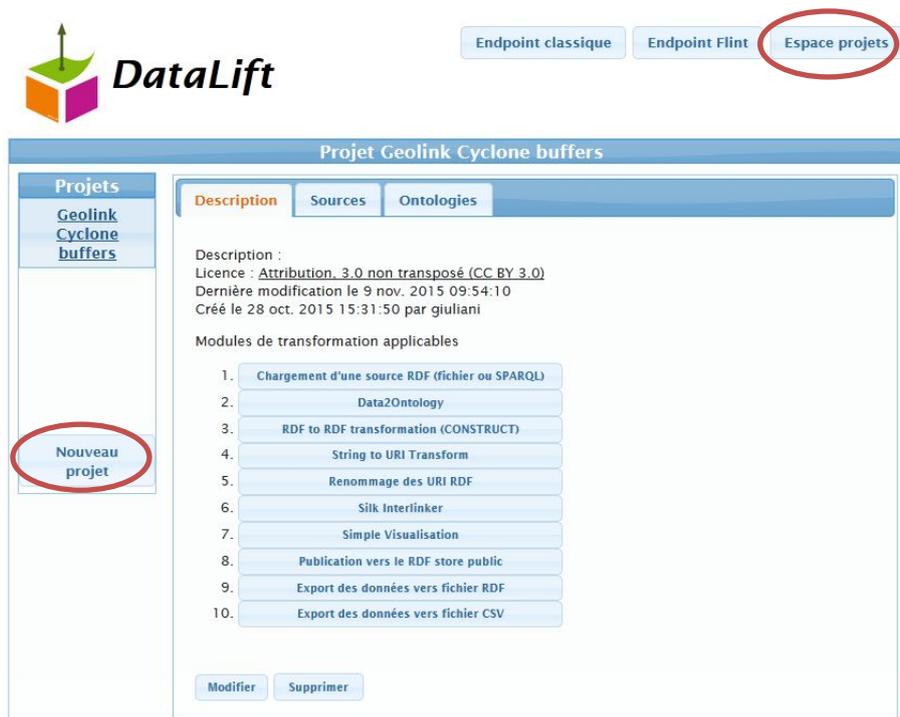
L'écriture d'une ontologie détaillée ne serait toutefois envisageable qu'à l'issue d'une modélisation complète et définitive des ressources du PREVIEW (voir chapitre 5.3)

3.5.3. Ecriture du graphe RDF des données PREVIEW

Utilisation de la plateforme DATALIFT

Plusieurs méthodes ont été testées pour écrire des graphes RDF dans le cadre de ce travail. Dans un premier temps, les outils de la plateforme Datalift ont été utilisés. L'installation de Datalift (http://wiki.datalift.org/index.php?title=Platform_installation#Stand-alone_Deployment) sur un serveur ou en local offre en effet accès à un grand nombre d'outils et permet de produire de simples graphes RDF à partir de données chargées par exemple en format CSV. La procédure pour charger un jeu de données au format CSV et le traduire en graphe RDF est la suivante :

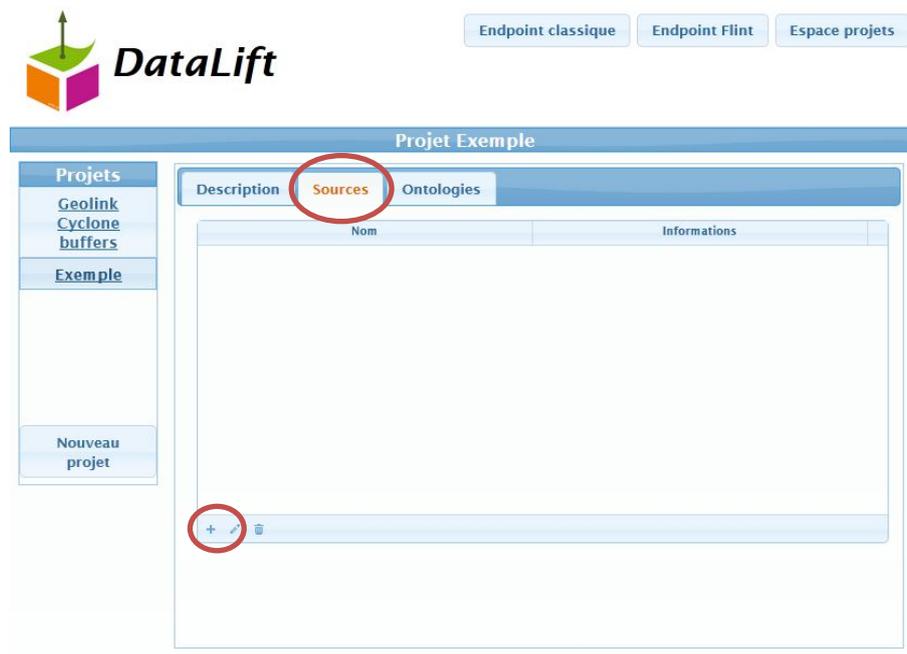
1) Aller sur l'espace projets et cliquer sur Nouveau projet :



2) Entrer le nom du projet et cliquer sur Ajouter :



3) Cliquer sur l'onglet Sources puis sur le signe + en bas à gauche de l'écran :



4) Cliquer sur CSV. Il est possible d'importer des données d'autre type, dont des fichiers shapefile ou GML :



5) Choisir la source de données :

Ajouter un jeu de données - Projet Exemple

[Retour à l'espace de travail](#)

▼ CSV

Fichier CSV à télécharger *:
 cyc2014Preview.csv

ou URL de téléchargement *:

Nom de la source *:

Description :

Encodage :

Séparé par * :

Position de la ligne de titres :

Première ligne de données :

Dernière ligne de données :

Délimiteur de texte multiligne :

6) Une fois la source chargée, elle est disponible dans la liste et peut être visualisée :



Projet Exemple

Projets

[Geolink](#)

[Cyclone](#)

[buffers](#)

[Exemple](#)

Description

Sources

Ontologies

Nom	Informations
cyc2014Preview.csv	Fichier : project/exemple/cyc2014Preview.csv Séparé par : point virgule, Ligne de titres : oui Créée le : 14 nov. 2015 15:41:45

7) Aperçu des données chargées au point précédent :



Endpoint classique Endpoint Flint Espace projets

Contenu de la source cyc2014Preview.csv (1000 premières entrées)

Rechercher dans Toutes les colonnes

	bufferID	ibtracsSer	evid	year	label	countryna	ISO3	ssCat	StartMont	StartDate	EndDate	WKT
1	1	1970003S1	tc0006	1970	ADA	Australia	AUS	0	1	19700110	19700110	MULTIPOLY
2	2	1970003S1	tc0006	1970	ADA	Australia	AUS	1	1	19700117	19700117	MULTIPOLY
3	3	1970003S1	tc0006	1970	ADA	Solomon Isl	SLB	0	1	19700110	19700110	MULTIPOLY
4	4	1970151N	tc0029	1970	ADELE	NoData	NoData	1	5	19700531	19700604	MULTIPOLY
5	5	1970151N	tc0029	1970	ADELE	NoData	NoData	0	6	19700602	19700602	MULTIPOLY
6	6	1970151N	tc0029	1970	ADELE	NoData	NoData	2	6	19700603	19700603	MULTIPOLY
7	7	1970138N	tc0028	1970	ALMA	Cayman Isl	CYM	0	5	19700519	19700519	MULTIPOLY
8	8	1970138N	tc0028	1970	ALMA	Colombia	COL	0	5	19700519	19700519	MULTIPOLY
9	9	1970138N	tc0028	1970	ALMA	Cuba	CUB	0	5	19700519	19700519	MULTIPOLY
10	10	1970138N	tc0028	1970	ALMA	United Stat	USA	0	5	19700519	19700519	MULTIPOLY
11	11	1970138N	tc0028	1970	ALMA	NoData	NoData	1	5	19700520	19700520	MULTIPOLY
12	12	1970227N	tc0054	1970	ANITA	Japan	JPN	1	8	19700820	19700822	MULTIPOLY
13	13	1970227N	tc0054	1970	ANITA	Japan	JPN	2	8	19700820	19700821	MULTIPOLY
14	14	1970227N	tc0054	1970	ANITA	Japan	JPN	0	8	19700821	19700821	MULTIPOLY

8) Pour traduire les données en RDF ; aller sur l'onglet Description et cliquer sur « Transformation directe CSV vers RDF » :



Endpoint classique Endpoint Flint Espace projets

Projet Exemple

Projets
Geolink
Cyclone
buffers
Exemple

Description Sources Ontologies

Description :
Licence : [Attribution 3.0 non transposé \(CC BY 3.0\)](#)
Dernière modification le 14 nov. 2015 15:41:45
Créé le 14 nov. 2015 15:29:02 par giuliani

Modules de transformation applicables

1. **Transformation directe CSV vers RDF**

Nouveau projet

Modifier Supprimer

9) Définir le nom du graphe et l'URI de base. Dans cet exemple, on laisse les valeurs par défaut. Il faut ensuite cliquer sur l'onglet « Typage des données » pour définir le format des différentes données :

DataLift

Transformation directe CSV vers RDF - Projet Exemple

Source à convertir :

Nom de la source à créer :

URI du graphe nommé associé :

URI de base :

Type RDF cible :

Identifiant d'objet :

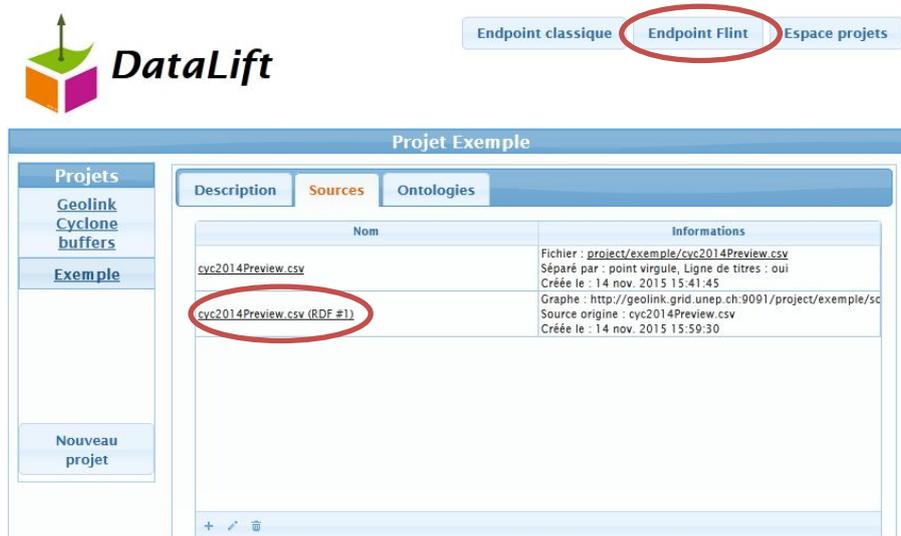
Typage des données

10) La valeur par défaut est « Chaîne de caractères », mais on veut ici par exemple que le champ « year » soit un nombre entier :

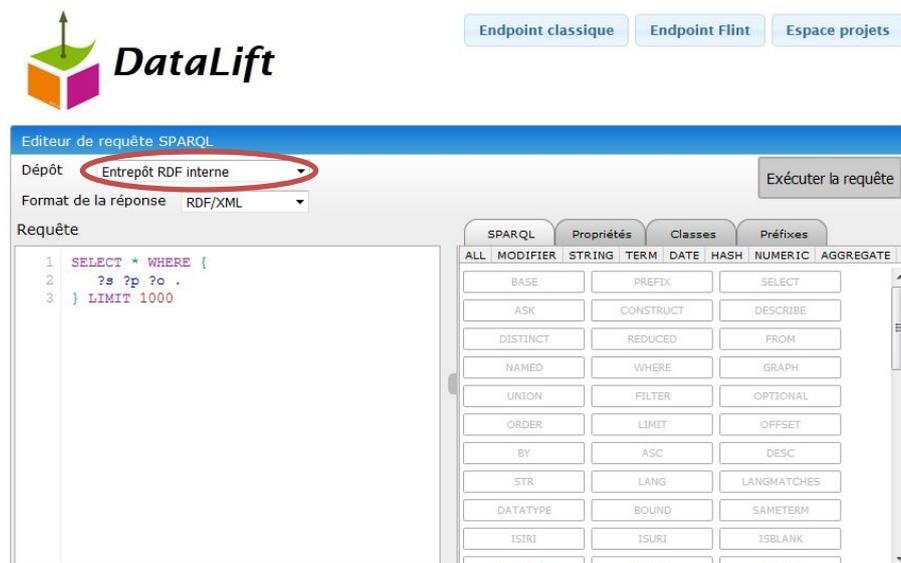
▼ Typage des données

Colonne "bufferID" :	<input type="text" value="Chaîne de caractères"/>
Colonne "ibtracsSerial" :	<input type="text" value="Chaîne de caractères"/>
Colonne "evid" :	<input type="text" value="Chaîne de caractères"/>
Colonne "year" :	<input type="text" value="Nombre entier"/>
Colonne "label" :	<input type="text" value="Chaîne de caractères"/>
Colonne "countryname" :	<input type="text" value="Chaîne de caractères"/>
Colonne "ISO3" :	<input type="text" value="Chaîne de caractères"/>
Colonne "ssCat" :	<input type="text" value="Chaîne de caractères"/>
Colonne "StartMonth" :	<input type="text" value="Chaîne de caractères"/>
Colonne "StartDate" :	<input type="text" value="Chaîne de caractères"/>
Colonne "EndDate" :	<input type="text" value="Chaîne de caractères"/>
Colonne "WKT" :	<input type="text" value="Chaîne de caractères"/>

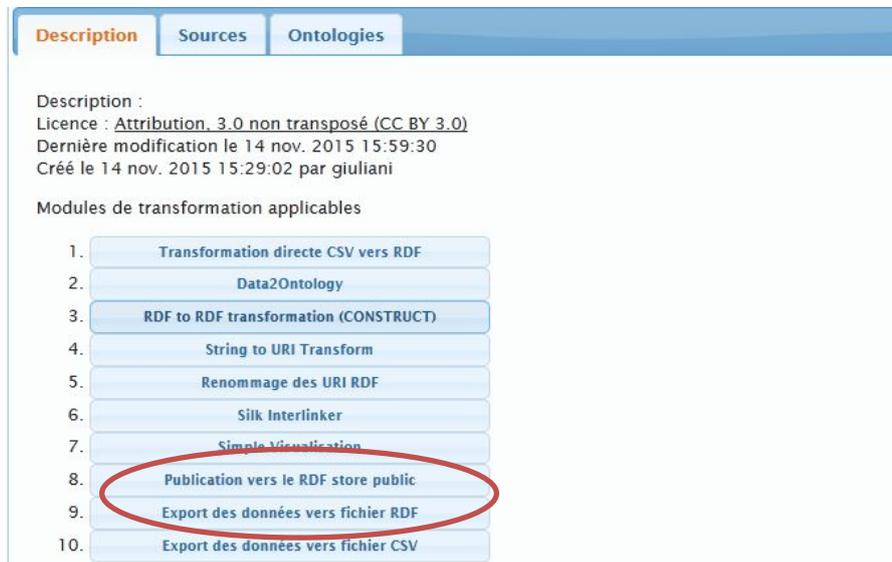
11) Une fois le graphe créé, il apparaît dans la liste des sources. Cliquer ensuite sur le bouton Endpoint Flint :



12) On accède alors à une version du endpoint SPARQL permettant d'inclure des commandes pré-enregistrées. Le graphe peut alors être interrogé avec une requête SPARQL. Il n'est pas encore publié, il faut donc préciser que l'on veut interroger l'entrepôt RDF interne :



13) En retournant à l'onglet Description, on peut ensuite, entre autres, publier le jeu de données ou l'exporter sous forme de fichier au format RDF, avec plusieurs syntaxes à choix (RDF/XML, Turtle, N-Triples, etc.). Cette option a notamment été employée pour réécrire proprement le graphe complet des données PREVIEW (voir 3.8) :



Datalift permet ainsi de publier très simplement et très rapidement des données structurées au format RDF. La plateforme intègre de nombreux autres outils, comme le Data2Ontology, qui permet de rechercher des termes d'ontologies existantes pour remplacer des prédicats internes. Si ces outils, très pratiques, ont été largement employés au début de ce travail pour écrire et publier des échantillons de données à tester, ils n'ont pas été sollicités lors de la préparation finale des graphes ni pour leur publication. L'outil de transcription RDF ne réalise en effet qu'une modélisation très simple à partir des champs d'un fichier CSV importé, et ne permettait notamment pas d'écrire directement le graphe RDF prenant en compte la modélisation des données PREVIEW telle que définie en 3.5.1.

Ecriture du graphe avec un script R

De nombreux autres outils d'écriture de RDF sont accessibles sur le web, comme par exemple Open Refine (<http://openrefine.org/>) et son extension RDF, ou encore Sheet to RDF (<http://art.uniroma2.it/sheet2rdf/>) ou Spread to RDF (<https://github.com/marcelotto/spread2rdf>).

