xBRL-CSV y datos granulares - El ejemplo de AnaCredit, una prueba de concepto de XBRL Europa



Publicado el 22 de septiembre de 2021 por [**Editor**](https://www.xbrl.org/xbrl-csv-and-granular-data-the-anacredit-example-a-proof-of-concept-from-xbrl-europe/)

*Esta es una publicación de invitado de Vincent Le Moal-Joubel, científico de datos y experto en XBRL en Banque de France, basada en su presentación en el*[*evento 28th XBRL Europe Digital Week sobre informes bancarios y de seguros*](https://www.xbrleurope.org/?p=4496)*, el 23 de junio de 2021. Ofrece una importante prueba de concepto sobre el uso de xBRL-CSV para informes europeos. xBRL-CSV es uno de los nuevos formatos que se ofrecen dentro del conjunto de especificaciones del*[*Modelo abierto de información (OIM)*](https://www.xbrl.org/the-standard/what/introducing-the-oim/)*, y es ideal para manejar grandes cantidades de información detallada. Siga leyendo (y haga clic para ver las imágenes a tamaño completo), mire junto con el video de la presentación*[*aquí*](https://www.xbrleurope.org/wp-content/uploads/2021/05/Bank-Insurance-digital-event-23-June-28th-XBRL-Europe-digital-week_ANACREDIT.mp4)*, o vea las diapositivas*[*aquí*](https://www.xbrleurope.org/wp-content/uploads/2021/05/4-28thXBRLEuropeDay_BankInsuranceWorkingGroup_IntegratedReportingAnacredit.pdf)*.*

En nombre del Grupo de Trabajo de Banca y Seguros de XBRL Europa, me gustaría presentar nuestro trabajo hasta ahora sobre cómo se puede usar el formato xBRL-CSV para manejar datos granulares, dentro de un enfoque de informes cada vez más integrado. Nuestra prueba de concepto sugiere que xBRL-CSV podría agilizar el proceso de generación de informes y facilitar a los usuarios la comparación y el análisis de información de diferentes países y requisitos de generación de informes, incluidos datos extensos y detallados.

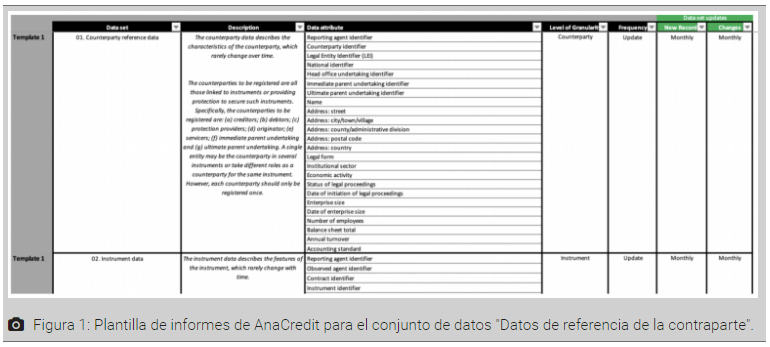
Queríamos ver una iniciativa de informes existente que es conocida por producir grandes volúmenes de datos y, por lo tanto, tomamos a AnaCredit como nuestro ejemplo. AnaCredit, que significa 'conjuntos de datos analíticos de crédito', es un proyecto del Banco Central Europeo (BCE) que requiere la presentación de información detallada sobre préstamos bancarios individuales en toda la zona del euro. Como era de esperar, esto representa una gran cantidad de datos.

AnaCredit se lanzó en 2011, aunque las primeras presentaciones no se realizaron hasta 2018. Incluso desde el principio, estábamos interesados ​​​​en usar AnaCredit como un caso de estudio, pero en ese momento no teníamos las herramientas de modelado de puntos de datos (DPM) que tenemos. hoy, y ni siquiera se pensó en xBRL-CSV. Por lo tanto, era demasiado pronto para adaptar AnaCredit a XBRL, pero ahora tenemos capacidades muy diferentes a nuestra disposición. Nuestro objetivo no es efectuar un cambio en los informes de AnaCredit, sino usarlos como un ejemplo de cómo podemos manejar esta cantidad de datos y probarlos con nuestros procesos.

*Modelo de punto de datos para conectar datos con significado*

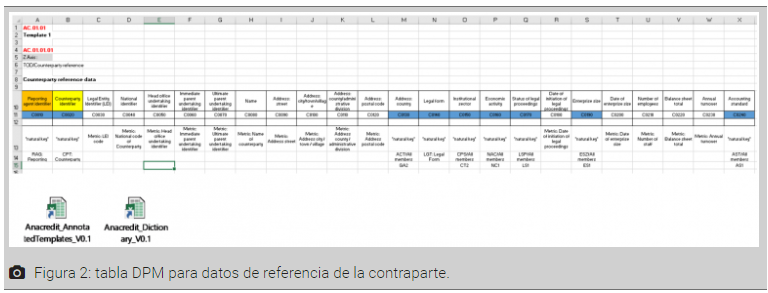
Nuestro primer paso fue capturar las definiciones utilizadas para AnaCredit. Una de las dificultades que enfrentamos fue que estas definiciones no son exactamente las mismas en todas partes, ya que cada país ha decidido diferentes procesos de presentación de informes. Eso significa que, para los proveedores de software, trabajar con AnaCredit en toda Europa es una pesadilla porque las reglas cambian de un país a otro.

