Project Lintol is officially launched with funding from the Open Data Institute.

We're delighted to announce that Belfast based **Project Lintol** has been awarded funding from the **Open Data Institute.**

Project Lintol: what are we building?

We're building an Open Data validating tool. Put simply, this is a bit like a grammar checker for Open Data (data which everyone can access, use or share). It's designed to be used just before the point of publishing and will detect errors and mistakes in a data set; alerting the data owner and allowing them to make any crucial fixes before their data is released to the public.

The validating tool has been created for an [Open Data Institute](#) R&D project exploring how better publishing tools can improve the quality, speed and cost-effectiveness of data publishing.

Many open data sets are published using Open Data platforms specifically designed to streamline the process of publishing, sharing and finding information. We will develop Lintol so that it can be integrated into such platforms, starting with CKAN (Comprehensive Knowledge Archive Network) which is used by Governments globally to publish their data. Users will simply upload their data to the CKAN platform as normal and select Lintol to validate their data sets before publishing.

Who's funding us?

The funding is from the Open Data Institute. Innovate UK, the UK's innovation agency, is providing £6 million over three years to the Open Data Institute, to advance knowledge and expertise in how data can shape the next generation of public and private services, and create economic growth. Some of this work will be carried out by the ODI itself, and some by external experts, as in the case of Lintol. Work on improving the conditions for data publishing is one of six projects, chosen by the ODI, in this first year of the funding. Lintol's validating tool is a key element of this data publishing project.

Offering solutions to a timely problem…

It's a pretty electrifying time for Open Data in the UK. An unprecedented surge in Open Data publishing is happening as both public and private sector organisations make the transformation towards data transparency. From healthcare and medicine, to the environment, economics, politics, agriculture and transport: Open Data holds within it the possibility to inspire innovation and change.

In order for Open Data to be of its utmost use to the data consuming community, it needs to be of the highest possible quality; free from errors and mistakes. Data sets which contain incorrect information can often be completely unusable. Data publishers go to great lengths to check their data manually; but this is time consuming. Despite thorough checks, inconsistencies in data are detected ***after*** the point of publishing: by then it's too late to fix.

> **"Poor quality data is a serious roadblock, restricting the flow of Open Data and cutting short its potential journey".**
> *Martin Naughton, Lead Developer, Project Lintol.*
>
> ***"Creating better data quality tools not only increases trust in the data being published, it also improves the confidence organisations have in the data they publish - removing a major barrier to the publication of data, and significantly helping realise the promises of open data."***
>
> *Olivier Thereaux, Head of Technology at the Open Data Institute.*

Who can benefit from using Lintol?

Basically, any public or private body involved in publishing Open Data. For example, Belfast City Council or NHS Digital both publish Open Data, as do Deutsche Bank and Arup. We carried out our own research, and realised

that no two Data Publishing teams are the same. Some team members are technical, some aren't. What's more, there's no one definitive publishing team structure. With this in mind, we are developing Lintol so it can be used by both technical and non-technical users, and can fit seamlessly into a team's supply chain or sequence of publishing.

We're hoping to help erase the "fear" described to us by Data Owners who are keen to avoid publishing flawed data.

Follow our progress, offer us your insights.

For us, it's key that the principles of Open Data and transparency are part of Lintol's genetic code, and we're using a Python based open source workflow engine. To keep up to date on our progress, or ask our team members direct questions, join us on GitHub github.com/lintol, or Twitter @projectlintol.

Editors' Notes / further information

A more techie overview…

Lintol will provide an automated system incorporating pre-existing, high quality data validation tools such as CSV Lint and Goodtables-py, and offer new plug-ins for additional data types. For example, Lintol will be distinct in offering the following functions and possibly more:
Checking that geospatial points (in, say a GeoJSON dataset are within boundaries...Analysing how many points in a 3D point set are outside a bounding sphere...Detecting whether common name or address strings, or column names, can be found in a dataset, as an anonymization re-confirmation...Validating data against gov.uk registers - for example, given a column containing council names, that all are on the relevant register of local government names...Scoring validation against relevant W3C and ISO standards.

The long term impact of using Lintol.

Our approach is intended to minimise errors, and improve quality and in turn we hope this will encourage data providers to publish and make a higher volume of their data publicly usable.

Saving time
For data providers, time is both scarce and precious. Inefficient manual checking can now be removed from the data publishing process. Using Lintol is repeatable. Once you've successfully used it to validate and publish, you're all set to go again. Lintol incorporates many validation tools within it, but you'll only ever need to familiarise yourself with Lintol itself. Lintol will help users identify patterns, and recognise datasets that present consistent challenges.

Saving money
By automating steps in the validation process, the need for data providers to receive in-depth training and repeatedly execute these steps is removed. Paying for extra tools to help streamline the publishing process won't be necessary.

The Project Lintol Team
We're a team of Collaboration Partners, largely drawn from the Open Business Industry Group, a tightly linked collaboration network being built in cooperation with InvestNI (the public Northern Ireland business development organisation). to demonstrate opportunities for collaborative business through open source and open data. Most of our team were involved in the Open Data project OurRagingPlanet, funded by OpenDataNI, as a competition challenge winner. OurRagingPlanet developed a stack of open source tools to perform job-based Python analysis (numerical simulations) using a range of geospatial open data sources, such as elevation maps, school locations, street lighting and places of interest. Several of the team are driving and supporting the Xpand Access project - a commercially-driven platform to bring together business data globally, particularly open data sources, to provide market intelligence.

About the Open Data Institute

The Open Data Institute (ODI) is independent, nonprofit and nonpartisan, founded in 2012 by Sir Tim Berners-Lee and Sir Nigel Shadbolt. From its headquarters in London and via its global network of startups, members and nodes, the ODI offers training, research and strategic advice for organisations looking to explore the possibilities of data. You can learn more about the ODI at theodi.org