Des solutions propriétaires existent aussi:

<http://www.topquadrant.com/tools/IDE-topbraid-composer-maestro-edition/>

<http://www.cambridgesemantics.com/products/anzo-express>

Un test d'Open Refine s'est avéré peu concluant, non pas à cause de l'outil lui-même qui fonctionnait correctement, mais en raison de sa dépendance pour l'écriture des fichiers de sortie à des outils externes sur URIBurner (<http://linkeddata.uriburner.com/>), qui étaient indisponibles lors de plusieurs essais réalisés à plusieurs jours d'intervalle au mois de juillet 2015. Il est intéressant de noter que ces mêmes outils en panne limitaient le fonctionnement de l'endpoint SPARQL de DBpedia au même moment.

Pour écrire le graphe RDF des données PREVIEW, il a finalement été choisi de créer un outil local et indépendant de ressources externes, consistant en un script R qui écrit tous les triplets du graphe à

partir des colonnes du fichier CSV préparé au chapitre 3.4. Le script intègre les prédicats choisis dans les ontologies existantes (voir 3.5.2.) ainsi qu'une fonction permettant d'écrire les prédicats internes à partir des en-têtes de colonnes du fichier. Il écrit ensuite les triplets au format Turtle, sans toutefois utiliser de préfixes pour abrégier les URIs, afin d'obtenir un fichier détaillé contenant des triplets complets. Le fichier résultant est enregistré en format Turtle, avec l'extension .ttl. Cette solution semi-automatique et artisanale est moins confortable que l'utilisation de Datalift, mais a permis beaucoup plus de souplesse ainsi que l'écriture directe sous forme de triplets du modèle de données défini à l'étape 3.5.1.

3.6. Préparation du graphe RDF des données EM-DAT

3.6.1. Modélisation et choix des vocabulaires

Dans le cadre de ce travail qui visait à tester l'établissement de liens entre les entités spatiales du PREVIEW et certains champs statistiques de l'EM-DAT, les données de l'EM-DAT ont fait l'objet d'une modélisation très simple et les URIs correspondant à chaque ressource ne sont pas déréférencées. Chaque ressource est donc simplement identifiée par l'URI de base <http://geolink.grid.unep.ch:9091/emdatCyc#> et un numéro d'identifiant, par exemple <http://geolink.grid.unep.ch:9091/emdatCyc#1>.

Chaque ressource est ensuite décrite en utilisant diverses propriétés correspondant aux champs du jeu de données utilisé. Comme pour le graphe du PREVIEW, des termes issus de plusieurs ontologies existantes ont été employés pour décrire les principales propriétés des ressources de l'EMDAT :

<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://sweet.jpl.nasa.gov/2.3/phenAtmoPressure.owl#TropicalCyclone>
<http://www.w3.org/2000/01/rdf-schema#label>
<http://aims.fao.org/aos/geopolitical.owl#nameListEN>
<http://aims.fao.org/aos/geopolitical.owl#codeISO3>
<http://dbpedia.org/property/year>

Les autres prédicats ont été générés automatiquement à partir des en-têtes de champs correspondant dans le document CSV préparé à l'étape 3.4.4 :

<http://geolink.grid.unep.ch/emdatCyc#region>
<http://geolink.grid.unep.ch/emdatCyc#continent>
<http://geolink.grid.unep.ch/emdatCyc#startYear>
<http://geolink.grid.unep.ch/emdatCyc#startMonth>
<http://geolink.grid.unep.ch/emdatCyc#startDay>
<http://geolink.grid.unep.ch/emdatCyc#endYear>
<http://geolink.grid.unep.ch/emdatCyc#endMonth>
<http://geolink.grid.unep.ch/emdatCyc#endDay>
<http://geolink.grid.unep.ch/emdatCyc#totalDeaths>
<http://geolink.grid.unep.ch/emdatCyc#noInjured>
<http://geolink.grid.unep.ch/emdatCyc#noAffected>
<http://geolink.grid.unep.ch/emdatCyc#noHomeless>
<http://geolink.grid.unep.ch/emdatCyc#totalAffected>
<http://geolink.grid.unep.ch/emdatCyc#totalDam1000US>

La modélisation des données EM-DAT n'inclut pas de dimension spatiale et présente une structure plus simple que celle du PREVIEW ne nécessitant pas l'utilisation du standard GeoSPARQL :

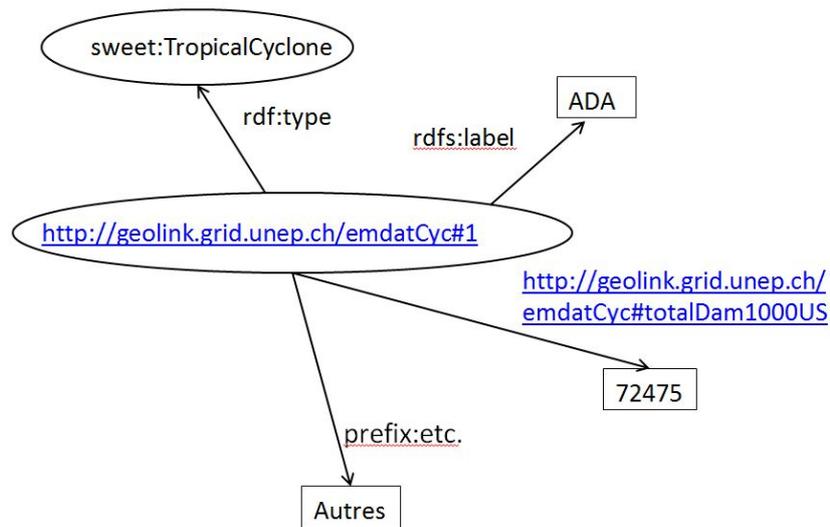


Fig. 13. Modélisation du graphe RDF des cyclones tropicaux de l'EM-DAT. Le préfixe « etc. » et le littéral « Autres... » résumant la présence dans le graphe de plusieurs autres propriétés et valeurs de propriétés correspondant aux différents champs du jeu de données de l'EM-DAT.

3.6.2. Ecriture du graphe RDF des données EM-DAT

Le graphe RDF des données EM-DAT a été produit avec la même méthode que celui du graphe PREVIEW, en utilisant un script R écrivant les triplets à partir du fichier CSV préparé à l'étape 3.4.4.

3.7. Etablissement des liens entre les graphes PREVIEW et EM-DAT

L'établissement de liens entre les données PREVIEW et EM-DAT traduites en RDF a été réalisé à partir des deux graphes produits aux chapitres 3.5 et 3.6, en utilisant l'outil Silk Workbench (<https://github.com/silk-framework/silk/tree/master/doc>), qui permet de définir à l'aide d'une interface graphique des règles de lien pour l'interconnexion des deux jeux de données. Les différentes étapes de l'établissement de liens sont les suivantes :

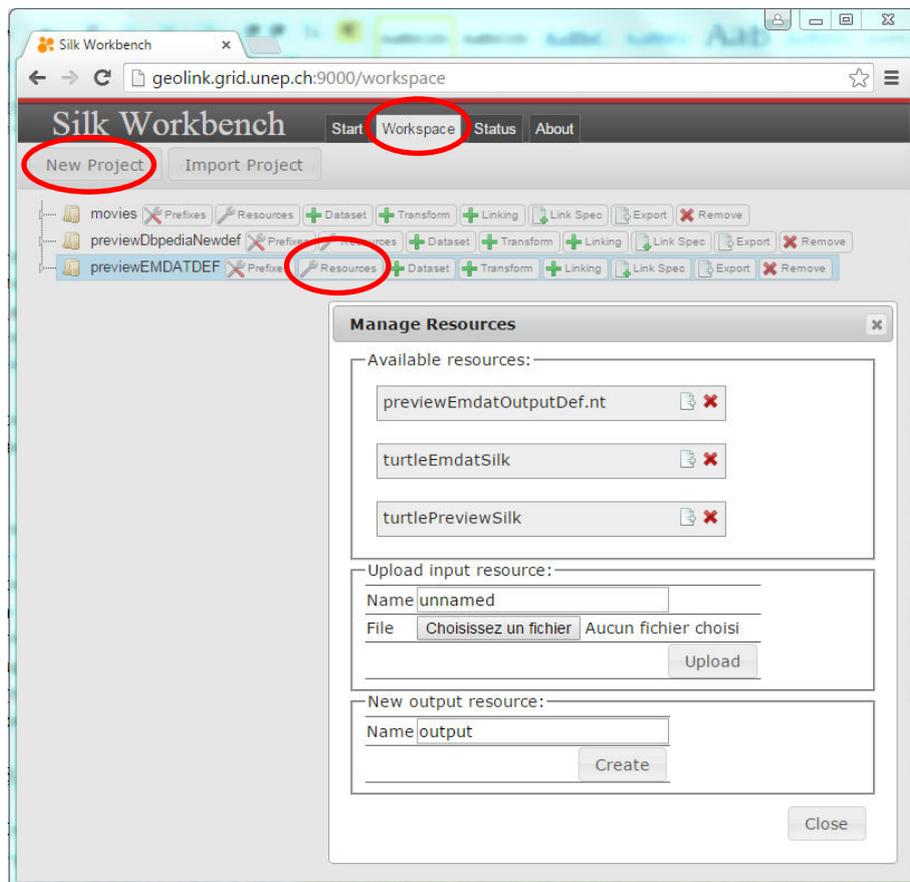
1) Création de deux graphes PREVIEW et EM-DAT simplifiés à partir de nouveaux scripts R dérivés de ceux utilisés en 3.5.3 et 3.6.2. Ces graphes ne comportent que les identifiants de ressources ainsi qu'une nouvelle propriété nommée « silk » et dont la valeur est une concaténation des champs label, year, StartMonth et ISO3. Cette approche, bien que moins précise que celle développée par Peduzzi et al. (2005) et ne permettant pas d'établir des liens entre tous les événements du PREVIEW et ceux de l'EM-DAT, a été retenue car elle suffisait à récupérer tous les liens entre événements dont les noms ont été uniformisés aux chapitres 3.4.3 et 3.4.4. Il aurait fallu dans l'idéal d'abord réaliser une étape de correspondance entre tous les événements à partir de leurs dates, notamment pour les événements non nommés. L'examen des données a toutefois révélé plusieurs problèmes nécessitant des corrections au cas par cas, ce qui nécessiterait un investissement en temps de travail largement supérieur à celui devant être consacré à ce stage. L'établissement de correspondances à partir du nom, de l'année, du mois et du pays touché a en outre permis de récupérer une quantité de liens largement suffisante pour tester l'application du concept de Linked Data aux données du PREVIEW et de l'EM-DAT.

2) Création d'un graphe simplifié à partir d'une liste des cyclones tropicaux décrits sur DBpedia. Les données ont été récupérées sur le endpoint de dbpedia (<http://dbpedia.org/sparql>) à l'aide de la requête SPARQL suivante :

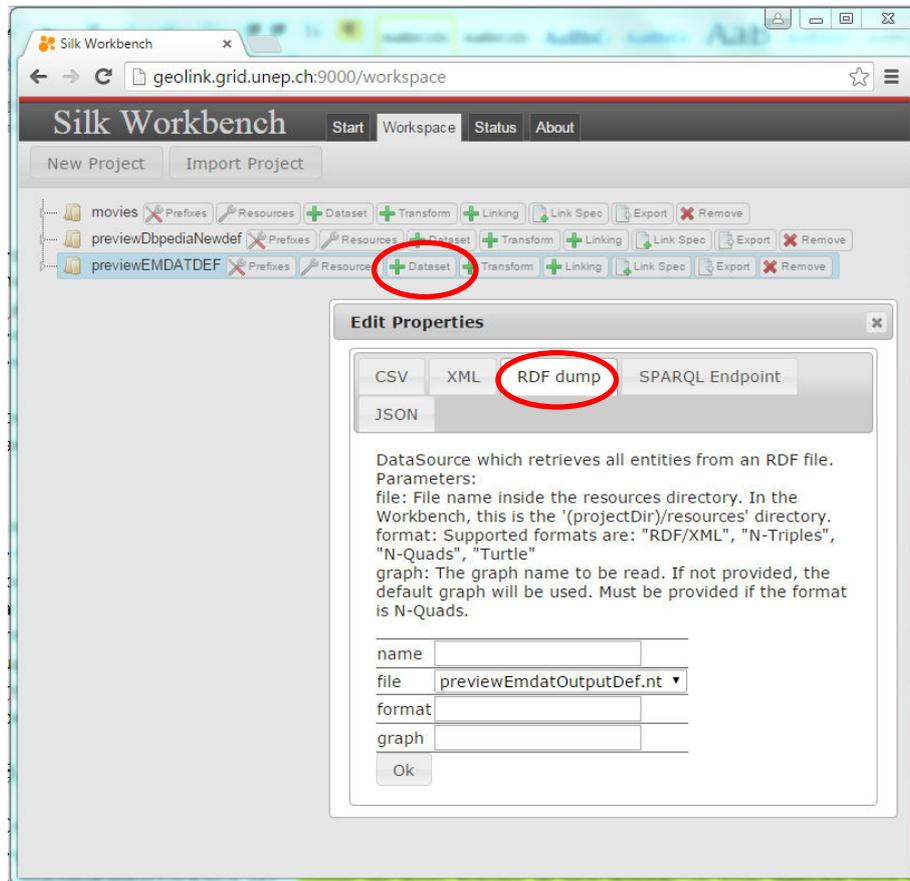
```
prefix dbo: <http://dbpedia.org/ontology/>
prefix dbprop: <http://dbpedia.org/property/>
select * where {
?event rdf:type <http://dbpedia.org/class/yago/Cyclone111443721>;
  rdfs:label ?label;
  dbp:year 2005;
  dbprop:formed ?debut;
  dbprop:dissipated ?fin;
  FILTER ( lang(?label) = "en" )
}
```

L'objectif était de lier les données du graphe PREVIEW aux ressources correspondantes sur DBpedia, afin de pouvoir accéder à des nombreuses informations supplémentaires sur les cyclones. Les noms des évènements récupérés sur ont également fait l'objet d'une uniformisation.

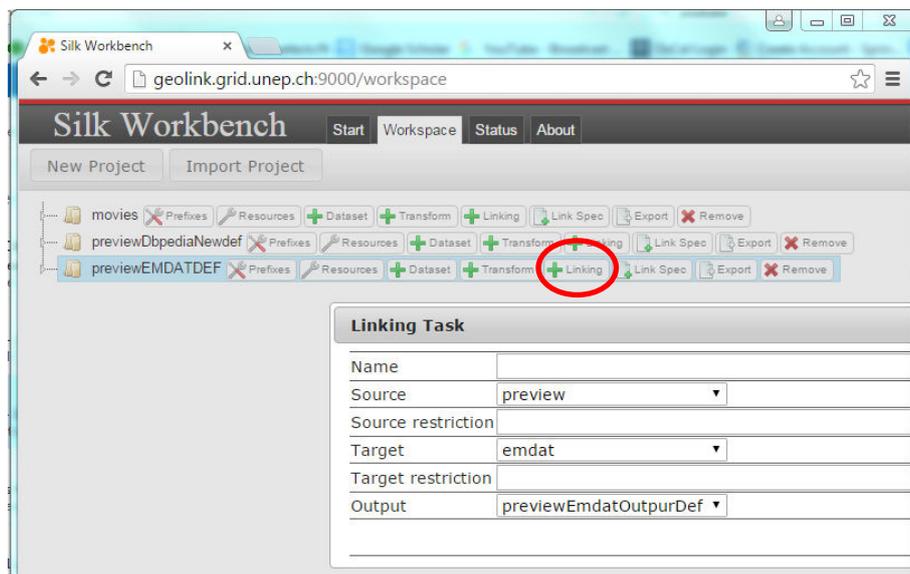
3) Création d'un nouveau projet (ici previewEMDATDEF) dans le workspace de l'application Silk workbench puis chargement des graphes préparés à l'étape 1 en cliquant sur l'onglet Resources. Pendant cette étape, il faut également créer une ressource d'output en format N-Triples (extension .nt) :



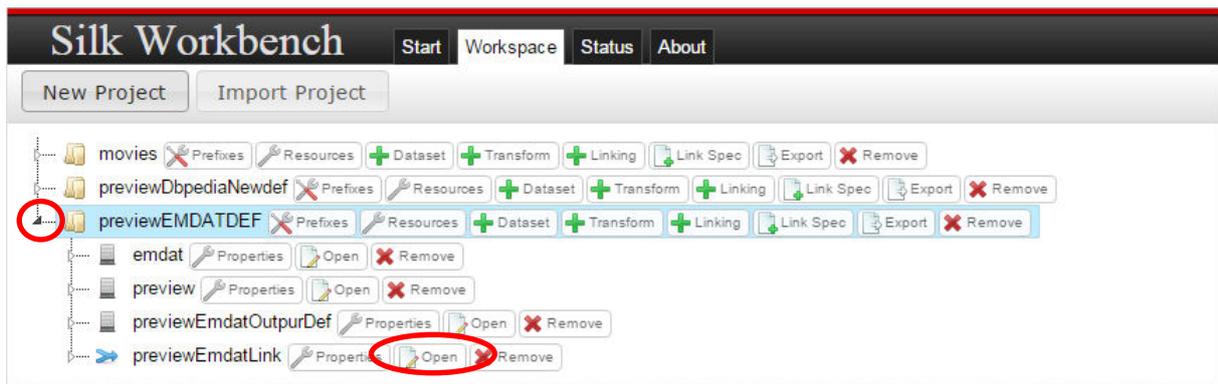
4) A partir des ressources précédemment créées, définir les jeux de données à utiliser en cliquant sur l'onglet Dataset puis en sélectionnant les graphes via l'onglet RDF dump. Tous les graphes chargés dans le cadre de ce travail sont au format Turtle, l'output créé à l'étape précédente étant pour sa part en format N-Triples :