Nuestra solución fue aplicar un modelo de puntos de datos a los datos y las reglas de archivo adjuntas. Un DPM conecta la plantilla de informes legibles por humanos con las definiciones técnicas en la taxonomía XBRL, dando significados legibles por máquina a los conceptos de informes.



La Figura 1, por ejemplo, muestra la plantilla de informes de AnaCredit para 'Datos de referencia de la contraparte', con una lista de campos para que los utilicen los contribuyentes. Nuestra tarea era identificar los datos y crear conceptos para cada hecho o, idealmente, tomar conceptos existentes utilizados en otros informes, por ejemplo, por la Autoridad Bancaria Europea (EBA).

La Figura 2 muestra los requisitos de información de datos de referencia de la contraparte convertidos al DPM, produciendo una plantilla anotada en formato de tabla. También viene con un diccionario de definiciones de conceptos y sus orígenes. Esto combina conceptos específicos que desarrollamos para AnaCredit con conceptos EBA importados, como la definición del código del Identificador de Entidad Legal (LEI), que ya está implementado en los informes EBA.



Una de las complicaciones es que la forma en que se reportan los datos de Anacredit varía considerablemente entre países. La Figura 3 muestra algunos ejemplos tomados de informes digitales reales de diferentes países. Podemos ver que si tomamos el elemento 'Número de empleados' (encerrado en un círculo rojo), en Alemania este concepto se denomina "NMBR\_EMPLYS", en Francia "NUMBR\_EMPLY" y en Italia "NumberOfEmployees". El DPM nos permite vincular con éxito todos estos al concepto subyacente.

Si bien todos estos países usan XML para generar informes, la estructura del código es muy diferente. Finalmente, algunos países también tienen requisitos de informes nacionales adicionales, reflejados en campos adicionales. Por ejemplo, en Francia, los declarantes también deben informar el SIREN, un identificador nacional, para cada contraparte (encerrado en un círculo morado). El objetivo del grupo de trabajo era ver cómo se podían abordar estos problemas.



*Ventajas de XBRL*

El beneficio más importante que aporta XBRL es la estandarización de los datos. El diccionario y los conceptos compartidos guían a los declarantes sobre cómo informar correctamente cada hecho y garantizar que los datos sean comparables. Es importante destacar que los datos se generan y consumen fácilmente utilizando las herramientas familiares de XBRL, que ya están disponibles para los informes actuales en la mayoría de los bancos y las autoridades nacionales competentes (NCA) de toda Europa.

XBRL también proporciona extensibilidad. Los países individuales o los reguladores pueden ampliar la taxonomía para agregar campos específicos para capturar sus propios requisitos de informes. Banque de France, por ejemplo, agrega campos adicionales a la plantilla de informes del Fondo Europeo de Pensiones. Con XBRL, estos se manejan fácilmente y los datos adicionales se filtran antes de enviarlos a la Autoridad Europea de Seguros y Pensiones de Jubilación (EIOPA). En otras palabras, XBRL nos brinda una forma estándar de agregar campos personalizados según sea necesario, sin afectar la integridad de los informes centrales compartidos. Al mismo tiempo, la capacidad de reutilizar conceptos existentes ayuda a facilitar la comparación y el análisis de datos de diferentes informes.

La Figura 4 muestra una visualización fácil de usar de algunas líneas de un informe XBRL de AnaCredit, que se muestra como se especifica en nuestra taxonomía de prueba de concepto. Una herramienta XBRL estándar puede producir esta visualización usando información incrustada en el paquete de informes. Otra ventaja de XBRL es la capacidad de los usuarios para determinar cómo se presentan los datos digitales, potencialmente utilizando plantillas diferentes o modificadas. De particular interés en Polyglot Europe es el potencial para integrar de forma nativa múltiples idiomas para mostrar etiquetas y conceptos de tablas, y permitir que los espectadores cambien automáticamente de un idioma a otro. Esto ya se está haciendo para los informes de la Junta Única de Resolución (SRB), por ejemplo, donde la taxonomía está disponible en diferentes idiomas.