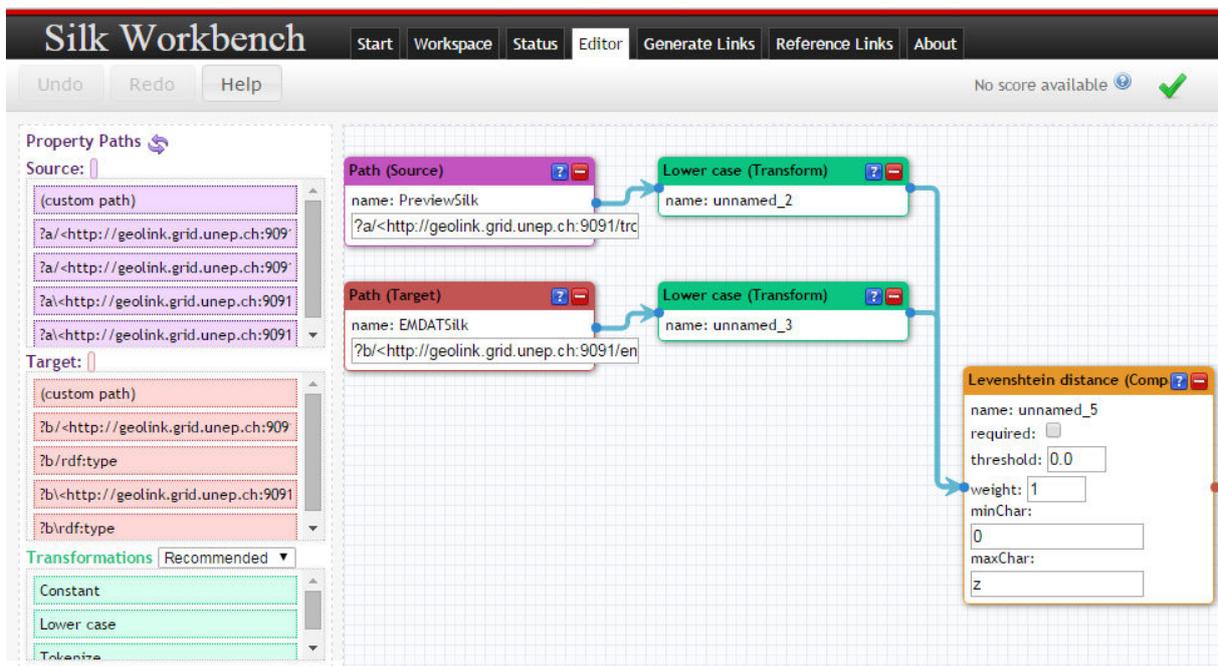
5) Créer et nommer une nouvelle règle de liens (Linking Task), avec comme fichier source le graphe PREVIEW et comme cible le graphe EM-DAT. Il faut également préciser le chemin vers le fichier d'output précédemment créé :



6) Ouvrir le Linkage Rule Editor et définir la procédure de lien :



La propriété #silk, obtenue à partir du champ concaténé créé à l'étape 1 dans les deux fichiers PREVIEW et EM-DAT, est alors utilisée pour lier les deux graphes. Les champs de chaque fichier font d'abord l'objet d'une uniformisation en passant toutes les lettres en caractères minuscules (Lower Case). La comparaison entre les champs des deux jeux de données se fait avec la distance de Levenshtein, qui est une distance mathématique donnant une mesure de la similarité entre deux chaînes de caractères. Elle est égale au nombre minimal de caractères qu'il faut supprimer, insérer ou remplacer pour passer d'une chaîne à l'autre. Les fichiers PREVIEW, EM-DAT et DBpedia ayant été déjà préparés en uniformisant les labels des cyclones, seuls des liens à 100% peuvent être obtenus dans ce cas :



7) Cliquer sur l'onglet Generate Links puis sur le bouton Start. Les liens sont créés et on peut les visualiser en détail en cliquant sur les flèches à gauche de chacun d'entre eux. Compte tenu de la méthode employée, seuls des liens à 100% sont effectivement créés. Il faut ensuite cocher les cases correspondant à l'état du lien constaté (correct, indéterminable, incorrect) :

The screenshot shows the Silk Workbench interface. At the top, there is a navigation bar with tabs: Start, Workspace, Status, Editor, **Generate Links** (circled in red), Reference Links, and About. Below the navigation bar, there is a control bar with a **Start** button (circled in red), a Stop button, and a green progress bar that says "Generate Links finished in 1s".

The main area displays a list of generated links. The first link is selected and expanded. The table below shows the source and target links, along with their scores and correctness status.

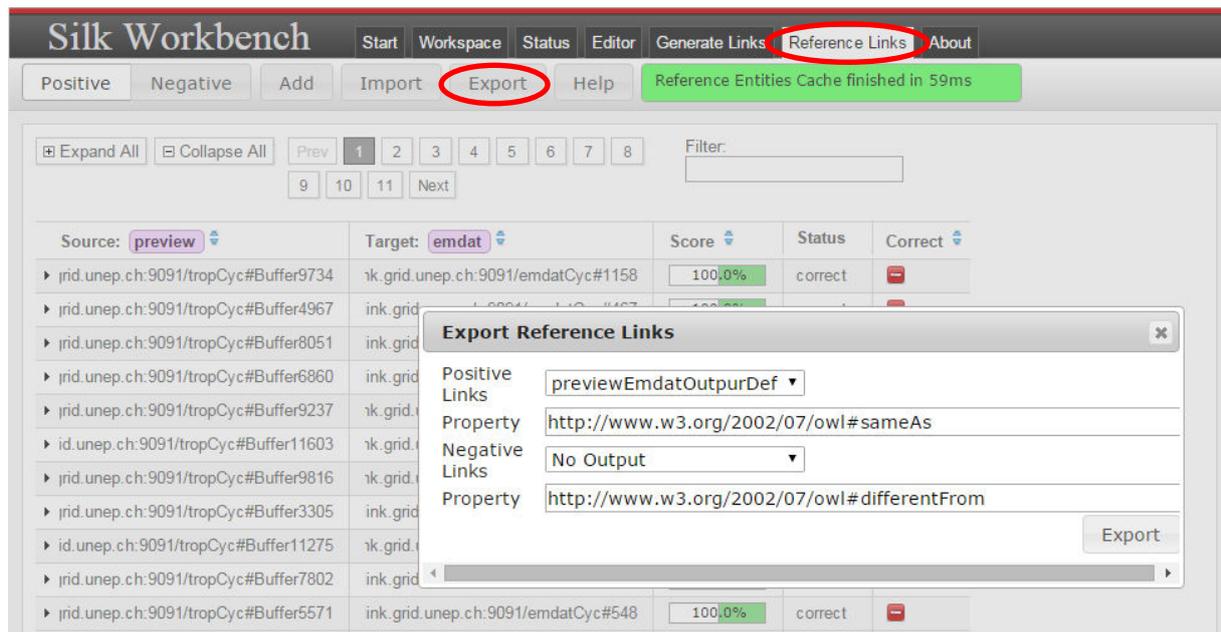
Source	Target	Score	Correct
grid.unep.ch:9091/tropCyc#Buffer5779	link.grid.unep.ch:9091/emdatCyc#565	100.0%	✓ ? ✕
grid.unep.ch:9091/tropCyc#Buffer5777	link.grid.unep.ch:9091/emdatCyc#563	100.0%	✓ ? ✕
grid.unep.ch:9091/tropCyc#Buffer5776	link.grid.unep.ch:9091/emdatCyc#563	100.0%	✓ ? ✕
grid.unep.ch:9091/tropCyc#Buffer5775	link.grid.unep.ch:9091/emdatCyc#563	100.0%	✓ ? ✕
grid.unep.ch:9091/tropCyc#Buffer5774	link.grid.unep.ch:9091/emdatCyc#563	100.0%	✓ ? ✕
grid.unep.ch:9091/tropCyc#Buffer5773	link.grid.unep.ch:9091/emdatCyc#563	100.0%	✓ ? ✕
grid.unep.ch:9091/tropCyc#Buffer5772	link.grid.unep.ch:9091/emdatCyc#563	100.0%	✓ ? ✕
grid.unep.ch:9091/tropCyc#Buffer5770	link.grid.unep.ch:9091/emdatCyc#562	100.0%	✓ ? ✕

The detailed view of the first link shows the following information:

- Comparison: levenshteinDistance (unnamed_5) 100.0%
- Transform: lowerCase (unnamed_2) angela199210vnm
- Input: ?a/<http://geolink.grid.unep.ch:9091/tropCyc#silk> (PreviewSilk) ANGELA199210VNM
- Transform: lowerCase (unnamed_3) angela199210vnm
- Input: ?b/<http://geolink.grid.unep.ch:9091/emdatCyc#silk> (EMDATSilk) Angela199210VNM

Silk Workbench permet bien entendu de définir des règles qui créeront des liens imparfaits entre deux chaînes de caractères, permettant d'explorer de nombreuses possibilités de liens entre les ressources des deux jeux de données. Les liens n'atteignant pas 100% de correspondance peuvent alors être inspectés et retenus s'ils sont jugés corrects. Une démarche plus exploratrice de ce genre a d'ailleurs été testée lors d'un premier essai. Toutefois, la taille des fichiers comparés étant importante, la démarche de création de liens à 100% entre champs pré-uniformisés a été retenue. Une vérification de l'ensemble de liens imprécis créés à partir de concaténation de champs ne correspondant pas parfaitement (par exemple en comparant les dates des événements, ce que permet Silk Workbench) aurait permis de récupérer plus d'informations mais aurait été extrêmement chronophage. A nouveau, une correspondance parfaite entre les données PREVIEW et EM-DAT passerait par une vérification détaillée de tous les événements, mais encore plus idéalement par la création d'un champ commun aux deux bases de données, par exemple en utilisant le numéro de série de l'Ibtracs. Ceci nécessiterait un travail considérable dépassant largement le cadre de ce stage, d'autant plus qu'au cours de l'aménagement des données, il a été constaté que des événements enregistrés dans la base de l'EM-DAT n'existent pas dans le jeu du PREVIEW et vice-versa.

8) Enfin, cliquer sur l'onglet Reference Links puis sur le bouton Export afin de référencer les liens et de les exporter dans le fichier output créé à l'étape 3. Dans la fenêtre de dialogue d'export, il faut alors définir le format d'output de la façon suivante :



Tous les liens positifs entre ressources des deux graphes sont alors créés, en écrivant dans le fichier d'output des triplets liant les ressources correspondantes à l'aide de la propriété owl:sameAs, qui indique que deux ressources décrivent la même chose. Par exemple, la première ressource (buffer) du jeu de données PREVIEW décrit le même évènement que la première ressource de l'EM-DAT, soit le cyclone baptisé ADA en 1970, le triplet s'écrivant de la façon suivante :

```
<http://geolink.grid.unep.ch/tropCyc#Buffer1>          <http://www.w3.org/2002/07/owl#sameAs>
<http://geolink.grid.unep.ch/emdatCyc#1>
```

3.8. Préparation d'un graphe PREVIEW lié complet

Les fichiers d'output obtenus au chapitre 3.7 contiennent respectivement 1801 triplets de liens entre PREVIEW et EM-DAT et 2262 triplets de liens entre PREVIEW et DBpedia. L'ensemble de ces triplets ont été ajoutés au graphe PREVIEW préparé au chapitre 3.5., afin de produire un graphe complet lié aux données de l'EM-DAT et aux données DBpedia.

Le fichier complet a ensuite été chargé sur la plateforme Datalift, laquelle réécrit automatiquement le graphe sous une forme plus propre associant dans l'ordre tous les triplets correspondant à chaque ressource. Ce graphe mis au net a ensuite été exporté au format Turtle (voir 3.5.3., étape 13).

3.9. Chargement des graphes PREVIEW et EM-DAT sur Parliament

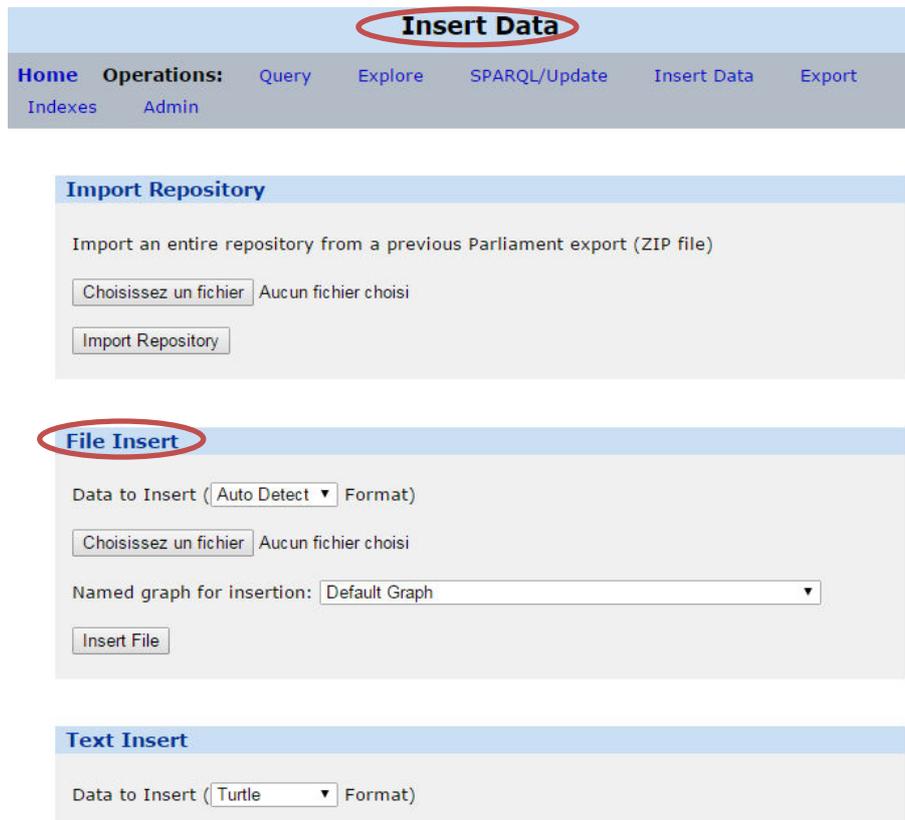
Les graphes PREVIEW et EM-DAT ont ensuite été chargés sur le serveur Parliament, accessible à l'URL <http://geolink.grid.unep.ch:9090/parliament/>, qui offre la possibilité d'interroger les données via un endpoint SPARQL supportant le standard GEOSPARQL, ce que ne permet pas Datalift (voir 3.3.).

1) Dans un premier temps, il est nécessaire de créer des index spatiaux en cliquant sur l'onglet Indexes⁵ puis sur Create All. Il est indispensable de créer ces index pour que les requêtes spatiales avec GEOSPARQL soient ensuite possibles :

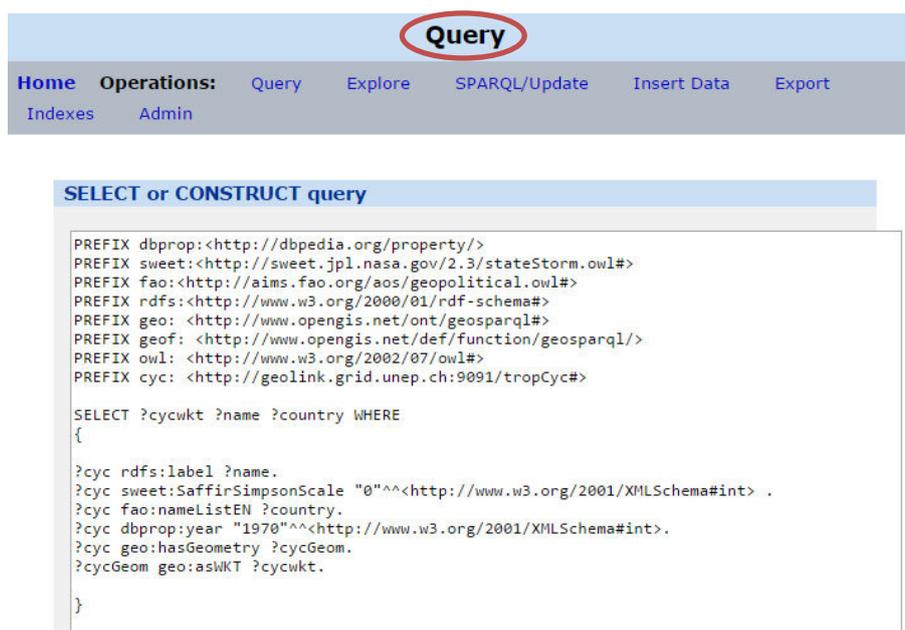


⁵ Pour des raisons indéterminées, la création ou la suppression des index spatiaux ne fonctionne pas avec Firefox mais parfaitement bien avec le navigateur Chrome.

2) Les graphes préparés sont ensuite chargés via l'onglet Insert Data et l'interface File Insert. Parliament offre aussi la possibilité de charger des triplets seuls via la fenêtre Text Insert :



3) Une fois les graphes PREVIEW et EM-DAT chargés, il est possible d'effectuer des requêtes SPARQL sur les données, y compris en utilisant des fonctions GEOSPARQL, via la fenêtre Query. Dans l'exemple ci-dessous, la requête renvoie tous les littéraux WKT correspondant aux buffers du PREVIEW représentant l'extension maximale des cyclones (catégorie de Saffir-Simpson =0) survenus en 1970 :



4. Démonstrateur

4.1. Site web

En complément à l'accès aux données via le endpoint SPARQL du serveur Parlement, une interface de démonstration des différentes possibilités d'exploitation du jeu de données des cyclones du PREVIEW au format RDF a été élaborée sous la forme d'un site web (fig. 14), accessible à l'URL <http://geolink.grid.unep.ch/>. La conception des pages a fait recours aux outils de Bootstrap (<http://getbootstrap.com/>) afin que le site soit consultable depuis tout type d'appareil fixe ou mobile.

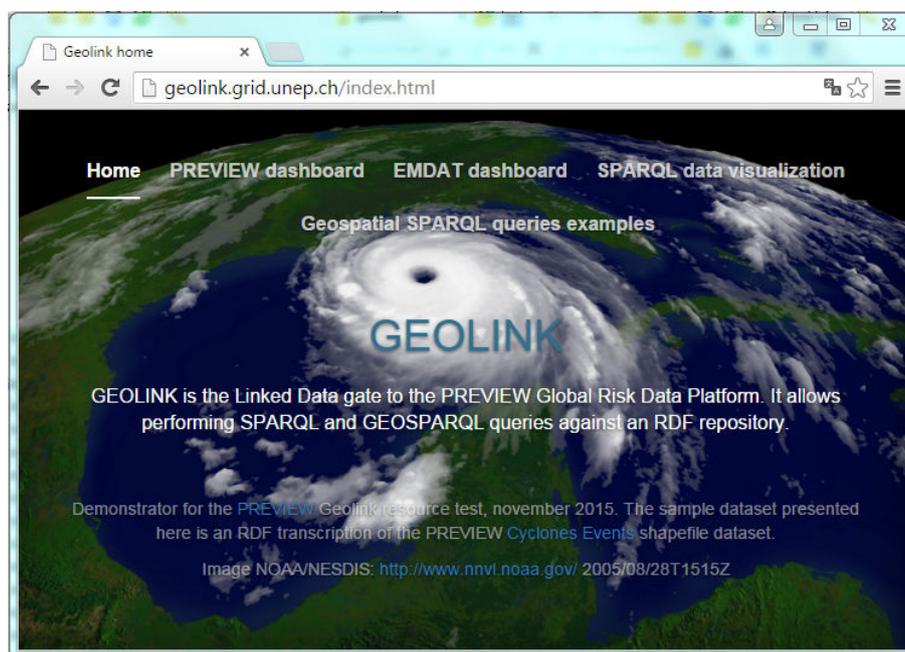


Fig. 14. Page d'accueil du démonstrateur web

Depuis la page d'accueil à l'URL geolink.grid.unep.ch, le site donne accès à deux tableaux de bord donnant un aperçu des données du [PREVIEW](#) et de [l'EM-DAT](#) publiées en RDF.

Une troisième page permet d'interroger et de visualiser les données publiées sur Parlement par l'intermédiaire d'un [formulaire de requête SPARQL](#), permettant de choisir un type de visualisation des données via une liste déroulante. Cette interface permet notamment de cartographier les polygones de cyclones du graphe RDF du PREVIEW, en choisissant la fonction MapWKT dans la liste déroulante Chart type. Enfin, la quatrième page donne accès à trois [exemples de requêtes](#) sur des données spatiales.

4.2. Applications de visualisation du site web

Toutes les applications du démonstrateur ont été réalisées à l'aide de la bibliothèque Javascript Sgvizler (<http://dev.data2000.no/sgvizler/>), qui permet de visualiser des résultats de requêtes SPARQL sous forme de divers graphiques ou tableaux (Skjæveland 2012). Le code de Sgvizler utilise lui-même largement le code du Google visualization API, mais plusieurs applications originales sont également disponibles. C'est le cas notamment de l'outil MapWKT, qui permet de cartographier des objets publiés sous forme de littéraux WKT en utilisant la bibliothèque d'OpenLayers 2 (<http://openlayers.org/api/OpenLayers.js>). Le code complet des différentes applications, contenant les requêtes SPARQL correspondant à chaque graphe ou tableau, est disponible dans les annexes.

La bibliothèque Sgvizler fonctionne très bien dans l'ensemble, seul un changement d'URL pour appeler OpenLayers 2 ayant dû être effectué (<http://openlayers.org/api/OpenLayers.js>). La requête par défaut a également été changée dans le code original de Sgvizler (ligne 1660). La nouvelle requête par défaut est celle qui apparaît dans la fenêtre du [formulaire SPARQL](#).

4.2.1. Tableaux de bord

Les tableaux de bord PREVIEW et EM-DAT font appel à des fonctions de Sgvizler utilisant l'API Google (Google Table, Pie Chart, Google, Column Chart et Area Chart).

Le tableau de bord PREVIEW affiche les résultats de requêtes SPARQL renvoyant, respectivement :

- sous forme de pie chart, les 30 pays les plus touchés par des cyclones tropicaux classés sur l'échelle de Saffir-Simpson et ayant donc atteint le statut d'ouragan ou de typhon, avec des vents d'au moins 118 km/h (fig. 15) ;

- sous forme de tableau, la liste de tous les buffers de cyclones tropicaux du jeu de données PREVIEW, y compris les précurseurs non classés comme ouragans et identifiés dans le jeu de données par une catégorie de Saffir-Simpson dont la valeur est 0. Si l'évènement a touché plus d'un pays, il existe autant de buffers que de pays concernés par le même évènement ;

- sous forme de column chart, le nombre annuel de buffers correspondant aux cyclones ayant touché au moins un pays ;

- sous forme d'area chart, le nombre annuel de buffers, y compris ceux ne touchant aucun pays, correspondant aux cyclones tropicaux majeurs, soit ceux ayant atteint une catégorie de Saffir-Simpson supérieure à 2 et des vents d'au moins 178 km/h.

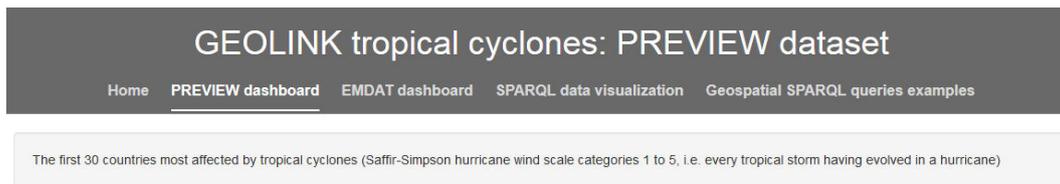


Fig. 15. Exemple de graphe du tableau de bord PREVIEW, affichant les résultats d’une requête SPARQL qui retourne les 30 pays les plus touchés par des cyclones tropicaux classés dans l’échelle de Saffir-Simpson.

Le tableau de bord EM-DAT affiche les résultats de requêtes SPARQL renvoyant, respectivement :

- sous forme de pie chart, les 10 pays comptant le plus de morts liés aux cyclones tropicaux pour l’ensemble du jeu de données, soit entre 1970 et 2011 ;
- sous forme de pie chart, le nombre de morts par région, également pour l’ensemble du jeu de données ;
- sous forme d’area chart, le nombre total de victimes par année (blessés et morts) (fig. 16) ;
- sous forme d’area chart, un indice des victimes, soit le rapport du nombre total de morts au nombre total de personnes affectées ;
- sous forme d’area chart, le montant total par année des dégâts estimés

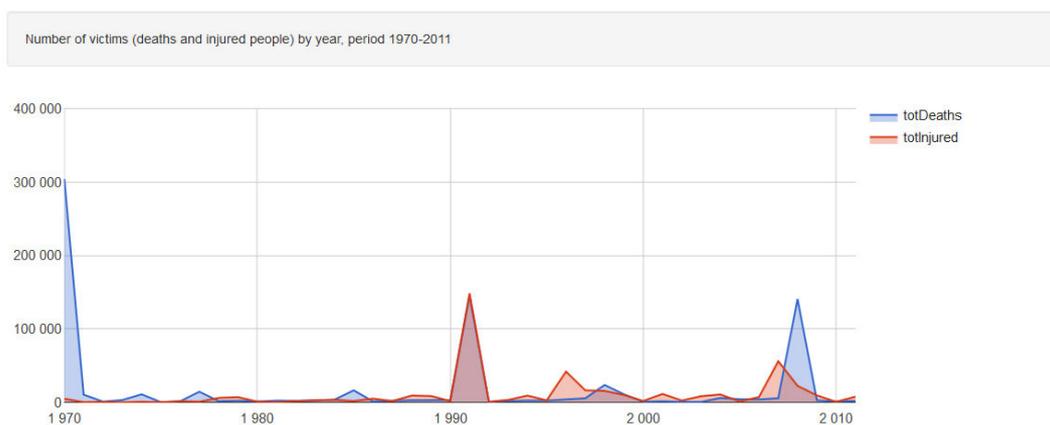


Fig. 16. Exemple de graphe du tableau de bord EM-DAT, affichant les résultats d’une requête SPARQL qui retourne le nombre de blessés et de morts par année (1970-2011).

4.2.2. Formulaire SPARQL

Le formulaire SPARQL (fig. 17) a été réalisé avec la fonction `sgvizler.formDraw()`. Les préfixes nécessaires aux requêtes ainsi que la requête par défaut qui est affichée à l'ouverture de la page sont pré-enregistrés dans le code. L'utilisateur a ensuite la possibilité de modifier la requête ou d'en saisir une nouvelle puis de choisir le type de représentation dans la liste déroulante Chart type. La requête par défaut, qui affiche tous les buffers représentant l'extension maximale des cyclones tropicaux survenus en 1970 (catégorie Saffir-Simpson=0) est conçue pour créer avec la fonction `MapWKT` une carte qui affiche les polygones correspondant (fig. 17).

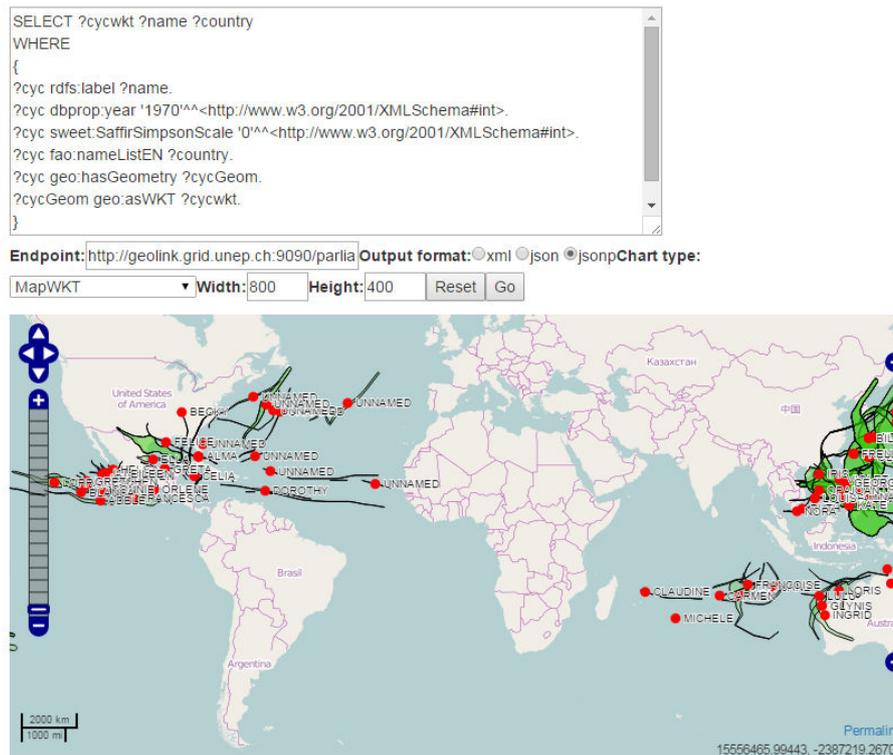


Fig. 17. Le formulaire SPARQL, affichant les résultats de la requête par défaut avec l'application MapWKT.

4.2.3. Exemples de requêtes sur des données spatiales

La page d'exemples affiche les résultats de trois requêtes impliquant des données spatiales. La première requête permet d'exploiter les liens entre les données PREVIEW et celles de l'EM-DAT (fig. 18). Elle renvoie le polygone représentant l'emprise de l'ouragan Katrina de 2005 lorsqu'elle atteignait la catégorie 2 de l'échelle de Saffir-Simpson, soit des vitesses de vents supérieures à 153 km/h. La seconde partie de la requête débutant par l'instruction `OPTIONAL` permet d'afficher, pour tout polygone lié à une ressource du jeu de données de l'EM-DAT, des propriétés supplémentaires non disponibles dans le jeu du PREVIEW. Dans cet exemple, la requête renvoie le nombre de morts ainsi que le montant total des dégâts causés par l'ouragan Katrina. Ces valeurs sont accessibles dans un pop-up activé en cliquant sur le centroïde rouge.

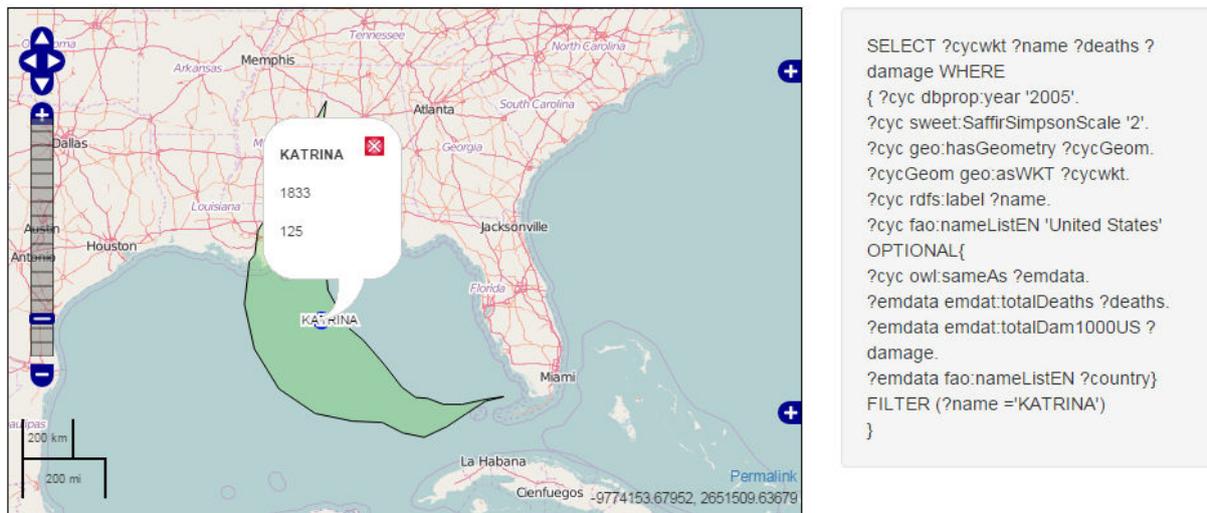


Fig. 18. Requête exploitant les liens entre PREVIEW et EM-DAT en affichant le nombre morts et le montant des dégâts en milliards de dollars causés par Katrina.

La seconde requête met en œuvre la fonction GEOSPARQL `sfWithin`. Elle retourne, à partir d'un jeu de données préalablement récupéré via le endpoint SPARQL de DBpedia puis chargé sur l'entrepôt Parlement, toutes les villes de plus de 10'000 habitants qui ont été touchées par l'ouragan Katrina alors qu'il était classé en catégorie 2 de Saffir-Simpson (fig. 19). La fonction `sfWithin`, introduite à la fin de la requête dans la commande `FILTER`, permet de ne renvoyer que les villes (qui sont des points dont les coordonnées sont écrites en littéraux WKT) se situant à l'intérieur du polygone du cyclone (qui est le même que celui affiché dans le premier exemple). La requête récupère également le site web de chaque ville ainsi que sa population. En l'état, la fonction `MapWKT` de Sgvizler ne permet malheureusement pas d'afficher en même temps les points et le polygone et l'ajout de cette fonctionnalité serait un plus.

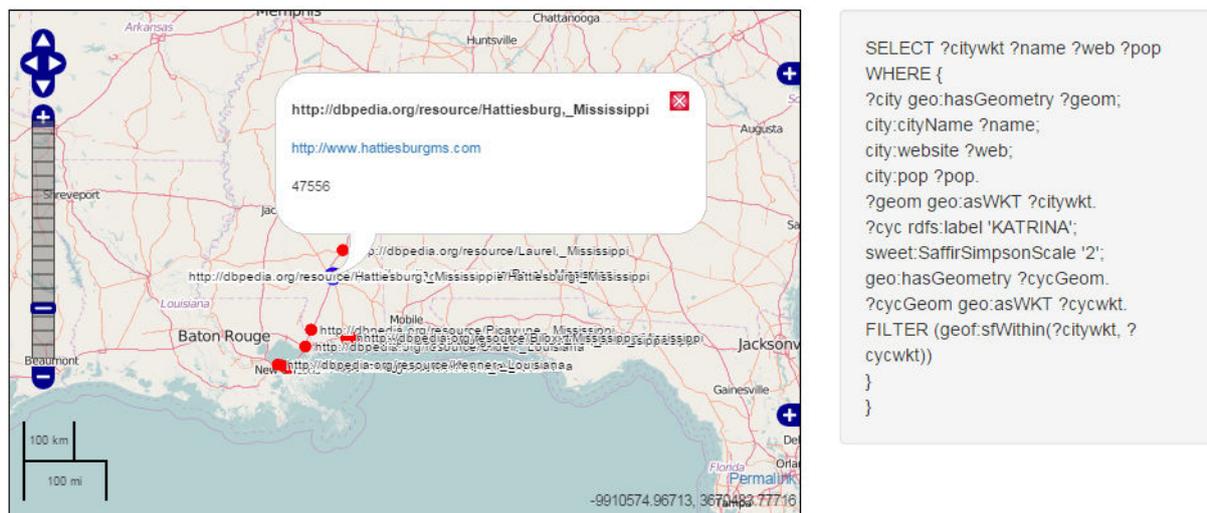


Fig. 19. Résultat de la requête utilisant la fonction `sfWithin` de GEOSPARQL, permettant de récupérer les villes de plus de 10'000 habitants (données DBpedia) impactées par l'ouragan Katrina lorsqu'il atteignait la catégorie 2 de l'échelle de Saffir-Simpson. Les pop-ups affichent le site web de chaque ville concernée ainsi que sa population, données également issues de DBpedia.