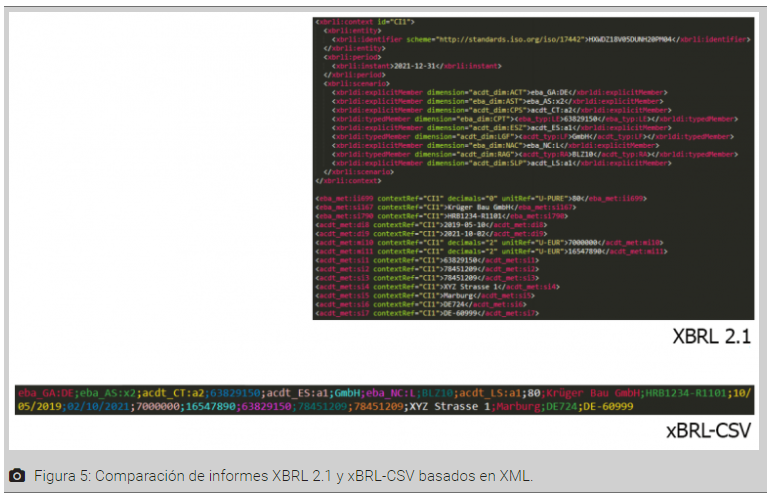


El último, pero no menos importante, beneficio clave de XBRL es el potencial de validación para verificar datos, detectar problemas y mejorar la calidad. Las reglas de validación se pueden definir e implementar de forma nativa para ejecutarse en herramientas XBRL estándar. Los resultados de la validación dependen de los niveles de precisión establecidos para los datos; la capacidad de ajustar estos añade sensibilidad a los controles de validación.

*Ventajas de xBRL-CSV*

Además de los beneficios generales de XBRL, la ventaja adicional más importante de xBRL-CSV es que produce archivos mucho más pequeños, lo que hace que los informes sean mucho más fáciles de producir, manejar, enviar y almacenar. La Figura 5 muestra una comparación de dos informes, con los mismos datos capturados en los formatos tradicionales XBRL 2.1 basados ​​en XML y xBRL-CSV. Incluso sin una gran cantidad de conocimientos técnicos, está claro que en el informe XML los metadatos ocupan mucho espacio y los datos informados, que se muestran en blanco, son solo una fracción del contenido del código. Los datos en xBRL-CSV son notablemente más compactos.

En general, el uso de xBRL-CSV para los datos de AnaCredit reduce el tamaño del archivo casi diez veces, en una proporción lineal. Por ejemplo, una tabla con 300.000 filas produce un archivo de 62 Mb en xBRL-CSV y un archivo de 574 Mb en XBRL 2.1. Para una tabla con 3.000.000 de filas, esos números son 622 Mb y 5.744 Mb respectivamente.



Los archivos xBRL-CSV también son fáciles de leer y editar. Todos los datos están contenidos en un archivo o archivos CSV simples, mientras que los metadatos están definidos en un archivo JSON separado muy pequeño, al igual que las propiedades de la tabla. Esto facilita que los declarantes y otros usuarios se concentren en los datos; es muy sencillo acceder a datos específicos y editar los archivos CSV sin conocimientos de XBRL, utilizando herramientas conocidas y extendidas. Este acceso rápido a los datos también facilita potencialmente el proceso de validación.

*Una base para la experimentación*

Esta prueba de concepto nos brinda una buena base para preparar informes integrados de datos más granulares, incluidos conocimientos sobre cómo proceder con experimentos en grandes volúmenes de datos reales y cómo cumplir con las necesidades y los desafíos de validación. Como mencioné, no estamos trabajando para cambiar AnaCredit *per se* , sino para comprender cómo aplicar xBRL-CSV a grandes conjuntos de datos para optimizar los informes.

Este es solo el comienzo de nuestro trabajo, y pretendemos que sirva como base para más experimentación. Algunos de nuestros próximos pasos podrían incluir agregar más tablas de AnaCredit al DPM, lo que también nos permitiría probar y analizar los controles de validación de AnaCredit existentes entre y dentro de las tablas, así como implementar y evaluar una validación adicional compatible. También nos gustaría agregar etiquetas en varios idiomas para ilustrar esta funcionalidad y, si es posible, experimentar con datos anónimos reales.

No hay duda de que las tendencias en la presentación de informes se están moviendo hacia la recopilación de grandes volúmenes de datos más granulares, pero estos datos no tienen sentido a menos que puedan manejarse y analizarse de manera efectiva. Hasta el momento, xBRL-CSV parece ofrecer una solución eficiente y práctica para la presentación de informes en el futuro, en particular para reunir y dar sentido a la información de diversas jurisdicciones.

[MODELO DE PUNTO DE DATOS](https://www.xbrl.org/tag/data-point-model/)[EUROPA](https://www.xbrl.org/tag/europe/)[DATOS GRANULARES](https://www.xbrl.org/tag/granular-data/)[INFORMES](https://www.xbrl.org/tag/reporting/)[XBRL](https://www.xbrl.org/tag/xbrl/)[XBRL EUROPA](https://www.xbrl.org/tag/xbrl-europe/)[XBRL-CSV](https://www.xbrl.org/tag/xbrl-csv/)