La troisième requête présente une autre possibilité très intéressante de l'utilisation du Linked Data. Il s'agit d'une requête fédérée entre le jeu de données du PREVIEW et DBpedia (fig. 20). Le jeu de

données PREVIEW étant lié aux ressources de cyclones tropicaux de DBpedia (voir 3.4), il est en effet possible d'accéder directement à d'autres informations sur DBpedia. Grâce à la clause SERVICE, qui permet d'appeler des ressources disponibles via des URIs externes à celles publiées sur le triplestore Parliament, et à l'utilisation du prédicat owl:sameAs, la requête permet de récupérer l'abstract décrivant le cyclone sur DBpedia ainsi que l'URL correspondante sur Wikipedia. Bien d'autres données liées, des photos du cyclone par exemple, pourraient être récupérées avec une requête similaire. A nouveau, l'interface cartographique offerte par la fonction MapWKT nécessiterait quelques aménagements, notamment pour pouvoir contrôler le format de l'abstract et afficher l'URL de wikipedia plus distinctement et en tant qu'hyperlien.

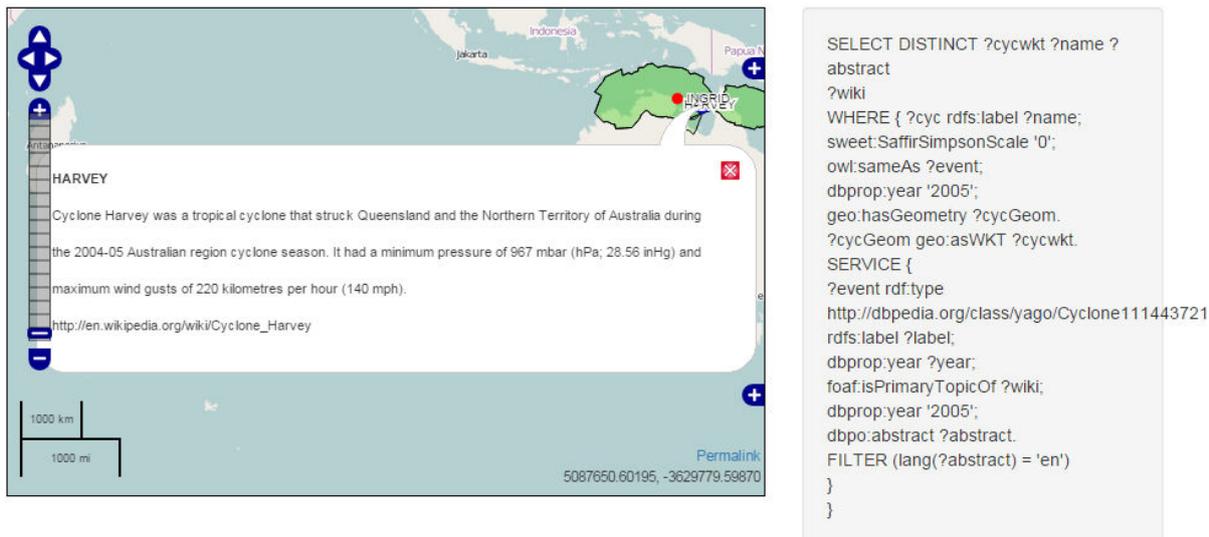


Fig. 20. Requête fédérée PREVIEW-DBpedia. La requête utilisant la clause SERVICE permet d'accéder sur DBpedia à des informations supplémentaires concernant les cyclones liés des deux jeux de données. Dans cet exemple, la carte permet d'afficher l'abstract DBpedia et les URL de wikipedia correspondant aux événements auxquels les polygones du PREVIEW sont associés.

4.4. Interrogation des données avec le package R SPARQL

Il est également possible d'interroger le graphe de cyclones publié dans le triplestore Parliament en utilisant R et le package SPARQL (<https://cran.r-project.org/web/packages/SPARQL/SPARQL.pdf>).

L'exemple de la fig. 21 présente les résultats d'une requête effectuée sur les graphes du PREVIEW et de l'EM-DAT, qui renvoie le nombre de morts par pays touché par des cyclones en 2005. Cet exemple d'accès aux données permet aussi d'illustrer une limite de l'utilisation d'entités spatiales pour obtenir une description exhaustive des événements passés. En effet, si on exécute la même requête uniquement sur les données de l'EM-DAT, on constate un plus grand nombre de pays touchés ainsi qu'un pic important de victimes au Guatemala, totalement invisible sur le graphique obtenu en croisant données du PREVIEW et de l'EM-DAT (fig. 21a). Un examen détaillé des données démontre notamment que le cyclone Stan, qui a touché la mer des Caraïbes et l'Amérique centrale en 2005, n'est en effet associé à des buffers ne touchant que le seul Mexique alors que les statistiques de l'EM-DAT révèlent que plusieurs autres pays ont subi des pertes humaines et matérielles (Costa Rica, Salvador, Guatemala, Haïti, Honduras et Nicaragua) (fig. 21b). Si le Mexique est effectivement le pays où le plus de personnes ont été affectées par Stan, le Guatemala déplore 1513 morts liés à cet événement contre 36 au Mexique. Le cyclone Stan n'a en effet jamais dépassé la catégorie 1 de l'échelle de Saffir-Simpson, mais de fortes pluies lui étant associées ont été à l'origine de nombreux glissements de terrain meurtriers au Guatemala, dont le territoire n'est effectivement pas touché par les buffers modélisés.

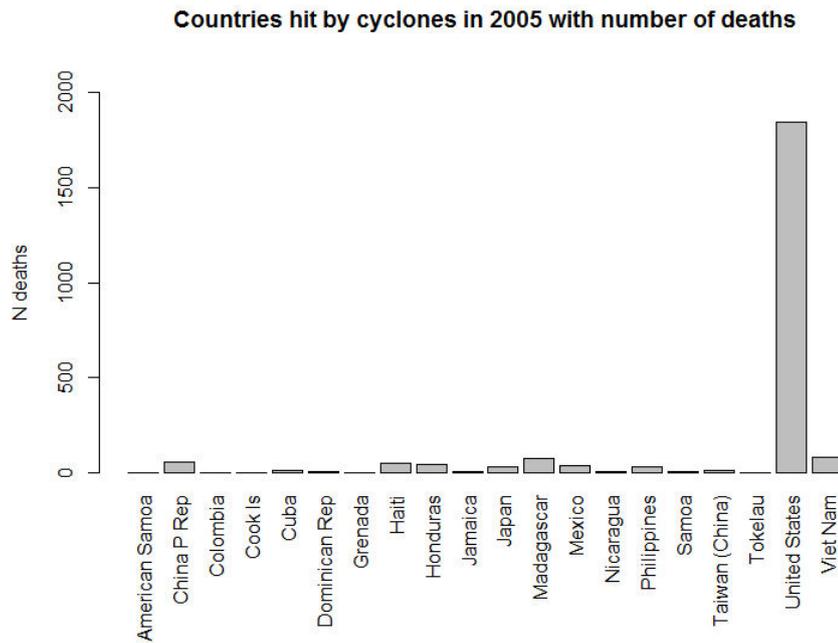


Fig. 21a. Résultats d’une requête effectuée sur les graphes PREVIEW et EM-DAT avec R et le package SPARQL.

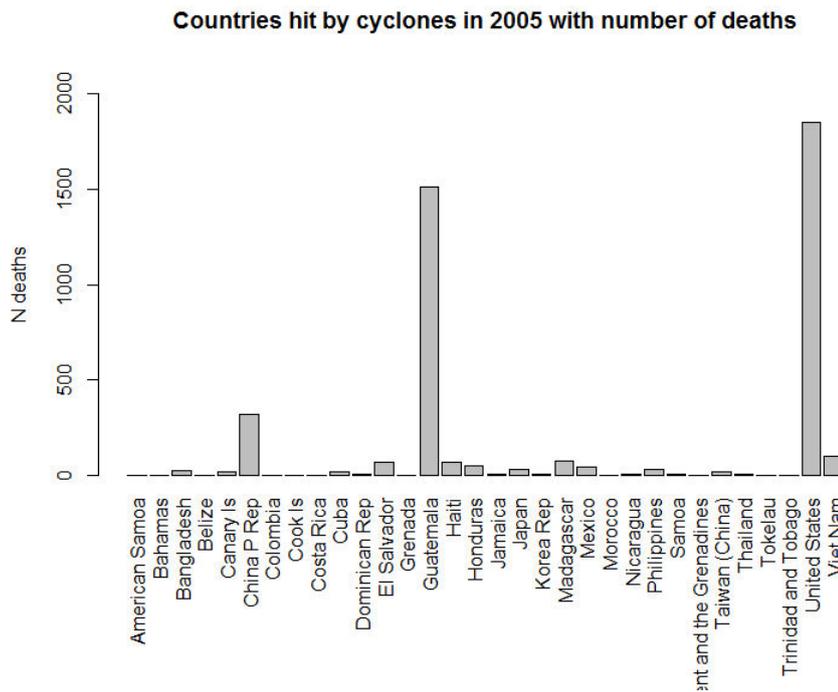


Fig. 21b. Résultats d’une requête similaire à celle de la figure 21a, effectuée avec la même méthode mais uniquement sur le graphe EM-DAT. Outre un nombre de pays plus élevés, on remarque l’important pic de victimes au Guatemala, absent sur le graphique de données liées PREVIEW-EMDAT.

5. Bilan et perspectives

5.1 Avantages et potentiel du Linked Data

A l'issue de ce travail, le bilan général est largement positif quant au potentiel d'application du concept de Linked Data à la publication et l'exploitation de données géospatiales. Après une mise en œuvre nécessitant un certain temps pour la prise en main des standards et outils décrits dans les chapitres précédents, les bénéfices et le potentiel de développement sont rapidement perceptibles.

L'effort de publication sous un nouveau format et d'apprentissage d'interrogation de ces données peut être justifié tout d'abord par le gain d'interopérabilité, chaque ressource possédant un identifiant unique à l'échelle du Web sous forme d'URI, tandis que les propriétés permettant de caractériser ces ressources sont elles-mêmes des URIs, provenant soit de vocabulaires déjà existants et publiés, soit de vocabulaires spécifiques à un jeu de données particulier pouvant faire l'objet d'une nouvelle ontologie. L'utilisation de standards, notamment le standard OGC GeoSPARQL, permet non seulement de partager et d'interroger des données spatiales interopérables sur le Web mais également d'utiliser ces données spatiales pour en interroger d'autres, comme le montre l'exemple d'utilisation de la fonction GeoSPARQL `sfWithin` présenté au chapitre 4.2.3.

Le plus grand bénéfice est bien entendu lié à l'essence même du Linked Data, soit la possibilité d'interconnecter plusieurs jeux de données spatiaux et statistiques, permettant d'accéder à de nombreuses informations supplémentaires. L'établissement de liens entre jeux de données, même en attribuant par exemple simplement un champ commun à deux documents CSV distincts et non liés, est un enjeu crucial ainsi que le démontre le cas illustré au chapitre 4.4. Dans cet exemple, les pertes humaines et matérielles qu'a subies le Guatemala lors du cyclone Stan en 2005 demeurent invisibles en liant l'emprise spatiale du cyclone sous forme de buffers aux statistiques des pays qu'il intersecte. Stan n'a en effet pas dépassé la catégorie 1 sur l'échelle de Saffir-Simpson et les buffers équivalents à son emprise ne touchent pas le Guatemala, le nombre important de victimes dans ce pays étant dû à d'importants glissements de terrain causés par des pluies abondantes en périphérie du cyclone. Un modèle de données permettant de lier les deux sources de données par un identifiant unique permettrait d'associer la ressource Stan du PREVIEW à toutes les statistiques de l'EM-DAT le concernant. Traduire ces deux sources d'information au format Linked Data serait un excellent moyen de résoudre le problème, en adoptant un modèle plus élaboré pour les données du PREVIEW (où chaque événement serait identifié par une URI, cette URI étant liée à son homologue de l'EM-DAT et possédant un ou plusieurs buffers).

Le modèle de graphe du RDF est aussi un avantage, puisqu'il permet de rapides mises à jour ponctuelles d'un jeu de données. Il est ainsi facile d'ajouter individuellement des triplets correspondant à de nouvelles ressources ou à de nouvelles propriétés pour des ressources existantes.

Les possibilités d'interrogation, de visualisation et d'exploitation des données sont en outre variées et permettent une grande souplesse d'utilisation. Un même échantillon d'un jeu de données spatiales décrit dans un graphe RDF sous forme de littéraux WKT peut être visualisé à l'aide d'applications de web mapping, permettre de réaliser une requête spatiale avec une fonction GeoSPARQL, ou encore être téléchargé au format CSV et réutilisé dans un SIG.

Rappelons qu'il est également possible d'exploiter des données raster au format RDF (voir 2.4.4), cet aspect n'ayant toutefois pas été abordé dans le cadre de ce travail.

L'utilisation du langage SPARQL est également très efficace pour accéder à un sous-ensemble précis d'un jeu de données et, nonobstant le nécessaire investissement en temps d'apprentissage, peut s'avérer plus simple et plus efficace que l'écriture d'une requête URL très détaillée à un service Web WFS, par exemple.

5.2 Principaux inconvénients

L'un des principaux inconvénients du Linked Data rencontré lors de la réalisation de ce travail est la grande instabilité des ressources Web associées à ce concept relativement récent. Ce problème a notamment été rencontré de façon récurrente lors de la recherche d'outils destinés à la traduction de données au format RDF ou à la visualisation, certains d'entre eux étant développés lors de projets d'étudiants puis disparaissant après quelques années. Si plusieurs plateformes comme Datalift ou Geoknow proposent des solutions intégrées permettant d'aborder les différentes étapes de préparation, de publication et d'exploitation de Linked Data, ces dernières sont toujours en cours de développement et certains outils restent difficiles à prendre en main. Par ailleurs, si de nombreuses solutions pour produire des graphes RDF à partir de données de différents formats sont assez facilement accessibles sur le Web, elles demeurent souvent au stade de développement et difficiles à prendre en main pour un nouvel utilisateur.

Dans le même ordre d'idées, le recours de certaines applications à des fonctionnalités externes situées sur des serveurs confrontés à des problèmes peut paralyser tout un ensemble d'outils intéressants (voir l'exemple d'Open Refine au chapitre 3.2.3).

Déjà évoqué en 5.1., le temps d'apprentissage et de prise en main des outils sera plus ou moins long en fonction du bagage informatique d'un nouvel utilisateur. L'autre problème éventuel lié aux utilisateurs potentiels de données géospatiales au format Linked Data est simplement leur quantité et leur accessibilité. Malgré l'essor évident du Linked Open Data depuis quelques années, quel est l'effectif réel d'utilisateurs et donc la demande de ce type d'accès aux données ? Réciproquement, la quantité et la nature de données existantes et susceptibles de motiver l'apparition de nouveaux utilisateurs sont également importantes. Le Linked Data étant destiné à faciliter la lecture des données par des machines, il est souvent difficile pour un utilisateur humain d'aborder la structure d'un jeu de données, l'immense quantité de données accessibles sur DBpedia étant par exemple peu aisée à prendre en main. Il serait donc utile de fournir un minimum d'informations « humaines » sur les ressources et les propriétés existantes dans un triplestore afin de faciliter leur prise en main par de nouveaux utilisateurs. Certains projets, comme LinkedGeoData, donnent des exemples de requêtes sur leurs données, qui apportent un excellent moyen d'apprendre à exploiter un jeu de données.

La maintenance et l'évolution des technologies mais aussi des standards est aussi essentielle. La multiplication de vocabulaires et d'ontologies soulève aussi de nombreuses questions. Quels vocabulaires utiliser pour produire un jeu de données et faciliter son interconnexion avec d'autres sources ? Comment exploiter un triplestore intéressant mais utilisant des prédicats ou des typages de données complètement différents que ceux de son propre jeu de données ?

Des questions techniques concernant la traduction d'autres formats de sources de données se posent également. Par exemple, comment mettre en place une automatisation de mises à jour d'un graphe RDF construit à partir de sources de données accessibles sous un autre format et structurées différemment ?

5.3 Conclusion et perspectives de développement

Les résultats de ce travail de stage tendent à démontrer le potentiel certain de l'utilisation du concept de Linked Data et du Web sémantique pour la publication et l'exploitation de données spatiales environnementales, malgré un certain nombre de limitations liées non seulement à l'hétérogénéité, au fonctionnement ou à la disponibilité des outils nécessaires mais aussi à une prise en main pouvant être ardue pour de nouveaux utilisateurs. Bien que moins conviviale et plus difficile d'accès qu'une interface de recherche facettée sous forme de listes de cases à cocher ou de listes déroulantes, une interface de requête directe en langage SPARQL, moyennant un temps d'apprentissage qui peut être court pour des requêtes simples, permet une exploration à la fois souple et précise des données. Le gain réside aussi dans l'interopérabilité des données, qui peuvent être rapidement réutilisées et liées à d'autres jeux de données. Le temps investi dans la prise en main du langage SPARQL est donc certainement rapidement compensé par les possibilités d'intégration de données offertes.

En guise de conclusion, les développements suivants, envisageables à plus ou moins court terme, sont proposés dans l'optique de la préparation d'une plateforme Linked Data pleinement opérationnelle pour le PREVIEW :

-Repenser la modélisation des données à publier sous forme de Linked Data. Dans le cadre de ce travail exploratoire, on s'est contenté de traduire le shapefile Cyclone Events sous format RDF, dans une version simplifiée. L'une des solutions serait de modéliser les données en utilisant le RDF Data Cube Vocabulary dans l'optique de mieux structurer l'association de nombreux champs statistiques aux entités spatiales et de faciliter les regroupements de données (voir 2.3.2) ;

-Publier une ontologie environnementale complète permettant de décrire les différentes catastrophes du PREVIEW. Il serait par exemple utile de créer un prédicat SaffirSimpsonMax permettant de classer directement les événements d'après leur plus forte catégorie atteinte. L'idéal serait d'écrire une ontologie calquée sur la terminologie et la classification de l'EM-DAT, afin de proposer un vocabulaire et une structure identiques favorisant la standardisation et facilitant la publication ultérieure de Linked Data. Ce travail devrait être fait idéalement en collaboration avec l'EM-DAT. Dans le même ordre d'idées, il serait souhaitable de développer une ontologie liée aux risques s'appuyant également sur une terminologie commune aux deux bases de données ;

-Développer un outil simple mais « User friendly » pour faciliter l'interconnexion des données ;

-Tester la traduction et l'exploitation de données raster du PREVIEW au format Linked Data ;

-Développer des applications plus complètes à partir du code de Sgvizler, en réécrivant notamment la fonction MapWKT en OpenLayers3. Cette fonction devrait également permettre de superposer deux types de géométries, soit des points et des polygones. Le rendu de requêtes spatiales utilisant les fonctions de GEOSPARQL serait ainsi plus parlant. La représentation des données sous forme de pop-ups s'ouvrant à partir d'un centroïde des polygones n'est pas forcément idéale et nécessiterait aussi quelques aménagements ;

-Mettre à disposition un ensemble plus grand de données du PREVIEW et de l'EDE sous forme de Linked Data afin de quantifier leur utilisation et d'évaluer la demande pour ce type d'accès aux données.

Remerciements

Je tiens à exprimer ici tous mes remerciements à Gregory Giuliani, superviseur de ce stage à l'ISE/GRID-Geneva, pour m'avoir confié un sujet exigeant mais passionnant, pour sa disponibilité, pour son aide plus que précieuse concernant les arcanes tortueux du Linux ainsi que pour ses très précieux encouragements. Toute ma gratitude va également à Claudine Métral et Gilles Falquet du Centre Universitaire d'Informatique (CUI) de l'Université de Genève, qui ont bien voulu encadrer et évaluer ce travail, pour leur intérêt et leur soutien lors de ma confrontation aux subtilités du Linked Data et du langage SPARQL. Je tiens aussi à remercier chaleureusement Pascal Peduzzi et tous les Griddies pour leur accueil et leurs coups de main. Un très grand merci à mes collègues du bureau A-106 Bruno Chatenoux et Jean-Philippe Richard qui ont répondu à nombre de mes questions et requêtes, même lorsqu'ils étaient confrontés à d'autres priorités d'histoires cartographiques. Un immense merci également à tous les enseignants du certificat de géomatique ainsi qu'aux membres du laboratoire enviroSPACE à l'Institut des Sciences de l'Environnement de l'Université de Genève pour leur accueil dans les locaux neufs et plein de surprises technologiques du boulevard Carl-Vogt. Et enfin, merci aux collègues du certificat de géomatique 2015 pour tous les bons géo-moments partagés à Battelle, Carl-Vogt, Châtelaine et ailleurs.

Références

- Athanasidou S., Bezat L., Giannopoulos G., Patroumpas K., Skoutas D. 2013. *Deliverable 2.1.1. Market and Research Overview*. Document du projet GeoKnow. URL: http://svn.aksw.org/projects/GeoKnow/Public/D2.1.1_Market_and_Research_Overview.pdf
- Battle R. & Kolas D. 2012. Enabling the Geospatial Semantic Web with Parliament and GeoSPARQL. *Semantic Web*, 3(4), 355-370.
- Bauer F. & Kaltenböck M. 2011. *Linked open data: The essentials*. Edition mono/monochrom, Vienna.
- Bereta K., Nikolaou C., Karpathiotakis M., Kyzirakos K. & Koubarakis M. 2013. SexTant: Visualizing Time-Evolving Linked Geospatial Data." In: *The 12th International Semantic Web Conference (ISWC 2013)*, Sydney, Australia, October 21-25, 2013. URL : <http://www.earthobservatory.eu/SexTant>
- Berners-Lee T. 2006. *Linked Data - Design Issues*. URL: <http://www.w3.org/DesignIssues/LinkedData.html>
- Berners-Lee T., Hendler J. & Lassila O. 2001. The semantic web. *Scientific american*, 284(5), 28-37.
- Bizer C., Heath T. & Berners-Lee T. 2009. Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems*, 5(3), Pages 1-22.
- Campbell L. M. & MacNeill S. 2010. The Semantic Web, linked and open data. A Briefing Paper. Centre for Educational Technology & Interoperability Standards. URL: http://wiki.cetis.ac.uk/images/1/1a/The_Semantic_Web.pdf
- Emanuel K. 2005. Increasing destructiveness of tropical cyclones over the past 30 years. *Nature*, 436 (7051), 686-688.
- Geophysical Fluid Dynamics Society. 2015. Global Warming and Hurricanes. *An Overview of Current Research Results*. URL: <http://www.gfdl.noaa.gov/global-warming-and-hurricanes>.
- Giuliani G. & Peduzzi P. 2011. The PREVIEW Global Risk Data Platform: a geoportal to serve and share global data on risk to natural hazards. *Natural Hazards and Earth System Science* 11 (1), 53-66.
- Heath T. & Bizer C. 2011. Linked data: Evolving the web into a global data space. *Synthesis lectures on the semantic web: theory and technology*, 1(1), 1-136.
- Holland G.J. 1980. An analytic model of the wind and pressure profiles in hurricanes. *Monthly Weather Review* 108, 1212-1218.
- Knutson T. R., McBride J. L., Chan J., Emanuel K., Holland G., Landsea C., ... & Sugi M. 2010. Tropical cyclones and climate change. *Nature Geoscience*, 3(3), 157-163.
- Kolas D. & Battle R. 2012. *GEOSPARQL user guide*. URL: <http://ontolog.cim3.net/file/work/SOCoP/Educational/>
- Lehmann J., Athanasidou S., Both A. et al. 2015. *The GeoKnow Handbook*. URL: <http://svn.aksw.org/projects/GeoKnow/Public/GeoKnow-Handbook.pdf>

OGC. 2012. *OGC GeoSPARQL - A Geographic Query Language for RDF Data*. URL: <http://www.opengis.net/doc/IS/geosparql/1.0>

Peduzzi P., Chatenoux B., Dao H., De Bono A., Herold C., Kossin J., ... & Nordbeck O. 2012. Global trends in tropical cyclone risk. *Nature climate change*, 2(4), 289-294.

Peduzzi P., Dao H. & Herold H. D. C. 2005. Mapping disastrous natural hazards using global datasets. *Natural Hazards*, 35(2), 265-289.

Pielke Jr. R.A., Landsea C., Mayfield M., Laver J. & Pasch R. 2005. Hurricanes and global warming. *Bulletin of the American Meteorological Society*, 86(11), 1571-1575.

Roda C., Navarro E. & Cuesta C.E. 2013. *Analyzing Linked Data tools for SHARK*. Technical report, University of Castilla – La Mancha.

Scharrenbach T., Bischof S., Fleischli S. & Weibel R. 2012. Linked Raster Data. *GIScience 2012, Extended Abstracts* (Columbus OH, USA, 19-21 September 2012).

Schloemer R. W. 1954. Analysis and synthesis of hurricane wind patterns over Lake Okeechobee, Florida. *Hydrometeorological Report* 31, 49.

Schüpfer F. 2014. *Linked Raster Data*. Unpublished bachelor thesis, Institut für Informatik, Universität Zürich.

Shadbolt N., Hall W. & Berners-Lee T. 2006. The semantic web revisited. *Intelligent Systems, IEEE*, 21(3), 96-101.

Skjæveland M. G. 2012. Sgvizler: A JavaScript wrapper for easy visualization of SPARQL result sets. In: *9th Extended Semantic Web Conference (ESWC 2012), workshop and demo proceedings* (Heraklion, Crete, Greece, 2012).

Tauberer J. 2006. *What is RDF ?* XML.com. URL: <http://www.xml.com/pub/a/2001/01/24/rdf.html>

Usery E. L. & Varanka D. 2012. Design and development of linked data from The National Map. *Semantic Web*, 3(4), 371-384.

Verborgh R. 2014. *Serendipitous Web Applications through Semantic Hypermedia*. PhD Thesis, Ghent University.

Webster P. J., Holland G. J., Curry J. A. & Chang H. R. 2005. Changes in tropical cyclone number, duration, and intensity in a warming environment. *Science*, 309(5742), 1844-1846.

Wiegand N., Grove R., Wilson J. & Kolas D. 2014. Querying Geospatial Data over the Web: a GeoSPARQL Interface. In: *Proceedings of Workshop on Managing and Mining Enriched Geo-Spatial Data*. ACM. URL:

http://delivery.acm.org/10.1145/2620000/2619116/a4-Wiegand.pdf?ip=129.194.8.73&id=2619116&acc=ACTIVE%20SERVICE&key=FC66C24E42F07228%2E1F81E5291441A4B9%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35&CFID=731499207&CFTOKEN=71820935&acm_s=1447926816_4f73a1bea2b306da162ac7e5ef09e44d

Wood D., Zaidman M., Ruth L. & Hausenblas M. 2014. *Linked Data. Structured data on the Web*. New York: Manning.

W3C. 1999. *Resource Description Framework (RDF) Model and Syntax Specification*. W3C Recommendation, 22 February 1999. URL: <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>

Annexes : code du démonstrateur web

Le code des pages du site de démonstration geolink.grid.unep.ch, contenant notamment les requêtes SPARQL envoyées au endpoint Parliament et leur visualisation avec Sgvizler, est présenté dans cette annexe dans l'ordre de pages telles qu'elles apparaissent de gauche à droite sur la page d'accueil du site. Voir le chapitre 4.2. pour une description du site.

A. Index

```
<!DOCTYPE html>
<html lang="en">
  <head>
    <meta charset="utf-8">
    <meta http-equiv="X-UA-Compatible" content="IE=edge">
    <meta name="viewport" content="width=device-width, initial-scale=1">
    <!-- The above 3 meta tags *must* come first in the head; any other
head content must come *after* these tags -->
    <meta name="description" content="">
    <meta name="author" content="">

    <title>Geolink home</title>

    <!-- Bootstrap core CSS -->
    <link href="./css/bootstrap.min.css" rel="stylesheet">

    <!-- Custom styles for this template -->
    <link href="./cover.css" rel="stylesheet">

    <script src="../../assets/js/ie-emulation-modes-warning.js"></script>

    <!-- HTML5 shim and Respond.js for IE8 support of HTML5 elements and
media queries -->
    <!--[if lt IE 9]>
      <script
src="https://oss.maxcdn.com/html5shiv/3.7.2/html5shiv.min.js"></script>
      <script
src="https://oss.maxcdn.com/respond/1.4.2/respond.min.js"></script>
    <![endif]-->
  </head>

  <body>

    <div class="site-wrapper">

      <div class="site-wrapper-inner">

        <div class="cover-container">

          <div class="masthead clearfix">
            <div class="inner">

              <nav>
                <ul class="nav masthead-nav">
                  <li class="active"><a href="index">Home</a></li>
```

```

        <li><a href="./GeoLinkSite_PREVIEW.html">PREVIEW
dashboard</a></li>
        <li><a href="./GeoLinkSite_EMDAT.html">EMDAT
dashboard</a></li>
        <li><a href="./GeoLinkSite_SPARQL.html">SPARQL
data visualization</a></li>
        <li><a href="./QueriesExamples.html">Geospatial
SPARQL queries examples</a></li>

    </ul>
</nav>
</div>
</div>

<div class="inner cover">
    <h1 class="text-info" >GEOLINK </h1>
    <h2 class="lead">GEOLINK is the Linked Data gate to the PREVIEW
Global Risk Data Platform. It allows performing SPARQL and GEOSPARQL
queries against an RDF repository.</h2>

</div>

<div class="mastfoot">
    <div class="inner">
        <p>Demonstrator for the <a
href="http://preview.grid.unep.ch/">PREVIEW</a> Geolink resource test,
november 2015. The sample dataset presented here is an RDF transcription of
the PREVIEW

        <a
href="http://preview.grid.unep.ch/index.php?preview=data&events=cyclones&ev
cat=1&lang=eng"> Cyclones Events </a>
shapefile dataset.</p>
        <p>Image NOAA/NESDIS:<a href =
"http://www.nnvl.noaa.gov/"> http://www.nnvl.noaa.gov/ </a>
2005/08/28T1515Z</p>
    </div>
</div>

</div>

</div>

</div>

<!-- Bootstrap core JavaScript
===== -->
<!-- Placed at the end of the document so the pages load faster -->
<script
src="https://ajax.googleapis.com/ajax/libs/jquery/1.11.3/jquery.min.js"></s
cript>
<script src="./js/bootstrap.min.js"></script>
<!-- IE10 viewport hack for Surface/desktop Windows 8 bug -->
<script src="./js/ie10-viewport-bug-workaround.js"></script>
</body>
</html>

```

B. Dashboard PREVIEW

```
<!DOCTYPE HTML>
<html>
  <head>

    <title>GEOLINK Dashboard </title>
    <meta charset="UTF-8">

    <meta name="viewport" content="width=device-width, initial-scale=1">
      <!-- Bootstrap core CSS -->
    <link href="./css/bootstrap.min.css" rel="stylesheet">
    <style>
#logo
{
  background-color:white;
  color:white;
  text-align:right;
}
header {
  background-color:dimgray;
  color:white;
  text-align:center;
  padding:3px;
}
/*
 * Header
 */
.masthead-brand {
  margin-top: 10px;
  margin-bottom: 10px;
}
.masthead-nav > li {
  display: inline-block;
}
.masthead-nav > li + li {
  margin-left: 20px;
}
.masthead-nav > li > a {
  padding-right: 0;
  padding-left: 0;
  font-size: 16px;
  font-weight: bold;
  color: #fff; /* IE8 proofing */
  color: rgba(255,255,255,.75);
  border-bottom: 2px solid transparent;
}
.masthead-nav > li > a:hover,
.masthead-nav > li > a:focus {
  background-color: transparent;
  border-bottom-color: #a9a9a9;
  border-bottom-color: rgba(255,255,255,.25);
}
.masthead-nav > .active > a,
.masthead-nav > .active > a:hover,
.masthead-nav > .active > a:focus {
```

```
    color: #fff;
    border-bottom-color: #fff;
}
```

```
#section1{
    line-height:30px;
    background-color:white;
    height:400px;
    width:100%;
    padding:5px;
}
```

```
#section2{
    line-height:30px;
    background-color:white;
    height:400px;
    width:100%;
    padding:5px;
}
```

```
#section3 {
    line-height:30px;
    background-color:white;
    height:400px;
    width:100%;
    padding:5px;
}
```

```
#section4 {
    line-height:30px;
    background-color:white;
    height:400px;
    width:100%;
    padding:5px;
}
```

```
footer {
    background-color:dimgray;
    color:white;
    clear:both;
    text-align:center;
    padding:5px;
}
```

```
</style>
```

```
    <link rel="stylesheet" type="text/css" href="examples.css"><script
type="text/javascript"
src="http://ajax.googleapis.com/ajax/libs/jquery/1.9.0/jquery.js"></script>
<script type="text/javascript"
```

```

src="https://www.google.com/jsapi"></script><script type="text/javascript"
src="sgvizler.js"></script>
  <!-- Déclaration des préfixes utilisés dans l'appli sgvizler -->
  <script type="text/javascript">
    sgvizler
      .defaultEndpointOutputFormat('jsonp')
      .prefix('dbpo', 'http://dbpedia.org/ontology/');
    sgvizler.prefix('geo', 'http://www.opengis.net/ont/geosparql#');
    sgvizler.prefix('dbprop', 'http://dbpedia.org/property/');
    sgvizler.prefix('sweet',
'http://sweet.jpl.nasa.gov/2.3/stateStorm.owl#');
    sgvizler.prefix('fao',
'http://aims.fao.org/aos/geopolitical.owl#');
    sgvizler.prefix('rdfs', 'http://www.w3.org/2000/01/rdf-schema#');
    sgvizler.prefix('owl', 'http://www.w3.org/2002/07/owl#');
    sgvizler.prefix('emdat', 'http://geolink.grid.unep.ch/emdatCyc#');

    $(document).ready(function() { sgvizler.containerDrawAll(); });
  </script>

</head>
<body>

<header>
  <div id="banner"><h1>GEOLINK tropical cyclones: PREVIEW dataset</h1>

  <div class="masthead clearfix">
    <div class="inner">

      <nav>
        <ul class="nav masthead-nav">
          <li><a href="./index.html">Home</a></li>
          <li class="active"><a
href="./GeoLinkSite_PREVIEW.html">PREVIEW dashboard</a></li>
          <li><a href="./GeoLinkSite_EMDAT.html">EMDAT
dashboard</a></li>
          <li><a href="./GeoLinkSite_SPARQL.html">SPARQL
data visualization</a></li>
          <li><a href="./QueriesExamples.html">Geospatial
SPARQL queries examples</a></li>
        </ul>
      </nav>
    </div>
  </div>
</div>

</header>

<div class="container" >
<br>

  <div class="row">
    <div class="col-sm-12">
      <div class="well">The first 30 countries most affected by
tropical cyclones (Saffir-Simpson hurricane wind scale categories 1 to 5,
i.e. every tropical storm having evolved in a hurricane) </div>
    </div>
  </div>

```

```

<div class="row">
  <div class="col-sm-12">

      <!-- Définition du endpoint et de la requête
SPARQL -->

      <div id="section2" class="chart" data-sgvizler-
endpoint="http://geolink.grid.unep.ch:9090/parliament/sparql" data-
sgvizler-query="
SELECT ?country ((count(?cyc)) AS ?NCyclones)

WHERE {
?cyc sweet:SaffirSimpsonScale ?sscat .
?cyc dbprop:year ?year.
?cyc rdfs:label ?name.
?cyc fao:nameListEN ?country.
FILTER (?country != 'NoData')
FILTER (?sscat > '0'^^<http://www.w3.org/2001/XMLSchema#int>)
}
GROUP BY ?country
ORDER BY DESC (?NCyclones)
LIMIT 30

"
data-sgvizler-chart="google.visualization.PieChart" data-sgvizler-
loglevel="2" ></div> <!-- Choix du type de graphe dans la bibliothèque
sgvizler -->

      </div>

  </div>

  <div class="row">
    <div class="col-sm-12">
      <div class="well">The list of tropical cyclones/country
buffers for the 1970-2014 period.</div>
    </div>
  </div>

  <div class="row">
    <div class="col-sm-12">
      <!-- Définition du endpoint et de la requête
SPARQL -->

      <div id="section1" data-sgvizler-
endpoint="http://geolink.grid.unep.ch:9090/parliament/sparql" data-
sgvizler-query="
SELECT DISTINCT ?year ?name ?country WHERE
{
?cyclone rdfs:label ?name.
?cyclone sweet:SaffirSimpsonScale ?sscat .
?cyclone fao:nameListEN ?country.
?cyclone dbprop:year ?year.
FILTER (?country != 'NoData')
FILTER (?sscat > '0'^^<http://www.w3.org/2001/XMLSchema#int>)
}

"
data-sgvizler-chart="google.visualization.Table" data-sgvizler-loglevel="2"
></div> <!-- Choix du type de graphe dans la bibliothèque sgvizler -->

```

```

    </div>
  </div>

  <div class="row">
    <div class="col-sm-12">
      <div class="well">Yearly number of tropical cyclones/country
buffers affecting at least one country (TCBuffers)</div>
    </div>
  </div>

  <div class="row">
    <div class="col-sm-12">
      <!-- Définition du endpoint et de la requête
SPARQL -->

      <div id="section3" data-sgvizler-
endpoint="http://geolink.grid.unep.ch:9090/parliament/sparql" data-
sgvizler-query="

SELECT ?year ((count(DISTINCT ?cyc)) AS ?TCBuffers)

WHERE {

  ?cyc sweet:SaffirSimpsonScale ?sscat .
  ?cyc dbprop:year ?year.
  ?cyc rdfs:label ?name.
  ?cyc fao:nameListEN ?country.
    FILTER (?country != 'NoData')

  FILTER (?sscat > '0'^^<http://www.w3.org/2001/XMLSchema#int>)
}

  GROUP BY ?year
  ORDER BY ?year

"
data-sgvizler-chart="google.visualization.ColumnChart" data-sgvizler-
loglevel="2" ></div> <!-- Choix du type de graphe dans la bibliothèque
sgvizler -->
    </div>
  </div>

  <div class="row">
    <div class="col-sm-12">
      <div class="well">Yearly number of all cyclones buffers,
including non-landing ones, representing Saffir-Simpson categories greater
than 2, i.e. with winds over 177 km/h, 110 mph or 95 knots, which are
considered as major hurricanes in the Atlantic or Eastern Pacific
basins</div>
    </div>
  </div>

  <div class="row">
    <div class="col-sm-12">
      <!-- Définition du endpoint et
de la requête SPARQL -->

```

```

        <div id="section4" data-sgvizler-
endpoint="http://geolink.grid.unep.ch:9090/parliament/sparql" data-
sgvizler-query="

SELECT ?year ((count(?cyc)) AS ?TCBuffers)

WHERE {

    ?cyc sweet:SaffirSimpsonScale ?sscat .
    ?cyc dbprop:year ?year.
    ?cyc rdfs:label ?name.
    ?cyc fao:nameListEN ?country.

    FILTER (?sscat > '2'^^<http://www.w3.org/2001/XMLSchema#int>)
}

GROUP BY ?year
ORDER BY ?year

"
data-sgvizler-chart="google.visualization.AreaChart" data-sgvizler-
loglevel="2" ></div> <!-- Choix du type de graphe dans la bibliothèque
sgvizler -->
    </div>
</div>

<br>
<br>
<br>
<br>

</body>

<footer>
    <!--Please leave a link to the Sgvizler homepage.-->
    <p>
        Visualization of SPARQL SELECT queries with <a
href="http://dev.data2000.no/sgvizler/">Sgvizler</a> ((c) 2011--2013 Martin
G. Skjæveland).
    </p>
</footer>
    <!-- Bootstrap core JavaScript
===== -->
    <!-- Placed at the end of the document so the pages load faster -->
    <script
src="https://ajax.googleapis.com/ajax/libs/jquery/1.11.3/jquery.min.js"></s
cript>
    <script src="./js/bootstrap.min.js"></script>
    <!-- IE10 viewport hack for Surface/desktop Windows 8 bug -->
    <script src="./js/ie10-viewport-bug-workaround.js"></script>

</html>

```

C. Dashboard EM-DAT

```
<!DOCTYPE HTML>
<html>
  <head>

    <title>GEOLINK Dashboard </title>
    <meta charset="UTF-8">

    <meta name="viewport" content="width=device-width, initial-scale=1">
      <!-- Bootstrap core CSS -->
    <link href="./css/bootstrap.min.css" rel="stylesheet">
    <style>
#logo
{
  background-color:white;
  color:white;
  text-align:right;
}
header {
  background-color:dimgray;
  color:white;
  text-align:center;
  padding:3px;
}
/*
 * Header
 */
.masthead-brand {
  margin-top: 10px;
  margin-bottom: 10px;
}
.masthead-nav > li {
  display: inline-block;
}
.masthead-nav > li + li {
  margin-left: 20px;
}
.masthead-nav > li > a {
  padding-right: 0;
  padding-left: 0;
  font-size: 16px;
  font-weight: bold;
  color: #fff; /* IE8 proofing */
  color: rgba(255,255,255,.75);
  border-bottom: 2px solid transparent;
}
.masthead-nav > li > a:hover,
.masthead-nav > li > a:focus {
  background-color: transparent;
  border-bottom-color: #a9a9a9;
  border-bottom-color: rgba(255,255,255,.25);
}
.masthead-nav > .active > a,
.masthead-nav > .active > a:hover,
.masthead-nav > .active > a:focus {
```

```
    color: #fff;
    border-bottom-color: #fff;
}
```

```
#section1{
    line-height:30px;
    background-color:white;
    height:400px;
    width:100%;
    padding:5px;
}
```

```
#section2{
    line-height:30px;
    background-color:white;
    height:400px;
    width:100%;
    padding:5px;
}
```

```
#section3 {
    line-height:30px;
    background-color:white;
    height:400px;
    width:100%;
    padding:5px;
}
```

```
#section4 {
    line-height:30px;
    background-color:white;
    height:400px;
    width:100%;
    padding:5px;
}
```

```
#section5{
    line-height:30px;
    background-color:white;
    height:400px;
    width:100%;
    padding:5px;
}
```

```
#section5{
    line-height:30px;
    background-color:white;
    height:400px;
    width:100%;
    padding:5px;
}
```

```

footer {
  background-color:dimgray;
  color:white;
  clear:both;
  text-align:center;
  padding:5px;
}
</style>

<link rel="stylesheet" type="text/css" href="examples.css"><script
type="text/javascript"
src="http://ajax.googleapis.com/ajax/libs/jquery/1.9.0/jquery.js"></script>
<script type="text/javascript"
src="https://www.google.com/jsapi"></script><script type="text/javascript"
src="sgvizler.js"></script>

<!-- Déclaration des préfixes utilisés dans l'appli sgvizler -->

<script type="text/javascript">
sgvizler
.defaultEndpointOutputFormat('jsonp')
.prefix('dbpo', 'http://dbpedia.org/ontology/');
sgvizler.prefix('geo', 'http://www.opengis.net/ont/geosparql#');
sgvizler.prefix('dbprop', 'http://dbpedia.org/property/');
sgvizler.prefix('sweet',
'http://sweet.jpl.nasa.gov/2.3/stateStorm.owl#');
sgvizler.prefix('fao',
'http://aims.fao.org/aos/geopolitical.owl#');
sgvizler.prefix('rdfs', 'http://www.w3.org/2000/01/rdf-schema#');
sgvizler.prefix('owl', 'http://www.w3.org/2002/07/owl#');
sgvizler.prefix('emdat', 'http://geolink.grid.unep.ch/emdatCyc#');

$(document).ready(function() { sgvizler.containerDrawAll(); });
</script>

</head>
<body>

<header>
<div id="banner"><h1> GEOLINK tropical cyclones. EMDAT disasters
dataset</h1>

<div class="masthead clearfix">
<div class="inner">

<nav>
<ul class="nav masthead-nav">
<li><a href="./index.html">Home</a></li>
<li><a href="./GeoLinkSite_PREVIEW.html">PREVIEW
dashboard</a></li>
<li class="active"><a
href="./GeoLinkSite_EMDAT.html">EMDAT dashboard</a></li>
<li><a href="./GeoLinkSite_SPARQL.html">SPARQL data
visualization</a></li>
<li><a href="./QueriesExamples.html">Geospatial SPARQL
queries examples</a></li>
</ul>
</nav>

```

```

        </div>
    </div>
</div>
</header>

<div class="container" >
<br>

    <div class="row">
        <div class="col-sm-12">
            <div class="well">The 10 countries with the largest numbers
of deaths for the whole dataset </div>
        </div>
    </div>

    <div class="row">
        <!-- Définition du endpoint et de la
requête SPARQL -->

        <div id="section1" data-sgvizler-
endpoint="http://geolink.grid.unep.ch:9090/parliament/sparql" data-
sgvizler-query="
SELECT ?country (sum(?deaths) AS ?totDeaths)
WHERE
{ ?cyclone dbprop:year ?year.
?cyclone fao:nameListEN ?country.
?cyclone emdat:totalDeaths ?deaths.
?cyclone emdat:noInjured ?injured.
?cyclone emdat:noHomeless ?homeless.
?cyclone emdat:totalDam1000US ?damages.
?cyclone emdat:totalAffected ?affected.
}
GROUP BY ?country
ORDER BY DESC(?totDeaths)
LIMIT 10
"
data-sgvizler-chart="google.visualization.PieChart" data-sgvizler-
loglevel="2" >
        </div>
    </div>

<div class="row">
    <div class="col-sm-12">
        <div class="well">Number of deaths by region for the whole
dataset </div>
    </div>
</div>

<div class="row">
    <!-- Définition du endpoint et de la requête SPARQL -->
    <div id="section5" data-sgvizler-
endpoint="http://geolink.grid.unep.ch:9090/parliament/sparql" data-
sgvizler-query="
SELECT ?region (sum(?deaths) AS ?totDeaths)
WHERE
{ ?cyclone dbprop:year ?year.
?cyclone fao:nameListEN ?country.
?cyclone emdat:totalDeaths ?deaths.
?cyclone emdat:region ?region.

```

```

?cyclone emdat:noInjured ?injured.
?cyclone emdat:noHomeless ?homeless.
?cyclone emdat:totalDam1000US ?damages.
?cyclone emdat:totalAffected ?affected.
}
GROUP BY ?region
ORDER BY DESC(?totDeaths)

"
data-sgvizler-chart="google.visualization.PieChart" data-sgvizler-
loglevel="2" > <!-- Choix du type de graphe dans la bibliothèque sgvizler -
->
    </div>
</div>

<div class="row">
    <div class="col-sm-12">
        <div class="well">Number of victims (deaths and injured
people) by year, period 1970-2011</div>
    </div>
</div>
<div class="row">
    <div class="col-sm-12">
        <!-- Définition du endpoint et de la requête
SPARQL -->

        <div id="section2"
data-sgvizler-
endpoint="http://geolink.grid.unep.ch:9090/parliament/sparql" data-
sgvizler-query="
SELECT ?year (sum(?deaths) AS ?totDeaths) (sum(?injured) AS ?totInjured)

WHERE
{ ?cyclone dbprop:year ?year.
?cyclone fao:nameListEN ?country.
?cyclone emdat:totalDeaths ?deaths.
?cyclone emdat:noInjured ?injured.
?cyclone emdat:noHomeless ?homeless.
?cyclone emdat:totalDam1000US ?damages.
?cyclone emdat:totalAffected ?affected.
}
GROUP BY ?year
ORDER BY ?year

"
data-sgvizler-chart="google.visualization.AreaChart" data-sgvizler-
loglevel="2" > <!-- Choix du type de graphe dans la bibliothèque sgvizler -
->
    </div>
</div>
</div>

<div class="row">
    <div class="col-sm-12">
        <div class="well">Number of affected people by year, period
1970-2011</div>
    </div>
</div>

<div class="row">
    <div class="col-sm-12">

```

```

                                                                    <!-- Définition du endpoint et de la
requête SPARQL -->

        <div id="section3" data-sgvizler-
endpoint="http://geolink.grid.unep.ch:9090/parliament/sparql" data-
sgvizler-query="
SELECT   ?year ( (sum( ?deaths))/(sum(?affected) )AS ?VictimIndex)

WHERE
{ ?cyclone dbprop:year ?year.
?cyclone emdat:totalDeaths ?deaths.
?cyclone emdat:noInjured ?injured.
?cyclone fao:nameListEN ?country.
?cyclone emdat:totalDam1000US ?damages.
?cyclone emdat:totalAffected ?affected.
?cyclone emdat:noHomeless ?homeless.

}
GROUP BY ?year
ORDER BY ?year

"
data-sgvizler-chart="google.visualization.AreaChart" > <!-- Choix du type
de graphe dans la bibliothèque sgvizler -->
        </div>
        </div>
        </div>

        <div class="row">
            <div class="col-sm-12">
                <div class="well">Estimation of damages in billion USD by
year, period 1970-2011</div>
            </div>
        </div>

        <div class="row">
            <div class="col-sm-12">
                                                                    <!-- Définition du endpoint et de la
requête SPARQL -->

                <div id="section4" data-sgvizler-
endpoint="http://geolink.grid.unep.ch:9090/parliament/sparql" data-
sgvizler-query="
SELECT ?year ((sum(?damages)/1000000) AS ?totDamage)

WHERE
{ ?cyclone dbprop:year ?year.
?cyclone emdat:totalDeaths ?deaths.
?cyclone fao:nameListEN ?country.
?cyclone emdat:totalDam1000US ?damages.
?cyclone emdat:totalAffected ?affected.
}
GROUP BY ?year
ORDER BY ?year

"
data-sgvizler-chart="google.visualization.AreaChart" > <!-- Choix du type
de graphe dans la bibliothèque sgvizler -->
                </div>
            </div>
        </div>

```

```
</div>
<
</div>
<br>
<br>
</body>
<footer >
  <!--Please leave a link to the Sgvizler homepage.-->
  <p> Visualization of SPARQL SELECT queries with <a
href="http://dev.data2000.no/sgvizler/">Sgvizler</a> ((c) 2011--2013 Martin
G. Skjæveland).
  </p>
</footer>
</html>
```

D. Formulaire SPARQL

```

<!DOCTYPE HTML>
<html>
  <head>
    <meta http-equiv="Content-Type" content="text/html; charset=utf-8">
    <title>GEOLINK SPARQL</title>
    <meta charset="UTF-8">
    <meta name="viewport" content="width=device-width, initial-
scale=1">
    <!-- Bootstrap core CSS -->
    <link href="./css/bootstrap.min.css" rel="stylesheet">
    <link rel="stylesheet" type="text/css" href="examples.css"><script
type="text/javascript"
src="http://ajax.googleapis.com/ajax/libs/jquery/1.9.0/jquery.js"></script>
<script type="text/javascript"
src="https://www.google.com/jsapi"></script><script type="text/javascript"
src="sgvizler.js"></script>
    <!-- Déclaration des préfixes utilisés dans l'appli sgvizler -->
    <script type="text/javascript">

        sgvizler.prefix('geo', 'http://www.opengis.net/ont/geosparql#');
        sgvizler.prefix
('geof', 'http://www.opengis.net/def/function/geosparql/');
        sgvizler.prefix('dbpo', 'http://dbpedia.org/ontology/');
        sgvizler.prefix('dbprop', 'http://dbpedia.org/property/');
        sgvizler.prefix('sweet',
'http://sweet.jpl.nasa.gov/2.3/stateStorm.owl#');
        sgvizler.prefix('fao',
'http://aims.fao.org/aos/geopolitical.owl#');
        sgvizler.prefix('rdfs', 'http://www.w3.org/2000/01/rdf-schema#');
        sgvizler.prefix('owl', 'http://www.w3.org/2002/07/owl#');
        sgvizler.prefix('emdat', 'http://geolink.grid.unep.ch/emdatCyc#');
        sgvizler.prefix('city', 'http://geolink.grid.unep.ch:9091/test-
2005/dbpedia-cities-10000-csv#');
    <!--$(document).ready(function() { sgvizler.containerDrawAll(); });--
>
    </script>

    <style>
header {
  background-color:dimgray;
  color:white;
  text-align:center;
  padding:3px;
}

/*
 * Header
 */
.masthead-brand {
  margin-top: 10px;
  margin-bottom: 10px;
}

.masthead-nav > li {
  display: inline-block;
}
.masthead-nav > li + li {

```

```
    margin-left: 20px;
}
.masthead-nav > li > a {
    padding-right: 0;
    padding-left: 0;
    font-size: 16px;
    font-weight: bold;
    color: #fff; /* IE8 proofing */
    color: rgba(255,255,255,.75);
    border-bottom: 2px solid transparent;
}
.masthead-nav > li > a:hover,
.masthead-nav > li > a:focus {
    background-color: transparent;
    border-bottom-color: #a9a9a9;
    border-bottom-color: rgba(255,255,255,.25);
}
.masthead-nav > .active > a,
.masthead-nav > .active > a:hover,
.masthead-nav > .active > a:focus {
    color: #fff;
    border-bottom-color: #fff;
}

#section1{
    line-height:30px;
    background-color:white;
    height:400px;
    width:100%;
    padding:5px;
}

#section2{
    line-height:30px;
    background-color:white;
    height:400px;
    width:100%;
    padding:5px;
}

#section3 {
    line-height:30px;
    background-color:white;
    height:400px;
    width:100%;
    padding:5px;
}

footer {
    background-color:dimgray;
    color:white;
    clear:both;
}
```

```

    text-align:center;
    padding:5px;
}
</style>
</head>
<body>

<header>
  <div id="banner"><h1>GEOLINK tropical cyclones: SPARQL data
visualization </h1>

  <div class="masthead clearfix">
    <div class="inner">

      <nav>
        <ul class="nav masthead-nav">
          <li><a href="./index.html">Home</a></li>
          <li><a href="./GeoLinkSite_PREVIEW.html">PREVIEW
dashboard</a></li>
          <li><a href="./GeoLinkSite_EMDAT.html">EMDAT
dashboard</a></li>
          <li class="active"><a
href="./GeoLinkSite_SPARQL.html">SPARQL data visualization</a></li>
          <li><a href="./QueriesExamples.html">Geospatial
SPARQL queries examples</a></li>
        </ul>
      </nav>
    </div>
  </div>
</div>

</header>

<div class="container" >
<br>
  <div class="row">
    <div class="col-sm-12">

      <div class="well">The GEOLINK SPARQL query form allows querying
the Geolink tropical cyclones buffer dataset with <a
href="http://www.w3.org/TR/rdf-sparql-query/"
target="_blank"><b>SPARQL</b></a>. The default query below returns a
table of buffers in WKT format (Well Known text), showing the maximal
spatial extent (i.e. when reaching a Saffir-Simpson category of 0) of
tropical cyclones in 1970 .
      </div>

      <!-- Appel de la fonction sgvizler formDraw
qui affiche un formulaire de saisie de requêtes SPARQL et un liste de
visualisateurs à choix -->

      <div id="myForm"></div>
      <script type="text/javascript">
        $(document).ready(function() { sgvizler.formDraw("myForm");
});
      </script>
    </div>
  </div>
</div>
<br>

```

```
<br>

</body>
<footer>
  <!--Please leave a link to the Sgvizler homepage.-->
  <p>
    Visualization of SPARQL SELECT queries with <a
href="http://dev.data2000.no/sgvizler/">Sgvizler</a> ((c) 2011--2013 Martin
G. Skjæveland).

  </p>
</footer>
  <!-- Bootstrap core JavaScript
===== -->
  <!-- Placed at the end of the document so the pages load faster -->
  <script
src="https://ajax.googleapis.com/ajax/libs/jquery/1.11.3/jquery.min.js"></s
cript>
  <script src="./js/bootstrap.min.js"></script>
  <!-- IE10 viewport hack for Surface/desktop Windows 8 bug -->
  <script src="./js/ie10-viewport-bug-workaround.js"></script>
  </body>

</html>
```

E. Exemples de requêtes

```

<!DOCTYPE HTML>
<html>
  <head>
    <meta http-equiv="Content-Type" content="text/html; charset=utf-8">
    <title>GEOLINK examples</title>
    <meta charset="UTF-8">
    <meta name="viewport" content="width=device-width, initial-
scale=1">
    <!-- Bootstrap core CSS -->
    <link href="./css/bootstrap.min.css" rel="stylesheet">
    <link rel="stylesheet" type="text/css" href="examples.css"><script
type="text/javascript"
src="http://ajax.googleapis.com/ajax/libs/jquery/1.9.0/jquery.js"></script>
<script type="text/javascript"
src="https://www.google.com/jsapi"></script><script type="text/javascript"
src="sgvizler.js"></script>
    <!-- Déclaration des préfixes utilisés dans l'appli sgvizler -->
    <script type="text/javascript">

      sgvizler.prefix('geo', 'http://www.opengis.net/ont/geosparql#');
      sgvizler.prefix
('geof', 'http://www.opengis.net/def/function/geosparql/');
      sgvizler.prefix('dbpo', 'http://dbpedia.org/ontology/');
      sgvizler.prefix('dbprop', 'http://dbpedia.org/property/');
      sgvizler.prefix('sweet',
'http://sweet.jpl.nasa.gov/2.3/stateStorm.owl#');
      sgvizler.prefix('fao',
'http://aims.fao.org/aos/geopolitical.owl#');
      sgvizler.prefix('rdf', 'http://www.w3.org/1999/02/22-rdf-syntax-
ns#');
      sgvizler.prefix('rdfs', 'http://www.w3.org/2000/01/rdf-schema#');
      sgvizler.prefix('owl', 'http://www.w3.org/2002/07/owl#');
      sgvizler.prefix('emdat', 'http://geolink.grid.unep.ch/emdatCyc#');
      sgvizler.prefix('foaf', 'http://xmlns.com/foaf/0.1/');
      sgvizler.prefix('city', 'http://geolink.grid.unep.ch:9091/test-
2005/dbpedia-cities-10000-csv#');

      $(document).ready(function() { sgvizler.containerDrawAll(); });
    </script>

    <style>
header {
  background-color:dimgray;
  color:white;
  text-align:center;
  padding:3px;
}
/*
 * Header
 */
.masthead-brand {
  margin-top: 10px;
  margin-bottom: 10px;
}

```

```
.masthead-nav > li {
  display: inline-block;
}
.masthead-nav > li + li {
  margin-left: 20px;
}
.masthead-nav > li > a {
  padding-right: 0;
  padding-left: 0;
  font-size: 16px;
  font-weight: bold;
  color: #fff; /* IE8 proofing */
  color: rgba(255,255,255,.75);
  border-bottom: 2px solid transparent;
}
.masthead-nav > li > a:hover,
.masthead-nav > li > a:focus {
  background-color: transparent;
  border-bottom-color: #a9a9a9;
  border-bottom-color: rgba(255,255,255,.25);
}
.masthead-nav > .active > a,
.masthead-nav > .active > a:hover,
.masthead-nav > .active > a:focus {
  color: #fff;
  border-bottom-color: #fff;
}
```

```
#map1{
  line-height:30px;
  background-color:white;
  height:400px;
  width:100%;
  padding:5px;
  border:1px solid black;
}
```

```
#map2{
  line-height:30px;
  background-color:white;
  height:400px;
  width:100%;
  padding:5px;
  border:1px solid black;
}
```

```
#map3{
  line-height:30px;
  background-color:white;
  height:400px;
  width:100%;
  padding:5px;
  border:1px solid black;
}
```

```
#query1{
```

```

    line-height:30px;
    background-color:white;
    height:400px;
    width:100%;
    padding:5px;
    border:1px solid black;
}

#section2{
    line-height:30px;
    background-color:white;
    height:400px;
    width:100%;
    padding:5px;
}

#section3 {
    line-height:30px;
    background-color:white;
    height:400px;
    width:100%;
    padding:5px;
}

footer {
    background-color:dimgray;
    color:white;
    clear:both;
    text-align:center;
    padding:5px;
}

</style>
</head>
<body>

<header>
  <div id="banner"><h1>GEOLINK tropical cyclones: Geospatial queries
examples </h1>

  <div class="masthead clearfix">
    <div class="inner">

      <nav>
        <ul class="nav masthead-nav">
          <li><a href="./index.html">Home</a></li>
          <li><a href="./GeoLinkSite_PREVIEW.html">PREVIEW
dashboard</a></li>
          <li><a href="./GeoLinkSite_EMDAT.html">EMDAT
dashboard</a></li>
          <li><a href="./GeoLinkSite_SPARQL.html">SPARQL
data visualization</a></li>

```

```

        <li class="active"><a
href="./QueriesExamples.html">Geospatial SPARQL queries examples</a></li>
    </ul>
</nav>
</div>
</div>
</div>
</header>

<div class="container" >
<br>

<div class="row">
    <div class="col-sm-12">
        <div class="well">Hurricane Katrina (2005) when reaching
the Saffir-Simpson category 2, with some linked stats from EMDAT: the
number of deaths and the damages in billion USD in USA</div>
    </div>

    <div class="row">
        <div class="col-sm-8">
            <!-- Définition du endpoint et de la requête SPARQL -->

            <div id="map1"
data-sgvizler-
endpoint="http://geolink.grid.unep.ch:9090/parliament/sparql"
data-sgvizler-query="
SELECT ?cycwkt ?name ?deaths ((?damage)/1000000 AS ?damageBilionUSD)

WHERE {
?cyc dbprop:year '2005'^^<http://www.w3.org/2001/XMLSchema#int>.
?cyc sweet:SaffirSimpsonScale '2'^^<http://www.w3.org/2001/XMLSchema#int> .
?cyc geo:hasGeometry ?cycGeom.
?cycGeom geo:asWKT ?cycwkt.
?cyc rdfs:label ?name.
?cyc fao:nameListEN 'United States'

OPTIONAL{
?cyc owl:sameAs ?emdata.
?emdata emdat:totalDeaths ?deaths.
?emdata emdat:totalDam1000US ?damage.
?emdata fao:nameListEN ?country}
FILTER (?name ='KATRINA')

}
"
            data-sgvizler-chart="sgvizler.visualization.MapWKT" <!-- Choix du type
de graphe dans la bibliothèque sgvizler -->

            >
                </div>
            </div>
        <div class="col-sm-4">

            <!-- Affichage simple sur la page de la requête utilisée pour créer
la carte précédente -->

            <div class="well"> SELECT ?cycwkt ?name ?deaths ?damage

```

```

WHERE <BR>{
?cyc dbprop:year '2005'.<BR>
?cyc sweet:SaffirSimpsonScale '2'.<BR>
?cyc geo:hasGeometry ?cycGeom.<BR>
?cycGeom geo:asWKT ?cycwkt.<BR>
?cyc rdfs:label ?name.<BR>
?cyc fao:nameListEN 'United States' <BR>

OPTIONAL{<BR>
?cyc owl:sameAs ?emdata.<BR>
?emdata emdat:totalDeaths ?deaths.<BR>
?emdata emdat:totalDam1000US ?damage.<BR>
?emdata fao:nameListEN ?country}<BR>
FILTER (?name = 'KATRINA')<BR>
}<BR>

</div>

</div>

</div>
<br>
<br>
<div class="row">
  <div class="col-sm-12">
    <div class="well">A geospatial query using the GEOSPARQL
standard, against a DBpedia dataset of the cities with more than 10'000
inhabitants. The query returns every city affected by the Hurricane Katrina
in 2005, when reaching the Saffir-Simpson category 2. The popups show the
websites of affected cities and their population</div>
  </div>
</div>
<div class="row">
  <div class="col-sm-8">
    <!-- Définition du endpoint et de la requête SPARQL -->

    <div id="map3"
data-sgvizler-
endpoint="http://geolink.grid.unep.ch:9090/parliament/sparql"
data-sgvizler-query="
SELECT ?citywkt ?name ?web ?pop

WHERE {

  ?city geo:hasGeometry ?geom;
    city:cityName ?name;
    city:website ?web;
    city:pop ?pop.

  ?geom geo:asWKT ?citywkt .
?cyc rdfs:label 'KATRINA';
sweet:SaffirSimpsonScale '2'^^<http://www.w3.org/2001/XMLSchema#int>;
geo:hasGeometry ?cycGeom.
?cycGeom geo:asWKT ?cycwkt.

FILTER (geof:sfWithin(?citywkt, ?cycwkt))
}
"

```

```

data-sgvizler-chart="sgvizler.visualization.MapWKT"      <!-- Choix du
type de graphe dans la bibliothèque sgvizler -->

```

```

>
    </div>
  </div>
  <div class="col-sm-4">
    <div class="well"> SELECT ?citywkt ?name ?web ?pop <br>

WHERE {<br>

  ?city geo:hasGeometry ?geom;<br>
        city:cityName ?name;<br>
        city:website ?web;<br>
        city:pop ?pop.<br>
  ?geom geo:asWKT ?citywkt.<br>
  ?cyc rdfs:label 'KATRINA';<br>
  sweet:SaffirSimpsonScale '2';<br>
  geo:hasGeometry ?cycGeom.<br>
  ?cycGeom geo:asWKT ?cycwkt.<br>

FILTER (geof:sfWithin(?citywkt, ?cycwkt))<br>
}<br>
  }<br>
  </div>

  </div>

<br>
<br>

<div class="row">
  <div class="col-sm-12">
    <div class="well">A federated query (Geolink-DBpedia),
returning the maximal extent of tropical cyclones in 2005 with the
corresponding DBpedia abstracts and Wikipedia entries URLs </div>
  </div>

  <div class="row">
    <div class="col-sm-8">
      <!-- Définition du endpoint et de la requête SPARQL -->

      <div id="map2"
data-sgvizler-
endpoint="http://geolink.grid.unep.ch:9090/parliament/sparql"
data-sgvizler-query="
SELECT DISTINCT
?cycwkt ?name ?abstract ?wiki

WHERE {
?cyc rdfs:label ?name;
  sweet:SaffirSimpsonScale '0'^^<http://www.w3.org/2001/XMLSchema#int>;
  owl:sameAs ?event;
  dbprop:year '2005'^^<http://www.w3.org/2001/XMLSchema#int>;
  geo:hasGeometry ?cycGeom.
?cycGeom geo:asWKT ?cycwkt.

```

```

SERVICE <http://dbpedia.org/sparql> {
  ?event rdf:type <http://dbpedia.org/class/yago/Cyclone111443721>;
  rdfs:label ?label;
  dbprop:year ?year;
  foaf:isPrimaryTopicOf ?wiki;
  dbprop:year '2005'^^<http://www.w3.org/2001/XMLSchema#int>;
  dbpo:abstract ?abstract.

  FILTER (lang(?abstract) = 'en')
}

"
  data-sgvizler-chart="sgvizler.visualization.MapWKT" <!-- Choix du type
de graphe dans la bibliothèque sgvizler -->

>
  </div>
</div>
<div class="col-sm-4">
  <div class="well"> SELECT DISTINCT
?cycwkt ?name ?abstract <br>?wiki <br>

WHERE {
?cyc rdfs:label ?name;<br>
sweet:SaffirSimpsonScale '0'; <br>
owl:sameAs ?event;<br>
dbprop:year '2005';<br>
geo:hasGeometry ?cycGeom.<br>
?cycGeom geo:asWKT ?cycwkt.<br>

SERVICE <http://dbpedia.org/sparql> {<br>
  ?event rdf:type http://dbpedia.org/class/yago/Cyclone111443721<br>
  rdfs:label ?label;<br>
  dbprop:year ?year;<br>
  foaf:isPrimaryTopicOf ?wiki;<br>
  dbprop:year '2005';<br>
  dbpo:abstract ?abstract.<br>

  FILTER (lang(?abstract) = 'en')<br>
}<br>
}<BR>

  </div>

  </div>

</div>

<br>
<br>

</div>
  </body>

</html>

```